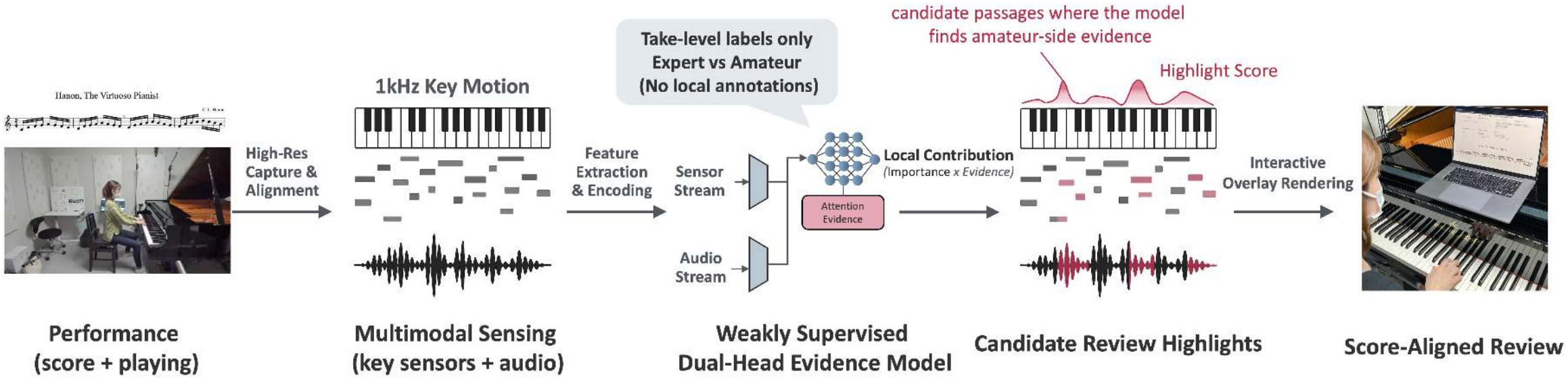


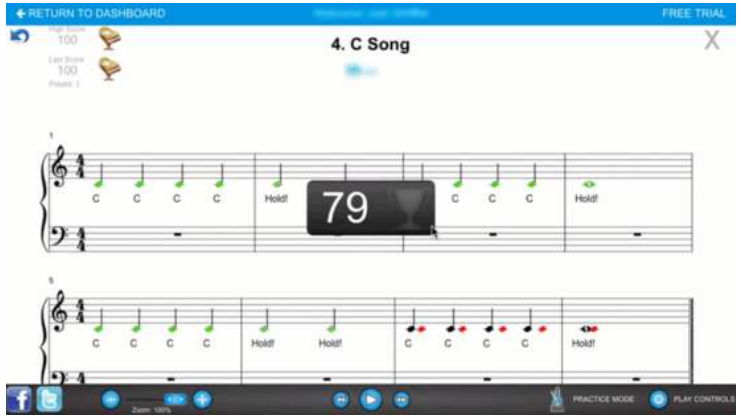
Profy: Interpretable Visualization of Expertise-Dependent Motor Skills Toward Supporting Piano Practice

Kazuki Kawamura, Fujiki Nakamura, Hayato Nishioka,
Momoko Shioki, Shinichi Furuya, Jun Rekimoto

UTokyo Sony CSL

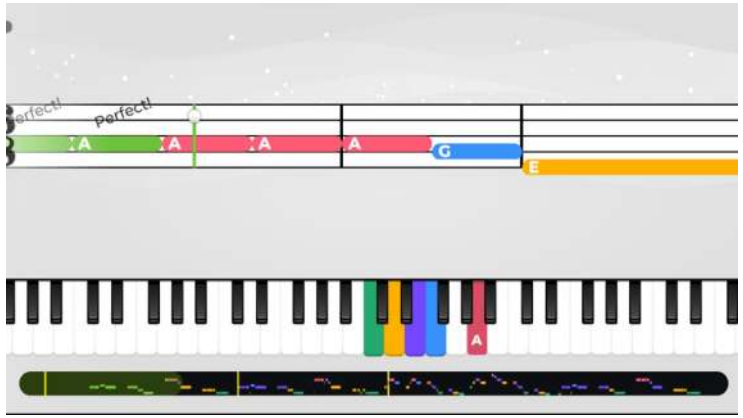
DIS 2026

Existing tools support solo piano practice



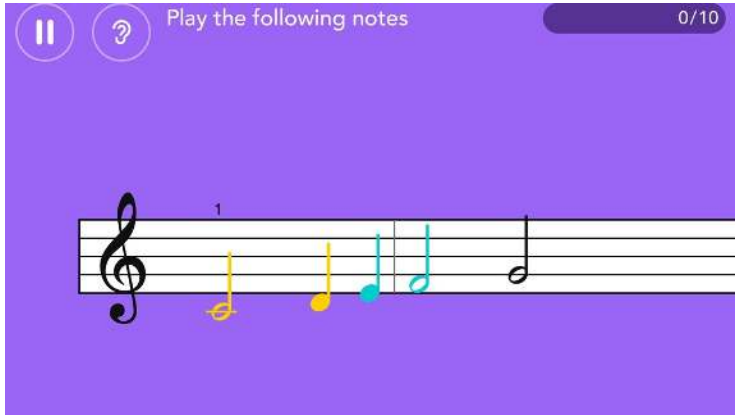
Performance scoring apps

Note and rhythm accuracy score



Practice apps

Pitch and timing feedback



Guided exercises

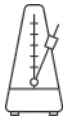
Correct / wrong note feedback

Practice systems can already show where the played sound differs from the written score or a reference performance.

