SOUNDS OUT OF PLÄCE? SCORE INDEPENDENT DETECTION OF CONSPICUOUS MISTAKES IN PIANO PERFORMANCES

Alia Morsi [1], Kana Tatsumi [2], Akira Maezawa [3], Takuya Fujishima [3], Xavier Serra [1]

[1] MTG, Universitat Pompeu Fabra [2] Nagoya Institute of Technology
[3] Yamaha Research Lab

ISMIR 2023

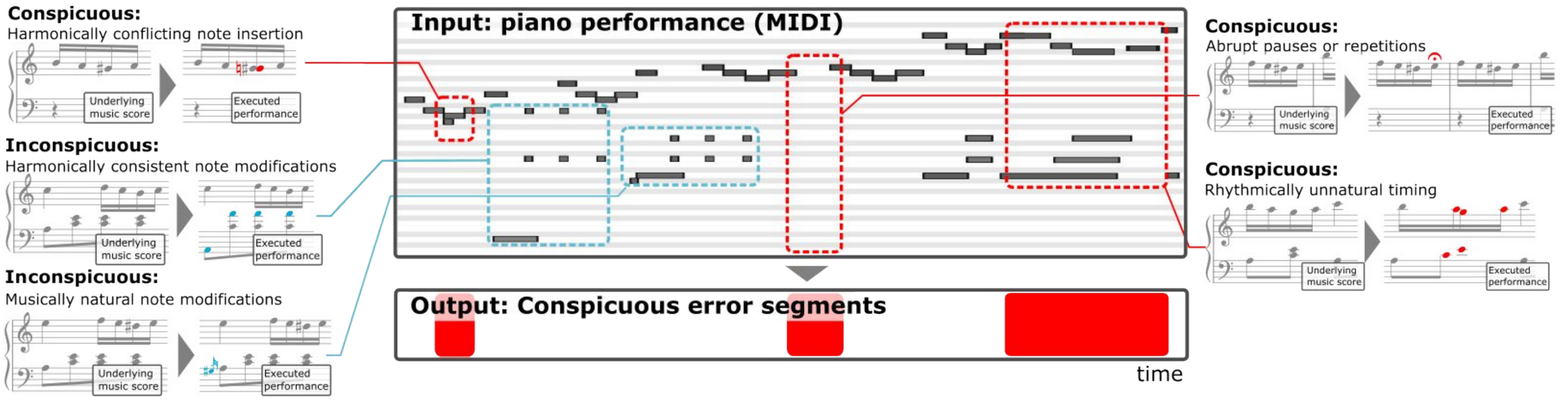
Paper Goals

Build a **score-independent** conspicuous error detector for **standard piano repertoire of beginner to intermediate students.**

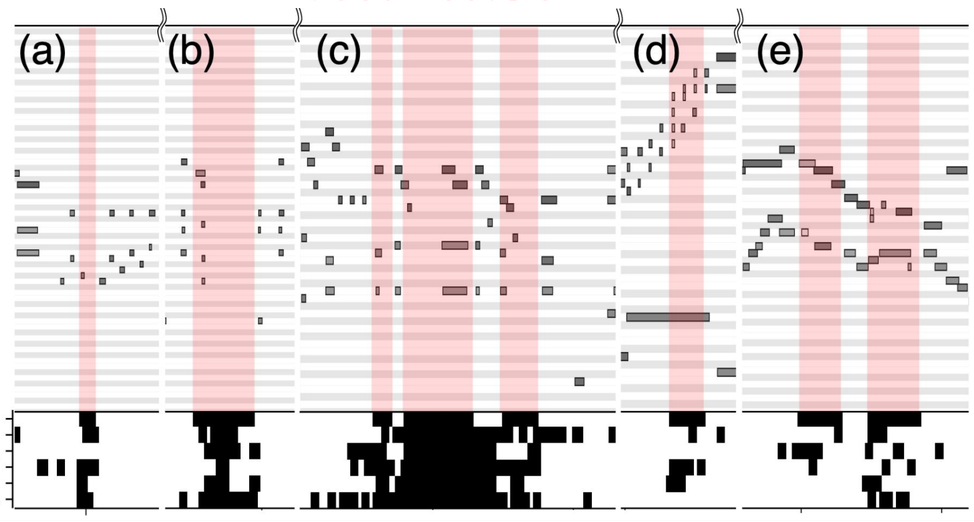
Premise

Not all piano performance mistakes are equally salient to a listener [1].

