Raga Recognition using Swara Intonation

M.Tech Thesis Stage-I Report

Submitted in partial fulfillment of requirements for the degree of

Master of Technology

By

Saurav Shrivastava
Roll Num: 163050028

under the guidance of

Prof. Rushikesh Joshi

Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Oct, 2017

Acknowledgement

I am extremely thankful to my M.Tech advisor **Prof. Rushikesh Joshi** for helping me through out the research project. All the research meetings and the discussions have been very crucial in resolving doubts and coming up with solutions related to the Tonic detection. I would like to thank **Mr. Joe Ross**, phD Scholar of CFILT Lab, IIT Bombay for giving insights throughout the thesis work. I also thankful to **Sankalp Gulalti**, Music Technology Group, Universitat Pompeu Fabra for suggestions and providing CompMusic.I can not thank enough my parents for their constant encouragement and support. I am also very grateful to my peers **Mohammad Kaleem** and **Lavudiya Goutam** for all the discussions held through out the semester. Finally, I would like to thank other members of Swarg Lab for their help.

Abstract

In this report we discuss the possibility of automating raga recognition using intonation information.Raga recognition has several interesting applications in digital music indexing, recommendation and retrieval. However, this problem is hard due to (i) the absence of a fixed frequency for a note (ii) relative scale of notes (iii)oscillations around a note, and (iv) improvisations.first, we briefly provide introduction of Hindustani music and compare it with western music.Also we study several important concepts in music theory which provide the better insight of the problem and then later explain the concept of tonic and discuss past methods used to identify the tonic from audio.In chapter 3,we discuss an approach to identify tonic using multi-pitch representation and detail description of this method and its implementation is provided.we are successful in recognizing tonic of 35 audio performance out of 40 using this method.

In chapter 4 , We attempt to solve the raga classification problem using swara's features extracted from pitch class profile.Accuracy reported using this method is approximately 76%.We also verify the result of this approach on CompMusic dataset.Finally we conclude our discussion with stating the results and incorporating the temporal information as well for classification in the next phase.

Contents

1 Introduction **1**
1.1 Comparison between Western and Classical music 1
1.2 Motivation and Goals 2

2 Musicological Background **3**
2.1 Introduction 3
2.2 Notes and Scale 3
2.3 Shruti 4
2.4 Consonance 4
2.5 Raga 5
2.6 Tonic 5
2.7 Tonal structure of tanpura 5

3 Tonic Detection **7**
3.1 Tonic in Indian Art music 7
3.2 Methodology 7
3.2.1 Sinusoid Extraction 8
3.2.2 Salience Function computation 9
3.2.3 Tonic Candidate generation 9
3.2.4 Classification & Results 11

4 Raga Classification **12**
4.1 Introduction 12
4.2 Details of dataset 13
4.3 Predominant pitch melody extraction 14
4.4 Folded Pitch Distribution 14
4.5 Pitch Class Distribution 14
4.6 Experiment Details 15
4.7 Classification 16
4.8 Results and Analysis 17

5 Conclusion and Future work **18**
5.1 Conclusion . . . 18
5.2 Future work . . . 18

Bibliography **19**

List of Figures

2.1 Scale 4
2.2 Shruti distribution 4
2.3 Tanpura with its all body components 6

3.1 Block diagram of Tonic detection process 8
3.2 Pitch histogram 10
3.3 Multi-pitch histogram 11

4.1 Process of Raga Classification 12
4.2 Folded Pitch distribution 15

List of Tables

2.1 Ratio of Shruthi centers with respect to Tonic 5

4.1 Dataset Description 13
4.2 Swaras used in various Ragas 13
4.3 Swara features for segments of Bhairav and Puria Dhanashri Raga by same artist 16
4.4 Swara features for segments of Bhoopali and Hamsadhwani Raga by same artist . 16
4.5 Experiment results 17

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.

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 task of identifying scale in quite simple in western music but becomes particularly challenging 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 a 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 while citing a raga as long as the characteristics of raga are intact.

1.1 Comparison between Western and Classical music

In Western music the two defining attribute in any performance are key and time signature.The key is set of notes which are used while performing.This may vary from part to part of performance.The time signature contains information about the rhythm about the performance. The two distinctive attributes in any Indian Classical music performance are Raga and taal.The raga is melodic framework within which the performer stays and improvise.We may say that raga in Indian classical music is analogous to key in western music.the tala is the rhythm cycle that is used in the performance and consists of a fixed number o counts or time units and is analogous to time signature in western music.

