Representation and Segmentation of Melodies in Indian Art Music

3rd CompMusic Workshop, IIT-Madras, Chennai, India
13th Dec, 2013

Sankalp Gulati
Supervisor: Prof. Xavier Serra
PhD Candidate, MTG-UPF, Barcelona, Spain
Block diagram of the proposed semi-supervised methodology for melodic motif discovery
Block diagram of the proposed semi-supervised methodology for melodic motif discovery
Block diagram of the proposed semi-supervised methodology for melodic motif discovery
**Tonic Identification: JNMR article**

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>Feature Distribution</th>
<th>Tonic Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS (Sengupta et al., 2005)</td>
<td>Pitch (Datta, 1996)</td>
<td>N/A</td>
<td>Error minimization</td>
</tr>
<tr>
<td>RH1/2 (Ranjani et al., 2011)</td>
<td>Pitch (Boersma &amp; Weenink, 2001)</td>
<td>Parzen-window-based PDE$^1$</td>
<td>GMM fitting</td>
</tr>
<tr>
<td>JS (Salamon et al., 2012)</td>
<td>Multi-pitch salience (Salamon, Gómez, &amp; Bonada, 2011)</td>
<td>Multi-pitch histogram</td>
<td>Decision tree</td>
</tr>
<tr>
<td>SG (Gulati et al., 2012)</td>
<td>Multi-pitch salience (Salamon et al., 2011)</td>
<td>Multi-pitch histogram</td>
<td>Decision tree</td>
</tr>
<tr>
<td></td>
<td>Predominant melody (Salamon &amp; Gómez, 2012)</td>
<td>Pitch histogram</td>
<td>Decision tree</td>
</tr>
<tr>
<td>AB1 (Bellur et al., 2012)</td>
<td>Pitch (De Cheveigné &amp; Kawahara, 2002)</td>
<td>GD$^2$ histogram</td>
<td>Highest peak</td>
</tr>
<tr>
<td>AB2 (Bellur et al., 2012)</td>
<td>Pitch (De Cheveigné &amp; Kawahara, 2002)</td>
<td>GD histogram</td>
<td>Template matching</td>
</tr>
<tr>
<td>AB3 (Bellur et al., 2012)</td>
<td>Pitch (De Cheveigné &amp; Kawahara, 2002)</td>
<td>GD histogram</td>
<td>Highest peak</td>
</tr>
</tbody>
</table>

Table 1: Summary of existing tonic identification approaches.


Tonic Identification: Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Avg. length (min)</th>
<th>#Excerpts</th>
<th>Hi.(%)</th>
<th>Ca.(%)</th>
<th>Voc. (M/F)(%)</th>
<th>Inst. (%)</th>
<th>#Usong</th>
<th>#Uartists</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM1</td>
<td>3</td>
<td>271</td>
<td>41</td>
<td>59</td>
<td>0</td>
<td>100</td>
<td>169</td>
<td>33</td>
</tr>
<tr>
<td>CM2</td>
<td>3</td>
<td>935</td>
<td>45</td>
<td>55</td>
<td>100 (68 / 32)</td>
<td>0</td>
<td>547</td>
<td>81</td>
</tr>
<tr>
<td>CM3</td>
<td>14.8</td>
<td>428</td>
<td>45</td>
<td>55</td>
<td>100 (72 / 28)</td>
<td>0</td>
<td>428</td>
<td>71</td>
</tr>
<tr>
<td>IITM1</td>
<td>144.6</td>
<td>38</td>
<td>0</td>
<td>100</td>
<td>89 (79 / 21)</td>
<td>11</td>
<td>N/A</td>
<td>22</td>
</tr>
<tr>
<td>IITM2</td>
<td>12.3</td>
<td>472</td>
<td>0</td>
<td>100</td>
<td>92 (77 / 23)</td>
<td>8</td>
<td>472</td>
<td>22</td>
</tr>
<tr>
<td>IISC1</td>
<td>7.4</td>
<td>55</td>
<td>0</td>
<td>100</td>
<td>100 (80 / 20)</td>
<td>0</td>
<td>55</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2: Dataset summary, including average excerpt length (Avg. length), number of excerpts (#Excerpts), percentage of Hindustani music (Hi), Carnatic music (Ca), vocal excerpts (Voc.), instrumental excerpts (Inst.), number of unique songs (#Usong) and number of unique artists (#Uartists) in each dataset. For vocal excerpts we also provide the breakdown into male (M) and female (F) singers. Percentage (%) values are rounded to the nearest integer.

