Large-Scale Content-Based Matching of MIDI and Audio Files

Colin Raffel and Daniel P. W. Ellis
ISMIR 2015
October 27, 2015
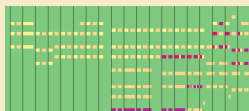
Lab
ROSA

NSF · NATIONAL · SCIENCE · FOUNDATION

IGERT Integrative Graduate
Education and Research Traineeship

Ground Truth from MIDI

0.0s: F major, 53.2s: D minor, ...
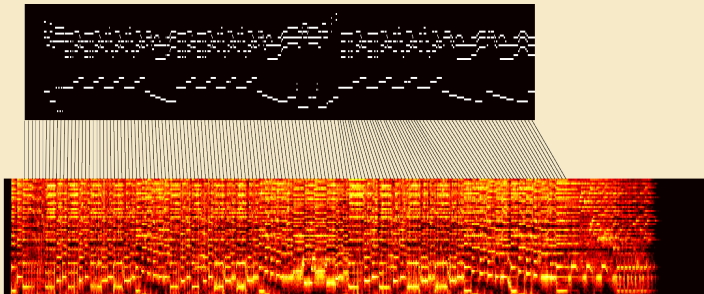

0.0s: 110 bpm, 22.2s: 103 bpm, ...
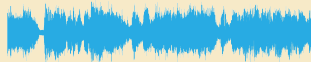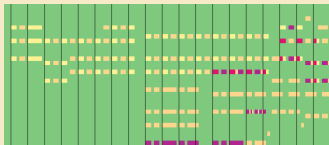
Matching and Aligning

Beatles/hello, goodbye.mid ↔ The Beatles/Hello.mp3

The MSD: An Ideal Large Collection

Matching by Metadata Won't Work

```
J/Jerseygi.mid
V/VARIA18O.MID
Carpenters/WeveOnly.mid
2009 MIDI/handy_man1-D105.mid
G/Garotos Modernos - Bailanta De Fronteira.mid
Various Artists/REWINDNAS.MID
GoldenEarring/Twilight_Zone.mid
Sure.Polyphone.Midi/Poly 2268.mid
d/danza3.mid
100%sure.polyphone.midi/Fresh.mid
rogers_kenny/medley.mid
2009 MIDI/looking_out_my_backdoor3-Bb192.mid
```
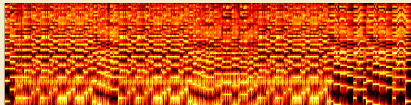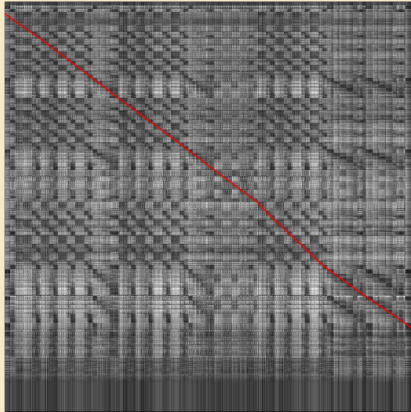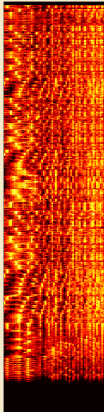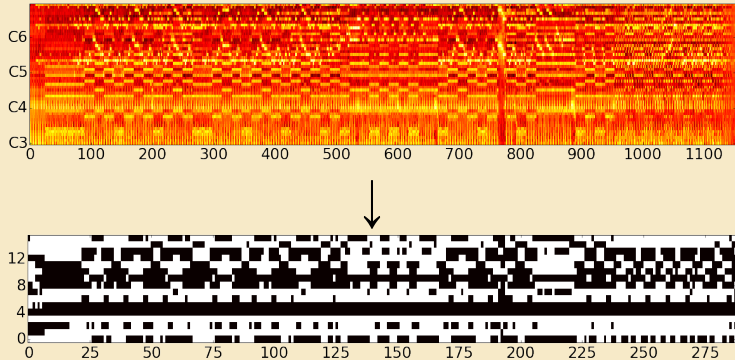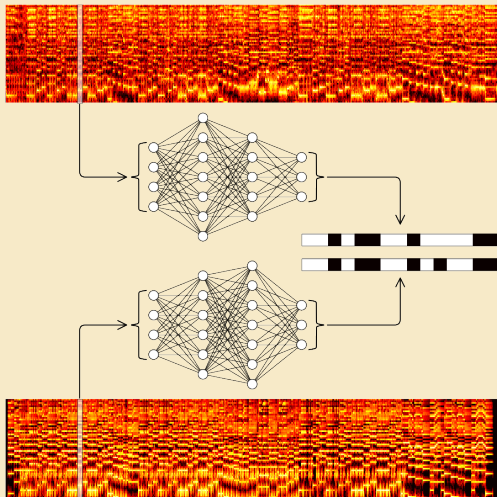
DTW: Natural, and Too Slow

