Analysis of Classical Raga based music

A Seminar Report

Submitted by
Saurav Shrivastava
Roll No- 163050028

under the guidance of
Prof. Rushikesh K. Joshi

Master of Technology
in
Computer Science Engineering

Indian Institute of Technology
Bombay, India

April, 2017

Contents

Chapter 1

Introduction

Raga is very important concept in Indian classical music and used to capture the emotion of the performance.Music performer are bound to restrict the musical composition under the rules defined by the Raga framework. Automatic Raga identification can provide a basis for creating play list of songs which are similar to each other.It can also be used to evaluate the performance of singer to check how accurately a performer singing 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.

In this study we first gave brief ideas of terminology used in Hindustani classical music then we discuss Hidden markov model based raga detection method then we move to more sophisticated model which use pitch features of audio waveforms for classification of ragas.Finally we dig more deep into details of how these pitch class features are implemented and used to for classification process.

1.1 Notes and Scale

There are total 12 musical notes(7 shuddha + 4 komal + 1Tivra swara's) namely defines as shuddha Shadja(Sa),komal Rishabha(ri),shuddha Rishabha(Ri), komal Gandhaara(ga), shuddha gandhar, komal madhyam , shuddha madhyam, Tivra Panchama, komal Dhaivata, shuddha Dhaivata, komal Nishaada, shuddha Nishaada formed an octave.Every musical melody is composed using collection of subsets of these 12 notes.

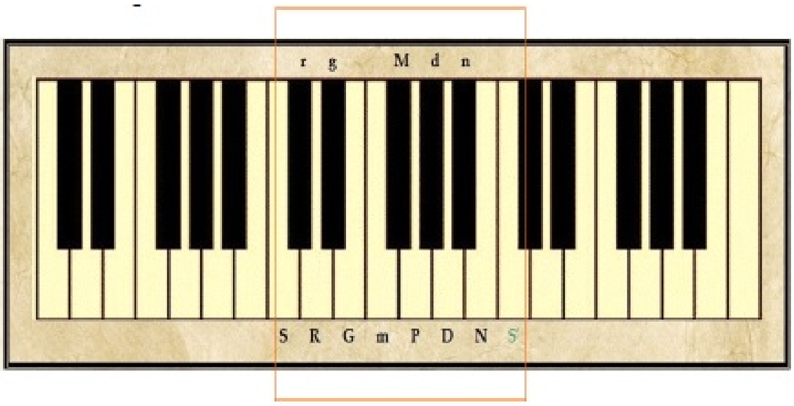

Figure 1.1: Scale

A scale is collection of notes(discrete values of pitch or pitch interval with respect to tonic) that is used in the representation of a piece of musical work.Usually the pattern of a scale in one octave is repeated across all the other octaves.

1.2 Shruti

According to Muni Bharata(200 BC), the author of Natyashastra,a shruti is the smallest possible interval that can distinguish one sound from another as lower or higher pitch[1].

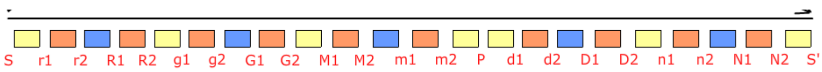

Figure 1.2: Shruti distribution

As Shown in fig.1.2 there are total 22 shruthis,namely Shadj-4,rishabh-3,gandhar-2,madhyam-4,pancham-4,dhaivat-3 and nishad-2.

1.3 Raga

Raga is a melodious framework which a performer sets for his performance and improvises over it with his own style.it also specifies a set of allowed notes (minimum five) that can be played in that raga. A Raga is characterised by several attributes,like its Vaadi-Samvaadi, Aarohana-Avrohana and Pakad, besides the sequence of notes which denotes it. It is important to note here that no two performances of the same Raga, even two performances by the same artist, will be identical.A certain music piece is considered a certain Raga as long as the attributes associated with it are satisfied[2].

1.4 Tonic

"Tonic" is the base pitch selected by an artist and it serves as the foundation on which the artist builds his performance. The entire performance will be relative to the tonic. A performer chooses a tonic which is most comfortable for him/her to fully explore their vocal (or instrumental) pitch range. All accompanying instruments are tuned in relation to the tonic chosen by the lead performer.