## Tonic Identification: Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>CM1</th>
<th>CM2</th>
<th>CM3</th>
<th>IISCB1</th>
<th>IITM1</th>
<th>IITM2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
<td>TPC</td>
<td>TP</td>
<td>TPC</td>
<td>TP</td>
<td>TPC</td>
</tr>
<tr>
<td>JS</td>
<td>-</td>
<td>88.9</td>
<td>87.4</td>
<td>90.1</td>
<td>75.6</td>
<td>77.5</td>
</tr>
<tr>
<td>SG</td>
<td>-</td>
<td>92.2</td>
<td>87.8</td>
<td>90.9</td>
<td>79.8</td>
<td>85.3</td>
</tr>
<tr>
<td>RH1</td>
<td>-</td>
<td>81.4</td>
<td>69.6</td>
<td>84.9</td>
<td>81.8</td>
<td>83.6</td>
</tr>
<tr>
<td>RH2</td>
<td>-</td>
<td>63.2</td>
<td>65.7</td>
<td>78.2</td>
<td>83.6</td>
<td>83.6</td>
</tr>
<tr>
<td>AB1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AB2</td>
<td>-</td>
<td>88.9</td>
<td>74.5</td>
<td>82.9</td>
<td>72.7</td>
<td>76.4</td>
</tr>
<tr>
<td>AB3</td>
<td>-</td>
<td>86</td>
<td>61.1</td>
<td>80.5</td>
<td>72.7</td>
<td>72.7</td>
</tr>
</tbody>
</table>

Table 3: Accuracies for tonic pitch (TP %) and tonic pitch-class (TPC %) identification by seven methods on six different datasets using only audio data. The best accuracy obtained for each dataset is highlighted using bold text. The dashed horizontal line divides the methods based on supervised learning (JS and SG) and those based on expert knowledge (RH1, RH2, AB1, AB2 and AB3). TP column for CM1 is marked as ‘-’, because it consists of only instrumental excerpts for which we not evaluate tonic pitch accuracy.

Tonic Identification: Analysis

Performance as a function of length

Performance as a function of category

Error type as a function of dataset

Block diagram of the proposed semi-supervised methodology
Melody Extraction and Representation

- Melody (lead artist/voice)
  - Pitch (Fundamental Frequency, F0)
  - Timbre
  - Loudness


Original Audio
Predominant Voice
F0 of the Predominant Voice
Loudness and Timbral Facets
Loudness and Timbre

Evaluation: Essentia Pitch Extraction

- Predominant melody extraction (F0)
  - 6 Hindustani music pieces ~45 mins

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>GT err. (%)</th>
<th>V/UV err. (%)</th>
<th>Oct. err. (%)</th>
<th>Oth. err. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Salamon &amp; Gómez, 2012)</td>
<td>9.3</td>
<td>2.7</td>
<td>1.7</td>
<td>4.9</td>
</tr>
</tbody>
</table>


Loudness and Timbre: Motif Detection

- Motif similarity: DTW
  - mnDP (#72 instances)
  - DnDP (#56 instances)

- # Combinations:
  - Postives: 4096
  - Negative: 4032

- Dataset: IIT Bombay
  - Raga: Alhaiya Bilawal
  - # Performances: 5
  - # Artists: 4

ROC: Pitch
ROC: Timbre
ROC: Loudness
Block diagram of the proposed semi-supervised methodology
Melody Segmentation

- Task, context and music style/tradition dependent
- Do we need it? (motif discovery)
  - At what stage of processing?

- Melodic motifs <> nyas svar

- Melodic segmentation: estimating boundaries of nyas svars

Nyas Svar Segmentation
Nyas Svar Segmentation
Nyas Svar Segmentation
Nyas Svar Segmentation

S. Gulati, J. Serrà, K. K. Ganguli, and Xavier Serra, “LANDMARK DETECTION IN HINDUSTANI MUSIC MELODIES” Submitted to ICASSP 2013
Tonic Identification
Rhythm Analysis

Melodic Feature Extraction

Melody Segmentation
Melody Similarity

Pattern Extraction
Redundancy reduction

Melodic motivic discovery

Melodic motives representation
Melodic motives similarity graph

Rāg characterization & Music similarity measures

Data extracted knowledge

• Rhythmic Variations (overall timing, non-linear time variations)
• Extent of pitch variations
• Melodic segmentation information

Expert knowledge

• Nyas swaras
• Motif transpositions
• Salient melodic movements

Listening tests

Dunya integration

Evaluation

Block diagram of the proposed semi-supervised methodology
Nyas Identification

Predominant melody extraction \rightarrow Tonic identification

Histogram computation \rightarrow Svar identification \rightarrow Segmentation

Local Feature extraction \rightarrow Contextual Feature extraction \rightarrow Local + Contextual Feature extraction

Segment classification \rightarrow Segment fusion

Nyās svars

S. Gulati, J. Serrà, K. K. Ganguli, and Xavier Serra, “LANDMARK DETECTION IN HINDUSTANI MUSIC MELODIES” Submitted to ICASSP 2013
Nyas Identification

- Local Features (#9)
  - Segment length
  - Mean, variance of pitch values
  - Mean, variance of the differences in adjacent peak locations in pitch sequence
  - Mean, variance of peak amplitudes of pitch sequence
  - Temporal centroid
  - Flatness measure (output of segmentation method)
Nyas Identification

