Landmark Detection in Hindustani Music Melodies

Sankalp Gulati[1], Joan Serrà[2], Kaustuv K. Ganguli[3] and Xavier Serra[1]

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

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

SMC-ICMC 2014, Athens, Greece

upf. Music Technology Group

compmusic

INSTITUT D'INVESTIGACIÓ EN INTEL.LIGÈNCIA ARTIFICIAL - IIIA

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

Svaras

S	r	R	g	G	m	M	P	d	D	n	N
Do		Re		Mi	Fa		Sol		La		Ti

Melodies in Hindustani Music

- Rāg: melodic framework of Indian art music

Bhairavi Thaat

S	**r**		**g**		**m**		**P**	**d**		**n**	

Svaras

S	**r**	**R**	**g**	**G**	**m**	**M**	**P**	**d**	**D**	**n**	**N**
Do		Re		Mi	Fa		Sol		La		Ti

Melodies in Hindustani Music

- Rāg: melodic framework of Indian art music

Nyās translates to home/residence

*Nyās Svar (Rāg Bilaskhani todi)

S	r*		g*		m*		P	d*		n	

Bhairavi Thaat

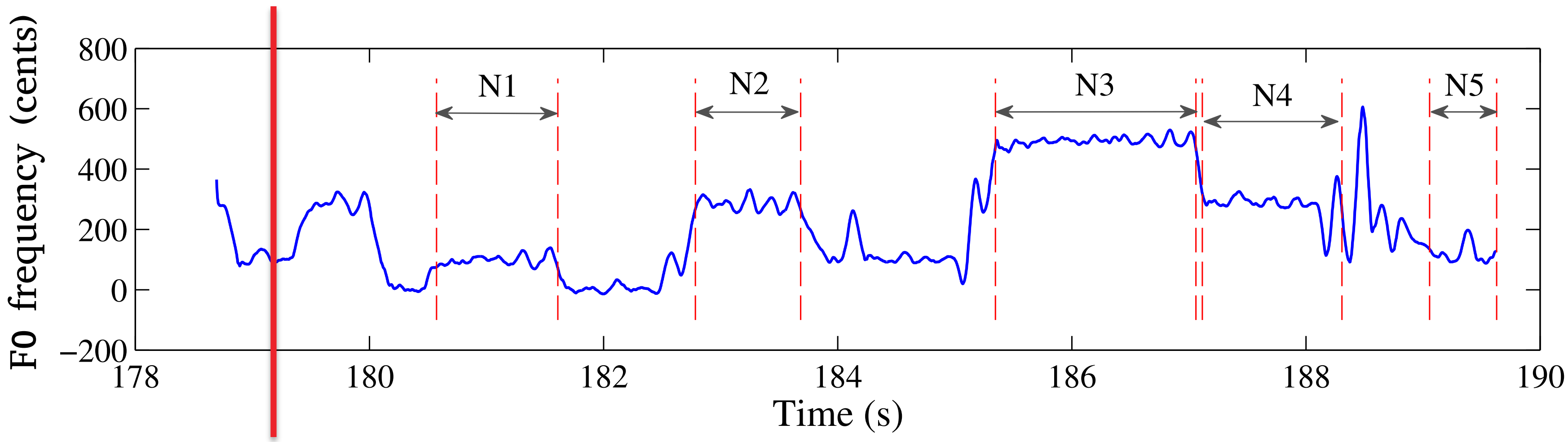
S	r		g		m		P	d		n	

Svaras

S	r	R	g	G	m	M	P	d	D	n	N
Do		Re		Mi	Fa		Sol		La		Ti

Melodic Landmark: Nyās Occurrences

- Example



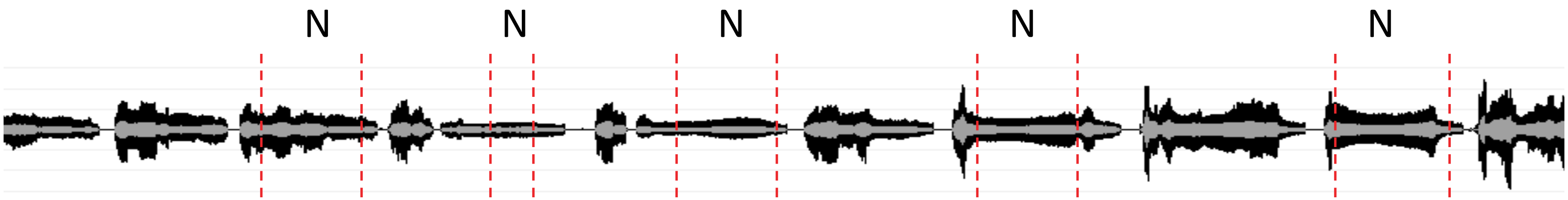
A. K. Dey, Nyāsa in rāga: the pleasant pause in Hindustani music. Kanishka Publishers, Distributors, 2008.

Goal and Motivation

- Methodology for detecting nyās occurrences

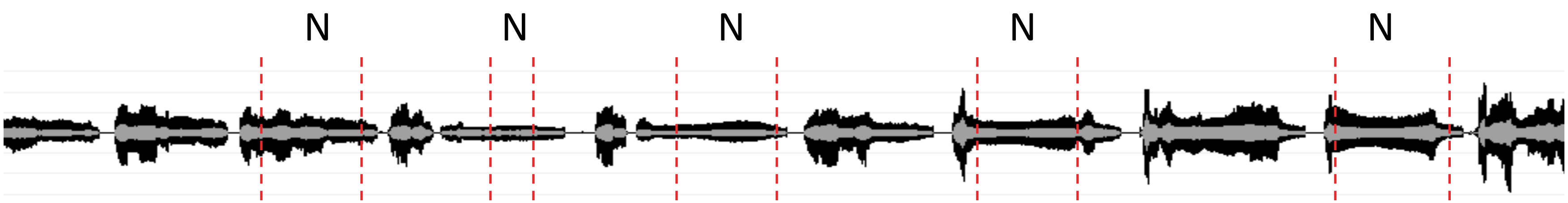
Goal and Motivation

- Methodology for detecting nyās occurrences



Goal and Motivation

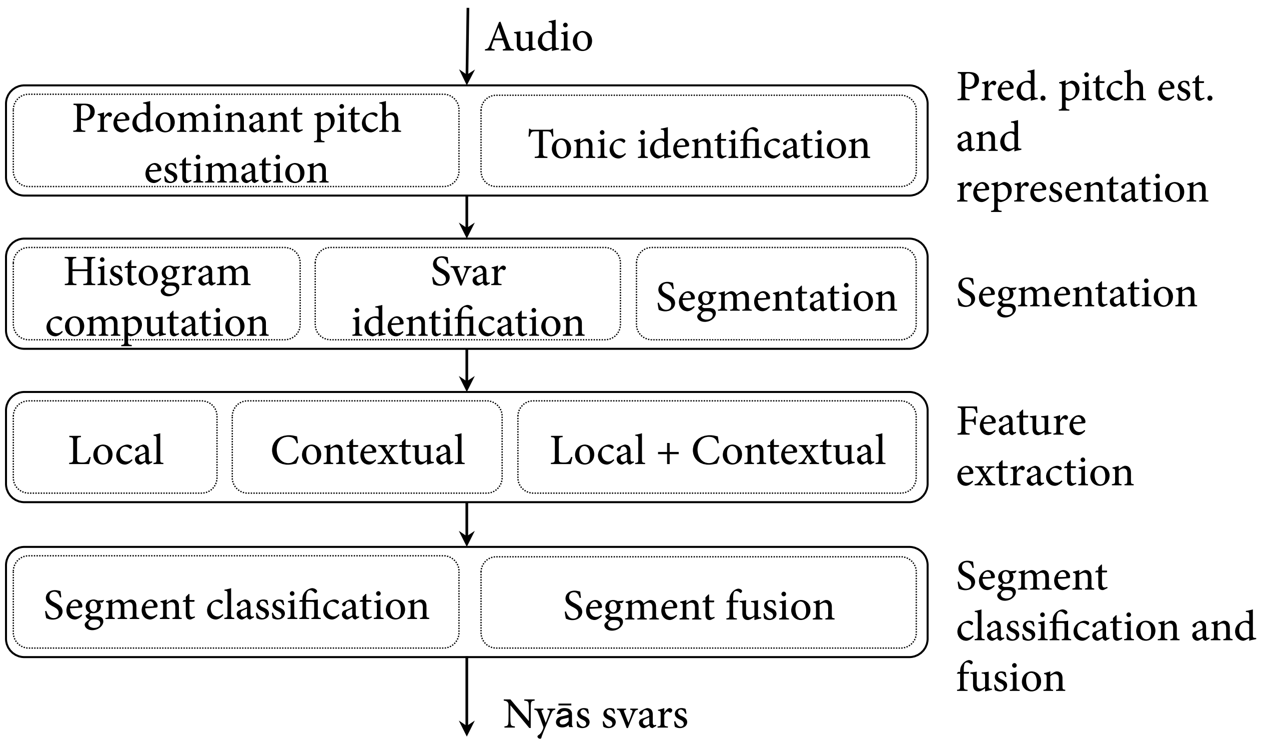
- Methodology for detecting nyās occurrences



