Optimizing DTW-Based Audio-to-MIDI Alignment and Matching

Colin Raffel and Daniel P. W. Ellis
41st IEEE International Conference on Acoustics, Speech and Signal Processing
March 23, 2016
Dynamic Time Warping
DTW for Audio-to-MIDI Alignment
System Design: Representation?

Chromagram?

Constant-Q Spectrogram?

Log Magnitude?

Z-scored?

L2-normalized?

Beat-synchronous?
System Design: Path Constraints?
System Design: Score Reporting?

\[
\text{score} = \frac{\sum_{i=1}^{\|p_m\|} D[p_m[i], p_a[i]] + \Phi(i)}{\max(p_m) \max(p_a) \sum_{i=\min(p_m)} \sum_{j=\min(p_a)} D[i,j]}
\]

\[
\frac{\|p_m\|}{\max(p_m) - \min(p_m)} \frac{\max(p_a) - \min(p_a)}{\|p_a\|}
\]
Bayesian Optimization

Expected Improvement
Idea: Synthetic Alignment Data

Original MIDI CQT

Note
C6  C5  C4  C3

Frame
0   200  400  600  800  1000  1200  1400

After corruption

Note
C6  C5  C4  C3

Frame
0   200  400  600  800  1000  1200  1400
Artificial Time Warping
Correcting Time Warping

Timing correction

Offset from original time

Original time

Ground-truth offset

Fixed corrupted offset
Measuring Error

Correction error

Original time

Error
Score Normalization Search

![Graph showing the relationship between Normalized DTW distance and Alignment error.](image-url)
Best System:

- Use log-magnitude constant-Q spectrograms
- Don’t beat synchronize
- L2 normalize spectra (cosine distance)
- Don’t z-score spectrograms
- Use median distance as non-diagonal penalty
- Force sequences to match up to 96% of shorter
- Don’t use a band path constraint
- Include penalties in confidence score
- Normalize by path length and submatrix mean
Real-World Test
Pointers

http://github.com/craffel/alignment-search
http://github.com/craffel/pretty-midi
http://github.com/craffel/djitw
http://github.com/bmcfee/librosa

craffel@gmail.com