Structural Segmentation Of Dhrupad Vocal Bandish Audio Based On Tempo
Rohit M. A., Vinutha T. P., Preeti Rao
Dept. of Electrical Engineering, I.I.T. Bombay

Dhrupad Vocal Concert

http://youtu.be/Icgcgya3qM4?t=2442

Lead vocals
Pakhawaj
Tanpura

A section – during which surface tempo of neither instrument changes

Rhythmic Structure in the Bandish
- Metric tempo (m.t.) - Underlying tempo of composition
  - Range: 30 – 85 BPM
- Surface tempo (s.t.) - Rate of sung syllables or played strokes
  - Range: 30 – 960 BPM!
  - Generally an integer multiple of m.t. – 1, 2, …, 8, 16
- Surface tempo multiple (s.t.m.) = s.t. ÷ m.t.

Concert structure

Tasks
- Track metric tempo and overall surface tempo across a concert
- Track the surface tempo of source separated vocals and pakhawaj and obtain section boundaries

System Overview

1. Source separation – Spleeter 2 stems model [2]
3. Surface tempo multiple estimation - modified tempo-cnn

Methods
- 14 concerts (Source: Dunya corpus [4], YouTube)
- 634 sections → 1127 8-second chunks (training examples)
- Data augmentation using time-scaling and overlap between examples

Results
- M.T. estimation (%) accuracy:
  - Vocals: 71.6, 74.7
  - Pakhawaj: 67.7, 71.0
- S.T.m. estimation (% accuracy): 70.4

Takeaways
- Better results observed on pakhawaj and mixture audios
- Imperfect source separation and melismatic singing are challenges in vocals
- Confusions in s.t.m. due to accents on alternate beats
- Using metric tempo as conditioning could help resolve confusions in s.t.m. prediction

References
[3] https://github.com/hendriks73/tempo-cnn