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

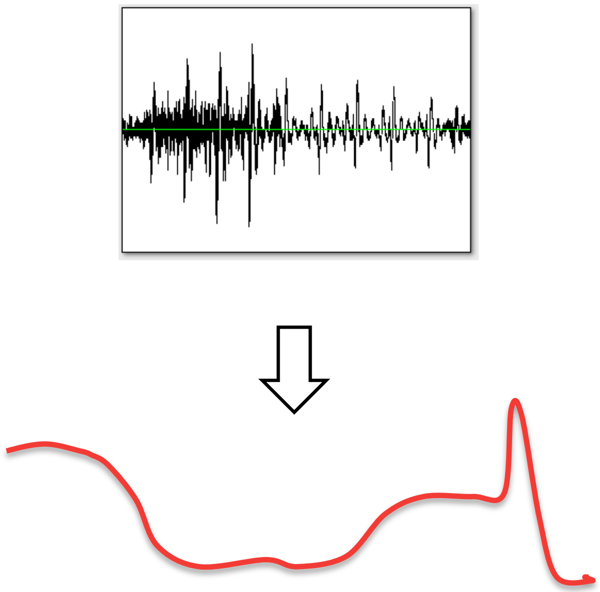
J. C. Ross and P. Rao, "Detection of raga-characteristic phrases from Hindustani classical music audio," in Proc. of 2nd CompMusic Workshop, 2012, pp. 133– 138.

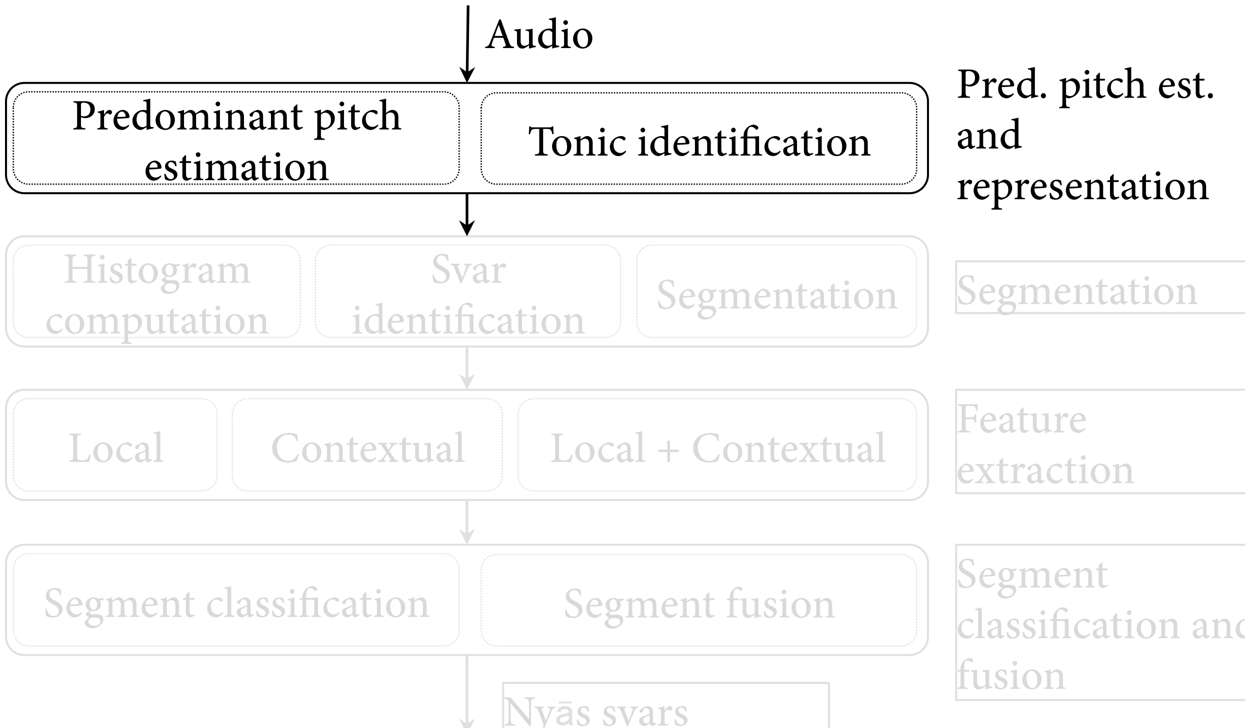
Methodology: Block Diagram



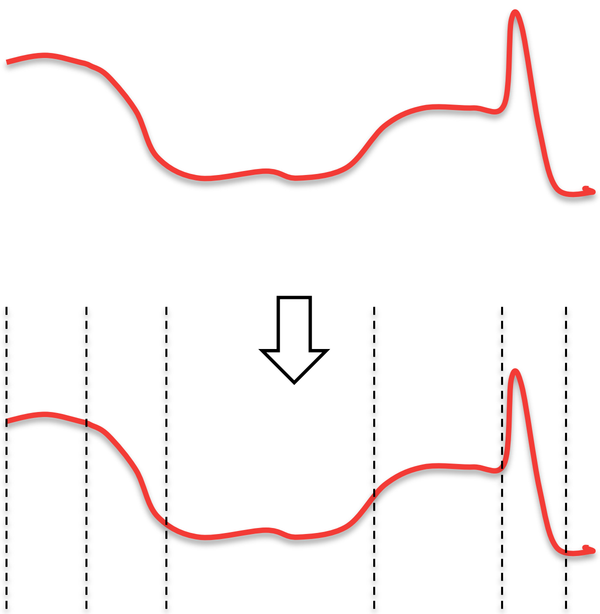
Block diagram of the proposed methodology

