Cover Song Identification with Timbral Shape Sequences

Chris Tralie. chris.tralie@gmail.com
Ph.D. Candidate, NSF Graduate Fellow, Electrical and Computer Engineering, Duke University
Paul Bendich, bendich@math.duke.edu
Assistant Research Professor, Mathematics, Duke University

Abstract
We introduce a novel low level feature for identifying cover songs which quantifies the relative changes in the smoothed frequency spectrum of a song. Our key insight is that a sliding window representation of a chunk of audio can be viewed as a time-ordered point cloud in high dimensions. For corresponding chunks of audio between different versions of the same song, these point clouds are approximately aligned. We treat MFCC embeddings as point clouds and cast the problem as a relative shape sequence. We are able to correctly identify 42/80 cover songs in the “Covers 80” dataset. By contrast, all other work to date on cover songs exclusively relies on matching note sequences from Chroma derived features.

Beat-Synchronous Blocking And Windowing

- B beats per block. Take all such blocks in the song
- Take MFCC sliding window features to summarize each block
- MFCC Window size average beat interval

MFCC Self-Similarity Matrix Computation

SSM_{ij} = \|X[i] - X[j]\|_2
- Computed for each block if B contiguous beats for each song
- Invariant to rotation/translation
- Point center and sphere-normalize windows within each block to help make invariant to scale

"Don't Let It Bring You Down" Different gender singer, different instruments, different vocal/instrument balance

Smith Waterman Alignment with Diagonal Constraints

- Allow gaps in Smith Waterman for extra beats, but promote near-diagonal paths
- Score of 1 for matching SSMs
- Affine gap penalty −0.5 ≤ γ ≤ 0.7 (γ=1)
- For gap of length g
- Similar to approach in [2]

Results: "Covers 80"

- 80 pop song benchmark [1]. Best results reported in [4]
- Most correct top rank with our scheme: 44/80
Results below for 50 x 50 SSMs with 10 windows per beat (mean/median rank of correct song shown in parentheses)

Results: "Blurred Lines" Cross-Similarity Matrices

MCFC Self-Similarity

Chroma Optimal Transposition Index?

- Every 4 beats rhythmic pattern repeats itself (many diagonals)
- Note sequences are different, so traditional chroma-based approaches fail to recognize similarities