Raga Recognition using Intonation Information

MTP Phase-1

Saurav Shrivastava

IIT Bombay

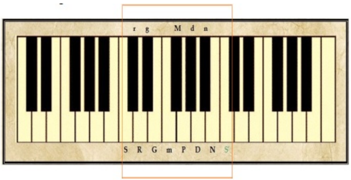
October 20, 2017

Table of Contents

1. Introduction to Music theory
2. Tonic Detection
3. Raga Recognition
4. Results and Analysis
5. Implementation
6. Future Work & Conclusions

Introduction to Music theory

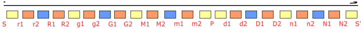
- Music consist of 3 major things sur(melody),taal(beat),laya(rhythm).



- **Swara(Notes):**There are total 12(7 Shudha + 4 komal+ 1tivra) musical notes.
- **Scale:** It 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.there are 2 types of scale in music:
    1. Equal tempered scale
    2. Natural tempered scale

Music Theory[CONTD.]

- **Shruti:**It is the smallest possible interval that can distinguish one sound from another as lower or higher pitch.



- **Raga:**It is a melodic framework within which the performer stays and improvise. Raga of performance captures its mood and emotions and it is used to express a feeling.A Raga is characterised by several attributes such as its Vaadi-Samvaadi, Aarohana-Avrohana and Pakad etc.

Raga	Swara sanchar	Aroha	Avaroha
Desh	SrGmPDnS'	SRmPNS'	S'nDPmGRG'NS
Kedar	SrmMPDnS'	SMP,MPDnDPS'	S'DnDP,MPDPm,SRS

Music Theory[CONTD.]

- **Tonic:** It is the base pitch selected by an artist and it serves as the foundation on which the artist builds his performance.
- **Consonance:** When two complex tones are played together and one or more of their partials(sinusoid with the lowest frequency which are multiples of the fundamental frequency) exactly coincide.
- **Drone:** Drone instrument acts as a reference of the music to a tonal background,reinforcing all the harmonic and melodic relationships.Tanpura is a 4 string instrument with following tuning:
  1. 1_{st} peg are tuned to fifth (pa) with respect to the tonic pitch, in the lower octave.
  2. 2_{nd}& 3_{rd} peg are tuned to the tonic pitch of the lead performer (Sa).
  3. 4_{th} peg is tuned to one octave below the tonic pitch (sa).

Tonic Identification

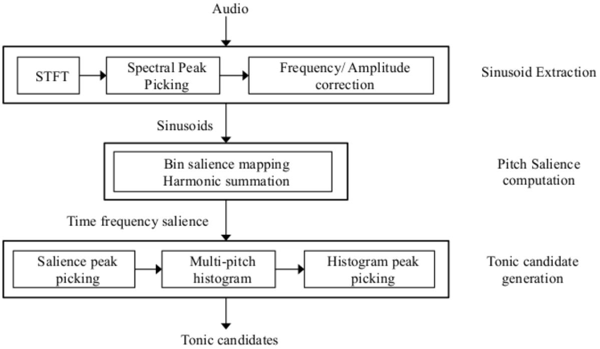
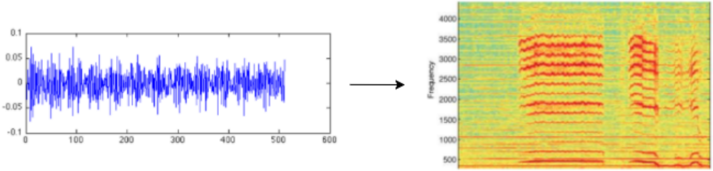


Figure: Block diagram of Tonic detection process

Sinusoid Extraction



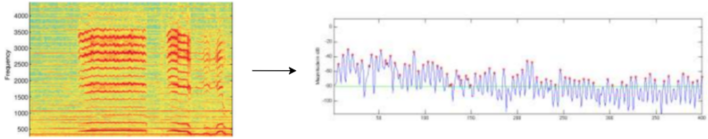
- Short-Time Fourier Transform (STFT) is given by:

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

- STFT Parameters:
  Hop size=11ms,Window length=46ms,Window type:Hamming

Sinusoid Extraction[CONTD.]

Spectral Peak Picking



- Energy threshold(T_s) is calculated as follows:

$$T_s = max(T_r, \alpha), \tag{2}$$
$$T_r = E_m + \beta$$

- Peak threshold:
Absolute threshold parameter(α) $= -70dB$
Relative threshold parameter(β) $= -40dB$

Salience Function computation

- Time-frequency representation indicating the salience of different pitches over time.
- Salience of a given frequency is computed as a weighted summation of energy found at all the integer multiples (harmonics) of that frequency.
- 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 :

$$S(j_f) = \sum_{h=1}^{N_h} \sum_{i=1}^{N_p} g(j, h, \hat{f}_i).(\hat{a}_i)^{\beta} \tag{3}$$

Salience Function computation[CONTD.]