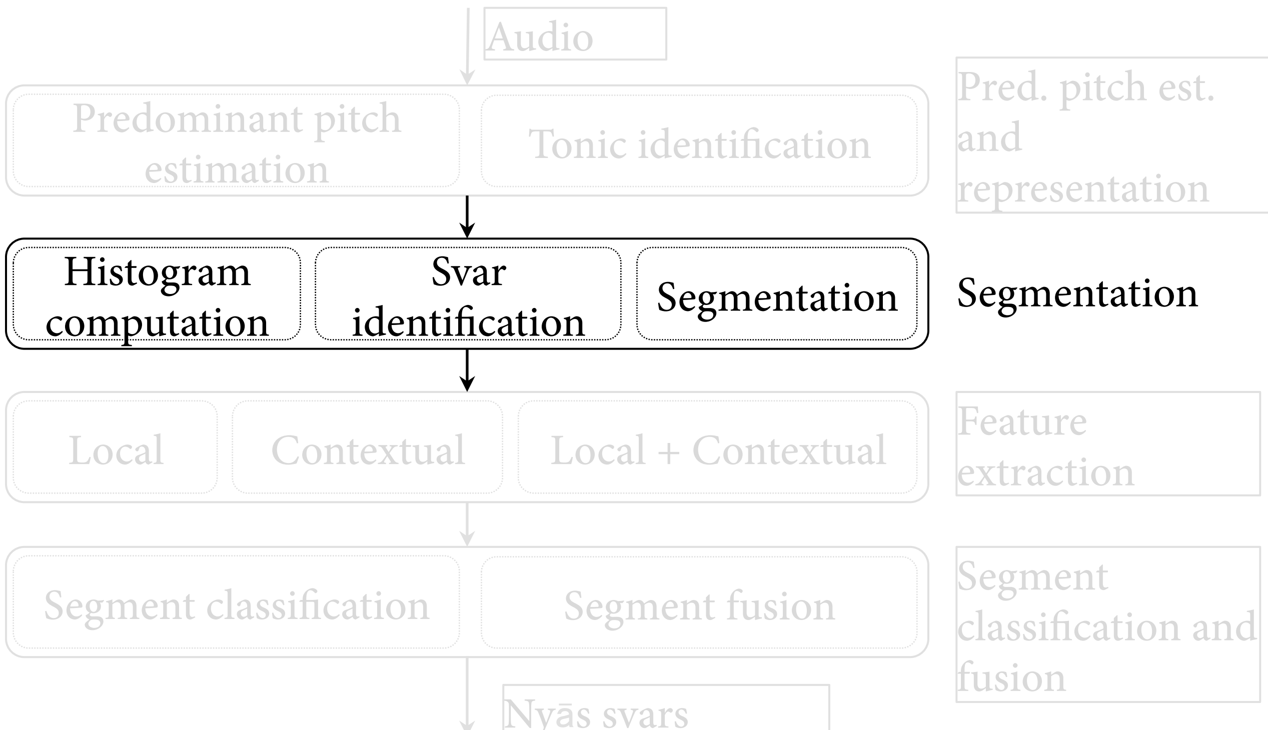
Methodology: Pred. Pitch Estimation



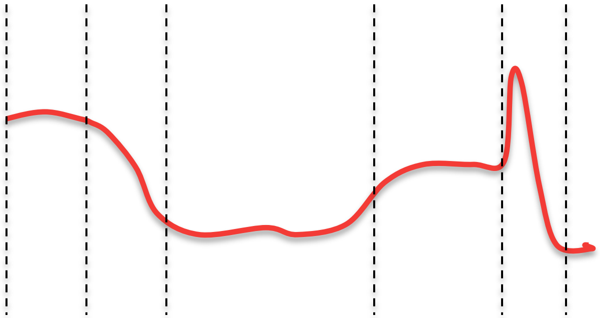


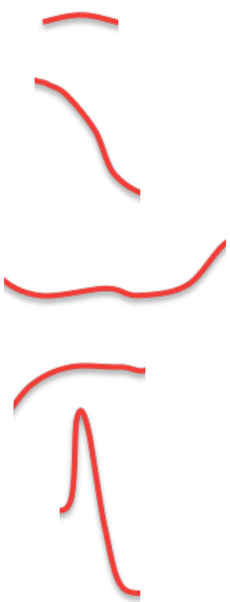
Methodology: Segmentation





Methodology: Feature Extraction





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

Audio

Predominant pitch estimation | Tonic identification — Pred. pitch est. and representation

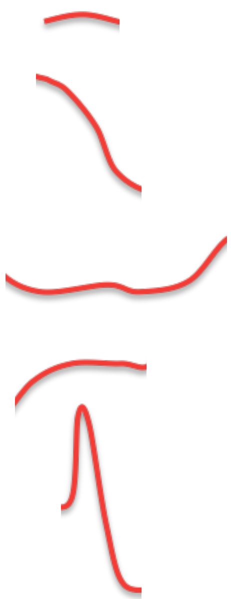
Histogram computation | Svar identification | Segmentation — Segmentation

