IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Survey Paper on Raga Identification and Classification

SAKSHI SONJE
Dept. of Electronics and Telecommunication
Pune Institute of Computer Technology
Pune, India

SWANAND KULKARNI
Dept. of Electronics and Telecommunication
Pune Institute of Computer Technology
Pune, India

DNYANADA BHOSALE
Dept. of Electronics and Telecommunication
Pune Institute of Computer Technology
Pune, India

Abstract — **This survey paper presents a survey of current research on identifying techniques of Raga. Music is an important aspect in the human life. The Raga is regarded as the fundamental building block of Indian classical music. The movement of the Swara (lyrical notes) reveals Raga's behavior. Hindustani (North Indian) classical music and Carnatic (South Indian) classical music are the two divisions of Indian classical music, which is based on the Raga. Our experiment is concentrated on Hindustani classical music. Traditional techniques for analyzing Ragas in the time and frequency domain are less effective, less accurate and some algorithms take a lot of time. Artificial neural networks can be used to improve the classification of Ragas overall. In this work we introduce audio signals in the form of Spectrogram and use of CNN models for processing data of audio signals.**

Keywords – **Raga, swara, Spectrogram, CNN (Convolutional Neural Network), ICM (Indian Classical Music), Chroma features, pakad.**

1. INTRODUCTION

Two primary branches of ICM are Hindustani classical music[1] ad Carnatic classical music. A vital aspect of Indian Classical Music (ICM) is the raag, which serves as a base on which compositions of songs and Improvisations are created and presented. Raag is a basic element for compositions and improvisations (or "Manodharma"[2]) and is defined as a pattern of notes having characteristic embellishments, rhythm and intervals. The primary seven notes are S (Sa), R (Re or Ri), G (Ga), M (Ma), P (Pa), D (Dha), N (Ni). Aaroh and Avroh. Each Raga or Raag is composed of a "Aaroh" that implies swaras scale up and a "Avroh" that implies swaras scale down.

A piece of music is considered a certain Raga as long as the attributes associated with it are fulfilled. This concept of Indian classical music, in this way, is very open. In this freedom lies the beauty of Indian classical music and also, the root of our problems of identifying correct raga. The problem we addressed was "Given an audio sample (with some constraints), predict the underlying Raga". More succinctly, By giving some audio sample find the underlying Raga for the input with complexity of Raga is highly variable in performance.

Another important feature of a raga is its anga, which depicts whether the first half of the saptak (or octave) is more important in its rendition, in which case it is Poorvanga, or the second half, in which it is Uttaranga [3].

Automatic raga identification can provide a basis77 for searching for similar songs and generating automated play-lists that are suited for a certain aesthetic theme. It can also be used by novice musicians who find it difficult to distinguish ragas which are very similar to each other. It might also evolve into a system which checks how accurately a person is performing a certain raga. The distinguishing characteristics of ragas are typically the scale (set of notes/swaras) that is used, the order and hierarchy of its swaras, their manner or intonation and ornamentation, their relative strength, duration and frequency of occurrence. The present work addresses the problem of raga identification from an audio recording of a Hindustani classical vocal performance. We extract information about how the swaras of the performance are intoned to achieve this.

Seven-tone scales are the foundation of ragas. Thats are the name given to these scales. Table 1 lists the connection between the seven notes and the Western solfege syllables of the No major diatonic scale. The remaining five notes can only be changed in one direction while keeping the tonic (Sa) and dominant (Pa) unchanged. Re, Ga, Ha, and Ni may be moved back one-half step, whereas Ma may be moved forward one-half step[8].

Table 1 : Scale notes/ swaras

Note	**Indian note**	**Ratio to Sa (C)**
C (Tonic)	Sa	1
D (Super Tonic)	Re (Shudh)	9/8
E (Mediant)	Ga (Shudh)	5/4
F(Sub Dominant)	Ma (Shudh)	4/3
G (Dominant)	Pa	3/2
A (Sub Mediant)	Dha (Shudh)	5/3
B (Leading note)	Ni (Shudh)	15/8

2. LITERATURE REVIEW

2.1 Raag Detection techniques

A vital aspect of Indian Classical music (ICM) is the Raag, which serves as a base on which improvisations and compositions (IAC) are created and presented [2]. Although many works have explored the problem of classification of Raag, they have several short comings owing to the fact that they assume a prior knowledge of the tonic of the audio.