- $g(j, h, \hat{f}_i)$ is the function that defines the weighting scheme defined as:

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

- δ is the distance in semitone between folded frequency $\hat{f}_i$ and center frequency of j^{th} bin and α is harmonic weighting parameter and β is a magnitude compression factor.
- We use $\alpha = 0.8, N_h = 20$ and $\beta = 1$ in the current implementation.
- This brings out the fundamental frequency component of the complex sinusoidal mixture, as it receives contributions from all its harmonics.

Tonic Candidate generation

- 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.

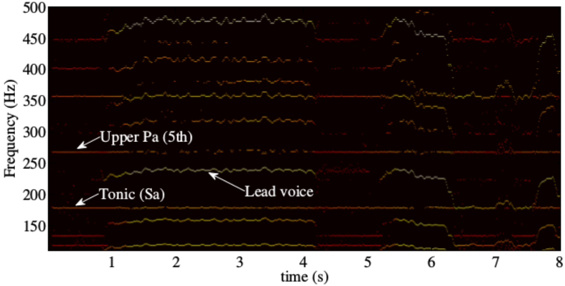


Figure: Pitch histogram

Salamon and Gómez [2012]

Tonic Candidate generation[CONTD.]

- We select the 10 most salient pitch values within the frequency range of 110-370 Hz from each frame.

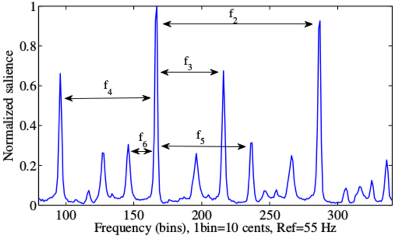


Figure: Multi-pitch histogram

Classification

- Set of features with candidate ($i = 1...10$) are as follows:
  1. **Pitch-Intervals(f_i) :** Distance in semitone between p_i and p_1
  2. **Amplitude features(a_i) :** Amplitude ratios of all the candidates with respect to the highest candidate.

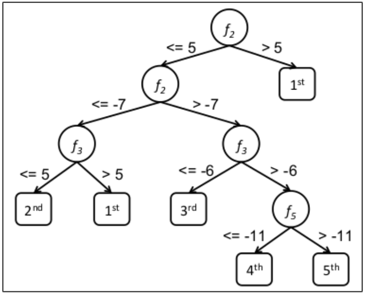


Figure: Decision tree for Tonic detection

- Accuracy reported is approximately 87.5% with 10Hz precision.

Raga Recognition

- **Non-Temporal Information:** Probabilities of all the frequency spread over one octave(FPD) or Probabilities of each swara across one octave(PCD) information is used for classification.

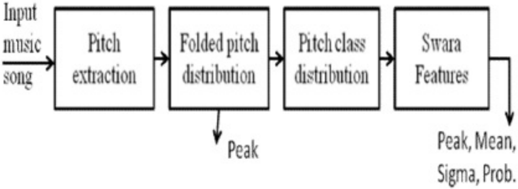


Figure: Process of Raga IdentificationBelle et al. [2009]

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

Pitch Extraction

1. Audio waveform consist of 8 segment of different length which is further divided into 30sec frame
2. Vocal performances from various artist in 4 ragas (Bhairavi,Puria Dhanashri,Bhoopali,Hamsadhwani) is used as training data. These performances have different segment length.
3. This data is pre-process by converting these performances into mono channel with sampling rate at 22050Hz.
4. Vocal pitch is extracted from each of these segment at different interval range from 100 Hz to 1000 Hz using salience based method Salamon and Gómez [2012] and then written to pitch contour file, this gives us pitch distribution. A pitch distribution provides the probability of occurrence of a pitch value over the segment duration

Folded Pitch distribution

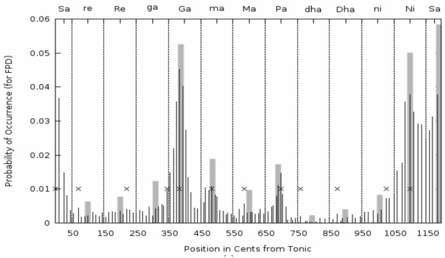
1. An arbitrary position (256Hz) was chosen for the initial bin of the FPD. Bins were logarithmically spaced at 5 cent intervals to give a total of 240 bins.
2. A pitch f in the pitch distribution was assigned to bin n in the FPD is given as :

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

3. 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 Nth bin were examined and the peak was found. The bin corresponding to the peak was considered to be the tonic bin.
4. The FPD was then rotated so that the tonic bin became the first bin.

Pitch class distribution

1. 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)
2. 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.



Features Extraction

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.