Local | Contextual | Local + Contextual — Feature extraction

Segment classification | Segment fusion — Segment classification and fusion

Nyās svars

Methodology: Segment Classification



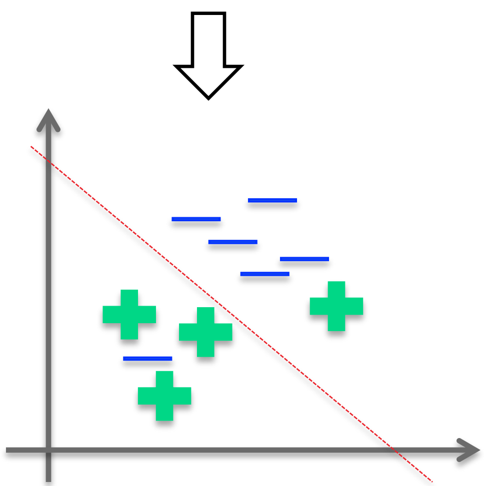
0.12, 0.34, 0.59, 0.23, 0.54

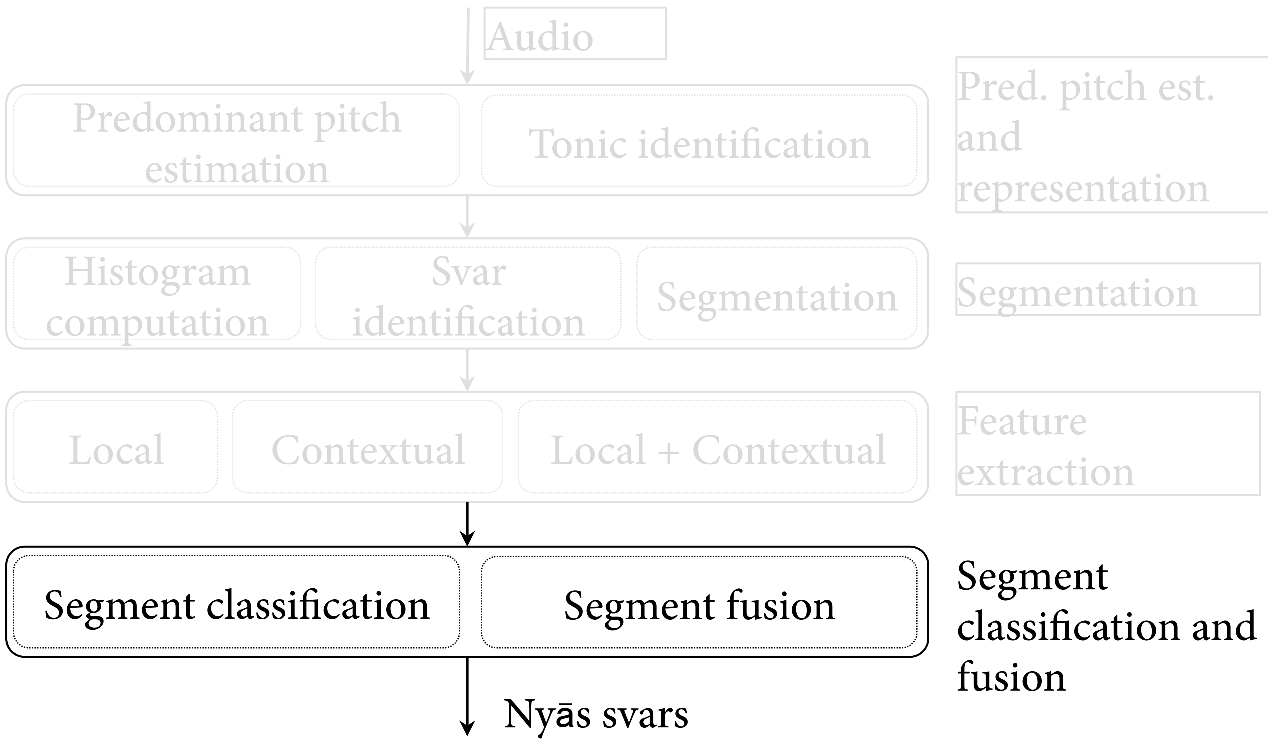
0.21, 0.24, 0.54, 0.54, 0.42

0.32, 0.23, 0.34, 0.41, 0.63

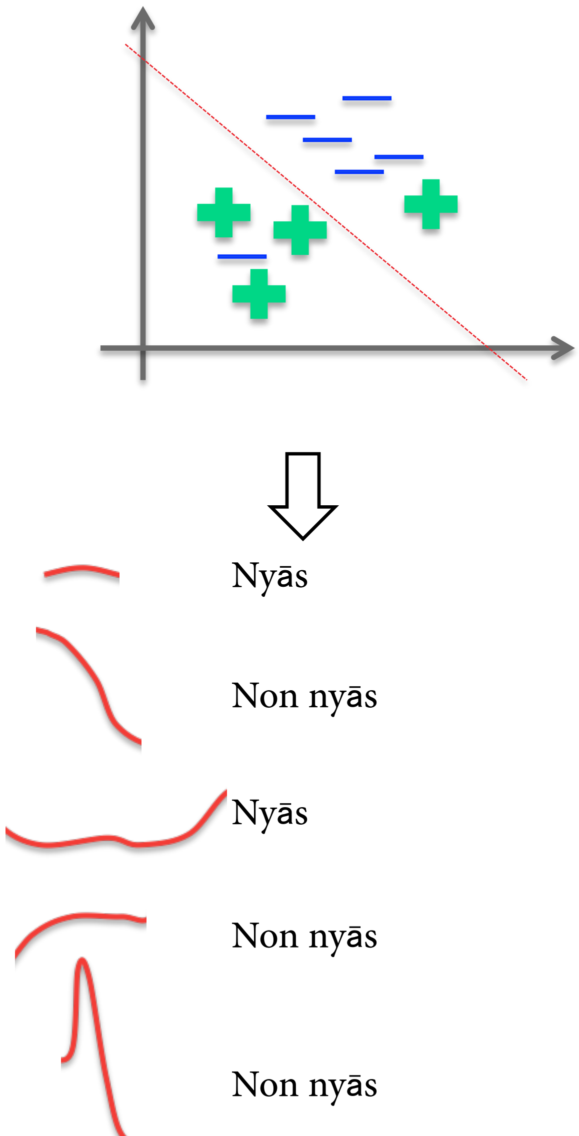
0.66, 0.98, 0.74, 0.33, 0.12

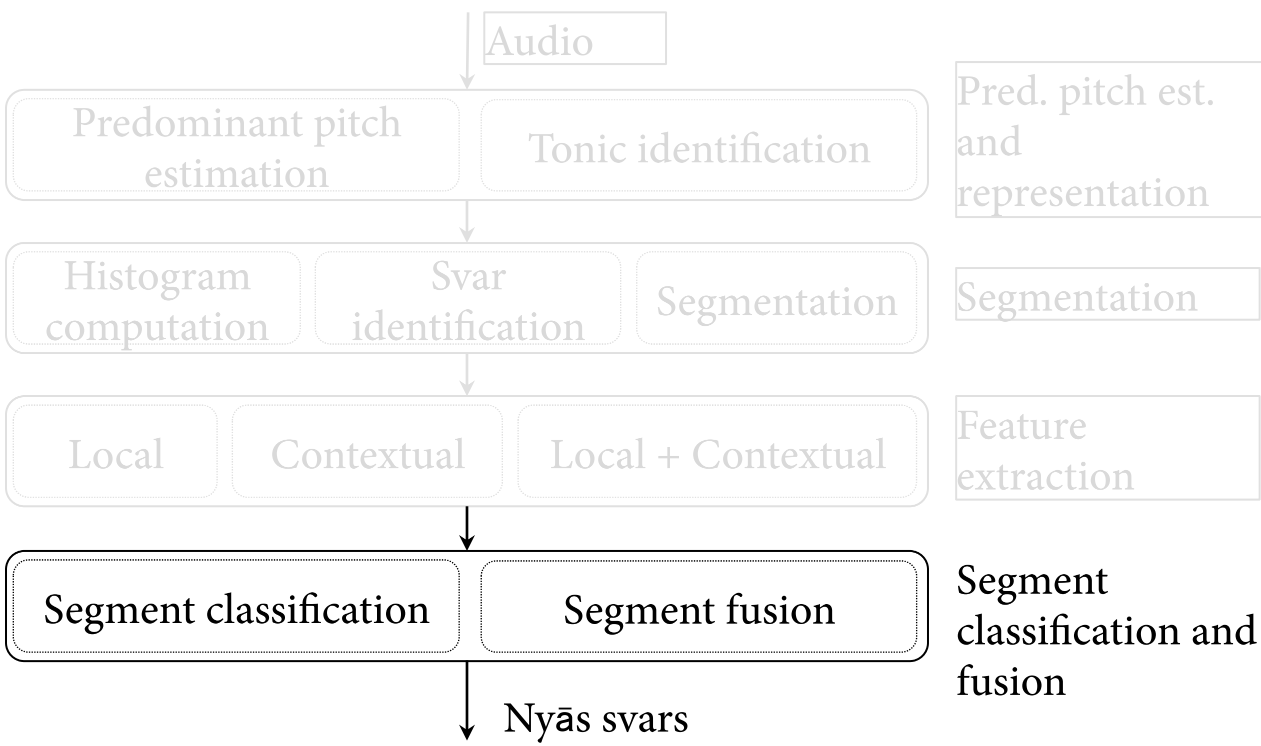
0.90, 0.42, 0.14, 0.83, 0.76



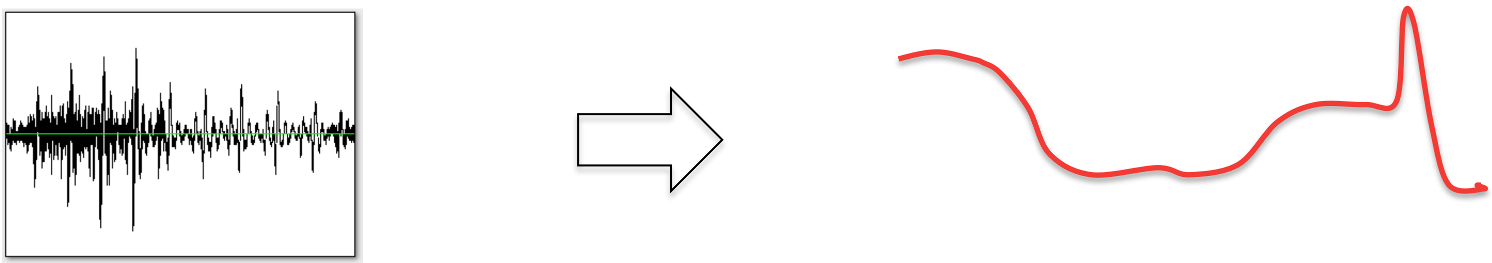


Methodology: Segment Classification





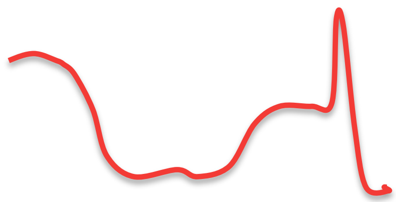
Pred. Pitch Estimation and Representation

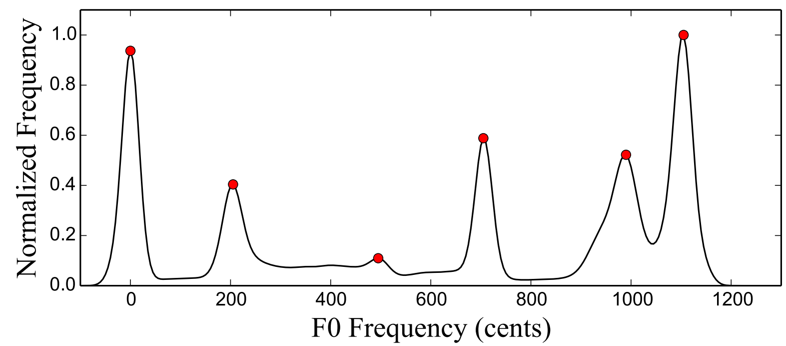


- Predominant pitch estimation
  - Method by Salamon and Gómez (2012)
  - Favorable results in MIREX'11
- Tonic Normalization
  - Pitch values converted from Hertz to Cents
  - Multi-pitch approach by Gulati et al. (2014)