Saumya Srivastava, Vandan Bhatia, Rhythm Bhatia and Manpreet Singh[26] has discussed the importance of feature extraction in classification of music. It was found that timbre feature provides meaningful information about audio signal. To represent the speech features, MFCC feature has been used and Group delay is used for phase information. In future, both these features can be combined and used together to train CNN models. They analyzed various Mel based features, Chroma prints, audio fingerprint features for music information retrieval process.

Harmonic product spectrum algorithm

Laura lee Bakliwal and William [4] conduct a study on perception of Raags and observe that an inexperienced listener is able to identify the emotion of a Raag effectively by relying only on psychological cues. Hence Raag can be used in a variety of music related tasks like organizing recordings into songs with similar emotional content, music recommendation systems [5] etc. Lee [11] extracted pitch and set of notes using harmonic algorithm and the raag is predicted. With pattern matching algorithms and binary-type chord templates, Lee [11] have introduced a new feature vector for automatic chord recognition from raw audio that performs better than the traditional chroma vector. To reduce the intensities at pitch classes inhabited by overtones of the chord tones, the new feature vector, the Enhanced Pitch Class Profile, or EPCP vector, was generated from the harmonic product spectrum of an input signal rather than the DFT.

Finite automata based detection using note sequences

There isn't a lot of literature accessible because, MIR in Indian Music is a very new topic. Sahasrabuddhe and Upadhye [7] made one of the first attempts when they employed finite state automata to represent a raga based on the pattern of its swara constituents. In [7] proposed a solution of word recognition problem where Raga compositions were treated as alphabetic formed words having notes using in ICM. This solution was enhanced by considering the Pakad of the Raga. To handle the problem of permitted variability found in the duration of particular note, Sahastrabuddhe and Upadhye designed two distinct heuristic methods for transcribing notes from any given audio sample.

Pitch-class Profiles and Note Bi-grams

Ragas were categorized using pitch class profiles and pitch bigrams by Tzanetakis et al. in [12]. His system uses a dataset of 72 minutes of monophonic instrumental data of Sarod performed in 17 ragas by a single musician. The pitch is extracted using the HPS method. The abrupt variations in phase and amplitude in the signal spectrum are used to identify note onsets. The automatic classification of audio signals into a hierarchy of musical genres is investigated in this work. More specifically, three feature sets are suggested for describing pitch content, rhythmic content, and timbral texture.

Hidden Markov Model

This method[13] is also based on the transition in note sequences and the arohona and avrohna patterns are used for raga detection. Hidden Markov models (HMMs) are mathematical representations of stochastic processes, or operations that produce random sequences of results with a given probability. A series of coin tosses is a straightforward illustration of such a procedure. More specifically, an HMM is a finite set of states, each of which is connected to a probability distribution (often multidimensional). A set of probabilities known as transition probabilities controls the transitions between the stages. According to the corresponding probability distribution, an outcome or observation can be produced in a specific state. An outside observer can only see the results, not the actual state. Thus, the term "hidden Markov model" refers to the "hidden" nature of the states. The sequences of notes for different Ragas are well defined and a model based on discrete states with transitions between them is the ideal representation for these note sequences. The notes are small in number, hence making the setup of an Hidden Markov Model easier than other methods [14].

Sheh and Ellis [19] proposed a statistical learning approach for segmentation of chord and recognition of chord, where they used HMM trained by Expectation Maximization algorithm, where they treated chord labels as hidden values in the EM framework.

Problems in MIR techniques for raag identification

Sreeparna Banerjee [3] addressed some major concerns to built a suitable identification techniques of the raga for Music Information Retrieval (MIR). Hindustani Classical music ragas use 10 micro tones in addition to the seven notes listed in the table 1, along with five semitones that correlate to the altered notes of Re, Ga, Ma, Dha, and Ni, creating a total of 22 Shrutis. An further factor in HCM is the employment of a drone instrument like the tanpura, which offers the tonic or absolute frequency in addition to the fifth (Pa) or fourth (Ma), which provides a continuous pitch reference.

