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

The Dhrupad Vocal Concert

Concert setting

https://youtu.be/IcgcgYa3qM4

Lead vocals

Tanpura

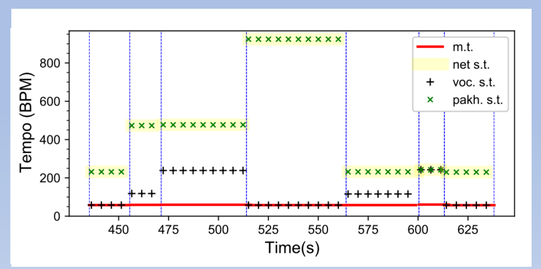
Pakhawaj

Alap (unmetered) ≈ 40 minutes → *Bandish* (metered) ≈ 15 mins. → *time*

Concert structure

Tempo and 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.

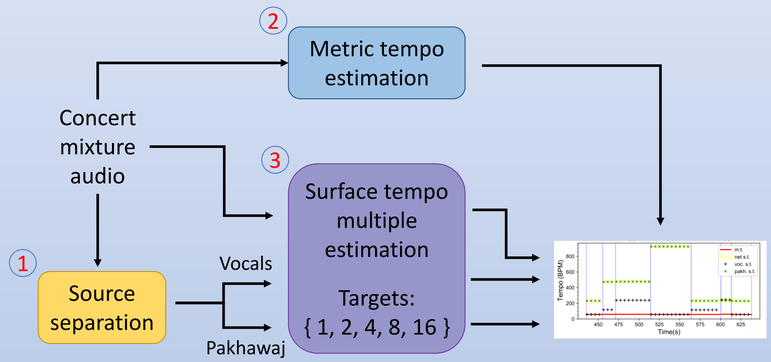


A section – during which surface tempo of neither instrument changes

Tasks

- Track metric and overall surface tempo across a concert
- Separate vocals and pakhawaj using source separation and track the surface tempi of each

System Overview



Methods

1. Source separation – Spleeter 2 stems model [2]
2. Metric tempo estimation – tempo-cnn with octave-error correction [3]
3. Surface tempo multiple estimation - modified tempo-cnn
   - Use of more dropout
   - Fewer multi-filter blocks with shorter filters

Layer	Dimensions
Input	40 x 400
(BN, Conv, ELU, DO) x3	16 x 1 x 5
AvgPool	5 x 1
BN, MF Conv, DO	12x {1x16, 1x32, 1x64, 1x96}
Concat, Conv	16 x 1 x 1
AvgPool	1 x 400
BN, DO, FC, Softmax	# output classes

Dataset

- 14 concerts (Source: Dunya corpus [4], YouTube)
- 634 sections → 1127 8-second chunks (training examples)
- Data augmentation using time-scaling and overlap between examples
- Input to tempo estimation models: log-mel spectrogram

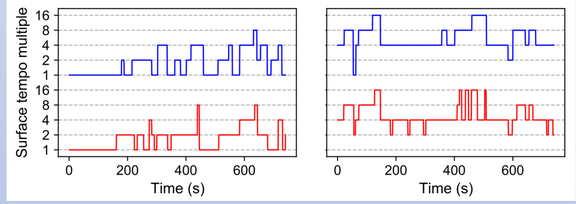
Results

S.t.m. estimation (% accuracy)

Vocals	Pakhawaj	Net
67.7	71.0	70.4

Boundary detection (± 3s tolerance)

Precision	Recall	F-score
0.40	0.53	0.45



Vocals Pakhawaj

Takeaways

- Better results on observed for 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

[1] Clayton, M., Time in Indian Music, 2000
[2] https://github.com/deezer/spleeter
[3] https://github.com/hendriks73/tempo-cnn
[4] https://dunya.compmusic.upf.edu/