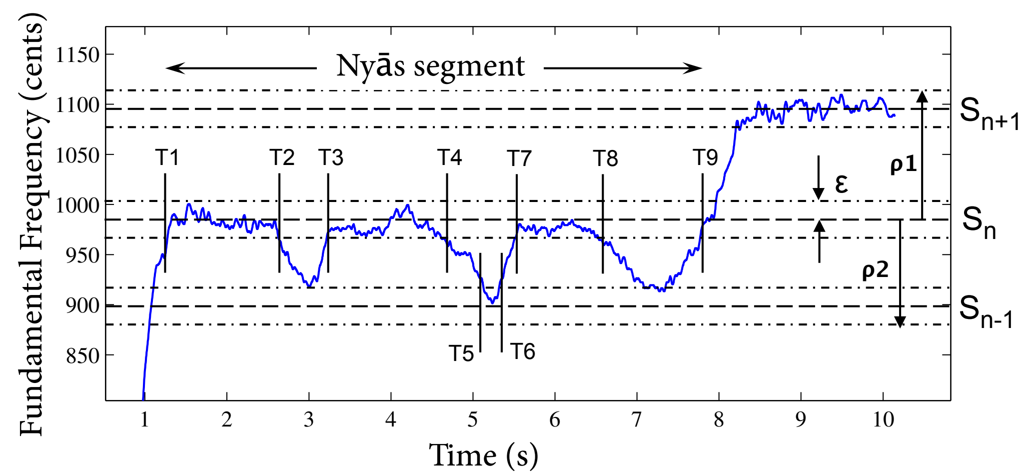
J. Salamon and E. Go ́mez, "Melody extraction from polyphonic music signals using pitch contour charac- teristics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 6, pp. 1759–1770, 2012.
S. Gulati, A. Bellur, J. Salamon, H. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra, "Automatic Tonic Identi- fication in Indian Art Music: Approaches and Evalua- tion," Journal of New Music Research, vol. 43, no. 01, pp. 55–73, 2014.

Melody Segmentation



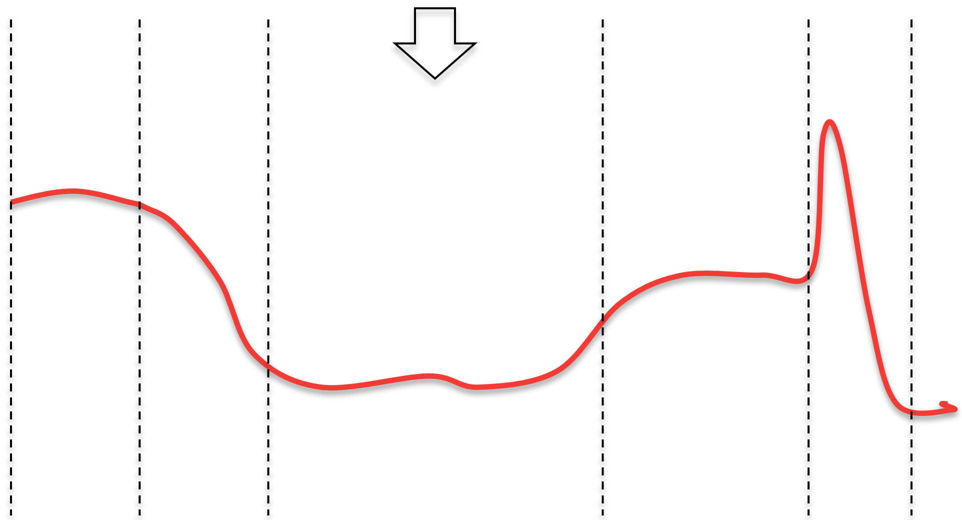


- Baseline: Piecewise linear segmentation (PLS)



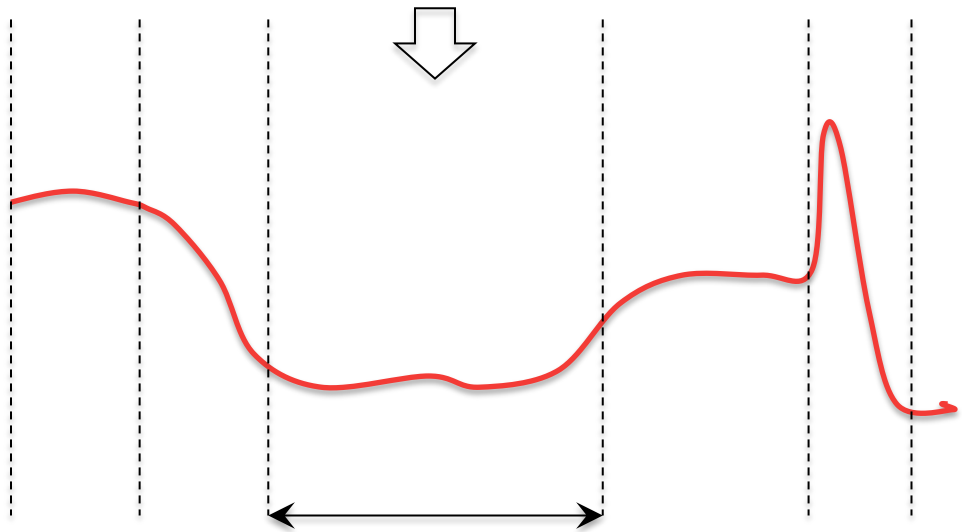
E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: A survey and novel approach," Data Mining in Time Series Databases, vol. 57, pp. 1–22, 2004.

Feature Extraction



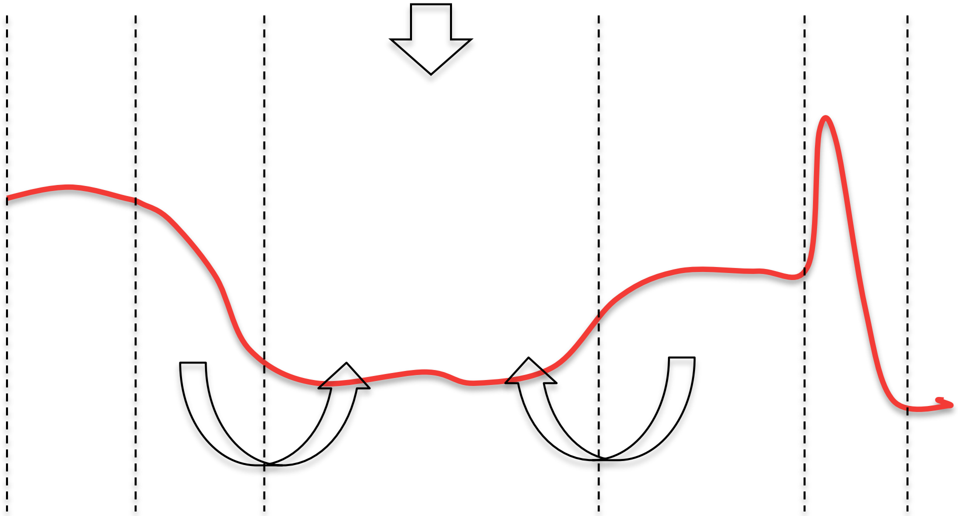
Feature Extraction

- Local (9 features)



