Raga Identification Using MFCC and Chroma Features

Kavita.M.Deshmukh
PG Student, Department of E&TC Engineering,
SES's RCPIT, Shirpur, India

Prof. Dr. Pramod. J. Deore
HOD, Department of E&TC Engineering,
SES's RCPIT, Shirpur, India

Abstract: Ragas are the heart of Indian classical music. Raga is one of the basic concepts in Indian music. Raga plays an important role in Indian classical music. It is a collection of swaras comprising of many features and is explained as melodic concept which is led to blossom by the musical artist. We have performed this approach on 3 ragas- Darbari, Khamaj, and Malhar. In this work, we propose a methodology to identify the ragas of an Indian music signal. This has several interesting applications in digital music indexing, recommendation and retrieval. In this work, we attempt the raga classification problem in a non-linear SVM (support vector machine) framework using a combination of two relevant features that represent the similarities of a music signal using two different features MFCC (Mel Frequency Cepstral Coefficient) and Chromagram. We assesses the proposed method on our own raga dataset achieve an improvement of 96.79% in accuracy by combining the information from two features relevant to Indian music.

Keywords: swaras; raga identification; SVM; MFCC and chromagram.

I. INTRODUCTION

Western music and Indian classical music differed from each other with respect to their timing, notes, and different characteristics associated with raga. Note of western classical music is similar to that of swaras of Indian classical music. A raga is the unique combination of swara, and their substrings. It plays vital role in an Indian classical music. Indian track has seven basic swaras (notes) namely Sa, Ri, Ga, Ma, Pa, Dha, Ni(Shadja, Rishab, Gandhar, Madhyam, Pancham, Dhaivatand, Nishad)..Indian classical music consists of different characteristics associated with particular raga that are not easily identified by using approach for identification of western music. Raga is a set of different specific notes which might be having a few unique properties (e.g. Arohana, avarohana, pakad, Taal, etc.).Notes of a raga organized in an ascending order called arohana of that raga and notes of a raga organized in a descending order called avarohana of that raga. Different specific notes are referred to as swaras in Indian classical music. Raga identity includes techniques that find notes from a music and as classify it into the proper raga. Ragas shape a very crucial idea in Hindustani classical song and seize the mood and emotion of performances [1].

It can also be utilized by novice musicians who find it tough to differentiate ragas which are very similar to each different and additionally useful for the beginners who examine this stunning artwork. For automatic identification, some of the characteristics of ragas have to be converted into appropriate features. This is very difficult for Indian music due to the following reasons which needs to be addressed while converting a music piece into swara strings. (i) A music piece may be composed from multiple instruments during a performance. (ii) Unlike Western music, the notes in Indian music are not on an absolute scale but on a relative scale (iii) there is no fixed starting swara in a raga. (iv)Notes in Indian music do not have a fixed frequency but rather band of frequencies (oscillations) around a note. (v) The sequence of swaras in the ragas are not fixed and various improvisations are allowed [2]. In this work, we attempt the raga classification problem using a non-linear SVM and a combination of two different features that is MFCC and Chromagram. In this approach, we further augment the MFCC features with chroma features to improve results. For the chromagram result, the MIR Toolbox [3], an open source toolbox for musical extraction was used to extract features.

II. RELATED WORK

There are different endeavours made in distinguishing the raga in an Indian music. One technique for raga classification is through the interpretation of raga straight forwardly into swaras at every intervals of time and order raga utilizing a classifier, for example K-NN or SVM.In [4], Vijay Kumar, Harit Pandya, C.V. Jawahar investigated the difficulty of raga recognizable proof in Indian Carnatic track. In mild of the belief that, contemporary strategies are either in light of pitch-class profiles or ngram histogram of notes yet not each, they attempted to fuse them in a multi-class SVM framework through linearly combining the 2 kernels. Each of these kernels capture the similarities of a raga based on Pitch-class profiles and ngram histogram of notes.Chordia and Rae [5] defined the consequences of the first massive-scale raga reputation experiment. Raga are the crucial structure of Indian classical music, every such as a unique set of complex melodic gestures. They have constructed a system to understand ragas based totally on pitch-class distributions (PCDs) and pitch-class dyad distribution (PCDDs) calculated at once from the audio signal. A massive, numerous database inclusive of 20 hours of recorded performances in 31 one-of-a-kind ragas by 19 different performers turned into assembled to educate and take a look at the system. Classification was finished the usage of support vector machines (SVM).

