Paper ID: 6126

A NOVEL 1D STATE SPACE FOR EFFICIENT MUSIC RHYTHMIC ANALYSIS

Mojtaba Heydari, Matthew McCallum, Andreas Ehmann and Zhiyao Duan

Audio Information Research Lab, University of Rochester, Rochester NY, USA
Pandora Media Inc., Oakland, CA, USA

IEEE International Conference on Acoustics, Speech & Signal Processing, Singapore , Singapore (ICASSP 2022)

Motivations / Contributions

➢ **Motivations:**

- **Music Rhythmic Analysis (MRA)** includes several essential Music Information retrieval (MIR) tasks e.g. beat, downbeat, tempo, meter, and rhythmic pattern detection.
- Many state of the art models for MRA tasks are computationally expensive and are not applicable for real life on-device or large scale industrial setups.
- Widely used Bar Pointer models are not optimal. They require many states leading to computationally expensive inferences.

➢ **Contributions:**

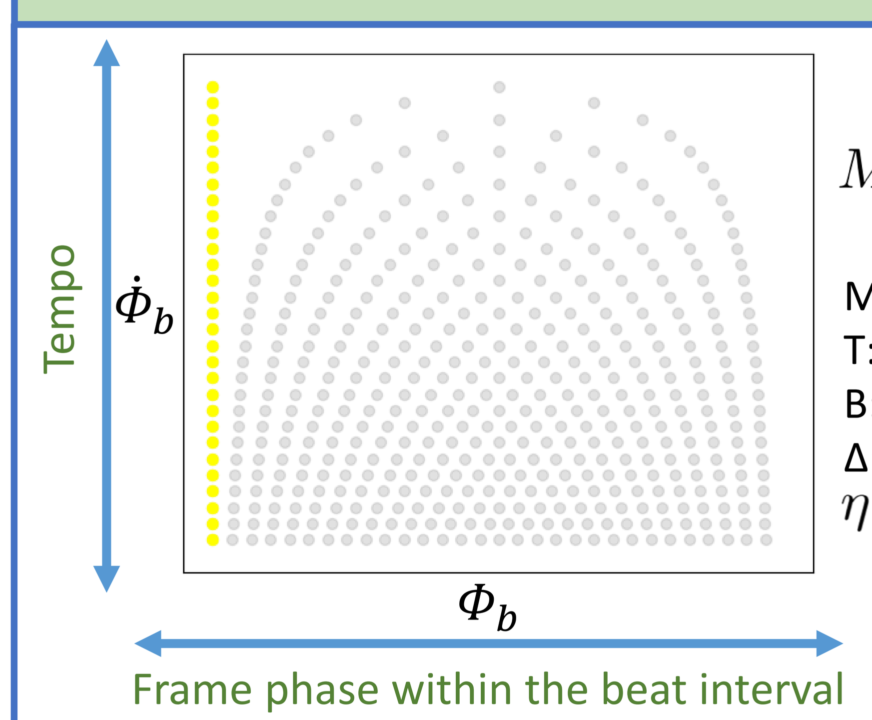
✓ We propose a novel 1D state space and transition model a super efficient alternative for widely used bar pointer models for music rhythmic analysis.

✓ We introduce a semi-Markov jump back reward inference technique.

✓ We Implement the proposed methods into a causal joint beat, downbeat, tempo and meter tracking system.

✓ Our model achieves more than 30X speed up, delivers state of the art beat tracking and comparable downbeat tracking results.