Morsi et al., "Sounds Out of Place?: Score-Independent Detection of Conspicuous Mistakes in Piano Performances," ISMIR, 2023.
Jeon et al., "Identifying Critical Segments Affecting Piano Performance Evaluation," CIKM, 2025.
Zhang et al., "LLaQo: Towards a Query-Based Coach in Expressive Music Performance Assessment," ICASSP, 2025.

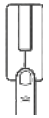
Piano expression is more than correct notes





Timing / coordination

Micro-timing, hand synchrony, and fingering stability shape how the same notes feel.



Touch / articulation

Legato, staccato, key release, and keystroke depth change the perceived expression.

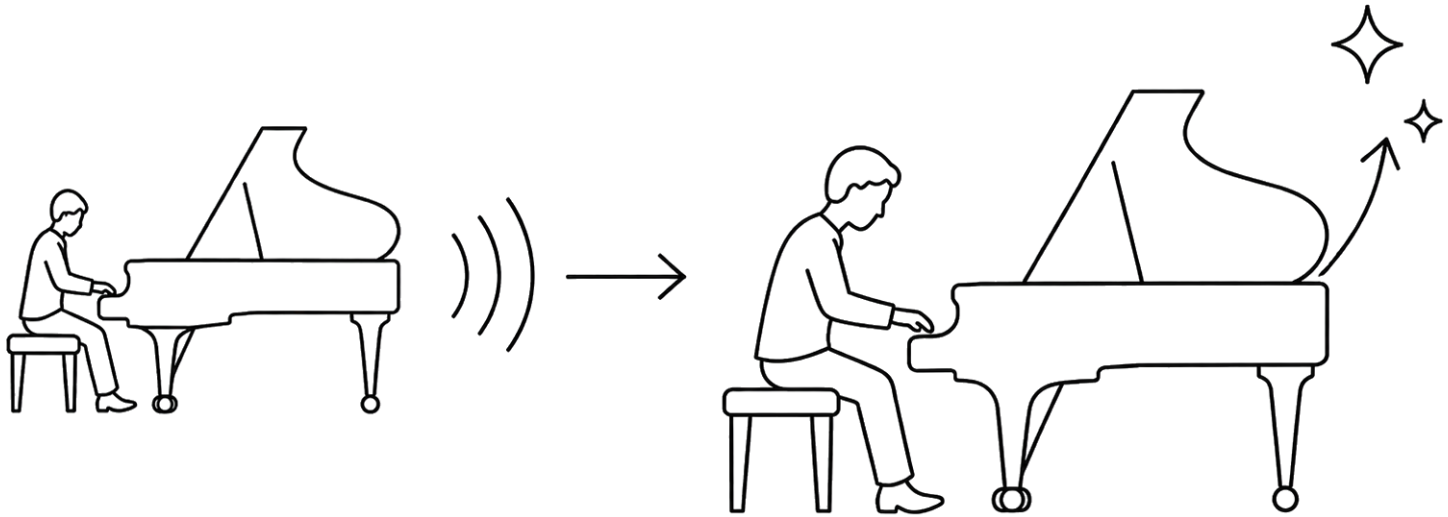


Sound / expression

Dynamics, tone color, voice balance, and phrasing emerge from motor control.

Lerch et al., "An Interdisciplinary Review of Music Performance Analysis," TISMIR, 2020.
Oku and Furuya, "Noncontact and High-Precision Sensing System for Piano Keys Identified Fingerprints of Virtuosity," Sensors, 2022.
Kuromiya et al., "Motor Origins of Timbre in Piano Performance," PNAS, 2025.

Advanced amateurs imitate professional expression



With a professional sound model (modeling), piano majors show short-term gains in audible expression—timbre, rubato, articulation, and dynamics.

Bae, "Independent practice approaches for expressive piano performance: modeling, structural understanding, and narrative imagery," Frontiers in Psychology, 2026.

Three gaps between evaluation and practice



Scores, not practice targets

For amateur pianists who can already play the written notes, tools rarely identify the few seconds to replay, loop, and practice.



Notes, not expression

They catch pitch, rhythm, and tempo errors, but often miss expressive differences—micro-timing, articulation, dynamics—that distinguish expert from amateur performances.



Hard to verbalize

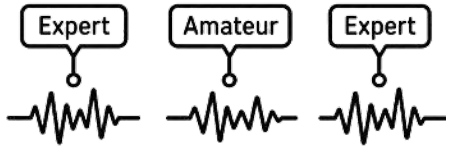
Even for expert pianists, it can be difficult to explain, in words, which moments sound less expressive and why.

Profy visualizes replayable expression gaps



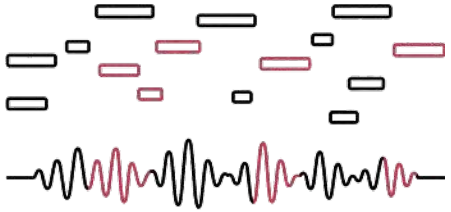
Input

Key motion + audio from one learner's performance
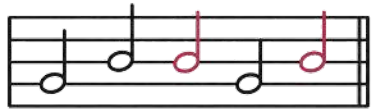


