A Survey on musical instrument Raag detection

Ekta Patel* and Savita Chauhan
Technocrats Institute of Technology (Excellence), Bhopal, MP, India

©2017 ACCENTS

Abstract
There are several research work is in progress in the direction of raga detection. Ragas are specific to the time of the day and it generates a specific sentiment only when played at the right time. It is the unique sequence in music which comprises of five to nine musical notes in melodic music. It depends on the pitch of musical notes and the mood in which they are conveyed rather than the sequence of notes. Its accurate detection is helpful in generating correct and accurate raga with the different musical instrument. There are several obstacles in accurate raga detection technique. The major challenges are the complex parameters like pitch and mood in the music, skipping extra tones, conversion of different data attributes and raga tempo. In this paper a study and analysis have been presented to stumble the gaps and finding the advantages of the previous approaches. The previous research suggests supervised and unsupervised learning both for raga detection. So this paper included the methods from above two for comparison. This study shows that the supervised learning is capable in improving the detection results.

Keywords
Raag, Thaats, Naïve bayes, Decision tree, Support vector machine (SVM).

1. Introduction
In the ancient times music is the heart of India and the other countries. The root of Indian classical music is very rich. It includes many gharana and the different style and tradition for those gharana. Bhatkhande [1] describes the culture of these gharana and their music forming methods. Indian classical music can be categorized into two main streams like North Indian and South Indian based music and styles. Raag is essential building blocks in Indian classical music. Melodic mode of music comprises of five to nine musical notes is also termed as raga.

In the recent past there are several works have been done on musical analysis and specially the Indian classical music, generating lot of new insight into this domain. The research related to musical information retrieval is thus attracting the interest of so many researchers. The music is categorized in different thaats based on which the ragas are derived. Different Distributions of notes making different note structures are called thaats. The Latest research methods and techniques are focusing on carnatic raga and its analysis. The music research and its analysis play an important role in finding the raga patterns on various ways. To identify their variety the thata categorization is available in [1].

It is a system that is very relevant with this type of categorization.

In 2013, Sharma et al. [2] proposed that thaats are classified in 10 different ways which are as follows: Bilawal, kalian, Todi, Bhairavi, Marwa, Kafi, Bhairav, Khamaj, Purvi, Asavari. These Thaats (raags) possess very different structural patterns so they can be distinguishingly identifiable [1]. Different ragas performed on different relevant timings for the largest impact and increased sentiments on humans mind. A thaat is a musical scale, conceived of as a Western musical scale can have, with the seven notes presented in their order of ascent (aroha). These thhats are dependent upon the aroho of the raga. In Indian Classical Music there are 10 thhats from each of which many ragas are created. The names of these ten thaats are - Kalyan, Bhairav, Kafi, Asavari, Bilabal, Khamaj, Bhairavi, Purbi and Torhi.

The formations of the thaats are described below in Table 1:

Table 1 Formation of thaats

<table>
<thead>
<tr>
<th>S.NO.</th>
<th>Thaats</th>
<th>Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kalyan</td>
<td>Sa Re Ga Ma Pa Dha Ni Sa</td>
</tr>
<tr>
<td>2</td>
<td>Bhairav</td>
<td>Sa re Ga ma Pa dha Ni SA</td>
</tr>
<tr>
<td>3</td>
<td>Kafi</td>
<td>Sa Re ga ma Pa Dha ni SA</td>
</tr>
</tbody>
</table>

*Author for correspondence
The main objective of the paper is to detect raag (rāga) in an audio signal. There exist many important contributions of [8, 9] for automatic detection of raaga. In one such module, 30 tunes in 3 raagas are presented. The information is captured by note bi-grams or HMMs. Based on supervised and unsupervised learning, the system can identify the raaga by the histogram approach and note segmentation and labeling. So property based discussion is presented so that the identification is easy. The following techniques are applied for raag detection:

**Scale matching**

In 2012, Shetty et al. [5] have followed an approach where the scale used in the tune is estimated, and compared with the scales in the database. The raaga corresponding to that scale in database which matches the estimated scale is output by the system. Their test data consisted of 30 tunes in 3 ragas sung by 4 artists. They use harmonic product spectrum algorithm [6-12] to extract the pitch. The tonic is manually fed. The other frequencies in the scale are marked down based on the respective ratio with the tonic. The set of notes which are used are matched against several sets of notes stored in the database for various raagas. Note that this is not the same as pitch-class profile. Here, the comparison is between the scale intervals, and not the pitch-class distribution.