Authors of [6] they investigate the problem of scale independent automated raga identification by means of accomplishing kingdom of the artwork results the usage of Gaussian mixture model (GMM) based Hidden Markov Models (HMM) and he combines three features i.e chromagram styles, mel-cepstrum coefficients and timbre features. We additionally carry out the above work using 1) discrete HMMs and 2) classification trees over swara based totally functions comprised of chromagrams using the idea of vadi of a raga. They perform their approach based on four

ragas- darbari, khamaj, malhar and sohini. They have carried out a median accuracy of 97%. [7] Sridhar and Geetha propose a strategy to recognize the raga of Carnatic music signal. The principle thought manner behind Raga identification is that it can be applied as a decent premise for music statistics healing of Carnatic track melodies or Film songs based totally on Carnatic track. The input polyphonic music signal was investigated and made to go through a signal separation set of rules to split the tool and the vocal signal. Utilizing their proposed artist identity set of rules they determined the singer with the help of fundamental frequency of artist.

The frequency component of the signal had been then decided and these frequency components, mapped into the swara sequence thereby therefore finding out the Raga of the particular melody. The raga that is coordinated with the raga placed away in the database is the following raga and output of the device. Their test information contains of 30 samples/tunes in 3 melakarta ragas sung via four artists, 175 Talam, raga database having raga name, arohana avarohana in swara phase structure. Pandey, Mishra, Paul [8] proposed the system 'Tansen' which relies upon on Hidden Markov Model and string, Pakad coordinating. Their test information comprises of result on just 2 ragas Yaman kalian and Bhupali. They applied HMM model seeing that notes are little in number and the grouping for raga is quite very much characterized. Baum-Welch getting to know algorithm is applied for identity of transition and preliminary state chance in HMM set of rules. Again to improve execution over HMM, Pakad coordinating method is utilized by becoming a member of gaining knowledge of into the framework. Prashanth T R, Radhika Venugopala [9] proposed a way for Note Identification in Carnatic Music from Frequency Spectrum. Instead of the usage of note transcription we can also perceive only notes in the input music. These system accept '.Wav' documents as input, frequency spectrum characteristics are analyzed and depends on that they mapped notes. They take a look at records consist of 15 raga alap with 3-8 min clip of different artists. They have achieved up to 90% of accuracy.

III. CHARACTERISTICS OF A RAGA

Raga is a collection of swaras and consists of sequential arrangement of swaras or notes. Different notes are called swaras in Indian classical music. The fundamental seven notes or swaras or symbols in classical music are S(Sa),R (Re or Ri), G (Ga), M(Ma), P (Pa), D (Dha), N(Ni) .Raga is a blend of various swaras that are having some exceptional properties (e.g. arohana, avarohana, Gamakas,Pakad, Taal etc.)[10].

A. Arohana and Avarohana: Raga consists of group of swaras or notes. Depending on sequence of notes or swara, characteristics of raga i.e. arohana and avarohana, it offers identity. Arohana is a collection of notes that are arranged in ascending order. Avarohana is a collection of notes that are arranged in descending order.

B. Gamakas: Each note in the swara sequence has a specific frequency value. Notes in a raga are arranged in a way that there is continuous oscillatory movement about the note, such arrangement of notes is referred to as gamakas.

D. Pakad: A Pakad is a characteristic phrase or set of swara which uniquely identify a raga. For each raga there is a unique and different Pakad from other raga. Pakad is a small sequence of swaras in a raga that acts as a signature for the raga and an artist often visits and revisits the pakad over a performance. It

is a major clue for human raga recognition. For some ragas, the pakad might be sincerely the avrohan wrapped over arohan and for some others; it is probably a totally different pattern of the constituent swaras.

