October 29th. 2024

Can AI Master Music Expression?

Al-Driven Transformation of Score MIDI into Expressive Piano Performance

Jingjing Tang <u>sjingjing.tang@qmul.ac.uk</u>

Supervisors: George Fazekas & Geraint Wiggins





What Elements Contribute to a Piano Performance?



Help AI Understand Music

Score MIDI

The MIDI generated directly from the composition's notations, featuring extremely accurate note timing and uniform velocity without dynamic changes.

Performance MIDI

A transcription or recorded MIDI of human performances, incorporating variations in timing and dynamics (velocity) that reflect musical interpretation. Selected From: Piano Sonata No. 9 in E major, Op. 14, No. 1, I. Allegro by Beethoven









Score MIDI

Performance MIDI

From Score to Audio

We aimed to create an integrated Al-driven system that transforms score MIDIs into expressive piano performance audios.



Lacking nuanced timing and velocity variations for expression

Containing expressive timing and velocity variations that mimic human pianists

Audio Performance

Capturing the authentic sound quality, acoustic nuances, and ambient effects of the live piano

Learn from Master Pianists

Create the AI pianist:

We used a large-scale dataset of piano performances by master pianists for training.

Through the training, the model learns to control the timing and velocity of each note, mimicking human interpretation.



ATEPP Dataset

A dataset of transcribed expressive piano performances (ATEPP) from master pianists, providing human performances in MIDI format

Transformer Model

A generative model adapted from the NLP domain for predicting timing and velocity of each note



Learn from Audio Recordings of Performances

Create the piano synthesiser:

We developed the synthesizer through transfer learning, using the same large-scale dataset to refine a model initially crafted for text synthesis.

This training enables the synthesiser to replicate piano sounds and capture the acoustic ambiance of environments like concert halls and recording studios.



ATEPP Dataset

Using both audios and performance MIDIs from the ATEPP dataset to train the synthesiser

Midi Synthesis Model

A generative model adapted from the Text-to-Speech domain for synthesising the performance MIDIs



Demo Presentation

Link to the demo page: https://tangjjbetsy.github.io/Showcase/

Thanks!

Questions are welcomed in the Q&A session!

