SIGIR 2013 Half-Day Tutorial

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Music Similarity and Retrieval

Peter Knees
Markus Schedl

{peter.knees, markus.schedl}@jku.at

Department of Computational Perception
Johannes Kepler University (JKU)
Linz, Austria

July 28th, Dublin, Ireland

http://www.cp.jku.at/tutorials/sigir2013.html

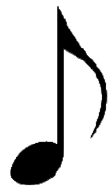
Overview

Goals of this tutorial

– Introduction to the field of music similarity estimation
– Approaches to music retrieval

Parts:

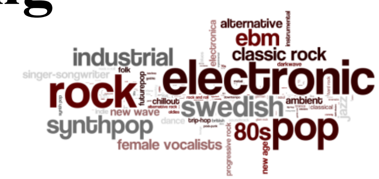
I. About Music Similarity



II. Content-Based Similarity and Retrieval

III. Music Context-Based Similarity and Indexing

IV. Personalization and User Adaptation



Department of Computational Perception

Who we are

Markus Schedl

Assistant Professor of the ***Department of Computational Perception*, JKU Linz**
M.Sc. in Computer Science from Vienna University of Technology
Ph.D. in Computational Perception from Johannes Kepler University Linz
M.Sc. in Int'l Business Administration from Vienna University of Economics and Business Administration

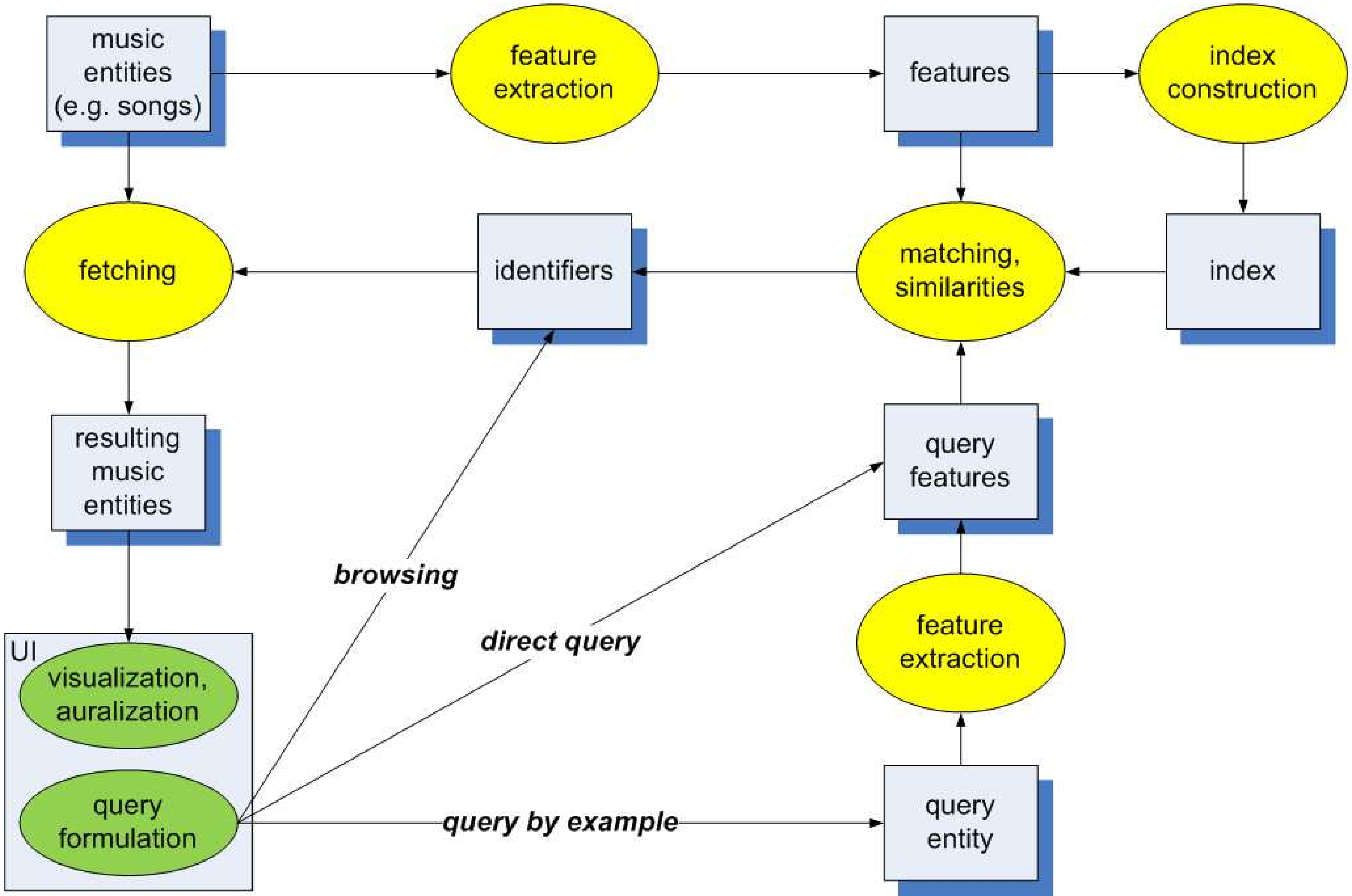
Research interests: social media mining, music and multimedia information retrieval, recommender systems, information visualization, and intelligent/personalized user interfaces

Peter Knees

Assistant Professor of the ***Department of Computational Perception*, JKU Linz**
M.Sc. in Computer Science from Vienna University of Technology
Ph.D. in Computer Science from Johannes Kepler University Linz

Research interests: music and web information retrieval, multimedia, user interfaces, recommender systems, digital media arts

What is MIR? An Information Retrieval view



Applications: Automatic Playlist Generation

"Personalized Radio Stations"

e.g.

- Pandora
- Last.fm
- Spotify Radio
- iTunes Radio

Continuously plays similar music

Based on content or collaborative filtering data

Optionally, songs can be rated for improved personalization

Pandora.com

Department of Computational Perception

Applications: Browsing Music Collections

Intelligent organization for "one-touch access"

- music collections become larger and larger (on PCs as well as on mobile players)
- most UIs of music players still only allow organization and searching by textual properties accoding to scheme (*genre-)artist-album-track*

→ novel and innovative strategies to access music are sought in MIR

„intelligent iPod" by CP@JKU
[Schnitzer et al., MUM 2007]

Applications: Audio Identification

Query-by-example/audio fingerprinting:
excerpt of a song (potentially recorded in low quality) used to identify the piece

Query-by-humming:
input is not excerpt of a song, but melody hummed by the user

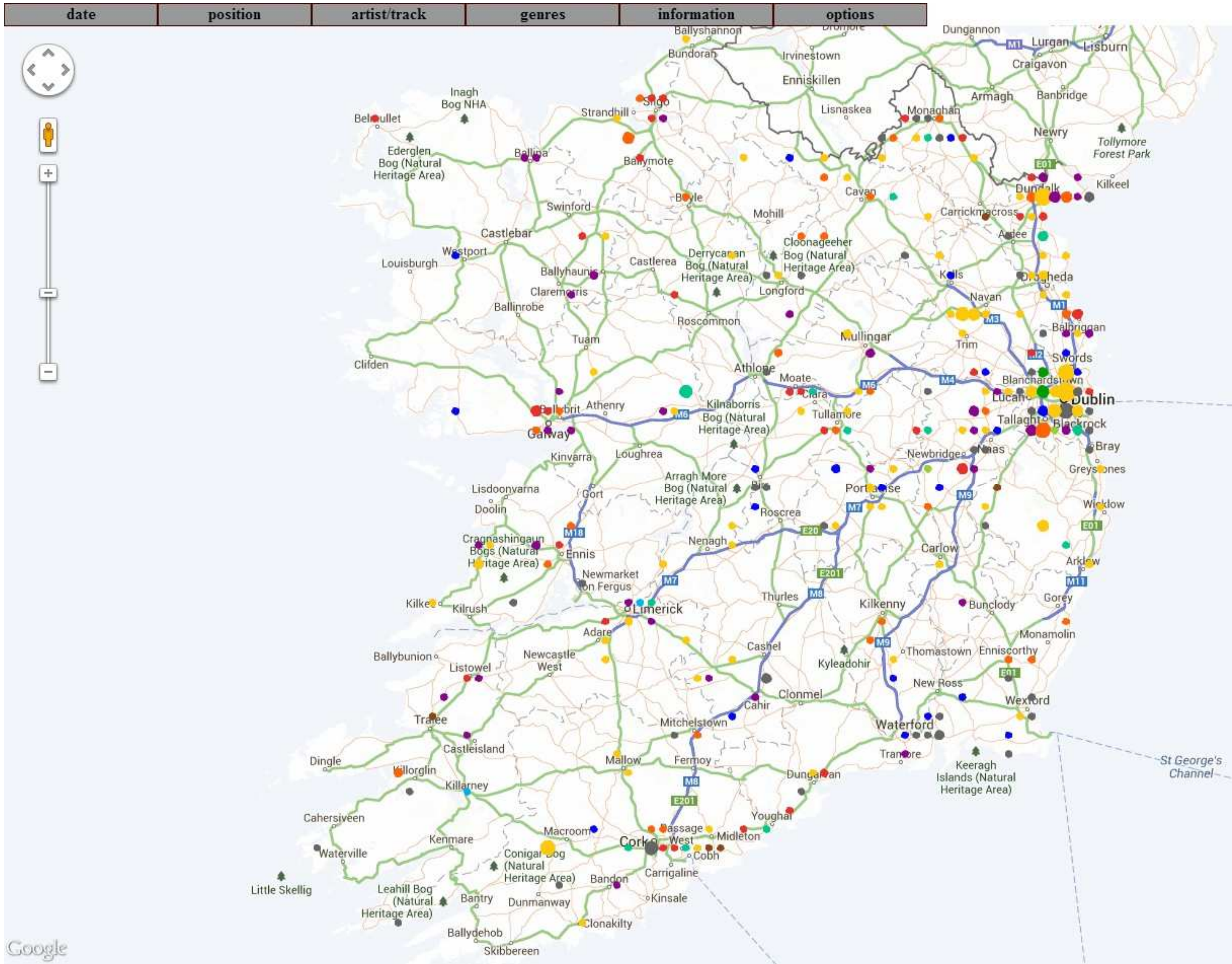
Examples:

www.shazam.com
www.soundhound.com
www.musicline.de/de/melodiesuche

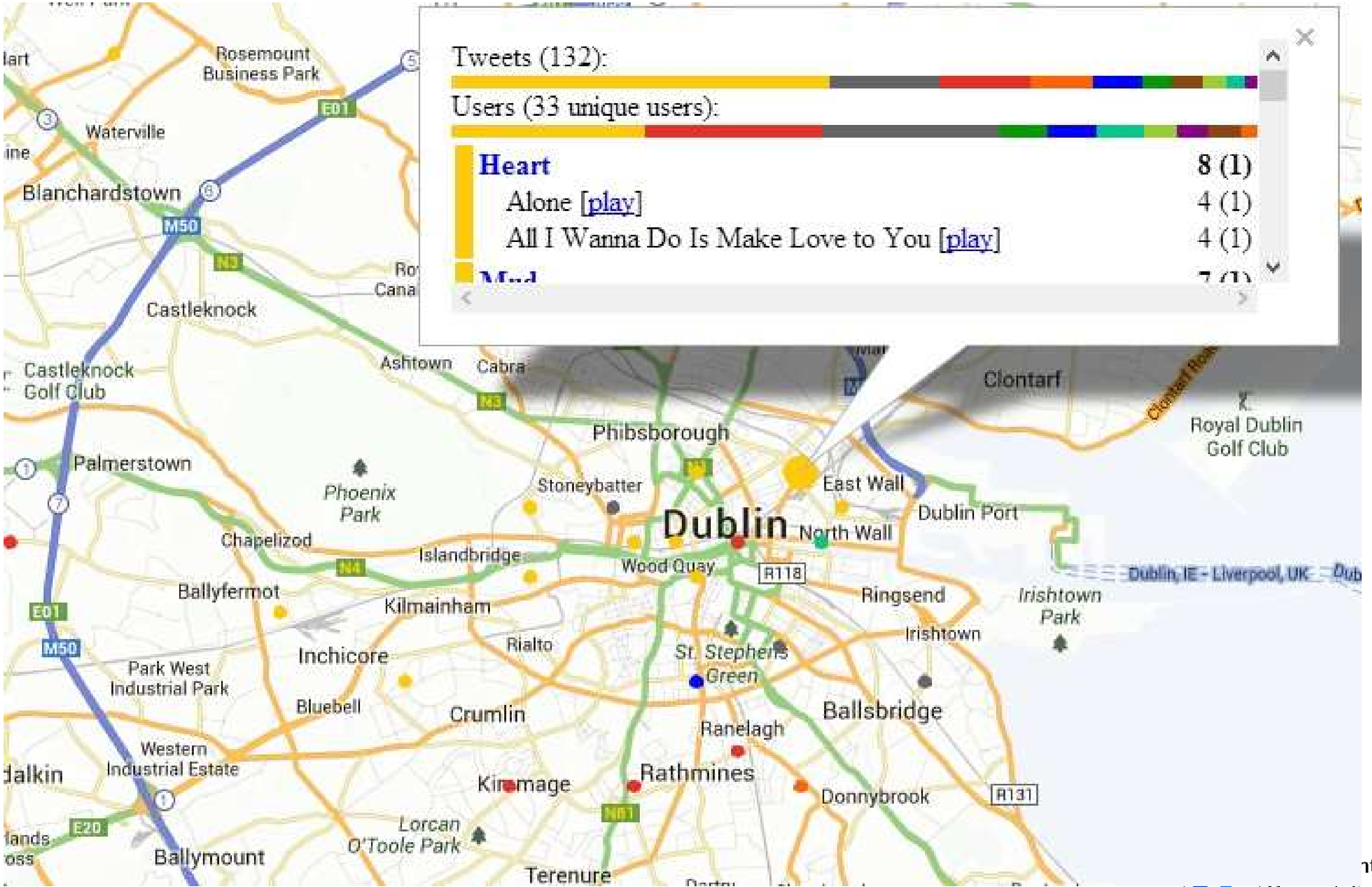
SoundHound

Melodiesuche
Query by Humming

Applications: Music Tweet Map



Applications: Music Tweet Map



Part I

ABOUT MUSIC SIMILARITY

Department of Computational Perception

Music Retrieval and Similarity

To retrieve music (query-by-example), we need to calculate how similar two music pieces are

What does similar mean?

- Sounding similar
- What does sounding similar mean?
  Genre (what is genre?), instruments, mood, melody, tempo, rhythm, singer/voice, … all of them? a combination?
- Any of that can contribute to two songs being perceived as similar, but describing sound alone falls short of grasping that phenomenon

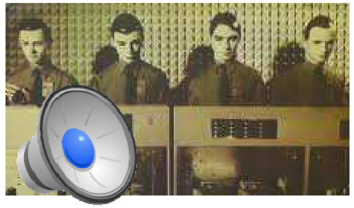
Music similarity is a multi-faceted task

Department of Computational Perception

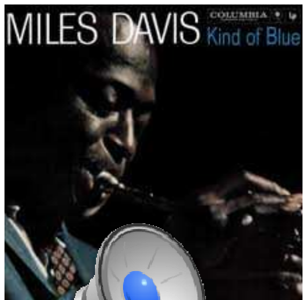
Music Similarity Examples

Which are similar?

Which go together?



Which are more similar?



Department of Computational Perception

The term “music similarity” is ill-defined

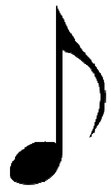
- Experiments show that humans only agree to about 80% when asked to assign music pieces to genres (Lippens et al.; 2004) (Seyerlehner et al.; 2010)
- Music similarity is highly subjective
- Contextual factors are also important (but not in the signal!)
  - artist/band context, band members, city/country, time/era, *lyrics*, *language*, genre, …
  - political views of artists, marketing strategies, …
  - also listening context, mood, peers (= user context)
- Optimally, similarity is calculated taking into account all influencing factors:

  audio content, music context, user context (difficult!), user properties (also difficult!)

Department of Computational Perception

Computational Factors Influencing Music Perception and Similarity

(Schedl et al.; JIIS 2013)



music content

Examples:
- rhythm
- timbre
- melody
- harmony
- loudness

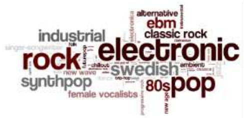
Examples:
- mood
- activities
- social context
- spatio-temporal context
- physiological aspects

user context

music perception and similarity

Examples:
- semantic labels
- song lyrics
- album cover artwork
- artist's background
- music video clips

music context



Examples:
- music preferences
- musical training
- musical experience
- demographics

user properties

Department of Computational Perception

Implications for Evaluation

If similarity is such a subjective concept, how can we evaluate algorithms that claim to find similar pieces?

What is the Ground Truth?

- Class labels (genres)? Often used, often criticized
- Multi-class labels (tags)?

How to obtain (ranked) relevance?

Best strategies so far:

- Use listening data as retrieval ground truth (playlists)
- Ask users directly about similarity (listening tests)

Evaluation Campaign: MIREX

Music Information Retrieval Evaluation eXchange

- Annual MIR benchmarking effort
- Organized by UIUC since 2005 (Prof. J.S. Downie + team)

~ 20 tasks in 2013

- Melody extraction, onset/key/tempo detection
- Score following
- Cover song detection
- Query-by-singing/humming/tapping
- etc.

Audio/signal-based tasks only so far

Department of Computational Perception

MIREX Audio Music Similarity and Retrieval Task

Evaluates query-by-example algorithms

Results evaluated by humans

> "*Evaluator question*: Given a search based on track A, the following set of results was returned by all systems. Please place each returned track into one of three classes (not similar, somewhat similar, very similar) and provide an indication on a continuous scale of 0 - 100 of how similar the track is to the query."

Each year: ~100 randomly selected queries, 5 results per query per algorithm (joined), "1 set of ears" per query

Friedman's test to compare algorithms

No "winners," but algorithm ranking

Department of Computational Perception

Other Evaluation Campaigns

MusiClef (e.g. @ MediaEval 2012)

Task: multi-modal tagging of songs

Data: audio, web, tag features, expert labels; 1355 songs

Evaluation measures: precision, recall, F1-measure

Million Song Dataset Challenge (McFee et al.; 2012)

Task: predicting songs a user will listen to

Data: user listening history playcounts (48M)

Evaluation: recall on ranking, MAP

KDD Cup 2011 (Dror et al.; 2012)

Task: predicting song ratings

Data: Yahoo! Music data set (260M ratings)

Evaluation: RMSE

Part II

CONTENT-BASED SIMILARITY

Department of Computational Perception

Categorization of Content-Based Features

Domain:

- **Time domain**
  consider signal in time/amplitude representation ("waveform")

- **Frequency domain**
  consider signal in frequency/magnitude representation

Transformation from time to frequency domain using, e.g., Fast Fourier Transform (FFT)

Department of Computational Perception

Categorization of Content-Based Features

Level of abstraction:

- **Low-level**
  properties of audio signal (e.g., energy, zero-crossing-rate)
- **Mid-level**
  aggregation of low-level descriptors,
  applies psycho-acoustic models (cf. MFCC, FP);
  typically the level used when estimating similarity
- **High-level**
  musically meaningful to listener, e.g., melody, themes, motifs;
  "semantic" categories, e.g., genre, time period, mood, …
  (cf. semantic tags learned from audio features)

Department of Computational Perception

How to Describe Audio Content?