**Statistical Modeling and Pakad Matching:**

In Chordia et al. [7] modeled the raaga as a finite automaton based on the rules set by the properties of each raaga. This idea is used to generate a number of note sequences for a raaga composition, which were technically correct and indistinguishable from human compositions. Inspired by this, in [8], used HMMs to capture the note transitions in their “Tansen” raaga recognition system. The rules to form a melodic sequence for a given raaga are well defined and the number of notes is finite. So, HMM model of a raaga proved to be good at capturing those rules in note transitions engraved by arohana and avarohana patterns of the respective raaga. They have complemented this system with scores obtained from two pakad matching modules. In one such module, pakad is identified with substring matching algorithm. In the other one, it is identified by counting the occurrences of n-grams of frequencies in the pakad. The other important contributions of [8] include two heuristics to improve the transcription of Indian classical music - the hill peak heuristic and the note duration heuristic. Unlike western music, Indian music has a lot of microtonal variations which makes even monophonic note transcription a challenging problem. The two heuristics try to get through these microtonal fluctuations in attaining a better transcription. The hill peak heuristic says that a significant change in the slope or the sign reversal in slope is closely associated with the presence of a unaccompanied instrument or voice, optionally followed by a step of note segmentation and labeling.

### S.NO. | Thaat | Formation
---|---|---
4 | Asavari | Sa Re ga ma Pa dha ni SA
5 | Bilabal | Sa Re Ga ma Pa Dha Ni SA
6 | Khamaj | Sa Re Ga ma Pa Dha ni SA
7 | Bhairavi | Sa re ga ma Pa dha ni SA
8 | Purbi | Sa re Ga MA Pa dha Ni SA
9 | Torhi | Sa re ga MA Pa dha Ni SA
10 | Marwa | Sa re Ga MA Pa Dha Ni SA

In this paper a survey is done on analytical approach used for estimating the raga which was based on supervised and unsupervised learning algorithms. The rest of this paper is organized as follows: In section 2, raga detection techniques are discussed along with the challenges associated to the raga recognition. In section 3, classification techniques based on supervised and unsupervised learning are discussed. In section 4, gap identification has been discussed. In section 5 overall discussions have been presented. Conclusions and directions for future research are identified in section 6. Finally references are given.

**1.1 Paper organization**

The rest of this paper is organized as follows: In section 2, raga detection techniques are discussed along with the challenges associated to the raga recognition. In section 3, classification techniques based on supervised and unsupervised learning are discussed. In section 4, gap identification has been discussed. In section 5 overall discussions have been presented. Conclusions and directions for future research are identified in section 6. Finally references are given.

**2. Raag detection techniques**

In 2013 Chordia et al. [3] found that how the raga and the tonic are both mutually attached to each other. In their study they introduced some technique to identify the raga by the histogram approach and the Hidden Markov model technique. The various studies in the same field discussed. There results suggest that the tonal features based on pitch distributions are robust, reliable features that can be applied to complex melodic music. In 2002, Tzanetakis [4] has also proposed various schemes in the English music classification based on their moods and styles of the performer as well as songs genre classification.

Past approaches to computer-based raaga recognition have based themselves on the properties of raga such as pitch class distributions or pitch sequence information as captured by note bi-grams or HMMs (Hidden Markov models) or swara intonations. The needed inputs are obtained by the pitch tracking of usually monophonic audio signals of an
The note duration heuristic assumes that a note is played for at least a certain constant span of time. Tansen is built to classify two raagas. The central idea in this approach, which is to model a raaga as HMM, was also used in [9]. The same idea is used in an attempt to automatically generate Hindustani classical music [10], but with less success.

