Melodic Style Detection in Hindustani Music

Amruta Vidwans
Prateek Verma
Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology
Bombay
OUTLINE

- Introduction
  - Objective
- Style Classification in Indian Classical Vocal Music
  - Literature Survey
  - Previous Work
  - Feature Description
  - Database and Listening Tests
  - Results and Extension to Turkish Music
- Style Classification In Flute Alap recordings
  - Database and Annotation
  - Signal Characteristics of Discriminatory Ornaments
  - Feature Design
- Conclusion and Future Work
- References
INTRODUCTION

Objective

- Melodic features to distinguish
  - Hindustani, Carnatic and Turkish music
  - Instrument playing styles

- Basis: Melody line alone suffices for listeners to reliably distinguish musical styles

- Applications: Extract important metadata automatically
Style Classification in Indian Classical Vocal Music
INTRODUCTION

Literature Survey – Similar work done

- **Timbral features**
  - Features based on timbre such as MFCC, delta-MFCC, and spectral features used by Parul et. al (ISMIR13) to distinguish Indian music genres using Adaboost, GMM etc.

- **Timbral + Rhythmic features**
  - Kini et. al. (NCC10) used these features to do genre classification of North Indian Classical Music into Quawali, Bhajan, Bollywood etc.
  - Liu et. al. (ICASSP09) used timbral, wavelet coefficients and rhythmic features to classify different styles viz. Arabic, Chinese, Japanese, Indian, Western classical
    - Emphasized that the diversity of Indian Classical Music is difficult to model

- **Timbral + Melodic features**
  - Salamon et al (ICASSP12) gave a large number of pitch based features in addition to timbral features which improved the performance.
Melody extraction
- Difficulty: Presence of multiple instruments along with voice
- Accuracy of the present state of the art automatic pitch trackers ~80%
- Semi automatic approach for pitch detection is used to achieve best possible performance

**Hindustani raga Jaijaiwanti by artiste Rashid Khan**

**Carnatic raga Dwijavanti by artiste R Vedavalli**

Department of Electrical Engineering, IIT Bombay
Localized contour shape-features

- Previous study used Steady Note (SN) and Gamak Measure (GM)
- A steady note: pitch segment of ‘N’ ms with standard deviation less than ‘J’ cents
- Data driven parameters gave highest accuracy (N=400ms J=20cents)
- SN: Normalized total duration of steady regions
- GM: ratio of number of non steady regions (1sec) having oscillations in 3-7.5Hz to the total number of regions
Primary Shape contour (PS)
- Contour typology proposed by C. Adams [6] for categorizing melodic shapes in 15 types
- Segments taken from silence to silence, assignment done using relation between Initial, Final, Highest and Lowest pitch values.

Contour types 12 and 13 in alap section in raga Hindolam by Carnatic artistes
(a) T. N. Sheshagopalan
(b) M. S. Subalaxmi
Feature Description

- **Distance of highest peak in unfolded histogram from Tonic (DistTonic)**
  - In unfolded histogram the Hindustani *alaps* are concentrated near the tonic, Carnatic *alap* pitch distribution is closer to the upper octave tonic.
  - Distance of the highest peak from the tonic in the unfolded histogram taken as feature.

![Histograms](https://via.placeholder.com/150)

- Rashid Khan raga Todi
- Sudha Raghunathan raga Subhapantuvarali
Melodic Transitions (MT)
- The overall progression of the melody can be characterized by this
- Haar wavelet basis function used to represent the melody
- Fifth level approximation is used
  - Lower level approximation captures very minute variation whereas higher level gives a coarse representation.
- Normalized size of upward jumps (>1 semitone) in concatenated pitch contour taken as feature.
Choice of *alap* section used for labeling a track

- Unmetered section always rendered in the start.