- Possible idea: get features that describe music the way humans do and compute similar songs based on this information
- Unfortunately we are are not able to extract most of these features reliably (or at all…)
  - even “simple” human concepts are difficult to model (**“semantic gap”**)
  - even tempo estimation is very hard…
  - NB: a human annotation approach is done in the Music Genome Project (cf. Pandora’s automatic radio station service)
- Furthermore some of these features are quite subjective (e.g., mood)
- Need to find computable descriptors that capture these dimensions somehow (…and work acceptably)

Descriptors of Content

Acoustic property to describe:

- **Loudness:** perceived strength of sound; *e.g., energy*
- **Pitch:** frequency, psychoacoustic ordering of tones (on scale; from low to high); *e.g., chroma-features*
- **Timbre:** "tone color", what distinguishes two sounds with same pitch and loudness; *e.g., MFCCs*
- **Chords and harmony:** simultaneous pitches
- **Rhythm:** pattern in time; *e.g., FPs*
- **Melody:** sequence of tones; combination of pitch and rhythm

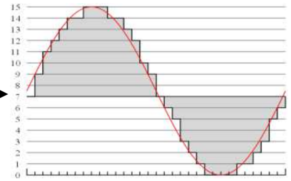
cf. (Casey et al.; 2008)

Department of Computational Perception

Scheme of Content-Based Feature Extraction

analog signal

sampling
quantization

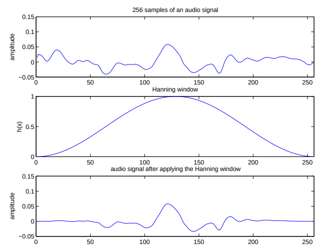


Pulse Code Modulation (PCM)

framing

frame 1: e.g. sample 1...256
frame 2: e.g. sample 129...384
frame 3: e.g. sample 257...512
...
frame n

windowing



FFT

time domain feature calculation

frequency domain feature calculation

aggregation, model building (mean, median, sum, GMM, HMM)

feature value, vector, or matrix

Department of Computational Perception

Low-Level Feature: RMS Energy

Root-Mean-Square Energy (aka RMS power, RMS level, RMS amplitude)

Scope: time domain

Calculation:

$$RMS_t = \sqrt{\frac{1}{K} \cdot \sum_{k=t\cdot K}^{(t+1)\cdot K-1} s(k)^2}$$

s(k)...amplitude of k^{th} sample in time domain
K...frame size (number of samples in each frame)

Remarks:

+ beat-related feature, can be used for beat detection

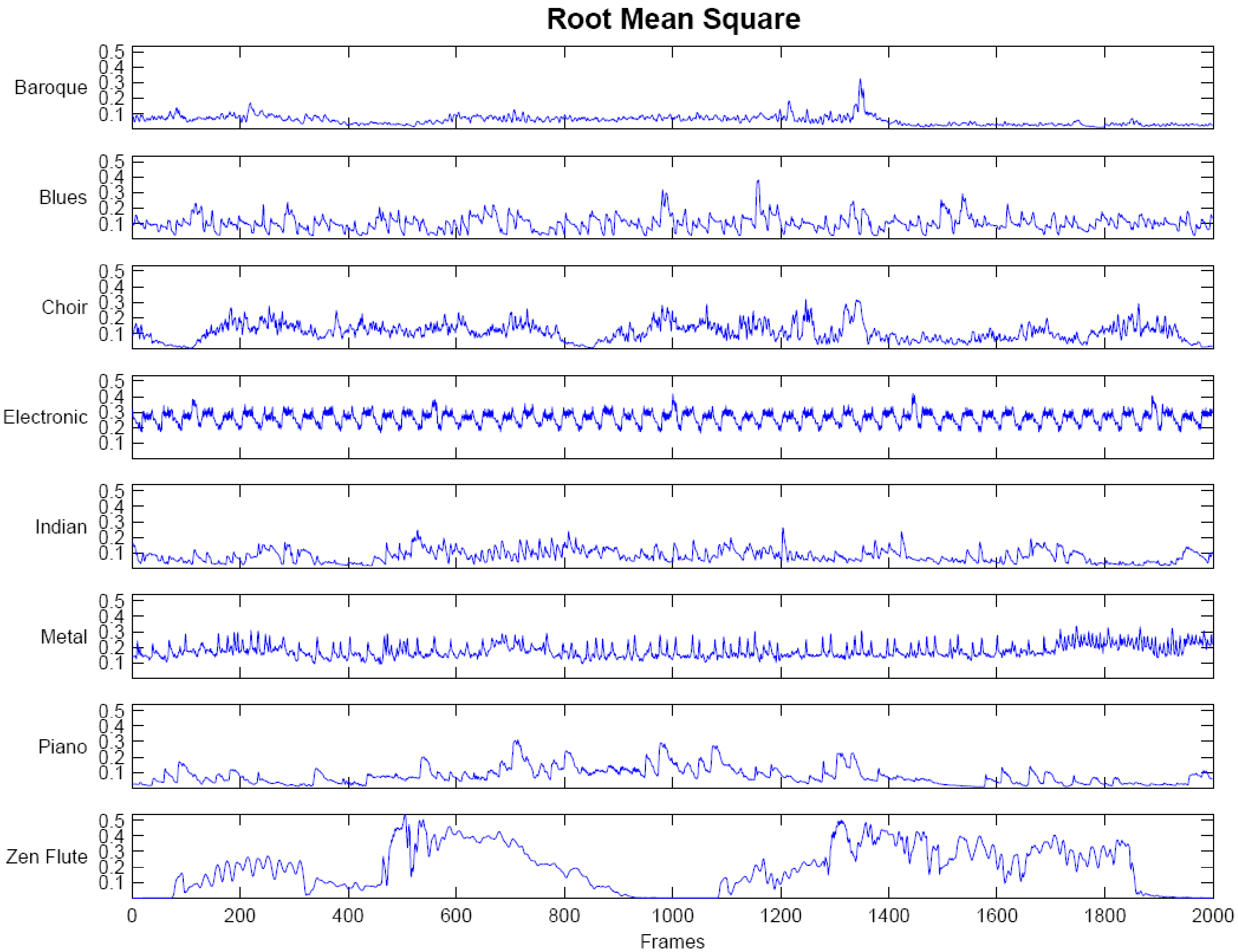
+ related to perceived intensity

+ good loudness estimation

– discriminative power not clear

Department of Computational Perception

RMS Energy: Illustration

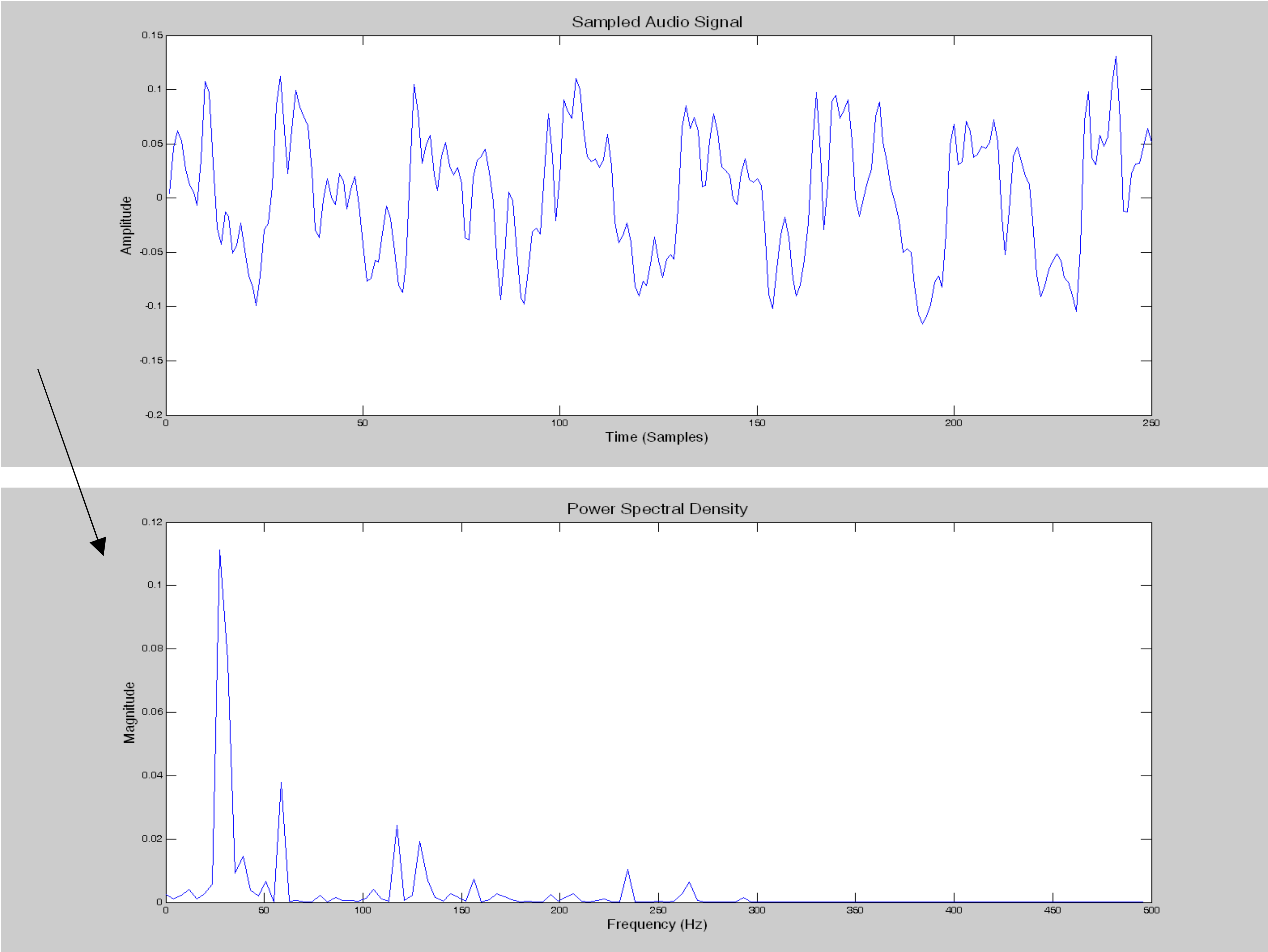


Fourier Transform

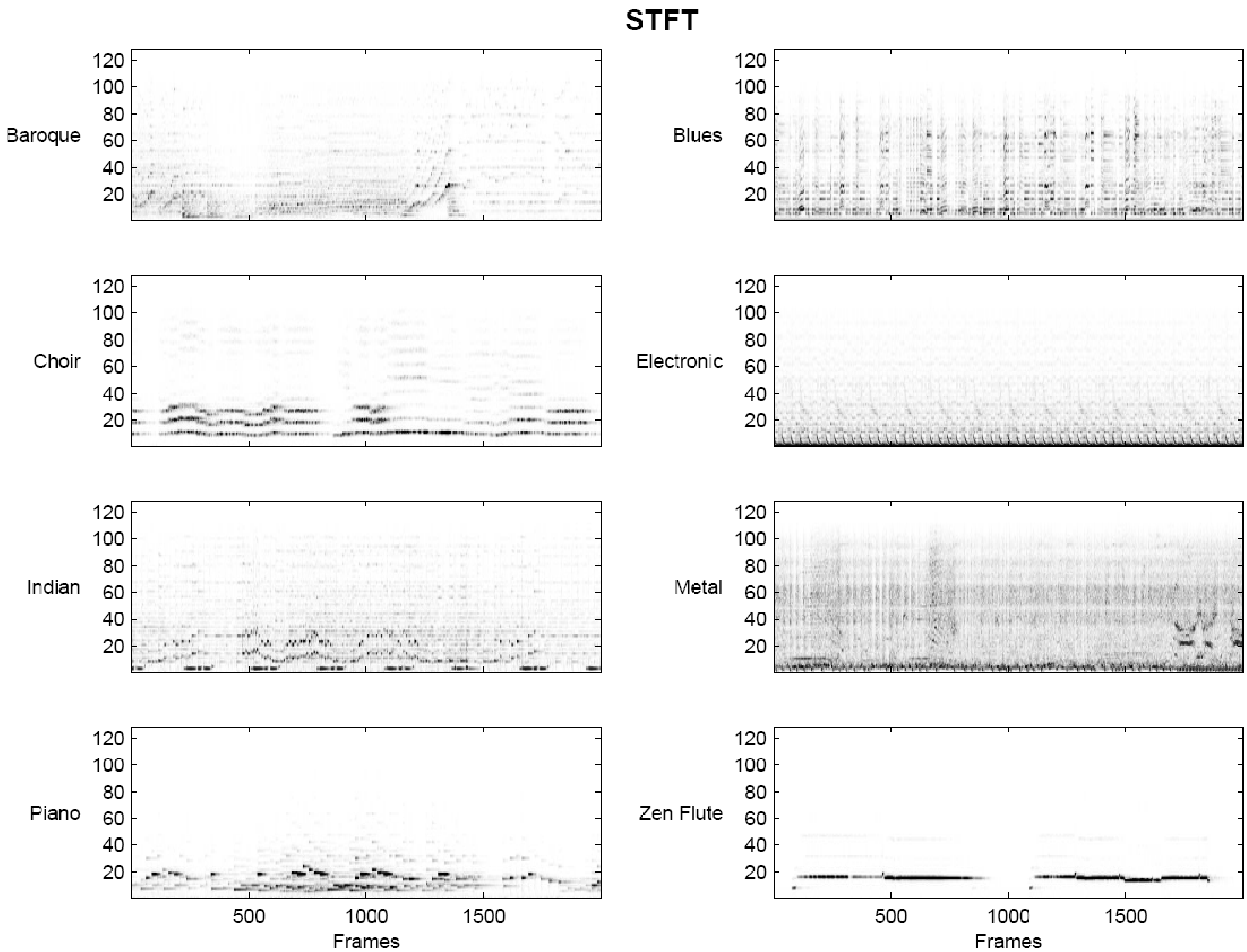
Transformation of the signal
from **time domain** (time vs. amplitude)
to **frequency domain** (frequency vs. magnitude)

Jean Baptiste Joseph Fourier

- Theorem: any continuous periodic function with a period of 2π can be represented as the sum of sine and/or cosine waves (of different frequencies)
- Implication: any audio signal can be decomposed into an infinite number of overlapping waves when periodic
- Periodicity is achieved by multiplying the PCM magnitude values of each frame with a suited function, e.g., a Hanning window (**windowing**)
- In our case: **Discrete Fourier Transform (DFT)**
- In practice efficiently calculated via **Fast Fourier Transform (FFT)** (Cooley, Tukey; 1965)


Sampled Audio Signal
Amplitude
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
0
50
100
150
200
250
Time (Samples)
Power Spectral Density
Magnitude
0.12
0.1
0.08
0.06
0.04
0.02
0
0
50
100
150
200
250
300
350
400
450
500
Frequency (Hz)

Representation as STFT



Low-Level Feature: Spectral Centroid

Scope: frequency domain

Calculation:

$$C_t = \frac{\sum_{n=1}^{N} M_t(n) \cdot n}{\sum_{n=1}^{N} M_t(n)}$$

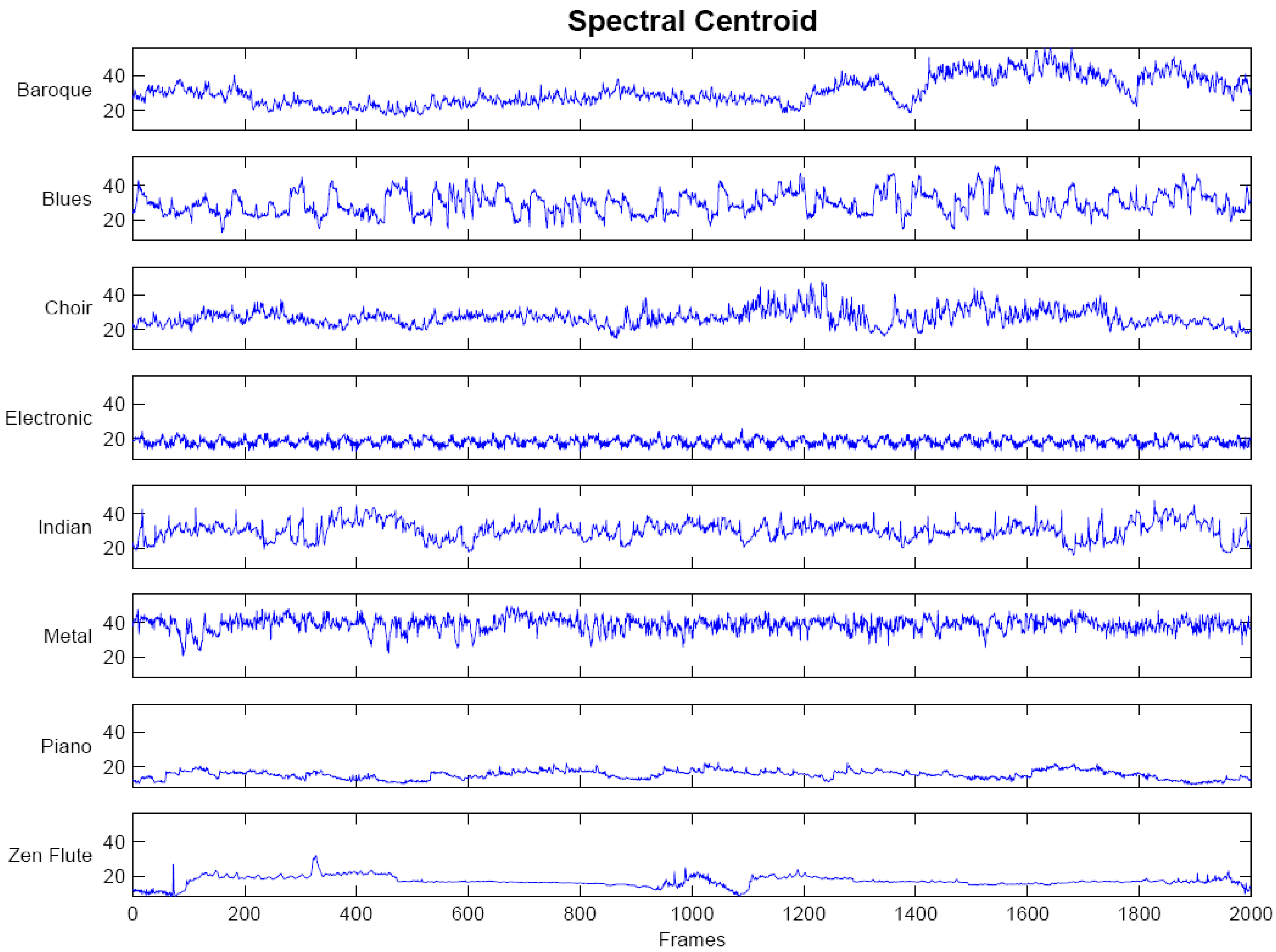
$M_t(n)$...magnitude in frequency domain at frame t and frequency bin n
N...number of highest frequency band

Description: center of gravity of the magnitude spectrum of the DFT, i.e. the frequency (band) region where most of the energy is concentrated

Remarks: used as measure of sound sharpness (strength of high frequency energy)

Department of Computational Perception

Spectral Centroid: Illustration



Advanced Content-Based Methods

In the following, we will look into…

Feature extraction

- MFCCs: to model timbral properties
- Fluctuation Patterns: to model rhythmic/periodic properties

Similarity calculation

- Statistical modeling ("Bag-of-frames")
- Vector Space Model

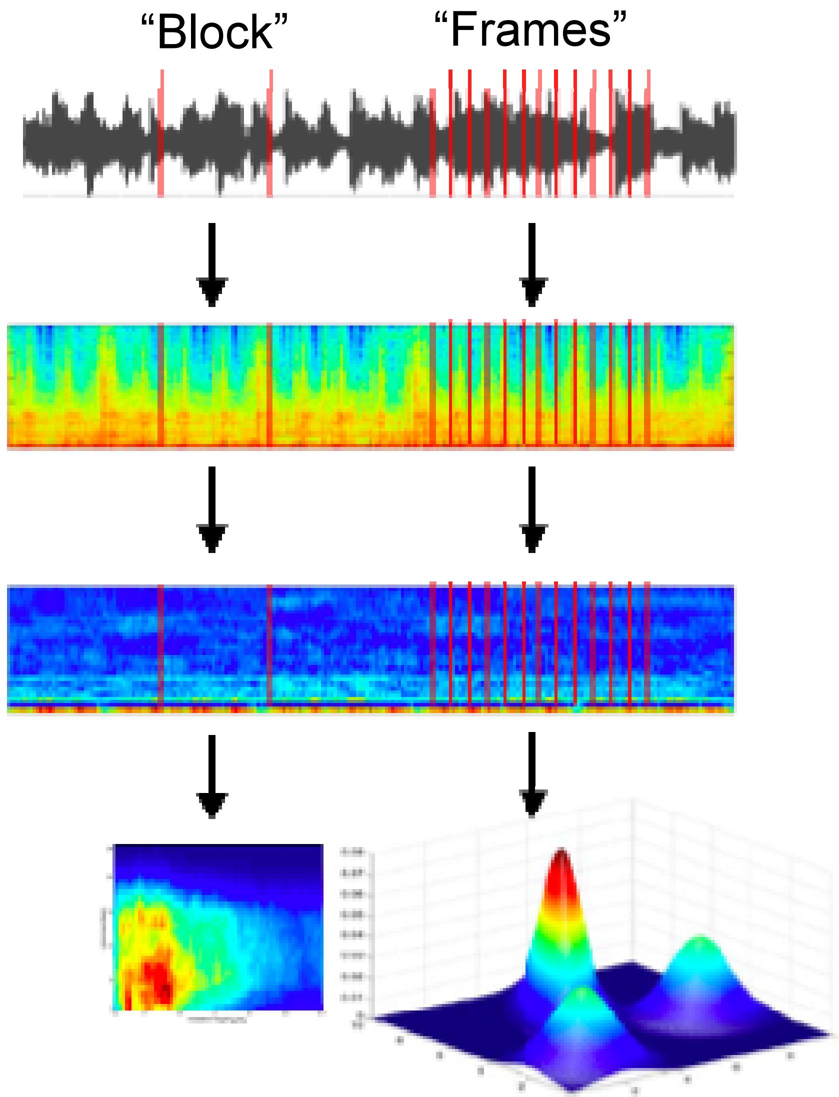
By means of two standard similarity approaches:

- Bag-of-frames modeling using MFCCs
- Comparing Fluctuation Patterns

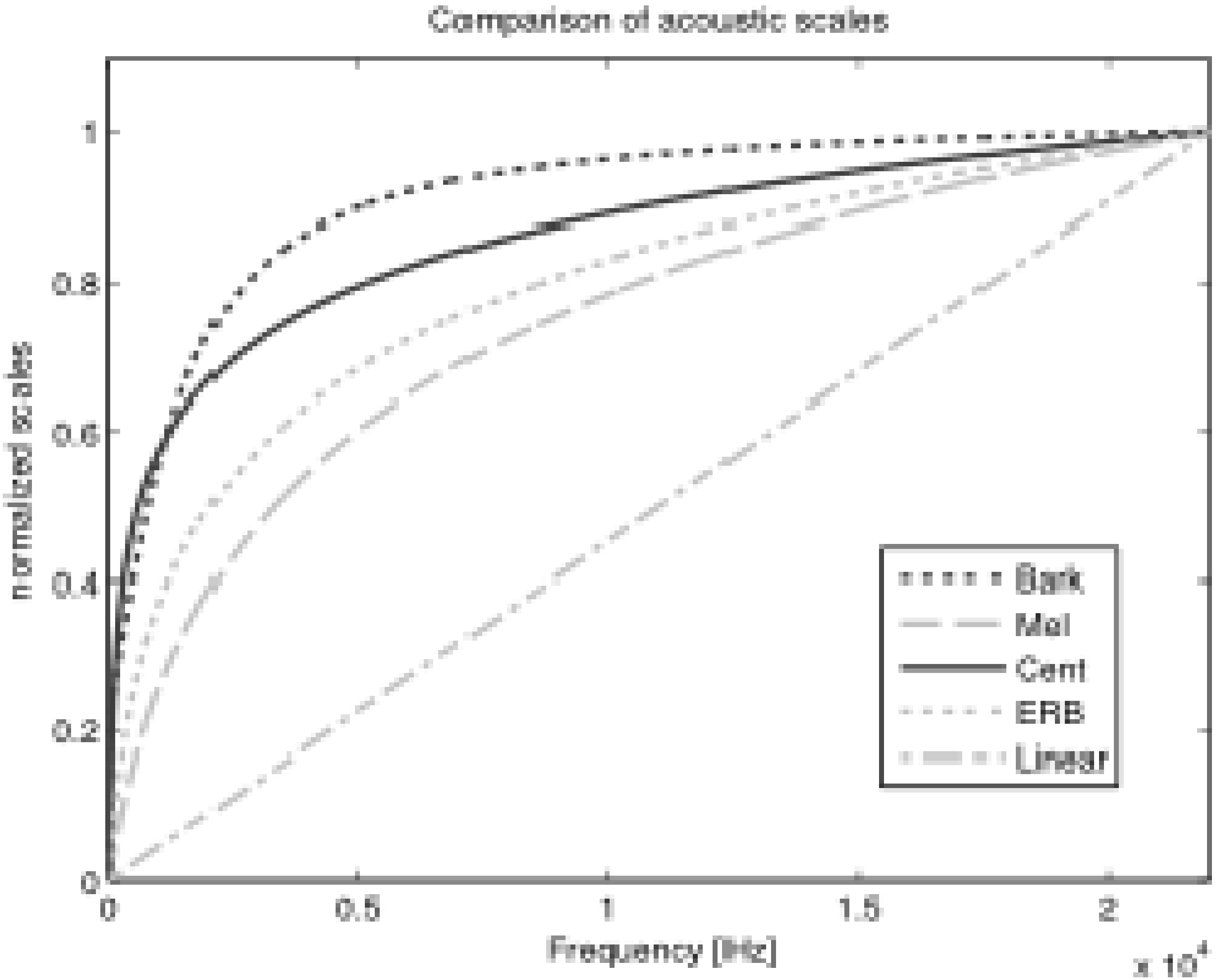
Department of Computational Perception

Processing Overview

- Convert signal to *frequency domain*, e.g., using an FFT
- *Psychoacoustic transformation* (Mel-scale, Bark-scale, Cent-scale, ...): mimics human listening process (not linear, but logarithmic!), removes aspects not perceived by humans, emphasizes low frequencies
- Extract features
  - *Block-level* (large time windows, e.g., 6 sec)
  - *Frame-level* (short time windows, e.g., 25 ms) needs feature distribution model