Weak label

Expert / amateur labels at the take level



Evidence

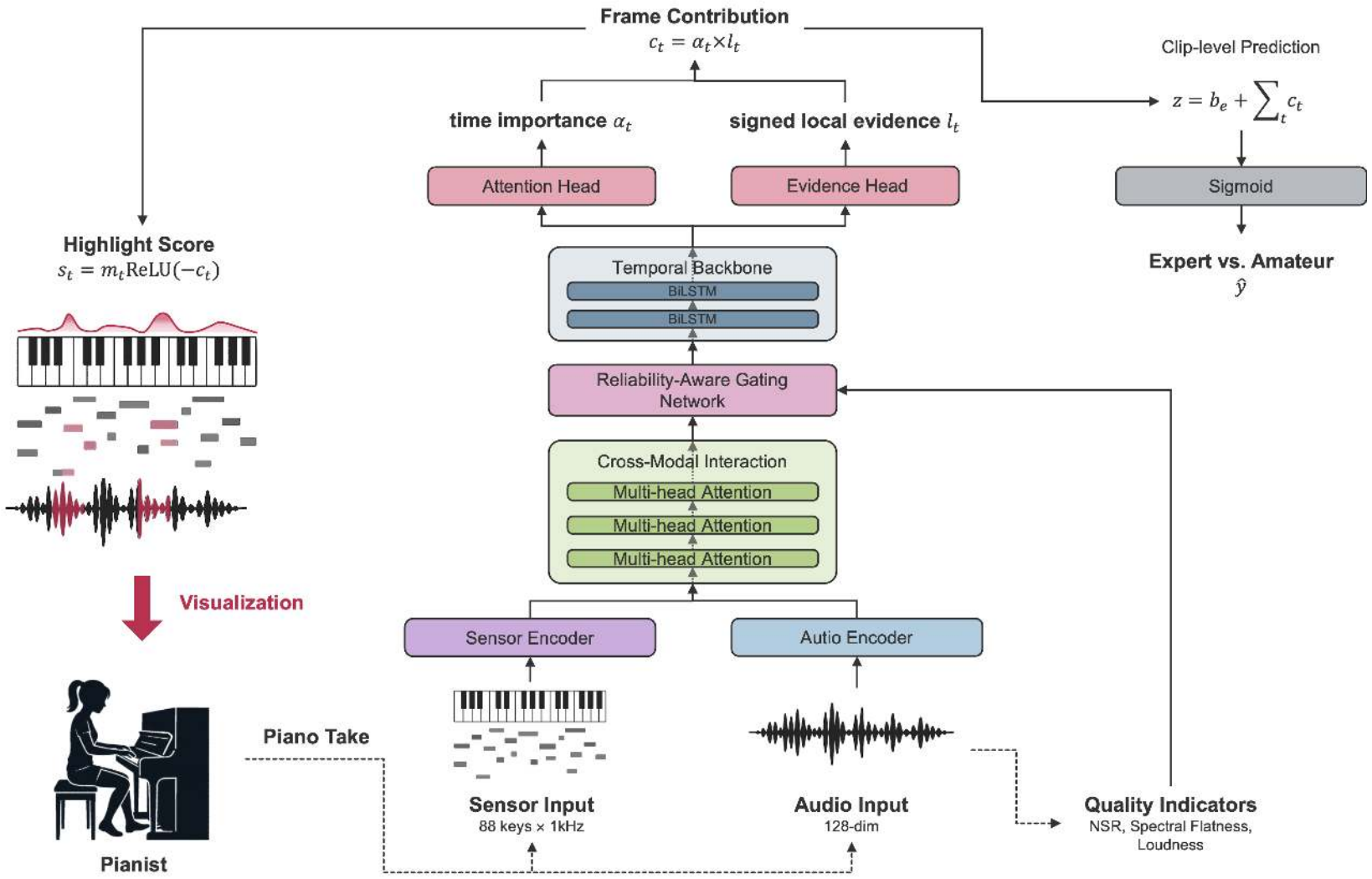
Time-local evidence for the amateur side



Output

Replayable highlights on the timeline and written score

Weak labels reveal time-local evidence



Highlights are direction-aware, not just attention

Score each moment by how much the model relies on it and which side it supports — highlight only moments that are important and amateur-leaning.

α_t : how much?

time importance: where the model attends

ℓ_t : which side?

signed evidence: expert (+) or amateur (−)

