Tutorial: Music Similarity

Abstract

The first part of this tutorial is an introduction to the computation of audio and web-based music similarity. This tutorial will cover low-level audio statistics related to timbre and rhythm as well as the application of text information retrieval techniques to MIR. In particular, the algorithm which won the ISMIR'04 genre classification contest will be described. We illustrate the use of these techniques for playlist generation and genre classification.

Elias Pampalk
Austrian Research Institute for Artificial Intelligence (OFAI)

General Context

- Available Technology

  cheap and fast broadband Internet (incl. e.g. UMTS), mass storage, computation power, encoding algorithms (MP3 etc.), …

- Market (digitized music)
  - online shops with > 1 million tracks
  - mobile audio players
  - …
  - additional non-mainstream opportunities
    - “creative commons”
    - “old” music where limited usage rights are expiring

- Problem
  - inefficient retrieval/browsing tools limit value of large collections
  - manual (e.g. genre) categorization is too expensive

- Solution?

  MIR in general and specifically similarity as core technology for retrieval/browsing applications

Tutorial Goals

1. What is music similarity? (Definition?)
2. What is it good for? (Applications?)
3. How (and from what) can similarity be computed?
4. How to evaluate the algorithms?
5. What are the limitations?
6. What are future directions?
7. What is happening at ISMIR'05?

Tutorial Outline

- **Source: Audio**

  → track level

  1. demo (playlist generation)
  2. techniques
  3. evaluation
  4. limitations and future directions

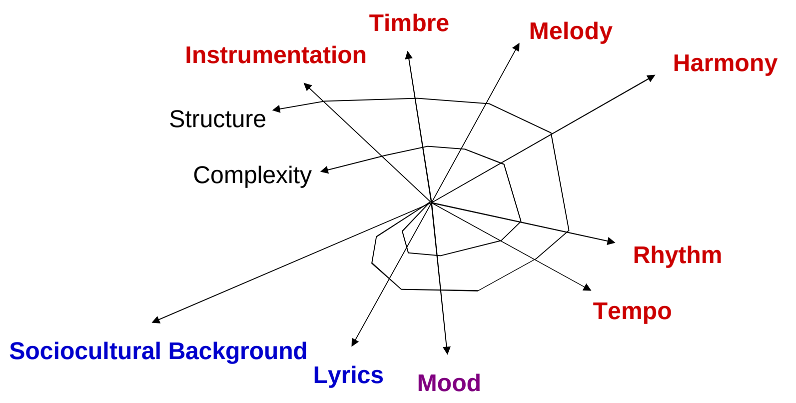
- **Source: Web-pages**

  → artist level (exception: e.g. lyrics)

  1. technique
  2. demo (hierarchical organization of music collection)

Music Similarity

Perception of similarity is subjective and context dependant
Important dimensions include:



Demonstration: Playlist Generation

Scenario

- Music: private collection (1,000 - 20,000 tracks)
- Hardware: e.g. mobile audio player
- User: minimal interaction (“lazy”)

Basic Idea

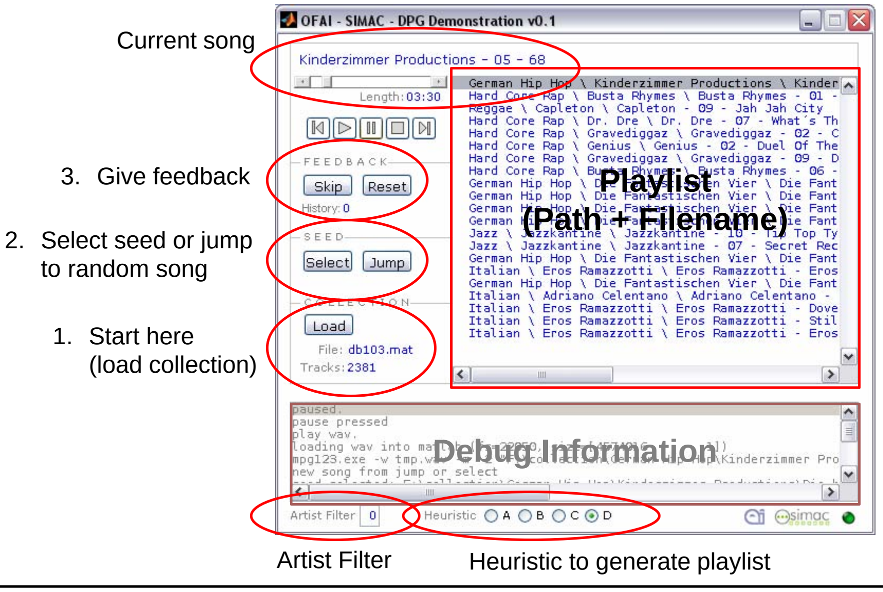
use audio-based similarity to create playlist given seed
(and take user feedback into account)

Details

see ISMIR’05 poster session Monday afternoon:
“Dynamic playlist generation based on skipping behaviour”

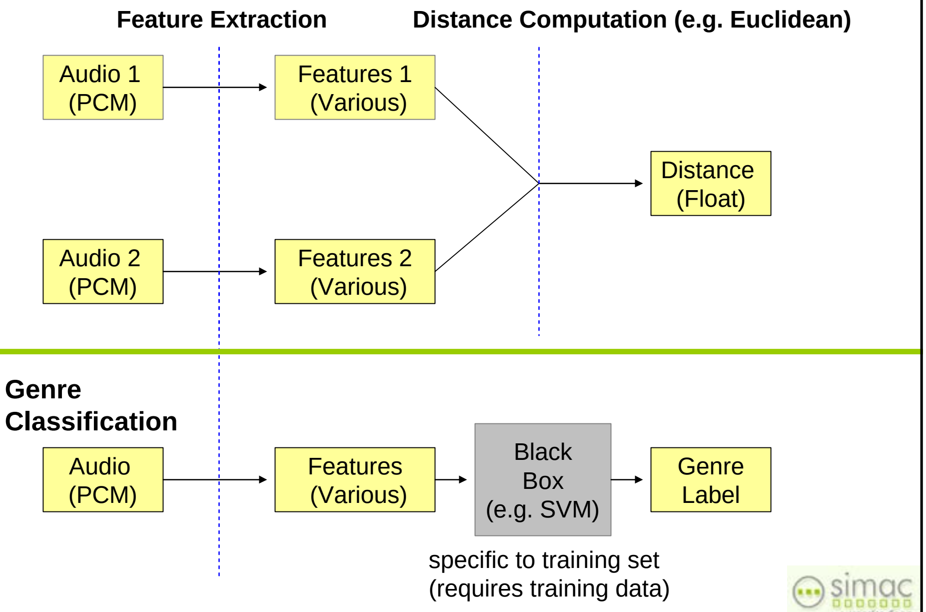
Demonstration: Playlist Generation

7



Music Similarity: Schema

8



Audio Features (Descriptors): Types and Scope

Data Types
- single numerical value (e.g. ZCR)
- vector (e.g. MFCCs)
- matrix or n-dimensional histograms (e.g. fluctuation patterns)
- multivariate probability distribution (e.g. spectral similarity)
- anything else (e.g. sequence of chords)