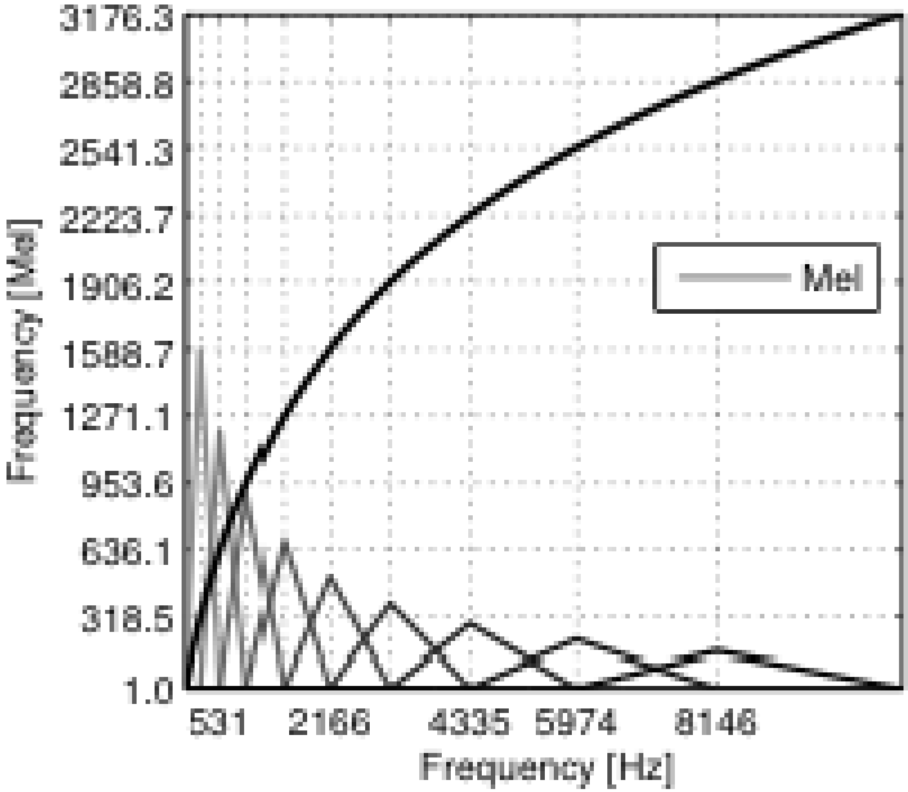


Acoustic Scales



Department of Computational Perception

The Mel Scale



Perceptual scale of pitches judged by listeners to be equal in distance from one another

Given Frequency *f* in Hertz, the corresponding pitch in Mel can be computed by

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

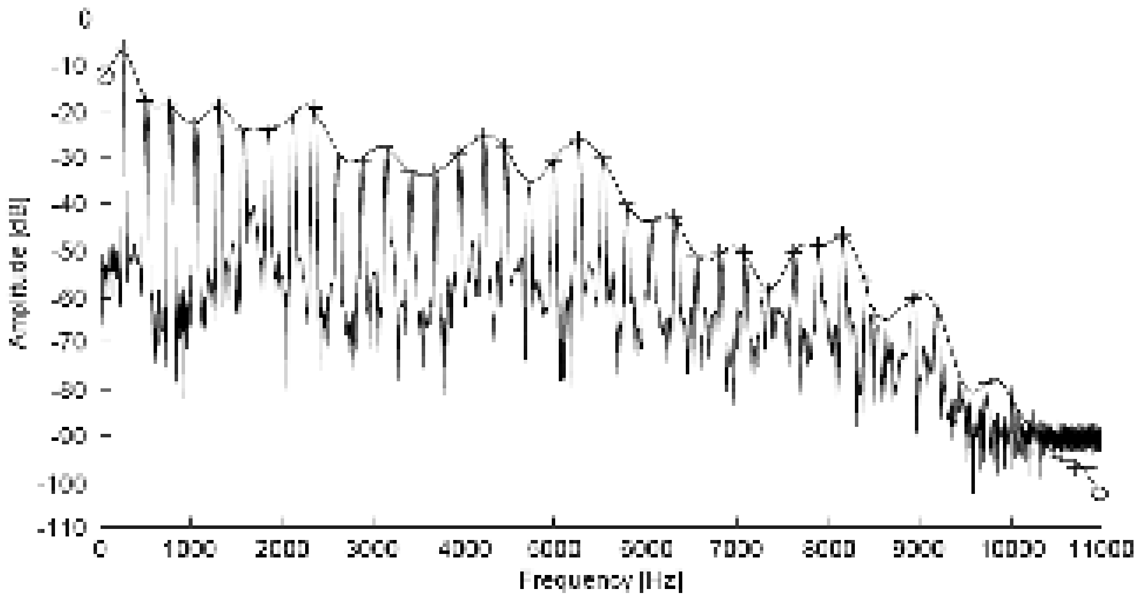
Normally around 40 bins equally spaced on the Mel scale are used

Department of Computational Perception

MFCCs

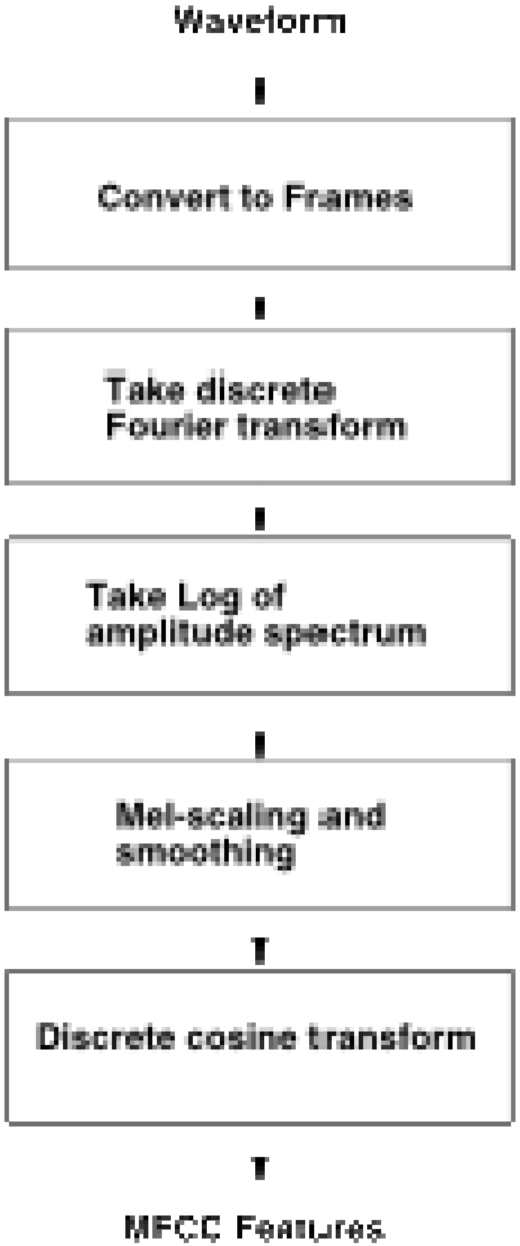
Mel Frequency Cepstral Coefficients (MFCCs) have their roots in speech recognition and are a way to represent the *envelope of the power spectrum* of an audio frame

- the spectral envelope captures perceptually important information about the corresponding sound excerpt (*timbral aspects*)
- most important for music similarity: sounds with similar spectral envelopes are generally perceived as similar.



Department of
Computational
Perception

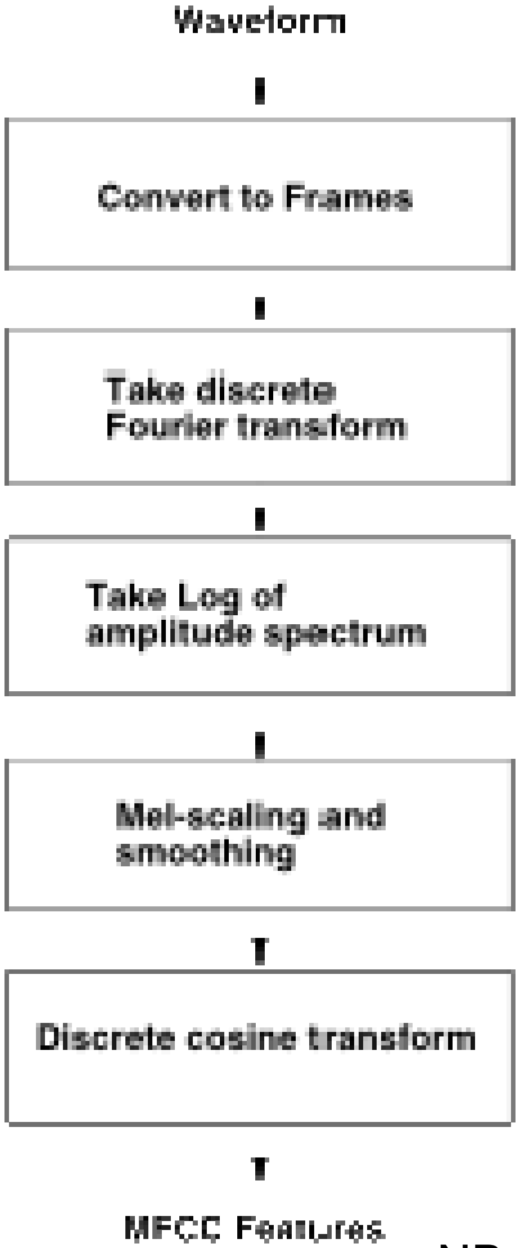
MFCCs



MFCCs are computed per frame

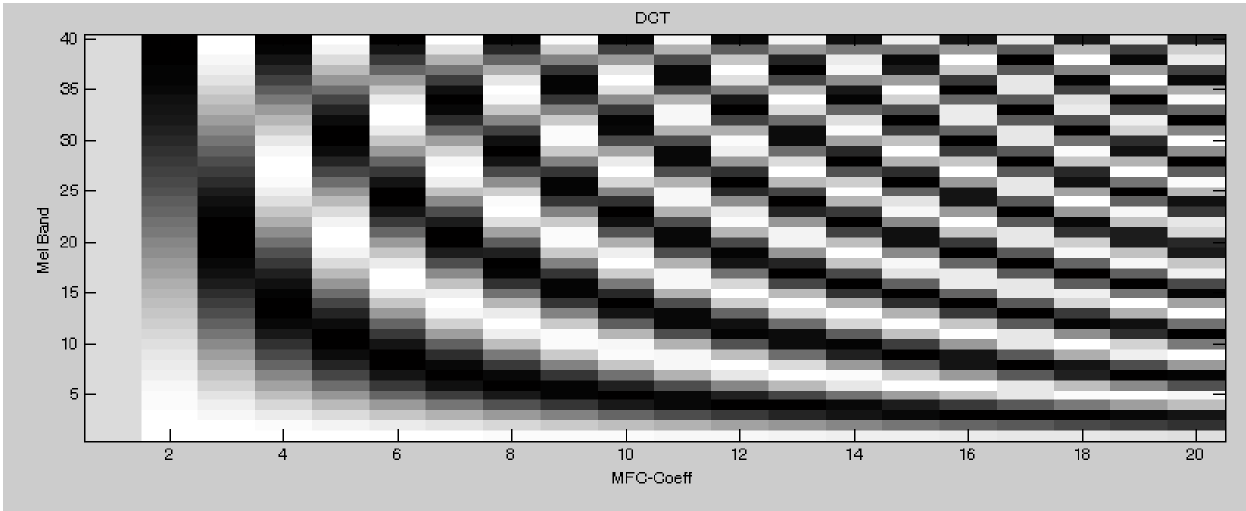
1. STFT: short-time Fourier transform
2. the logarithm of the amplitude spectrum is taken (motivated by the way we humans perceive loudness)
3. mapping of the amplitude spectrum to the Mel scale
4. quantize (e.g., 40 bins) and
   make linear (DCT doesn't operate on log scale)

Department of Computational Perception



5. perform Discrete Cosine Transform to de-correlate the Mel-spectral vectors
   - similar to FFT; only real-valued components
   - describes a sequence of finitely many data points as sum of cosine functions oscillating at different frequencies
   - results in n coefficients (e.g., $n = 20$)

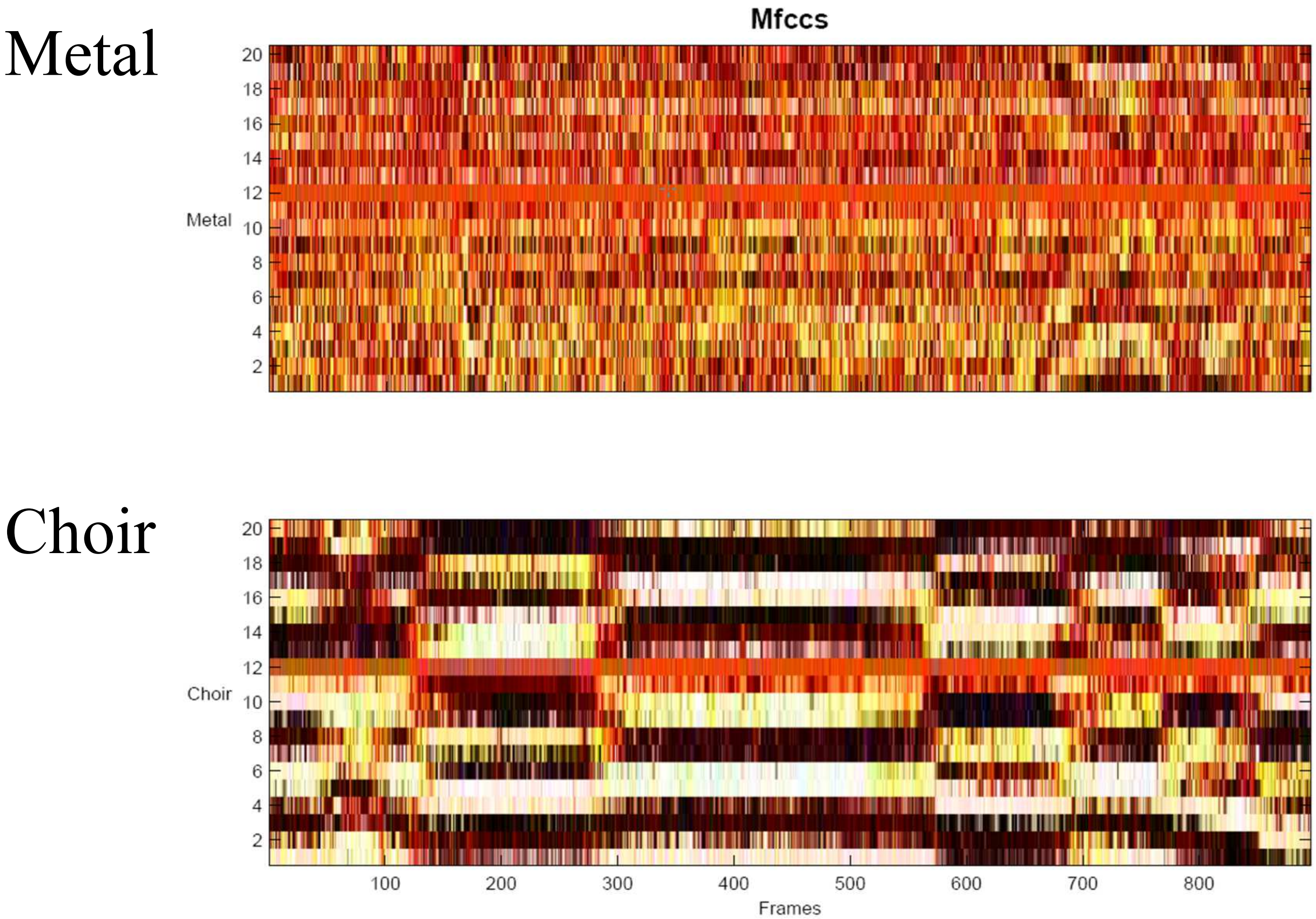
$$X_k = \sum_{n=0}^{N-1} x_n \cdot \cos\left(\frac{\pi}{N} \cdot \left(n + \frac{1}{2}\right) \cdot k\right) \qquad k = 0, \ldots, N-1$$



NB: performing (inverse) FT or similar on log representation of spectrum => "cepstrum" (anagram!)

Department of Computational Perception

MFCC Examples



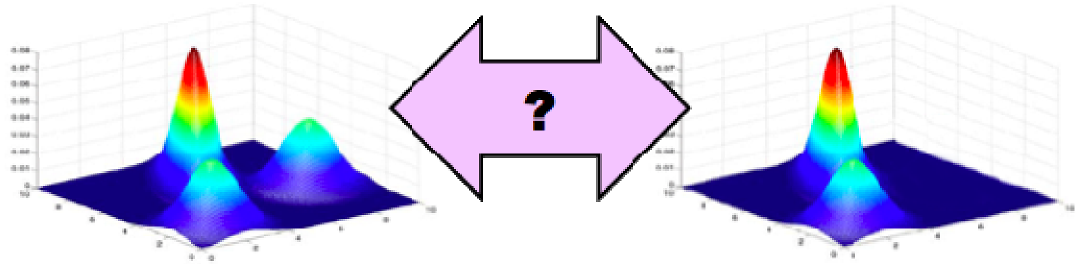
“Bag-of-frames” Modeling

- Full music piece is now a set of MFCC vectors; number of frames depends on length of piece
- Need summary/aggregation/modeling of this set
  - Average over all frames? Sum?
- Most common approach: Statistically model the distribution of all these local features
  - memory requirements, runtime and also the recommendation quality depend on this step
- Learn model that explains the data best
  - State-of-the-art until 2005: learn a Gaussian Mixture Model (GMM)
  - a GMM estimates a probability density as the weighted sum of M simpler Gaussian densities, called components of the mixture
  - each song is modeled with a GMM
  - the parameters of the GMM are learned with the classic Expectation-Maximization (EM) algorithm
    - this can be considered a shortcoming of this approach as this step is very time consuming

Department of Computational Perception

“Bag-of-frames” Modeling

Comparing two GMMs is non-trivial and expensive



- The Kullback-Leibler divergence can be used (approximated)

$$D_{KL}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx$$

- Basically, this requires to (Monte-Carlo) sample one GMM and calculate the likelihood of these observations under the other model and vice versa (non-deterministic, slow)

State-of-the-Art since 2005: Single Gaussian Model

Single Gaussian "Bag-of-frames" model

Describe the frames using the mean vector and a full covariance matrix

For single Gaussian distributions, a closed form of the KL-divergence exists (not a metric!)

$$D_{\mathrm{KL}}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma_1^{-1}\Sigma_0\right) + (\mu_1 - \mu_0)^\top \Sigma_1^{-1}(\mu_1 - \mu_0) - \ln\left(\frac{\det \Sigma_0}{\det \Sigma_1}\right) - k\right)$$

- μ ... mean, Σ ... cov. mat., *tr* ... trace, k ... dimensionality
- asymmetric, symmetrize by averaging

Alternatively, calculate Jenson-Shannon Divergence

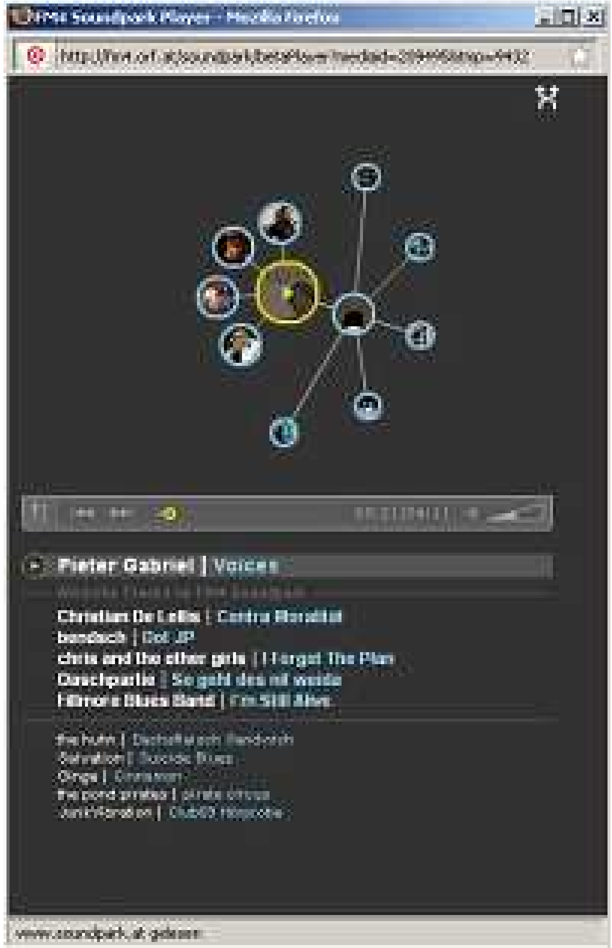
$$JSD(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M) \qquad M = \frac{1}{2}(P + Q) \quad (D = D_{KL})$$

- symmetric, square root is a metric!

