Structural Segmentation and Visualisation of Sitar and Sarod Concert Audio

T. P. Vinutha S. Suryanarayana K. K. Ganguli P. Rao

Department of Electrical Engineering , Indian Institute of Technology Bombay

ISMIR 2016

compmusic

Sitar and Sarod Music

->Sitar and sarod are plucked string instruments used in Hindustani classical music

->In sitar/sarod concerts, a tabla (percussion) accompanist plays to a certain metre (tala). The **metric tempo** (speed of the metre) increases gradually through the concert

->The **surface tempo** (perceived tempo), a multiple of the metric tempo, changes in certain sections. Over few cycles of the metre, one of the players plays at a faster rate

->Sitar and sarod concerts in Hindustani music have certain musicological sections, based on rhythm:
1) **Layakari:** the sitar/sarod plays in a fast, rhythmic manner
2) **Tabla solo:** the tabla takes center-stage, playing at a fast rate and improvising on the fixed meter

Sarod

Aim of the work

->To reproduce an expert's annotation of the metric tempo and surface tempo from the audio, as in Fig. 1 [1]

->To obtain the boundaries of layakari and tabla solo sections from these rhythmic features, as marked in Fig. 1

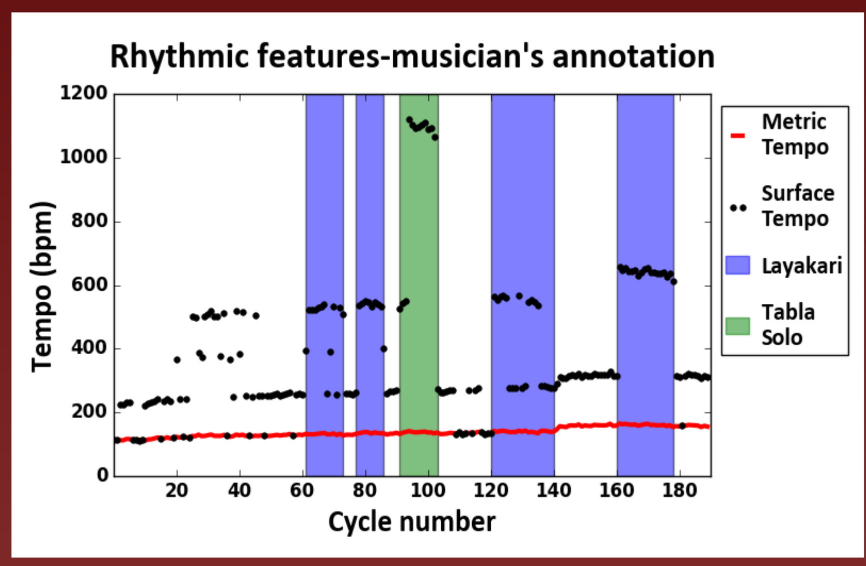

Figure 1

Onset Detection

->Standard onset detection functions (ODFs) like spectral flux [3] detect onsets of all instruments, and can be used to compute the surface tempo. However, they can't be used to:
1) Distinguish between layakari and tabla solo sections
2) Compute the metric tempo, which is set by the tabla alone

->The proposed strategy yields both a general ODF and a tabla-selective ODF. Together, they solve the above challenges

Performance of proposed ODFs:
->ROC of Fig. 2a: The proposed general ODF ~ spectral flux ODF in detecting all onsets

->ROC of Fig. 2b: The tabla ODF performs much better than the spectral flux ODF in selectively detecting tabla onsets