Key terms used for raga identification include the melodic themes known as bandish[9] in vocal music and asthayi in instrumental music. The homo-phonic character of HCM. A key component of music identification is matching the timbre [10] of the vocal or instrumental performance with the supporting elements.

2.2 Classification techniques based on supervised and unsupervised learning techniques.

2.2.1 Supervised Learning

Naïve Bayes:

It is technique for classification which is based on probability [15]. It considers all the elements free from each other with the computational likelihood of every component freely for a specific class name [16-18].

It is expressed as:

The probability of feature x in the feature set is shown as the class label 'y'. Then for all the features total probability will be:

$$P(x/y) = \pi_{k=1}^{d} p(x_k/y)$$

K Nearest Neighbor (KNN) :

Hemali Dodia, Shraddha Pandey and Chawande[21] identify raga using a classifier, and then connect this raga with its associated rasa to identify emotions in music. A larger dataset is used for the K-NN classifier to produce better results. Soft computing approaches such as fuzzy logic can only provide us with approximations. The SVM classifier is tough to handle on a large scale and with several instruments. K-NN may interfere with gammakas and pitch extraction.

Support Vector Machine :

In SVM[22] hyperplane that clearly divides the sampling points with various labels is identified. It separates sample points of labels as well as class on two different sides of hyperplane.

Random Forest:

This order computation employs group approaches to provide better predictive execution. It generates yield in the form of individual trees and is based on the choice tree computation. It is regarded to be a highly accurate classifier that can deal with a wide range of factors.

Convolutional Neural Network:

O. Abel, Hui Jiang and Dong Yu [27] discussed about image processing application to organize speech audio feature vectors into feature maps that suits for the CNN processing. The input image which can be in the form of spectrogram or chromagram bundle features into feature maps. CNN can be used as pattern recognition as the organized input data of feature maps is fed up to the pooling layers. The input is in 2D array, being the pixel values at the x and y axes of frequency and time. For RGB representation values can be seen as three different 2-D feature maps. CNNs run a small window over the input image at both training and testing time whereas, from various features of the input data, the weights of the network that looks through this window can learn regardless of their absolute position within the input. After the formation of the input feature maps, the convolution and pooling layers use their respective operations to generate the activations of the units in those layers in sequence. A single window of input to the CNN consists of a wide amount of context (9–15 frames). In this paper, they use log energy computed directly from the mel-frequency spectral coefficients (MFSC) features to represent each speech frame along with deltas and delta-deltas just to describe the acoustic energy distributions in each of several different frequency bands.

2.2.2 Unsupervised Learning

Unsupervised learning is the machine learning problem of inferring a function from unlabeled data to characterize hidden structure. Because the illustrations provided to the learner are unlabeled, there is no room for error or compensation when evaluating a potential arrangement. This acknowledges uncontrolled learning from regulated learning as well as assisted learning. If the objective yield is unknown, implies. Clustering methods such as k-means and c-means are examples of unsupervised learning approaches [23]. If the output of a neural network is known, it falls into the supervised category; otherwise, it falls into the unsupervised category. Supervised learning is a technique in which class labels are utilized to train the classifier. It is appropriate for huge classification issues such as Raag detection. Because the structure is always defined in terms of raag formats, there is a minor chance of using unsupervised learning.

3. Proposed Methodology

In this Project, Audio features are used to provide description of sound, different features capture different aspects of sound and to build intelligent audio systems. Raag Detection is performed on the musical file from which features are extracted. Audio data is nothing but amplitude and frequency where amplitude denotes volume and frequency denotes pitch.

Methodology

The method is based on extracting chroma information and feeding those features to a deep neural network. The distribution of 12 musical notes in an audio recording, all mapped to a single octave, makes up the chroma features.

Dataset

The dataset comprises of audio recordings with the "Aroha" and "Avroha" of the particular "Raga" as well as some well-known songs that were composed in particular "Ragas." Each audio clip is transposed on each of 12 different scales using a digital audio workstation for data augmentation purposes.