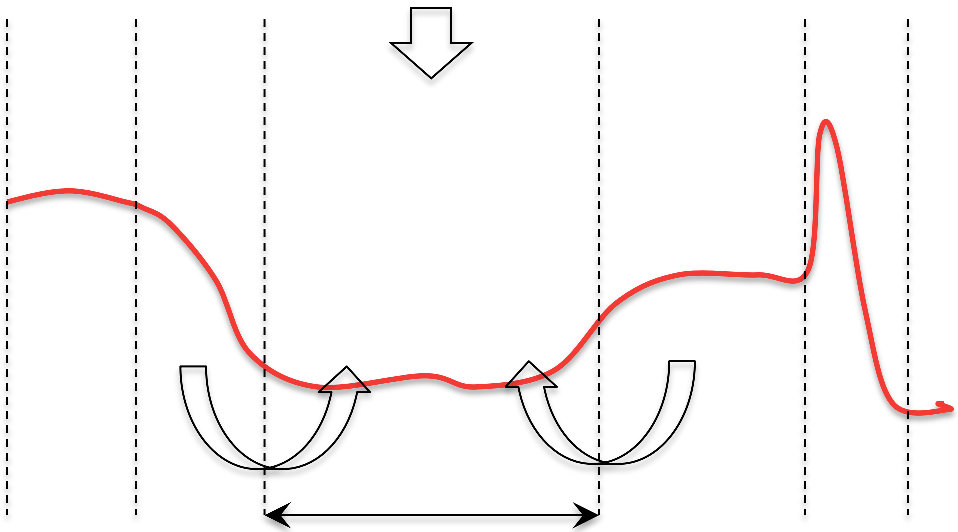
Feature Extraction

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



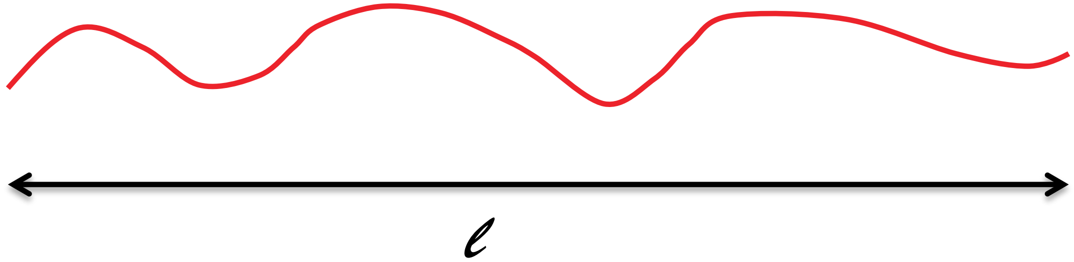
Feature Extraction

- Local (9 features)
- Contextual (24 features)
- Local + Contextual (33 features)



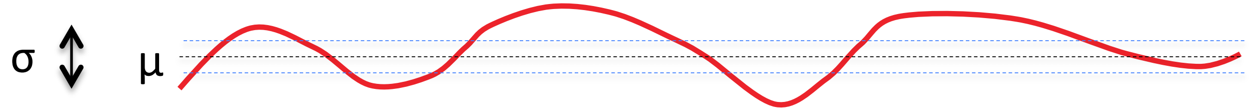
Feature Extraction: Local

- Segment Length



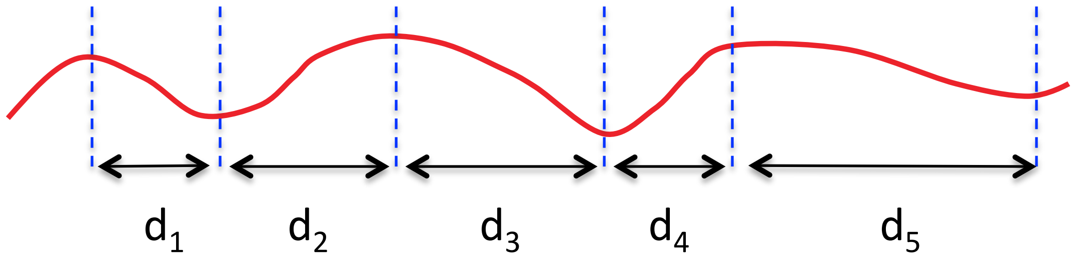
Feature Extraction: Local

- Segment Length
- μ and σ of pitch values



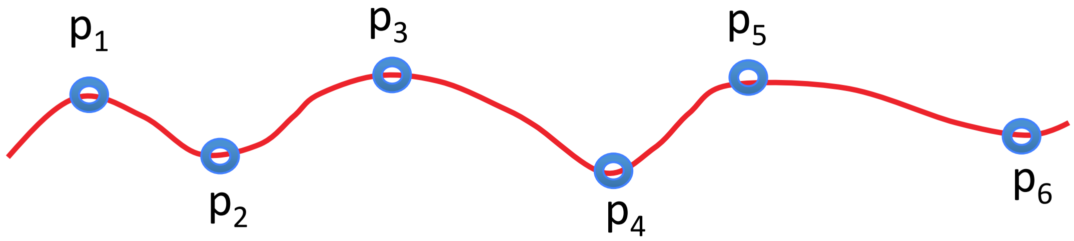
Feature Extraction: Local

- Segment Length
- μ and σ of pitch values
- μ and σ of difference in adjacent peaks and valley locations



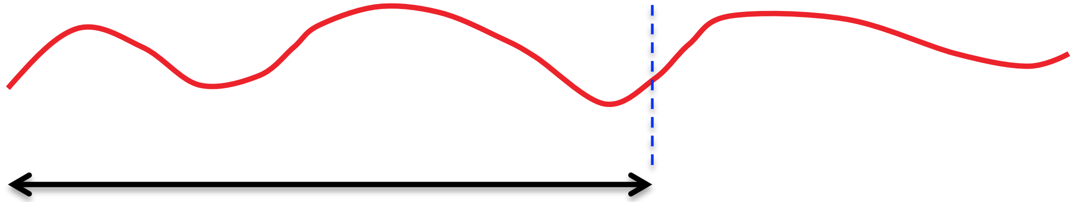
Feature Extraction: Local

- Segment Length
- μ and σ of pitch values
- μ and σ of difference in adjacent peaks and valley locations
- μ and σ of the peak and valley amplitudes



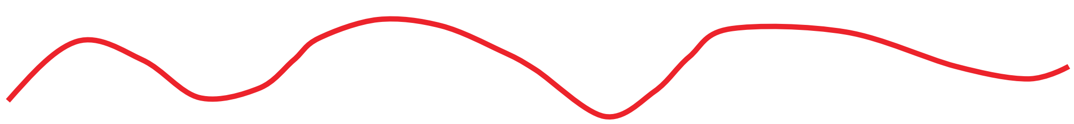
Feature Extraction: Local

- Segment Length
- μ and σ of pitch values
- μ and σ of difference in adjacent peaks and valley locations
- μ and σ of the peak and valley amplitudes
- Temporal centroid (length normalized)



Feature Extraction: Local

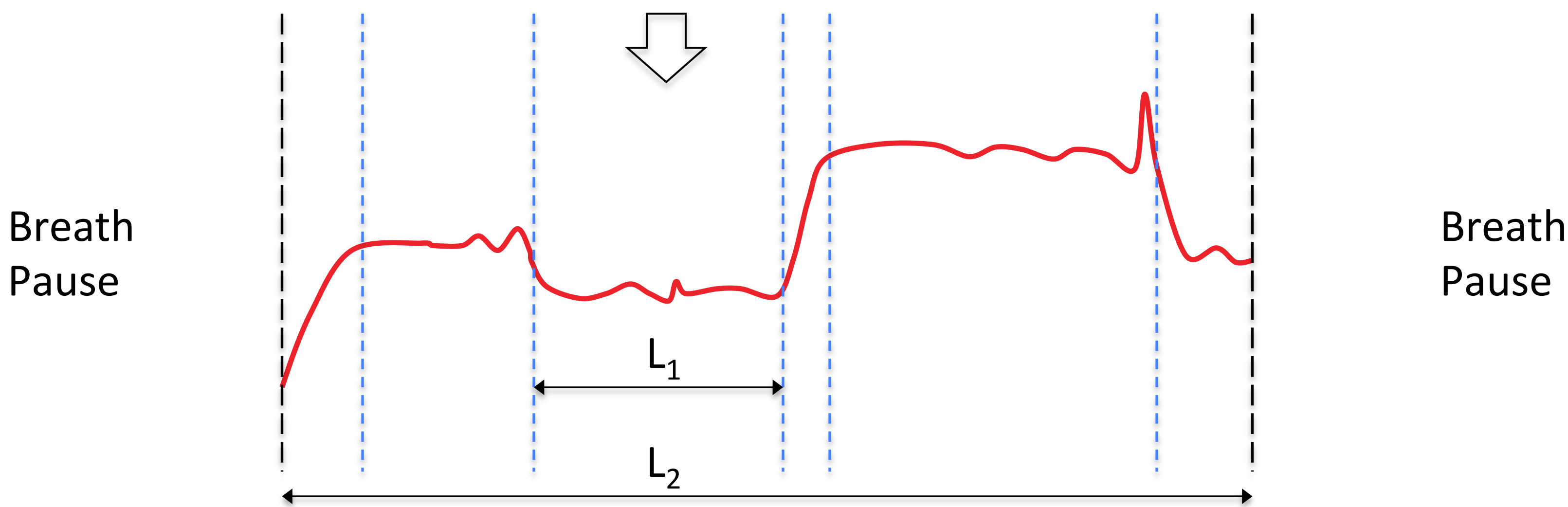
- Segment Length
- μ and σ of pitch values
- μ and σ of difference in adjacent peaks and valley locations
- μ and σ of the peak and valley amplitudes
- Temporal centroid (length normalized)
- Binary flatness measure