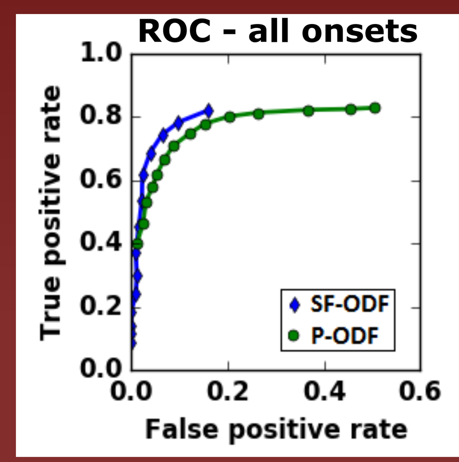

Figure 2a

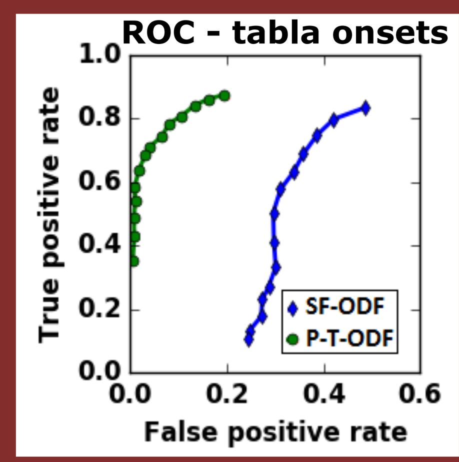

Figure 2b

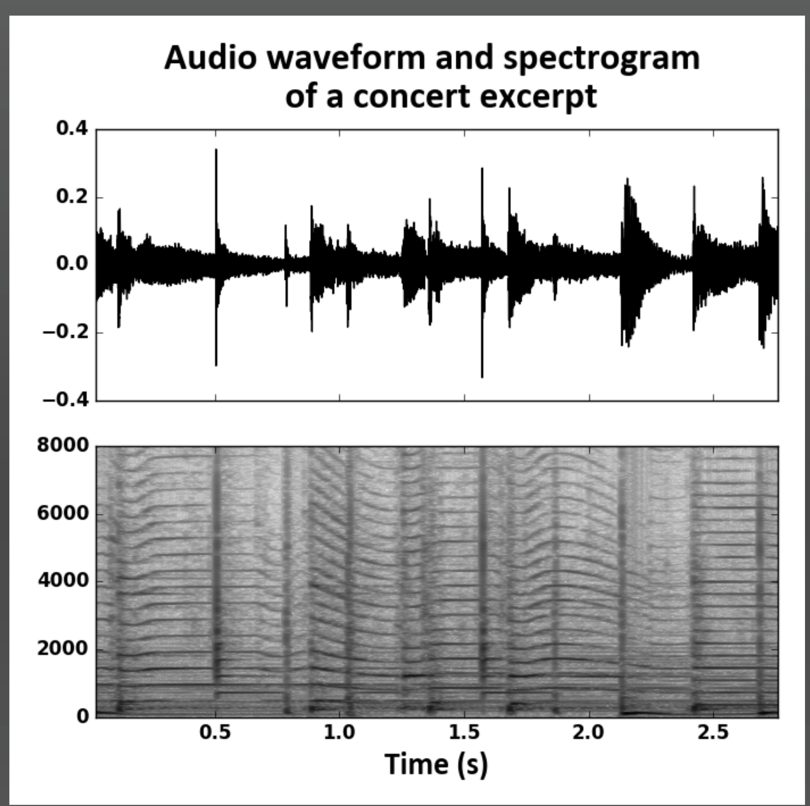

Figure 3

ONSET DETECTION

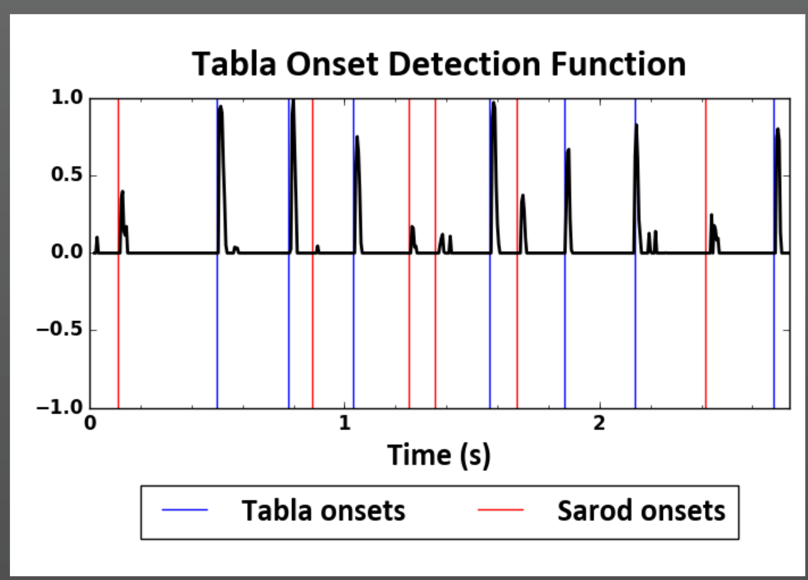

Figure 4a

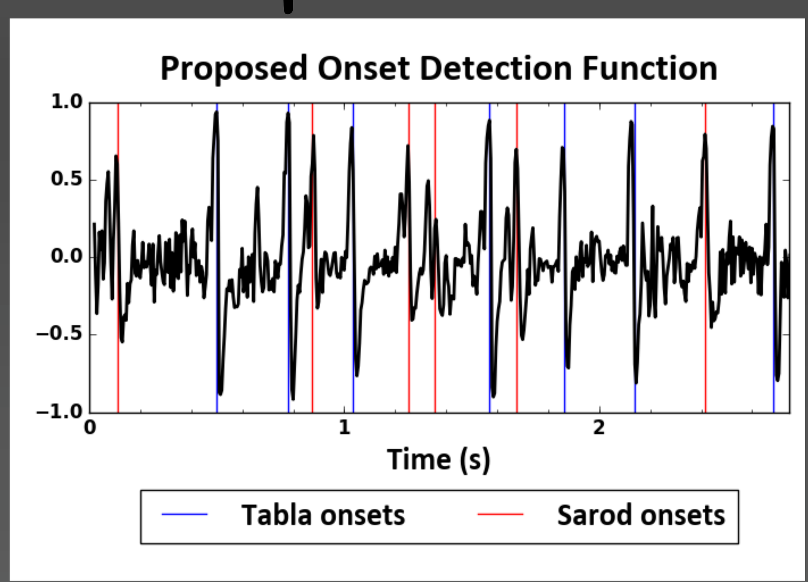

Figure 4b

ACF COMPUTATION

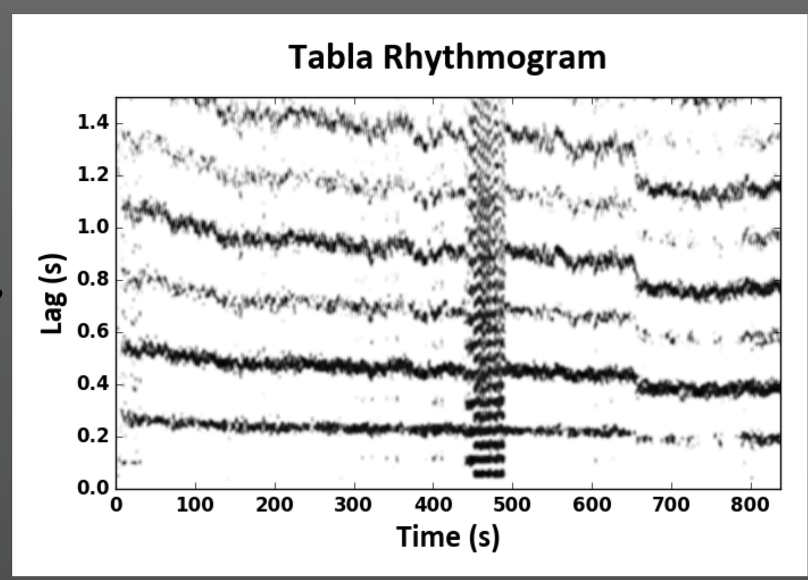

Figure 5a

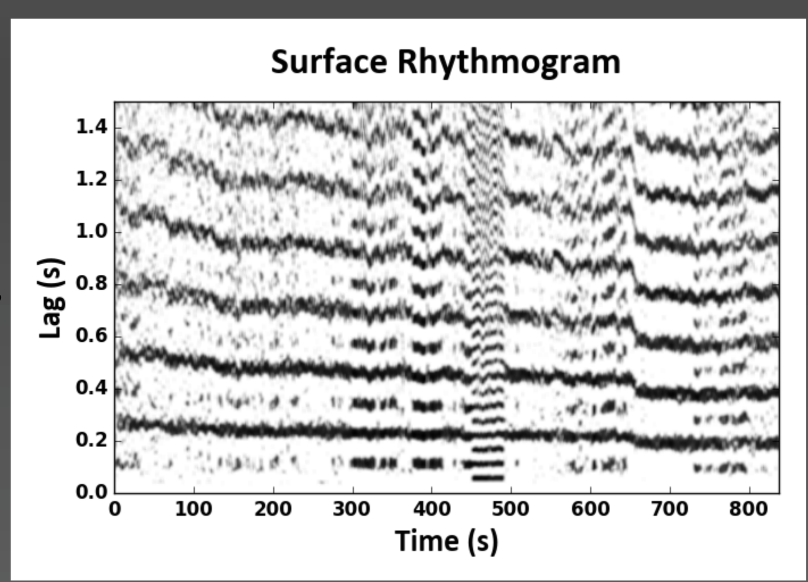

Figure 5b

TEMPO ESTIMATION

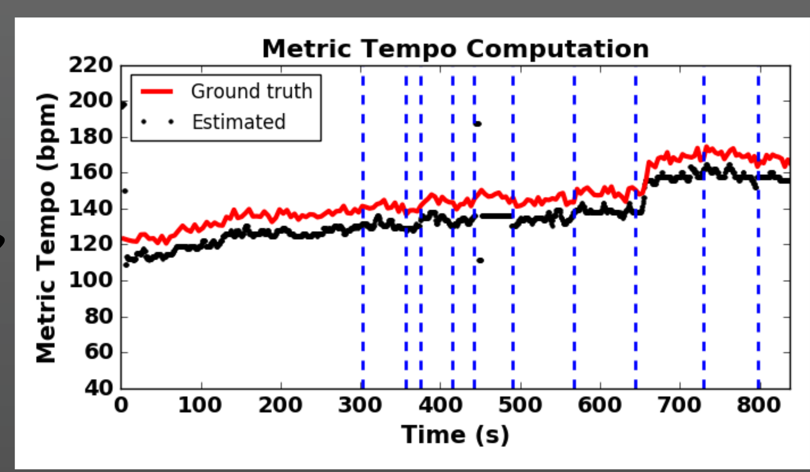

Figure 6a

Note: Computed and ground truth plots are offset for comparison

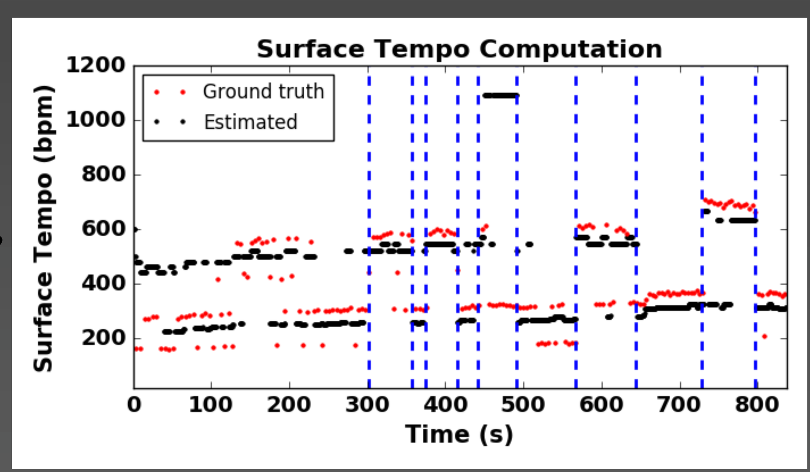

Figure 6b

The **proposed general ODF**:

$$P\text{-}ODF[n] = \sum_{k=0}^{N/2} \mathbb{1}\{|X[n,k]| > |X[n-1,k]|\}$$

The ODF counts the number of bins in a spectral frame where the energy increases from the previous frame

Features of this ODF:
1) Spike at every onset, due to increase in energy in all bins; tabla & sitar onsets percussive in nature
2) Downward lobe for tabla onsets alone, caused by sudden fall in energy after a tabla stroke

General ODF is normalised, inverted and thresholded at 0.3 to obtain tabla-sensitive ODF

Auto-correlation Function of ODFs computed piecewise
Texture window = 3 s Window hop = 0.5 s

Rhythmogram: Spectrogram like visualisation of ACF vectors, from which rhythm information can be obtained [2]

Metric tempo track: strongest band in the tabla rhythmogram

Surface tempo track: first peak in the surface rhythmogram

Layakari: seen distinctly only in the surface rhythmogram

Tabla solo: seen distinctly in both the rhythmograms

Tabla

Sitar

Metric Tempo and Surface Tempo

Metric tempo: from the tabla rhythmogram
-> maximize the *mean* of the ACF values at candidate lags and corresponding lag multiples

Surface tempo: from the surface rhythmogram
-> maximize the *sum* of the ACF values at candidate lags and corresponding lag multiples