Other notes used in the performance derive their meaning and purpose in relation to the Sa and the tonal context established by the particular raga

"Vadi(Sonant) is the most sonant or most important note of a Raga. It is usually the swara which is repeated the greatest number of times, and often it is the swara on which the singer can pause for a significant time[3].

1.5 Pakad

Pakad is a string of notes characteristic to a Raga to which a musician frequently returns while improvising in a performance. Each note in the Pakad can be embellished and improvised around to form new melodic lines. One common example of these embellishments is the splitting of Pakad into several substrings and playing each of them in order in disjoint portions of the composition, with the repetition of these substrings permitted.

In spite of such permitted variations, Pakad is a major identifying characteristic of a Raga and is used even by experts of Indian classical music for identifying the Raga been played. A Pakad is a catch-phrase of the Raga, with each Raga having a different Pakad. it is not necessary for a Pakad to be sung without any breaks in a Raga performance. Since the Pakad is a very liberal part of the performance in itself, standard string matching algorithms were not guaranteed to work. Approximate string matching algorithms designed specifically for computer musicology.

Chapter 2

Matching Using Hidden Markov models and pakads

An Automatic raga identification system called as *Tansen* describe the use of HMM along with pakad matching for identification of Raga[4].

2.1 Note Transcription

In the first stage pitch contour is extracted from audio file using Praat software,then transcription of note on the pitch contour is done using two given below methods

- **Hill Peak heuristic:**
  this method assume that whenever a note starts there is huge fluctuation in the pitch.Given a sequence of pitch$p_1, p_2, p_3, \ldots, p_{i-1}, p_i, p_{i+1}, \ldots, p_n$ at the instants $t_1, t_2, t_3, \ldots, t_{i-1}, t_i, t_{i+1}, \ldots, t_n$

- **Note duration heuristic:**
  This heuristic assumes that a note will last for atleast 25ms.First the were computed for all pitch values in the pitch contour.The history of the last k notes was maintained.The current note was accepted as new note only if current note is dominant note in the history. the dominant note is one which occur more than m times(fixed).

2.2 Hidden markov model

Let the HMM be λ with the following specifications:

- there are three octaves and each octave has 12 notes ,therefore they have created $N = 12 \times 3 = 36$
- the transition probabilities are $A = a_{ij}$ specify the probability that a note j occur after a note i.
- $\pi = \pi_i$ represents the initial state probability,i.e the probability that the first note in a performance was i.
- $B = B_i(j)$ is the probability of output a certain note j at state i.It is defined as follows:

$$B_i(j) = \begin{cases} 1 & \text{if i = j} \\ 0 & \text{if i} \neq \text{j} \end{cases}$$

2.3 Pakad Matching

Although there is fixed pakad for each raga,during performance only parts of the pakad may be played and even the pakad and even may be varied slightly.Two methods are used given as follows:

- To match an input note-sequence with the pakad of raga I a score γ_I was maintained which is defined as

$$\gamma = \frac{m_I}{n}$$

  where m_I = The maximum number of notes of the pakad of raga I identified in the input note-sequence.
  where n_I = The number of notes in the pakad of raga I.

- Another score was maintained which keep track of the number of times substring of the pakad of a raga appear in the input note-sequence.

$$score_I = \sum_n \sum_j freq_{j,n,I}$$

  $freq_{j,n,I}$ = the number of times the j^{th} substring of length n of the pakad of Raga I occurs in the input note-sequence.

2.4 Classification

Classification process consist of three steps:

1. The likelihood probability of the input note-sequence having been generated by λ_I was computed $\forall I,\ 1 \leq I \leq N_{Ragas}$.The value obtained then sorted in increasing order and then the indices were reordered accordingly.

   $$\text{if} \quad \frac{\text{prob}_{N_{\text{Ragas}}} - \text{prob}_{N_{\text{Ragas}}-1}}{\text{prob}_{N_{\text{Ragas}}-1}} \geq \eta_I \quad \text{then } Index = N_{Ragas}$$

2. Otherwise,the value of $\gamma_I$were sorted and indices reordered accordingly..

   $$\text{if} \quad \text{prob}_{N_{\text{Ragas}}} \geq \text{prob}_{N_{\text{Ragas}}-1} \quad and \quad \frac{\gamma_{N_{\text{Ragas}}} - \gamma_{N_{\text{Ragas}}-1}}{\gamma_{N_{\text{Ragas}}-1}} \geq \eta_\eta$$

   $$\text{then } Index = N_{Ragas}$$

3. Otherwise

   $$\text{Index} = argmax(logp(O|\lambda_I) + (K \times score_I))\ ,\ 1 \leq I \leq N_{Ragas}$$

Accuracy gain roughly by 10% using HMM with pakad matching compare to using only HMM.