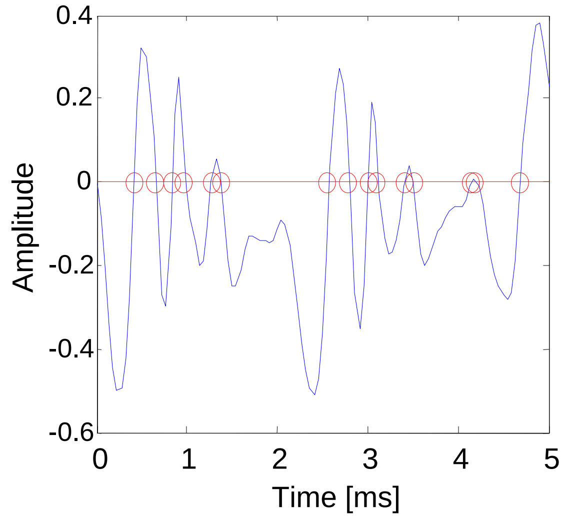
Temporal Scope
- frame (e.g. 20ms, usually: 10ms-100ms)
- segment (e.g. bar, alternatives: note, phrase, …)
- song
- set of songs (e.g. artist, …)

Audio Features in this Tutorial

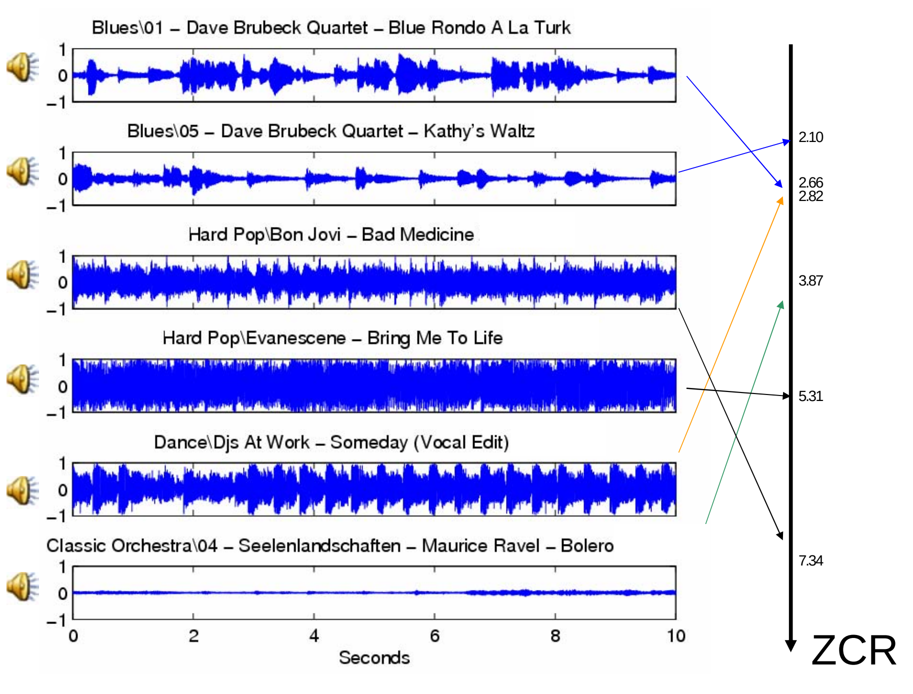
- Zero Crossing Rate (ZCR)
  - low-level audio statistic, time-domain descriptor
  - used by winner of MIREX'05 audio-based genre classification
- Timbre related
  - introduction to MFCCs
  - spectral similarity (won ISMIR'04 genre classification contest)
- Rhythm related
  - fluctuation patterns
- Harmony related
  - chroma complexity (preliminary)
  - higher level chord complexity outlook

Audio-based Music Similarity: Walkthrough

Zero Crossing Rate (ZCR) = 3/ms



= 15 / 5ms



- Similarity = **Feature Extraction** + Distance Computation
- Typical schema in feature extraction research (aka overfitting)
  1. find descriptor that works good on current set of music (e.g. 4 tracks)
  2. later on find out that there are other pieces where descriptor fails (go back to → 1)
- ZCR (and many other low-level audio statistics, incl. e.g. RMS)
  - \+ nice and simple
  - \+ interesting results (sometimes)
  - \- only weakly connected (if at all) to human perception of audio
  - \- generally not musically meaningful

→ meaningful descriptors require higher level analysis.

one typical intermediate representation is spectrogram …
(time domain → frequency domain)

Spectral Similarity (Timbre Related)

Spectrum

Mel Frequency Cepstrum Coefficients (MFCCs)

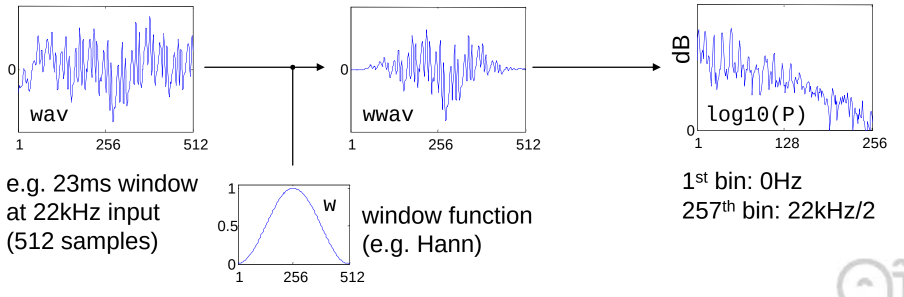
MFCCs are one of the most common representations used for Spectra in MIR

Given PCM signal

1. apply window function
2. compute power spectrum (with FFT)
3. …

```
01a w = hann(512);
01b wwav = wav.*w;
```

```
02a X = fft(wwav);
02b X = X(1:end/2+1);
02c P = abs(X).^2;
```



e.g. 23ms window at 22kHz input (512 samples)

window function (e.g. Hann)