E. Vadi: In each Hindustani classical song and Carnatic song, is the tonic (root) swara (musical be aware) of a given raga (musical scale). "Vadi is the most sonant or most essential note of a Raga. It does not check with the maximum performed note however it as an alternative refers to a notice of unique significance. It is normally the swara that's repeated the best number of times, and frequently it is the swara on which the singer can pause for a significant time. Vadi swara in a raga is most important swara. Specialty of any raga relies upon on vadi swara and because of this, the vadi swara is likewise called the Jeeva swara or the Ansha swara. A expert artist makes use of vadi swara in special ways like making a song vadi swara again and again, beginning a raga with vadi swara, to give up a raga with vadi swara, making a song vadi swara often in vital places with one of a kind swaras or someday making a song vadi swara for an extended time in one breath.

F. Tala: Tala refers to a fixed time cycle, set for a particular composition, that's constructed from groupings of beats.Talas have cycles of a described number of beats and infrequently trade inside a track. They have particular components, which in mixtures can provide upward thrust to the range to exist, allowing special compositions to have extraordinary rhythms. Carnatic track singers commonly hold the beat by means of moving their palms up and down in targeted patterns, and the use of their hands concurrently to maintain time.

IV. EXTRACTION OF FATURES

A. Mel Frequency Cepstral Coefficients (MFCCs):

The most usually used speech characteristic is the Mel Frequency Cepstral Coefficients (MFCC) features, MFCC is most commonly used feature, and because of its accurate estimate of the speech parameters and most effective results for speech [9]. Mel Frequency Cepstral Coefficients (MFCCs) are the most widely used features in the majority of the speaker and speech recognition applications. The typical process for feature extraction can be seen on Figure, with the assumption that it has been processed digitally and properly quantized. Extraction refers to procedure of transforming the speech signal into a number of parameters, while pattern matching is a task of obtaining parameter sets from memory closely matches the parameter set extracted from the input speech signal. In simple words, the essence a speech recognizer is to provide a powerful and accurate mechanism to transcribe speech into text [4]. Feature extraction is a crucial step of the raga identification process. The MFCC is the best method for feature extraction introduced in [4].
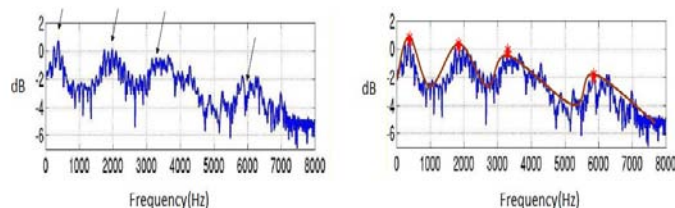

Fig1. Sample speech spectrum

- Peaks denote dominant frequency additives within the speech signal in fig (1).
- Peaks are referred to as formants.
- Formants convey the identity of the sound.

- Formants and a smooth curve connecting them.
- This Smooth curve is known as spectral envelope.
- Our aim: We want to split spectral envelope and spectral details from the spectrum shows in fig (2).

  Given log X[k], acquire log H[k] and log E[k],

  such that log X[k] = log H[k] + log E[k]

  Take Log on both sides

  log ||X[k]|| = ||log H[k] +|| log E[k]||

  Taking inverse FFT on both sides

  x[k] = h[k] + e[k]

- Mel-Frequency analysis of speech is based on human belief experiments.
- It is determined that human ear acts as clear out. It concentrates on only certain frequency components.
- These filters are non-uniformly spaced on the frequency axis extra filters in the low frequency regions. Less no. Of filters in high frequency regions Cepstral coefficients h[k] obtained for Mel spectrum are referred to as Mel-Frequency Cepstral Coefficients regularly denoted by MFCC.
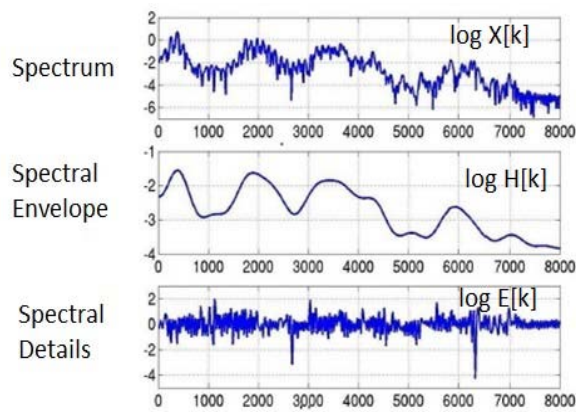

