MIREX 2026

MUSIC PERFORMANCE DIFFICULTY PREDICTION

A Transcription-Based System for MIREX 2026 Music Performance Difficulty Prediction

Qinyuan Liu · Zheyu Jiang · Xiaoyu Guan
KaraOrchee, Inc.



KaraOrchee
Intelligence Behind Music

2ND PLACE · **0.5864** Kendall's τ-c · **3.70 s** per song · **6.75×** faster than first-ranked[1] · **761 pieces** Official hidden test

01 FROM SOUND TO SKILL

Can a recording reveal how difficult a piano piece is to perform? We combine detected notes with the audio information retained inside the transcription model.

ONE TRANSCRIPTION PASS. TWO REPRESENTATIONS.

Solo piano audio
Mono conversion → resampling → loudness normalization

↓

Piano transcription model
MAESTRO v3.0.0 training split · 962 performances

↓

Note sequence N
Onset · offset · pitch · velocity

Audio embeddings E
Extracted in the same pass

↓

s = piece-level note statistics

ē = embedding summary

Gradient-boosted trees
Inputs: s, ē

Note-sequence Transformer
Inputs: N, s, ē

Audio-embedding Transformer
Inputs: E, s

THREE SCORES → ONE MEAN

ROBUSTNESS Unseen soundfonts: τ-c drops of 0.001 / 0.003. Hidden Voices: τ-c = 0.759 (34 pieces).
COMPUTE RTX 5090: ~100 s audio/s; ~7 s model load. Peak GPU memory: 2.3 GB (<5 min); 2.7 GB (87 min). These are paper benchmarks, separate from official runtime.

02 THE RANKING ENGINE

THE ENSEMBLE / CONTINUOUS OUTPUT

$$\hat{d}(x) = \frac{g(s, \bar{e}) + f_N(N, s, \bar{e}) + f_E(E, s)}{3}$$

x: input audio; N = {(onset$_i$, offset$_i$, pitch$_i$, velocity$_i$)}; E: audio-embedding sequence; s: piece-level note statistics; ē: piece-level embedding summary.

g, f_N, f_E: instance-averaged tree, note-Transformer and embedding-Transformer predictions, respectively.

7,254 / 7,901

USABLE PSYLLABUS RECORDINGS / LISTED ENTRIES

Scorers train only on PSyllabus, including our audio renderings of its transcribed MIDI files. Targets are grades 0–10; tree targets may exceed 10 for source grades above that level.

RANK FIRST; MAP TO GRADES ONLY FOR EVALUATION

Continuous scores preserve ordering for Kendall's τ-c. Categorical metrics use thresholds fitted to cross-validated training predictions.

COMPOSER-DISJOINT DEVELOPMENT

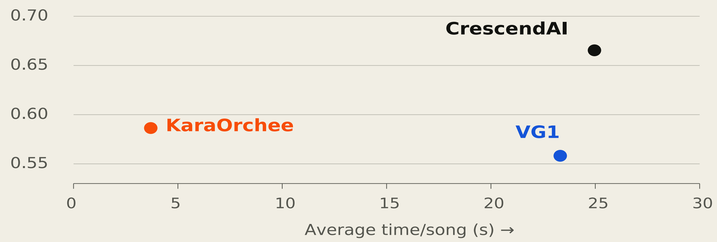
Five-fold development after withholding 15% of composers (184). Withheld audio: 1,393 pieces / 165 composers; τ-c = 0.789 for recording + rendering averages. Submitted models use all 7,254 entries.

DETERMINISTIC FALLBACK → OUTPUT 0
Silence · fewer than five detected notes · pipeline failure

03 EVIDENCE AT SCALE

A / OFFICIAL MIREX 2026 HIDDEN TEST

Ranking quality × processing time



SYSTEM	τ-c	TIME/SONG
CrescendAI	0.6654	24.96 s
KaraOrchee	**0.5864**	**3.70 s**
VG1	0.5582	23.32 s

[1] 24.96 / 3.70 = 6.75× (rounded); official averages.

B / PUBLISHED PSYLLABUS — NOT MIREX HIDDEN TEST

Five published splits · mean ± standard deviation

METRIC	KaraOrchee	MM baseline
Kendall's τ-c ↑	0.820 ± 0.004	0.775 ± 0.004
Macro-MSE ↓	1.427 ± 0.085	1.766 ± 0.090
Balanced accuracy ↑	0.401 ± 0.009	0.377 ± 0.016
Acc±1 ↑	0.826 ± 0.005	0.787 ± 0.006
Spearman's ρ ↑	0.919 ± 0.004	0.895 ± 0.004

τ-c / COMMON ZERO-BASED SCALE



MM: early fusion of CQT and piano roll [2].

LIMITATIONS Published splits are not composer-disjoint; ~80% of test pieces appeared in development. Our scorers use train + validation pieces; MM uses validation for early stopping. MIDI renderings approximate score renderings; multiple components changed together, so this is not a causal ablation.

[1] Liu, Jiang & Guan. MIREX 2026 technical report.
[2] Ramoneda et al. PSyllabus. IEEE TASLP 33, 2025.