Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching

Colin Raffel

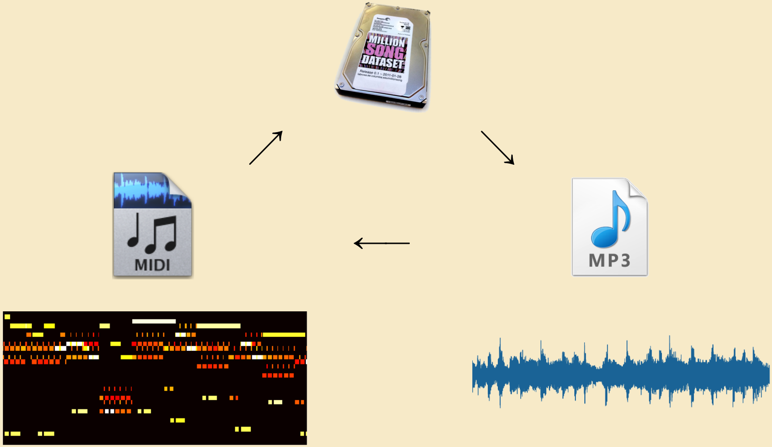
May 16, 2016

Lab ROSA

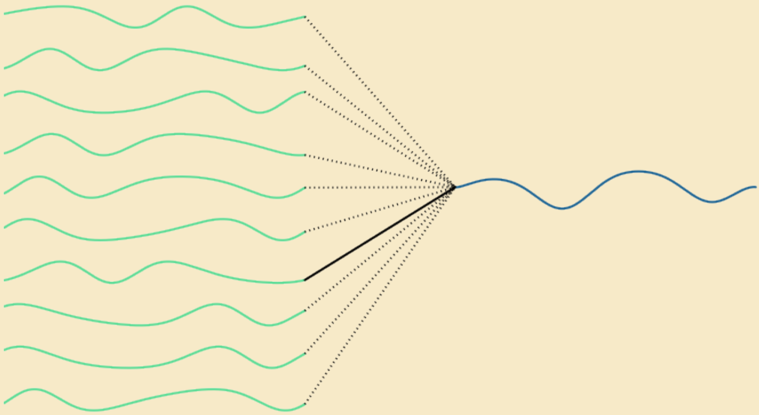
NATIONAL SCIENCE FOUNDATION NSF IGERT Integrative Graduate Education and Research Traineeship

The Goal

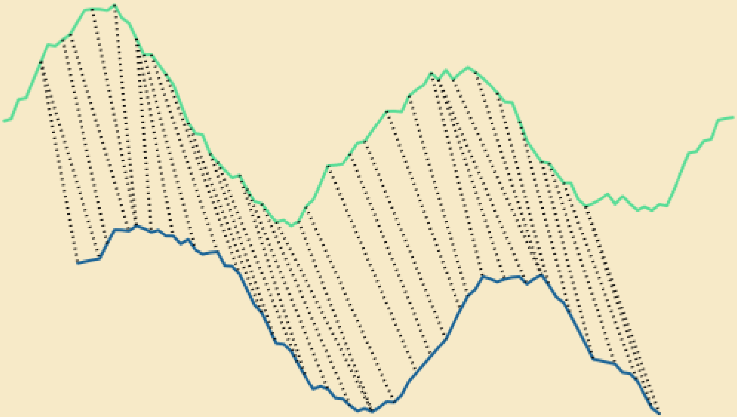
```
     artist: 'Tori Amos'
    release: 'LIVE AT MONTREUX'
      title: 'Smells Like Teen Spirit'
         id: 'TRKUYPW128F92E1FC0'
   duration: 216.4502
sample_rate: 22050
  audio_md5: '8'
 7digitalid: 5764727
       year: 1992
```