Baseline Efficient Bar Pointer Model



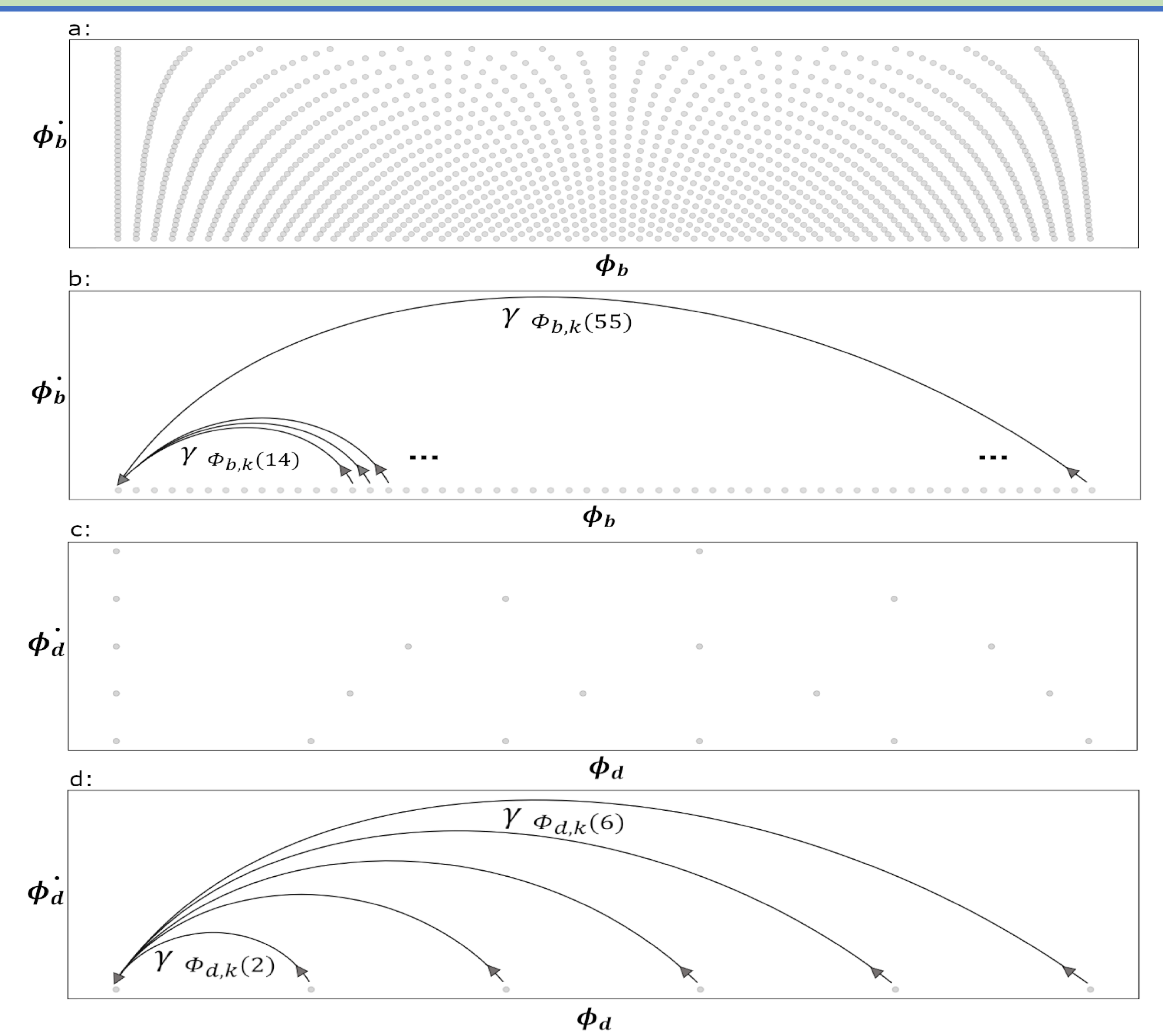
$$M(T) = Round(\frac{B \times 60}{T \times \Delta}), \quad (1)$$

M: Number of positions within a bar
T: Tempo in BPM
B: Number of beats per bar
Δ: Frame length in seconds
η: Diffuse Constant

$$\phi_k = (\phi_{k-1} + \dot{\phi}_{k-1}) \mod (\phi^{max} + 1), \quad (2)$$

$$p(\dot{\phi}_k|\dot{\phi}_{k-1}) = \begin{cases} \exp\left(-\eta\left|\frac{\dot{\phi}_k}{\dot{\phi}_{k-1}}\right|\right) & \text{if } \phi_k = 0 \\ \mathbb{1}(\dot{\phi}_k = \dot{\phi}_{k-1}) & \text{if } \phi_k > 0 \end{cases}, \quad (3)$$