**Pitch-class Profiles and Note Bi-grams**

In Tzanetakis et al. [4] has used the pitch class profiles and the bigrams of pitches to classify raagas. The dataset used in his system consists of 72 minutes of monophonic instrumental (sarod) data in 17 raagas played by a single artist. The HPS algorithm is used to extract the pitch. Note onsets are detected by observing the sudden changes in phase and amplitude in the signal spectrum.

**Swara Intonation**

It is often said that, in Indian classical music, a swarasthana 2 does not correspond to a fixed frequency value (with its octave equivalents). It is a region [11]. So, although two raagas share the same scale, the precise intonation of specific notes can vary significantly. In [12] they have used this clue to differentiate ragas that share the same scale intervals. They evaluated the system on 10 tunes, with 4 raagas evenly distributed in 2 distinct scale groups. They showed that the use of swara intonation features improved upon the accuracies achieved with straightforward pitch class distributions. In all the above attempts, most of the approaches which have mentioned in the beginning of the section, have been made use of. Ideally speaking those approaches should be capable of building a perfect raaga recognition system. In the following section, identify few problems that make this task difficult.

**Pitch Extraction**

Pitch detection is carried out at 10 ms intervals throughout the sampled audio file using a predominant pitch detection algorithm designed to be robust to pitched accompaniment. The pitch detector tracks the predominant melodic voice in polyphonic audio accurately enough to preserve fast pitch modulations. This is achieved by the combination of harmonic pattern matching with dynamic programming based smoothing. Analysis parameter settings suitable to the pitch range and type of polyphony are available via a graphical user interface thus facilitating highly accurate pitch tracking with minimal manual intervention across a wide variety of audio material. While the violin usually follows the melodic line, it plays held notes in this particular segment. Low amounts of reverberation were audible as well. The detected pitch track faithfully captures the vocal melody unperturbed by interference from the accompanying instruments observed by the analysis.

The comparison based on the analysis is shown in Table 2. Table 3 shows the comparison based on classifier techniques.

<table>
<thead>
<tr>
<th>S.NO</th>
<th>Authors</th>
<th>Technique</th>
<th>Features Used</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sridhar et al.[13]</td>
<td>Scale based raag detection</td>
<td>Scale is extracted from the music.</td>
<td>A training dataset of different scales with corresponding raag class is used. It is used in testing phase raag by matching scale only. Set of notes are also extracted along with the pitch and then raag is predicted. So this technique is not just pitch based detection.</td>
</tr>
<tr>
<td>3.</td>
<td>Sahasrabuddhe et al. [15]</td>
<td>Finite automata based detection</td>
<td>Note sequences are used for raag detection</td>
<td>Finite automata are build according to the rules framed based on the different note sequences in different raags. This method is also based on the transition in note sequences and the arohona and avrohna patterns are</td>
</tr>
<tr>
<td>4.</td>
<td>Pandey et al. [16]</td>
<td>Hidden Markov Model</td>
<td>Transitions in notes are used</td>
<td></td>
</tr>
</tbody>
</table>
Ekta Patel et al.

<table>
<thead>
<tr>
<th>S.NO</th>
<th>Authors</th>
<th>Technique</th>
<th>Features Used</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>Chakraborty et al. [17]</td>
<td>Scale interval based raag detection</td>
<td>Interval between scale frequencies are used</td>
<td>This method is based on intervals between the scale frequency.</td>
</tr>
</tbody>
</table>

3. Classification techniques based on supervised and unsupervised learning techniques

3.1 Supervised learning

**Naive Bayes**

It is a classification technique where the classification is based on probability [18]. It considers all elements free of each other. It computes likelihood of every component freely for a specific class name [19-21].

It is expressed as following:

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:

$$P(x/y) = \prod_{k=1}^{n} p\left(\frac{x_k}{y}\right)$$

Then the posterior probability of class ‘y’ given that x feature is in the feature set is given by:

$$P\left(\frac{y}{x}\right) = \frac{p\left(\frac{x}{y}\right) P(y)}{P(x)} = \frac{P\left(\frac{x}{y}\right) P(y)}{\sum_{j} p\left(\frac{x}{y_j}\right) P(y_j)}$$

The features for which P(y/x) is more are the most deciding features and can also be considered as principle components.