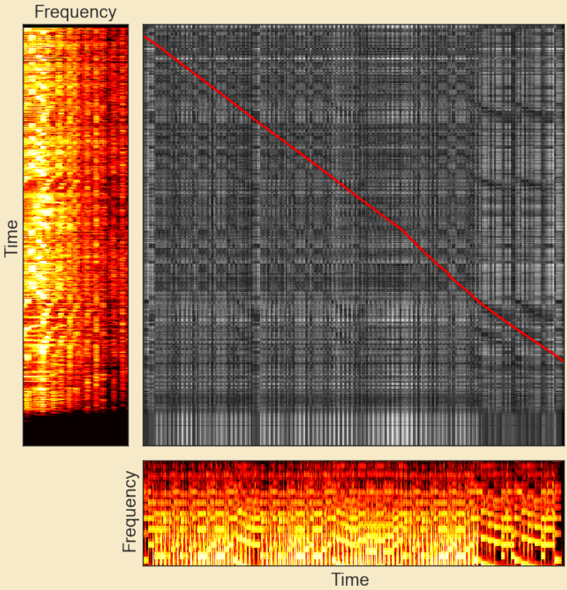
Sequence Matching



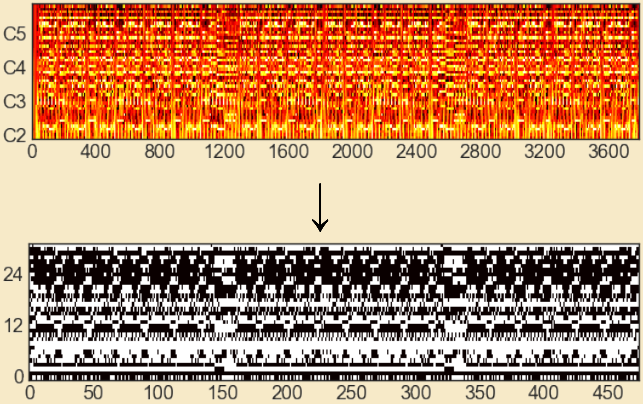
Dynamic Time Warping

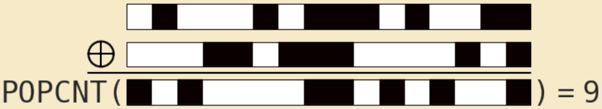


Comparing MIDIs with DTW



Downsampled Hash Sequences

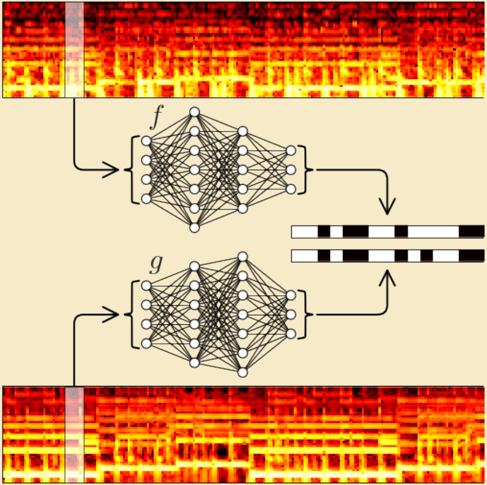




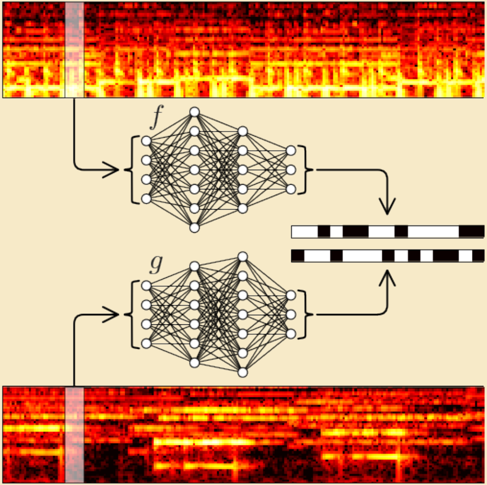
⊕

POPCNT( ) = 9

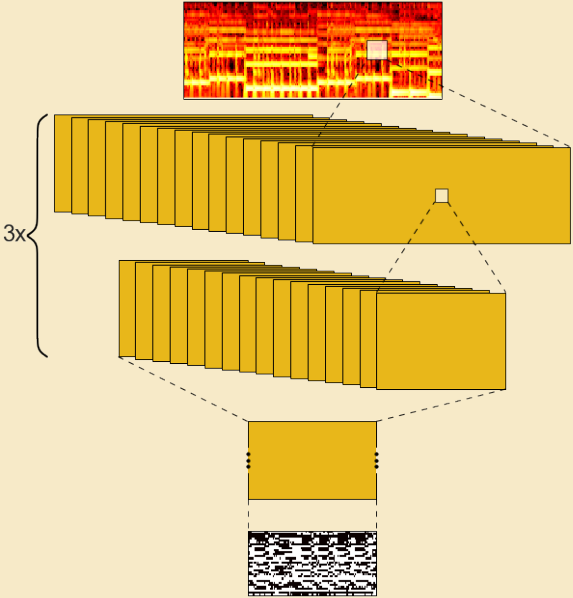
Similarity-Preserving Hashing