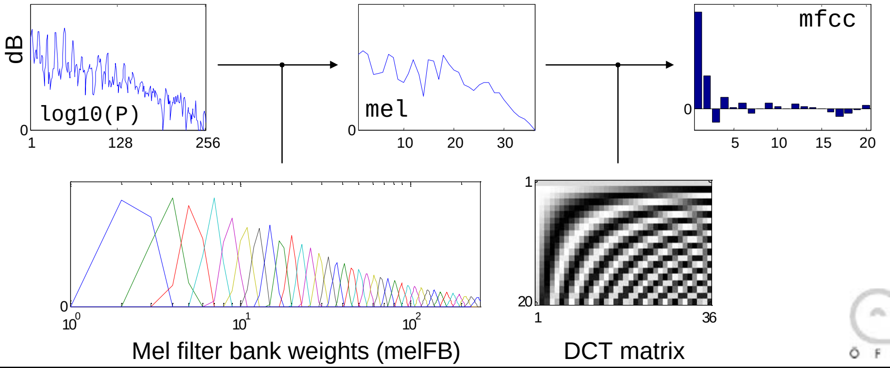
1^{st} bin: 0Hz
257^{th} bin: 22kHz/2

Mel Frequency Cepstrum Coefficients (MFCCs)

2. …
3. apply Mel filter bank
4. apply Discrete Cosine Transform (DCT) → MFCCs

```
03 mel = melFB * P;
%% size(melFB) == [36 257]
```

```
04 mfcc = DCT * log10(mel);
%% size(DCT) == [20 36]
```



Mel filter bank weights (melFB)

DCT matrix

Mel Frequency Cepstrum Coefficients (MFCCs)

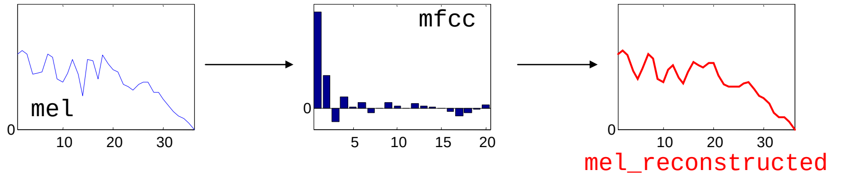
Advantages

- simple and fast (compared to other auditory models)
- well tested, many implementations available (thx2 speech processing)
- compressed representation, yet easy to handle
  (e.g. Euclidean distance can be used on MFCCs)

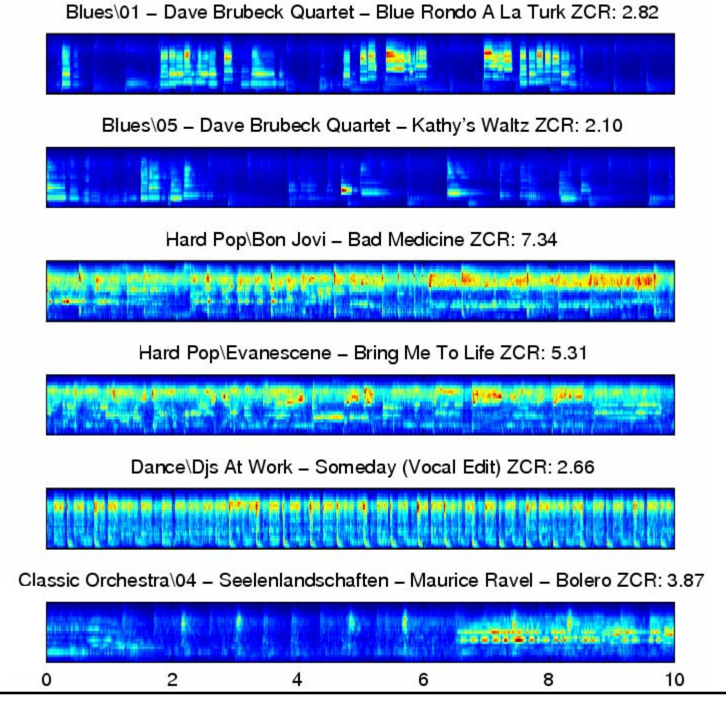
Important characteristics

- non-linear loudness (usually dB)
- non-linear filter bank (Mel scale)
- spectral smoothing (DCT; depends on number of coefficients used)
  simple approximation of psychoacoustic spectral masking effects

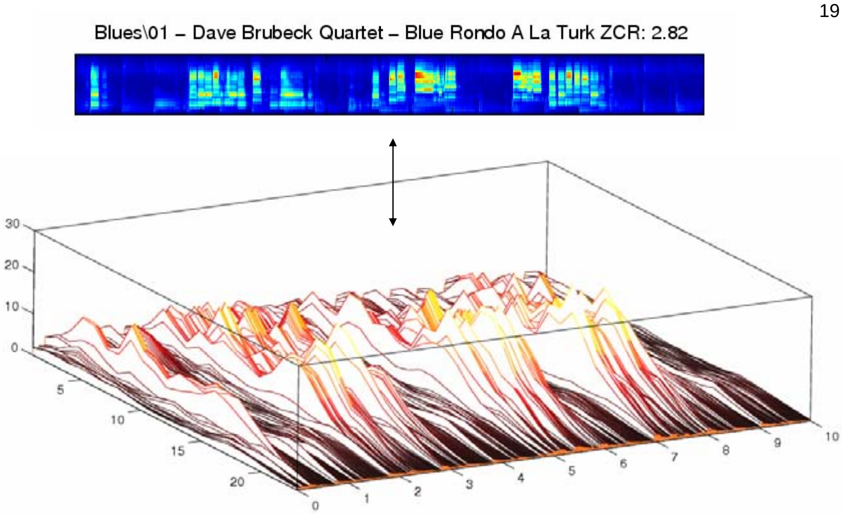
```
05 mel_reconstructed = DCT' * mfcc;
```



Spectrograms

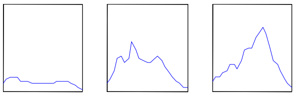


Spectrograms

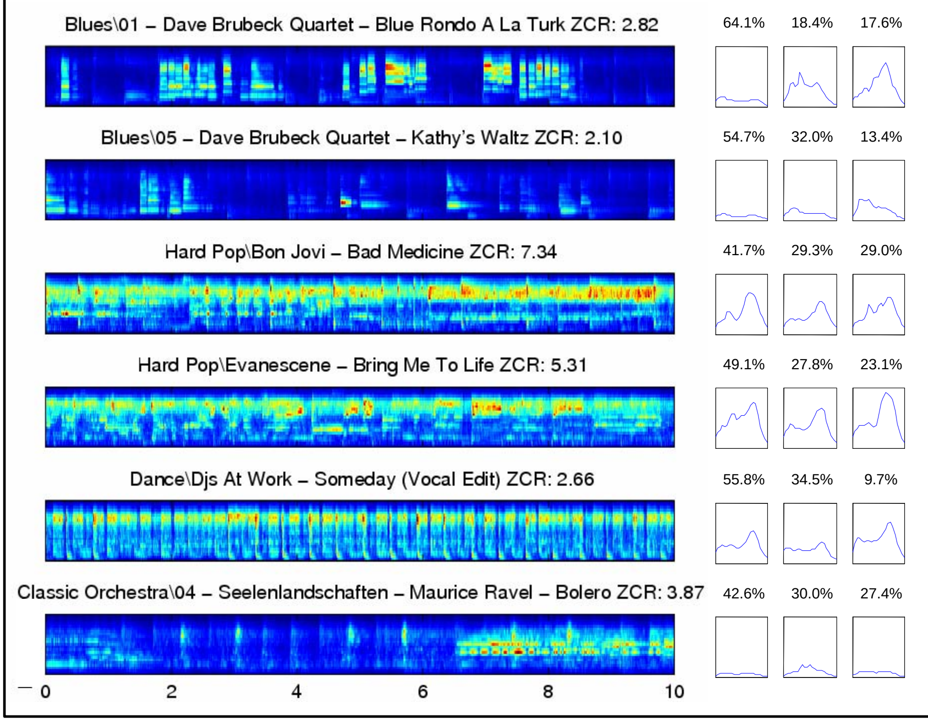


Typical Spectra

64.1% 18.4% 17.6%



Clustering
e.g. k-means or GMM-EM
(Logan & Salomon, ICME'01)