**Database Description**

- Widely performed raga pairs of same scale interval from Hindustani and Carnatic
- Ragas belonging to different scales chosen
- Renowned artists of various schools of music chosen

Total of 120 *alap* sections equally distributed across both the styles

<table>
<thead>
<tr>
<th>Scale</th>
<th>Hindustani <em>Raga</em></th>
<th>Carnatic <em>Raga</em></th>
<th>(No. of clips)</th>
<th>(No. of clips)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heptatonic</td>
<td>Todi (12)</td>
<td>Subhapanthuvarali (14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentatonic</td>
<td>Malkauns (18)</td>
<td>Hindolam (12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonatonic</td>
<td>Jaijaiwanti (10)</td>
<td>Dwijavanthy (14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Octatonic</td>
<td>Yaman and Yaman Kalyan (20)</td>
<td>Kalyani (20)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Department of Electrical Engineering, IIT Bombay
STYLE CLASSIFICATION IN INDIAN CLASSICAL VOCAL MUSIC
Listening Tests

- **Test Design**
  - Hypothesis: Melody alone is sufficient to carry out style distinction
  - Audio clips re-synthesized with constant timbre sound.
    - Removes bias towards artist identity, pronunciation, voice quality.
    - Volume dynamics retained (Sum of vocal harmonics).
  - **Interface Description**
    - User information in terms of training of subject as well as familiarity
    - Audios divided in 5 sets -- 12 clips of each Hindustani (H)-Carnatic (C) in each set.
    - Listening of 10 sec audio mandatory with option of pause, Skipping an audio not allowed
    - Decision label as H, C or NS (Not Sure) asked for each clip

<table>
<thead>
<tr>
<th>Category</th>
<th>Accuracy (no of participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>trained &lt;3yrs</td>
<td>74.6% (10)</td>
</tr>
<tr>
<td>trained 3-10yrs</td>
<td>79.7% (8)</td>
</tr>
<tr>
<td>trained &gt;10yrs</td>
<td>89.6% (2)</td>
</tr>
<tr>
<td>Amateur</td>
<td>75.2% (18)</td>
</tr>
<tr>
<td>Listener</td>
<td>77.5% (13)</td>
</tr>
<tr>
<td>Overall Accuracy</td>
<td>76.9% (51)</td>
</tr>
</tbody>
</table>
STYLE CLASSIFICATION IN INDIAN CLASSICAL VOCAL MUSIC
Automatic classification and feature selection

- Quadratic classifier used on feature sets
  - 5 fold CV on 5 different random partitions to avoid any raga bias
  - Exhaustive search on all possible feature combinations for feature selection
  - Separability of parameterized distribution taken into account for small dataset
  - Distribution of the Log Likelihood ratio (LLR) found for a feature set.
    \[ LLR = \ln \left( \frac{L(x/H)}{L(x/C)} \right) \]
  - F-ratio computed on distribution LLR as a confidence measure.
    \[ F-Ratio = \frac{(\mu_H - \mu_C)^2}{\sigma_H^2 + \sigma_C^2} \]
  - Highest accuracy achieved is 96% for SN, GM, DistTonic, and PS feature set.
Confusion matrix

(a) Confusion matrix for (a) listening tests (b) classifier output
The horizontal labels correspond to true labels.

<table>
<thead>
<tr>
<th>Obs.</th>
<th>C</th>
<th>H</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>True C</td>
<td>45</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>8</td>
<td>48</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs.</th>
<th>C</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>True C</td>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>3</td>
<td>57</td>
</tr>
</tbody>
</table>

Correlation with the listeners

- Subjective test label across all listeners decided by 60% threshold for labeling H, C or NS
- For objective measure region around 10% of LLR values around 0 taken as NS
- Cost value of -1, 1 or 0 is assigned according to the agreement between subjective and objective labels.
- Highest value of correlation was found to be 0.79 across all the feature combinations
STYLE CLASSIFICATION IN INDIAN CLASSICAL VOCAL MUSIC

Extension to Turkish Music

Hindustani *raga* Jaijaiwanti by artiste Rashid Khan

Carnatic *raga* Dwijavanti by artiste R Vedavalli

Turkish *makam* Nihavent by artiste Hafiz Kemal Bey

Department of Electrical Engineering, IIT Bombay
Experimental Setup and Results

- 60 Taksims considered for classification using the melodic features
- The features discussed before were applied to 3 classes.
- Highest accuracy of 81.6% was obtained using SN, GM, MT, DistTonic.

