Classification of Timbre Similarity

Corey Kereliuk
McGill University

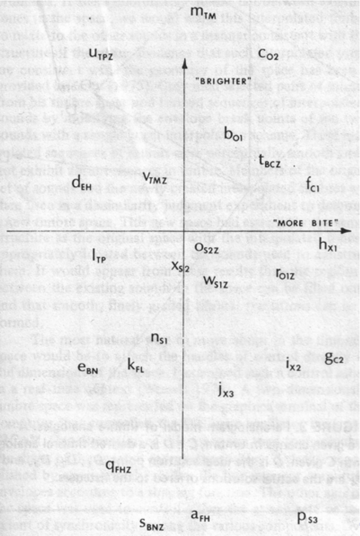
March 15, 2007

What Timbre is Not

- ⇒ A unified definition of timbre seems elusive
- ⇒ "Timbre tends to be the psychoacoustician's multidimensional waste-basket category for everything that cannot be labeled as pitch or loudness" (McAdams79)
- ⇒ OED definition: "The character or quality of a musical or vocal sound (distinct from its pitch and intensity) depending upon the particular voice or instrument producing it, and distinguishing it from sounds proceeding from other sources"
- ⇒ "Timbre refers to the 'color' or quality of sounds, and is typically divorced conceptually from pitch and loudness" (Wessel79)
- ⇒ All of these definitions describe timbre by saying what it is *not*

What Timbre is

- ⇒ "Perceptual research on timbre has demonstrated that the spectral energy distribution and temporal variation in this distribution provide the acoustical determinants of our perception of sound quality" (Wessel79)
- ⇒ Wessel collected perceptual dissimilarities through a series of listening tests:
  - Listeners were played two sounds and asked to rate how similar (on a scale [0-9]) the two sounds were
  - This produced $n(n-1)/2$ observations (in this case $n = 24$ orchestra instruments) which were organized into a 24x24 *dissimilarity matrix*
  - A multi-dimensional scaling algorithm was used to create a 2-dimensional *timbre space*, in which the dissimilarity between instrument timbres was proportional to their euclidean


"BRIGHTER"
"MORE BITE"
m_{TM}
u_{TPZ}
c_{O2}
b_{O1}
t_{BCZ}
d_{EH}
v_{TMZ}
f_{C1}
h_{X1}
o_{S2Z}
l_{TP}
x_{S2}
r_{O1Z}
w_{S1Z}
n_{S1}
e_{BN}
k_{FL}
i_{X2}
g_{C2}
j_{X3}
q_{FHZ}
s_{BNZ}
a_{FH}
p_{S3}

Motivation

- ⇒ Psychoacoustic studies
- ⇒ Musicological analyses
- ⇒ Source separation
- ⇒ Instrument identification
- ⇒ Content-based management systems for the navigation of large catalogues
- ⇒ Composition
- ⇒ Identifying bird calls from the same species
- ⇒ Speaker identification
- ⇒ etc.

Considerations

⇒ Whether to focus on *monophonic* or *polyphonic* timbres?

⇒ Whether to use *local* or *global* features?

⇒ Which local/global features to use (infinite possibilities)

⇒ Perceptual relevance of results

Common Approaches

⇒ Monophonic timbre similarity is relatively well understood

⇒ There is still much to be discoverd about polyphonic timbre similarity

⇒ Commonly used tools:

- Mel-Frequency Cepstrum Coefficients (MFCCs)
- Spectral Centroid
- Log-attack-time
- Principle Component Analysis (PCA)
- Spectral Flatness (Degree of noisy-ness)
- k-NN
- GMMs, HMMs, GAs, NNs

Long-term Statistics

⇒ In order to get a sense of the *global* spectral envelope of a signal:

- Compute the MFCC on N sequential frames
- Average the N frames together

⇒ One might expect the result to be flat or noisy, however, it turns out that a *global* shape emerges, which tends to be quite specific to a given texture

Long-term Statistics

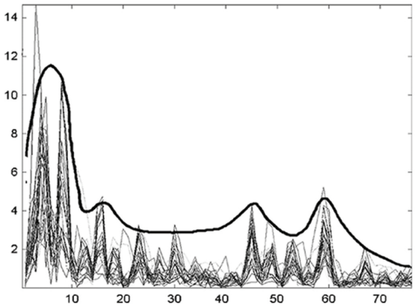


Figure: Global Spectral Shape(Aucouturier 2005)

Modeling the Global Spectrum

⇒ Aucouturier (2005) proposes modeling the MFCCs as a mixture of Gaussians:

$$p(F_t) = \sum_{m=1}^{M} \pi_m \aleph(F_t, \mu_m, \Gamma_m) \qquad (1)$$

⇒ Here the feature vector F_t at time t (MFCCs in this case) is modeled as the sum of M Gaussians with mean μ_m and variance Γ_m

⇒ The GMM is initialized by k-mean clustering and trained using the classic EM algorithm

Modeling the Global Spectrum

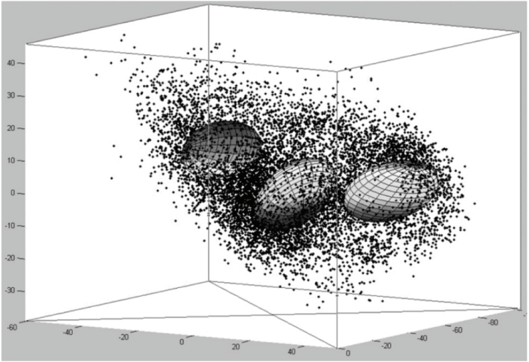


Figure: GMM Clustering (Aucouturier 2005)

⇒ In order to compare the timbral similarity of two songs:

- A GMM is computed for each song
- A large number of sampling points are evaluated to compute the likelihood that they could have come from the song under comparison
- This is illustrated by the following equation:

$$D(A,B) = \sum_{i=1}^{N} logP(S_i^A|A) + \sum_{i=1}^{N} logP(S_i^B|B) - \sum_{i=1}^{N} logP(S_i^A|B) - \sum_{i=1}^{N} logP(S_i^B|A)$$

⇒ N is the number of sampling points used

⇒ D is a probabilistic distance measure assessing the similarity between song A and song B

Timbral Similarity Results

⇒ Global Timbral Similarity Implemented in CUIDADO music browser

⇒ A query for "Ahmad Jamal - L'instant de Verite" —a jazz piano recording returns similarity results which all contain *romantic-styled* piano. For example, New Orleans Jazz (*G. Mirabassi*), Classical Piano (*Schumann, Chopin*)

⇒ Some of the most interesting results are unexpected (different genres and cultural backgrounds)

Evaluation

- ⇒ Finding an evaluation metric for this type of system would be difficult
- ⇒ The MIR community has ‘hotly’ debated the subject of evaluation
- ⇒ At this time standard test databases need to be developed in order to compare different techniques
- ⇒ There is also the question of what exactly defines similarity?
- ⇒ Comparing to hand segmented/clustered results might not be adequate since unexpected results (false-negatives) might be missed

The End