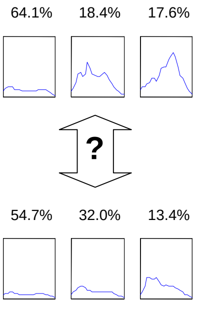


Computing Distances between Typical Spectra

21

1. **Earth Mover's Distance**
   Logan & Salomon, ICME'01
2. **Monte Carlo sampling**
   Aucouturier & Pachet, ISMIR'02

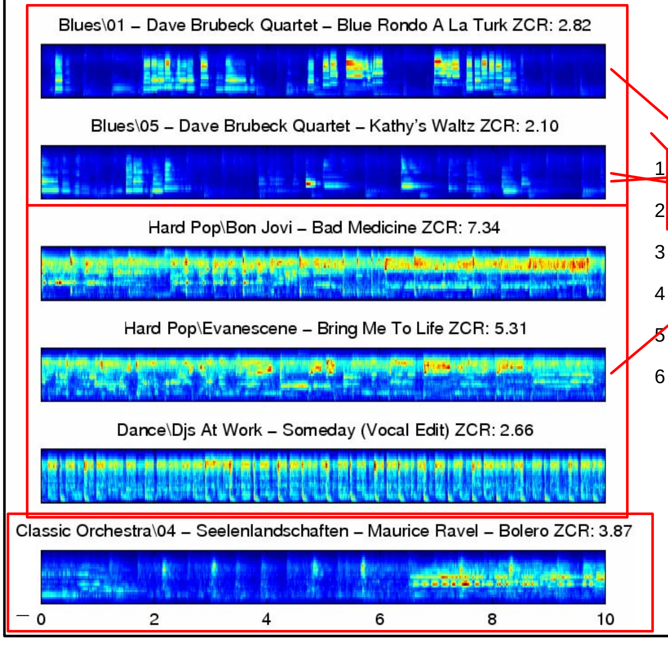
   won ISMIR'04 genre classification contest
   (30 GMM centers, 20-1 MFCCs, 2000 samples)
3. **Combination of 1+2**
   Pampalk, MIREX'05 (at ISMIR'05)
   (about factor 100 faster than last year's code)



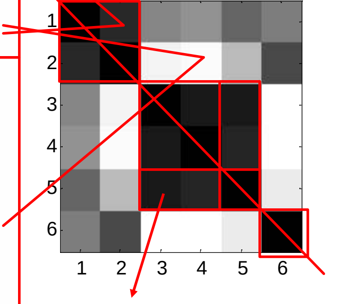
Recommended article
Aucouturier & Pachet: "Improving timbre similarity: How high is the sky?"
Journal of Negative Results in Speech and Audio Sciences, 1(1), 2004.

Spectral Similarity Distance Matrix

22

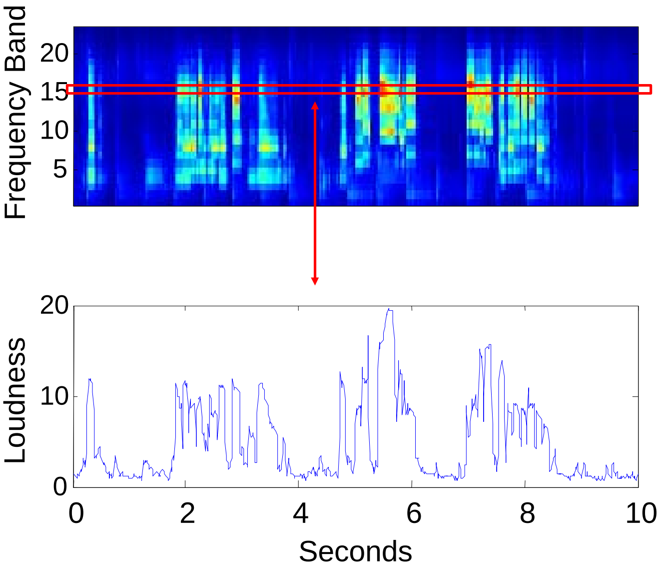


Distance Matrix

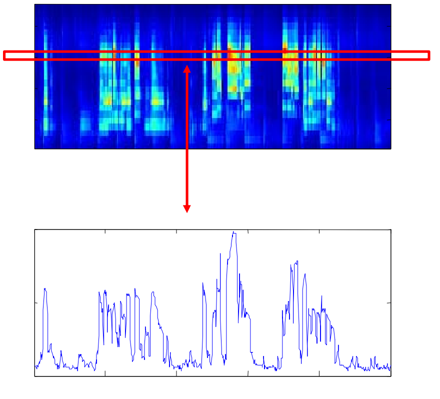


Problem:
the beats don't seem to have enough impact on the similiarity measure

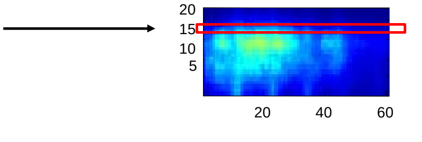
Fluctuation Patterns (Rhythm Related)



Fluctuation Patterns (Rhythm Related)



analyze peridocities
remove phase information
with e.g. FFT
(or autocorrelation,
or comb-filter)