Concept

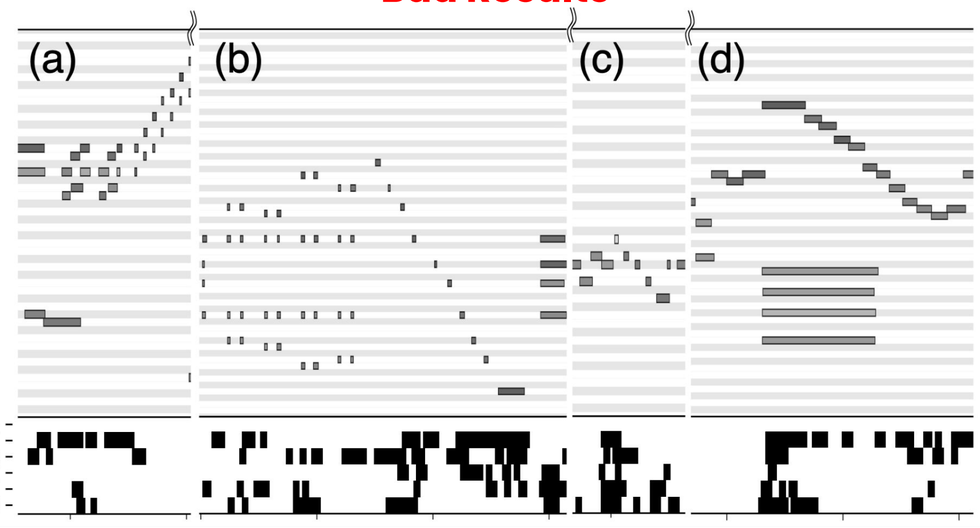


Good Results



For More Examples

Bad Results



Data Collection

Train Data

- **Sight Reading Data (SR)**: 103 sight reading performances. (379 minutes).
- **Performance Data (PF):** 245 performances of 3 minutes each (723 minutes).

Annotation procedure

- 2 annotators with music background.
- Asked to only annotate obvious mistakes.

Eval Eval

- **Burgmüller Data (BM):** 50 performances of from Op 100 (25 recorded twice).

Annotation procedure

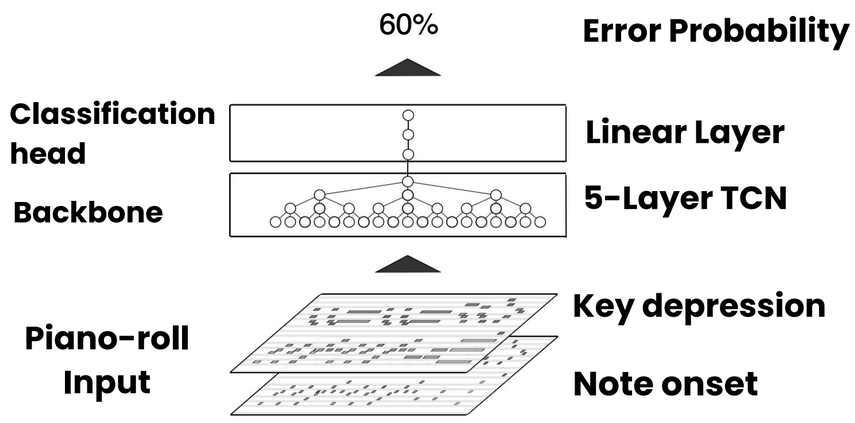
- Alignment to music score conducted first.
- 1 annotator with a music background asked to manually review the labels wrt the sheet music and make corrections.

Augmentation Strategies

Systematic adjustments resembling the mistakes of adult beginner learners were applied to a set of **mistake-free** performances. **(AUG)**

- Note mistakes: **Omissions**, **insertions**, **substitutions**
- Stops and hesitations: silences and repetitions of the last played note.

Architecture



Models

- **Baseline:** SR and PF data
- **SYNTH:** SR, PF, and AUG
- **SYNTH (FT):**
  - Pretrain with AUG,
  - FT with SR and PF
- **AE:**
  - Train TCN Autoencoder with data similar to PF
  - Use encoder as classifier backbone, and fine-tune with SR and PF
- **AE+SYNTH:** same as AE, but fine tune with SR, PF, and AUG.

Results

Method	Precision	Recall	F-measure
Baseline	**0.79**	**0.80**	**0.78**
SYNTH	0.65	0.76	0.69
SYNTH(FT)	0.61	0.69	0.62
AE	0.55	0.59	0.55
AE+SYNTH	0.44	0.65	0.51

(a) **SR** Data

Method	Precision	Recall	F-measure
Baseline	0.28	0.46	0.33
SYNTH	0.27	0.54	0.34
SYNTH(FT)	**0.30**	0.61	**0.38**
AE	0.28	0.52	0.34
AE+SYNTH	0.27	**0.63**	0.36

(b) **PF** Data

Method	Precision	Recall	F-measure
Baseline	0.26	0.36	0.26
SYNTH	0.26	**0.69**	0.35
SYNTH(FT)	0.26	0.49	0.32
AE	0.27	0.46	0.31
AE+SYNTH	**0.28**	0.52	**0.35**

(c) **BM** Data

[1] B. H. Repp, “The art of inaccuracy: Why pianists’ errors are difficult to hear”, Music Perception: An Interdisciplinary Journal, vol. 14, p. 161–183, 1996