Efficient (instantaneous retrieval of 10Ks of pieces)

Query-by-Example in the Real World

- Single Gaussian MFCC music similarity measure used in FM4 Soundpark Player
- For each played song, 5 similar sounding songs are recommended
- Retrieval in real-time
  - full database ~20K songs (?)
  - played song model compared to all whenever played
  - no caching necessary



http://fm4.orf.at/soundpark/

Limitations of Bag-of-Frames Approaches

Loss of Temporal Information:

- temporal ordering of the MFCC vectors is completely lost because of the distribution model (bag-of-frames)
- possible approach: calculate delta-MFCCs to preserve difference between subsequent frames

Hub Problem ("Always Similar Problem")

- depending on the used features and similarity measure, some songs will yield high similarities with many other songs without actually sounding similar (requires post-processing to prevent, e.g., recommendation for too many songs)
- general problem in high-dimensional feature spaces

Department of Computational Perception

Wrapping up MFCCs and BoF

- Similarity model applicable to real-world tasks
- Satisfactory results ("world's best similarity measure" for several years)
- Extensions make it applicable to search within millions of songs in real-time
  - approximate searching in lower-dimensional projection
- Possible Alternatives to BoF:
  - Hidden Markov Models
  - Vector Quantization Models ("Codebook")
  - …

Department of Computational Perception

Block-Level Features

Instead of processing single frames, compute features on larger blocks of frames

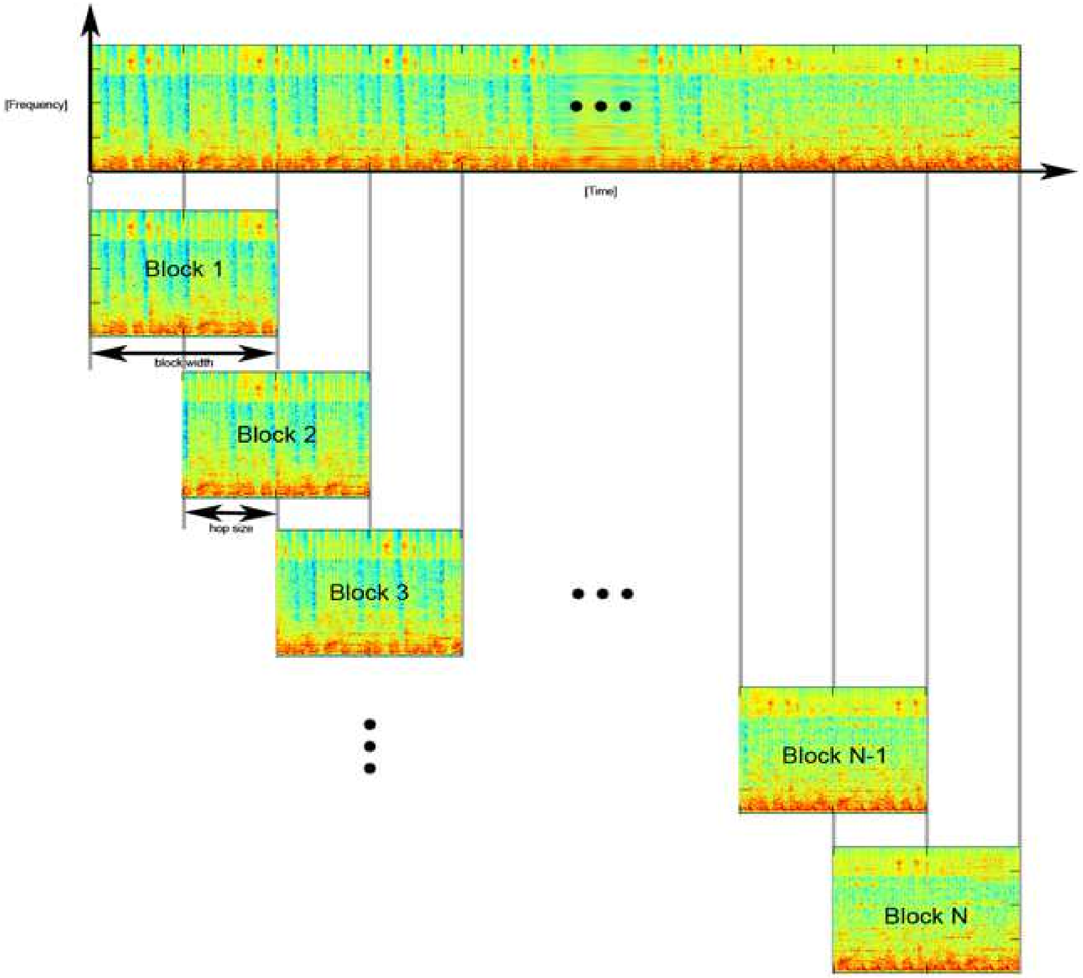
- blocks are defined as consecutive sequences of audio frames
- thus features are (to some extent) able to capture local temporal information

Afterwards the blocks are summarized to form a generalized description of the piece of music

Two systems:

- Fluctuation Patterns (Pampalk; 2001)
- Block Level Framework (Seyerlehner; 2010)

Department of Computational Perception



Block Processing

- The whole spectrum is processed in terms of blocks
- Each block consists of a fixed number of frames (block size W)
- Number of rows H is defined by the frequency resolution
- Blocks may overlap (hop size)
- Main advantage of processing in blocks:
  - blocks allow to perform some (local) temporal processing

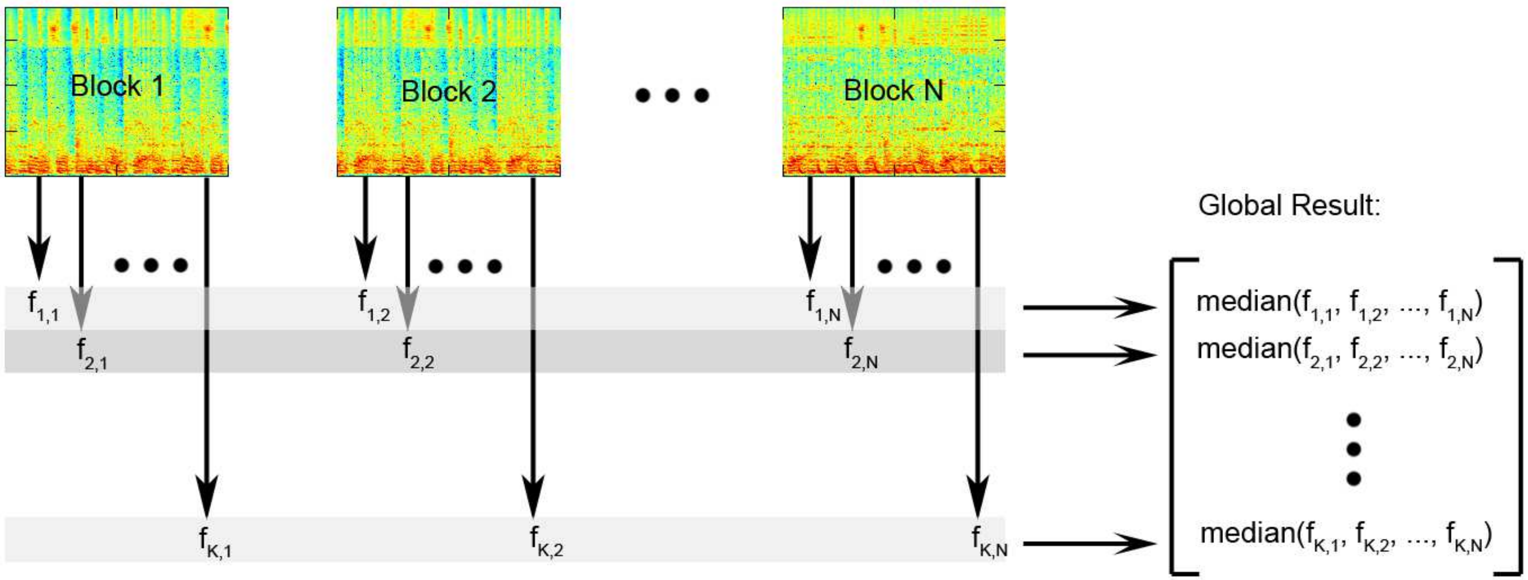
Department of Computational Perception

Generalization

To come up with a global feature vector per song, the local feature vectors must be combined into a single representation

This is done by a summarization function (e.g., mean, median, certain percentiles, variance, …)

The features in the upcoming slides will be matrices, however in these cases the summarization function simply is applied component by component



Department of Computational Perception

Fluctuation Patterns

Idea: measure how strong and fast beats are played within certain perceptually adjusted frequency bands

Aims at capturing periodicities in the signal (“rhythmic properties“)

Incorporates several psychoacoustic transformations

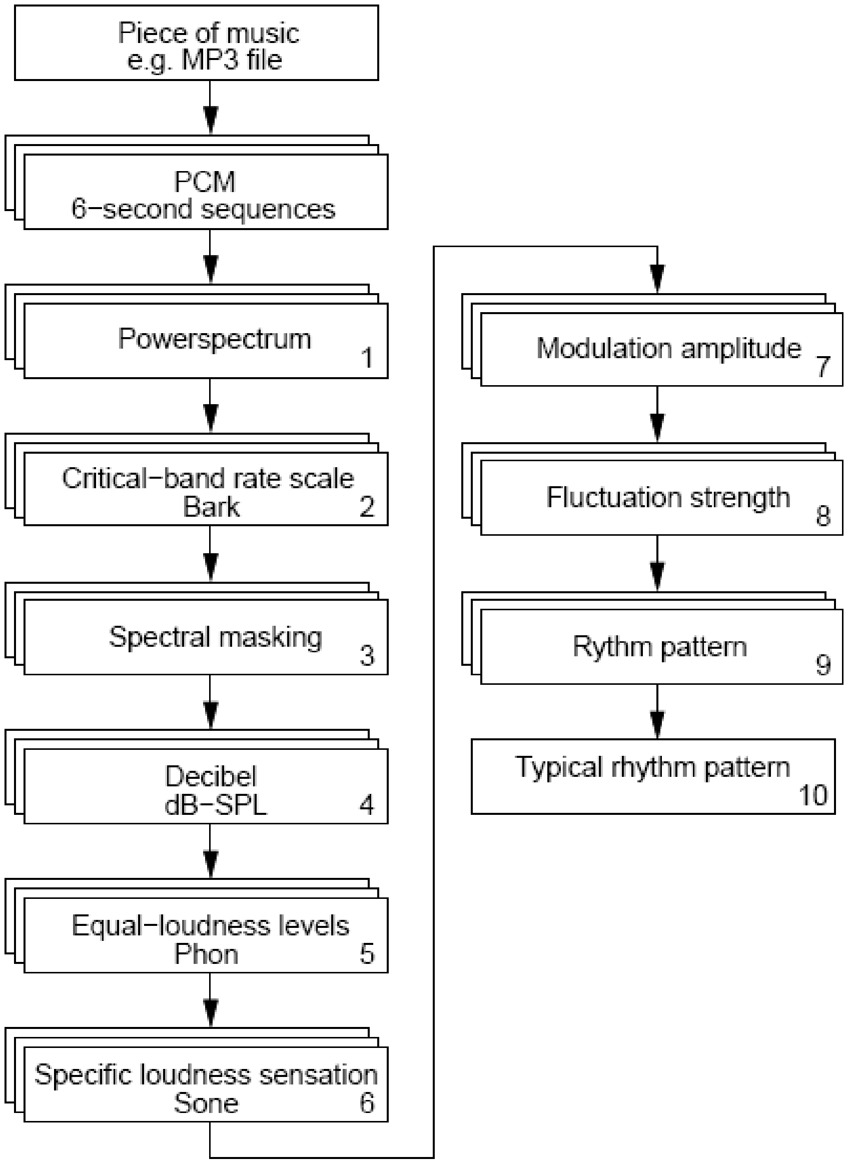
- Logarithmic perception of frequencies (Bark scale)
- Loudness
- Periodicities

Results in a vector description for each music piece

- Vector Space Model
- Favorable for subsequent processing steps and applications: classification, clustering, etc.

Department of Computational Perception

Fluctuation Patterns



Extract 6 sec blocks

– discard beginning and end

In each block:

FFT on Hanning-windowed frames (256 samples)

Convert spectrum to **20 critical bands** according to *Bark scale*

Calculate Spectral Masking effects

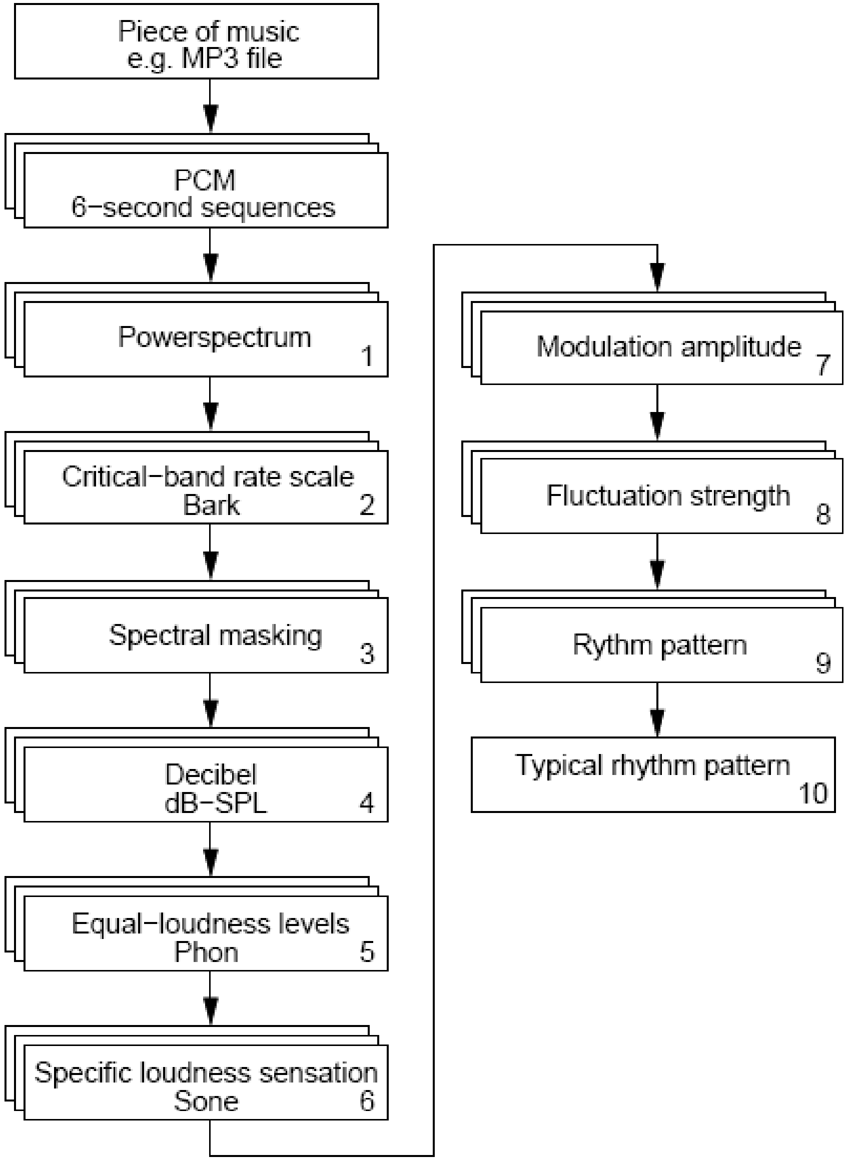
– (i.e. occlusion of a quiet sound when a loud sound is played simultaneously)

Several loudness transformations:

1. to dB (sound intensity)
2. to phon (human sensation: log)
3. to sone (back to linear)

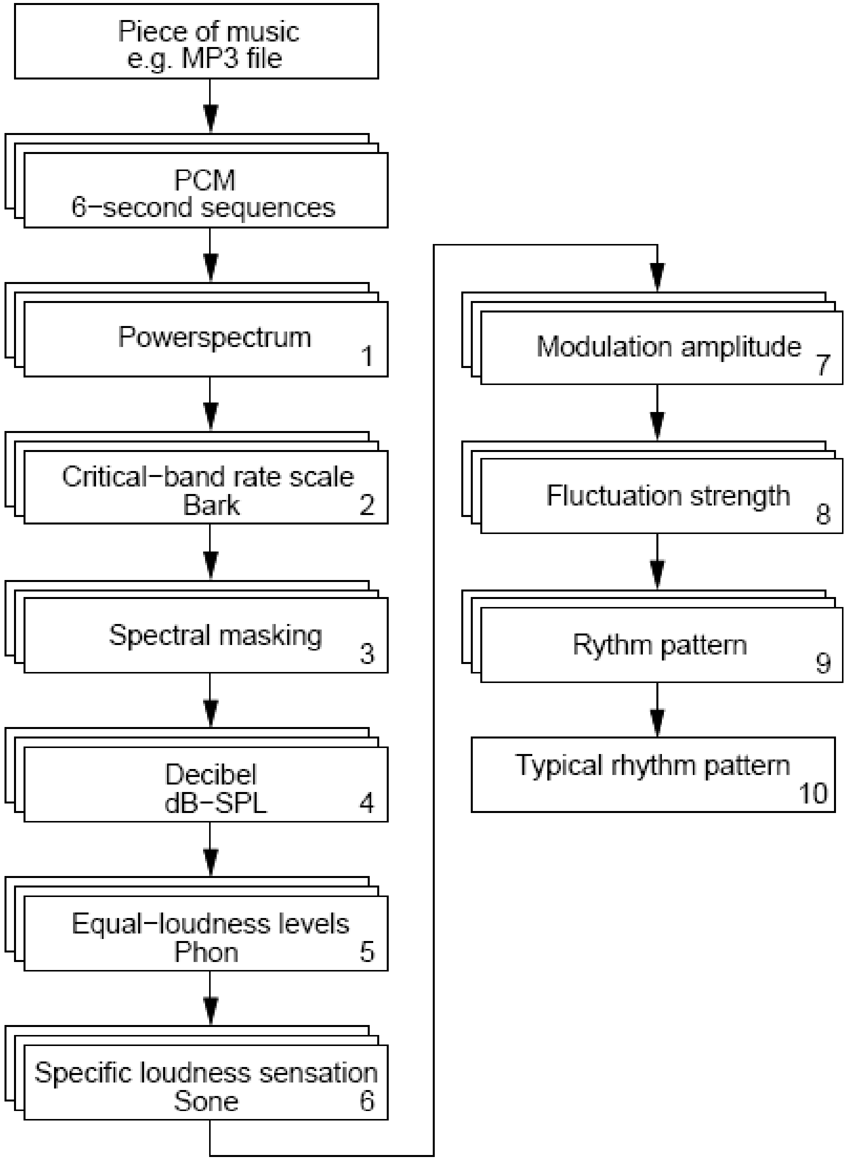
Department of Computational Perception

Fluctuation Patterns



- **Second FFT** reveals information about amplitude modulation, called *fluctuations*.
  - Fluctuations show how often frequencies reoccur at certain intervals within the 6-sec-segment
  - "frequencies of the frequencies"
- Psychoacoustic model of fluctuation strength
  - perception of fluctuations depends on their periodicities
  - reoccurring beats at 4Hz perceived most intensely
  - 60 levels of modulation (per band) (ranging from 0 to 600bpm)
- Emphasize distinctive beats

Fluctuation Patterns



Each block is now respresented as a matrix of fluctuation strengths with 1,200 entries (20 critical bands x 60 levels of modulation)

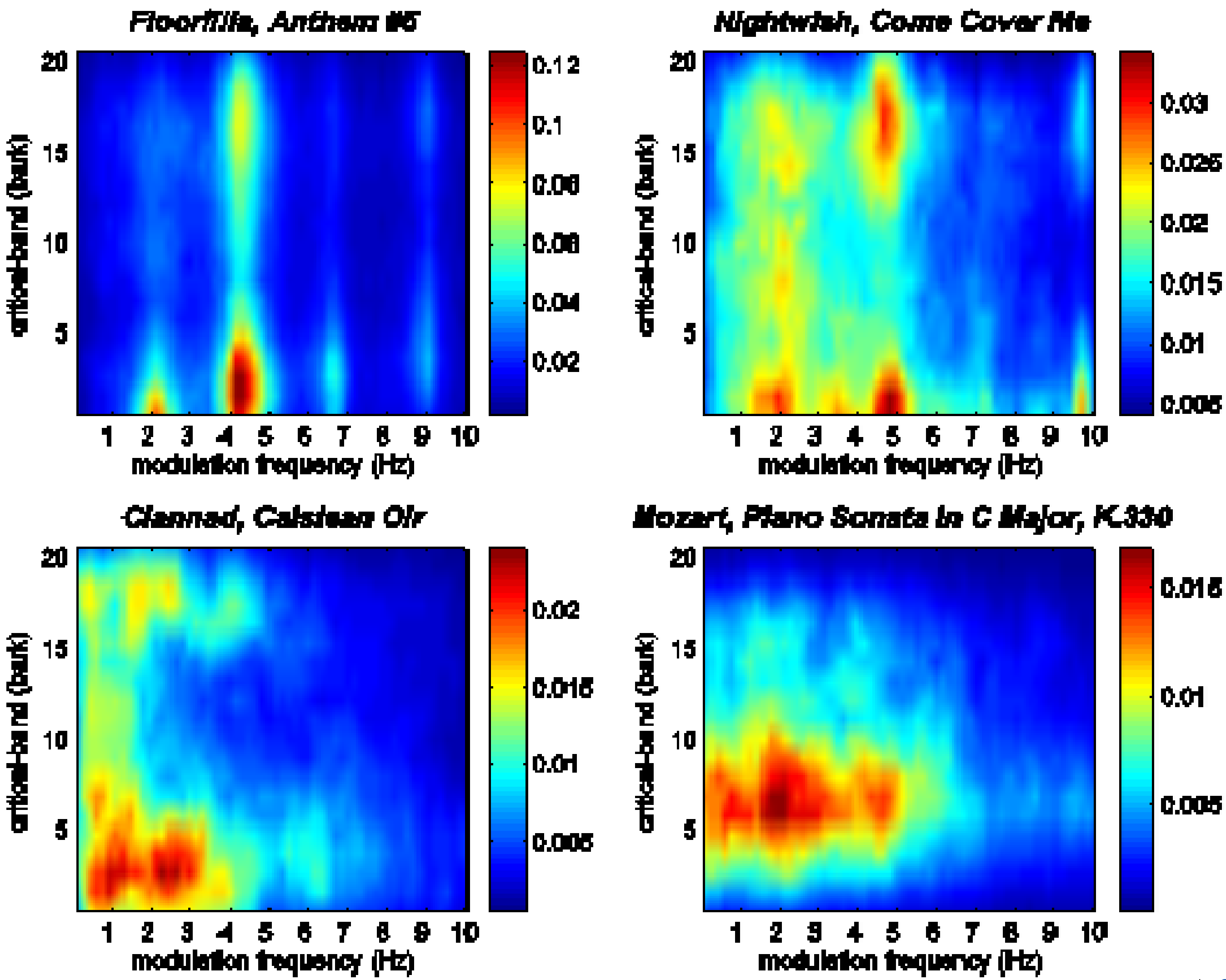
Aggregation of all blocks by taking *median* of each component

This results in a **1,200 dimensional feature vector** for each music piece

Comparison of two music pieces is done by calculating the *Euclidean distance* between their feature vectors

Department of Computational Perception

Examples



FloorFilla, Anthem #5
Nightwish, Come Cover Me
Clannad, Caislean Oir
Mozart, Piano Sonata in C Major, K.330
critical-band (bark)
modulation frequency (Hz)

Wrapping up FPs and VSM

(Some) temporal dependencies are modeled within segments of 6 second length

Properties:

+ Vector Space Model: The whole mathematical toolbox of vector spaces is available.
+ easy to use in classification
+ song models can be visualized

– high dimensional feature space (often a PCA is applied to reduce dim.)