Chapter 3

Pitch-class Distribution and Dyad Distribution Matching

Pitch class distribution represent the probability distribution of the various pitches on the chromatic scale,whereas Pitch-class Dyad Distribution(PCDDs) represent the probability of the dyads as feature for raga recognition[5].

3.1 Pitch detection

Harmonic product of spectrum (HPS) is used to detect pitch at sampling rate of 40ms.Each segment of audio is divided into frames using Gausian window with 75% overlapping.

Also onsets are required for generating PCDDs.they were detected using thresholding complex detection function.The DFT of 128 samples are overlapped 50% by using rectangular window.

3.2 Pitch class distribution

Pitch contour generated by above step is used to compute the probability distribution of pitch divided into fixed range bins,here no of bins depend on the range of pitch.The bins of histogram correspond to the chromatic notes(12 semitones or 7 shudha swara + 5 ashudha swara) of 5 octaves centered around tonic(vaadi).It is important to note here that music in performance may varies over multiple octaves but still notes will belong to be consider from one octave,this is due to very less significance of difference of octaves for raga detection, then folding of 5 octave to one is done to provide pitch class distribution.The onset detection is not required for PCD computation because onset does not affect the probability of pitch occurring.

3.3 Pitch class Dyad distribution

Pitch classed are arranged into 2 groups,which are called dyads in musical term. The pitch class dyad distribution is probability of transition from one note to another.For extracting PCDDs,The detected onset points are used to segment the pitch contour into notes.the next step was to assign each note a pitch class label,Since each note can last over many sample,The pitch estimates over these samples are discretized by assigning center value of the bin to each note.The mode of the resulting pitch value for a note was assigned to that note.All the octaves are folded into one.it must be noted that short glide from detected note to another note and the again forth back would not be detected as separate note because mode of the note is considered.

In cross validation experiment using SVM classifier 78% accuracy was achieved and 75.3% on unseen data.In case of PCDDs 97.1% and 57.1% respective accuracies are achieved.

Chapter 4

Folded Pitch Distribution Matching

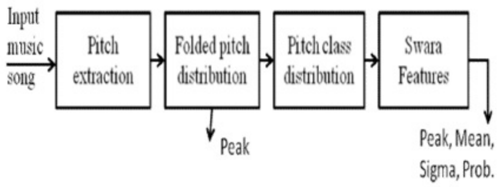

Figure 4.1: Process of Raga Identification

4.1 Preprocessing and feature extraction:

Vocal performances from various artist in 4 ragas (Desh,Tilak,Kamod,Bihag) is used as training data and each performances have different segment length this data is preprocess by converting these performances into mono channel with sampling rate at 22050Hz then vocal pitch is extracted from each of these segment at different interval range from 100 Hz to 1000 Hz, thereafter written to contour file,this gives us pitch distribution.Pitch distribution basically provides the probability of occurrence of a pitch value over the segment duration.

4.2 Folded Pitch distribution:

Process to fold pitch distribution(FPD) into a octave as follows:

- An arbitrary position (256Hz) was chosen for the initial bin of the FPD.
- Bins were logarithmic-ally spaced at 5 cent intervals to give a total of 240 bins.
- A pitch f in the pitch distribution was assigned to bin n in the FPD is given as :

$$n = \left(240 \log_2 \frac{f}{256}\right) \mod 240 \tag{4.1}$$

- For a given input tonic pitch F, and the corresponding FPD bin number computed as N, all the bins in a 100 cent window around the N th bin were examined and the peak was found.
- The bin corresponding to the peak was considered to be the tonic bin. The FPD was then rotated so that the tonic bin became the first bin.