- Local Features (#3)
  - Segment length
  - Mean, variance of pitch values
  - Mean, variance of the differences in adjacent peak locations in pitch sequence
  - Mean, variance of peak amplitudes of pitch sequence
  - Temporal centroid
  - Flatness measure (output of segmentation method)
Nyas Identification

- Contextual Features (#24)
  - Segment length / (longest segment length in breath phrase)
  - Segment length / (length of the breath phrase)
  - Segment length / (length of the previous segment)
  - Segment length / (length of the following segment)
  - Duration between the ending and succeeding silence
  - Duration between the starting and preceding silence
  - All local features of adjacent segments
Nyas Identification

- Contextual Features (#15)
  - Segment length / (longest segment length in breath phrase)
  - Segment length / (length of the breath phrase)
  - Segment length / (length of the previous segment)
  - Segment length / (length of the following segment)
  - Duration between the ending and succeeding silence
  - Duration between the starting and preceding silence
  - All local features of preceding segments
Nyas Identification: Baselines

- Segmentation
  - Piece-wise linear segmentation

- Classification
  - DTW + kNN classification

- Several random baselines
Nyas Identification: Dataset

- 20 recordings of Hindustani music
  - 15 Polyphonic: CompMusic collection
  - 5 Monophonic: Kaustuv’s recordings
- Unique artists: 8
- Unique Rāgs: 16
- Number of nyās segments: 1257
- Duration of nyās segments
  - Range: 150 ms – 16.7 s
  - Mean: 2.4 s
  - Median: 1.4 s
### Nyas Identification: Results

Local features with Proposed Segmentation

<table>
<thead>
<tr>
<th></th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>0.356</td>
<td>0.407</td>
<td>0.447</td>
<td>0.248</td>
<td>0.449</td>
<td>0.453</td>
</tr>
<tr>
<td>C</td>
<td>0.284</td>
<td>0.394</td>
<td>0.387</td>
<td>0.383</td>
<td>0.389</td>
<td>0.406</td>
</tr>
<tr>
<td>L+C</td>
<td>0.289</td>
<td>0.414</td>
<td>0.426</td>
<td>0.409</td>
<td>0.432</td>
<td>0.437</td>
</tr>
<tr>
<td>L</td>
<td>0.524</td>
<td>0.672</td>
<td>0.719</td>
<td>0.491</td>
<td>0.736</td>
<td>0.749</td>
</tr>
<tr>
<td>C</td>
<td>0.436</td>
<td>0.629</td>
<td>0.615</td>
<td>0.641</td>
<td>0.621</td>
<td>0.673</td>
</tr>
<tr>
<td>L+C</td>
<td>0.446</td>
<td>0.682</td>
<td>0.708</td>
<td>0.591</td>
<td>0.725</td>
<td>0.735</td>
</tr>
</tbody>
</table>

Table 1. F-scores for nyās boundary annotations using PLS method (upper half) and the proposed segmentation method (lower half). The best random baseline F-score is 0.184 obtained using RB2.

S. Gulati, J. Serrà, K. K. Ganguli, and Xavier Serra, “LANDMARK DETECTION IN HINDUSTANI MUSIC MELODIES” Submitted to ICASSP 2013
Nyas Identification: Results

Local features with Proposed Segmentation

<table>
<thead>
<tr>
<th></th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>0.553</td>
<td>0.685</td>
<td>0.723</td>
<td>0.621</td>
<td>0.727</td>
<td>0.722</td>
</tr>
<tr>
<td>C</td>
<td>0.251</td>
<td>0.639</td>
<td>0.631</td>
<td>0.690</td>
<td>0.688</td>
<td>0.674</td>
</tr>
<tr>
<td>L+C</td>
<td>0.389</td>
<td>0.694</td>
<td>0.693</td>
<td>0.708</td>
<td>0.722</td>
<td>0.706</td>
</tr>
<tr>
<td>L</td>
<td>0.546</td>
<td>0.708</td>
<td>0.754</td>
<td>0.714</td>
<td>0.749</td>
<td>0.758</td>
</tr>
<tr>
<td>C</td>
<td>0.281</td>
<td>0.671</td>
<td>0.611</td>
<td>0.697</td>
<td>0.689</td>
<td>0.697</td>
</tr>
<tr>
<td>L+C</td>
<td>0.332</td>
<td>0.672</td>
<td>0.710</td>
<td>0.730</td>
<td>0.743</td>
<td>0.731</td>
</tr>
</tbody>
</table>

Table 2. F-scores for nyās and non-nyās label annotations using PLS method (upper half) and the proposed segmentation method (lower half). The best random baseline F-score is 0.153 obtained using RB2.

S. Gulati, J. Serrà, K. K. Ganguli, and Xavier Serra, “LANDMARK DETECTION IN HINDUSTANI MUSIC MELODIES” Submitted to ICASSP 2013
Block diagram of the proposed semi-supervised methodology for melodic motif discovery
Melodic Motif Discovery: Brute-force
Block diagram of the proposed semi-supervised methodology for melodic motif discovery