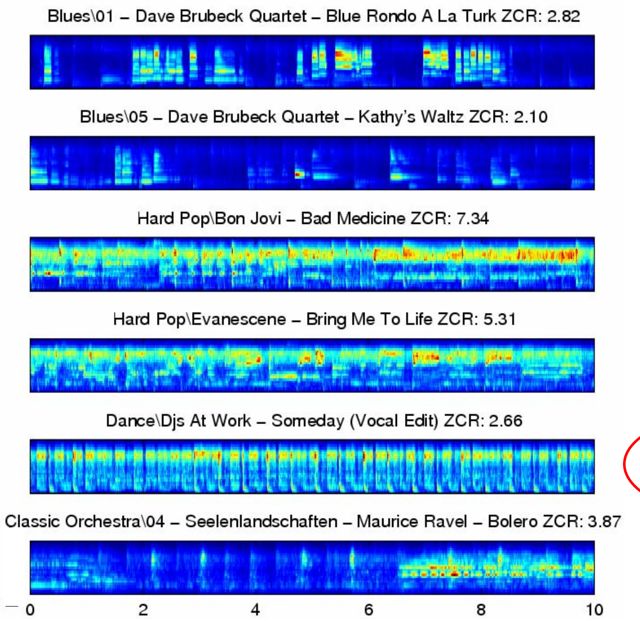


Details
e.g. Pampalk, 2001 MSc Thesis

see also e.g. ISMIR'05 presentation
Monday 12pm (Session 2): Lidy & Rauber

Fluctuation Patterns (Rhythm Related)

25

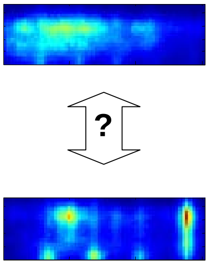


Fluctuation Patterns (Rhythm Related)

26

Distance computation

FP1



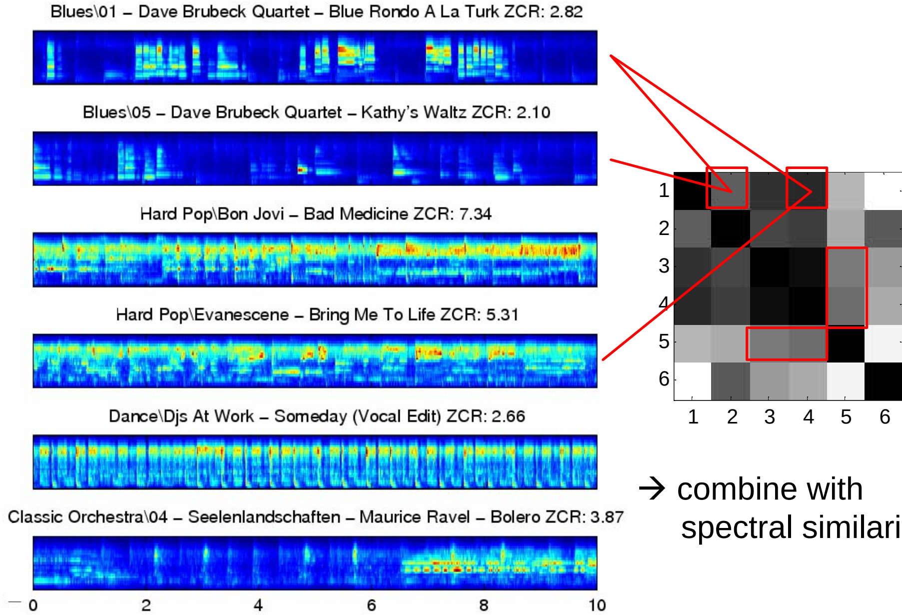
FP2

Euclidean distance (L2 norm)

```
d = sqrt(sum((FP1(:)-FP2(:)).^2));
%% e.g. size(FP1)    == [24 60]
%%      size(FP1(:)) == [1440 1]
```

simac
www.semanticaudio.org

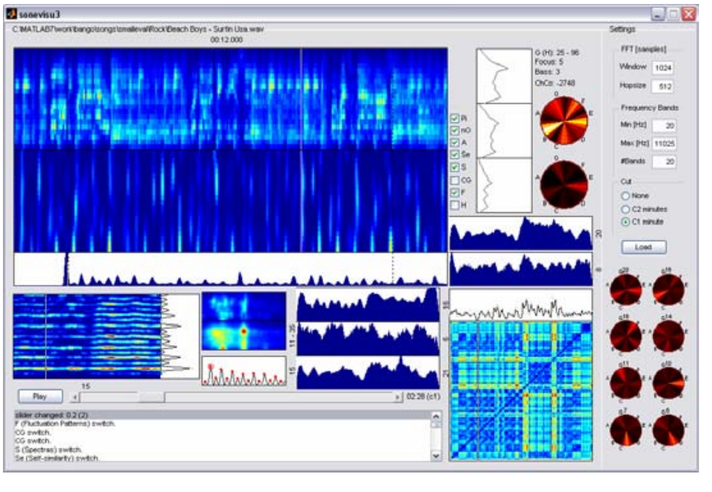
Fluctuation Patterns (Rhythm Related)



→ combine with spectral similarity

Demo: Tool used to study Feature Extraction

Searching for other features to use in combination …



Evaluation

- **User studies**
  - Logan: rate quality of playlist (application based)
  - Ellis et al., Berenzweig et al. (artist level): rate similarity of artists (given only the name), use ratings to evaluate artist similarity measures
  - Vignoli: user driven similarity (evaluate similarity in context of GUI)

- **Genre classification** (with nearest neighbor classifier!)
  assumption: similar pieces belong to the same genre

  typical genres used include
  rock, classic, jazz, blues, german hip hop, gangsta rap, pop, electronic, heavy metal, death metal, a capella, bossa nova, …

  \+ Advantages
  genre labels easy to collect, cheap, fast

  \- Disadvantages
  genre taxonomies are inconsitent,
  target function is only measured indirectly, …

Evaluation: MIREX'05

Genre classification (audio-based)

	Participant	Score	Classifier*
1	Begstra, Casagrande & Eck (1)	82	AdaBoost?
2	Mandel & Ellis	77	SVM
3	West	75	CART (+ LDA)
4	Lidy & Rauber (SSD+RH)	75	SVM
5	Pampalk	75	**Nearest Neighbor**
6	Lidy & Rauber (RP+SSD)	75	SVM
7	Lidy & Rauber (RP+SSD+RH)	75	SVM
8	Scaringella	73	Mixture of Experts (SVMs)
…	...	...	...

*) based on abstracts circulated on 22.07.2005

Genre classification results: **Magnatune (1005 + 510 tracks, 7? genres)**

	Participant	Hierarch.	Norm. Hierarch.	Raw	Norm. Raw	Time [hh:mm]	CPU Type
1	Begstra et al. (1)	**77.25**	**72.13**	**74.71**	**68.73**	06:30	B
2	Mandel & Ellis	**71.96**	69.63	**67.65**	63.99	02:25	A
3	West	**71.67**	68.33	**68.43**	63.87	12:02	B
4	Lidy & Rauber (RP+SSD)	71.08	70.90	**67.65**	**66.85**	01:46	B
5	Lidy & Rauber (RP+SSD+RH)	70.88	70.52	67.25	66.27	01:46	B
6	Lidy & Rauber (SSD+RH)	70.78	69.31	**67.65**	65.54	01:46	B
7	Scaringella	70.47	**72.30**	66.14	**67.12**	06:19	A
8	Pampalk	69.90	**70.91**	66.47	66.26	**00:55**	B
9	Ahrendt	64.61	61.40	60.98	57.15	**01:22**	B
10	Burred	59.22	61.96	54.12	55.68	03:28	B
11	Tzanetakis	58.14	53.47	55.49	50.39	**00:22**	B
12	Soares	55.29	60.73	49.41	53.54	06:38	A
…	...						