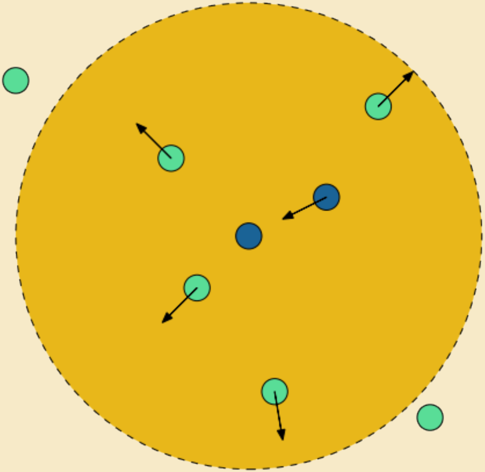
Similarity-Preserving Hashing



Network Structure

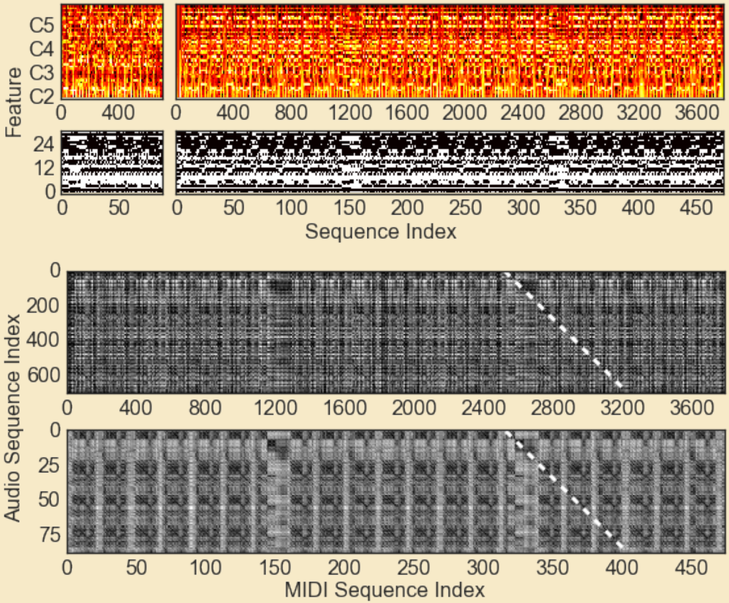


Loss Function

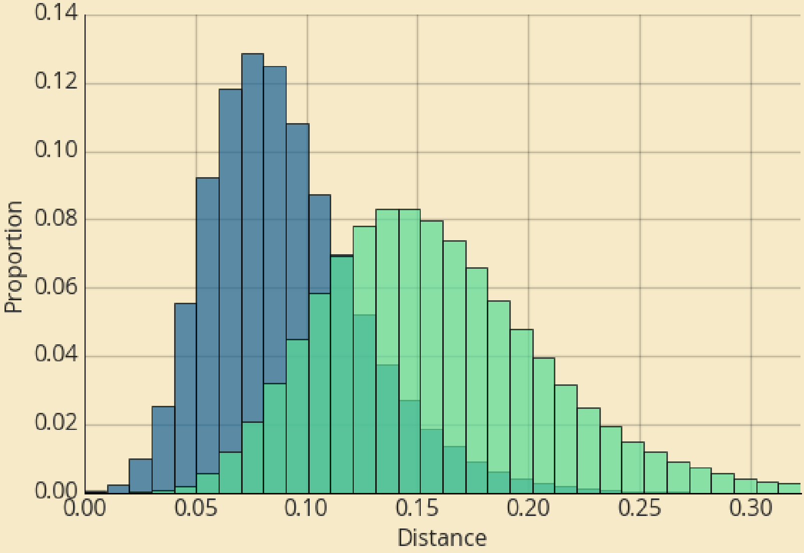


$$\mathcal{L} = \frac{1}{|\mathcal{P}|} \sum_{(x,y)\in\mathcal{P}} \|f(x) - g(y)\|_2^2 + \frac{\alpha}{|\mathcal{N}|} \sum_{(x,y)\in\mathcal{N}} \max(0, m - \|f(x) - g(y)\|_2)^2$$