Note: Swara feature for a segment were represented by a 48(12 swara × 4 feature each) dimensional vector

Classification

1. We 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.

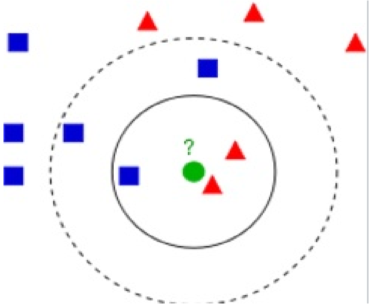


Figure: Nearest neighbour classifier

Wikipedia [2017]

Classification[CONTD.]

Distance between two swara features vector $swara_i$ and $swara_j$ is calculated as:

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

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

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

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

and KL(p||q) defined as

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

Results

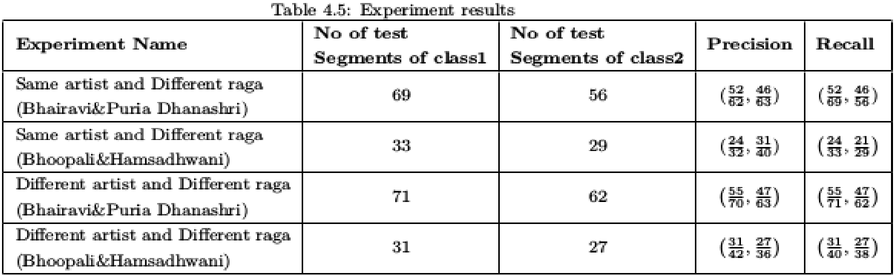


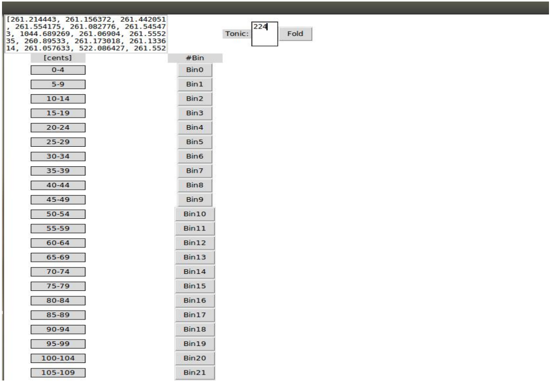
Table 4.5: Experiment results

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	$(\frac{31}{42}, \frac{27}{36})$	$(\frac{31}{40}, \frac{27}{38})$

$$\text{Precision} = \frac{tp}{tp + fp} \qquad \text{Recall} = \frac{tp}{tp + fn}$$

Implementation

1. Extraction of vocals
2. Extraction of pitch from vocals
3. Forming FPD from pitch



```
[261.214443, 261.156372, 261.442051
, 261.554175, 261.082776, 261.54547
3, 1044.689269, 261.06904, 261.5552
35, 260.89533, 261.173018, 261.1336
14, 261.057633, 522.086427, 261.552
```

Tonic: 224 Fold

[cents]	#Bin
0-4	Bin0
5-9	Bin1
10-14	Bin2
15-19	Bin3
20-24	Bin4
25-29	Bin5
30-34	Bin6
35-39	Bin7
40-44	Bin8
45-49	Bin9
50-54	Bin10
55-59	Bin11
60-64	Bin12
65-69	Bin13
70-74	Bin14
75-79	Bin15
80-84	Bin16
85-89	Bin17
90-94	Bin18
95-99	Bin19
100-104	Bin20
105-109	Bin21

Figure: Folded pitch distribution

Implementation[CONTD.]

Forming PCD using FPD

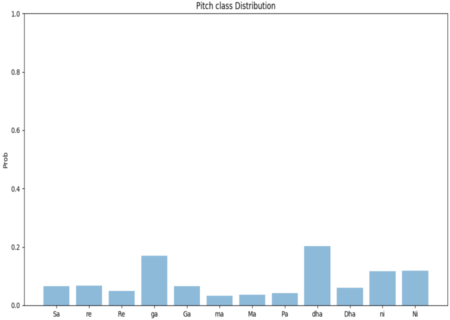


Figure: Pitch class distribution

Future Work

1. Pitch class profile method completely ignores the temporal information present in the notes.
2. Ragas usually contain repetitive Characteristic-phrases or motifs which provide a complementary information in identifying a raga.
3. 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.
4. 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.

Conclusions

1. Studied various musicological concepts such as shruti,raga etc.
2. Discussed about tonic identification using multi-pitch histogram and observe 87.5% accuracy with 10Hz precision.
3. Studied the swara features can be used to capture intonation information in raga
4. Also observed from previous discussion that only non temporal information is not sufficient .

References

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

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.

Wikipedia. K-nearest neighbors algorithm — wikipedia, the free encyclopedia, 2017. URL `https://en.wikipedia.org/w/index.php?title=K-nearest_neighbors_algorithm&oldid=797013672`. [Online; accessed 20-October-2017].

THANK YOU !