An Evaluation of Methodologies for Melodic Similarity in Audio Recordings of Indian Art Music

Sankalp Gulati*, Joan Serrà¥ and Xavier Serra*

sankalp.gulati@upf.edu, jserra@iiia.csic.es and xavier.serra@upf.edu

*Music Technology Group, UPF, ¥Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Spain

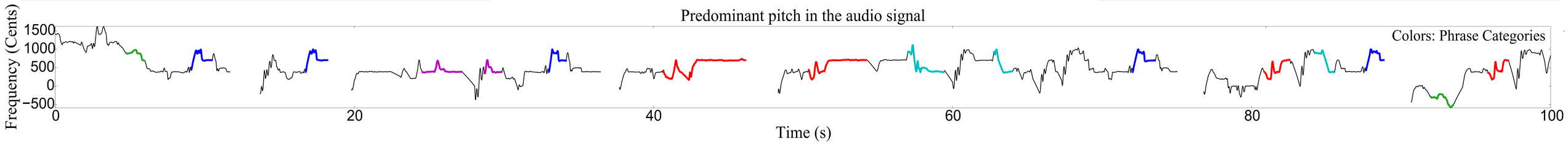
40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015

Indian art Music

- Hindustani (North-Indian), Carnatic (South-Indian) music.
- Rāg melody framework, Tāl rhythm framework
- Rāg: Svaras, Aroh-Avroh, **Characteristic phrases**
- Oral pedagogy, essentially audio music repertoire
- Practically no written music (descriptive) scores

Goals

- Detect the occurrences of phrases of a rāg
- Optimal methodology: approach and parameter values
- Evaluation of 560 combinations: sampling rates (5), normalization techniques (8), time-scaling (2) and distance measures (7)



Challenges

- Variability in the overall duration
- Large non-linear timing variations
- Added Melodic ornamentations

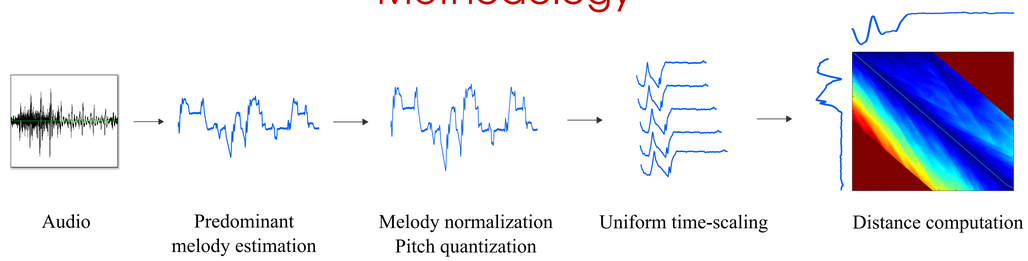
Music collection

- 5 hours of polyphonic audio recordings
- 23 Carnatic (CMD), 9 Hindustani (HMD)
- 10 Phrase categories, 6 rāgs, 21 artists
- Mean duration 1.15 (CMD), 1.5 (HMD)
- Annotations: 2 performing musicians

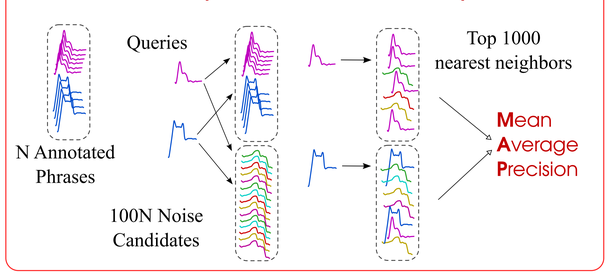
Evaluation

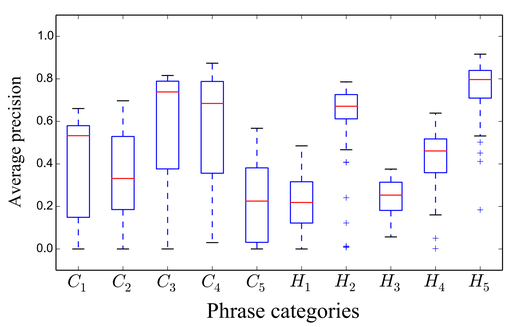
- Select top 1000 nearest neighbors
- Mean Average Precision (MAP)
- Wilcoxon signed rank-test ($p < 0.01$)
- Holm-Bonferroni adjustment

Methodology



Experimental setup



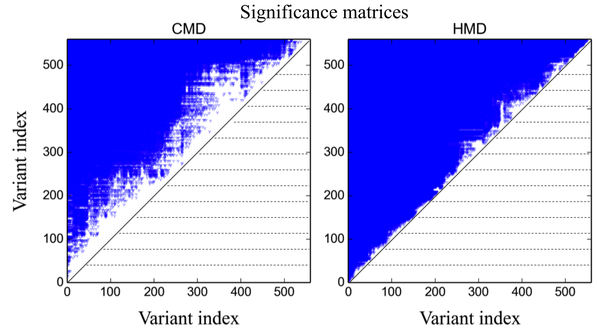


Results

Best performing variants

Dataset	MAP	Sample Rate (Hz)	Normalization	Time-Scaling	Distance
CMD	0.413	67	Mean	Off	$D_{DTW\_L1\_G90}$
	0.412	67	Mean	On	$D_{DTW\_L1\_G10}$
	0.411	100	Mean	Off	$D_{DTW\_L1\_G90}$
HMD	0.552	100	Tonic	Off	$D_{DTW\_L0\_G90}$
	0.551	67	Tonic	Off	$D_{DTW\_L0\_G90}$
	0.547	50	Tonic	Off	$D_{DTW\_L0\_G90}$

Significance matrices



References

- X. Serra, "A multicultural approach to music information research," in Proc. of ISMIR, 2011, pp. 151–156.
- J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 6, pp. 1759–1770, 2012.
- D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an audio analysis library for music information retrieval," in Proc. of ISMIR, 2013, pp. 493–498.
- S. Gulati, A. Bellur, J. Salamon, H. G. Ranjani, V. Ishwar, H. A. Murthy, and X. Serra, "Automatic tonic identification in Indian art music: approaches and evaluation," Journal of New Music Research, vol. 43, no. 1, pp. 55–73, 2014.
- A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by humming: musical information retrieval in an audio database," in Proc. of the third ACM Int. Conf. on Multimedia. ACM, 1995, pp. 231–236.

Conclusions

- DTW-based distance preferred over Euclidean
- Local constraint (DTW) improves the performance
- Time-scaling is beneficial for constrained DTW
- Carnatic melodies need higher sampling rate
- Mean normalization is beneficial in Carnatic music
- Tonic normalization is beneficial in Hindustani music