More comprehensive block-level features by (Seyerlehner; 2010) currently best performing similarity measure according to MIREX:

- Spectral Pattern (SP): frequency content
- Delta-Spectral Pattern (DSP): SP on delta frames
- Variance Delta-Spectral Pattern (VDSP): *variance* used to aggregate DSP
- Logarithmic Fluctuation Pattern (LFP): more tempo invariant
- Correlation Pattern (CP): temporal relation of frequency bands
- Spectral Contrast Pattern (SCP): estimate "tone-ness"
- Block **aggregation** via *percentiles*; **similarity** via *Manhattan distance*

Department of Computational Perception

Demo: Content-Based Music Browsing

Department of
Computational
Perception

nepTune – Structuring the Music Space

(Knees et al.; MM 2006)

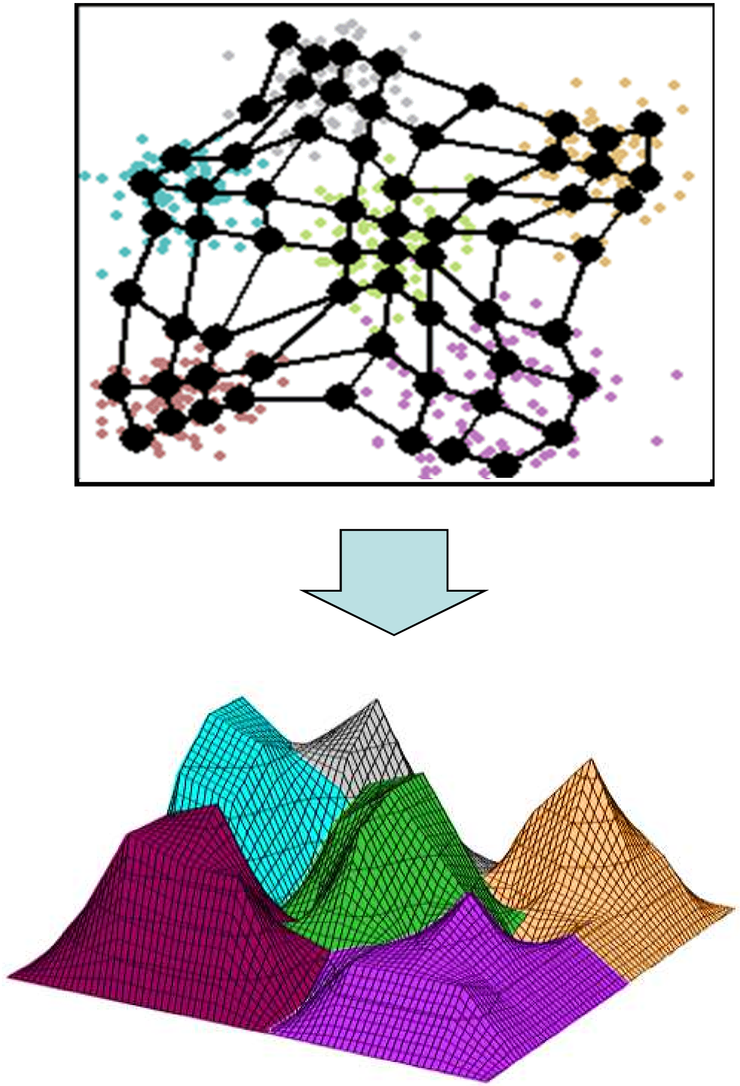
Clustering of music pieces

Each song corresponds to point in feature (similarity) space

Self-organizing Map

High-dimensional data (content-based features) is projected to 2-dim. plane

Number of pieces per cluster
→ landscape height profile



nepTune – Web-based Augmentation

(Knees et al.; MM 2006)

Automatic description of landscape via Web term extraction

artist names (ID3)

Music dictionary

$$G_{w,j} = \frac{F_{w,j}^2}{\sum_i F_{w,i}}$$

Term goodness

Department of Computational Perception

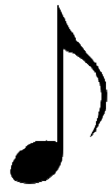
Part III

MUSIC CONTEXT BASED SIMILARITY

Department of
Computational
Perception

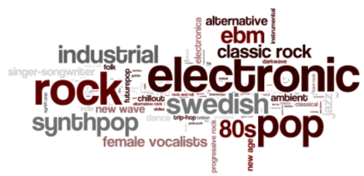
Music Content vs. Music Context

Advantages of Content Analysis



- Features can be extracted from any audio file
- No other data or community necessary
- No cultural biases (i.e., no popularity bias, no subjective ratings etc.)

Advantages of Context Analysis



- Captures aspects beyond pure audio signal
- No audio file necessary
- Usually, user-based features are closer to what users want

Department of Computational Perception

Music Content vs. Music Context

Challenges for Context-Based Feature Extractors

- Dependence on availability of sources (Web pages, tags, playlists, ...)
- Popularity of artists may distort results
- Cold start problem of community-based systems (newly added entities do not have any information associated, e.g. user tags, users' playing behavior)
- Hacking and vandalism (cf. last.fm tag "brutal death metal")
- Bias towards specific user/listener groups (e.g., young, Internet-prone, metal listeners in last.fm)
- (Reliable) data often only available on artist level

Challenge for both Content and Context Analysis

- Extraction of relevant features from *noisy signal*

Department of Computational Perception

Context- and Web-Based Methods

In the following, contextual data refers to extended meta-data, usually

- Generated by users
- Unstructured data-sources
- Accessible via the Web

Two main classes of approaches covered in the following

- Text processing
- Co-occurrence analysis

As for content-based methods, similarity is the central concept for retrieval

Department of Computational Perception

Text-Based Approaches

Data sources:

- Web pages retrieved via Web search engines
- microblogs on *Twitter*
- product reviews
- semantic tags
- lyrics

Text-Based Similarity and Genre Classification

Use Web data to transform the music similarity task into a text similarity task

Allows to use the full armory of IR methods, typically…

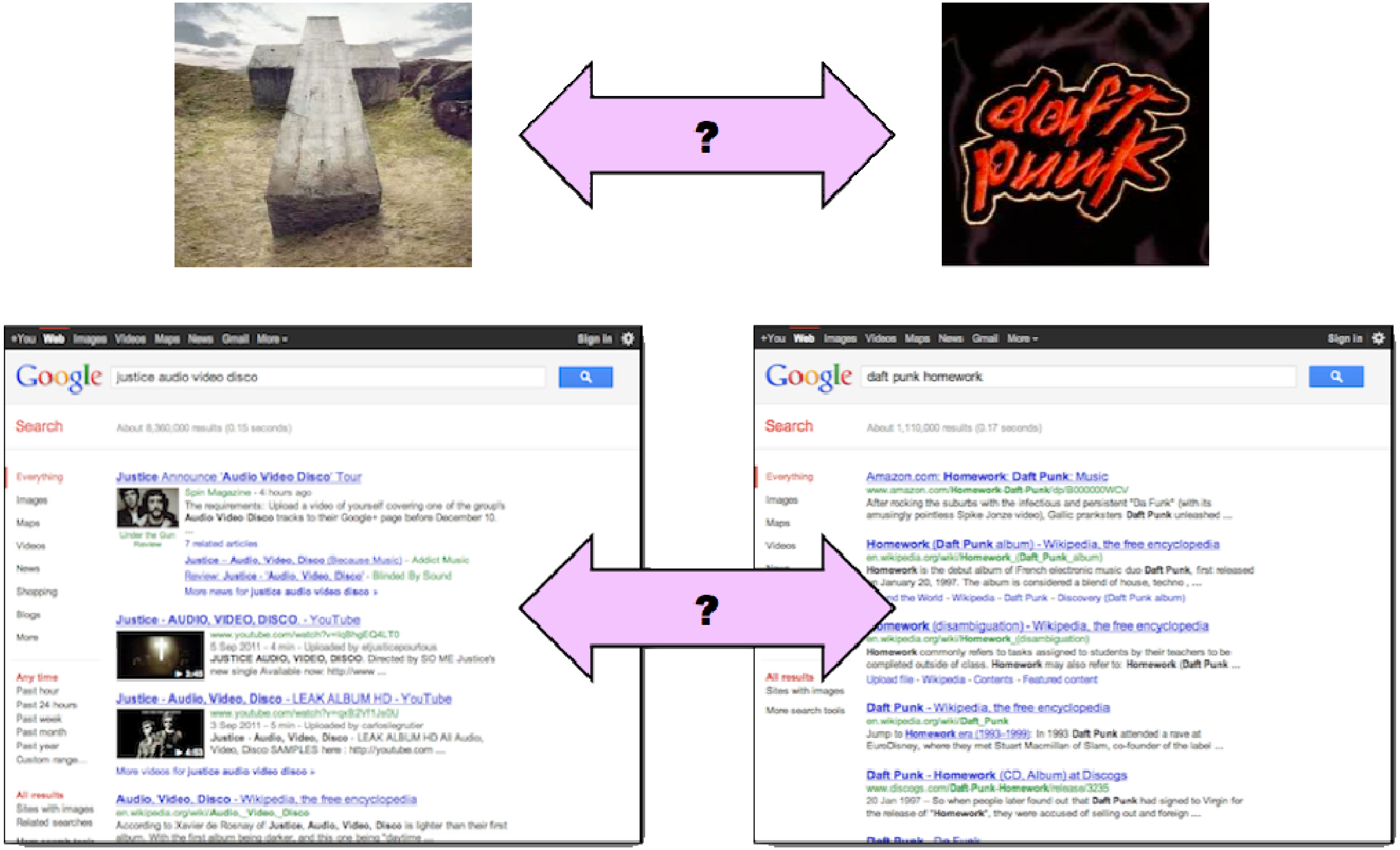
- Bag-of-words, Vector Space Model
- Stopword removal, dictionaries, term selection
- TF·IDF
- Latent Semantic Indexing
- Part-of-Speech tagging
- Named Entity Detection
- Sentiment analysis

Large range of possible similarity measures

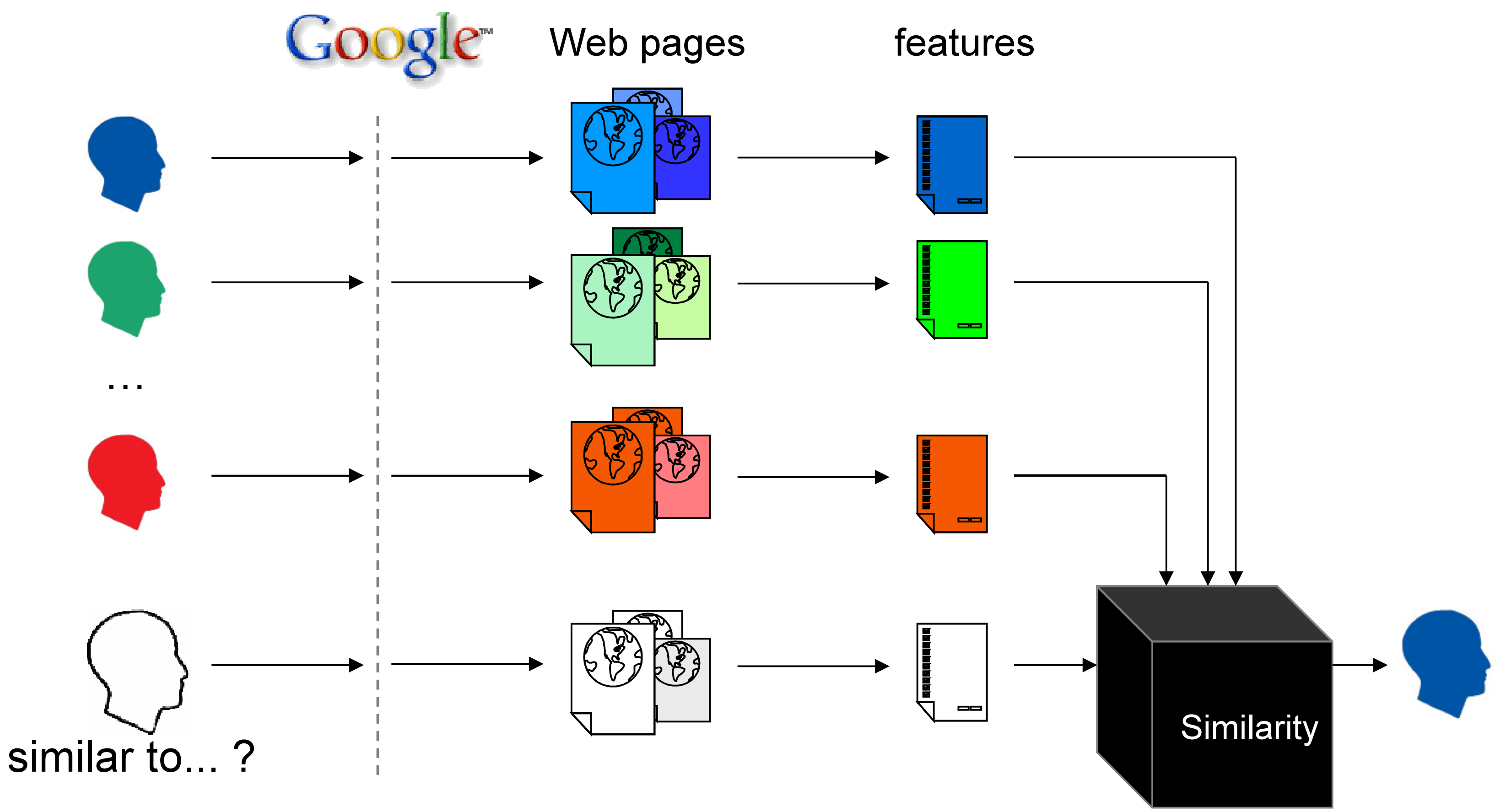
- Overlap, Manhattan, Euclidean, Cosine, etc.

Department of Computational Perception

Related Web Pages as Text Source



Related Web Pages as Text Source



Department of Computational Perception

Related Web Pages as Text Source

- Using search engines and queries such as

```
“artist” +music
“artist” +music +review
```

(Whitman, Lawrence; 2002) (Baumann, Hummel; 2003) (Knees et al.; 2004)

- Analyze
  - result page directly or
  - download up to top 100 Web pages (combine into one “virtual document” or analyze separately)
- Apply “IR magic”
- Applicable for similarity estimation, classification, retrieval, annotation

(NB: Most discriminating terms between genres are artist names and album/track titles)

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- term frequency

Abbr.	Description	Formulation
TF_A	Formulation used for binary match SB = b	$r_{d,t} = \begin{cases} 1 & \text{if } t \in \mathcal{T}_d \\ 0 & \text{otherwise} \end{cases}$
TF_B	Standard formulation SB = t	$r_{d,t} = f_{d,t}$
TF_C	Logarithmic formulation	$r_{d,t} = 1 + \log_e f_{d,t}$
TF_C2	Alternative logarithmic formulation suited for $f_{d,t} < 1$	$r_{d,t} = \log_e(1 + f_{d,t})$
TF_C3	Alternative logarithmic formulation as used in *ltc* variant	$r_{d,t} = 1 + \log_2 f_{d,t}$
TF_D	Normalized formulation	$r_{d,t} = \frac{f_{d,t}}{f_d^m}$
TF_E	Alternative normalized formulation. Similar to [55] we use $K = 0.5$. SB = n	$r_{d,t} = K + (1 - K) \cdot \frac{f_{d,t}}{f_d^m}$
TF_F	Okapi formulation, according to [55, 36]. For W we use the vector space formulation, i.e., the Euclidean length.	$r_{d,t} = \frac{f_{d,t}}{f_{d,t} + W_d / av_{d \in D}(W_d)}$
TF_G	Okapi BM25 formulation, according to [35].	$r_{d,t} = \frac{(k_1+1) \cdot f_{d,t}}{f_{d,t} + k_1 \cdot \left[(1-b) + b \cdot \frac{W_d}{av_{d \in D}(W_d)}\right]}$ $k_1 = 1.2, b = 0.75$

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- **term frequency**
- **inverse document frequency**

Abbr.	Description	Formulation
IDF_A	Formulation used for binary match SB $= x$	$w_t = 1$
IDF_B	Logarithmic formulation SB $= f$	$w_t = \log_e\left(1 + \frac{N}{f_t}\right)$
IDF_B2	Logarithmic formulation used in *ltc* variant	$w_t = \log_e\left(\frac{N}{f_t}\right)$
IDF_C	Hyperbolic formulation	$w_t = \frac{1}{f_t}$
IDF_D	Normalized formulation	$w_t = \log_e\left(1 + \frac{f_m}{f_t}\right)$
IDF_E	Another normalized formulation SB $= p$	$w_t = \log_e \frac{N - f_t}{f_t}$
	The following definitions are based on the term's noise n_t and signal s_t.	$n_t = \sum_{d \in \mathcal{D}_t} \left(-\frac{f_{d,t}}{F_t} \log_2 \frac{f_{d,t}}{F_t}\right)$ $s_t = \log_2(F_t - n_t)$
IDF_F	Signal	$w_t = s_t$
IDF_G	Signal-to-Noise ratio	$w_t = \frac{s_t}{n_t}$
IDF_H		$w_t = \left(\max_{t' \in \mathcal{T}} n_{t'}\right) - n_t$
IDF_I	Entropy measure	$w_t = 1 - \frac{n_t}{\log_2 N}$
IDF_J	Okapi BM25 IDF formulation, according to [35, 31]	$w_t = \log \frac{N - f_t + 0.5}{f_t + 0.5}$

Department of Computational Perception

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- term frequency

- inverse document frequency

- virtual document modeling: *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.

Department of Computational Perception

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- term frequency

- inverse document frequency

- virtual document modeling: *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.

- normalization with respect to document length

Abbr.	Description	Formulation
NORM_NO	No normalization.	
NORM_SUM	Normalize sum of each virtual document's term feature vector to 1.	$\sum_{t \in \mathcal{T}_d} r_{d,t} = 1$
NORM_MAX	Normalize maximum of each virtual document's term feature vector to 1.	$\max_{t \in \mathcal{T}_d} r_{d,t} = 1$

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- **term frequency**
- **inverse document frequency**
- **virtual document modeling:** *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.
- **normalization with respect to document length**
- **similarity measure**

Abbr.	Description	Formulation
SIM_INN	Inner Product	$S_{d_1,d_2} = \sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})$
SIM_COS	Cosine Measure	$S_{d_1,d_2} = \frac{\sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}{W_{d_1} \cdot W_{d_2}}$
SIM_DIC	Dice Formulation	$S_{d_1,d_2} = \frac{2 \sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}{W_{d_1}^2 + W_{d_2}^2}$
SIM_JAC	Jaccard Formulation	$S_{d_1,d_2} = \frac{\sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}{W_{d_1}^2 + W_{d_2}^2 - \sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}$
SIM_OVL	Overlap Formulation	$S_{d_1,d_2} = \frac{\sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} \cdot w_{d_2,t})}{\min(W_{d_1}^2, W_{d_2}^2)}$
SIM_EUC	Euclidean Similarity	$D_{d_1,d_2} = \sqrt{\sum_{t \in \mathcal{T}_{d_1,d_2}} (w_{d_1,t} - w_{d_2,t})^2}$ $S_{d_1,d_2} = \left(\max_{d_1',d_2'} (D_{d_1',d_2'})\right) - D_{d_1,d_2}$
SIM_JEF	Jeffrey Divergence-based Similarity	$S_{d_1,d_2} = \left(\max_{d_1',d_2'} (D_{d_1',d_2'})\right) - D_{d_1,d_2}$ $D(F,G) = \sum \left(f_i \log \frac{f_i}{m_i} + g_i \log \frac{g_i}{m_i} \right)$

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- **term frequency**
- **inverse document frequency**
- **virtual document modeling:** *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.
- **normalization with respect to document length**
- **similarity measure**
- **index term set**

Abbr. / Term Set	Cardinality	Description
`TS_A - all_terms`	`C224a, QS_A`: 38,133 `C224a, QS_M`: 19,133 `C3ka, QS_A`: 1,489,459 `C3ka, QS_M`: 437,014	All terms (stemmed) that occur in the corpus of the retrieved `Twitter` posts.
`TS_S - scowl_dict`	698,812	All terms that occur in the entire `SCOWL` dictionary.
`TS_N - artist_names`	224 / 3,000	Names of the artists for which data was retrieved.
`TS_D - dictionary`	1,398	Manually created dictionary of musically relevant terms.
`TS_L - last.fm_toptags`	250	Overall top-ranked tags returned by `last.fm`'s *Tags.getTopTags* function.
`TS_F - freebase`	3,628	Music-related terms extracted from `Freebase` (genres, instruments, emotions).