Range of tempo considered: 80 bpm to 1200 bpm

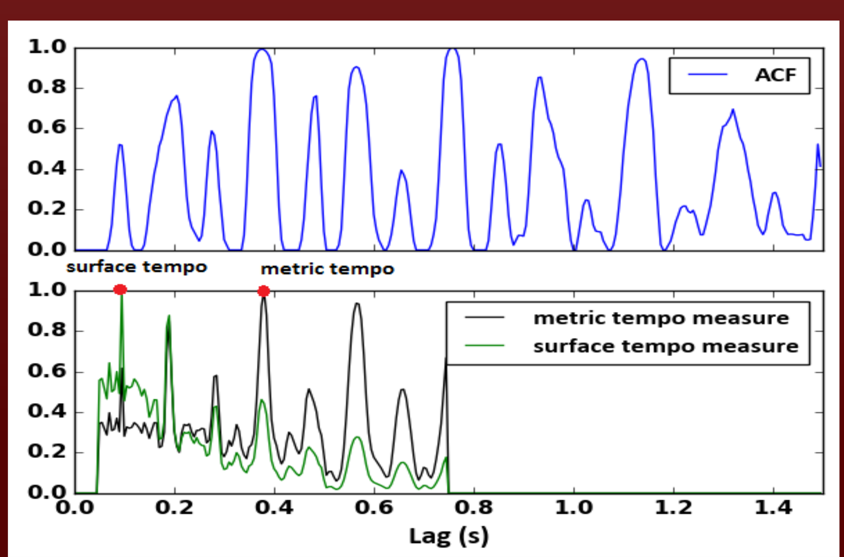

Figure 7

Segmentation by Surface Rhythmogram

Surface rhythmogram
-> Self-distance matrix
-> Novelty function
-> Thresholding
-> Boundaries [4]

Drawback:
->Surface rhythmogram captures rhythmic patterns, not just tempo
-> Segmentation by surface rhythm leads to spurious peaks

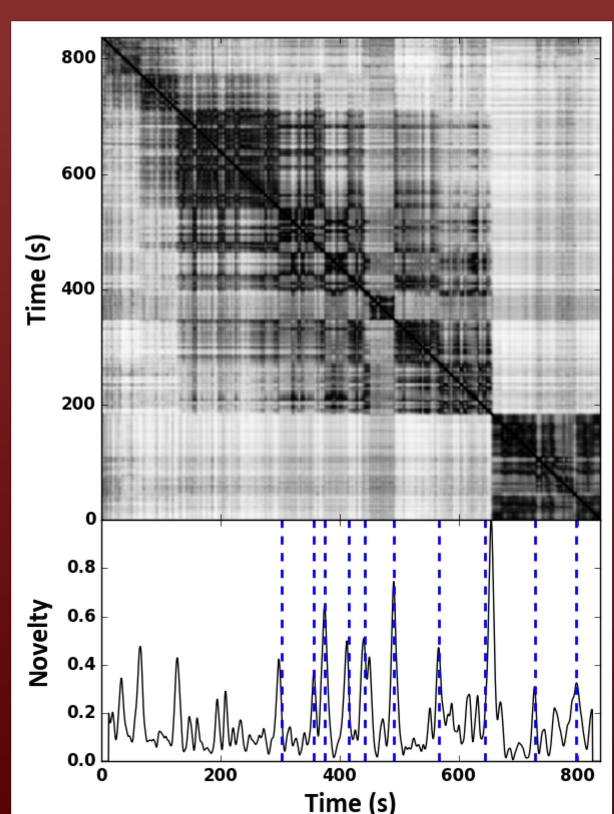

Figure 8

Future work

Effective alternate ways for segmentation can be:
->Section boundaries of a gat form a subset of the change points of either the metric or surface tempo or the ratio between the two. Hence, these reduced vectors can be used for segmentation
->Tempo features combined with
1) Short-time energy feature distinguishing strokes in layakari
2) Chroma variance feature characterising rapid *chikari* (drone string) plucks in faster sections of layakari

References

[1] Martin Clayton. *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*, Chapter 11
[2] Kristoffer Jensen et. al. *Rhythm-Based Segmentation of Popular Chinese Music*, ISMIR, 2005
[3] J. P. Bello et. al. *A Tutorial on Onset Detection of Music Signals*, IEEE Trans. on Speech and Audio Processing, 2005
[4] J. Foote. *Automatic audio segmentation using a measure of audio novelty*, IEEE Intnl. Conf. on Multimedia and Expo, 2000