4.3 Pitch class Distribution

PCDs are distributions with 12 bins that represent the probability of occurrence of the 12 swaras over one octave.these 12 bins are corresponding mapped with 12 swaras(7 shudha + 4 komal + 1 Tivra)

The PCDs were constructed from tonic aligned FPDs as follows.

- The boundary between two bins was defined as the arithmetic mean of the centre of the two bins
- All the FPD bins which fell within the boundaries of a PCD bin contributed to that PCD bin.

These four features for each swara were extracted from the FPD of each performance

1. Peak:The most likely position of the swara (in cents)
2. Mean: The mean position of the swara (in cents)
3. Sigma: The standard deviation of a swara (in cents)
4. Prob: Overall probability of a swara.

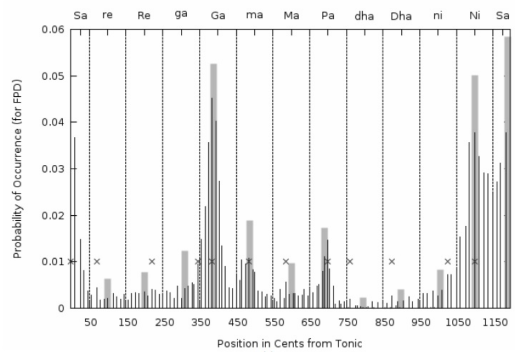

Figure 4.2: Pitch class distribution

4.4 Classification:

They used a Nearest Neighbour Classifier with leave-one-out cross validation for classification of raga with combination of Euclidean distance and KL distance to measure the distance between them.

$$d(swara_{k_i}, swara_{k_j}) = KLdist(prob_{k_i}, prob_{k_j})\times$$

$$\sqrt{(peak_{k_i} - peak_{k_j})^2 + (mean_{k_i} - mean_{k_j})^2 + (sigma_{k_i} - sigma_{k_j})^2 + (prob_{k_i} - prob_{k_j})^2}$$

kullback leibler(KLdist) distance defined as follows:

$$KLdist(p, q) = KL(p||q) + KL(q||p)$$

where KL(p,q) is

$$KL(p||q) = \sum_f p(f)\log_2 \frac{p(f)}{q(f)}$$

Chapter 5

Implementation

Our dataset consist of 4(Raga Deshi,Tilak-Kamod,Bhairavi,) audio wavefile recorded on monochannel

5.1 Extraction of vocals

Extraction of vocals require open source of audacity tool,However there is no fixed general method to extract vocals from every raw audio waveform.it is challenging to extract vocals in polyphonic melody because instrumental are recorded on sidetrack while vocal are recorded on mono-track and both are overlap with each other.Various approach are followed to extract vocals depending on type of vocals are present in raw audio

5.2 Extraction of pitch from vocals

Open source tool Praat is used to extract pitch contour from audio vocals. Following steps are performed for pitch extraction:

1. Open the audio waveform with vocals
2. Select the audio file and click on View panel on right
3. Change pitch using option (Pitch → Pitch setting) from 0 to 1200Hz
4. Select the waveform and extract pitch using option(Pitch→Pitch listing) and save as text file[6].

5.3 Forming PCD and FPD

FPD is formed using as described in section4.2 where pitch contour file is used to distribute all the pitches in the segment into bins.Since pitch ranges from 1200Hz range and each bin consist of 5 cent window, therefore we have total 240 number of bins.these bins are distributed as per the equation 4.1.