Flat or non-flat

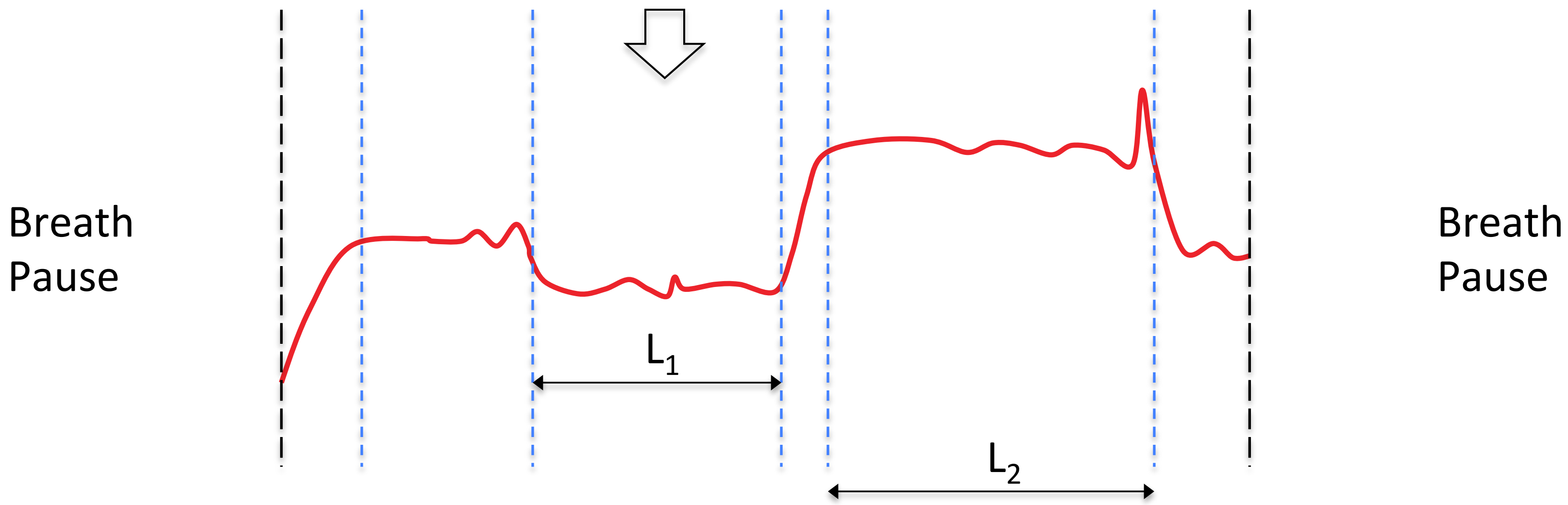
Feature Extraction: Contextual

- 4 different normalized segment lengths



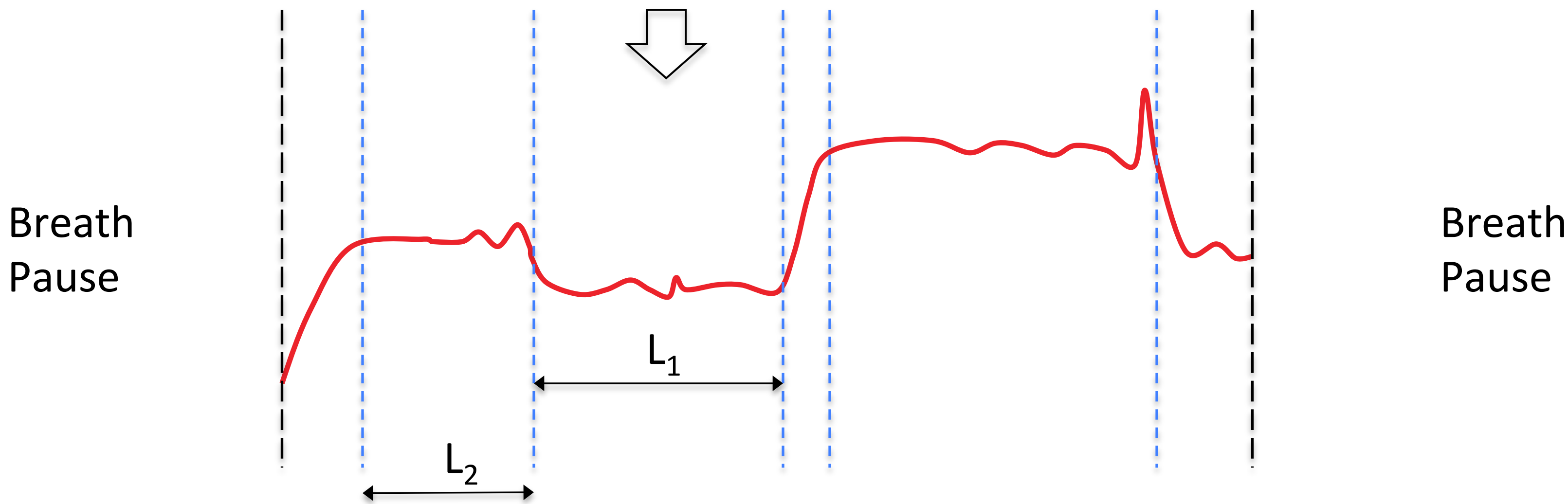
Feature Extraction: Contextual

- 4 different normalized segment lengths



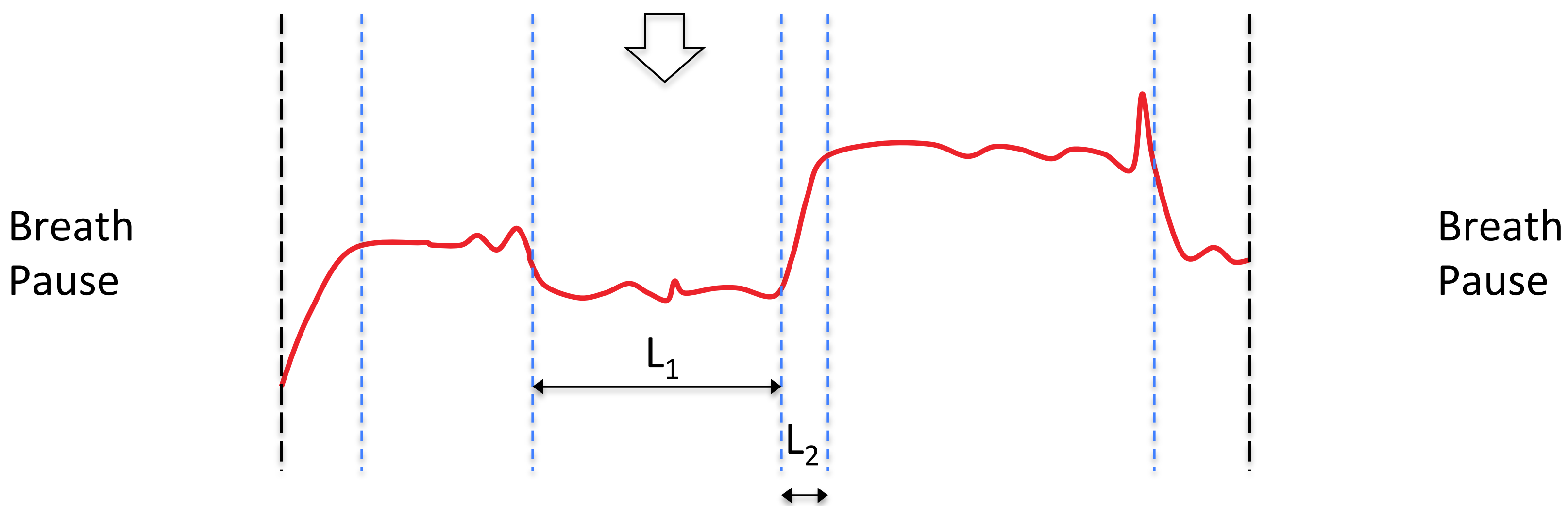
Feature Extraction: Contextual

- 4 different normalized segment lengths



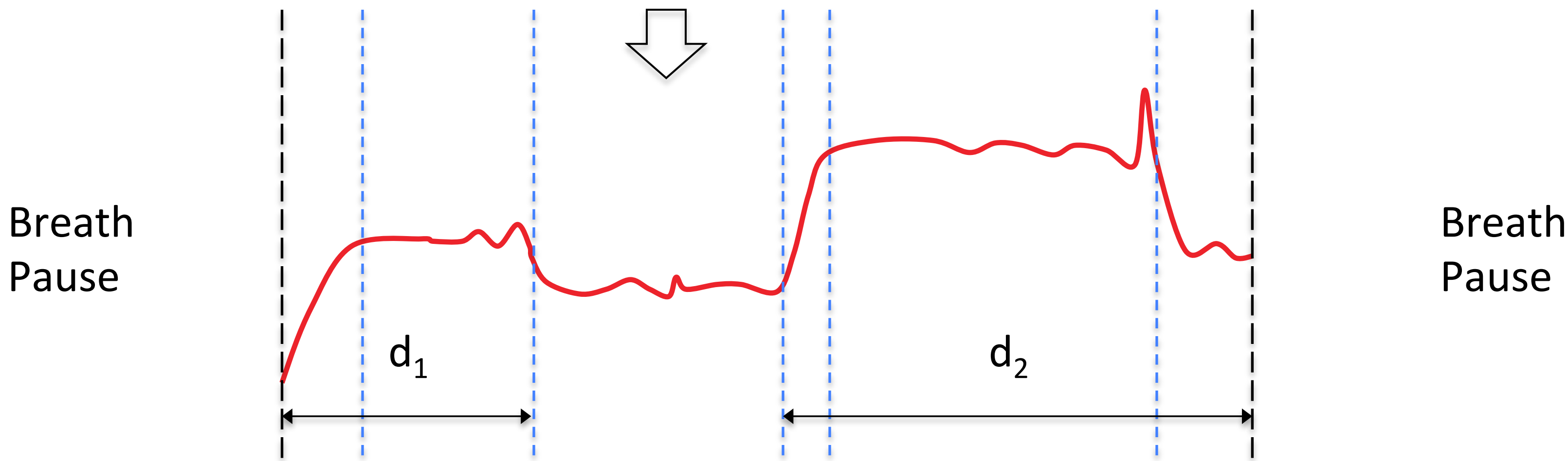
Feature Extraction: Contextual

- 4 different normalized segment lengths



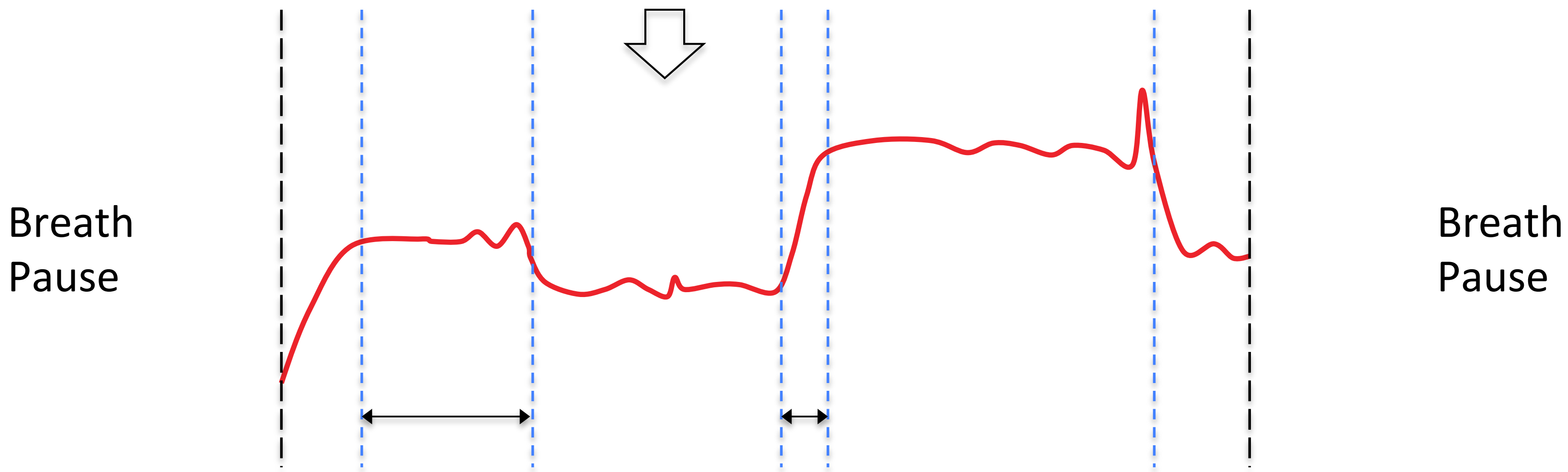
Feature Extraction: Contextual

- 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*2=18)



Segment Classification

- Class: Nyās and Non-nyās
- Classifiers:
  - Trees (min_sample_split=10)
  - K nearest neighbors (n_neighbors=5)
  - Naive bayes (fit_prior=False)
  - Logistic regression (class_weight=‘auto’)
  - Support vector machines (RBF)(class weight=‘auto’)
- Testing methodology
  - Cross-fold validation
- Software: Scikit-learn, version 0.14.1

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duches- nay, “Scikit-learn: machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825– 2830, 2011

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 to16.7 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

M. Levy and M. Sandler, "Structural segmentation of musical audio by constrained clustering," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 2, pp. 318–326, 2008.
H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," The annals of mathematical statistics, vol. 18, no. 1, pp. 50–60, 1947.
S. Holm, "A simple sequentially rejective multiple test procedure," Scandinavian journal of statistics, pp. 65– 70, 1979.