Since this methodology depends on the likelihood it can be connected to a wide assortment of spaces and results can be utilized as a part of numerous ways [22]. It doesn't require expansive measure of information before preparing to start. These calculations are computationally quick to settle on choices [23].

**Support vector machine**

It is a classifier which finds a hyper plane that clearly separates the sample points of different labels [17, 24]. It divides such that sample points of both labels and class are on different sides of hyper plane.

**Decision tree**

This sort of classifier models information with the assistance of a tree. Tree is having highlights as the inward hubs and edges show the estimations of components [25]. What's more, edges isolated hubs in light of the qualities. All the leaf hubs of the choice tree speaks to a class which is relied upon to be acquired in the event that have every one of the elements having individual qualities which are in the way from the root to that class having moderate component hubs [26].

- **Function classifier**: It sends the idea of relapse and neural system. Information is mapped to the yield. It utilizes the iterative parameter estimation plan.
- **Lazy classifier**: This requires the capacity of the whole preparing cases and backings incorporation of new information simply after characterization time. The prime point of interest of this order plan is the neighborhood estimate of the objective capacity [17]. For every inquiry to the framework, the target capacity is approximated locally consequently empowering languid learning frameworks to take care of different issues simultaneously. In any case, the detriment is that it expends an immense measure of storage room to store the whole preparing occurrence on the double. It is tedious too.
- **Meta Classifier**: These sets of classifiers are essential to find the optimal set of attributes which can be used for training the base classifier [24]. New adaptive machine learning algorithms can be constructed using these classifiers and those new models can be further used for making predictions.
- **Mi Classifier**: It is denoted as multi-instance classifiers. It comprises of numerous occasions in an illustration, yet perception of one class is conceivable just for every one of the examples [27]. Along these lines, it is an efficient learning method.
- **Rules Classifier**: Some sort of affiliation tenet is utilized for right expectation of class among every one of the qualities [28]. The measure of right forecast is characterized by the term scope and is communicated in rate or exactness structure. The associated items are may be unrelated. More than one conclusion can be anticipated [29].
- **Bayes Net**: It is a widely used technique which takes at the essential Bayes hypothesis and structures a Bayesian system in the wake of computing restrictive likelihood on every hub. It is a graphical model which is probabilistic in nature and depicts a gathering of discretionary variables
alongside their restrictive conditions through a coordinated non-cyclic chart [30].

- **Logistic:** This system utilizes relapse to anticipate the likelihood of a result which can have just two qualities. One or a few indicators are utilized to make the expectation [31]. Logistic relapse delivers a logistic bend that is limited to values somewhere around 0 and 1. The bend is built utilizing the regular logarithm of the chances of the objective variable and not the likelihood.

- **IBK:** It remains for occurrence based information representation of the preparation cases and does not close or foresee a standard set or a choice tree [32]. After an arrangement of preparing examples has been put away, the memory is hunt down the new preparing occasion. So it is tedious and requires space too.

- **JRip:** This system executes a proposed guideline learner and aggregate blunder pruning strategy to diminish mistake. It depends on affiliation rules with diminished blunder pruning methods, in this way making it a powerful strategy [33].

- **PART:** It utilizes a separation and vanquish way to deal with build a C4.5 choice tree in part for every cycle indicating the ideal guideline affiliation. Utilizing an entropic separation measure strategy, it performs occurrence based learning.

- **J48:** It is an upgraded variant of C 4.5 which spins on the ID3 calculation with some additional usefulness to determine issues that ID3 was clumsy [34]. Be that as it may, this strategy is time and space expending. At first, it constructs a tree utilizing the gap and vanquishes calculation and after that applies heuristic criteria. The standards as indicated by which the tree is created are exact and natural.

- **Random Forest:** This order calculation utilizes group techniques to acquire better prescient execution [35]. It produces yield as individual trees and is based upon the choice tree calculation [36]. It is thought to be a profoundly exact classifier and can deal with numerous variables.