Fig 2.Spectral Envelope

B. Chromagram: In the music context, chroma feature or chromagram closely relates to the twelve different pitch classes. Chroma-based features, these are also referred to pitch class profiles, are a powerful tool for analyzing tune whose pitches can be meaningfully categorized (often into twelve categories) and whose tuning approximates to the equal-tempered scale. One predominant property of chroma features is that they capture harmonic and melodic characteristics of track [11], while being sturdy to modifications in timbre and instrumentation. A chromagram [6] is a visual illustration of energies within the 12 semitones (or chromas) of the western musical octave particularly C, C#, D, D#, E, F, F#, G, G#, A, A# and B. So, it basically depicts the distribution of energy in the twelve pitch classes. The western semitones are such that they are fixed with respect to absolute frequency values and the musical octave has the property that the semitone one octave below or above is equivalent to the current semitone. Figure (3),(4),(5) shows the chromagram features of raga Darbari, Khamaj and Malhar.
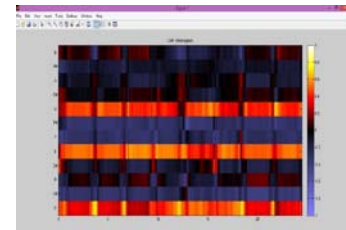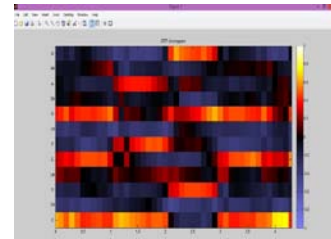

Fig.3Chromagram for Raga Darbari

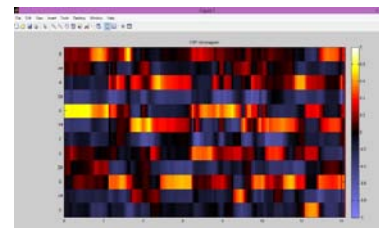**Fig.4 Chromagram for Raga Khamaj**

Figure.5 Chromagram for Raga Malhar

If one notice has a frequency of 440 Hz, the be aware an octave above it is at 880 Hz, and the note an octave below is at 220 Hz. Since the semitones get repeated in every octave above and below, the energies within the chromagram for each semitone (chroma) is computed by means of wrapping and adding it up over different octaves. Above figure shows depicts a chromagram generated from an arohan of raga Darbari, Khamaj and Malhar. The arohan for raga bhairavi is: SA Re-Kom Ga-Kom Ma Pa Dh-Kom Ni-Kom SA. The Sa swara of Darbari coincides with the semitone G and the rest of the swaras in the arohan get aligned with the semitone pattern in the chromagram. These observations and previous use of chromagrams in audio analysis and chord recognition [12] motivated us to use the chromagram to extract information about swaras. From the chromagram, we extract the semitone with maximum energy in each frame and get a sequence of semitones for the raga. Though these sequences of semitones might have some identifying information about the raga, using them for raga identification is not appropriate since ragas are defined over MFCC [13]. In our approach, we assume that we do not have information about the tonic frequency of the raga performance. We must therefore find the mapping from the absolute frequency scale employed by the chromagram to the relative scale of the musical piece, so that the swara sequence can be identified. To do so, we use the concept of vadi discussed earlier to convert the semitone sequence to a swara sequence. We compute the most frequently occur ring semitone from the semitone sequence and associate it with the vadi of a raga which is known for each raga. For example vadi for raga Khamaj is swara "Ga". In an audio of Khamaj, if the semitone C# is most prominent, we label swara "Ga" at semitone C# and then convert the rest of the semitone sequence into a swara sequence. The above procedure is raga

specific, i.e. the conversion from semitone sequence to swara sequence utilizes the identity of the raga-specific vadi swara. Assume that we are building a system for n ragas and the actual raga for the test audio is not known, then for the given audio, we must compute separate swara transcriptions for each of the n ragas [14].