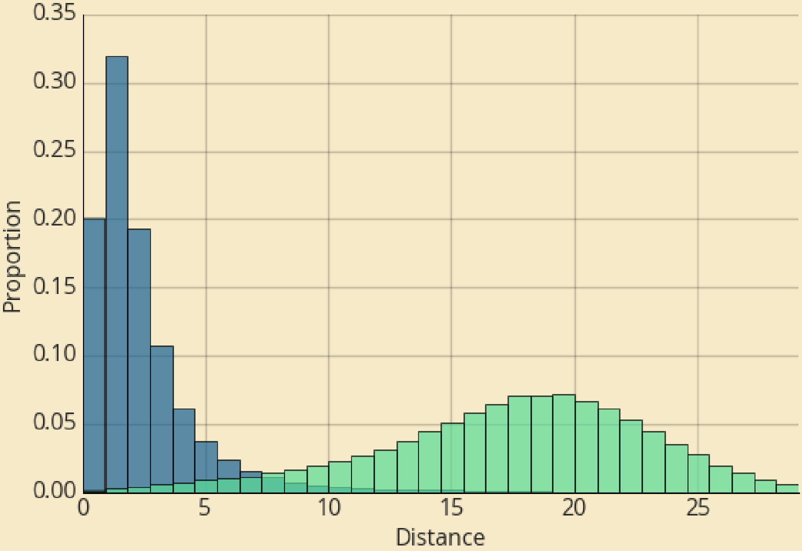
Example Output



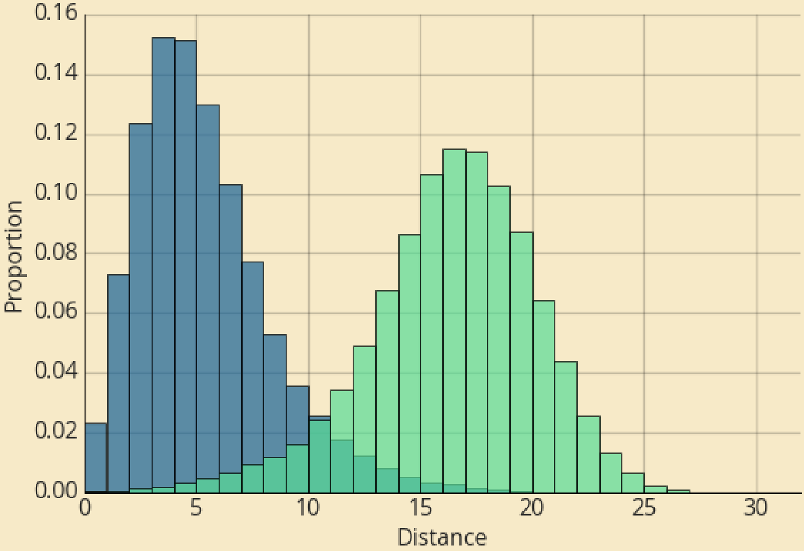
Raw Distance Distributions



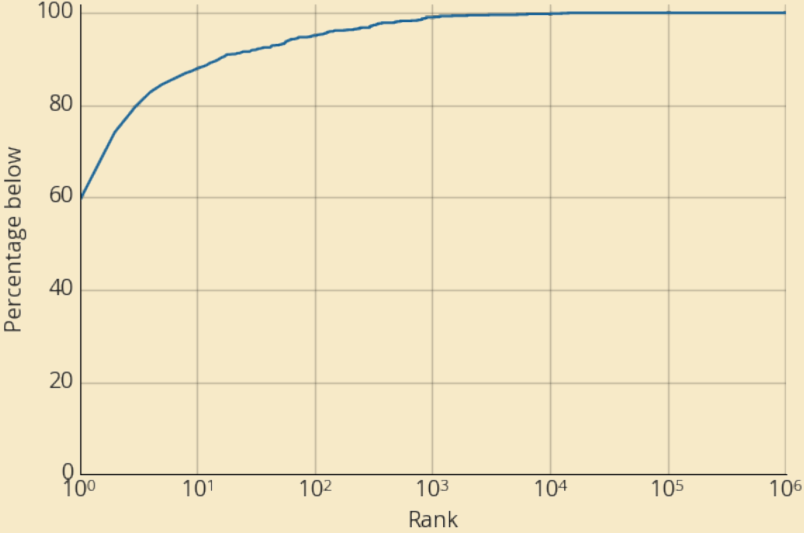
Output Distance Distributions



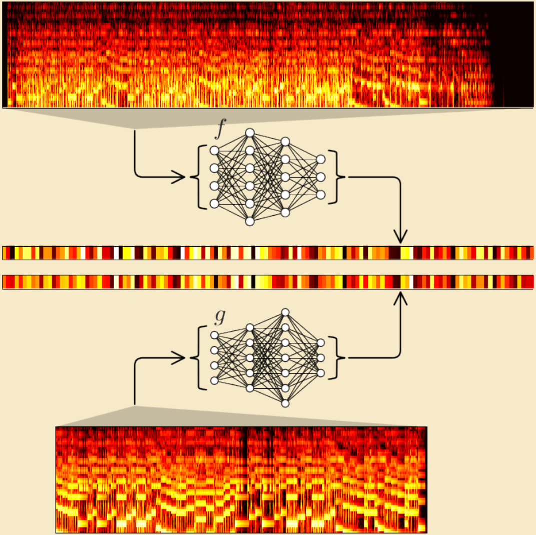