Then, chroma features are extracted from these audio clips through processing. These chroma features are transformed into a data frame that has musical notes as its features.

Model

An artificial neural network with several layers is fed the newly produced data frame. The network is trained and utilized to make predictions.

4. System Architecture

Fig 1 shows the complete architecture of raga identification and classification system. Chroma feature extractions, data augmentation and training are three key steps to achieve this system. The initial step is to give the input in the form of audio which can be vocal or instrumental to identify its type of raga. Based on the give input audio, chroma features will be extracted with the help of chromagram approach in deep learning and then it will be passed to the Classifier and then raga will be recognized. Training dataset will go through the data augmentation which is based on scale. Chroma features will be extracted on each data so that feature training can be performed as shown in Fig 2.

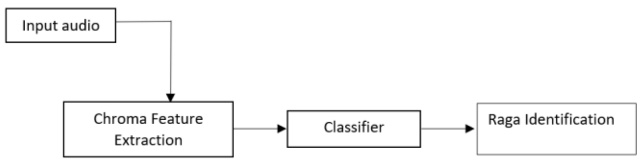

Fig 1. Architecture diagram

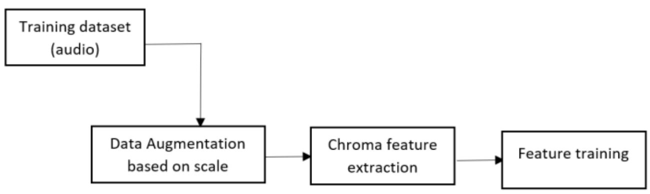

Fig 2. Classifier

5. Discussion

In our experiment, we have chosen three ragas having 12 songs each. We also have used vocal instrument datasets of ragas like Bhup, Yaman, Bhairavi.

According to the findings of this survey, for correct classification of Raag in music pitch, scale, and dataset must be clean and clear. It has also been discovered that supervised learning algorithms are effective in detecting Raag in music.

Sathwik Madhusadan, G. Chowdhary have presented a new data augmentation technique that enables a model to learn the semantics of different Raags in numerous different tonics apart from the tonic present in the audio. The model, which is having 5 layer deep Convolutional Neural Network (CNN), achieves an accuracy of about 72-77% on testing datasets. Pandey [13] presented the system Tansen for automatic Raga identification, which is based on hidden Markov models and string matching. A very important part of Tansen is its note transcription, which is proposed heuristic strategies based on the pitch of sound.

According to Koduri et al. [24], the tonic identification itself contains some degree of error, which could have a negative impact on the execution of a raga acknowledgement system. Because neither the Carnatic nor the Hindustani frameworks adhere to any absolute tonic recurrence, it is advantageous to create a framework that can disregard the total area of the tonic. According to Katte [25], the major challenges are a limited database containing a limited number of raga, incorrect pitch extraction, manual tonic detection, assumptions made for different parameters of the algorithm, and different constraints on inputs, such as singers, number of swaras, time length, and monophonic type. As a result, increasing these characteristics may aid in better detection.

6. Conclusion

A well defined sequence of lyrical notes of musical swara with a particular pitch and mood is termed Raga. Raga detection is very challenging task in terms of machine learning. In this paper as survey is done on various raga detection techniques which are based on supervised learning techniques. Techniques based on low-level features known as timbre, scale, pitch and notes is studied. Several classification techniques like CNN, naïve bayes, SVM and many others are studied. In this survey paper most frequently used supervised learning approaches like CNN are studied to do efficient Raag detection with high accuracy.

7. Acknowledgement

We would like to thank our guide Mrs. Anagha A. Bidkar for helping us throughout the process of selection of the topic and providing us the opportunity to work on this project "Raga Identification and Classification using deep learning". We are really thankful to our guide and reviewers for their support, advice and helpful critique, all of which have greatly aided the development of our project related ideas.

References

[1] Hindustani sangeet paddhati: Kramik pustak maalika vol. i-vi. Sangeet Karyalaya, 72, 1990.