The fundamental difference between Indian classical and western classical music is that in the former,the raga remains constant through out a performance,whereas in western music ,the key may change during the performance,The raga itself is far more complex than a key.A key just defines the set of notes that may be used in the performance whereas the raga has rules about when a certain note may be played.

1.2 Motivation and Goals

Music Content Analysis(MCA) refers to automated methods for extracting the musical information from audio files.These techniques can be used in music search engines that use humming or singing as the query.Another area where it may be used is in detecting how accurately a performer is maintaining timing or key/raga. MCA can also be used for generating the score or transcription of the music from given audio file.

Our goal is to develop a fully automatic raga recognition system which extract the features from raw audio waveform and these features are used for training the raga recognition model.However this involves multiple challenges such as identifying the singer pitch ,removing noise(Instruments),Motifs detection etc.

In further discussion we have proposed the solution to overcome these challenges. In Chapter 2 we provide background about the theory of Hindustani Music and thereafter in Chapter 3 we have provide an approach to detect the vocal pitch of singer(tonic).Later in subsequent chapter we have make use of this tonic pitch to extract swara features from audio waveform which in turn used to classify raga.

Chapter 2

Musicological Background

2.1 Introduction

In this chapter we will provide detail description of selective musicological concepts such as notes,scale,tonic and consonance etc.Further we will discuss certain Hindustani musical concepts which will help us to gain better insight of methods that we will discuss in chapter 3 for tonic detection and in chapter 4 for classification of raga.

2.2 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.[9]

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.

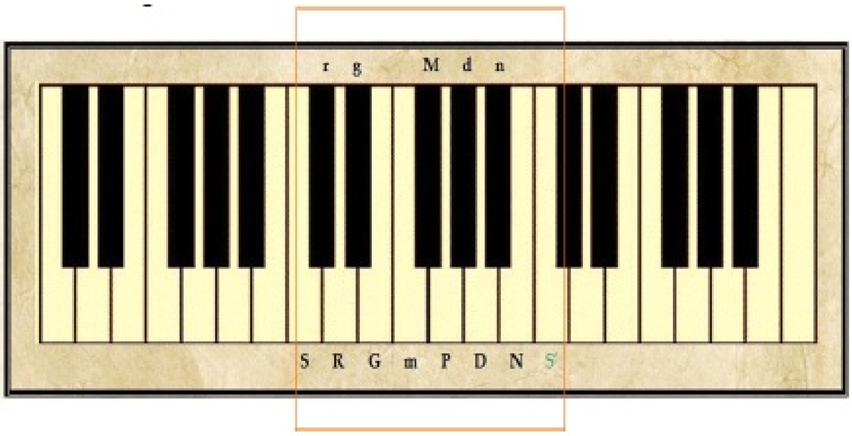

Figure 2.1: Scale

2.3 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.[9]

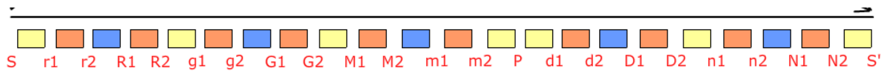

Figure 2.2: Shruti distribution

As Shown in fig.2.2 there are total 22 shruthis,namely Shadj-4,rishabh-3,gandhar-2,madhyam-4,pancham-4,dhaivat-3 and nishad-2.

2.4 Consonance

Two sounds that are pleasant when played together are said to be in consonance.A complex tone is a combination of two or more sinusoids that are uniformly spaced on the frequency scale.The sinusoid with the lowest frequency which are multiples of the fundamental frequency are called partials.When two complex tones are played together and one or more of their partials exactly coincide,they sound pleasant and are considered as consonant. For example,if two complex tones with frequencies $f_1 = 100Hz$ and $f_2 = 150Hz$ are played together then they will be consonant because they share a common partials at 300Hz i.e $3f_1 = 2f_2$.The human ear has a frequency range of about 10Hz to 20kHz and is very sensitive up to 5kHz so if common partials lies below 5kHz,then consonance will be clearly perceived.[2]

Shruti:	S	R_1	R_2	R_3	R_4	G_1	G_2	G_3	G_4	M_1	M_2
Ratio:	1/1	21/20	16/15	10/9	9/8	32/27	6/5	5/4	81/64	21/16	4/3