![Confusion matrix for (a) Listening test output (b) Classifier output](image)

- Most confusion occurred for T and C which was validate via subjective tests
- No category of primary shape was discriminating Turkish-Carnatic clips
Style Classification In Flute Alap recordings
INTRODUCTION
Example: Gayaki vs Tantrakari
**INTRODUCTION**

**Tantrakari vs Gayaki Style**

- **Gayaki vs Tantrakari styles**
  - Gayaki: Vocal characteristics incorporated into instrument play.
    - Singing voice is fluid, glides used often
  - Tantrakari: Instrumental characteristics (particularly plucked string)
    - Melodic Leaps eg. Discreteness
    - Elements synonymous with pluck adapted in Flute: Tonguing
    - Fast Stroke pattern while playing melody adapted in Flute: Fast Tonguing

- **Ornament Production**
  - Glide: Slowly lifting the finger - Rate controlling the glide shape
  - Vibrato: Periodic Pulsations in Air Flow
  - Fast Tonguing: Tongue movement
    - Using tongue to articulate notes differently in a melody
    - Synonymous with Sitar-rapid movements of right hand (**Tantrakari**) (Tantrakari)
  - Blow Stop:
    - Blowing with stops while rendering a note
    - Periodic blowing patterns (1-2 or 1-1-2) with same/different note
## STYLE CLASSIFICATION IN FLUTE ALAP AUDIOs
### Database and Annotation

<table>
<thead>
<tr>
<th>Artist</th>
<th>Raga</th>
<th>Duration (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hari Prasad Chaurasia</td>
<td>Jhinjhoti (T-Metadata)</td>
<td>18:54</td>
</tr>
<tr>
<td></td>
<td>Hansadhwani</td>
<td>1:50</td>
</tr>
<tr>
<td></td>
<td>Yaman</td>
<td>2:29</td>
</tr>
<tr>
<td></td>
<td>Sindh Bhairav</td>
<td>3:30</td>
</tr>
<tr>
<td>Rupak Kulkarni</td>
<td>Hemant (T-Metadata)</td>
<td>6:50</td>
</tr>
<tr>
<td>Ronu Majumdar</td>
<td>Vibhas</td>
<td>5:39</td>
</tr>
<tr>
<td>Pannalal Ghosh</td>
<td>Yaman</td>
<td>2:18</td>
</tr>
<tr>
<td></td>
<td>Shri</td>
<td>2:20</td>
</tr>
<tr>
<td></td>
<td>Pili</td>
<td>1:30</td>
</tr>
<tr>
<td>Nityanand Haldipur</td>
<td>ShuddhaBasant</td>
<td>2:46</td>
</tr>
</tbody>
</table>

Four Music Experts unanimously agreed on the two clips as tantrakari
STYLE CLASSIFICATION IN FLUTE ALAP AUDIOs

Annotation Rules

- Five categories used for annotation
  - Tantrakari
    - T1: Blowing same notes with stops
    - T2: Steady notes with abrupt movements
    - T3: Discreteness with silences
    - T4: Discreteness with rapid jumps
    - T5: Fast Tonguing with any pitch movement
  - Gayaki
    - G1: Oscillatory repetitive movements
    - G2: Non-Oscillatory repetitive movements
    - G3: Glides
    - G4: Non-specific continuous pitch
    - G5: Vibrato

- Final characteristic feature taken as the consensus of the feature agreed by all the musicians
### STYLE CLASSIFICATION IN FLUTE ALAP AUDIOs