V. CLASSIFICATION

We identify a raga by combining the information from two different and relevant features, MFCC and Chromagram. We incorporate this systematically into an SVM framework. In machine learning, the (Gaussian) radial basis function kernel, or RBF(Radial Basis Function) kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification [4]. In machine learning, support vector machines (SVMs, also support vector networks [4]) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In addition to acting linear category, SVMs can correctly perform a non-linear classification which is called the kernel trick, implicitly mapping their inputs into excessive-dimensional feature spaces. Kernel Trick: The main idea behind the kernel trick" is to map the data into a different space, called feature space, and to construct a linear classier in this space. It can also be seen as a way to construct non-linear classifiers in the original space the classier function f(x) can be expressed as a sum of inner products with support vectors. An important result, called Mercer's theorem, states that any symmetric positive semi-denite function K(x;z) is an inner product in some space (and vice-versa). In other words, any such function K(x;z) implicitly denes a mapping into so called feature space

$\emptyset$: x$\rightarrow$ $\emptyset(x)$ such that K(x;z) =< $\emptyset$(x), $\emptyset$(z) >.

Such functions K are called kernels. The unique most-margin hyperplane algorithm proposed by Vapnik in 1963 and constructed a linear classifier. However, in 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a manner to create nonlinear classifiers through applying the kernel trick (originally proposed by Aizerman et al.[15]) to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; although the classifier is a hyperplane in the transformed feature space, it may be nonlinear in the original input space.

VI. RESULT

We evaluate the performance of our proposed approach on our dataset. Our data set is small consisting of only 3 ragas with limited instruments .To evaluate our method, we created a dataset comprising of 4 ragas namely Darbari, Khamaj and Malhar. All audio files are of type instrumental of type flute, harmonium and sitar recordings. We initially conducted experiment on our dataset. We implemented the feature extraction procedure for ragas as described in [4]. Polyphonic audio signals are converted to predominant melody using melody extraction software. Pitch-class profiles and n-grams

are extracted as explained in the [4]. We randomly create the dataset and evaluate raga so that create data base used for training and evaluate the selected raga. We created approximately 90 tonnes in data base and test the ragas report the accuracy. We compare our approach with the approach proposed by Mr.Vijay Kumar. Results are shown in Table I and the best results for both methods are shown. It is clear that, our approach which combines MFCC and Chromagram feature extraction methods achieves superior performance compared to [4] where only Pitch-class profiles and n-gram histogram of notes is used. Results in the table clearly demonstrates the superiority of our approach. The best accuracy obtained by our approach is 96.79% which is higher than their best reported accuracy 91.20%.

Number of Ragas in Data Base	Ragas Checked	Correctly Identified by Existing Method	Correctly Identified by Proposed Method	Accuracy % (Existing Method)	Accuracy % (Proposed Method)	Delay (Existing Method)	Delay (Proposed Method)
5	4	4	4	100	100	12	14
5	5	5	5	100	100	12	14
10	6	6	6	100	100	18	20
10	8	7	8	87.5	100	18	20
10	10	9	10	90	100	18	20
15	8	7	8	87.5	100	21	24
15	12	11	11	91.66	91.66	21	24
15	14	12	13	85.71	92.85	21	24
15	15	13	14	86.66	93.33	21	24
20	8	7	8	87.5	100	30	35
20	12	11	11	91.66	91.66	30	35
20	16	14	15	87.5	93.75	30	35
20	20	18	19	90	95	30	35

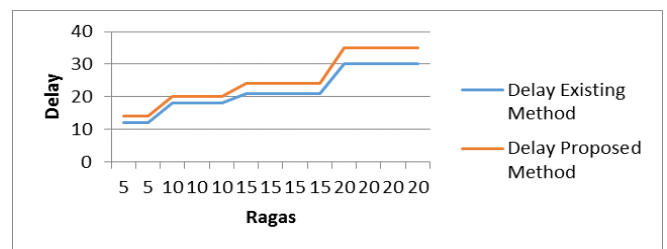**Table I: Comparison of performance of approach [2] with our approach on our dataset.**