CPU Types
A: WinXP, Intel P4 3.0GHz , 3GB RAM
B: CentOS, Dual AMD Opteron 64 1.6GHz, 4GB RAM

http://www.music-ir.org/evaluation/mirex-results

Genre classification results: **uspop2002 (940 + 474 tracks, 4? genres)**

	Participant	Raw	Norm. Raw	Time* [hh:mm]	CPU Type
1	Begstra et al. (1)	**86.29**	**82.50**	06:30	B
2	Mandel & Ellis	**85.65**	76.91	02:11	A
3	Pampalk	**80.38**	**78.74**	**00:52**	B
4	Lidy & Rauber (SSD+RH)	79.75	75.45	**01:26**	B
5	West	78.90	74.67	05:09	B
6	Lidy & Rauber (RP+SSD)	78.48	77.62	**01:26**	B
7	Ahrendt	78.48	73.23	02:42	B
8	Lidy & Rauber (RP+SSD+RH)	78.27	76.84	**01:26**	B
9	Scaringella	75.74	**77.67**	06:50	A
10	Soares	66.67	67.28	03:59	A
11	Tzanetakis	63.29	50.19	**00:22**	B
12	Burred	47.68	49.89	02:34	B
13	Chen & Gao	22.93	17.96	N/A	A
…	...				

***) CPU Types**
A: WinXP, Intel P4 3.0GHz , 3GB RAM
B: CentOS, Dual AMD Opteron 64 1.6GHz, 4GB RAM
http://www.music-ir.org/evaluation/mirex-results

Evaluation: MIREX'05

Artist identification (audio-based)

	Participant	Score	Classifier*
1	Mandel & Ellis	72	SVM
2	Pampalk	61	**Nearest Neighbor**
3	West & Lamere	47	Bagging, LDA
4	Tzanetakis	42	SVM?
5	Logan (ICME'01)	26	**Nearest Neighbor**
…	...	...	...

*) based on abstracts circulated on 22.07.2005 and 24.08.2005 (West & Lamere)

MIREX'05 Genre Classification: Critical Remarks

1. **Overfitting?**
   uspop2002 and Magnatune collections were used by
   some participants to optimize their algorithms (e.g. I used Magnatune)
   → necessary to be careful when generalizing

2. **Genre classification = artist identification?**
   no artist filter used!
   → pieces from same artist in test and trainingset
   e.g. training: classifier is given 5 Eminem songs
   and told that all Eminem songs belong to genre "rap"
   testing: classifer is given another song by Eminem
   and asked to classify it … (not knowing it is by Eminem,
   but if it can recognize that it is from the same artist it will
   score 100% on genre classification)
   → application scenario for genre classification results
   not as clear as it might appear
   (true performance is lower, at least for the my own results)

[MUSIC-IR] Mailing list: thread on genre classification

G. Tzanetakis (2 Sept): “To me genre classification has always been an easy way to compare audio content features.”

M. Sandler (3 Sept): “most of my favourite music does not fit comfortably into any single "genre". what does that tell us? something about me and/or something about the whole concept of genre?”

J. Pickens (4 Sept): “The presumption here, I think, is that if a user likes a particular song or set of songs, and is looking for "similar" songs (whether to buy or to add as the next items in their playlist or whatever), songs from the same "genre" will meet that information need. Am I correct that this, roughly, is the justification for working on the genre problem? […]
In other words, suppose we *could* do "genre“ classification with 100% accuracy. *What*, then, would we do with that information?”

D. Eck (5 Sept): “Genre prediction by itself is not a good end goal. We should be careful not to turn it into one. We don't want genre to become the next Query By Humming. […] Thus in my mind two interesting goals are (a) collaborative and/or content-based filtering and (b) automatic playlist generation by example are interesting goals.”

Limitations: Audio-Based Similarity

100% accuracy is not possible because …

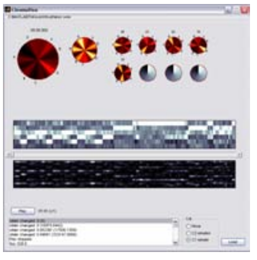
- Genre taxonomies are inconsistent
  even human experts do not agree 100%
- Some important aspects are not in the audio signal or difficult (if not impossible) to extract: sociocultural background, lyrics, mood
  (→ web-based similarity)
- Extracted features are too low-level (i.e. not meaningful enough)
  → higher level analysis (future work)
  e.g. rhythm, harmony, etc.

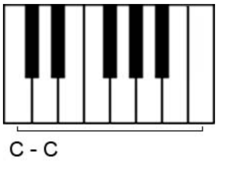
(Do we need a perfect similarity measure for applications?)

Chroma und Harmony

- Chroma: “color” of a note
- Chords: combination of several notes
- Harmony: combination (sequence) of chords

Demo:

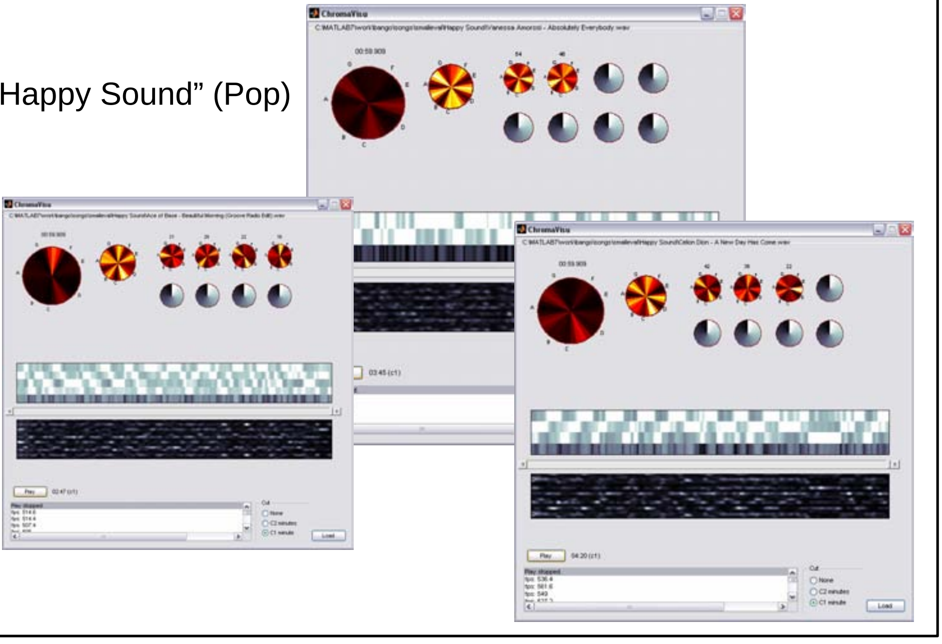




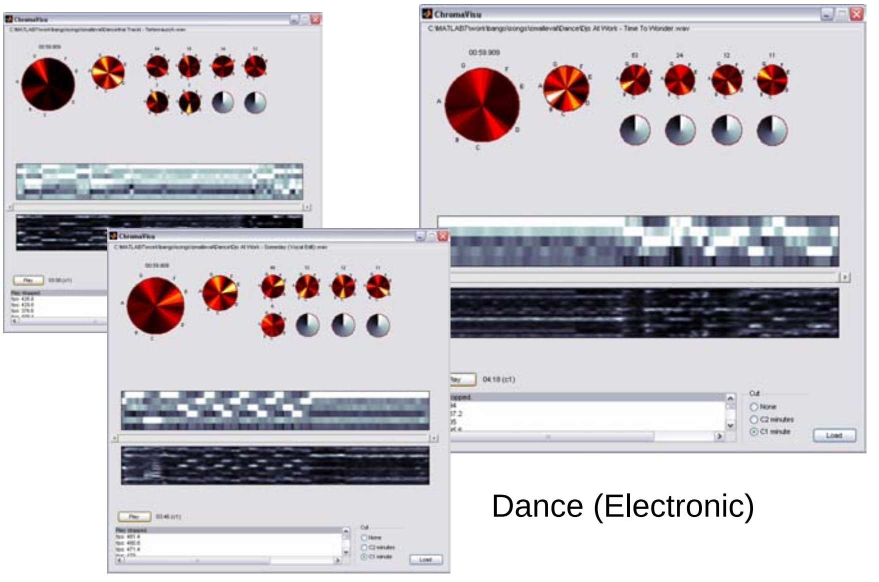
Recently a lot of work on chromagrams and higher level respresentations has been going on, e.g. ISMIR’05: Bello & Pickens:
“A robust mid-level representation for harmonic content in music signals”

Chroma Complexity

“Happy Sound” (Pop)



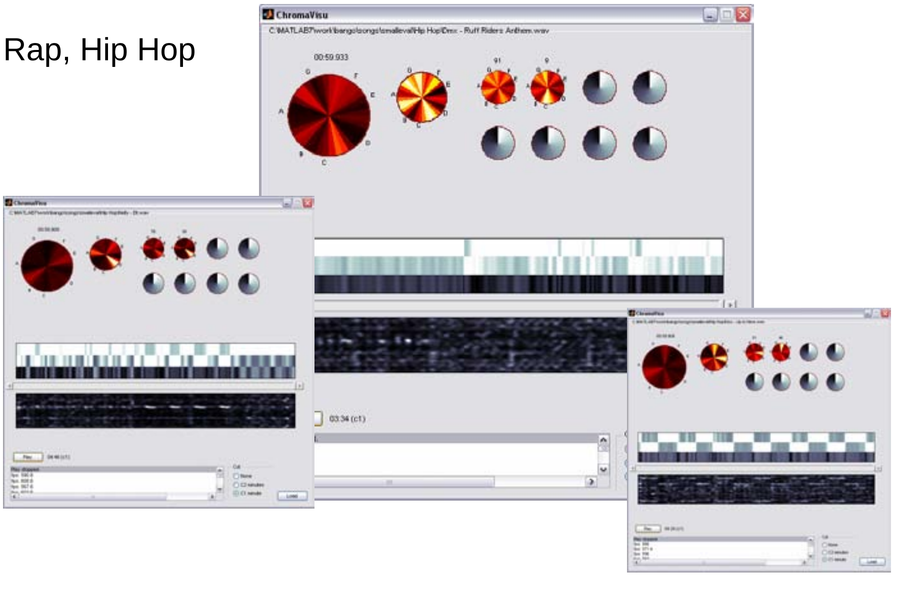
Chroma Complexity



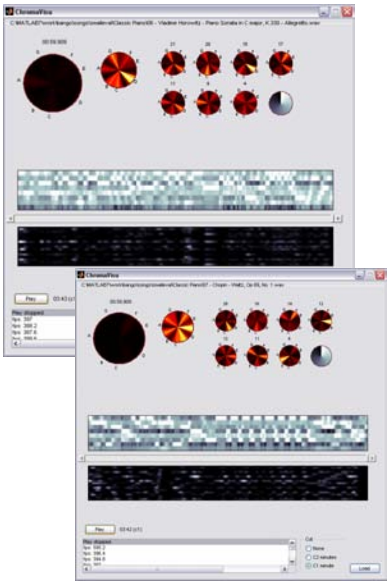
Dance (Electronic)

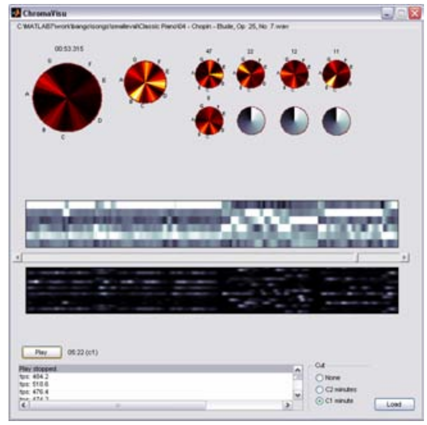
Chroma Complexity

Rap, Hip Hop



Chroma Complexity

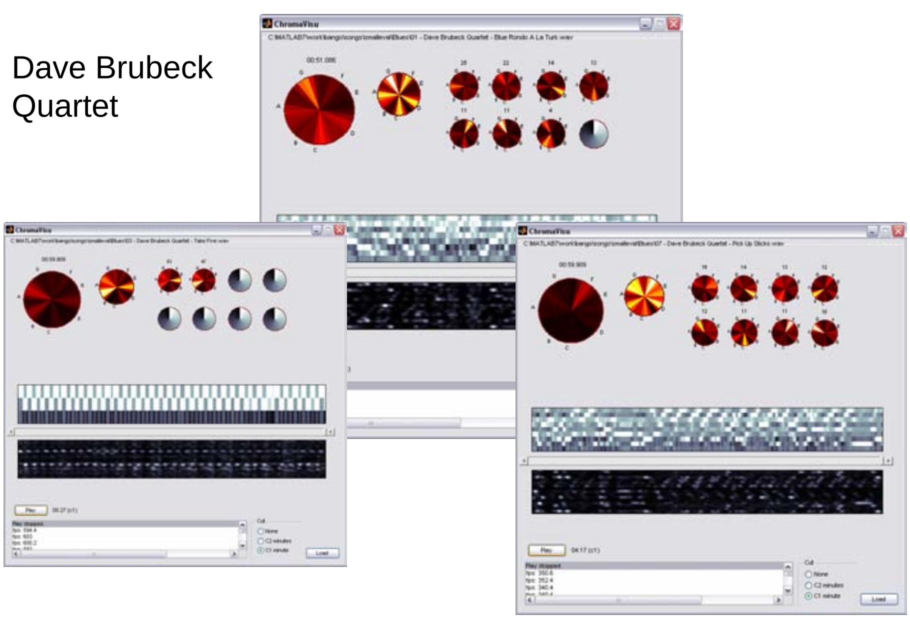




Classical Piano (Chopin, Mozart)

