

Music Information Retrieval

Andreas Rauber

<http://www.ifs.tuwien.ac.at/~andi>

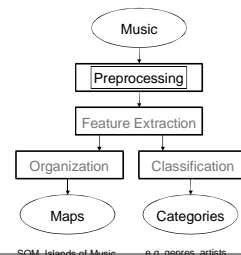
Feature extraction

- FFT-based: means and variance of
 - Spectral Centroid: center of gravity
 - Spectral rolloff: frequ. below which 85% of energy
 - Spectral flux: squared difference btw. spectral histograms
 - Low energy: % of windows with less than avg. energy
 - Zero crossings: zero-line crossings of signal

Feature extraction

- MFCCs: 5 FFT bins grouped by Mel frequ. scale
- Mpitch: folded/unfolded pitch histograms
- Beat histogram: DWT-based
- System: Marsyas (George Tzanetakis)

RP Feature Extraction



RP Feature Extraction

- PCM Audio Signal
- Power Spectrum
- Frequency Bands
- Masking Effects
- Phon
- Sone

RP Feature Extraction

- Loudness Modulation Amplitude
- Fluctuation Strength
- Filter (Gradient, Gauss)
- Median (opt.)
- 1,440-dim feature vec.

Rps Feature Extraction

- What do the features capture? (FAQ!)
- It is not
 - Rhythm
 - Pitch/melody
 - Energy
- It is all of the above to some degree: complex rhythmic patterns

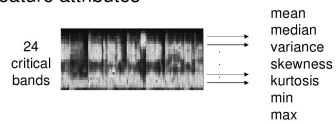
Feature Extraction Summary

- 1,440-dim feature vector
- representing fluctuation patterns
- loudness modulation frequ. in different frequency bands
- complex representations of rhythmical characteristics
- combined with other feature sets
 - energy
 - roughness
 - zero-crossings
 - ...
- basis for classification, clustering and retrieval

New Feature Sets (1/2)

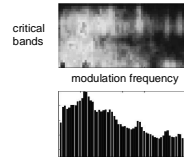
Statistical Spectrum Descriptor (SSD)

- description of each of the 24 critical bands of the Sonogram by statistical moments
- 168 feature attributes



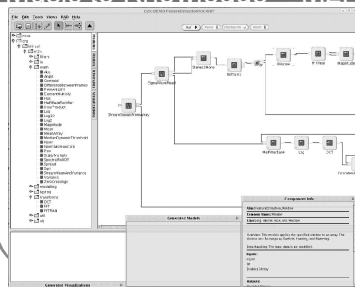
New Feature Sets (2/2)

Rhythm Histogram (RH)



- histogram of modulation magnitude per modulation frequency
- 60 bins -> 60 feature attributes

Music to Knowledge - M2K



Evaluation method

- Evaluation of psycho-acoustic transformations on 3 standard databases:
 - GT: 1000 instances, 10 genres (George Tzanetakis, 2002)
 - IR: 698 instances, 8 genres (ballroom dances) (ISMIR 04 Audio description contest, rhythm classification)
 - IG: 1458 instances, 6 genres (ISMIR 04 Audio description contest, genre classification)
- music genre classification, using SVM
- 10-fold cross-validation, measuring:
 - macro-averaged precision
 - macro-averaged recall
 - F1 measure
 - Accuracy

Comparison of Feature Sets

	databases (% Acc.)		
	GT	IR	IG
■ Rhythm Patterns' baseline	58.5	81.7	71.0
■ Rhythm Patterns' best variant	64.4	82.8	75.0
■ SSD (168 attributes)	72.7	54.7	78.5
■ RH (60 attributes)	44.1	79.9	63.2

(* RP: 1400 attributes)

Combination of Feature Sets

	databases (% Acc.)		
	GT	IR	IG
■ Rhythm Patterns baseline	58.5	81.7	71.0
■ Rhythm Patterns best variant	64.4	82.8	75.0
■ SSD (168 attributes)	72.7	54.7	78.5
■ RH (60 attributes)	44.1	79.9	63.2
■ RP + SSD	72.3	83.5	80.3
■ RP + RH	64.2	83.7	75.5
■ SSD + RH	74.9	82.7	79.6
■ all 3	72.4	84.2	80.0

ISMIR 2004 Audio Description

- Genre Classification Contest
 - 70.4 % Accuracy
 - 4th rank
- Rhythm Classification Contest
 - 82 % Accuracy
 - 1st rank

2005 Audio Description Contest

- MIREX 05
Music Information Retrieval Evaluation
eXchange
- Intended to establish as *the* Music IR
evaluation platform
 - Submission June 2005
 - Evaluation/Presentation September 2005 @ ISMIR
 - results:
<http://www.music-ir.org/evaluation/mirex-results/audio-genre/index.html>