SCALE

MIREX 2026
Music Structure Analysis

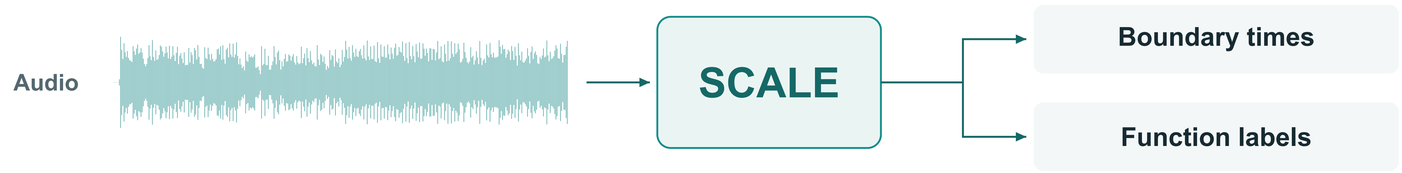
Structural Context from Audio and Lyric Evidence for Music Structure Analysis Using the New SheetSage-MSA Dataset

Hongbing Li[1] **Yaolong Ju**[2,*]
1 Queen Mary University of London, UK 2 Great Bay University, China * Corresponding author

Code

01 Introduction

Goal: split a song at its section boundaries and assign a function label to each section.

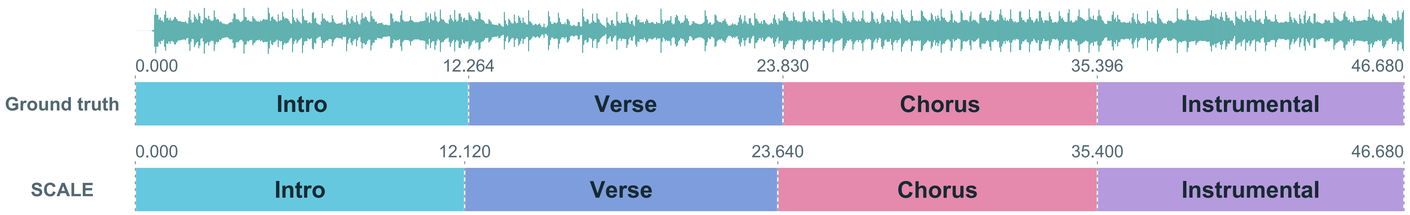


All seven output labels

Intro · Verse · Chorus · Bridge · Instrumental (inst) · Outro · Silence

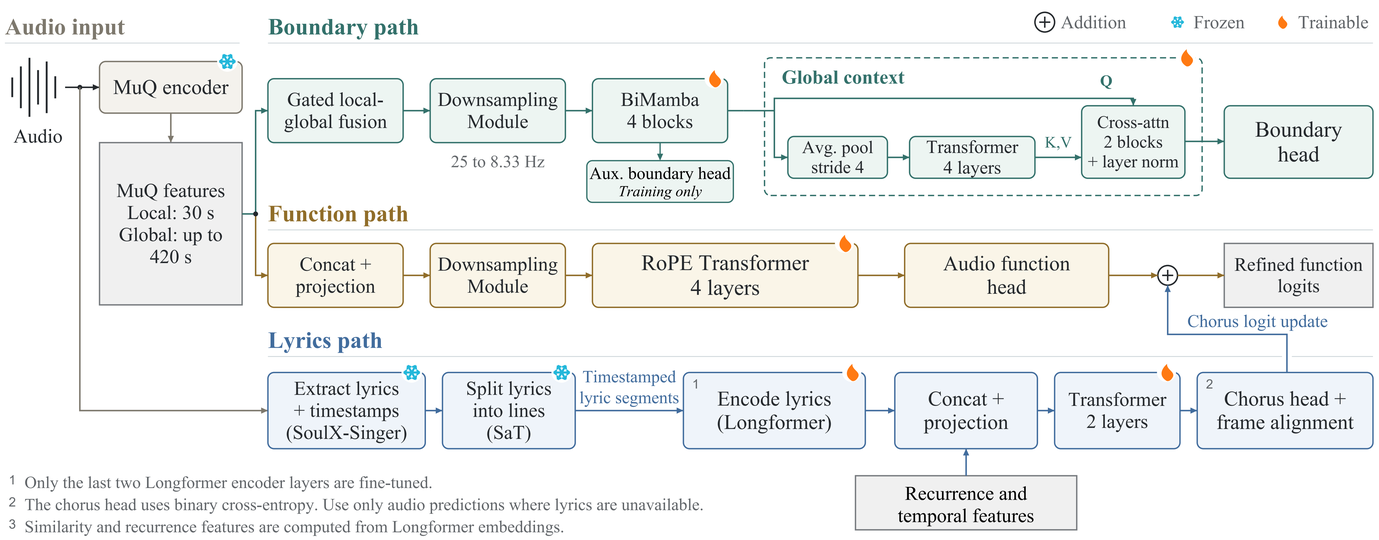
Example: Someday - Tegan and Sara

First 46.680 s; ground truth and SCALE prediction



03 Model: SCALE

Separate boundary and function pathways; lyric recurrence refines chorus predictions.



[1] Only the last two Longformer encoder layers are fine-tuned.
[2] The chorus head uses binary cross-entropy. Use only audio predictions where lyrics are unavailable.
[3] Similarity and recurrence features are computed from Longformer embeddings.

02 Datasets

HarmonixSet: fully labeled training and validation data.
SheetSage-MSA (new): additional, partially labeled training data.

SheetSage-MSA — **13,880** cleaned annotations · **9,010** songs · **82.5 h** annotated audio

All examples below: 42 by Coldplay.

Alignment and Function Labels

Group excerpt annotations
Match excerpts by artist and title.
Group their section annotations.
Example annotations (IDs simplified)
Song: Coldplay - 42
`coldplay_42_id1 (Verse)`
`coldplay_42_id2 (Solo)`
`coldplay_42_id3 (Pre-Outro)`
Grouped function labels for this song.

Update excerpt boundaries
Replace with refined beat times.
Example: Verse boundaries (s)
`alignment.user.times`
40.810-94.910
`alignment.refined.times`
40.964, ..., 95.044
Updated: 40.964-95.044 s

Attach refined times to grouped labels

Clean labels
Map to 7 classes; deduplicate.
Example mappings

Original		Result
Verse	→	*verse*
Solo	→	*inst*
Pre-Outro	→	*outro*

Export annotations
One song, partially labeled.
Example output

start	end	label
40.964	95.044	verse
122.660	150.280	inst
164.050	177.470	outro

Times in s; gaps unlabeled.

Unlabeled regions are masked; pre-chorus maps to verse for evaluation.

04 Results

SongFormBench-HX

Model	ACC	HR.5F	HR3F
All-In-One	74.0	59.6	73.0
LinkSeg	78.0	63.0	76.2
SongFormer	80.7	69.6	78.0
SCALE (HX + SheetSage-MSA)	**81.6**	**73.1**	**82.2**

Scores in %; bold: best in each table.

SCALE outperforms prior models across all three metrics, achieving state-of-the-art results.

MIREX 2026 | Official HarmonixSet test results

Model	ACC	HR.5F	HR3F
MusicFM (2025, reported)	70.5	64.4	71.0
SongFormer	**80.2**	70.4	79.8
SongMamba	77.6	**73.9**	**84.0**
SCALE	79.9	72.8	82.8

Scores in %; bold: best in each table.

High boundary scores with stable label accuracy suggest robust performance across evaluation settings.

Hongbing Li: hongbingli96@gmail.com
Yaolong Ju: juyaolong@gbu.edu.cn