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- **term frequency**
- **inverse document frequency**
- **virtual document modeling:** *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.
- **normalization with respect to document length**
- **similarity measure**
- **index term set**
- **query scheme**

Abbr.	Query Scheme
QS_A	"artist name"
QS_M	"artist name"+music

Department of Computational Perception

Large-Scale Study

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments): (Schedl et al.; 2011)

- **term frequency**
- **inverse document frequency**
- **virtual document modeling:** *concatenate* all Web pages/posts of the artist or perform *aggregation* via mean, max, etc.
- **normalization with respect to document length**
- **similarity measure**
- **index term set**
- **query scheme**

implemented in our ***CoMIRVA*** framework available from http://www.cp.jku.at/comirva

Department of Computational Perception

Interesting Findings

(Schedl et al.; 2011)

- modeling artists as ***virtual documents*** is preferable
- using query scheme ***"artist" +music*** outperforms ***"artist"***
- ***normalization*** does not yield a statistically significant difference
- standard ***cosine similarity*** measure does not yield the very best results, but the most stable ones (varying other parameters)
- ***consistent results*** among the (top-ranked) variants for two collections
- ***minor change in one parameter*** can have a huge impact on performance
- overall winners in terms of term weighting functions:

  `TF_C3.IDF_I`

  `TF_C3.IDF_H` → logarithmic formulations for TF and IDF

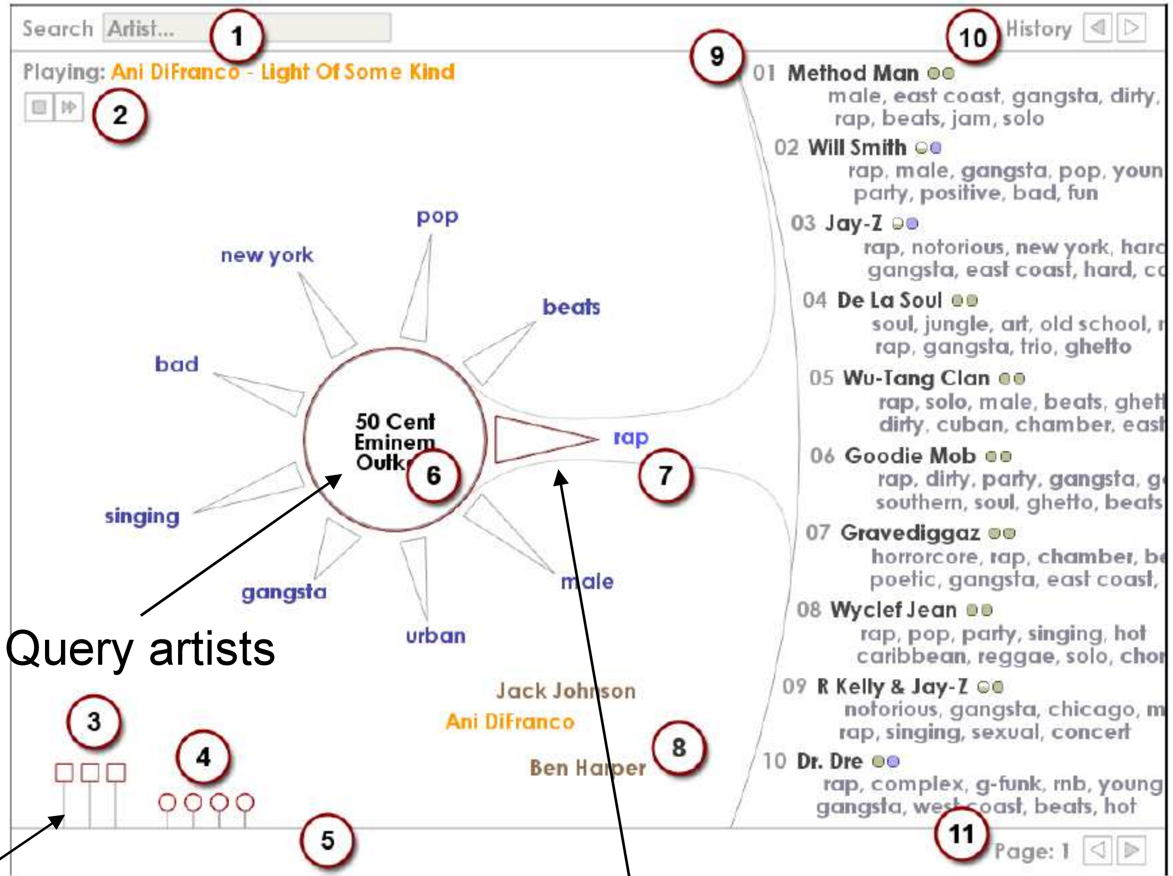
  `TF_C2.IDF_I`

Department of Computational Perception

Web-Based Descriptions for Browsing

"MusicSun"
(Pampalk, Goto; 2007)

- Interactive "Artist Recommender"
- Recommendation is influenced/directed by selecting relevant similarity dimensions
- Combines different similarity measures



Query artists

3 types of similarity: audio, web-based, word
overall similarity = weighted average of ranks

Relevant word dimension

Department of Computational Perception

Web-Based Texts for Indexing and Retrieval

- Use Web data to transform music retrieval into a text retrieval task
- Find associated (or associable) texts and use them instead of music
- Allows for diverse and semantic queries
  (e.g, “chilled music”, “great riffs”)

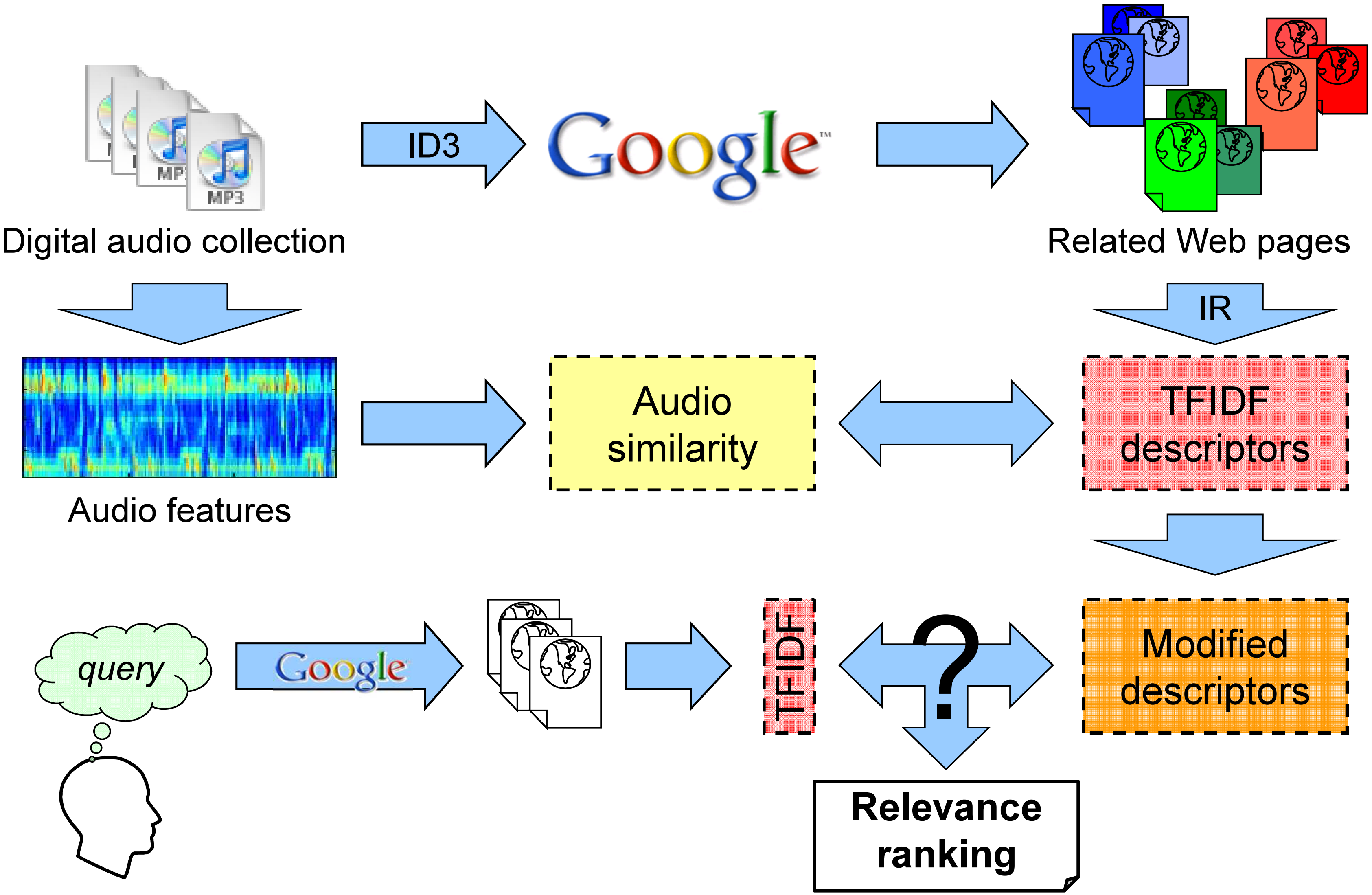
Search Sounds (Celma et al.; 2006)
Crawl lists of RSS feeds and use Weblog entries to index pieces

Squiggle (Celino et al.; 2006):
Combine meta-data databases (like MusicBrainz) for rich indexing

Gedoodle (Knees et al.; SIGIR 2007):
Query Google and combine Web pages to index pieces

Department of
Computational
Perception

Gedoodle



ID3
Google
Digital audio collection
Related Web pages
IR
Audio features
Audio similarity
TFIDF descriptors
query
Google
TFIDF
?
Modified descriptors
Relevance ranking

Gedoodle

(Knees et al.; SIGIR 2007)

- For each track: join 100 Google results of
  - “artist” music
  - “artist” “album” music review
  - “artist” “title” music review -lyrics
- Combine all pages into one virtual document
- Create normalized TFIDF vector for each track
- Include audio similarity for vector modification and dimensionality reduction

Department of Computational Perception

Gedoodle (Example queries)

damon albarn | Search for Music

Results **18** - **27** of **1691** for **damon albarn**. (**0.02** seconds)

Music Is My Radar
by blur
from the album: *blur: the best of*
Genre: Alternative - 192 kBit/s - length: 5:29 min. Listen

On Your Own
by blur
from the album: *blur: the best of*
Genre: Alternative - 192 kBit/s - length: 4:27 min. Listen

Girls & Boys
by blur
from the album: *blur: the best of*
Genre: Alternative - 192 kBit/s - length: 4:19 min. Listen

There's No Other Way
by blur
from the album: *blur: the best of*
Genre: Alternative - 192 kBit/s - length: 3:14 min. Listen

Feel Good Inc.
by Gorillaz
from the album: *FM4 Soundselection 12*
Genre: Alternative - 192 kBit/s - length: 4:20 min. Listen

Say So What
by Graham Coxon
from the album: *Uncut - 2006.05*
Genre: Rock - 192 kBit/s - length: 3:05 min. Listen

Slash Dot Dash
by Fatboy Slim
from the album: *The Greatest Hits: Why Try Harder*
Genre: Electronic - 192 kBit/s - length: 2:55 min. Listen

smooth and relaxing | Search for Music

Results **1** - **10** of **1774** for **smooth and relaxing**. (**0.02** seconds)

Joy And Pain
by Count Basic
from the album: *Moving In The Right Direction*
Genre: Acid Jazz - 168 kBit/s - length: 6:25 min. Listen

Higher
by Count Basic
from the album: *Bigger & Brighter*
Genre: Acid Jazz - 192 kBit/s - length: 4:00 min. Listen

Sweet Luis
by Count Basic
from the album: *Moving In The Right Direction*
Genre: Acid Jazz - 158 kBit/s - length: 5:11 min. Listen

Got To Do
by Count Basic
from the album: *Moving In The Right Direction*
Genre: Acid Jazz - 167 kBit/s - length: 4:58 min. Listen

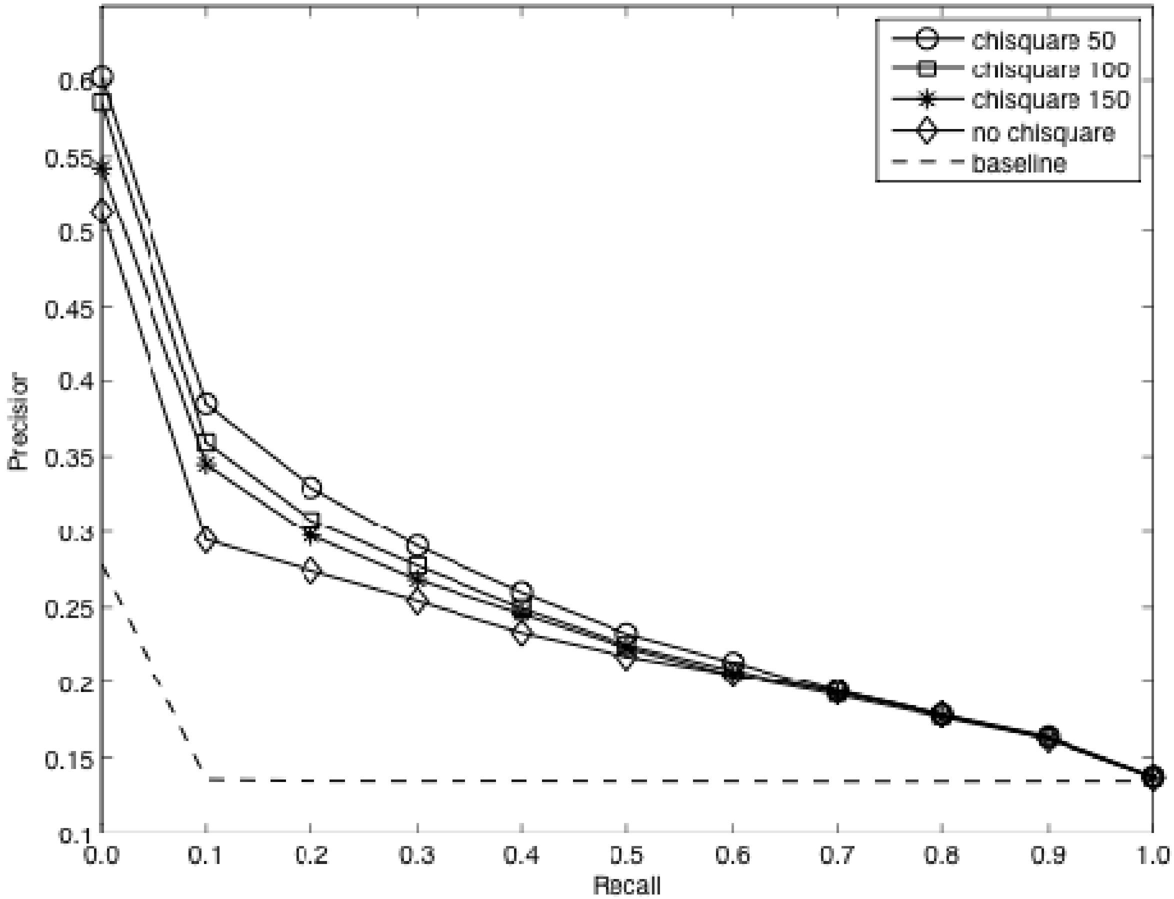
John Lee Huber
by Tosca
from the album: *J.A.C.*
Genre: Electronica/Dance - 192 kBit/s - length: 4:33 min. Listen

No More Olives
by Tosca
from the album: *J.A.C.*
Genre: Electronica/Dance - 192 kBit/s - length: 6:02 min. Listen

Naschkatze
by Tosca
from the album: *J.A.C.*
Genre: Electronica/Dance - 192 kBit/s - length: 4:34 min. Listen

Gedoodle Results

Effects of TFIDF feature space pruning using content-similarity-based χ^2-test (Knees et al.; SIGIR 2007)



Department of
Computational
Perception

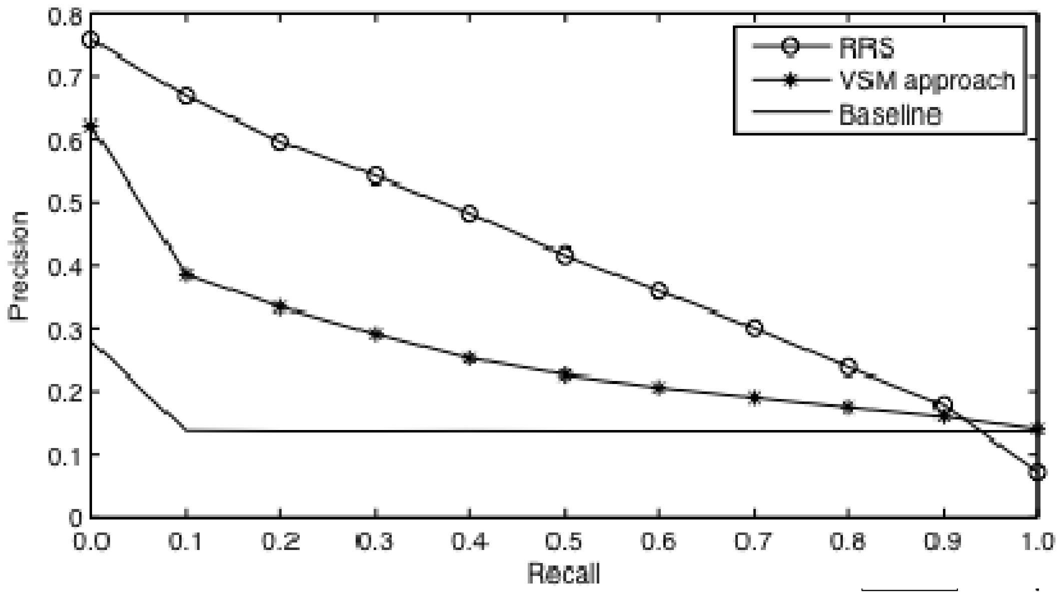
Gedoodle Results

Alternative: Document-centered ranking (Knees et al.; ECIR 2008)

- Indexing of all web documents in standard index
- Music query addresses this index
- Music ranking calculated from web doc ranking according to

$$RRS(m,q) = \sum_{p \in D_m \cap D_q} 1 + |D_q| - rank(p, D_q)$$

Comparison with vector space model



Semantic Querying via Auto-Tagging

- Use machine learning techniques to predict tags (labels) based on song features (content, context, or combination)
- Automatic description of music (browsing) and automatic generation of indexing terms for retrieval
- Mitigates "cold-start problem" in social tagging

Automatic Record Reviews (Whitman, Ellis; 2004)

Regularized least squares learning on TFIDF-Web and cepstral features

Autotagger (Bertin-Mahieux et al.; 2008)

Ensemble classifier to map MFCCs, autocorrelation, Const-Q. to Web tags

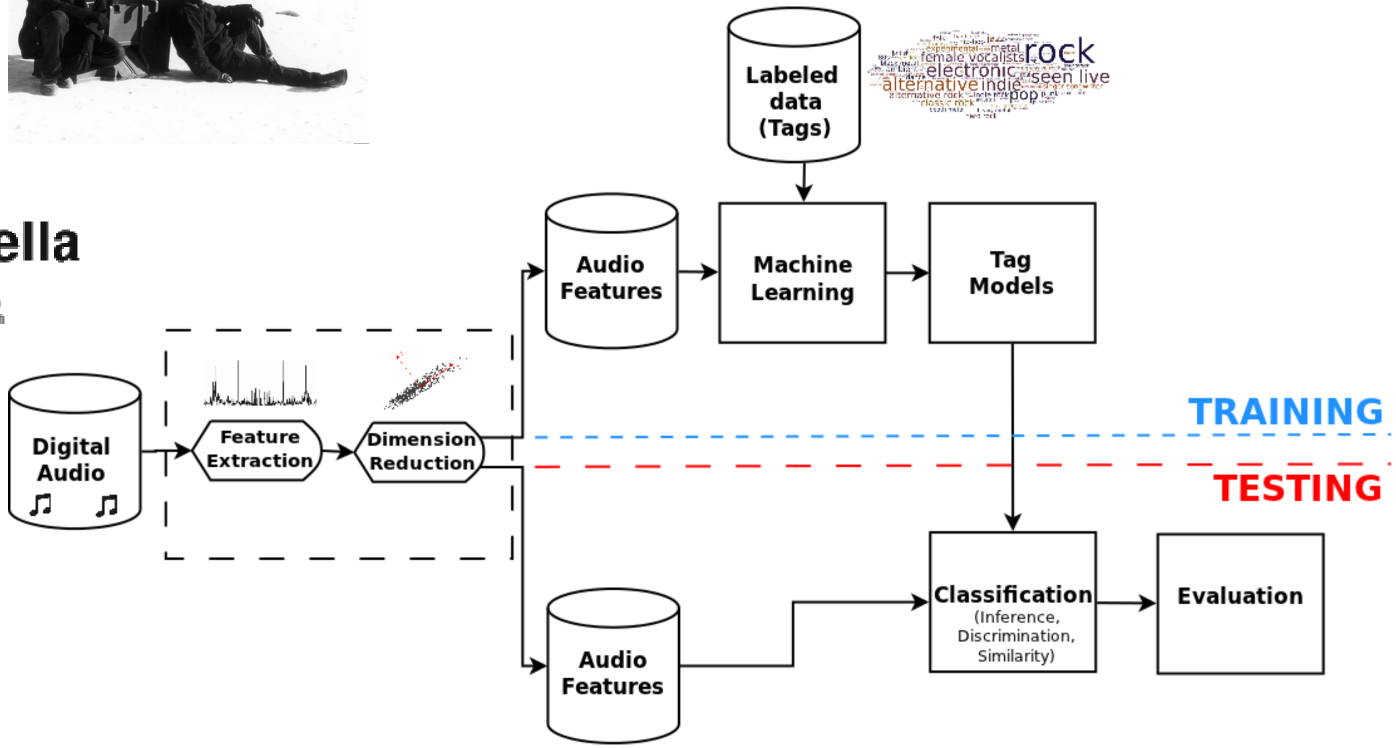
Semantic Music Discovery (Turnbull et al.; SIGIR 2007, 2009):
Combines timbre, harmony, Web texts, and Web tags to predict user labels

Semantic Annotation of Music Collections (Sordo; 2012)
Propagation of tags through audio similarity

Auto-Tagging/Retrieval by Tag

Learning indexing labels from content features

(Sordo; 2012)



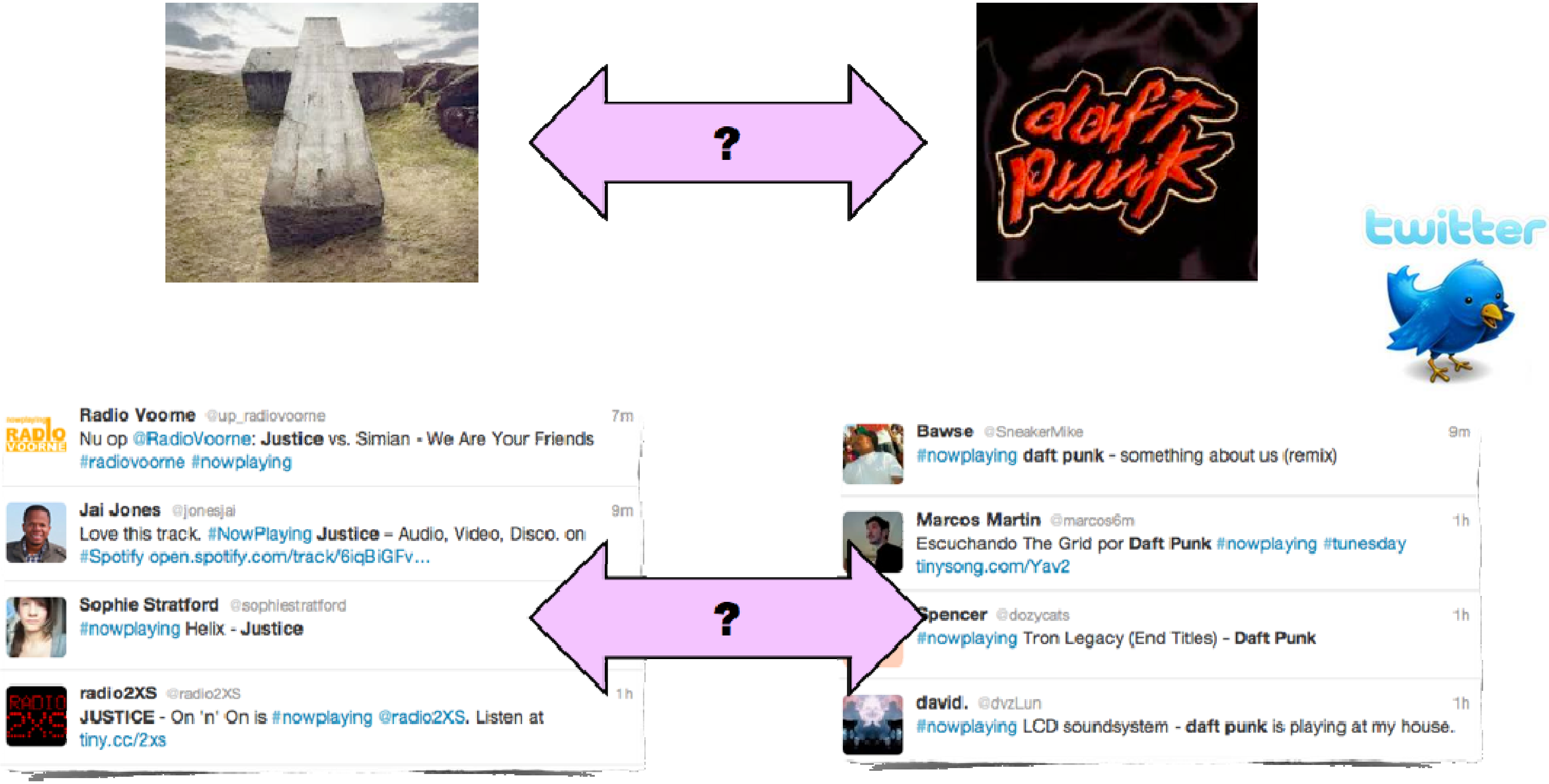
Music Information Extraction from Web Pages

