ROBUST ON-LINE ALGORITHM FOR REAL-TIME AUDIO-TO-SCORE ALIGNMENT BASED ON A DELAYED DECISION AND ANTICIPATION FRAMEWORK

Ryuichi YAMAMOTO, Shinji SAKO, Tadashi KITAMURA (Nagoya Institute of Technology, JAPAN)

1. Motivation

Audio-to-score alignment has been well studied but still has room for improvement on the robustness of its ON-LINE algorithm.

Most popular dynamic programming approaches such as DTW and HMMs rely on their OFF-LINE optimization algorithms.

Some approximations must be required in ON-LINE settings.

2. The Idea

How to improve the robustness against polyphonic signals?

Delayed decision improves the reliability of estimating past score position compared with an instant decision.

How to find a reliable current score position?

Future anticipation using tempo will be useful.

How to find both score position and tempo?

We propose a delayed decision and anticipation framework that jointly optimizes score position and tempo based on Segmental Conditional Random Fields and Linear Dynamical System.

3. Segmental CRFs

Conditional Random Fields (CRFs)

Features

- Segment-level Markov chain (NOT frame-level)
- Flexible feature design than classical Hidden Markov Models
- Chroma features, onset features, duration features, etc.

Advantages

- Score position and tempo are jointly optimized minimizing the anticipation error
- Delayed decision can find reliable score position
- Reliable future position is anticipated using adaptively estimated tempo

Our algorithm is intermediate between ON-LINE and OFF-LINE

4. SCRFs with Tempo Model

Chord transition model

based on SCRFs

Audio observation

Chromagram

Tempo model

based on LDS

5. Proposed Algorithm

Delayed decision Viterbi algorithm to find the reliable past score position

Tempo induction from the estimated path (Kalman Filter)

Future anticipation using the past-decided score position and tempo

6. Experiments

We test with various delay-time 0 s, 0.5 s, 1.0 s, 1.5 s

The result with delay-time 0 s is a baseline

Database

MAPS: Classic: Piano

60 recordings (about 4-hours)

RWC: Jazz: Multiple instruments include percussions

50 recordings (about 3.5-hours)

Settings

Evaluation measure

Onset recognition rate

Onset detection tolerance δ

100 ms, 300 ms

Hop-size

10 ms

Model parameters

Tuned by a grid search

7. Results and Discussion

EFFECT OF DELAY-TIME (CLASSIC)

EFFECT OF DELAY-TIME (JAZZ)

8. Conclusion

Robust ON-LINE algorithm for score alignment based on a delayed decision and anticipation framework

The combined model of SCRFs and LDS provides an unified framework to find both score position and tempo

Our framework with a large delay time get better results in general

9. Future Direction

Dynamic optimization of delay-time

Learning the model from real-data

Application to automatic accompaniment

1 "Ryry: Real-time Score Following and Automatic Accompaniment," Demonstration movies are found on YouTube