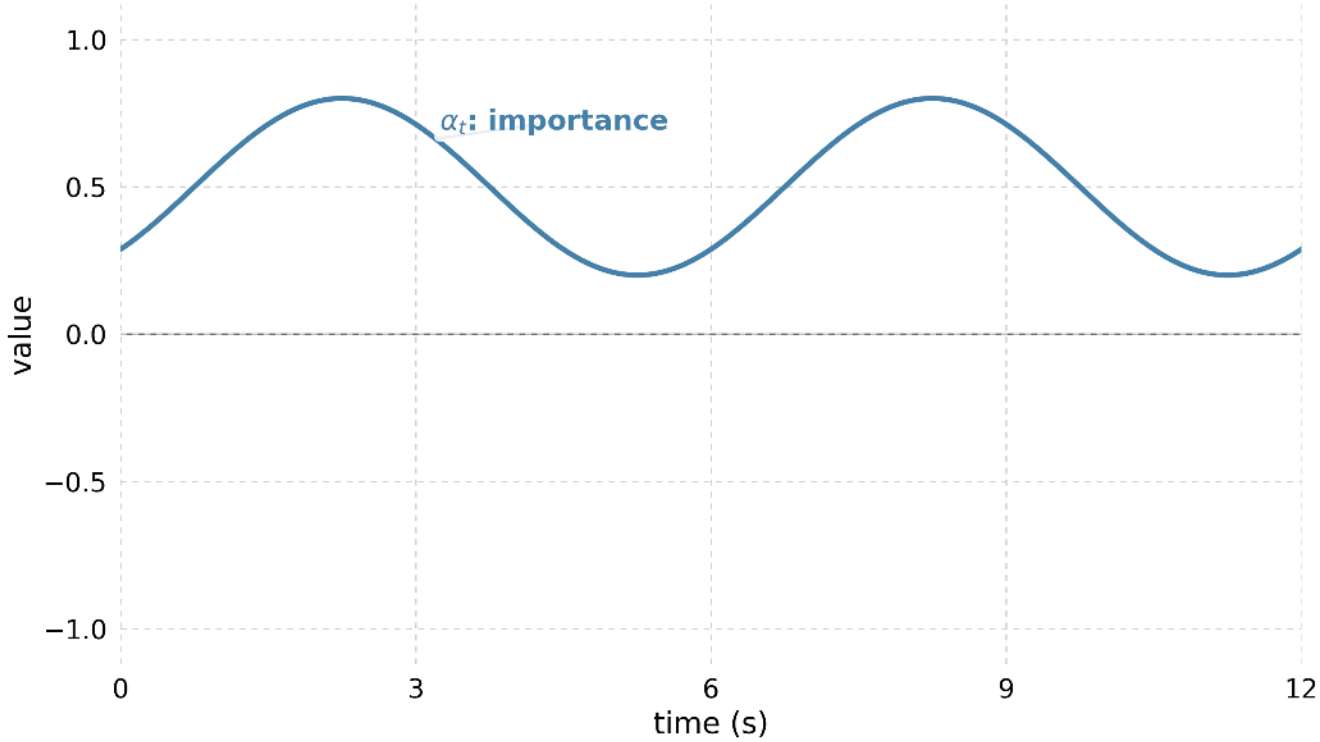


c_t : contribution

how much × which side: $c_t = \alpha_t \cdot \ell_t$

s_t : highlight

keep amateur-leaning moments: $s_t = \mathrm{ReLU}(-c_t)$



Highlights are direction-aware, not just attention

Score each moment by how much the model relies on it and which side it supports — highlight only moments that are important and amateur-leaning.

α_t : how much?

time importance: where the model attends

ℓ_t : which side?

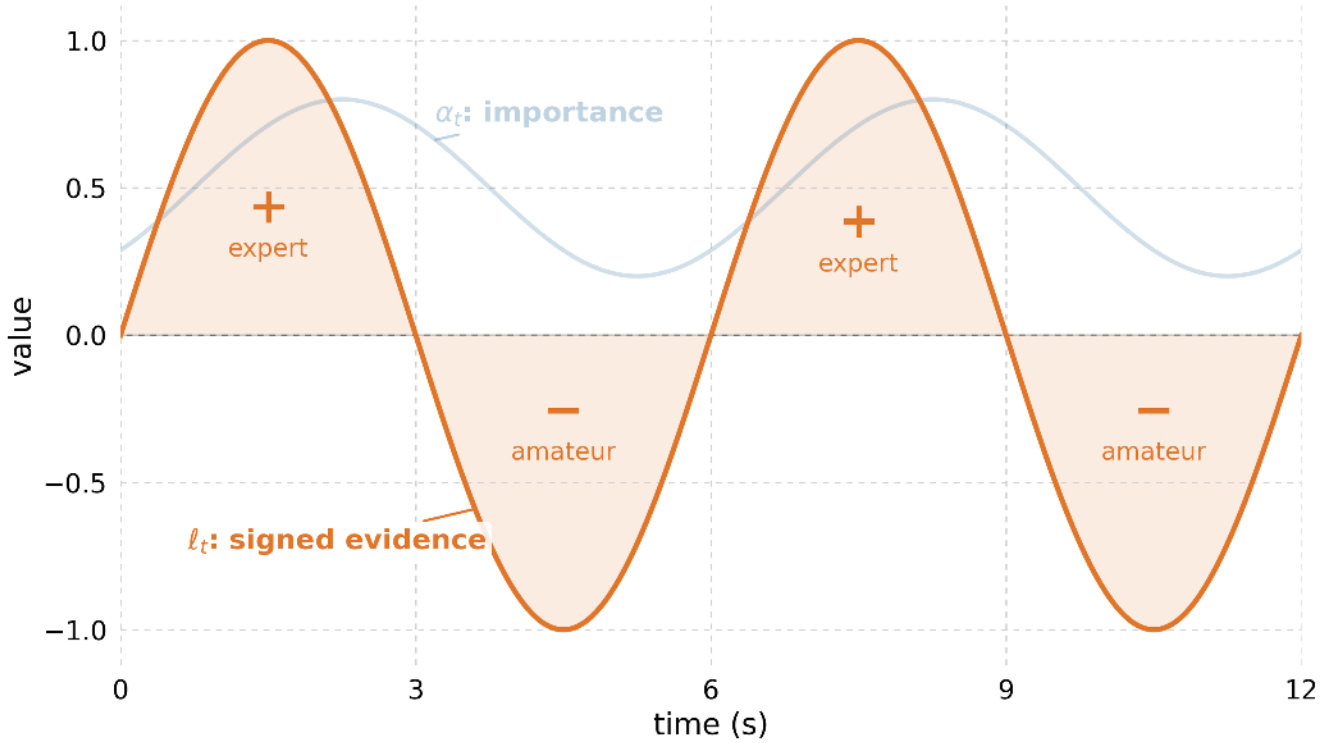
signed evidence: expert (+) or amateur (−)

c_t : contribution

how much × which side: $c_t = \alpha_t \cdot \ell_t$

s_t : highlight

keep amateur-leaning moments: $s_t = \mathrm{ReLU}(-c_t)$



Highlights are direction-aware, not just attention

Score each moment by how much the model relies on it and which side it supports — highlight only moments that are important and amateur-leaning.