Hash Sequences

$$\texttt{distance}[m, n] = \texttt{bits\_set}[x[m] \oplus y[n]]$$

Similarity-Preserving Hashing

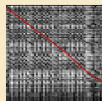# Similarity-Preserving Hashing
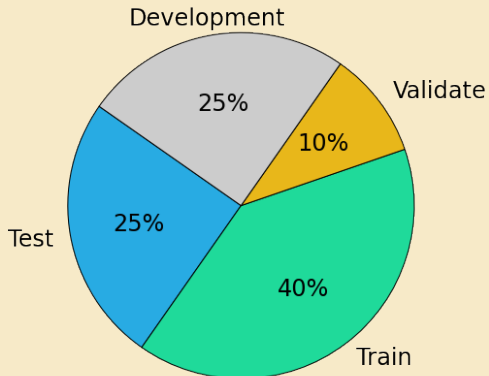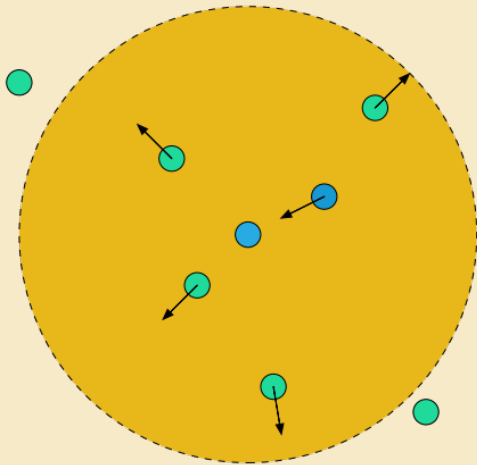
Collecting Data

140,910

24,850

17,243

26,311

10,035

Development 25%

Validate

Test 25%

10%

Train 40%

Network Structure

5×12 convolution, 2×2 max-pool

3×3 convolution, 2×2 max-pool
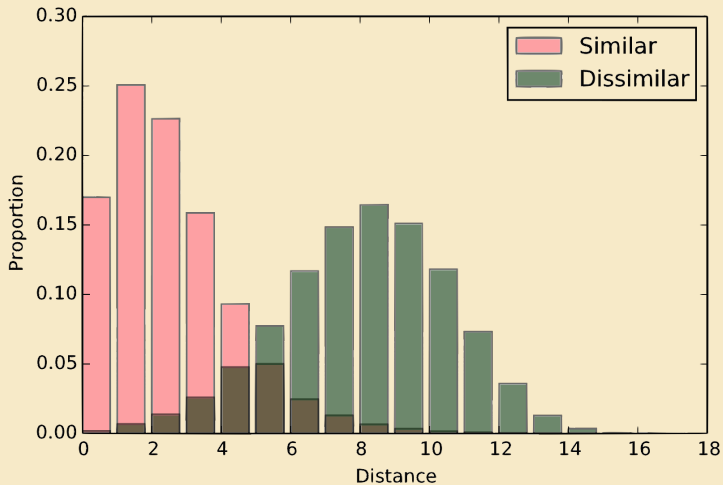
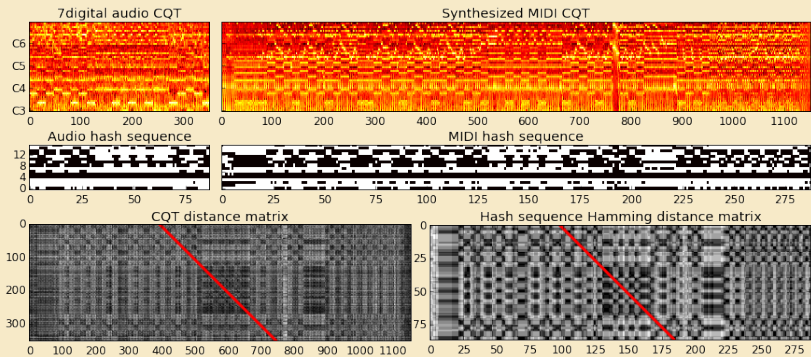Dense ReLU