### Table 3 Comparison based on classifier techniques

<table>
<thead>
<tr>
<th>S.NO.</th>
<th>Authors</th>
<th>Classifier</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Chakraborty et al. [17]</td>
<td>Naive Bayes</td>
<td>Probability based classifier</td>
<td>This is a probability based classifier based on Naive Bayes conditional probability</td>
</tr>
<tr>
<td>2.</td>
<td>Chakraborty et al. [17]</td>
<td>Bayesian Net</td>
<td>Probability based classifier</td>
<td>This is a probability based classifier based on Naive Bayes conditional probability</td>
</tr>
<tr>
<td>3.</td>
<td>Roy et al. [8]</td>
<td>J48</td>
<td>Tree based approach</td>
<td>It is enhanced version of C 4.5 algorithm and used ID3.</td>
</tr>
<tr>
<td>4.</td>
<td>Roy et al. [8]</td>
<td>Random Forest</td>
<td>Tree based approach</td>
<td>It is also a decision tree based approach but have more accuracy as compared to J48.</td>
</tr>
<tr>
<td>5.</td>
<td>Roy et al. [8]</td>
<td>Random Tree</td>
<td>Tree based approach</td>
<td>It generates a tree by randomly selecting branches from a possible set of trees.</td>
</tr>
<tr>
<td>7.</td>
<td>Ross et al. [37]</td>
<td>Phrases</td>
<td>Segmentation of phrases</td>
<td>Their proposed method was efficient in segmentation of phrases through identification of nyas and computes similarity with the reference characteristic phrase.</td>
</tr>
<tr>
<td>8.</td>
<td>Priya et al. [38]</td>
<td>C4.5 decision tree algorithm, Random Tree and Rule Induction algorithm</td>
<td>Hybrid approach</td>
<td>The training sets considered for the work are Avarohanam notes of 72 melakartha raga and 212 Janya raga. C4.5 decision tree algorithm, Random Tree and Rule Induction algorithm were utilized to classify the Melakartha raga and the Janya raga.</td>
</tr>
<tr>
<td>9.</td>
<td>Kumari et al. [39]</td>
<td>KNN classifier and SVM classifier</td>
<td>Hybrid approach</td>
<td>For data classification they used different types of classifier just like</td>
</tr>
<tr>
<td>S.NO.</td>
<td>Authors</td>
<td>Classifier</td>
<td>Category</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
<td>------------------------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>10.</td>
<td>Chordia et al. [40]</td>
<td>Bayesian decision rule</td>
<td>Probability based</td>
<td>Their system computes the pitch-class distribution and uses a Bayesian decision rule to classify the resulting twelve dimensional feature vectors, where each feature represents the relative use of each pitch class.</td>
</tr>
<tr>
<td>11.</td>
<td>Rao et al. [41]</td>
<td>Machine learning</td>
<td>Learning and training</td>
<td>They used machine learning methods on labeled databases of Hindustani and Carnatic vocal audio concerts to obtain phrase classification on manually segmented audio. Dynamic time warping and HMM based classification are applied on time series of detected pitch values used for the melodic representation of a phrase.</td>
</tr>
<tr>
<td>12.</td>
<td>Sell et al. [42]</td>
<td>logistic regression, K-NN and SVM</td>
<td>Hybrid approach</td>
<td>They have applied machine learning techniques (logistic regression, K-NN, SVM). It creates the best overall classifier.</td>
</tr>
</tbody>
</table>

3.1 Unsupervised learning approach
Unsupervised leaning is the machine learning task of inferring a function to describe hidden structure from unlabeled data. Since the illustrations given to the learner are unlabeled, there is no mistake or compensate sign to assess a potential arrangement. This recognizes unsupervised gaining from regulated learning and support learning. Implies if the objective yield is not known. Clustering like k-means, c-means are the examples of unsupervised learning approach. Neural network are categorized in both of the category if the output is known then it lies in supervised otherwise it is lies in unsupervised category. Supervised learning is used where class labels are given for training the classifier. It is suitable for large classification problems like Raag detection. The structure is always defined in terms of raag formats so there is a minor chance to use the unsupervised learning.