α_t : how much?

time importance: where the model attends

ℓ_t : which side?

signed evidence: expert (+) or amateur (−)

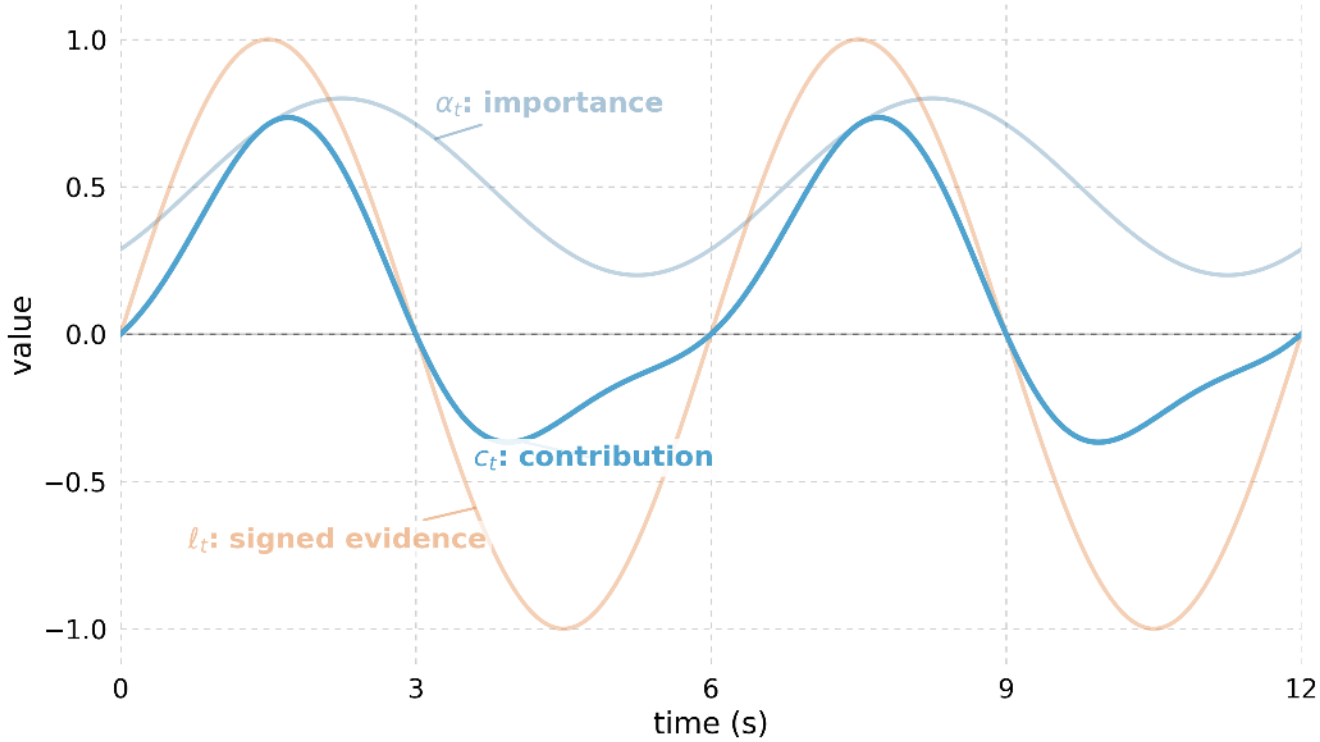


c_t : contribution

how much × which side: $c_t = \alpha_t \cdot \ell_t$

s_t : highlight

keep amateur-leaning moments: $s_t = \mathrm{ReLU}(-c_t)$



Highlights are direction-aware, not just attention

Score each moment by how much the model relies on it and which side it supports — highlight only moments that are important and amateur-leaning.

α_t : how much?

time importance: where the model attends

ℓ_t : which side?

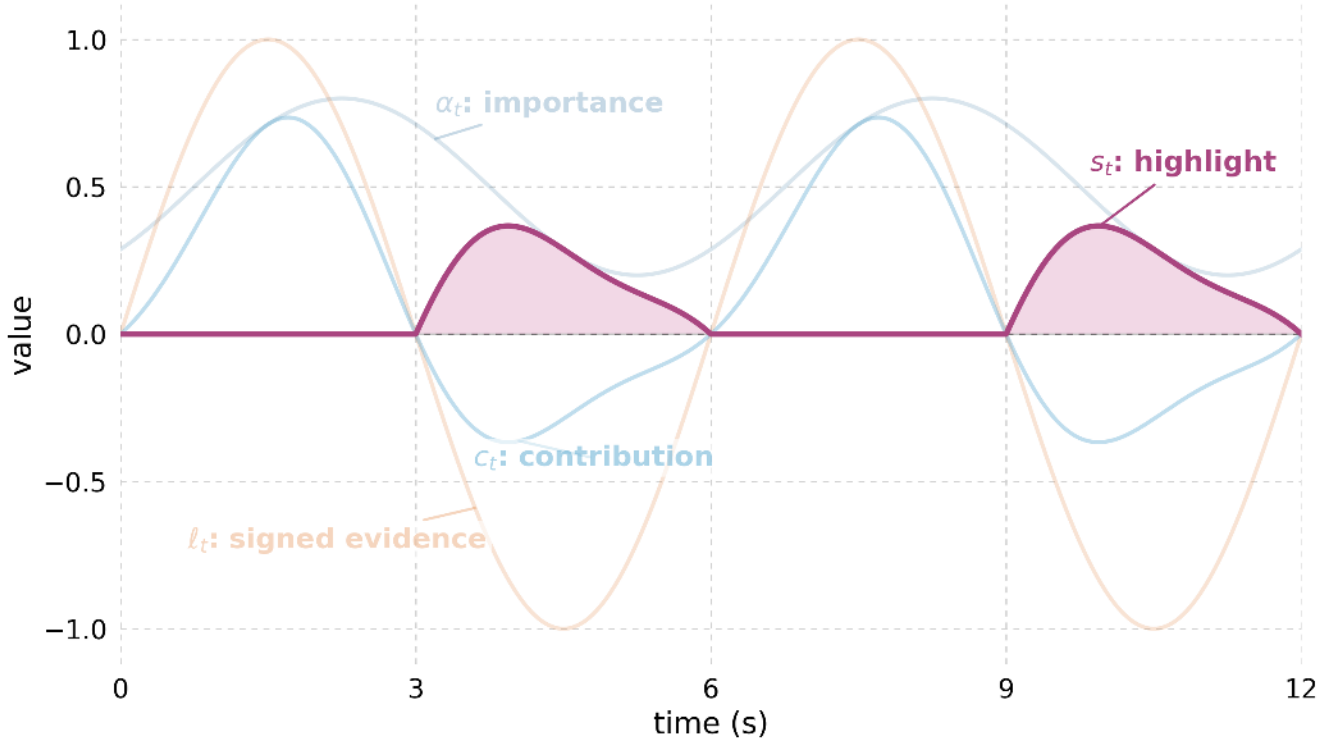
signed evidence: expert (+) or amateur (−)

