PhraseSplitGen: Two-Stage Piano Continuation with Development and Ending Awareness for MIREX 2026

MIREX 2026 Symbolic Music Generation — Piano Music Continuation

Di Lu, Xinran Zhang, Qingwen Zhou, Xiaobing Li, Maosong Sun.

ISMIR 2026
LBD / MIREX 2026
Symbolic Music Generation

1 Task

MIREX 2026 Task

Symbolic music generation / piano continuation.

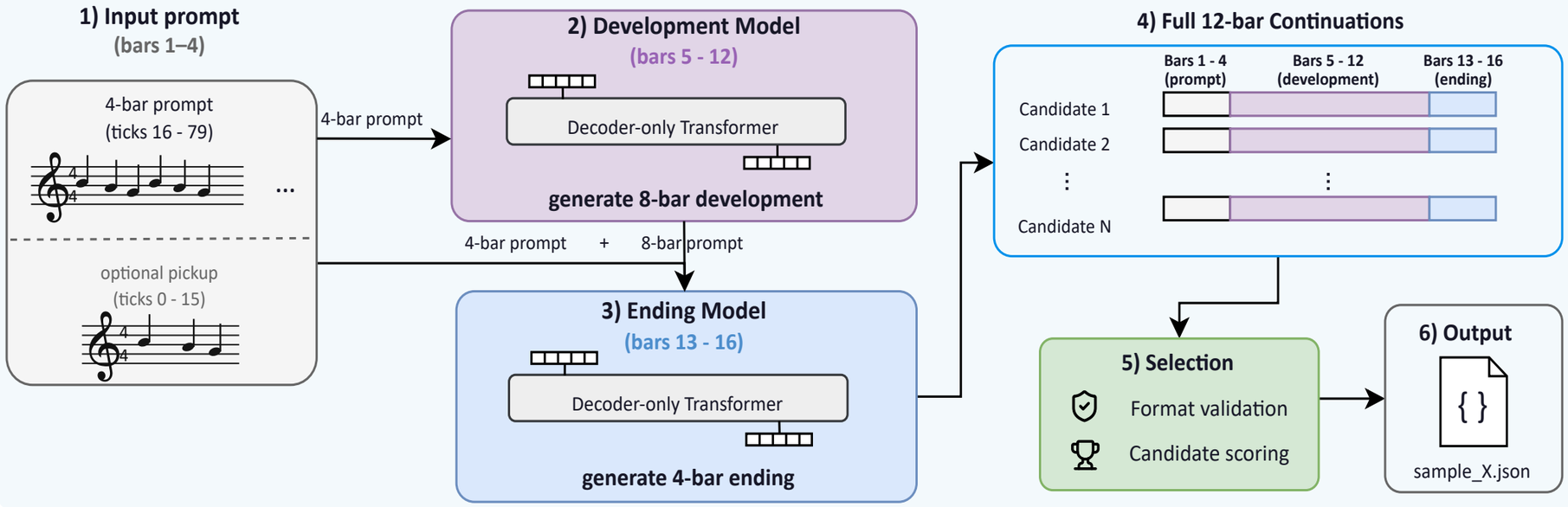
Goal

Generate a 12-bar piano continuation from a 4-bar piano prompt.

Evaluation Dimensions

- **Coherency** — Musical consistency
- **Structure** — Long-term organization
- **Creativity** — Novelty and inventiveness
- **Musicality** — Overall musical quality

2 System Overview: Two-Stage Development- and Ending-Aware Generation



3 Training Data

Classical piano data (score-derived)

- PDMX (Piano e-Competition Dataset)
- ASAP (Automatic Score Alignment Dataset)
- DCML (Der Deutsche Musikarchiv)
- KernScores
- Humdrum keyboard collection

Pop piano data (filtered)

- POP909
- RWC (RWC Music Database)
- Lakh Piano (from Lakh MIDI Dataset)

Data preparation

- Bar-aligned 16-bar windows
- Split into development subset (bars 5–12) and ending subset (bars 13–16)
- Used for training specialized models

4 Method Summary

Base continuation model
Decoder-only Transformer for symbolic piano generation.

Two specialized checkpoints
- Development model for bars 5–12.
- Ending model for bars 13–16.

Candidate sampling
Generate multiple full 12-bar candidates.

Musical reranking
Format validation and music-aware scoring.

Diversity-aware selection
Choose a high-quality and diverse continuation.

5 Official MIREX 2026 Results

Method	Coherency (↑)	Structure (↑)	Creativity (↑)	Musicality (↑)
DAMN	**3.68 ± 0.16** [a]	**3.39 ± 0.18** [a]	3.18 ± 0.16 [a]	**3.47 ± 0.14** [a]
PhraseSplitGen	2.95 ± 0.18 [b]	2.74 ± 0.15 [b]	2.89 ± 0.15 [a]	2.61 ± 0.14 [b]
MuseCoco (BL-1)	2.79 ± 0.17 [b]	2.76 ± 0.15 [b]	3.42 ± 0.12 [a]	2.68 ± 0.16 [b]
Anticipatory Music Transformer (BL-2)	**3.47 ± 0.20** [a]	**3.34 ± 0.16** [a]	3.34 ± 0.16 [a]	**3.61 ± 0.17** [a]

Note: Different superscript groups indicate significant differences after Holm-corrected Wilcoxon tests. PhraseSplitGen shares the same group as MuseCoco on Coherency, Structure, and Musicality, while all systems are in the same group for Creativity.

6 Observations

- PhraseSplitGen explicitly models phrase development and phrase ending.
- Official evaluation shows competitiveness with MuseCoco baseline on several dimensions.
- Results suggest room for improvement in data scale and stylistic coverage.

7 Takeaway

PhraseSplitGen frames piano continuation as two complementary subproblems: development and closure.

The system combines structure-aware data curation, specialized generation, and candidate post-selection.