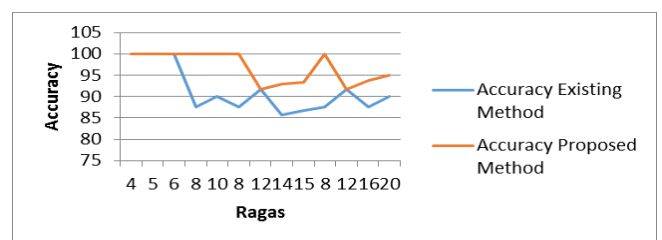**Fig 6.Final Delay for Test Ragas**

Fig 7. Accuracy for Test Ragas

VII. CONCLUSION

The feature analysis component of an Raga Identification system plays a crucial role in the overall performance of the system.There are many feature extraction techniques available, but ultimately we want to maximize the performance of these

systems The objective of this method investigates the results that can be obtained when you combine Mel-Frequency Cepstral Coefficients(MFCC) and Chromagram features as feature components for the front-end processing of an Raga Identification .The MFCC and Chroma feature components combined are suggested to improve the reliability of a Raga Identification system. The MFCC are typically the "de facto" standard for Raga Identification systems because of their high accuracy and low complexity; however they are not very robust at the presence of additive noise. The Chroma features in recent studies have shown very good robustness against noise and acoustic change. The main idea is to integrate MFCC & Chroma features to improve the overall Raga Identification performance in low signal to noise ratio (SNR) conditions. We achieved average accuracies of 98.30%.These are the best current results for scale independent raga identification and compare closely with the results by [4].Overall, the combined feature set consisting of chroma and MFCC consisting the best results.

VIII. ACKNOWLEDGEMENT

IX. REFERENCES

[1]   Ranjani, H.G, Arthi,S. and T.V. Sreenivas, "Carnatic Music Analysis: Shadja, Swara identification and raga verification in AlApana using stochastic models", IEEE Workshop on Application of Signal Processing to Audio and Acoustics October 16-19, 2011,New Paltz,NY.

[2]   Bruno Nettl, Melinda Russell, In the Course of Performance: Studies in the World of Musical Improvisation, Chapter 10, Page 219.

[3]   Bruno Nettl, Melinda Russell, In the Course of Performance: Studies in the World of Musical Improvisation, chapter 10.

[4]   Vijay Kumar, Harit Pandya, C.V. Jawahar, "Identifying Ragas in Indian Music" International Institute of Information Technology, Hyderabad, India , IEEE International Conference 2014.

[5]   Parag Chordia and Alex Rae, "Raag Recognition using Pitch-Class and Pitch-Class Dyad Distributions", International Symposium/Conference on Music Information Retrieval, 2007.

[6]   Pranay Dighe, Parul Agrawal, Harish Karnick, Siddartha Thota and Bhiksha Raj, Scale independent raga identification using chromagram patterns and swara based features, IEEE International Conference on Multimedia and Expo Workshops, 2013.

[7]   Rajeswari Sridhar and T.V.Geetha, "Raga Identification of Carnatic music for Music Information Retrieval", International Journal of Recent Trends in Engineering, Vol.1, May 2009.

[8]   Gaurav Pandey, Chaitanya Mishra, and Paul Ipe, "Tansen: A System For Automatic Raga Identification", In proc. 1st Indian Intl. Conf. on Artificial Intelligence, pages 1-9,2003.

[9]   Prashanth T R, Radhika Venugopala, "Note Identification In Carnatic Music From Frequency Spectrum", In proc. of IEEE, page 87, 2011.

[10] Pranay Dighe, Harish Karnick, Bhiksha Raj Indian Institute of Technology, Kanpur, India. 2. Carnegie Mellon University, Pittsburgh PA, USA.

[11] Rajshri Pendekar, S. P. Mahajan, Rasika Mujumdar, Pranjali Ganoo "Harmonium Raga Recognition" International Journal of Machine Learning and Computing, Vol. 3, No. 4, August 2013.

[12] Springer, 2011. [D. A. Reynolds, "An overview of automatic speaker recognition].

[13] H. Sahasrabuddhe and R. Upadhye, "On the computational model of raag music of India." in Workshop on AI and Music: European Conference on AI, 1992.

[14] J airazbhoy, N.A.-The Rags of North Indian Music: Their Structure & Evolution; Popular Prakashan Bombay 1995.

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