Audio-to-score alignment for piano performances

Laboratory:

Swiss Data Science Center

Type:

Semester project

Description:

A music score is a symbolic notation of instructions and recommendations given by the composer. While pitches, rhythms and other chords are part of the original design, some freedom of interpretation is given to the performer; each performance is a unique display of skills and choices. As part of an educational tool, this project goal is to align a piano recording of Western classic music to the original score, in order to relate each expected note to its physical waveform.

Various related topics have been discussed in the signal processing and computational musicology domain. Usually, approaches rely on Fourier analysis (e.g., Short-time Fourier transform, Constant-Q transform) to extract frequency features, and then apply sequence alignment methods (e.g., Hidden Markov Model, Fast Dynamic Time Warping [1]). Onset detection – locating the attack time of each note – can be seen as a by-product of such alignment, which can be refined for more accuracy. Furthermore, following recent advances in deep learning, researchers have successfully applied transformer-based models to recover the onsets from the audio [4], without the score.

In this project, we assume that the score is known. The main objective is to infer accurate onsets for each note. This fine-grained alignment is key for further analysis, in order to provide detailed feedback to the performer. As a first step, the student will also be working on a symbolic alignment [2], focusing on the sequence alignment challenge. The MAESTRO dataset [3] is a good starting point; further data collection could be envisioned.

Goals/benefits:

- Working with signal processing and machine learning libraries in Python
- Hands-on experience, practical application

Prerequisites:

- Python (advanced skills)
- Machine learning (intermediate skills)
- Linear algebra and signal processing are a plus.
- Interest in sound processing and computational musicology recommended.

Deliverables:

- Well-documented code.
- Written report and oral presentation.

References:

[1] Salvador, Stan, and Philip Chan. "FastDTW: Toward accurate dynamic time warping in linear time and space." KDD workshop on mining temporal and sequential data. State College, PA, USA: Citeseer, 2004.

[2] Nakamura, Eita, Kazuyoshi Yoshii, and Haruhiro Katayose. "Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment." ISMIR. 2017.

[3] Hawthorne, Curtis, et al. "Enabling factorized piano music modeling and generation with the MAESTRO dataset." arXiv preprint arXiv:1810.12247 (2018).

[4] Hawthorne, Curtis, et al. "Sequence-to-sequence piano transcription with Transformers." arXiv preprint arXiv:2107.09142 (2021).

Contact:

Johan Berdat (johan.berdat@epfl.ch)