Landmark Detection in Hindustani Music Melodies

Sankalp Gulati\textsuperscript{1}, Joan Serrà\textsuperscript{2}, Kaustuv K. Ganguli\textsuperscript{3} and Xavier Serra\textsuperscript{1}

sankalp.gulati@upf.edu, jserra@iiia.csic.es, kaustuvkanti@ee.iitb.ac.in, xavier.serra@upf.edu

\textsuperscript{1}Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
\textsuperscript{2}Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain
\textsuperscript{3}Indian Institute of Technology Bombay, Mumbai, India

SMC-ICMC 2014, Athens, Greece
Indian Art Music

- Hindustani music (North Indian music)

- Carnatic music
Melodies in Hindustani Music

- Rāg: melodic framework of Indian art music
Melodies in Hindustani Music

- **Rāg**: melodic framework of Indian art music

<table>
<thead>
<tr>
<th>S</th>
<th>r</th>
<th>R</th>
<th>g</th>
<th>G</th>
<th>m</th>
<th>M</th>
<th>P</th>
<th>d</th>
<th>D</th>
<th>n</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do</td>
<td>Re</td>
<td>Mi</td>
<td>Fa</td>
<td>Sol</td>
<td>La</td>
<td>Ti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Melodies in Hindustani Music

- Rāg: melodic framework of Indian art music

---

**Bhairavi Thaat**

<table>
<thead>
<tr>
<th>S</th>
<th>r</th>
<th>g</th>
<th>m</th>
<th>P</th>
<th>d</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do</td>
<td>Re</td>
<td>Mi</td>
<td>Fa</td>
<td>Sol</td>
<td>La</td>
<td>Ti</td>
</tr>
</tbody>
</table>

**Svaras**
Melodies in Hindustani Music

- Rāg: melodic framework of Indian art music

Nyās translates to home/residence

*Nyās Svar (Rāg Bilaskhani todi)

<table>
<thead>
<tr>
<th>S</th>
<th>r*</th>
<th>g*</th>
<th>m*</th>
<th>P</th>
<th>d*</th>
<th>n</th>
</tr>
</thead>
</table>

Bhairavi Thaat

<table>
<thead>
<tr>
<th>S</th>
<th>r</th>
<th>g</th>
<th>m</th>
<th>P</th>
<th>d</th>
<th>n</th>
</tr>
</thead>
</table>

Svaras

<table>
<thead>
<tr>
<th>S</th>
<th>r</th>
<th>R</th>
<th>g</th>
<th>G</th>
<th>m</th>
<th>M</th>
<th>P</th>
<th>d</th>
<th>D</th>
<th>n</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do</td>
<td>Re</td>
<td>Mi</td>
<td>Fa</td>
<td>Sol</td>
<td>La</td>
<td>Ti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Melodic Landmark: Nyās Occurrences

- Example

Goal and Motivation

- Methodology for detecting nyās occurrences
Goal and Motivation

- Methodology for detecting nyās occurrences
Goal and Motivation

- Methodology for detecting nyās occurrences

- Motivation
  - Melodic motif discovery [Ross and Rao 2012]
  - Melodic segmentation
  - Music transcription

Methodology: Block Diagram

Block diagram of the proposed methodology

- Audio
  - Predominant pitch estimation
  - Pred. pitch est. and representation
  - Tonic identification
  - Segmentation
    - Segmentation
    - Feature extraction
    - Local
      - Local
      - Segment classification and fusion
      - Segment classification
    - Contextual
    - Local + Contextual
    - Svar identification
    - Svar
    - Nyās svars
Methodology: Pred. Pitch Estimation

Audio

Predominant pitch estimation
Tonic identification

Histogram computation
Svar identification
Segmentation

Local
Contextual
Local + Contextual

Segment classification
Segment fusion

Nyās svars

Pred. pitch est. and representation

Segmentation
Feature extraction
Segment classification and fusion
Methodology: Segmentation