Shruti:	M_3	M_4	P	D_1	D_2	D_3	D_4	N_1	N_2	N_3	N_4
Ratio:	45/32	64/45	3/2	63/40	8/5	5/3	27/16	16/9	9/5	15/8	243/128

Table 2.1: Ratio of Shruthi centers with respect to Tonic

2.5 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 characterized 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.[8]

2.6 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.[7]

2.7 Tonal structure of tanpura

The presence of the drone in the background is a characteristic feature of Indian art music and plays a very crucial role. The drone acts as a reference of the music to a tonal background, reinforcing all the harmonic and melodic relationships. The presence of the drone brings out the issues of intonation and consonance more than it otherwise would have been.

Tanpura is a long-necked plucked lute, which comes in different sizes that corresponds to the different pitch ranges it can produce.It usually has 4 strings which are plucked serially in a regular pattern to create a rounded resonant sound.

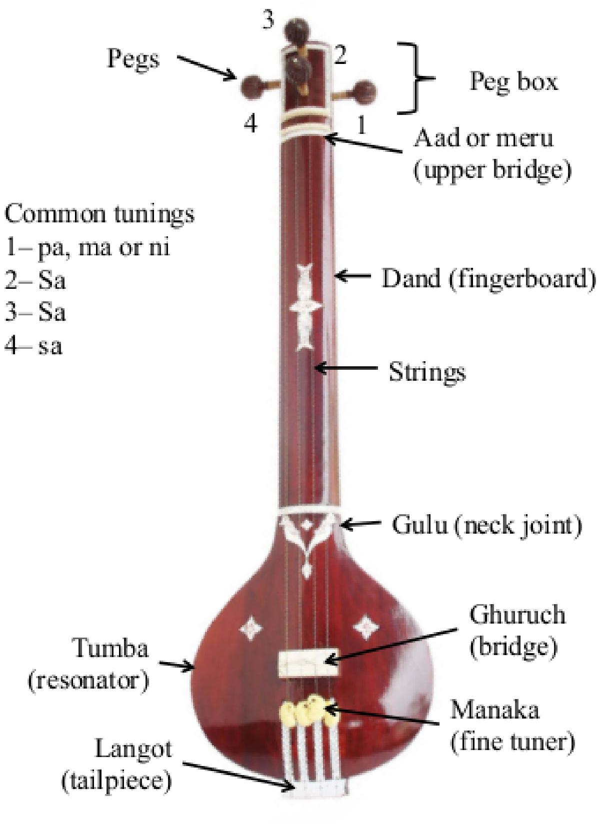

Figure 2.3: Tanpura with its all body components

Figure 2.3 shows a 4 string tanpura with its important body parts duly labelled and the frequently used tuning configurations listed on the side. The pegs corresponding to 4 strings are marked with the numbers 1 to 4 respectively. The two middle strings of the tanpura (corresponding to pegs 2 and 3) are tuned to the tonic pitch of the lead performer (Sa), while the fourth string (corresponding to peg 4) is tuned an octave below the tonic pitch (sa). In addition to reinforcing the tonic pitch, tanpura also produces secondary pitch classes. The first string of the tanpura (attached to peg 1) is frequently tuned to the fifth (pa) with respect to the tonic pitch, in the lower octave, resulting in pa-Sa-Sa-sa type of tuning. For rags which omit pa (fifth), the first string is tuned to the natural fourth (ma) as ma-Sa-Sa-sa.[4]

Chapter 3

Tonic Detection

3.1 Tonic in Indian Art music

It is the base pitch of a performer, carefully chosen in order to explore the full pitch range effectively in a given rāga rendition. The tonic acts a reference and the foundation for the melodic integration throughout the performance i.e. all the tones in the musical progression are constantly referred and related to the tonic pitch. All the accompanying instruments such as tablā, violin and tānpūrā are tuned using the tonic of the lead performer. It should be carefully noted that tonic in Indian art music refers to a particular pitch value not to a pitch-class. The frequency range of the tonic pitch for male and female singers spans more than one octave (roughly 110-260 Hz).