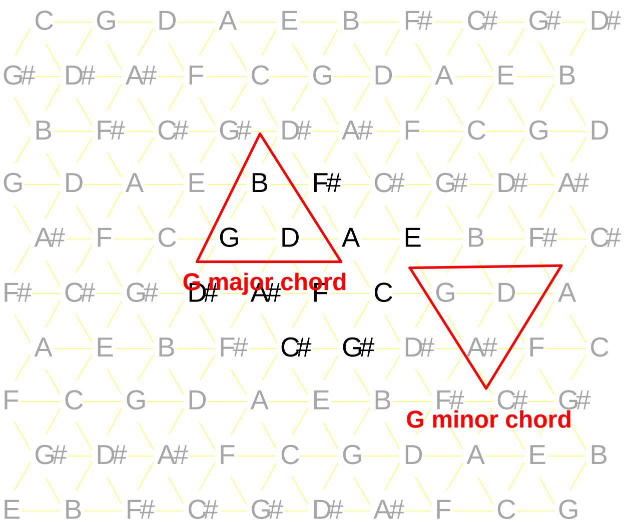
Chroma Complexity

Dave Brubeck Quartet



Chroma und Harmony

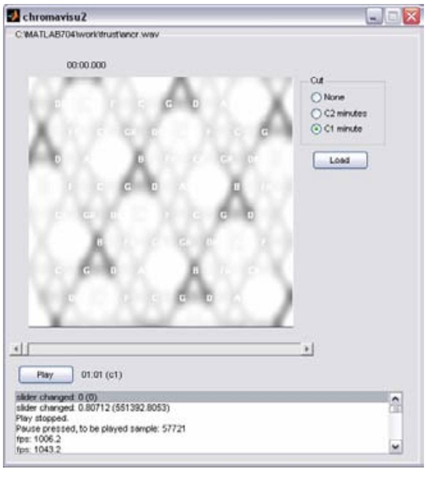
Tonnetz



special thanks to Chris Harte

Demo: Chroma und Harmony

Towards higher level harmonic complexity analysis…



In collaboration with Chris Harte and Juan Bello from QMUL

Web-Based Similarity

- Complements audio-based approaches
  e.g. sociocultural information

- Various approaches exist
  e.g. Google-based (Whitman & Lawrence, ICMC'02)

- Idea
  use artist name to find related web-pages (e.g. fan sites or reviews)
  compare words occurring on web-pages to estimate similarity

- Problems
  unknown artists (e.g. creative commons, Magnatune)
  names not unique (e.g. "War", "Slayer", "Saints")

Web-Based Similarity

Idea (Withman & Lawrence, ICMC'02)

"Robbie Williams" +music +review → Google
→ 50 top ranked web-pages
→ word occurrences (TFxIDF)
(remove stop words: and, or, is, that, etc. and typos)

TFxIDF = Term Frequency * Inverse Document Frequency

high term frequency (TF) e.g.:
music, review, sing, song, album, pop, …

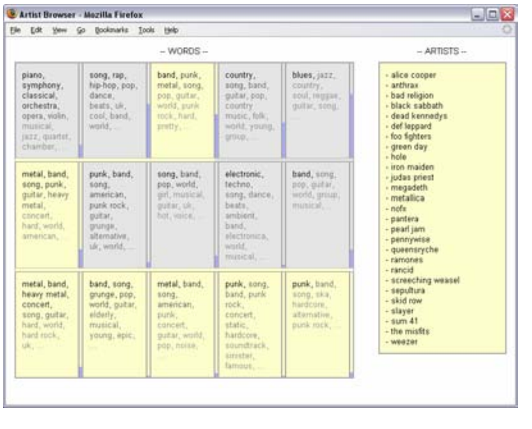
high document frequency (DF) e.g.:
music, review, album, …

Demo: Web-Based Artist Similarity

- Idea
  - hierarchical organization of artists
  - automatically find clusters (don’t use predefined genres)
- Problem?
  - how to describe clusters?
    (assume user does not know the artist names)
- Solution!
  - label clusters with words found on web pages
    (using a “music” dictionary)

Details see: Pampalk, Flexer, Widmer (ECDL’05)
“Hierarchical organization and description of music collections on the artist level”

Demo: Web-Based Artist Similarity



Data: 224 artist names from 14 genres

http://www.ofai.at/~elias.pampalk/wa/wordart-lk

Combining Audio and Web-Based Approaches
49

Examples

- Whitman & Smaragdis: ISMIR'02 "Combining musical and cultural features for intelligent style detection"
- Whitman & Ellis: ISMIR'04 "Automatic record reviews"
- Baumann, Pohle, Shankar: ISMIR'04 "Towards a socio-cultural compatibility of MIR systems"

Problem: need a lot of data for reliable results.
e.g. 1000 artists, 2 albums per artist, 15 tracks per album → 30'000 tracks

Related Sessions at ISMIR'05
50

- **Genre classification** [Mon #2]
  the same features can directly be used for similarity computations
- **MIR systems** [Tue #1]
  often based on some concept of "similarity"
- **Melody** [Tue #2], **Harmony** [Wed #2], and **Rhythm** [Thu #1]
  meaningful features
  often, however the primary goal is not a similarity measure
- **Optimized and efficient methods** [Tue #3b]
  necessary when dealing with huge collections
  "speed up nearest neighbor search"!
- **MIREX**! [Wed]
- **Music similarity** [Wed #1]
  "user driven similarity"!
- **Voice/Instrument analysis** [Wed #3]
  → timbre similarity
- various **posters** (and demos) …

Tutorial Goals

51

1. What is music similarity? (Definition?)
2. What is it good for? (Applications?)
3. How (and from what) can similarity be computed?
4. How to evaluate the algorithms?
5. What are the limitations?
6. What are future directions?
7. What is happening at ISMIR'05?

52

simac
www.semanticaudio.org

EU research project
Semantic Interaction with Music Audio Contents
QMUL, IUA/MTG, OFAI, Philips Research, MD
http://www.semanticaudio.org

ÖFAI

Austrian Research Institute for Artificial Intelligence (OFAI)
Intelligent Music Processing and Machine Learning Group
http://www.ofai.at/music