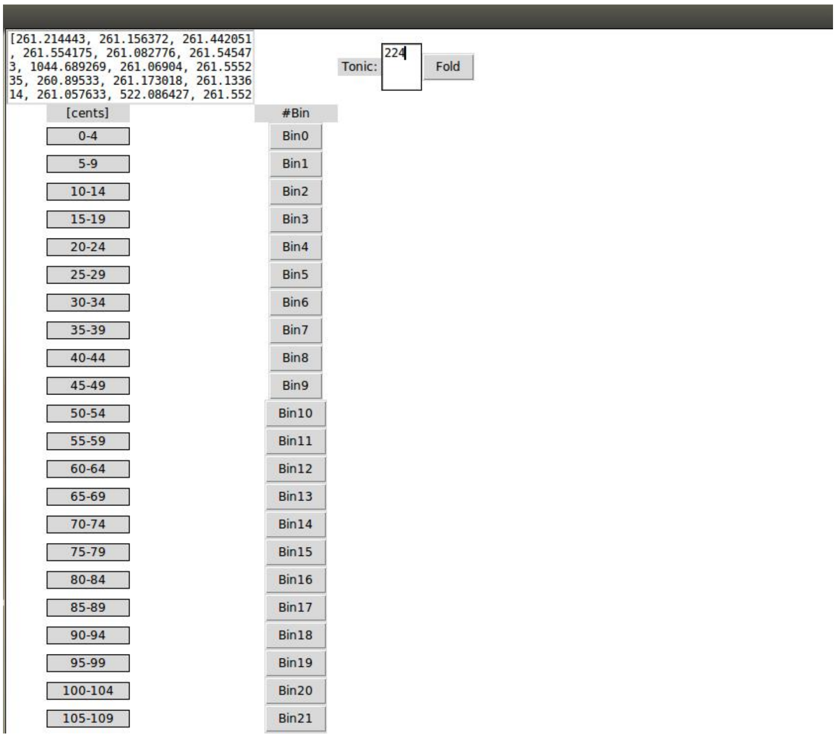

Figure 5.1: Folding Pitch Distribution

As shown in figure 5.1 UI also provide the capability to input tonic pitch,this can be used to fold the pitch distribution based on the tonic frequency i.e This folding ensure that tonic bin becomes the last bin.

PCD is formed by folding the 240 bins FPD to equally spaced 12 PCD bins of size 100 cents.Probabilty distribution of these PCD bins are formed by taking mean of pitches fall into the range of a bin.

this PCD can be depicted in the figure4.3,and can be used to retrieve features such as prob,meam,variance of swaras for training the model for a particular raga.

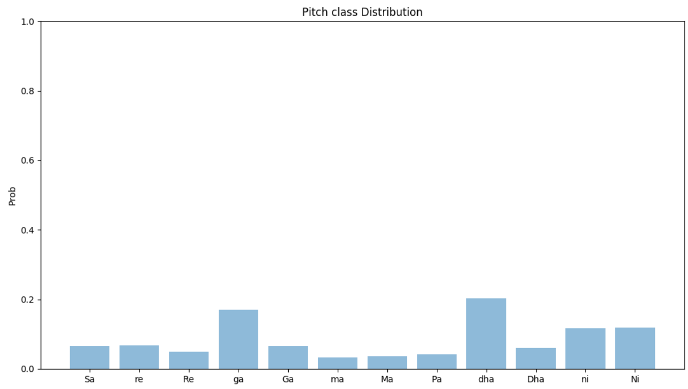

Figure 5.2: Pitch Class Distribution

Chapter 6

Conclusions and future work

We discussed about musicological studies of various terms and methodology such as HMM and PCD,PCDD's for Raga detection.Also we have study the swara feature can be used to capture intonation information in raga.We observed from chapter 2 that only notes sequence information is not sufficient.Later in chapter 3 when multiple feature such as pitch class and dyad distribution helps to improve performance of Automatic raga recognition system.Finally we implemented the folded pitch distribution and formed Pitch class distribution and study the behavior of 4 ragas Desh,tilak,kamod and bhairavi.

For the music to be pleasing,consonance must occur amongst the notes.In Hindustani classical music,the tonic is constantly present in the drone.As a result,all the notes should be consonant with the tonic note.It might be possible to make use of this information to detect the tonic.Also with recent advancement in computation capabilty,DNN based models such as CNN and RNN can be used to extract Ragas features and with more availability of more data we can expect improvement in results.

Bibliography

[1] Wikipedia, "Shruti (music) — wikipedia, the free encyclopedia," 2017. [Online; accessed 30-April-2017].

[2] Wikipedia, "Raga — wikipedia, the free encyclopedia," 2017. [Online; accessed 30-April-2017].

[3] Wikipedia, "Vadi (music) — wikipedia, the free encyclopedia," 2016. [Online; accessed 30-April-2017].

[4] G. Pandey, C. Mishra, and P. Ipe, "Tansen: A system for automatic raga identification.," in *IICAI*, pp. 1350–1363, 2003.

[5] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions.," in *ISMIR*, pp. 431–436, Citeseer, 2007.

[6] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.