Timbre Similarity

By Daniel McEnnis

Outline

- Timbre similarity by MDS
  - Grey and Gordon
- Timbre Similarity and Classification
  - Fujinaga
  - Fujinaga and MacMillan
  - Brown
  - Herrera, Yeterian, and Gouyon
  - Tindale, Kapur, Tzanetakis, Fujinaga
- Concluding thoughts

Grey and Gordon 78

- Similarity of Timbre by Psychoacoustics
  - Attempts to find acoustically relevant features
  - Utilizes similarity judgments by participants for raw data
  - Correlate results using MDS
  - 16 tones from 14 instruments
  - Each subject listens to every possible combination of tones

Grey and Gordon 78

- MDS – Multi Dimensional Scaling
  - Take a set of distance measurements and calculate the smallest possible number of dimensions that can 'adequately' explain results
  - Take guesses as to what the features are and test for fit
- Finds 3 dimensions 'adequately' explain results
- Finds spectral centroid as 1 feature with a high correlate, but none found for other dimensions

Fujinaga 98

- Data source
  - McGill sound source – 39 instruments at different pitches for 1338 individually played notes
- Features
  - First 10 Moments of spectrum
  - Fundamental frequency
  - Amplitude
  - Partials above fundamental

Fujinaga 98

- Features Used
  - Fundamental Frequency
  - 0th - 3rd order moment
  - Standard deviation
  - First 2 partials
- Method
  - K Nearest Neighbor with Genetic Algorithms
- Results
  - 64% over all 39 instruments with high of 90% and low of 8%

MacMillan and Fujinaga

- Real time system
- Data Source
  - SHARC database
- Features
  - First 10 Moments of spectrum
  - Fundamental frequency
  - Amplitude
  - Partials above fundamental
  - Velocity of centroid
  - Variance of centroid
  - Spectral irregularity
  - Tristimulus

MacMillan and Fujinaga

- Method Used
  - K Nearest Neighbor with Genetic Algorithms
- Results
  - 68% correct classification across all 39 instruments.

Brown 99

- Data Source
  - Music from Wellesley College Music Library Collection (not samples)
  - Music from the personal collection of Jay Panetta (not samples)
- Training set of 1 minutes sax and 1 minute oboe
  - Chosen by which sample gives best results
- Test set of 52 sax and 28 oboe excerpts

Brown 99

- Features
  - Cepstrum Coefficients
- Model
  - Gaussian classifiers
- Results
  - Not easily summarized

Herrera, Yeterian, Gouyon 2002

- Data Source
  - 234 drum sounds from various commercial cd's
- Features
  - Attack
    - Energy
    - Log attack time
    - Zero crossing rate
    - Temporal centroid
    - TC/EA (temporal centroid / attack length)
  - Decay
    - Spectral flatness
    - Spectral centroid + variance
    - Strong peak
    - Spectral kurtosis

Herrera Yeteran Gouyan 2002

- Decay cont.
  - Zero crossing rate + variance
  - Skewness
- Other
  - 8 empirically chosen bands
  - 13 MFCC

- Model
  - Pre-filtered by none, Correlation Based Feature Selection, or ReliefF
  - K Nearest Neighbor, Canonical Discriminant analysis, and C4.5

Herrera Yeterian Gouyan 2002

- Results
  - KNN 99.2%
  - C4.5 97.2%
  - Canonical Discriminant Analysis 99.1%

Tindale et al. 2004

- Data Source
  - Self collected drum hit sounds
- Features
  - Spectral Flux
  - Spectral Rolloff
  - Spectral Centroid
  - Spectral Kurtosis
  - Spectral Skewness
  - MFCC (12)
  - Linear Prediction Coding Coefficients
  - 9 wavelet bands and their variances

Tindale et al. 2004

- Model
  - Neural Net (6 hidden and 7 output nodes)
  - K Nearest Neighbor
  - Support Vector Machine
- Results
  - Neural Nets 89%
  - Nearest Neighbor 94.9%
  - Support Vector Machine 68.8%

Concluding Thoughts

- Not nearly everything, but its a start
- Most early work is psychoacoustic studies, not classification experiments