Using Bass-line Features for Content-based MIR

1. Our motivation

Content-based MIR

- To retrieve musical pieces based on the analysis of the content of them
- Many attempts by many researchers
  e.g. Automatic genre classification by G. Tzanetakis (2002)
  Computational models of music similarity by A. Berenzweig (2003), E. Pampalk (2006)
- Common problem in most existing studies: *Use of low-level features*
  e.g. Spectral centroid, rolloff, flux, MFCCs, zero crossing rates
  by many researchers including G. Tzanetakis, E. Pampalk, J.-J. Aucouturier, A. Berenzweig
  Chroma features (a.k.a. pitch-class profile)
  by J. P. Bello (2005), D. P. W. Ellis (2007), and so on…

Spectral / cepstral features

- Roughly represent timbres but directly represent a frequency characteristic
- *Do not clearly or directly correspond to any of musical aspects* (melody, harmony, rhythm)

Chroma features

- Approximately represent harmony
- Useful for chord recognition (T. Yoshioka 2004), Chorus detection (M. Goto 2003), etc.

We believe:

We need new audio features that are *directly related to a musical aspect* for further improvement of content-based MIR

E. Pampalk also says, using only low-level statistics, there is a clear limit to improve the quality of musical similarity.

2. Strategy for designing new features

Strategy I

Mid-level representation

- To transform low-level audio features to mid-level representation using pattern recognition
  e.g. Instrogram (T. Kitahara, 2006)

Strategy II

Features of a particular instrument part

- To extract a particular instrument part from poly-phony and then extract features from that part
  e.g. MIR based on vocal timbre (H. Fujihara, 2007)

Why the bass part?
- It plays important roles for both rhythm and harmony
- A method for extracting bass lines from polyphony exists.

We adopt *Strategy II* and focus on the *bass part*

Appendix: Low-level features used in our experiments

I. Timbral (spectral) features

- Features same as L. Lu's spectral features

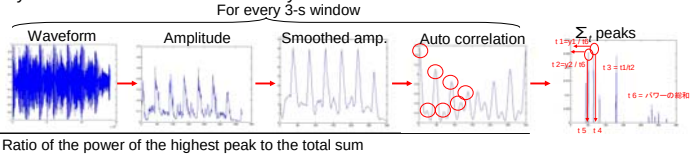
Intensity	Sum of intensities for all frequency bins
Sub-band intensity	Intensity of each sub-band (7 sub-bands prepared)
Spectral centroid	Centroid of the short-time amplitude spectrum
Spectral rolloff	85th percentile of the spectral spectrum
Spectral flux	2-norm distance of the frame-to-frame spectral amplitude difference
Bandwidth	Amplitude weighted average of the differences between the spectral components and the centroid
Sub-band peak	Average of the percent of the largest amplitude values in the spectrum of each sub-band
Sub-band valley	Average of the percent of the lowest amplitude values in the spectrum of each sub-band
Sub-band contrast	Difference between "peak" and "valley" in each sub-band

- MFCCs (up to 5th)

II. Rhythmic features

Mostly same as G. Tzanetakis et al.'s rhythmic features

For every 3-s window

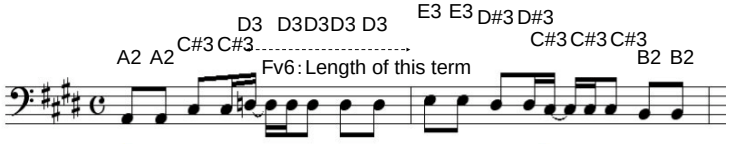


Fr1: Ratio of the power of the highest peak to the total sum
Fr2: Ratio of the power of the second-highest peak to the total sum
Fr3: Ratio of Fr1 and Fr2
Fr4: Period of the first peak in BPM
Fr5: Period of the second peak in BPM
Fr6: Total sum of the power for all frames in the window

Yusuke Tsuchihashi[*1], ○*Tetsuro Kitahara*[*1,*2], *and Haruhiro Katayose*[*1,*2]
[*1]*Kwansei Gakuin University, Japan;* [*2]*CrestMuse Project, CREST, JST, Japan*
t.kitahara@kwansei.ac.jp, http://ist.ksc.kwansei.ac.jp/~kitahara/

3. Design of Bass-line Features

I. Features of pitch variability



Fv1: # of different pitches that appear in at least one frame
= 6 (A2,C#3,D3,E3,D#3,B2)

Fv2: # of pitches from Fv1 excluding those with an appearance rate of less than 10% or 20%
= 5 (A2,C#3,D3,E3,D#3) for 10%

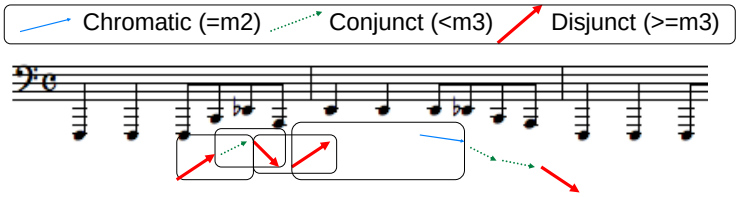
Fv3: Temporal mean of # of pitches within a sliding short-term window = 3.5 (if the unit time is 1 measure)
(1st meas.: {A2,C#3,D3}, 2nd meas: {E3,D#3,C#3,B2})

Fv4: % of appearing freq. of the i-th most frequently appearing pitches (i=1,…,5) = 0.53125 for i = 2

Fv5: Pitch interval between the 2 most frequently appearing pitches = 1 (1st: D3, 2nd: C#3)

Fv6: Period of the most frequently appearing pitch

II. Features of pitch motions



Ft1: Mean of # of pitch motions per unit time
= 4 (if the unit time is 1 measure)

Ft2: % of each motion pattern
Chromatic = 1/8, Conjunct = 3/8, Disjunct = 4/8

Ft3: % of each pattern of successive motions
Conj. + Conj.= 2/7, Conj. + Disj. = 2/7
Disj. + Conj. = 2/7, Disj. + Disj. = 1/7

These features are extracted from both pitch sequence and pitch class sequence (i.e. octave ignored) of the bass line

After extraction, the dimensionality is reduced to remove highly correlated features

4. Application to Genre Classification

- After bass-line, timbral, and rhythmic features are extracted, the genre is identified based on the Mahalanobis distance.
- Leave-one-out cross validation with data for ISMIR'04 Audio Description Contest (300 songs in total)

Without bassline

	Pop &Rock	Metal &Punk	Electronic	Jazz &Blues	Classical	World
Pop&Rock	21	6	11	0	0	3
Metal&Punk	13	42	8	0	0	4
Electronic	3	1	5	0	0	0
Jazz&Blues	7	1	12	48	4	7
Classical	2	0	9	0	31	20
World	4	0	5	2	15	16
Accuracy	42%	84%	10%	98%	62%	32%

Avg: **54.3%**

With bassline

	Pop &Rock	Metal &Punk	Electronic	Jazz &Blues	Classical	World
Pop&Rock	23	5	10	5	0	7
Metal&Punk	10	42	1	1	0	1
Electronic	8	0	23	1	0	6
Jazz&Blues	5	2	6	37	2	10
Classical	0	0	3	1	45	8
World	4	1	7	5	3	18
Accuracy	46%	84%	46%	74%	90%	36%

Avg: **62.7%**

- Feature distribution after PCA

5. Application to Music Collection Visualization

- Our bass-line, timbral, and rhythmic features are applied to Music Islands (E. Pampalk, 2006)
- Users can enhance a particular musical aspect by giving different weights to different-type features