MIREX 2026 SUBMISSION – MUSIC STRUCTURE ANALYSIS (MSA) TRACK

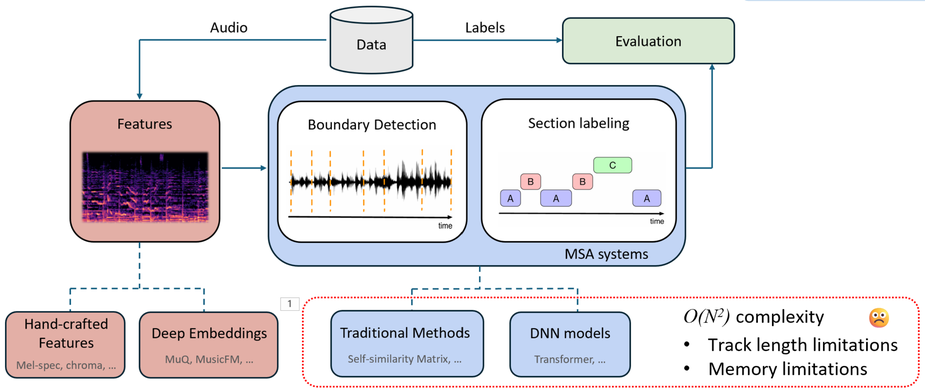
SongMamba: Efficient Music Structure Analysis via Bi-directional State-Space Modeling

Yuan Wang[1,3], Shenyang Xu[2], Yan Han[1,3], Wei Li[1,3,*]

[1] Central Conservatory of Music, China
[2] Xinghai Conservatory of Music, China
[3] Fudan University, China
[*] Corresponding Author

What is MSA

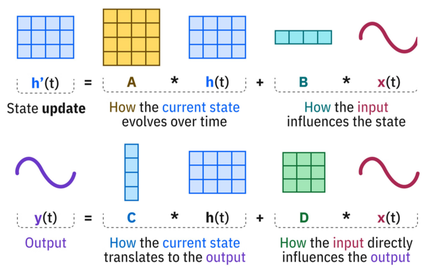
0 Backgrounds



What is State-Space Models

It takes an input sequence $x(t)$,
then maps it to a latent state representation $h(t)$,
finally predict an output sequence $y(t)$.

- **State update,** $h'(t) = A * h(t) + B * x(t)$
- **Sequence update,** $y(t) = C * h(t) + D * x(t)$



cr. newsletter.maartengrootendorst.com/p/a-visual-guide-to-mamba-and-state

😊 $O(N)$ complexity
- Long context
- Efficient

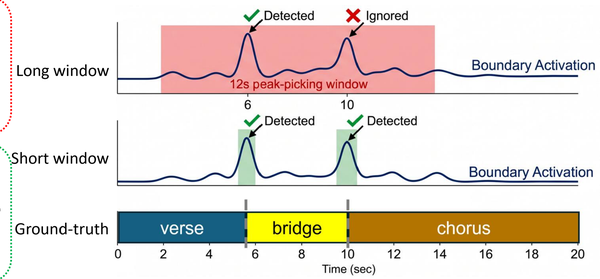
Temporal Continuity
- Sampling-invariant

Post-processing

2 Findings

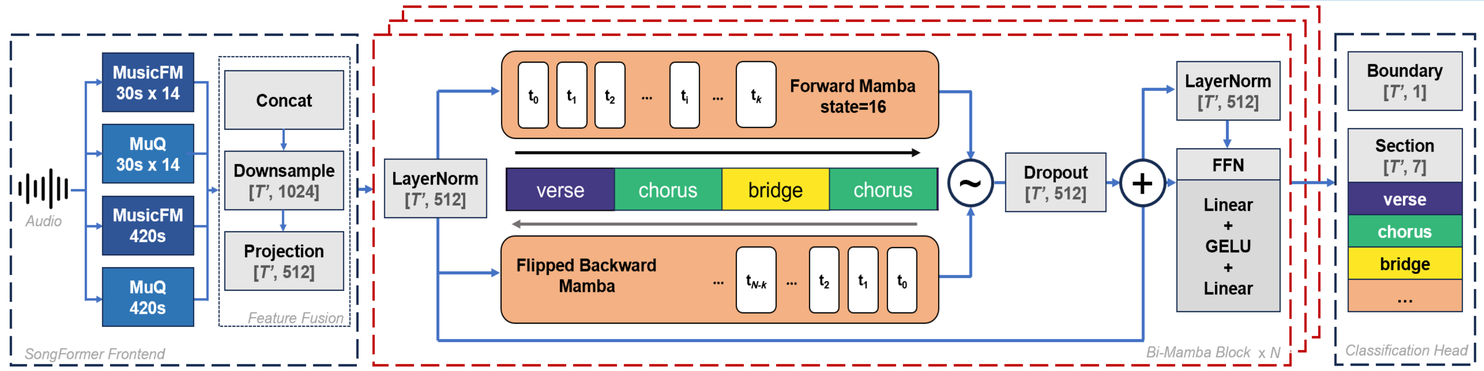
The default 12-second flaw: 😟
Prior MSAs universally by default use 12s peak picking window However, in HarmonixSet, over 20% of music segments are shorter than 12 seconds.