Both the performer and the audience need to hear the tonic pitch throughout the concert. This is accomplished by a constantly sounding drone instrument in the background of the performance, which reinforces the tonic. Along with tonic, the drone also emphasizes other notes like the fifth, fourth or sometimes the seventh, depending on the choice of the rāga. Essentially, the drone is the reference sound that establishes all the harmonic and melodic relationships between the pitches used during a given performance. Typically the drone is produced by either the tanpura or sruti box for the case of vocal music and by the empathetic strings of instruments such as sitar, sarangi and vina for the case of instrumental performances.

3.2 Methodology

Various modules used in the process of tonic identification is shown in figure Subsequent subsection will briefly discuss about each of these module in details.

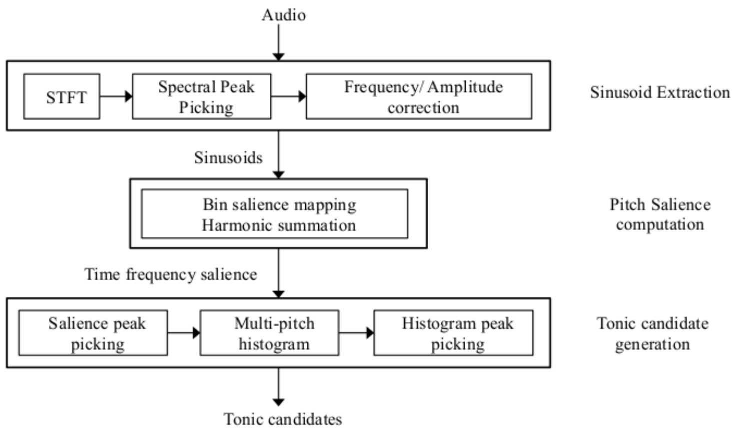

Figure 3.1: Block diagram of Tonic detection process

3.2.1 Sinusoid Extraction

We start off by extracting the sinusoidal components of the audio signal.This process is divided into three parts spectral transform, peak picking and sinusoid frequency and amplitude correction.

We use Short-Time Fourier Transform (STFT) to transform the audio signal from a time domain to a time-frequency domain representation. STFT is given by:

$$X_l(k) = \sum_{n=0}^{M-1} w(n).x(n + lH)e^{-j\frac{2\pi}{N}kn}, \qquad (3.1)$$
$$l = 0, 1... \text{ and } k = 0, 1, ..., N - 1$$

where x(n) is the time domain signal, w(n) the windowing function, l the frame number, M the window length, N the FFT length and H the hop size.Given the FFT of a single frame $X_l(k)$, spectral peaks p_i are selected by finding all the local maxima k_i of the magnitude spectrum $|X_l(k)|$.

Note that not all the spectral peaks correspond to valid sinusoids, there are many spurious peaks (relatively low energy) generated as a result of the windowing.In order to extract true sinusoids and filter out these noise we apply simple energy threshold to discard these spurious spectral peaks,Energy threshold(T_s) is calculated as follows:

$$T_s = max(T_r, \alpha), \qquad (3.2)$$
$$T_r = E_m + \beta$$

where T_r is the relative threshold w.r.t the maximum spectral peak (E_m) for each frame, α is the an absolute threshold and β is a relative threshold parameter. For the experiment we use $\alpha = -70$ dB and $\beta = -40$ dB.[5]

3.2.2 Salience Function computation

The extracted sinusoids are used to compute a salience function, a time-frequency representation indicating the salience of different pitches over time.basically,salience of a given frequency is computed as a weighted summation of energy found at all the integer multiples (harmonics) of that frequency. This brings out the fundamental frequency component of the complex sinusoidal mixture, as it receives contributions from all its harmonics. The peaks of the salience function at a given time instance represent the prominent pitches present in that frame.[6]

The mapping between a given frequency value f_i in Hz to its corresponding bin index $b(f_i)$ is given by:

$$b(f_i) = 1200\frac{\log_2(f_i/f_r)}{\eta} + 1 \tag{3.3}$$

where f_r is reference frequency and η is the bin resolutions in cents.we use $f_r = 55$ Hz and $\eta = 10$

for each frame,the salience pitch (s_j) for the j^{th} bin is computed using N_p number of sinusoid with frequency f_i and amplitude a_i and given as :

$$S(j) = \sum_{h=1}^{N_h}\sum_{i=1}^{N_p} g(j, h, \hat{f}_i).(\hat{a}_i)^{\beta} \tag{3.4}$$