Hash Distance Distributions



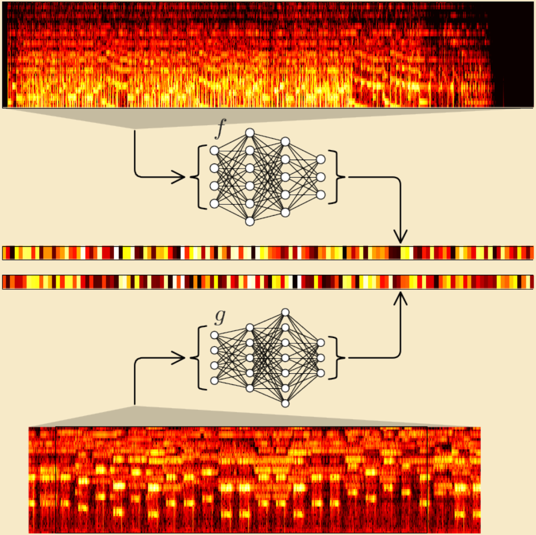
Match Ranks


100
80
60
40
20
0
Percentage below
10^0
10^1
10^2
10^3
10^4
10^5
10^6
Rank

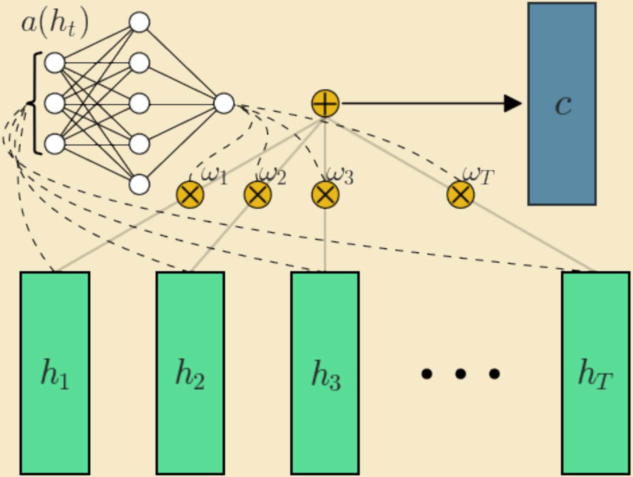
Pairwise Sequence Embedding