Domain Grounding: 😊
Here we propose to use 1.4-second window, derived from musical tempo (2 beats@85BPM, or 4 beats@170 BPM)



SongMamba Architecture

1 Novel System



Experiments & Results

3 MIREX

Dataset:
Following the strict protocol, we use **SongFormHX** only for training and evaluation. It consists of:
512 training samples, 200 validation samples, and 200 refined test samples.

Configuration:
To obtain the best practice of SongMamba, we conducted ablation studies in terms of bi-directional fusion, Mamba block depth, and state dimensions..

Configuration	$F_{.5}$	F_3	PWF
Depth: 2 blocks	.7380	.8409	.7717
Depth: 3 blocks	.7366	.8385	**.7851**
Depth: 4 blocks	**.7419**	**.8470**	.7832
State: 16	**.7412**	**.8411**	.7748
State: 32	.7113	.8253	.7651
State: 64	.7324	.8401	**.7782**

Model	$F_{.5}$	F_3	PWF
SongMamba, forward	.7045	.8014	.7557
SongMamba, sum	.7402	.8395	.7788
SongMamba, gate	**.7419**	**.8470**	**.7832**
SongMamba, concat	.7364	.8424	.7742

Best Practice
Bi-direction fusion:
- sum
- gate
- concat

Block depths:
- 2
- 3
- 4

State dimension
- 16
- 32
- 64

Unified MIREX Evaluation

Group	ACC	HR.5F	HR3F
SCALE	0.799	0.728	0.828
songmamba	0.776	**0.739**	**0.840**
Baseline: SongFormer	**0.802**	0.704	0.798
Baseline: MusicFM (MIREX 2025, reported)	0.705	0.644	0.710

music-ir.org/mirex/wiki/2026:Music_Structure_Analysis_Results

Experiments & Results

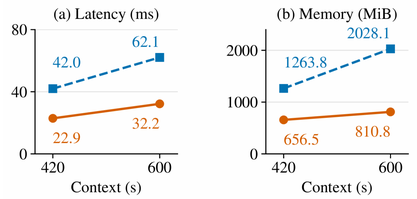
4 JSD

Dataset:
Extended experiments are conducted on the Jazz Structure Dataset, which features **longer track durations, wide length variance, and diverse tempi**. JSD consists of:
244 training samples, 28 validation samples, and 68 test samples. The average track duration is **~320 seconds**, about 1.5 times the average duration in HarmonixSet.

Context Experiments:
To thoroughly investigate SongMamba in long tracks, we use two context embedding frontend, namely **420s and 600s**. The 420s context follows strictly SongFormer, while the 600s context expand the pipeline to 600s. The MSA systems uses best practice in MIREX experiment. Results are shown below.

Model	Context	$F_{.5}$	F_3	PWF	S_f
SongFormer	420	.4301	.6131	.7764	.7939
SongMamba	420	.4278	.6179	.7829	.7968
SongFormer	600	**.4508**	.6015	.7776	.8001
SongMamba	600	.4490	**.6236**	**.7942**	**.8087**

Better Performance 😊
- Longer the context, better the performance.
- Outperforms SongFormer, with large headroom



Less consuming 😊

32.2 vs 62.1 ms
~ 2x faster inference

810.8 vs 2028.1 MiB
~ 60% less GPU memory

Conclusions

5 Takeaway

We present SongMamba, an efficient MSA system with linear complexity. The results on MIREX and JSD demonstrate a favorable balance of performance, efficiency and scalability, highlighting the potential of State-Space Models in complex MSA scenarios.

MIREX Page: *https://music-ir.org/mirex/wiki/2026:Music_Structure_Analysis*

Contact: Yuan Wang *yvan_w@126.com*