Score Following for Piano Performances with Sustain-pedal Effects

Bochen Li, Zhiyao Duan
Department of Electrical and Computer Engineering, University of Rochester
bochenli@rochester.edu, zhiyao.duan@rochester.edu

Summary

- Sustain pedal is commonly used in piano music since the Romantic era.
- The usage of sustain pedal extends notes and causes mismatch between audio and score, which may fail score following systems.
- Proposes a spectral peak removal operation to reduce the sustain pedal effect.
- Evaluations on 50 MAPS pieces show significant improvements on both accuracy and robustness.

1. System Framework [1]

- View score following as an online inference problem of the hidden state $s_t$ (current position and tempo) from current and previous audio observations $y_{t-1}, \ldots, y_t$ of a Hidden Markov Model.

- **Process model** defines how states transition using dynamical equations.

- **Observation model** evaluates the match between an audio frame and hypothesized state on the pitch content.

- Multi-Pitch Estimation (MPE) likelihood model proposed in [2] evaluates the likelihood of a hypothesized pitch set in explaining the magnitude spectrum of an audio frame.

2. Improved System

An onset detection method and peak removal operation is applied in the feature extraction part to locate unfaithful frames and remove sustain-pedal effects.

- **Spectra-based Onset Detection Method**
  - Spectrogram from STFT
  - Logarithmic compression on spectrogram $Y(n, k) = \log(1 + \gamma \cdot Y(n, k))$
  - summing positive temporal differences $\Delta v(n) = \sum \{|Y(n, k) - Y(n - 1, k)|, k\}$
  - normalize by audio energy $\Delta v(n) = \Delta v(n)/E(n)$

Peak Removal May Introduce New Mismatch

...but, it actually helps score following

1. *Observation model* evaluates the match between an audio frame and hypothesized state on the pitch content.

2. **Improved System**

An onset detection method and peak removal operation is applied in the feature extraction part to locate unfaithful frames and remove sustain-pedal effects.

- **Spectra-based Onset Detection Method**
  - Spectrogram from STFT
  - Logarithmic compression on spectrogram $Y(n, k) = \log(1 + \gamma \cdot Y(n, k))$
  - summing positive temporal differences $\Delta v(n) = \sum \{|Y(n, k) - Y(n - 1, k)|, k\}$
  - normalize by audio energy $\Delta v(n) = \Delta v(n)/E(n)$

Peaks Removal Criteria

A peak in the $n$-th frame $(a)$ whose frequency is very close to $(b)$ whose amplitude is smaller than $\theta_3$. $\Omega$ of a peak in the $m$-th frame $\{\Omega, \omega'\}$ such that $|f(m) - f(n)| < \theta_1 < \omega' < \omega_2$

Without sustain pedal:

Pressing and holding the key will yield an impulse-like articulation, and releasing the key will let the damper touch the string quickly, thus the sound ceases.

With sustain pedal:

- All dampers leave all the strings, no matter the key is released or not.
- Played sound continues until string naturally ceases.
- This can cause potential mismatch between audio and score especially after note onsets.

Method

1. **Observation model** evaluates the match between an audio frame and hypothesized state on the pitch content.

2. **Multi-Pitch Estimation (MPE) likelihood model** proposed in [2] evaluates the likelihood of a hypothesized pitch set in explaining the magnitude spectrum of an audio frame.

3. **Improved System**

An onset detection method and peak removal operation is applied in the feature extraction part to locate unfaithful frames and remove sustain-pedal effects.

- **Spectra-based Onset Detection Method**
  - Spectrogram from STFT
  - Logarithmic compression on spectrogram $Y(n, k) = \log(1 + \gamma \cdot Y(n, k))$
  - summing positive temporal differences $\Delta v(n) = \sum \{|Y(n, k) - Y(n - 1, k)|, k\}$
  - normalize by audio energy $\Delta v(n) = \Delta v(n)/E(n)$

Experiments

**Dataset:** 50 music pieces from MAPS dataset (25 synthetic, 25 acoustic)

**Measurement:**

- **Average Time Deviation:** Average of time deviation (ms) between all aligned and ground-truth onsets.
- **Align Rate:** Percentage of correctly aligned notes, with different levels of deviation tolerance.

**Results:**

- **On synthetic performances**

- **On acoustic performances (YAMAHA Disklavier)**

References