Web data is a rich source for all types of meta-data and semantic relations

Methods from NLP, IE, Named Entity Detection for data extraction

- **Genres, Moods, Similarities** using Rule Patterns (Geleijnse, Korst; 2006)
- **Band Members and Line-Up** using Rule Patterns (Schedl, Widmer; 2007)
- **Band Members, Discography, Artist Detection (**rule based) (Krenmair; 2010)
- **Band Members, Discography** using Supervised Learning (Knees, Schedl; 2011)
- **Album cover** detection and extraction (Schedl et al., ECIR 2006)

Microblogs as Text Sources



Extracting and Indexing Tweets on Music Listening

(Schedl, ECIR 2013)

(a) Filter Twitter stream (#nowplaying, #itunes, #np, …)
(b) Multi-level, rule-based analysis (artists/songs) to find relevant tweets (MusicBrainz)
(c) Last.fm, Freebase, Allmusic, Yahoo! PlaceFinder to annotate tweets

MusicBrainz

„Alice Cooper“
„BB King“
„Prince“
„Metallica“
…

{"id_str":"142338125895696385","place":null,"text":"**#NowPlaying Christmas Tree-Lady Gaga**","in_reply_to_user_id":null,"favorited":false,"geo":null,"retweet_count":0,"in_reply_to_screen_name":null,"in_reply_to_status_id_str":null,"source":"web","retweeted":false,"in_reply_to_user_id_str":null,"coordinates":null,"created_at":"Thu Dec 01 20:23:48 +0000 2011","in_reply_to_status_id":null,"contributors":null,"user":{"id_str":"20209983","profile_link_color":"2caba5","screen_name":"tamse77","follow_request_sent":null,"geo_enabled":false,"favourites_count":26,"location":"Maryland ","following":null,"verified":false,"profile_background_color":"e80e0e","show_all_inline_media":true,"profile_background_tile":true,"followers_count":309,"profile_image_url":"http:\/\/a1.twimg.com\/profile_images\/1647613274\/392960_10150559294659517_793614516_11700077_1689597400_n_normal.jpg","description":"being awesome since 1990. ","is_translator":false,"profile_background_image_url_https":"https:\/\/si0.twimg.com\/profile_background_images\/359728130\/frames.gif","friends_count":148,"profile_sidebar_fill_color":"ffffff","default_profile":false,"listed_count":3,"time_zone":"Central Time (US & Canada)","contributors_enabled":false,"created_at":"Fri Feb 06 01:51:10 +0000 2009","profile_sidebar_border_color":"f5f8ff","protected":false,"notifications":null,"profile_use_background_image":true,"name":"Katie","default_profile_image":false,"statuses_count":22172,"profile_text_color":"615d61","url":null,"profile_image_url_https":"https:\/\/si0.twimg.com\/profile_images\/1647613274\/392960_10150559294659517_793614516_11700077_1689597400_n_normal.jpg","id":20209983,"lang":"en","profile_background_image_url":"http:\/\/a2.twimg.com\/profile_background_images\/359728130\/frames.gif","utc_offset":-21600},"truncated":false,"id":142338125895696385,"entities":{"hashtags":[{"text":"NowPlaying","indices":[0,11]}],"urls":[],"user_mentions":[]}}

Department of Computational Perception

Extracting and Indexing Tweets on Music Listening

(Schedl, ECIR 2013)

{"id_str":"142338125895696385","place":null,"text":"#NowPlaying Christmas Tree-Lady Gaga","in_reply_to_user_id":null,"favorited":false,"geo":null,"retweet_count":0,"in_reply_to_screen_name":null,"in_reply_to_status_id_str":null,"source":"web","retweeted":false,"in_reply_to_user_id_str":null,"coordinates":null,"created_at":"Thu Dec 01 20:23:48 +0000 2011","in_reply_to_status_id":null,"contributors":null,"user":{"id_str":"20209983","profile_link_color":"2caba5","screen_name":"tamse77","follow_request_sent":null,"geo_enabled":false,"favourites_count":26,"location":"Maryland ","following":null,"verified":false,"profile_background_color":"e80e0e","show_all_inline_media":true,"profile_background_tile":true,"followers_count":309,"profile_image_url":"http:\/\/a1.twimg.com\/profile_images\/1647613274\/392960_10150559294659517_793614516_11700077_1689597400_n_normal.jpg","description":"being awesome since 1990. ","is_translator":false,"profile_background_image_url_https":"https:\/\/si0.twimg.com\/profile_background_images\/359728130\/frames.gif","friends_count":148,"profile_sidebar_fill_color":"ffffff","default_profile":false,"listed_count":3,"time_zone":"Central Time (US & Canada)","contributors_enabled":false,"created_at":"Fri Feb 06 01:51:10 +0000 2009","profile_sidebar_border_color":"f5f8ff","protected":false,"notifications":null,"profile_use_background_image":true,"name":"Katie","default_profile_image":false,"statuses_count":22172,"profile_text_color":"615d61","url":null,"profile_image_url_https":"https:\/\/si0.twimg.com\/profile_images\/1647613274\/392960_10150559294659517_793614516_11700077_1689597400_n_normal.jpg","id":20209983,"lang":"en","profile_background_image_url":"http:\/\/a2.twimg.com\/profile_background_images\/359728130\/frames.gif","utc_offset":-21600},"truncated":false,"id":142338125895696385,"entities":{"hashtags":[{"text":"NowPlaying","indices":[0,11]}],"urls":[],"user_mentions":[]}}

134243700380401664	127821914	11	2	106.83	-6.23	1	1	202085	3529910	0	1	...
134243869201154048	174194590	11	2	-0.142	51.52	2	2	330061	5762915	1	0	...

twitter-id user-id month weekday longitude latitude country-id city-id artist-id track-id <tag-ids>

Extracting and Indexing Tweets on Music Listening

(Schedl, ECIR 2013)

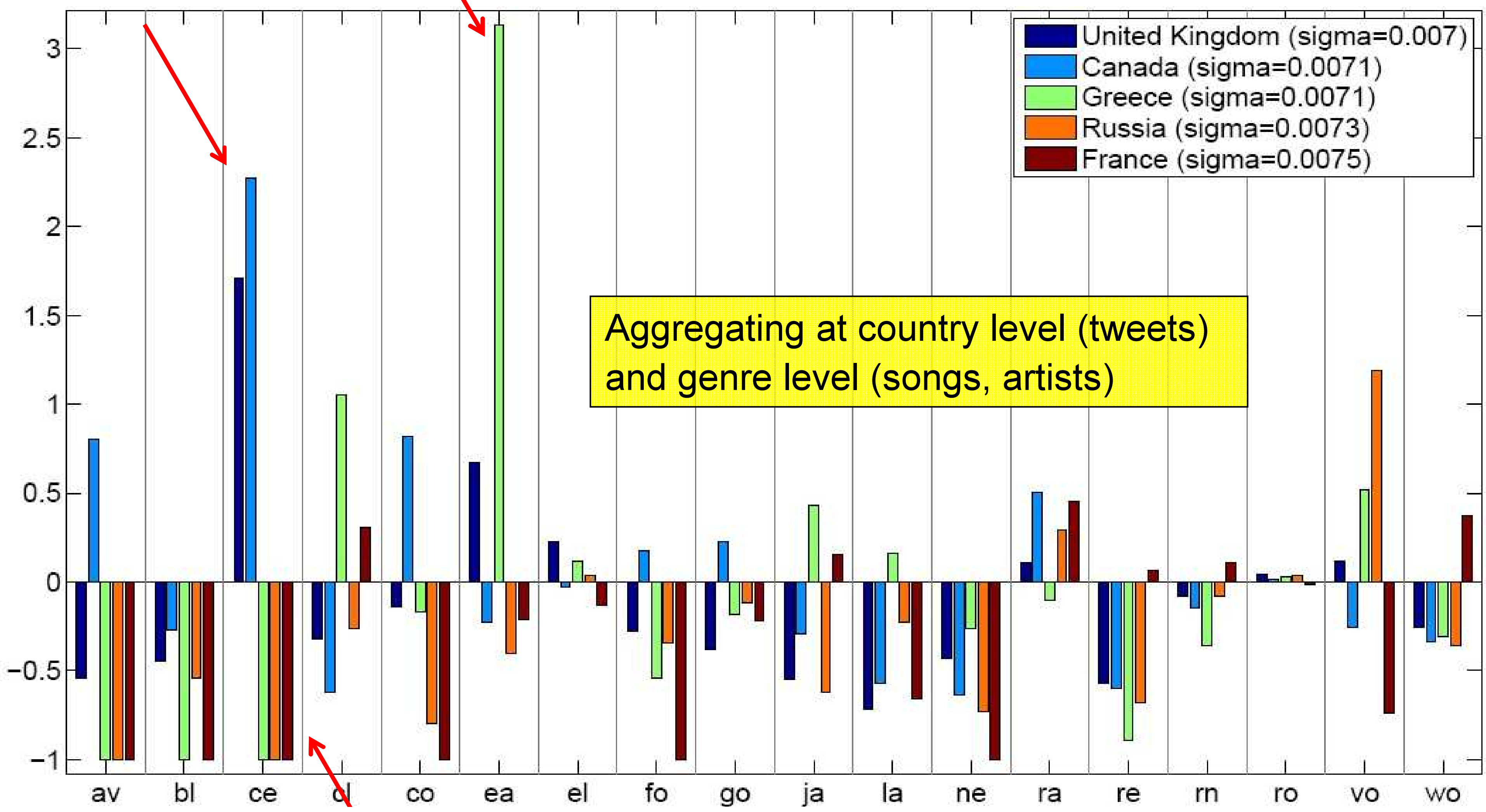
#nowplaying		#itunes	
country	tweets	country	tweets
Brazil	725,389	USA	78,460
USA	673,839	Japan	30,932
Japan	458,558	Mexico	23,047
Mexico	419,584	Brazil	16,390
Indonesia	284,082	UK	15,134
South Korea	251,132	Canada	11,266
China	183,178	South Korea	8,652
UK	128,744	Australia	5,119
Netherlands	121,134	China	4,492
Venezuela	110,336	Germany	3,157

most active countries

Department of Computational Perception

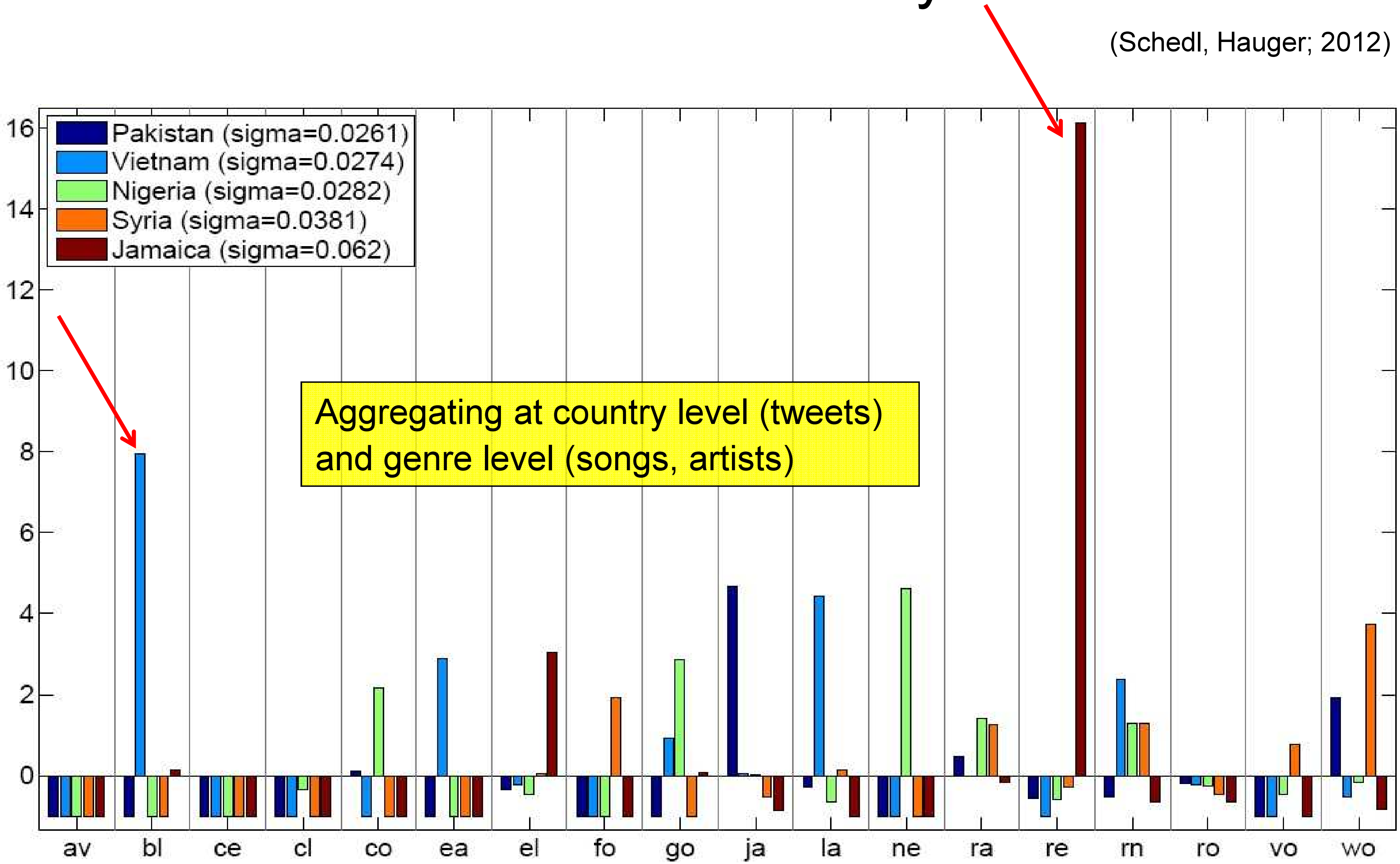
Geospatial Music Taste Analysis: Most Mainstreamy

(Schedl, Hauger; 2012)



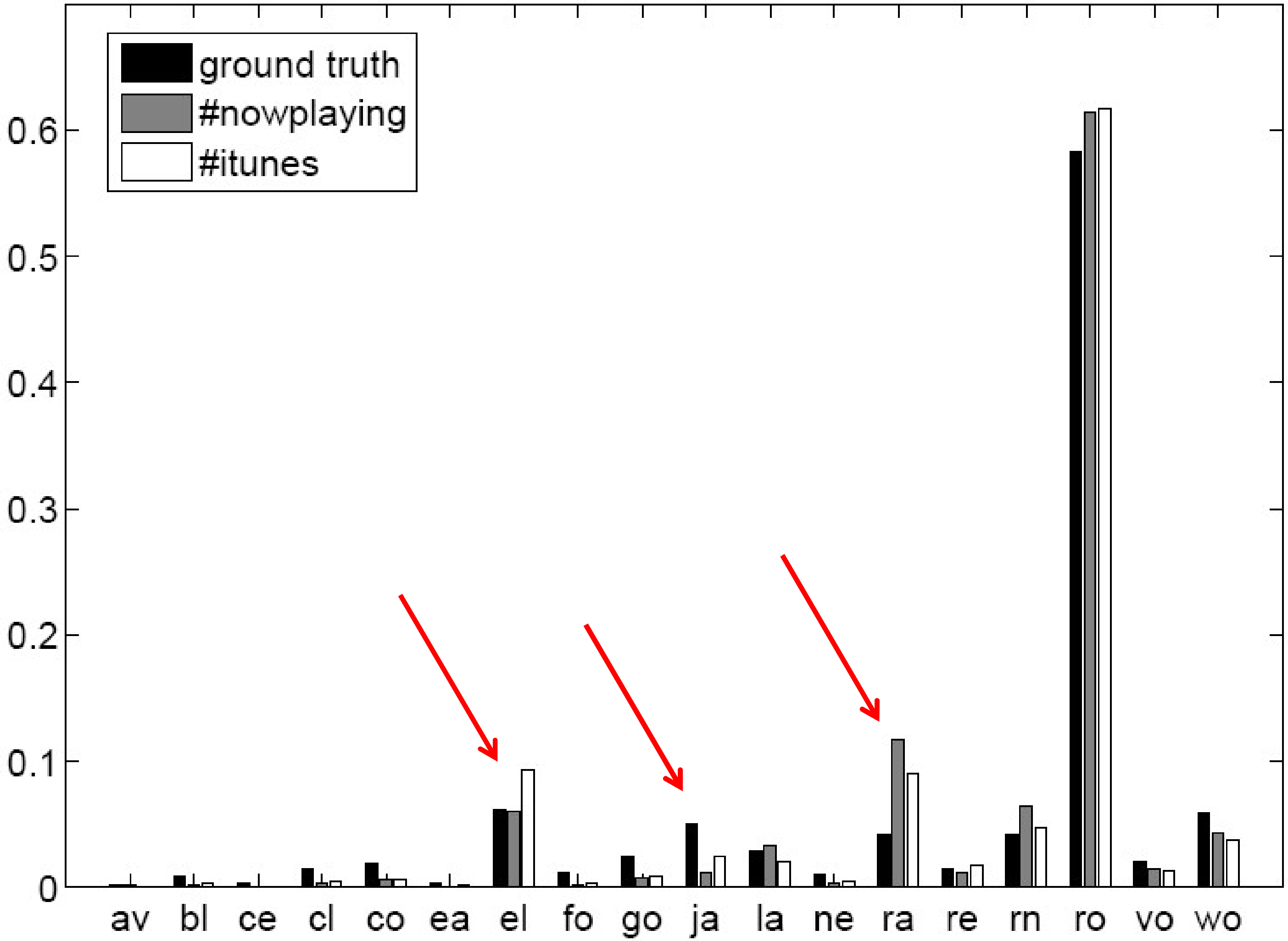
Geospatial Music Taste Analysis: Least Mainstreamy

(Schedl, Hauger; 2012)



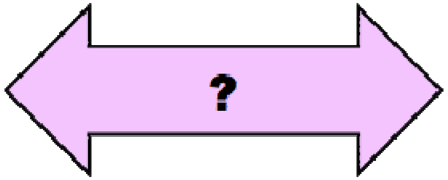
Geospatial Music Taste Analysis: Usage of Specific Products

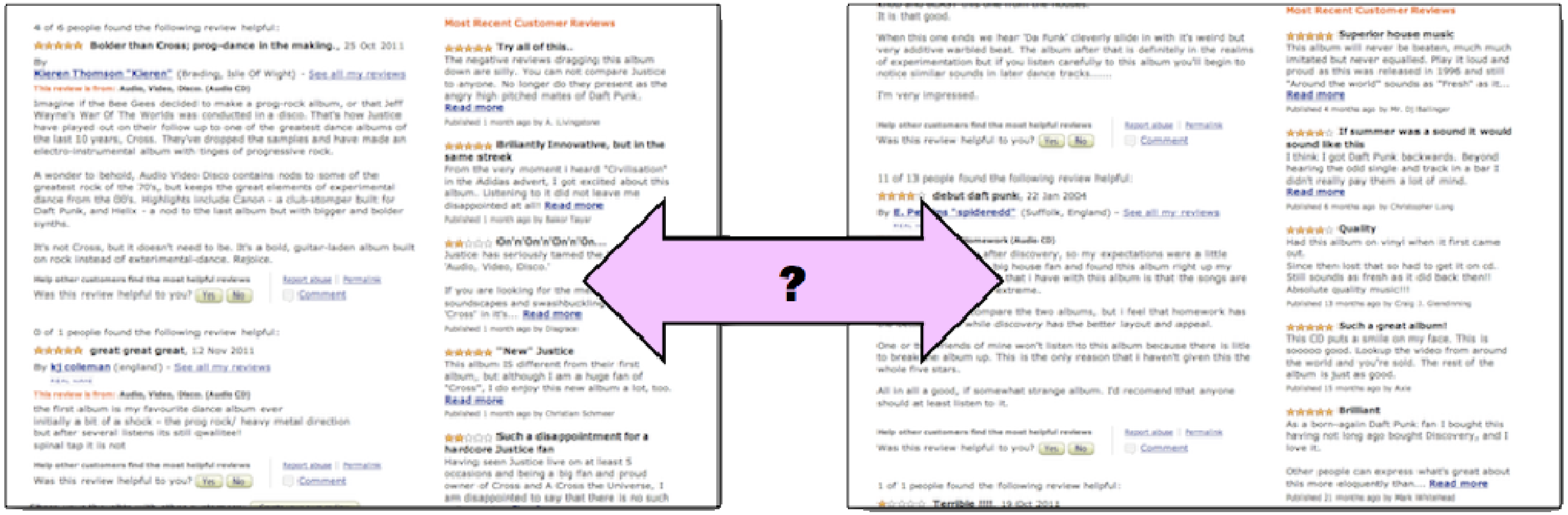
(Schedl, Hauger; 2012)



Department of Computational Perception

Product Reviews as Text Sources





Department of Computational Perception

Product Reviews as Text Sources

Exploiting sources such as *Amazon.com* or *Epinions.com* (Hu et al.; 2005)

This review is from: Ray of Light (Audio CD)

This is Madonna's work of art. And this CD is the very best collection of any music she has ever produced since "Erotica." Madonna's lyrics are beautiful and strong because even after 9 years it still stands the test of time. It's completely impossible for this CD to be dated; with the electronica feel to it and fast moving dance numbers, such as the title-track this CD was way ahead of its time. Even in the double-00's "Ray of Light" is still very important as both a dance record and a record of reflection and interpersonal renewal.