2048

Dense
ReLU

2048

Dense
tanh
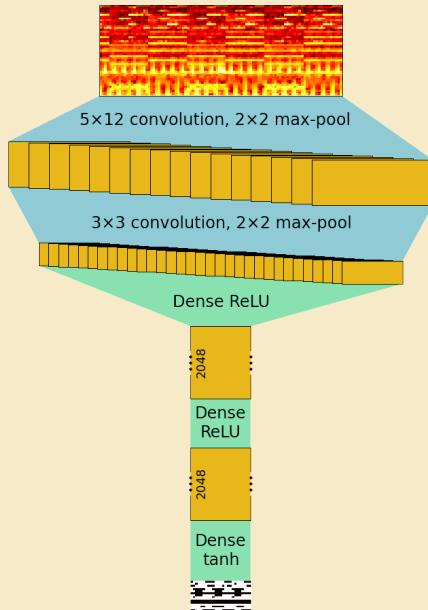
Loss Function

$$\mathcal{L} = \frac{1}{|\mathcal{P}|} \sum_{(x,y)\in\mathcal{P}} \|f(x)-g(y)\|_2^2 - \frac{\alpha}{|\mathcal{N}|} \sum_{(x,y)\in\mathcal{N}} \max(0, m - \|f(x)-g(y)\|_2)^2$$
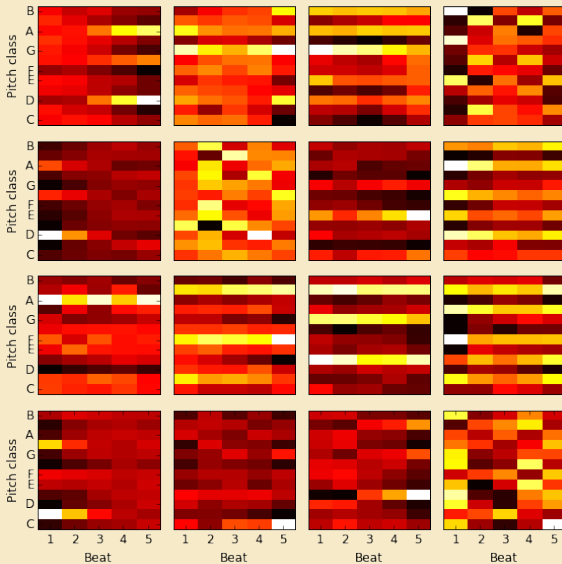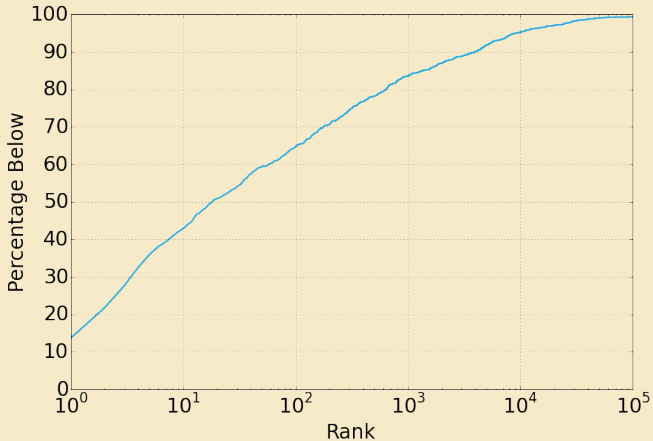
Validation Distance Distribution

Example Sequence
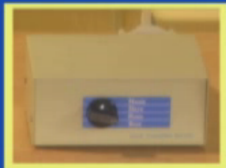
First Layer Filters

Test Set Matching Results

Code

http://github.com/craffel/midi-dataset

Related Work?