Audio

Predominant pitch estimation

Tonic identification

Histogram computation

Svar identification

Segmentation

Local

Contextual

Local + Contextual

Segment classification

Segment fusion

Nyās svars

Pred. pitch est. and representation

Segmentation

Feature extraction

Segment classification and fusion
Methodology: Feature Extraction

Audio

Predominant pitch estimation → Tonic identification

Pred. pitch est. and representation

Histogram computation → Svar identification → Segmentation

Segmentation

Local Feature extraction → Contextual → Local + Contextual

Segment classification → Segment fusion

Segment classification and fusion

Nyās svars

0.12, 0.34, 0.59, 0.23, 0.54

0.21, 0.24, 0.54, 0.54, 0.42

0.32, 0.23, 0.34, 0.41, 0.63

0.66, 0.98, 0.74, 0.33, 0.12

0.90, 0.42, 0.14, 0.83, 0.76
Methodology: Segment Classification

Audio

Predominant pitch estimation
Tonic identification

Histogram computation
Svar identification
Segmentation

Local
Contextual
Local + Contextual

Segment classification
Segment fusion

Segment classification and fusion

Nyās svars
Methodology: Segment Classification

Audio

Predominant pitch estimation → Tonic identification

Histogram computation → Svar identification → Segmentation

Local + Contextual

Segment classification

Segment fusion

Nyās svars

Pred. pitch est. and representation

Segmentation

Feature extraction

Segment classification and fusion
Pred. Pitch Estimation and Representation

- Predominant pitch estimation
  - Method by Salamon and Gómez (2012)
  - Favorable results in MIREX’11

- Tonic Normalization
  - Pitch values converted from Hertz to Cents
  - Multi-pitch approach by Gulati et al. (2014)


Melody Segmentation

- Baseline: Piecewise linear segmentation (PLS)

Feature Extraction
Feature Extraction

- Local (9 features)
Feature Extraction

- Local (9 features)
- Contextual (24 features)
Feature Extraction

- Local (9 features)
- Contextual (24 features)
- Local + Contextual (33 features)
Feature Extraction: Local

- Segment Length
Feature Extraction: Local

- Segment Length
- $\mu$ and $\sigma$ of pitch values
Feature Extraction: Local

- Segment Length
- \( \mu \) and \( \sigma \) of pitch values
- \( \mu \) and \( \sigma \) of difference in adjacent peaks and valley locations
Feature Extraction: Local

- Segment Length
- \( \mu \) and \( \sigma \) of pitch values
- \( \mu \) and \( \sigma \) of difference in adjacent peaks and valley locations
- \( \mu \) and \( \sigma \) of the peak and valley amplitudes
Feature Extraction: Local

- Segment Length
- $\mu$ and $\sigma$ of pitch values
- $\mu$ and $\sigma$ of difference in adjacent peaks and valley locations
- $\mu$ and $\sigma$ of the peak and valley amplitudes
- Temporal centroid (length normalized)
Feature Extraction: Local

- Segment Length
- $\mu$ and $\sigma$ of pitch values
- $\mu$ and $\sigma$ of difference in adjacent peaks and valley locations
- $\mu$ and $\sigma$ of the peak and valley amplitudes
- Temporal centroid (length normalized)
- Binary flatness measure

Flat or non-flat
Feature Extraction: Contextual

- 4 different normalized segment lengths
Feature Extraction: Contextual

- 4 different normalized segment lengths
Feature Extraction: Contextual

- 4 different normalized segment lengths
Feature Extraction: Contextual

- 4 different normalized segment lengths
Feature Extraction: Contextual

- 4 different normalized segment lengths
- Time difference from the succeeding and preceding breath pauses
Feature Extraction: Contextual

- 4 different normalized segment lengths
- Time difference from the succeeding and proceeding unvoiced regions
- Local features of neighboring segments ($9 \times 2 = 18$)
Segment Classification

- **Class: Nyās and Non-nyās**
- **Classifiers:**
  - Trees \((\text{min\_sample\_split}=10)\)
  - K nearest neighbors \((n\_neighbors=5)\)
  - Naive bayes \((\text{fit\_prior}=\text{False})\)
  - Logistic regression \((\text{class\_weight}='\text{auto}')\)
  - Support vector machines (RBF) \((\text{class\_weight}='\text{auto}')\)
- **Testing methodology**
  - Cross-fold validation
- **Software: Scikit-learn, version 0.14.1**

Evaluation: Dataset

- **Audio:**
  - Number of recordings: 20 Ālap vocal pieces
  - Duration of recordings: 1.5 hours
  - Number of artists: 8
  - Number of rāgs: 16
  - Type of audio: 15 polyphonic commercial recordings, 5 in-house monophonic recordings**

- **Annotations:** Musician with > 15 years of training

- **Statistics:**
  - 1257 nyās segments
  - 150 ms to 16.7 s
  - mean 2.46 s, median 1.47 s.