Evaluation: Baseline Approach

compmusic

- 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

X. Xi, E. J. Keogh, C. R. Shelton, L. Wei, and C. A. Ratanamahatana, “Fast time series classification using numerosity reduction,” in Proc. of the Int. Conf. on Ma- chine Learning, 2006, pp. 1033–1040.
X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. J. Keogh, “Experimental com- parison of representation methods and distance mea- sures for time series data,” Data Mining and Knowl- edge Discovery, vol. 26, no. 2, pp. 275–309, 2013.

Results: Nyās Boundary Annotation

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	0.356	0.407	0.447	0.248	0.449	0.453
	C	0.284	0.394	0.387	0.383	0.389	0.406
	L+C	0.289	0.414	0.426	0.409	0.432	0.437
B	L	**0.524**	0.672	**0.719**	0.491	**0.736**	**0.749**
	C	0.436	0.629	0.615	**0.641**	0.621	0.673
	L+C	0.446	**0.682**	0.708	0.591	0.725	0.735

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

	Feat.	DTW	Tree	KNN	NB	LR	SVM
A	L	**0.553**	0.685	0.723	0.621	0.727	0.722
	C	0.251	0.639	0.631	0.690	0.688	0.674
	L+C	0.389	0.694	0.693	0.708	0.722	0.706
B	L	0.546	**0.708**	**0.754**	0.714	**0.749**	**0.758**
	C	0.281	0.671	0.611	0.697	0.689	0.697
	L+C	0.332	0.672	0.710	**0.730**	0.743	0.731

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 Gulati[1], Joan Serrà[2], Kaustuv K. Ganguli[3] and Xavier Serra[1]

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

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

SMC-ICMC 2014, Athens, Greece

upf. Music Technology Group

compmusic

INSTITUT D' INVESTIGACIÓ EN INTEL.LIGÈNCIA ARTIFICIAL - IIIA