where N_h is the number of harmonics considered , β is a magnitude compression factor and $g(j, h, \hat{f}_i)$ is the function that defines the weighting scheme defined as:

$$g(j, h, \hat{f}_i) = \begin{cases} \cos^2(\delta.\frac{\pi}{2}.\alpha^{h-1}) & |\delta| \leq 1 \\ 0 & if|\delta| > 1 \end{cases}$$

where δ is the distance in semitone between folded frequency $\hat{f}_i$ and center frequency of bin j and $alpha$ is harmonic weighting parameter.We use $\alpha = 0.8$,$N_h = 20$ and $\beta = 1$ in the current implementation.

3.2.3 Tonic Candidate generation

We proceed to extract the potential tonic pitch candidates using the salience function computed in the previous step. Each candidate is represented by a frequency and an amplitude value. The process of generating the tonic candidates includes three sub-tasks

1. Detecting peaks of the salience function.
2. Computing a pitch histogram.
3. Extracting candidates as the peaks of this histogram.

Detecting peaks of the salience function

Peaks of the salience function represent the prominent pitches of the lead instrument, voice and other predominant accompanying instruments present in the audio recording at every point in time.

We start by selecting the peaks of the salience function at each frame.We chose a lenient frequency range of 110-370 Hz to select the peaks from the salience function. This ensures that the range of the tonic pitch for both male and female singers is covered. Moreover, the range spans nearly 2 octaves, and therefore the system must be able to identify not only the correct tonic pitch-class but also the octave in which it is played. [4]

Computing a pitch histogram.

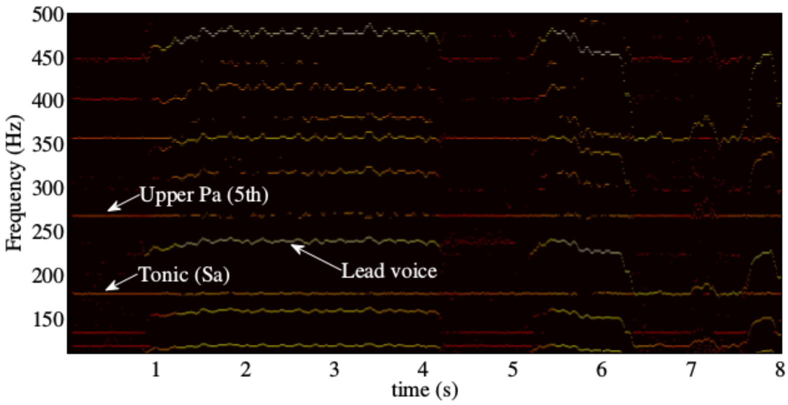

Figure 3.2: Pitch histogram

Extracting candidates as the peaks of this histogram

We select the 10 most salient pitch values within the frequency range of 110-370 Hz from each frame. The selected peaks are used to construct a multi-pitch histogram, which represents the cumulative occurrences of different pitches.[4]

We compute the distance between every tonic candidate (p_i) and the most salient candidate in the histogram (p_1). This gives us a set of features f_i ($i = 1...10$) (pitch-interval features), where f_i is distance in semitone between p_i and p_1 . Another set of features a_i ($i = 1...10$) (amplitude features) include the amplitude ratios of all the candidates with respect to the highest candidate.

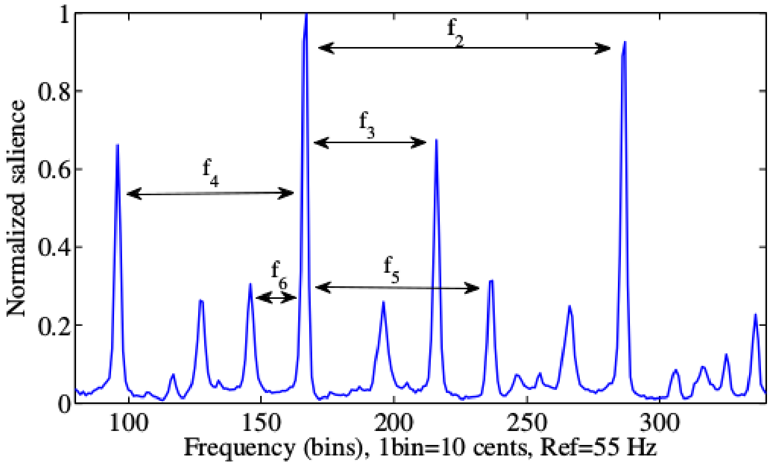

Figure 3.3: Multi-pitch histogram

3.2.4 Classification & Results

We have annotated each audio excerpt with a class label and use 20 features (f_i, a_i) to train a classifier in order to predict the class label.In this method the class label of an instance is assigned as the rank of the tonic pitch in the ordered list of all the candidates, arranged in descending order of their peak magnitude.