Pairwise Sequence Embedding

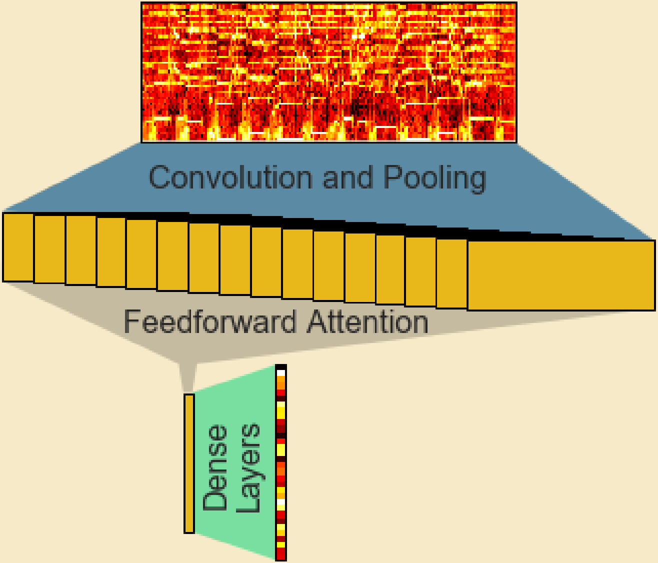


Feed-Forward Attention

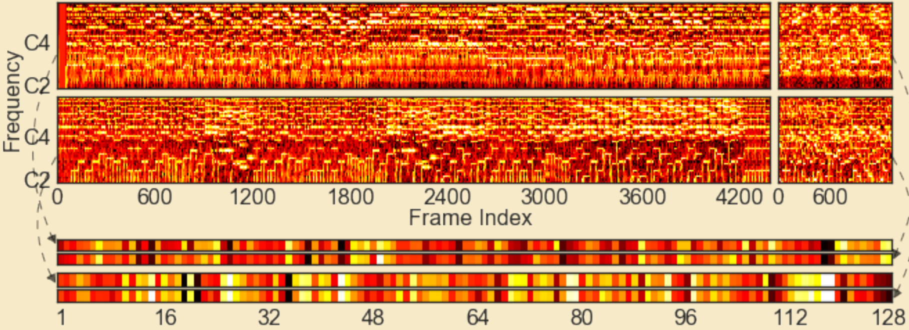


Raffel & Ellis, "Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems", ICLR 2016

