Aligning Semi-improvised Audio with Its Lead Sheet

Zhiyao DUAN & Bryan PARDO
Northwestern University, EECS Department, Evanston, IL, USA.
zhiyaoduan00@gmail.com • pardo@northwestern.edu • http://music.cs.northwestern.edu

Introduction

- Existing audio-score alignment methods assume that the audio performance is faithful to a fully-notated score.
- Semi-improvised music (e.g., jazz) strongly violates this assumption.
- We propose a system for aligning a semi-improvised music audio performance with its score, i.e., a lead sheet.
- Requires no prior training on the lead sheet to be aligned.
- Handles structural changes (e.g., jumps, repeats) in the performance.
- Obtains promising results on 24 piano performances and 12 full-band commercial recordings of 3 jazz lead sheets.

Problem Analysis

- Harmony is rendered in free rhythmic patterns.
- Melody may be significantly altered (as in Improv 2).
- Performers often make unexpected jumps and repeats.

Essential Information for Alignment

- Performed notes correspond to score harmony at scale of two beats.
- Structural changes only happen at section boundaries.

Experiments

- Dataset and Measure
  - 3 jazz lead sheets: Dindi by Antonio Carlos Jobim, Nicas’s Dream by Horace Silver and Without A Song by Vincent Youmans.
  - 12 performances for each lead sheet (4 easy piano, 4 medium piano and 4 commercial jazz band recordings).
  - View alignment as a classification problem: assign each audio frame (46ms long) a score measure number.
  - Accuracy: % of frames that are correctly assigned a measure number.
  - Ground-truth measure numbers are obtained manually.

Results:

- Tracking Audio Beat: we use the method in [1].
- Extracting Chromagrams: 1 for audio and 36 for MIDI
  - Calculate an audio chromagram for segments of length 1 = 2 beats and hop h = 1/4 beats.
  - Calculate a MIDI chromagram for each of the 3 scales of segments: (l, h), (1/2l, 1/2h) and (2l, 2h).
  - Transpose MIDI chromagrams 12 times for possible key transpositions.
- Aligning Chromagrams: Let A = (a1, a2, ..., an) be the audio chromagram, S = (s1, s2, ..., sk) be the score chromagram.
  - Audio may start from anywhere on the score:
    \[ C(0, 0) = 0, C(i, 0) = i \cdot c1, C(0, j) = 0 \] (1)
  - Audio may jump at possible jumping points:
    \[ C(i, j) = \begin{cases} 
    C(i, j - 1) + c1, & \text{skip audio} \\
    C(i - 1, j) + c2, & \text{skip score} \\
    \min_{k \in P(j)} C(i - 1, k) + d(a_i, s_j), & \text{transition} 
    \end{cases} \] (2)
  where \( P(j) \) is the set of segments (at the scale of 2 beats), from which a performance might transition to \( j \).
  - Audio may end at anywhere on the score: Trace back from \( C(m, j_1) \), where \( j_1 = \arg \min C(m, j) \).

More Examples: http://www.cs.northwestern.edu/~zdu459/ismir2011/examples


This work was supported by NSF grant IIS-0643752.