4. Gap identification
After studying several research works the following gaps have been analyzed in the previous techniques.
1. Key phrases identification is important as it is capable in extracting the maximum instances.
2. The attributes considered should be compared with social behaviors also.
3. Pitch and mood identification can be used as the training subset.

4. Compositions with similar patterns and dissimilar patterns should be identified separately.
5. Segmentation of the signal should be detected at the same frequency.

5. Discussion
In this survey it is found that for accurate classification of Raag in music pitch, scale and dataset should be clean and clear. It is also found that supervised learning approaches are efficient in Raag detection in music.

According to Koduri et al. [43] the tonic identification itself includes some measure of mistake; this could unfavorably affect the execution of a raaga acknowledgment framework. Neither the Carnatic nor Hindustani frameworks hold fast to any outright tonic recurrence, in this way it bodes well to manufacture a framework that can disregard the total area of the tonic. According to Katte [44] suggested that the limited database containing limited number of raga, incorrect pitch extraction, manual tonic detection, assumption made for different parameters of algorithm, different constraints on inputs, such as limitation to singers, number of swaras, time length, monophonic type are the major challenges. So improving in these attributes may help in better detection. In [19] authors suggested that the pitch tracked and pitch-class sequences can be segmented into notes using a complex-domain detection.
function. In [45] authors suggested that the melody types, method of organizing tunes based on certain natural principles can be efficient in detection. In [46] suggested that the pattern extraction from music strings is a complex problem. The main problem is to compare the strings. Some works according to the authors are only based on repeating patterns. So it may sound similar. So time frequency masking may be useful. In [47] authors suggested that pitch and raag information based detection much more useful. In [48-53] different mechanism and methods had suggested in the process of data classification and categorization and the data being utilized for the raag classification. In [54-59] there are several mood based and rhymes based recognition is suggested and recognized. This can be utilized in the several data recognition mechanism for the raag recognition and detection system.

Overall study and analysis indicate that the supervised learning techniques are capable in proper detection of raag classification. There are several factors which can improve the efficiency like it depends on the pitch of musical notes and the mood in which they are conveyed rather than the sequence of notes. The parameters like pitch and mood in the music, skipping extra tones, conversion of different data attributes and Raag tempo is important in this classification.

6. Conclusion and future work
A well-defined sequence of musical notes with a particular pitch and mood is termed as Raag. Raag detection in music is a challenging task. In this paper a survey is done on Raag detection techniques based on supervised learning. Techniques based on scale, pitch and notes sequence is studied. Several classification techniques like Naïve bayes, Support vector machine, J48, Random forest, PART and many other are studied. Several challenges in raag detection from music signal are also addressed. In this paper Raag detection mechanism out of raw music signal is studied and the basic and most frequently used supervised learning approaches are studied. It is found that different approaches are used based on the requirements may be proved to be efficient in Raag detection. It is also found that to accurately classify the Raag in addition to the sequence of notes, pitch and mood and gender detection plays an important role. This work is extended in the future considering the parameters which influence the conversion in the classification and data mapping. This may be tuned with pitch and mood for improving the detection efficiency.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References
Ekta Patel et al.


[28] Dubey AK, Dubey AK, Agarwal V, Khandagre Y. Knowledge discovery with a subset-superset approach for mining heterogeneous data with dynamic support. In CSI sixth international conference on software engineering 2012 (pp. 1-6). IEEE.


[46] Sivaranjani V, Umamaheswari J. Comparison of repeating pattern extraction techniques for audio pitch detection.


[51] Sudha R. A tool for identifying ragas using MIDI (musical instrument devices) for CMIR (classical music information retrieval). In proceedings eighth IEEE international conferences on dependable, autonomic and secure computing 2009 (pp. 837-42).


[56] Hainsworth SW, Macleod MD, Wolfe PJ. Analysis of reassigned spectrograms for musical transcription. In IEEE Workshop on the applications of signal processing to audio and acoustics 2001 (pp. 23-26). IEEE.