**Openly available under CC license in freesound.org**
Evaluation: Measures

- Boundary annotations (F-scores)
  - Hit rate
  - Allowed deviation: 100 ms
- Label annotations (F-scores):
  - Pairwise frame clustering method [Levy and Sandler 2008]
- Statistical significance: Mann-Whitney U test (p=0.05)
- Multiple comparison: Holm-Bonferroni method

Evaluation: Baseline Approach

- DTW based kNN classification (k=5)
  - Frequently used for time series classification

- Random baselines
  - Randomly planting boundaries
  - Evenly planting boundaries at every 100 ms
  - Ground truth boundaries, randomly assign class labels


Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th>Feat.</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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th></th>
<th>Feat.</th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<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></td>
<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></td>
<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>B</td>
<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></td>
<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></td>
<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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th>Feature Set</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><strong>0.524</strong></td>
<td>0.672</td>
<td><strong>0.719</strong></td>
<td>0.491</td>
<td><strong>0.736</strong></td>
<td><strong>0.749</strong></td>
</tr>
<tr>
<td>C</td>
<td>0.436</td>
<td>0.629</td>
<td>0.615</td>
<td><strong>0.641</strong></td>
<td>0.621</td>
<td>0.673</td>
</tr>
<tr>
<td>L+C</td>
<td>0.446</td>
<td><strong>0.682</strong></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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th></th>
<th>Feat.</th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<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></td>
<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></td>
<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>B</td>
<td>L</td>
<td>0.524</td>
<td>0.672</td>
<td><strong>0.719</strong></td>
<td>0.491</td>
<td><strong>0.736</strong></td>
<td><strong>0.749</strong></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.436</td>
<td>0.629</td>
<td>0.615</td>
<td><strong>0.641</strong></td>
<td>0.621</td>
<td>0.673</td>
</tr>
<tr>
<td></td>
<td>L+C</td>
<td>0.446</td>
<td><strong>0.682</strong></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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th></th>
<th>Feat.</th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<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></td>
<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></td>
<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>B</td>
<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></td>
<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></td>
<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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Boundary Annotation

<table>
<thead>
<tr>
<th></th>
<th>Feat.</th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<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>A</td>
<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></td>
<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></td>
<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>B</td>
<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></td>
<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 detection using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. F-score for the random baseline obtained using RB2 is 0.184.
Results: Nyās Label Annotation

<table>
<thead>
<tr>
<th></th>
<th>Feat.</th>
<th>DTW</th>
<th>Tree</th>
<th>KNN</th>
<th>NB</th>
<th>LR</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<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></td>
<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></td>
<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>B</td>
<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></td>
<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></td>
<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 task using PLS method (A) and the proposed segmentation method (B). Results are shown for different classifiers (Tree, KNN, NB, LR, SVM) and local (L), contextual (C) and local together with contextual (L+C) features. DTW is the baseline method used for comparison. The best random baseline F-score is 0.153 obtained using RB2.
Conclusions and Future work

- Proposed segmentation better than PLS method
- Proposed methodology better than standard DTW based kNN classification
- Local features yield highest accuracy
- Contextual features are also important (maybe not complementary to local features)

Future work
- Perform similar analysis on Bandish performances
- Incorporate Rāga specific knowledge
Landmark Detection in Hindustani Music Melodies

Sankalp Gulati¹, Joan Serrà², Kaustuv K. Ganguli³ and Xavier Serra¹

sankalp.gulati@upf.edu, jserra@iiiia.csic.es, kaustuvkanti@ee.iitb.ac.in, xavier.serra@upf.edu

¹Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
²Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Barcelona, Spain
³Indian Institute of Technology Bombay, Mumbai, India

SMC-ICMC 2014, Athens, Greece