Proposed 1D State Spaces



a: Efficient Beat pointer for Beat/Tempo level analysis
b: Proposed 1D state space for Beat/Tempo level analysis
c: Efficient Beat pointer for Bar/time-signature level analysis
d: Proposed 1D state space for Bar/time-signature level analysis

Jump Back Reward Inference

❖ Motion Step:

$$p(\phi_{k+1}|y_{1:k}) = \sum_{\phi_k} p(\phi_{k+1}|\phi_k)p(\phi_k|y_{1:k}), \quad (4)$$

$$p(\phi_{k+1}|\phi_k) = \begin{cases} \gamma(\phi_k) & \text{if } \phi_{k+1} = 1 \\ 1-\gamma(\phi_k) & \text{if } \phi_{k+1} = \phi_k + 1 \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

❖ Correction Step:

$$p(\phi_{k+1}|y_{1:k+1}) = \frac{1}{Z_{k+1}} p(y_{k+1}|\phi_{k+1})p(\phi_{k+1}|y_{1:k}), \quad (6)$$

$$p(y_{k+1}|\phi_{k+1}) = \begin{cases} b_{k+1} & \text{if } \phi_{k+1} = 1 \text{ and } b_{k+1} \geq T \\ \epsilon & \text{otherwise} \end{cases}, \quad (7)$$

Where b_{k+1} is the beat activation at frame k+1 and ϵ is a small constant.

Jump Update Using The Reward Function

❖ Jump back weights' Update :

$$\gamma(\phi_{k+1}) = \lambda\gamma(\phi_k) + (1-\lambda)\Gamma(\phi_{k+1}), \quad (8)$$

Where $\gamma(\phi_{k+1})$ and $\Gamma(\phi_{k+1})$ are the jump back weight and state reward at frame k+1 and λ is the forget factor.

❖ Reward function :

$$\Gamma(\phi_{k+1}) = \begin{cases} p(\phi_{k+1}|y_{1:k}) - p(\phi_{k+1}|y_{1:k+1}) & \text{if } b_{k+1} \geq T \\ -(p(\phi_k|y_{1:k}) - p(\phi_{k+1}|y_{1:k})) & \text{if } b_{k+1} < T \text{ and } \phi_{k+1} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Evaluation

Table 1. Performance and speed comparison of several online beat and downbeat tracking models on the GTZAN.

Method	*F-Measure Beats*	*F-Measure Downbeats*	*Comp. Time (Seconds)*
Aubio [19]	57.09	—	0.1
BeatNet [8]	75.44	<u>46.49</u>	8.87
Böck ACF [22]	64.63	—	7.01
Böck FF [20]	74.18	—	2.19
DLB [5]	73.77	—	21.30
IBT [6]	68.99	—	4.89
1D state space	<u>76.48</u>	42.57	0.29

Advantages :

- ☑ General approach (Applicable to multiple rhythmic analysis tasks)
- ☑ More than 30X computationally efficient than the previous SOTA model
- ☑ Robust against tempo fluctuation
- ☑ Delivering SOFA Beat tracking results

Drawbacks :

- ☒ Worse downbeat tracking results
- ☒ Not the fastest model

References

1. N. Whiteley, A. Cemgil, and S. Godsill, "Bayesian modelling of temporal structure in musical audio," in Proc. of the 7th Intl. Conf. on Music Information Retrieval (ISMIR), 2006..
2. F. Krebs, S. Bock, and G. Widmer, "An efficient state- ¨ space model for joint tempo and meter tracking," in Proc. of the 16th Intl. Conf. on Music Information Retrieval (ISMIR), 2015, pp. 72–78.
3. M. Heydari, F. Cwitkowitz, and Z. Duan, "BeatNet: CRNN and particle filtering for online joint beat downbeat and meter tracking," in Proc. of the 22th Intl. Conf. on Music Information Retrieval (ISMIR), 2021.
4. A. Cont, "A coupled duration-focused architecture for real-time music-to-score alignment," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 6, pp. 974–987, 2009.
5. P. M. Brossier, "Automatic annotation of musical audio for interactive applications," pp. 58–102. Queen Marry University, London, UK, August 2006.