c_t : contribution

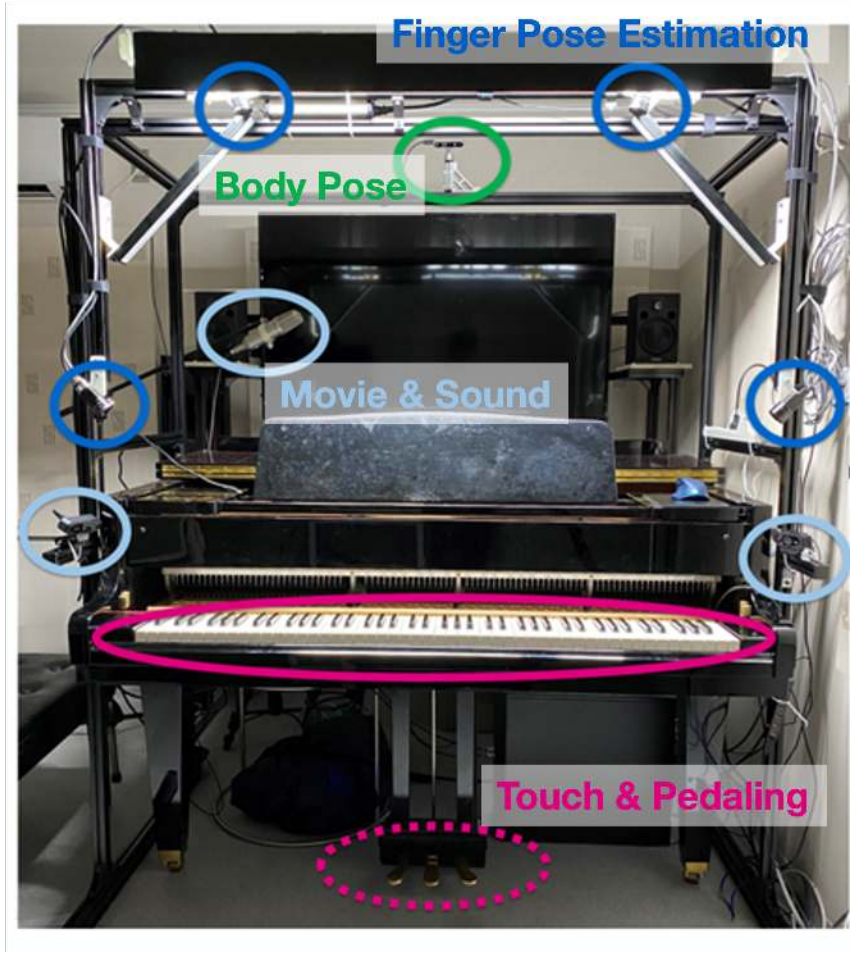
how much × which side: $c_t = \alpha_t \cdot \ell_t$

s_t : highlight

keep amateur-leaning moments: $s_t = \mathrm{ReLU}(-c_t)$



A controlled multimodal corpus for training



Cohort & takes

- **80** pianists (73 valid), amateur → professional
- **2,433** recorded → **1,083** valid takes
- **15** tasks · 9 scales + 6 arpeggios
- ~**10.7 s** median per take