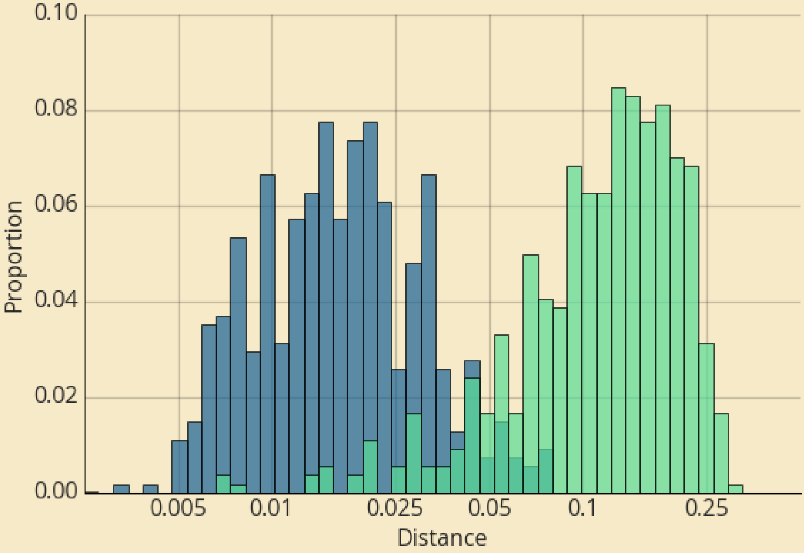
Embedding Network



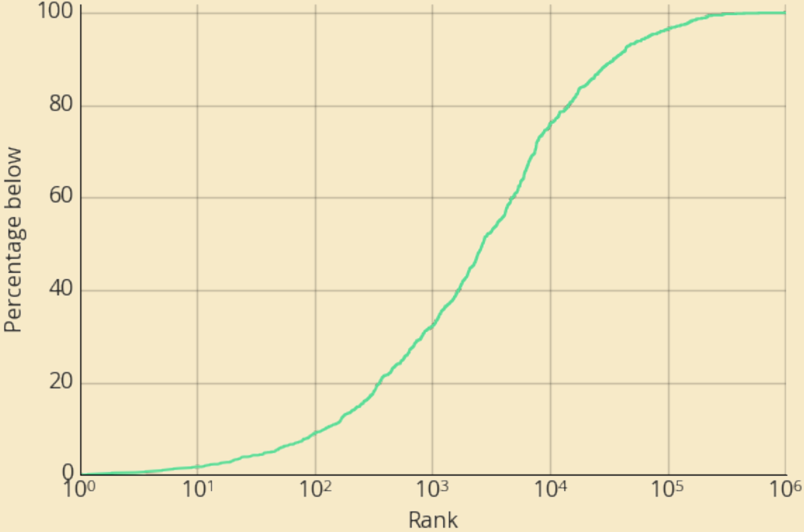
Example Embeddings



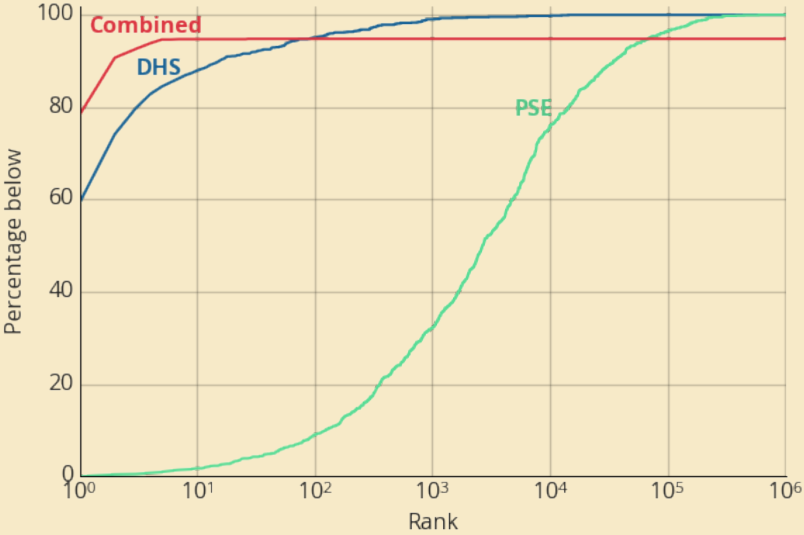
Embedding Distances



Match Ranks


100
80
60
40
20
0
Percentage below
10^0
10^1
10^2
10^3
10^4
10^5
10^6
Rank

Combined Match Ranks



References

[1] Raffel, “Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching”, PhD Thesis

[2] Raffel & Ellis, “Large-Scale Content-Based Matching of MIDI and Audio Files”, ISMIR 2015

[3] Raffel & Ellis, “Pruning Subsequence Search with Attention-Based Embedding”, ICASSP 2016