#### Database and Annotation

<table>
<thead>
<tr>
<th>Artist</th>
<th>Raga</th>
<th>Duration (mins)</th>
<th>f/T/G (in s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hari Prasad Chaurasia</td>
<td>Jhinjhoti (T-Metadata)</td>
<td>18:54</td>
<td>352/465/267</td>
</tr>
<tr>
<td></td>
<td>Hansadhwani</td>
<td>1:50</td>
<td>43/14/33</td>
</tr>
<tr>
<td></td>
<td>Yaman</td>
<td>2:29</td>
<td>57/12/40</td>
</tr>
<tr>
<td></td>
<td>Sindh Bhairav</td>
<td>3:30</td>
<td>50/20/66</td>
</tr>
<tr>
<td>Rupak Kulkarni</td>
<td>Hemant (T-Metadata)</td>
<td>6:50</td>
<td>124/102/111</td>
</tr>
<tr>
<td>Ronu Majumdar</td>
<td>Vibhas</td>
<td>5:39</td>
<td>47/11/44</td>
</tr>
<tr>
<td>Pannalal Ghosh</td>
<td>Yaman</td>
<td>2:18</td>
<td>39/3/66</td>
</tr>
<tr>
<td></td>
<td>Shri</td>
<td>2:20</td>
<td>45/0/29</td>
</tr>
<tr>
<td></td>
<td>Pilu</td>
<td>1:30</td>
<td>133/35/105</td>
</tr>
<tr>
<td>Nityanand Haldipur</td>
<td>ShuddhaBasant</td>
<td>2:46</td>
<td>94/0/41</td>
</tr>
</tbody>
</table>

- G: Gayaki
  - T: Tantrakari
  - f: Neither Tantrakari or Gayaki

- Alap-Jod-Jhala composition derived from Sitar. (It mainly contains Tantrakari elements)

- Two different style artists and their disciple chosen
## Presence of Tantrakari and Gayaki elements

<table>
<thead>
<tr>
<th>Artist</th>
<th>Raga</th>
<th>Vibrato</th>
<th>Fast Tounging</th>
<th>Blow - Stop</th>
<th>T3/T4</th>
<th>G3/G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hari Prasad Chaurasia</td>
<td>Jhinjhoti(T)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Hansadhwani</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Yaman</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Sindh Bhairav</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rupak Kulkarni</td>
<td>Hemant(T)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ronu Majumdar</td>
<td>Vibhas</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pannalal Ghosh</td>
<td>Yaman</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Shri</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Pilu</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Nityanand Haldipur</td>
<td>ShuddhaBasant</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>
STYLE CLASSIFICATION IN FLUTE ALAP AUDIOS
Production and Acoustic Characteristics

- Blow-Stop & Fast Tonguing – Two of the most distinctive features
  - Energy Contour different due to difference in articulation
  - Higher rate of energy variation due to use of tongue.
  - Pitch movements do not influence energy movements.
  - Intensity of Tanpura almost constant
Features capturing acoustic characteristics of Fast Tonguing

- **Sub-band Peak Ratio (SPR)** = \[
\frac{\max (|E(k)|_{k_{10\text{Hz}} \leq k \leq k_{20\text{Hz}}})}{\max (|E(k)|_{k_{20\text{Hz}} \leq k \leq k_{30\text{Hz}}})}
\]

- **Sub-band FFT Ratio (SFR)** = \[
\frac{\sum_{k_{10\text{Hz}}}^{k_{20\text{Hz}}} |E(k)|}{\sum_{k_{20\text{Hz}}}^{k_{30\text{Hz}}} |E(k)|}
\]

- 2s segments of data across the two styles taken as data point
- Ground Truth obtained using manual annotation

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic</td>
<td>74.07</td>
</tr>
<tr>
<td>Linear</td>
<td>73.15</td>
</tr>
</tbody>
</table>
Conclusion

- Successfully proposed melodic features to distinguish vocal music styles
- Importance of energy based features brought out for flute style classification.
- Automatic classification results given with degree of separability

Future Work

- Extension of database to non-alap recordings in flute style classification study.
- Addition of new features to distinguish C and T to improve the classification accuracy
References

8] C. Clemments, “Pannalal Ghosh and the Bānsurī in the Twentieth Century”, City University of New York, New York City, September 2010