[2] Laudan Nooshin and Richard Widdess. Improvisation in iranian and indian music. Journal of the Indian Musicological Society, 36:104–119, 2006.

[3] Banerjee, Sreeparna. (2017). A Survey of Prospects and Problems in Hindustani Classical Raga Identification Using Machine Learning Techniques.

[4] Laura-Lee Balkwill and William Forde Thompson. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music perception: an interdisciplinary journal, 17(1):43–64, 1999

[5] Yading Song, Simon Dixon, and Marcus Pearce. A survey of music recommendation systems and future perspectives. In 9th International Symposium on Computer Music Modeling and Retrieval, volume 4, 2012.

[6] Basili, R., Serafini, A., Stellato, A.: Classification of musical genre: a machine learning approach. In: ISMIR (2004)

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

[8] Mary A. Castellano, J. J. Bharucha, Carol L. Krumhansl, Tonal Hierarchies in the Music of North India, Journal of Experimental Psychology: General, 113 (3), pp. 394–412 (1984).

[9] Joe Cheri Ross*, Vinutha T. P. and Preeti Rao, Detecting melodic motifs from audio for Hindustani classical music, 13th International Society for Music Information retrieval conference (ISMIR 2012) 2012.

[10] Saurabh H. Deshmukh and S. G. Bhirud, North Indian Classical Musics Singer Identification by Timbre Recognition using MIR Toolbox, Intl. J. Computer Applications, 91(4) 1–4 (2014).

[11] Lee K. Automatic chord recognition from audio using enhanced pitch class profile. In ICMC 2006.

[12] Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing. 2002; 10(5):293-302

[13] Pandey G, Mishra C, Ipe P. TANSEN: a system for automatic raga identification. In IICAI 2003 (pp. 1350-63).

[14] H. V. Sahasrabuddhe: "Searching for a Common Language of Ragas": Proc. Indian Music and Computers: Can 'Mindware' and Software Meet?, August 1994.

[15] Yadav RM, Bhagel RK. Web based malware detection using important supervised learning techniques on online web traffic. International Journal of Computer Applications. 2015; 130(17).

[16] Chordia P. Automatic raag classification of pitch tracked performances using pitch-class and pitch-class dyad distributions. In ICMC 2006.

[17] Gupta M, Shrivastava V. Review of various techniques in clustering. International Journal of Advanced Computer Research. 2013; 3(2):134-7.

[18] Vaska JS, Sowjanya AM. Clustering diabetics data Using M-CFICA. International Journal of Advanced Computer Research. 2015; 5(20):327-33.

[19] Sheh, A. and D. P. Ellis (2003). Chord segmentation and recognition using em-trained hidden Markov models. In Proceedings of the International Symposium on Music Information Retrieval, Baltimore, MD.

[20] Kalyani C. Waghmare and Balwant A. Sonkamble, "Raga Identification Techniques for Classifying Indian Classical Music: A Survey", International Journal of Signal Processing Systems Vol. 5, No. 4, December 2017

[21] Hemali dodia, Shraddha Pandey and. N. Chawande "Identification of Raga by machine learning with chromagram" International Research Journal of Engineering and Technology (IRJET) Volume 07 May 2020.

[22] Chakraborty S, De D. Pattern classification of Indian classical ragas based on object oriented concepts. International Journal of Advanced Computer engineering & Architecture. 2012; 2:285-94.

[23] Ekta Patel, S. Chauhan "A Survey on musical instrument Raag detection" International Journal of Advanced Technology and Engineering Exploration, Vol 4(29)

[24] Koduri GK, Gulati S, Rao P. A survey of raga recognition techniques and improvements to the state of-the-art. Sound and Music Computing. 2011.

[25] Katte T. Multiple techniques for raga identification in indian classical music. International Journal of Electronics and Computer Engineering. 2013; 4(6):82- 7.

[26] Rhythm Bhatia, Saumya Srivastava, Vandan Bhatia and Manpreet Singh, "Analysis of Audio Features for Music Representation," IEEE, 2018.

[27] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu "Convolutional Neural Networks for Speech recognition" IEEE /ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 22, NO. 10, OCTOBER 2014