For example, if the candidate corresponding to the tonic pitch is the second highest peak of the histogram, we assign a label "Second(2)".J48 decision tree is used for classification task and Among 40 performances in our dataset we are able to detect tonic of 35 audio performance with a precision of 10Hz.

Chapter 4

Raga Classification

4.1 Introduction

Various methods of Raga classification can be broadly categorize into two main approach based on the type of information it use to classify raga:

1. **Non-Temporal Information:** In this approach probabilities of all the frequency spread over one octave(FPD) or Probabilities of each swara across one octave(PCD) information is used for classification.We will follow this approach in our further discussion.

2. **Temporal Information:** Information about the sequence in which swaras occur.PCDD and HMM are used to capture this information.

The main building block of Raga classification process is shown in fig 4.1. First step is to extract vocal pitch from polyphonic audio waveform.Output of this block is pitch contour which is further used to create pitch distribution.In subsequent steps pitch distribution is logarithmically distributed over 240 bins with tonic frequency as reference pitch.Further this FPD is processed and divided into 12 equal size partition where each partition represent pitch class of each swara .Some features are extracted from these partition which is further used in raga classification.

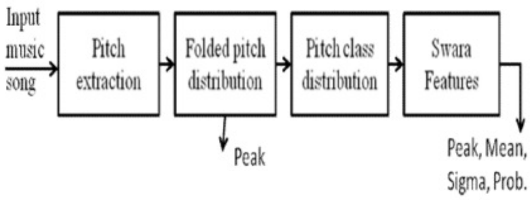

Figure 4.1: Process of Raga Classification

4.2 Details of dataset

We provide brief details of dataset in this section and further motivation behind choosing set of ragas is provided. Each audio recording is pre-process by converting performances into mono channel with sampling rate at 22050Hz.Later each audio waveform is segmented to 30 seconds frames.[1]

Table 4.1: Dataset Description

Raga	**Artist**	**Duration (sec)**	**Class Label**	**Length of segment (sec)**	**Tonic Pitch (Hz)**
Bhairavi	Abdul Rashid Khan	30	1	5138	138
Bhairavi	Hirabai Barodekar	30	1	3410	209
Puria Dhanashri	Bhimsen Joshi	30	2	3740	156
Puria Dhanashri	Pandit Jasraj	30	2	2908	142
Bhoopali	Ghulam Ali Khan	30	3	1440	140
Bhoopali	Kesarbai Kerkar	30	3	1829	196
Hamsadhwani	Veena Sahasrabudhe	30	4	1676	120
Hamsadwani	Rajan Mishra	30	4	1277	99

The motivation behind choosing these set of ragas (Bhairavi and Puriya dhanashri) and (Bhoopali and Hamsadhwani) is that they belong to the same tonic scale and share same Jaati.therefore to verify the robustness of method we deliberately perform classification on pair of ragas which are structurally similar.

Table 4.2: Swaras used in various Ragas

Swara Name/ Raga	**Sa**	**re**	**Re**	**ga**	**Ga**	**ma**	**Ma**	**Pa**	**dha**	**Dha**	**ni**	**Ni**
Bhairav	✓	✓			✓	✓		✓	✓			✓
Puria Dhanashri	✓	✓			✓	✓		✓	✓			✓
Bhoopali	✓		✓		✓			✓		✓		
Hamsadhvani	✓		✓		✓			✓				✓

4.3 Predominant pitch melody extraction

This task aims at estimating the main or predominant melody line from a polyphonic or heterophonic audio music recording. Vocal pitch of singer is extracted from each of these segment at different interval range from 100 Hz to 1000 Hz then written to contour file using salience based method viz. the state of art algorithm for predominant melody extraction.However the details description of the method is beyond the scope of this thesis,we encourage readers to refer [5] for detailed description of this method.

