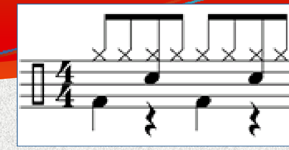
Calculating Musical Rhythm Similarity



Eric Battenberg

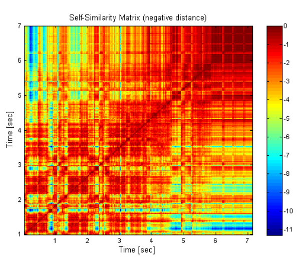
Center for New Music and Audio Technologies, University of California, Berkeley

Motivation

Automatic music information retrieval is an important topic in today's world due to the increasing prevalence of digital media in our daily lives. More and more people are amassing countless gigabytes of compressed digital audio on their hard drives and portable players. Most listeners are limited to sorting their music by artist, album, genre and other types of metadata tagged to their music files. This type of organization makes it very difficult to find songs that actually sound similar or have a common "feel".

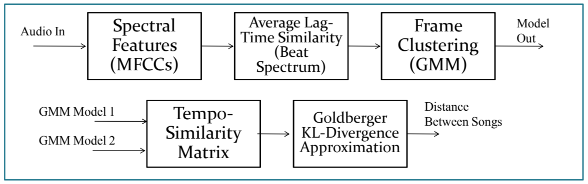
Music information retrieval attempts to solve this problem by providing supplemental information on a song that can be used to compare it to others. There are a number of techniques that compare songs by the timbre, or quality, of the audio; however, rhythmic similarity, though very important perceptually, hasn't seen as much action. Some authors have pursued methods to calculate individual characteristics of rhythm, such as tempo , meter, and "swing", but systems which characterize the patterns and overall rhythmic feel of a song are lacking.

We attempt to make progress in this area by modeling and comparing the self-similarity in songs.



Self similarity matrix from 7 seconds of AC/DC's Back in Black

Methods



Feature Extraction

1. Mel-band spectral energies are extracted in 23ms frames and compressed using the DCT.
2. The similarity between all near-by frames is calculated using Euclidean distance, and the results are stored in a similarity matrix.
3. Sums along the diagonals of the matrix are computed to yield the similarity for a particular lag-time (the beat spectrum [1]).
4. The resulting beat spectra are clustered using EM on a gaussian mixture model (GMM).

Feature Comparison

The model parameters belonging to two songs are then compared using the Goldberger approximation [2] of the KL-divergence of a GMM. In this project, the diagonal covariance matrix in each model is first transformed using a tempo-similarity matrix, which is explained in the following section. The modification is as follows:

$$\Sigma' = \Sigma^{1/2} S_T^{-1} \Sigma^{1/2}$$

This results in a distance measure that is hopefully more meaningful to the human auditory system.

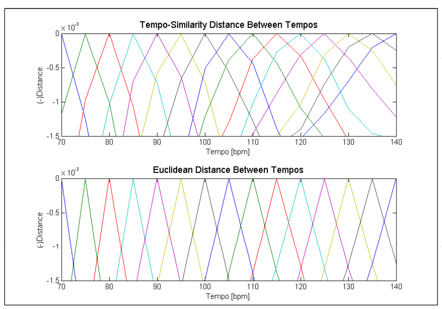
Tempo-Similarity Distance

Other work [1] has proposed using the Euclidean distance between beat spectra to compare rhythm. This approach fails to capture the perceptual similarity of neighboring tempos, especially the increasing similarity of neighboring tempos as the rate increases. To address this issue, we propose the Tempo-Similarity Distance (TSD) :

$$(x-y)^T S_T (x-y)$$

Where S_T is a tempo-similarity matrix made up of triangular kernels with widths proportional to the "just noticeable difference" [3] at a particular tempo (about ± 8%).

Below, the (negative) tempo-similarity distance is demonstrated using a basic rock beat at various tempos. Compared to results of the Euclidean distance below, it shows increasing similarity at faster tempos.

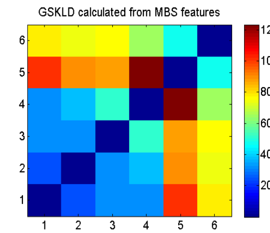


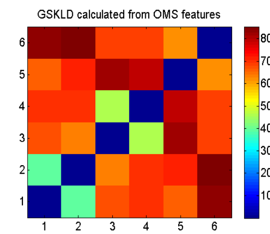
The tempo-similarity distance can be adjusted to accurately model how the human auditory system perceives tempos.

Song Comparisons

Real-World Effectiveness

- The system was tested on a small sampling of popular music (listed below).
- The Goldberger-Similarity KL-Divergence (GSKLD) was calculated for each pair of songs.
- The upper matrix shows the distance between models generated from MFCC beat spectrum (MBS) features.
- The lower matrix was calculated using the modulation spectral note onset energies (OMS), cf. [4].
- The MBS features yield markedly better discrimination between genres and match well with perceptual judgments.



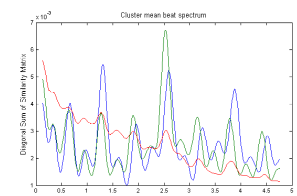


Example Songs

1. Bob Marley – Is This Love
2. Bob Marley – Buffalo Soldier
3. Santana – Black Magic Woman
4. Santana – Evil Ways
5. AC/DC – Back in Black
6. AC/DC – Deep in the Hole

Conclusions

- The tempo-similarity distance more accurately models human tempo perception.
- The beat spectrum is an effective feature to be used in rhythmic comparisons.
- There are many parameters used in this implementation that can be optimized to yield better results.
- A much larger database of test music needs to be used to gauge the robustness of this system.



Mean beat spectra from the first three clusters (of five) from AC/DC's Back in Black

References

1. Foote, J. et. al, "Audio retrieval by rhythmic similarity", in *Proceeding of the International Conference on Music Information Retrieval*, 2002, vol. 3, pp. 265-266.
2. Goldberger, J. et. al, "An efficient image similarity measure based on approximation of KL-divergence between two gaussian mixtures," in *Proceeding of ICCV 2006*, Nice, October 2003, vol. 1, pp. 487-493.
3. Thomas, K., "Just noticeable difference and tempo change", in *Journal of Scientific Psychology*, May 2007, pp. 14-20.
4. Klapuri, A. et. al, "Analysis of the meter of acoustic musical signals", in *IEEE Transactions on Audio, Speech, and Language Processing*, January, 2006, vol. 14, no. 1.

Eric Battenberg, eric@cnmat.berkeley.edu