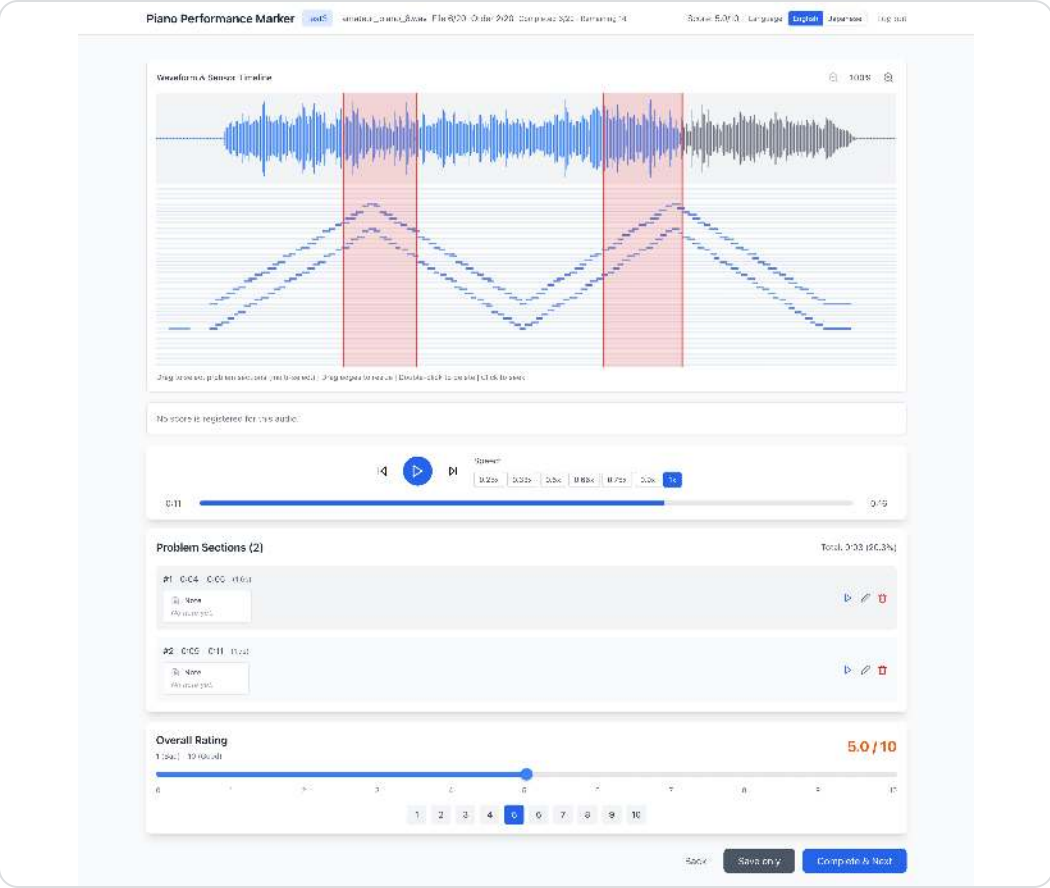
Synchronized modalities

- **1 kHz** per-key motion · 88 keys (non-contact optical, mm)
- **44.1 kHz** stereo audio

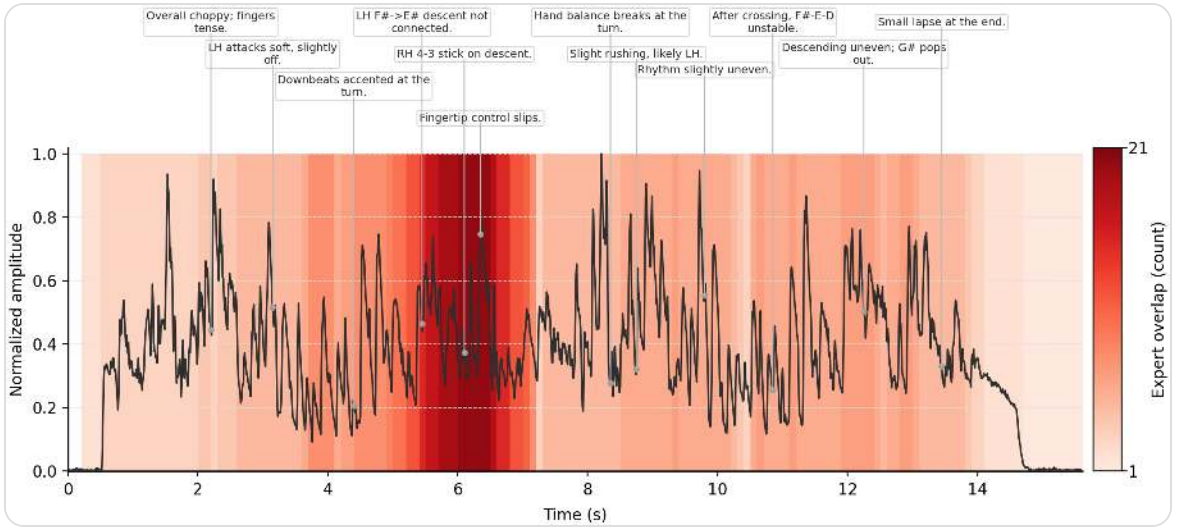
Labels — expert vs. amateur

- **6,517** listener ratings from **53** raters
- **597** expert / **486** amateur takes (median-thresholded)

Collecting expert review spans for evaluation



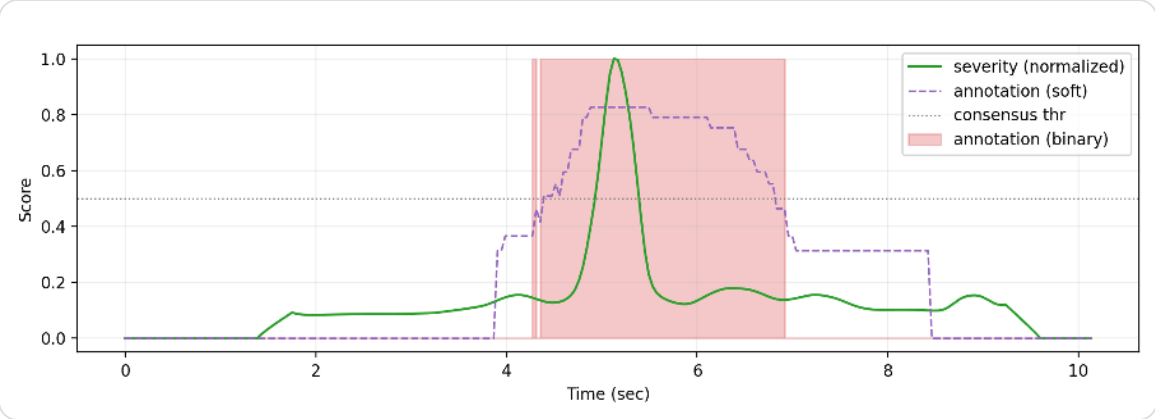
Annotation tool. Experts mark review spans on a waveform + key-motion timeline — loop, slow playback, optional notes per span.