This provide us the pitch distribution which is further used to extract swara's features for classification.A pitch distribution provides the probability of occurrence of a pitch value over the segment duration.Next section explain the steps to fold the pitch distribution into 1200 bins.[3]

4.4 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 logarithmically 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 :
  $n = \left(240 \log_2 \frac{f}{256}\right) \mod 240$
- 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.

4.5 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

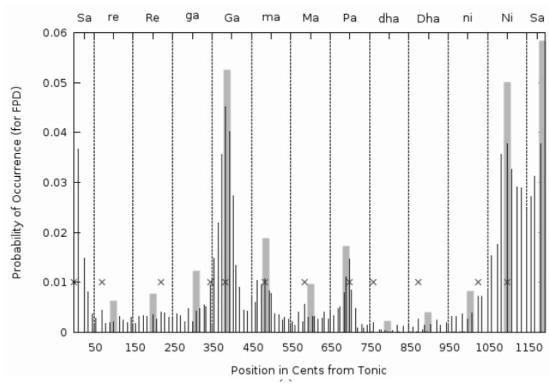

Figure 4.2: Folded Pitch distribution

- 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.

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

4.6 Experiment Details

Tables 4.3 and 4.4 depicts the swaras features of various raga extracted out from previous stage.These features are used to perform classification task. We have divided the classification task into four different experiments listed as below:

1. Same artist and Different raga(Bhairavi & Puria Dhanashri)
2. Same artist and Different raga(Bhoopali & Hamsadhwani)
3. Different artist and Different raga(Bhairavi & Puria Dhanashri)
4. Different artist and Different raga(Bhoopali & Hamsadhwani)

Table 4.3: Swara features for segments of Bhairav and Puria Dhanashri Raga by same artist

Raga/	Bhairav					Puria Dhanashri			
Swara	Peak	Mean	Sigma	Prob		Peak	Mean	Sigma	Prob
Sa	0	11	13	0.08		0	17	16	0.06
re	167	155	24	0.10		175	142	17	0.06
Re	273	213	17	0.05		217	208	22	0.07
ga	327	367	33	0.085		356	342	43	0.04
Ga	420	431	18	0.14		497	415	21	0.04
ma	561	542	32	0.10		524	521	11	0.05
Ma	619	611	55	0.18		618	654	23	0.08
Pa	780	775	41	0.02		732	721	31	0.10
dha	846	833	23	0.04		816	821	22	0.14
Dha	981	974	31	0.02		976	965	24	0.17
ni	1039	1031	25	0.09		1058	1049	12	0.10
Ni	1138	1117	35	0.08		1175	1163	25	0.08

Table 4.4: Swara features for segments of Bhoopali and Hamsadhwani Raga by same artist

Raga/	Bhoopali					Hamsadhwani			
Swara	Peak	Mean	Sigma	Prob		Peak	Mean	Sigma	Prob
Sa	0	14	16	0.07		0	12	14	0.06
re	178	164	19	0.20		148	153	28	0.12
Re	286	264	24	0.04		225	229	16	0.07
ga	345	336	26	0.05		378	385	35	0.06
Ga	485	425	47	0.06		497	415	21	0.04
ma	552	546	28	0.10		573	565	11	0.15
Ma	604	617	42	0.18		619	674	53	0.08
Pa	727	735	32	0.03		728	746	29	0.10
dha	883	872	24	0.04		853	823	34	0.09
Dha	990	956	27	0.02		976	965	24	0.35
ni	1021	1011	33	0.10		1043	1028	16	0.05
Ni	1167	1120	35	0.04		1175	1127	31	0.13

4.7 Classification

Nearest Neighbor 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.Suppose we are given two instances of swara features:

$$S_i = \{swara_{1_i}...swara_{12_i} = \{peak_{k_i}, mean_{k_i}, prob_{k_i}\}\}$$
$$S_j = \{swara_{1_j}...swara_{12_j} = \{peak_{k_j}, mean_{k_j}, prob_{k_j}\}\}$$

The distance between the two instances is computed as:

$$D(S_i, S_j) = \sum_{k=1}^{12} d(swara_{k_i}, swara_{k_j})$$

where

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

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

where KLdist(p,q) is kullbuck leiber distance between 2 probability distribution p,q and calculated as follws:

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

4.8 Results and Analysis

Table 4.5: Experiment results

Experiment Name	**No of test Segments of class1**	**No of test Segments of class2**	**Precision**	**Recall**
Same artist and Different raga (Bhairavi&Puria Dhanashri)	69	56	$(\frac{52}{62}, \frac{46}{63})$	$(\frac{52}{69}, \frac{46}{56})$
Same artist and Different raga (Bhoopali&Hamsadhwani)	33	29	$(\frac{24}{32}, \frac{31}{40})$	$(\frac{24}{33}, \frac{21}{29})$
Different artist and Different raga (Bhairavi&Puria Dhanashri)	71	62	$(\frac{55}{70}, \frac{47}{63})$	$(\frac{55}{71}, \frac{47}{62})$
Different artist and Different raga (Bhoopali&Hamsadhwani)	31	27	$(\frac{31}{42}, \frac{27}{36})$	$(\frac{31}{40}, \frac{27}{38})$

As shown in Table 4.5,Experiment 1 provide precision of $(\frac{52}{62}, \frac{46}{63})$ means that out of 62 segments are classify as raga bhairavi by the classifier out of which only 52 are correctly classified i.e 10 are misclassified segments with class raga bhairavi.

Apart from the experiment mention in Section 4.6 ,We perform another experiment when all raga's segments are present.The overall accuracy observed is roughly 67% i.e. only 67% of segment are classified correctly. Its also been observed that as we increase the no of Raga(K-class) the performance of model swiftly decreases.

Chapter 5

Conclusion and Future work

5.1 Conclusion

In this report we have looked into the problem of Raga Identification using Pitch class profile of raga.First we have made comparison between western and classic music and determine that tonic identification plays a crucial role in identification of raga.In chapter 2,we have studied various musicology concepts which act as foundation for understanding the approach of tonic identification using multi-pitch histogram.We reported 87.5% accuracy using this approach.Later we have folded pitch distribution into 1200 bins which represent FPD.thereafter we have mapped these bin into one octave which represent Pitch class distribution of raga(PCD).Classification is done using features extracted from PCD.We found that pitch class profile is sophisticated enough to recognize ragas with similar structure also.Another important observation is that as we increase the no of class the performance of model decreases.

5.2 Future work

This method provide us fairly sophisticated results however it completely ignores the temporal information present in the notes.Ragas usually contain repetitive Characteristic-phrases or motifs which provide a complementary information in identifying a raga.In next phase we are planning to incorporate an approach which allows us to learn a decision boundary in the combined space of Pitch-class profile and n-gram note distribution, where different ragas are linearly separable.Later We will define a kernels for pitch-class profile and n-gram distribution of notes that gives a measures of similarity between two music pieces.

Further we are planning to design a web based tool which enables user to detect raga of any audio waveform in real-time.Also it will help the research community to analyze the directly retrieve these features which reduce the implementation overhead.

Bibliography

[1] Music Technical Group, compmusic,xavier serra from the music technology group of the universitat pompeu fabra in barcelona (spain). http://http//www.compmusic.upf.edu/. Accessed: 14-October-2017.

[2] Shreyas Belle, Rushikesh Joshi, and Preeti Rao. Raga identification by using swara intonation. *Journal of ITC Sangeet Research Academy*, 23, 2009.

[3] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, O. Mayor, Gerard Roma, Justin Salamon, J. R. Zapata, and Xavier Serra. Essentia: an audio analysis library for music information retrieval. In *International Society for Music Information Retrieval Conference (ISMIR'13)*, pages 493–498, Curitiba, Brazil, 04/11/2013 2013.

[4] S. Gulati. A tonic identification approach for Indian art music. Master's thesis, Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain, 2012.

[5] Justin Salamon and Emilia Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.

[6] Justin Salamon, Emilia Gómez, and Jordi Bonada. Sinusoid extraction and salience function design for predominant melody estimation. In *Proc. 14th Int. Conf. on Digital Audio Effects (DAFX-11)*, pages 73–80, 2011.

[7] Wikipedia. Vadi (music) — wikipedia, the free encyclopedia, 2016. [Online; accessed 14-October-2017].

[8] Wikipedia. Raga — wikipedia, the free encyclopedia, 2017. [Online; accessed 14-October-2017].

[9] Wikipedia. Shruti (music) — wikipedia, the free encyclopedia, 2017. [Online; accessed 14-October-2017].