Classifying music audio with timbral and chroma features

Dan Ellis • Columbia University • dpwe@ee.columbia.edu

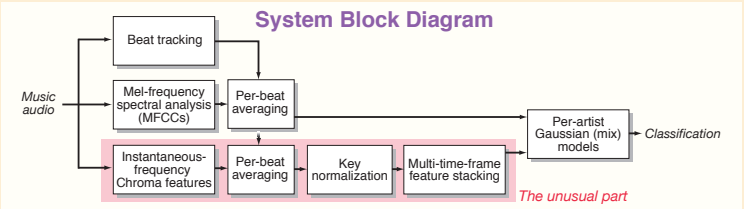
Columbia University

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

Summary: In addition to conventional cepstra, we model the **covariance of chroma** (melodic/harmonic) features and gain a small improvement in a 20-way pop music artist identification.

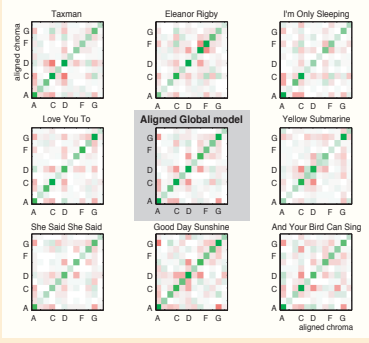
Introduction

- "Classic" approaches to music audio classification model the covariance of **spectral features** (e.g. MFCCs) as distributions [1] or discriminatively [2].
  These appear to work by reflecting the instrumentation, which correlates well with genre or band [3].
- **Chroma features** [4] attempt to represent the pitch content (melody and harmony) while minimizing the influence of instrumentation.
- We investigate using the **covariance of chroma features** as a basis for music classification by artist. Although weak on their own, chroma features can improve classification when combined with spectral features
- This suggests that artists have **particular harmonic combinations** or motifs that can be automatically recognized.



Key Normalization

- Similar melodic/harmonic gestures occur 'relative' to the **different keys** of individual songs.
- **Key normalization** attempts to transpose (rotate) the chroma features to a canonical key prior to modeling.
- We do this by:
  - build a **chroma covariance** matrix from **all songs**
  - **transpose** each song to maximize likelihood under global model
  - **re-estimate** global model from transposed songs and repeat.



Data - "artist20" dataset

- 1413 tracks, from 120 albums (20 artists x 6 albums) [5]
- Contemporary pop music drawn from uspop2002 [6] and others
- Studio albums, chosen for chronological & stylistic consistency
- MFCCs, Chroma features, etc. available for **download** at
  `http://labrosa.ee.columbia.edu/projects/artistid/`

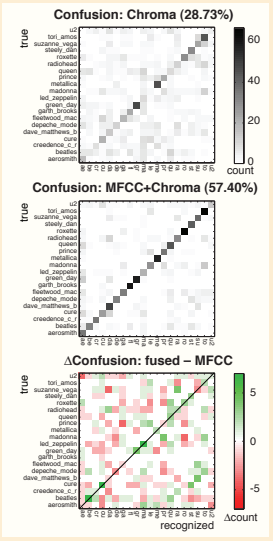
Results

Feature	Model	T win	Acc	Exec time
MFCC20	FullCov	1	**56%**	127 s
MFCC20	64 GMM	1	56%	563 s
Chroma	FullCov	1	14%	21 s
Chroma	FullCov	4	20%	57 s
Chroma	64GMM	1	25%	337 s
Chroma	64GMM	4	29%	1060 s
ChromaKN	FullCov	1	23%	70 s
ChromaKN	FullCov	4	28%	197 s
ChromaKN	64GMM	1	32%	516 s
ChromaKN	64GMM	4	**33%**	1238 s
MFCC + Chroma fusion			**59%**	

- MFCC-based classification is much more accurate (56%) than using Chroma features (33% at best), but **combining the two** does give further gains (to 59% correct; McNemar $p < .001$).
- A single, full-covariance Gaussian is adequate to model the MFCC data, but Chroma data need a **64-mix GMM**.
- **Key normalization** (ChromaKN) is important for Chroma features.
- Stacking Chroma features from up to 4 adjacent beat-times (T win = 4) **helps a little**.

Discussion and Conclusions

- Covariance of single beat-chroma vectors **carries some information** about artist, beyond that captured by MFCCs. This could relate to e.g. an artist's **preferred chords**.
- Used alone, chroma features do quite well on **a few artists** (tori_amos, metallica) but learn almost nothing about others (madonna, beatles).
- Our best performance came from a simple **weighted sum of likelihoods** from separate MFCC and Chroma models. Differences were small, but gains were on different artists than those classified best by Chroma alone (beatles, led_zeppelin).



References

[1] J.-J. Aucouturier & F. Pachet, "Music similarity measures: What's the use?" ISMIR, Paris, 2002.
[2] M. I. Mandel &D. P. W. Ellis, "Song-level features and support vector machines for music classification," ISMIR, London, 2005.
[3] J. H. Jensen, M. G. Christensen, S. H. Jensen, "A framework for analysis of music similarity measures," EUSIPCO, 2007.
[4] M. A. Bartsch & G. H. Wakefield, "To catch a chorus: Using chroma-based representations for audio thumbnailing," WASPAA, Mohonk, 2001.
[5] D. P. W. Ellis, "Music artist identification: artist20 baseline system in Matlab," web resource, 2007. `http://labrosa.ee.columbia.edu/projects/artistid/`
[6] D. P. W. Ellis, A. Berenzweig, B. Whitman, "The uspop2002 Pop Music data set," web resource, 2003. `http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html`

MATLAB code to run this system is available at:
`http://labrosa.ee.columbia.edu/projects/timbrechroma/`

for ISMIR'07 • 2007-09-15, revised/corrected 2007-09-25 dpwe@ee.columbia.edu