Consensus + comments. Overlap of expert-marked spans (darker = more agreement) with attached free-text notes.

20 amateur takes · 21 expert pianists · 407 review notes

Highlights align with expert consensus

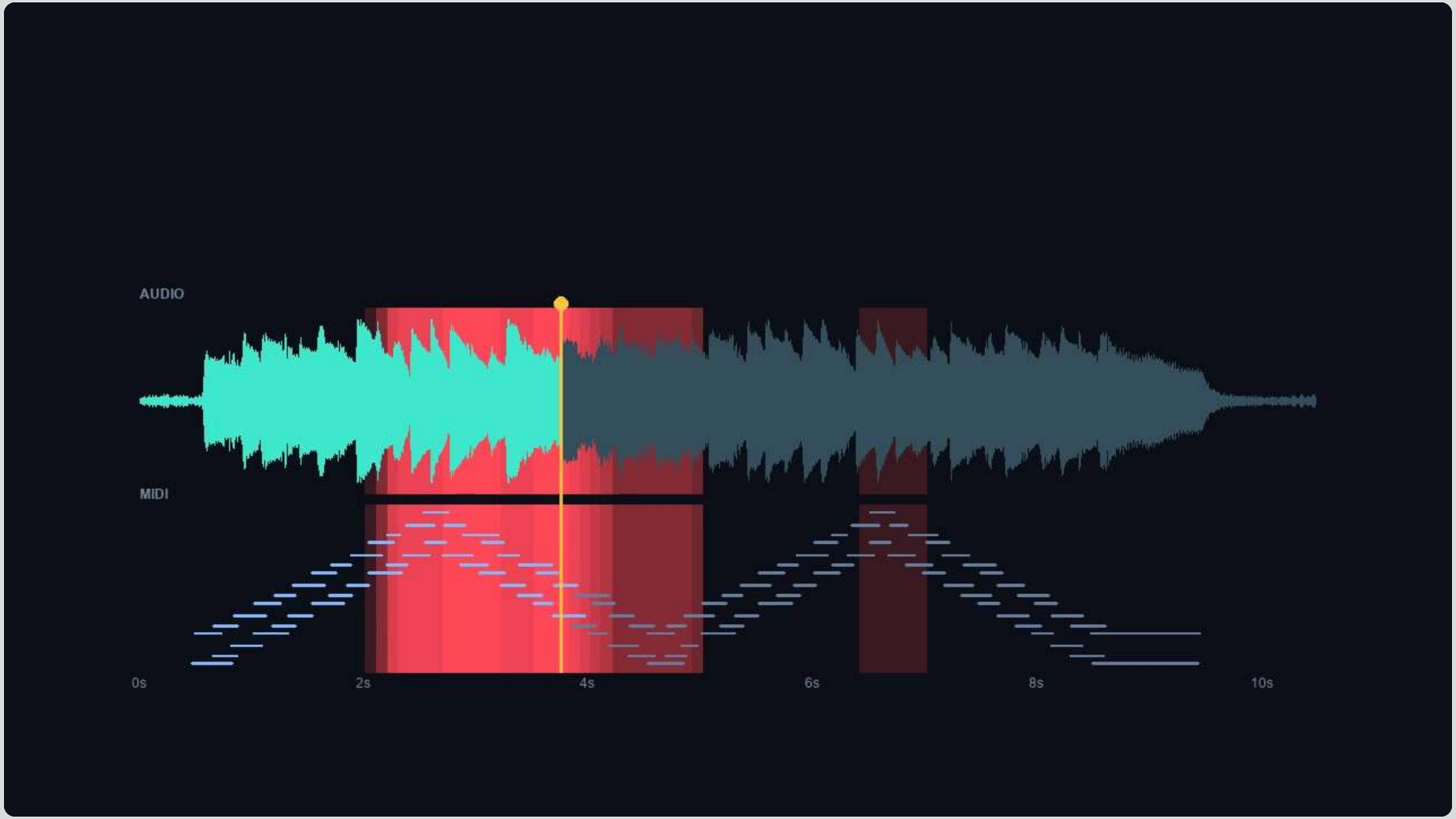


Model highlight (green) peaks **inside the span experts agreed to review** (red) — one amateur take.

Model	Pearson	AP	ROC-AUC
Random	≈0	0.20	0.50
Audio-only	0.590	0.539	0.720
Sensor-only	0.606	0.546	0.732
Multimodal (ours)	**0.612**	**0.567**	**0.753**

No localized training labels — yet highlights show some agreement with expert review spans. Multimodal outperforms every baseline: **r = 0.61**, **ROC-AUC 0.75**.

Profy visualization demo



Target performance → Profy visualization

Three directions for future work
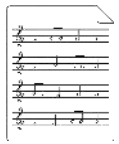


Validate practice outcomes

Run long-term, interventional studies—Profy vs. no tool vs. global score vs. teacher—measuring real improvement, retention, and efficiency, not just expert-span agreement.



From where to what & how

Move from review cues toward diagnosis: multi-channel issues (timing, tone, legato…) and hypotheses like “the right hand may be rushing,” refined with a teacher in the loop.



Scale to real repertoire

Generalize from short technique studies to long, expressive pieces and other genres—without treating interpretation as an error just because it looks amateur-like.