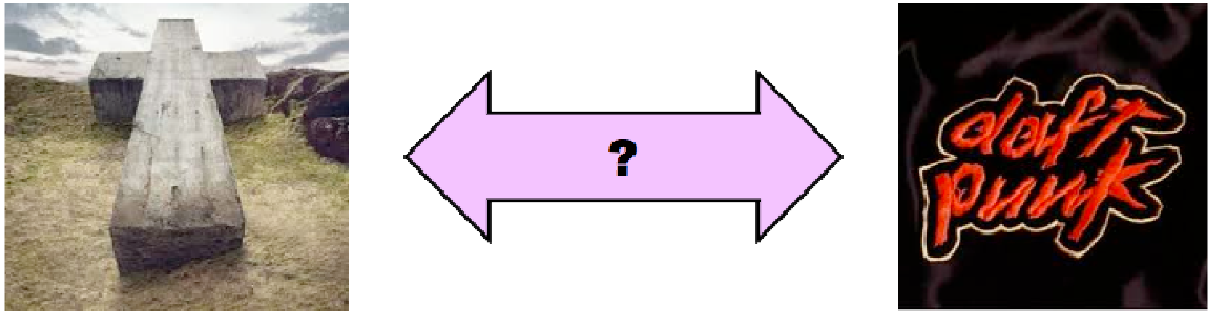
This review is from: **Never Gonna Give You Up (MP3 Download)**

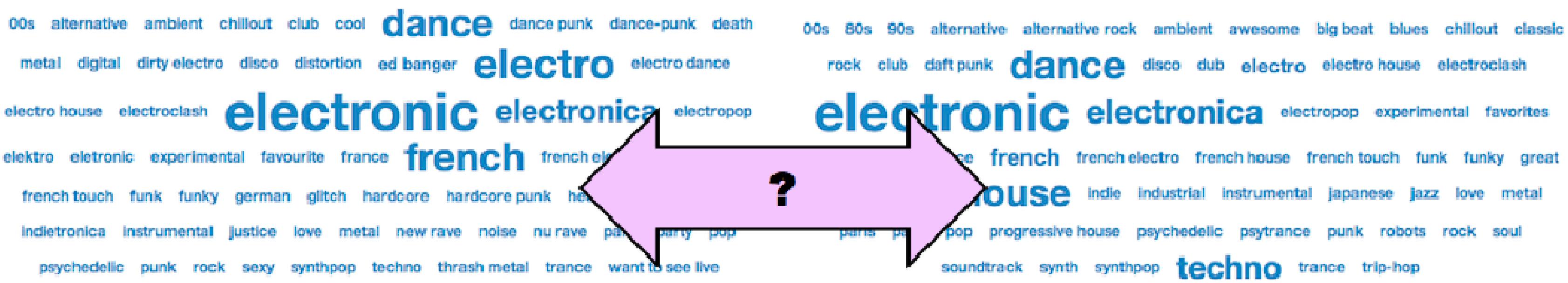
This is truly Astley's greatest opus.
The track is flawless. It is instantly accessible, but features many hidden layers and pleasures that cannot be discovered upon the first listen alone. With this and all of his other fantastic work, it's no wonder that Radiohead calls Astley their "greatest inspiration."

Allows for sentiment analysis and associated rating prediction

Very prone to attacks (remedy: consider “helpfulness” ratings)

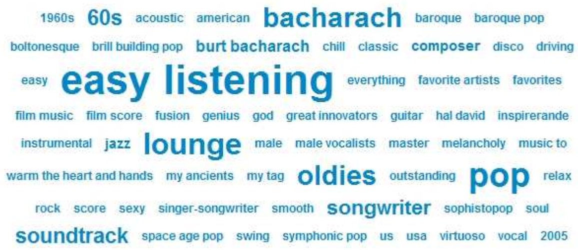
Community Tags as Text Sources



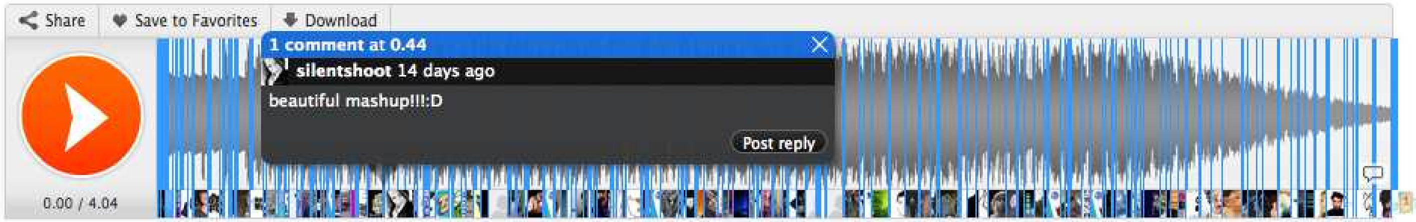


Tag Sources

- **Community**
  e.g., Last.fm



  e.g., Soundcloud (annotations along timeline)



- **Games with a purpose** (GWAP)
  e.g., Tag-a-Tune
  (Law, von Ahn; 2009)

- **Autotags** (see before)

Department of Computational Perception

Community Tags as Text Sources

Treating collections of tags (e.g., from Last.fm) as documents

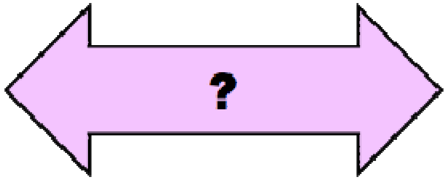
(Pohle et al.; 2007) (Levy, Sandler; 2008) (Hu et al.; 2009)

- Retrieve tags for artist or track from Last.fm
- Cleaning of noisy and redundant tags:
  manually or automatically (Geleijnse et al.; 2007)
- List of collected terms is treated as text document and TF·IDF'd
  (Levy, Sandler; 2007)
- Optionally, LSA to reduce dimensionality
- Comparison of vectors via cosine similarity (or overlap score)

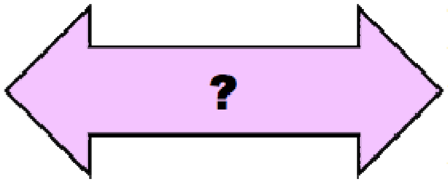
- Data often available in standardized fashion, dedicated terms for music
- Lower dimensionality
  e.g., 13,500 tags vs. >200,000 Web terms (Levy, Sandler; 2007)
- Depends on community, needs annotators
- Hacking and Attacks!

Department of Computational Perception

Lyrics as Text Source



Before day break there was none
And as it broke there was one
The Moon, the sun, it goes on 'n' on
The winter battle was won
The summer children were born
And so the story goes on 'n' on
Come woman if your life beats
Those we buried with the house keys
Smoke and feather where the fields are green
From here to eternity
Come woman on your own time



Around the world, around the world
Around the world, around the world
Around the world, around the world

Around the world, around the world
Around the world, around the world
Around the world, around the world

Around the world, around the world
Around the world, around the world
Around the world, around the world

Department of Computational Perception

Lyrics as Text Source

Topic Features (Logan et al.; 2004)

- Typical topics for lyrics are distilled from a large corpus using (P)LSA ("Hate", "Love", "Blue", "Gangsta", "Spanish")
- Lyrics are transformed to topic-based vectors, similarity is calculated via L_1 distance

- Alternative approaches use TF·IDF with optional LSA and Stemming for **Mood Categorization** (Laurier et al.;2009) (Hu et al.; 2009)

Rhyme Features (Mayer et al.; 2008) (Hirjee, Brown; 2009)

- Phonetic transcription is searched for patterns of rhyming lines (AA, ABAB, AABB)
- Frequency of patterns + statistics like *words per minute, punctuation freq.* etc.

Other Features (Mahedero et al.; 2005) (Hirjee, Brown; 2009)

- Language, structure

Department of Computational Perception

Text-based Similarity Approaches: Summary

	Web-Terms	Microblogs	Reviews	Tags	Lyrics
Source	Web pages	platform	shops, platform	Web service	portal
Community-based	depends	depends	yes	yes	no
Level	artists	artists (tracks)	albums	artists (tracks)	tracks (artists)
Feature Dimensionality	very high	high	possibly high	moderate	possibly high
Specific Bias	low	low	personal	community	none
Potential Noise	high	high	low	moderate	low

Department of Computational Perception

Similarity from Co-Occurrences

Idea: expect entities that occur frequently in the same context to be similar

Data sources considered:

- Page count estimates from Web search engines
- Shared folders/search queries on the *Gnutella* file sharing network
- Collaborative filtering on playcounts from *Last.fm*
- Occurrences in playlists

Department of Computational Perception

Search Engine Page Count Estimates

(Schedl et al.; 2005)

For all pairs of artists: query **`“artist 1” “artist 2” +music +review`**
For each artist: query **`“artist” +music +review`**

Use **page counts** for sim. (results in quadratic page count matrix)

$$sim_{pc\_cp}(A_i, A_j) = \frac{1}{2} \cdot \left(\frac{pc(A_i, A_j)}{pc(A_i)} + \frac{pc(A_i, A_j)}{pc(A_j)} \right)$$

To avoid quadratic number of queries: download top 100 pages for each artist and parse for occurrences of other artists (linear complexity)

NB: asymmetry of *pc* matrix can be used to identify prototypical artists!

Department of Computational Perception

Shared Folders in a P2P Network

Make use of meta-data transmitted as files names or ID3 tags in P2P network OpenNap (Whitman, Lawrence; 2002) (Ellis et al.; 2002)

Information gathered from users' shared folders (no file downloads!)

Similarities via artist co-occurrences in collections (cond. prob.)

Sparse co-occurrence matrix

Experiments on Gnutella network (Shavitt, Weinsberg; 2009):

- meta-data highly inconsistent
- can be used as song-based similarity measure and to estimate localized popularity/trends (matching IP addresses difficult!)

Department of Computational Perception

Last.fm Playcounts

Use *explicit* or *implicit* ratings of users or interpret number of plays of a song as a “rating”

Results in a user-track rating matrix

Use standard **collaborative filtering** approaches to predict similarities (or to recommend unknown music)
e.g., (Resnick et al.; 1994)

Item-based: compare tracks by calculating similarity on vectors over all users

User-based: find similar users by comparing listening pattern vectors; use to find relevant/similar tracks yet unknown to user

Department of Computational Perception

Playlist Co-Occurrences

Analysis of co-occurrences of artists and songs on radio station playlists and compilation CD databases (CDDB) (Pachet et al.;2001)

$$sim_{pl\_cooc}(A_i, A_j) = \frac{1}{2} \cdot \left[\frac{cooc(A_i, A_j)}{cooc(A_i, A_i)} + \frac{cooc(A_j, A_i)}{cooc(A_j, A_j)} \right]$$

Analysis of 29K playlists from "Art of the Mix" (Cano, Koppenberger;2004): artists similar if they co-occur in playlist (highly sparse)

Analysis of >1M playlists from "MusicStrands" (Baccigalupo et al.; 2008):

- distance in playlists taken into account $\beta_0 = 1, \beta_1 = 0.8, \beta_2 = 0.64$

$$dist_{pl\_d}(A_i, A_j) = \sum_{h=0}^{2} \beta_h \cdot [d_h(A_i, A_j) + d_h(A_j, A_i)]$$

- playlist prediction using case-based reasoning

Department of Computational Perception

Co-occurrence-based Approaches: Summary

	Web Co-Ocs	Playcounts	P2P nets	Playlists
Source	search engines, Web pages	listening service	shared folders	radio, compilations, Web services
Community-based	no	yes	yes	depends on source
Level	artists	tracks	artists (tracks)	artists (tracks)
Specific Bias	"wikipedia"-bias	popularity	community	low
Potential Noise	high	low	high	low

Department of Computational Perception

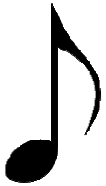
Part IV

PERSONALIZATION AND USER ADAPTATION

Department of Computational Perception

Computational Factors Influencing Music Perception and Similarity

(Schedl et al., JIIS 2013)



music content

Examples:
- rhythm
- timbre
- melody
- harmony
- loudness

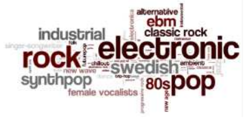
Examples:
- mood
- activities
- social context
- spatio-temporal context
- physiological aspects

user context

music perception and similarity

Examples:
- semantic labels
- song lyrics
- album cover artwork
- artist's background
- music video clips

music context



Examples:
- music preferences
- musical training
- musical experience
- demographics

user properties

Department of Computational Perception

Geospatial Music Recommendation

(Schedl, Schnitzer; SIGIR 2013)

- Combining music content + music context features
  - audio features: PS09 award-winning feature extractors (rhythm and timbre)
  - text/web: TFIDF-weighted artist profiles from artist-related web pages

- Using collection of geo-located music tweets (cf. (Schedl; ECIR 2013))

- Aims:

  (i) determining ideal combination of music content and –context

  (ii) ameliorate music recommendation by user's location information

Department of Computational Perception

Ideal combination of music content and context

(Schedl, Schnitzer; SIGIR 2013)

ξ	$K=1$	$K=3$	$K=5$
web only – 0.00	**.5829**	**.5753**	**.5774**
.05	.6421	.6280	.6257
.15	.6432	**.6286**	**.6261**
.25	**.6433**	.6275	.6258
.35	.6430	.6275	.6257
.45	.6408	.6266	.6252
.55	.6394	.6259	.6244
.65	.6379	.6255	.6232
.75	.6368	.6234	.6221
.85	.6330	.6202	.6188
.95	.6215	.6083	.6059
audio only – 1.00	**.5436**	**.5302**	**.5247**

Adding user context (different approaches)

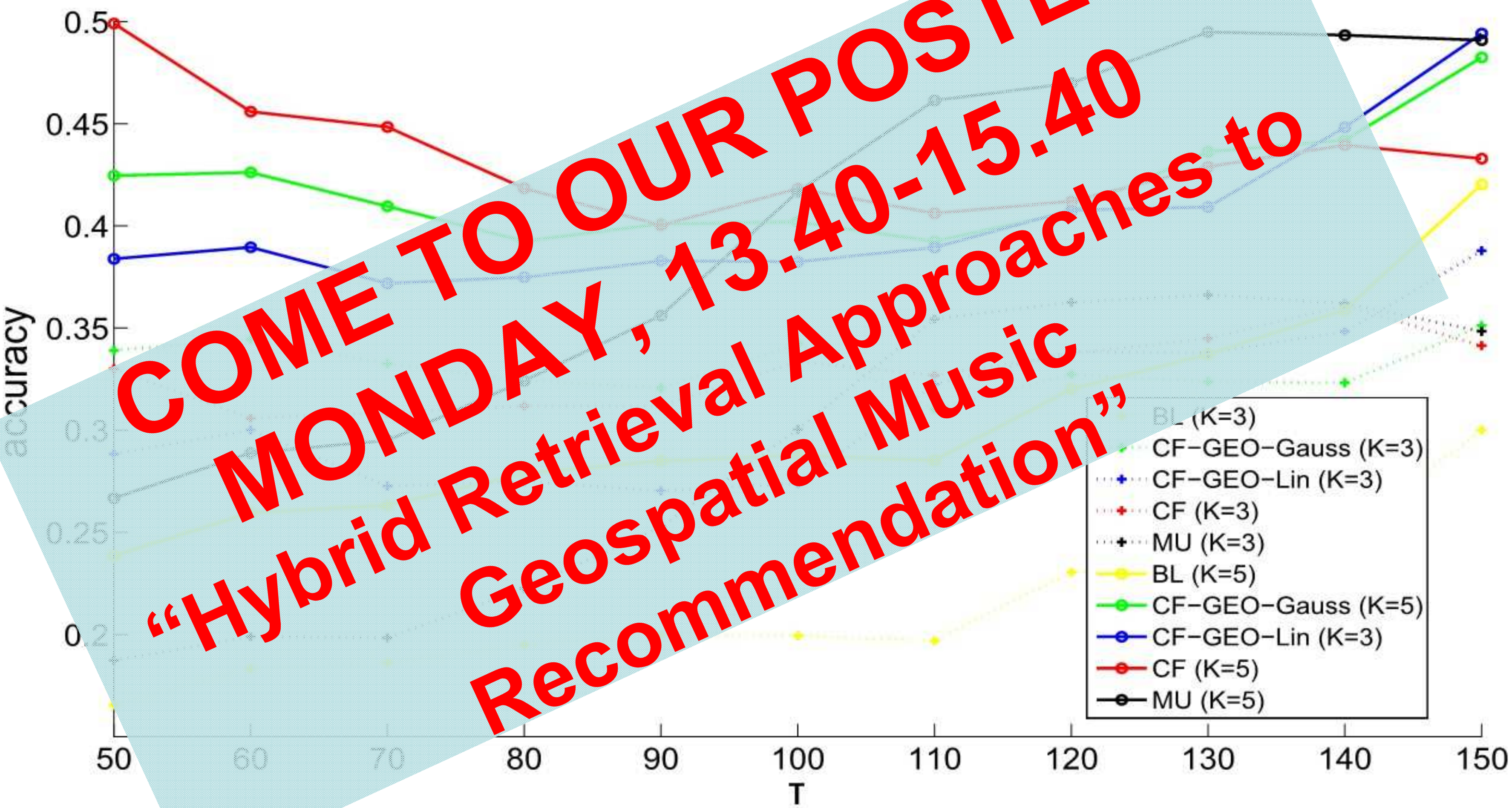
(Schedl, Schnitzer; SIGIR 2013)

Abbreviation	Description
BL	random baseline
MU	hybrid music model
CF	collaborative filtering model
CF-GEO-Lin	CF model: geospatial user weighting using linear spatial distances
CF-GEO-Gauss	CF model: geospatial user weighting weighting using a Gauss kernel

Department of Computational Perception

(Schedl, Schnitzer; SIGIR 2013)

Evaluation Results



COME TO OUR POSTER MONDAY, 13.40-15.40

“Hybrid Retrieval Approaches to Geospatial Music Recommendation”

T: minimum number of distinct artists a users must have listened to to be included

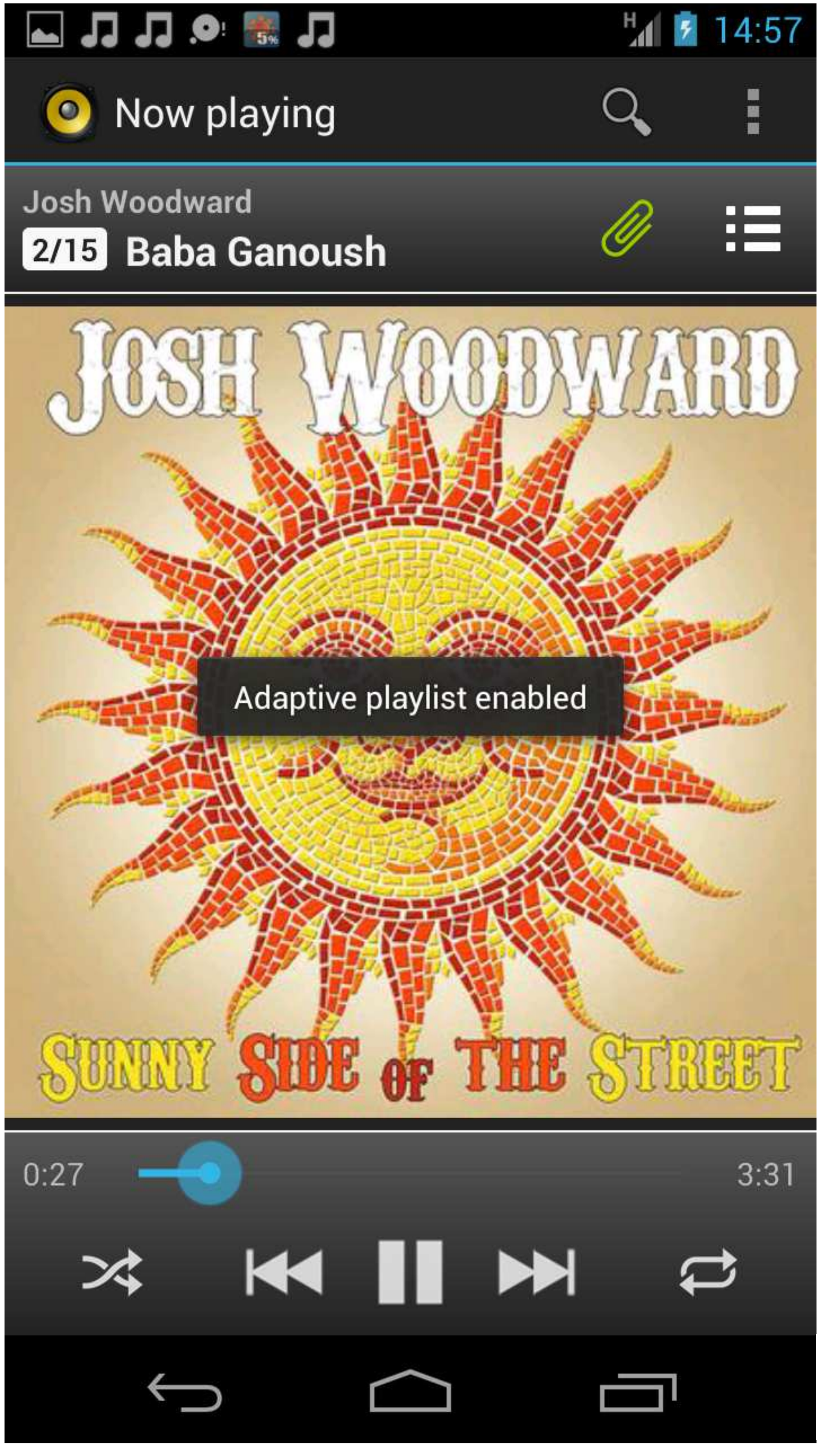
Department of Computational Perception

User-Aware Music Recommendation on Smart Phones

"Mobile Music Genius": music player for the Android platform

- collecting user context data while playing
- adaptive system that learns user taste/preferences from implicit feedback (player interaction: play, skip, duration played, playlists, etc.)
- ultimate aim: dynamically and seamlessly update the user's playlist according to his/her current context

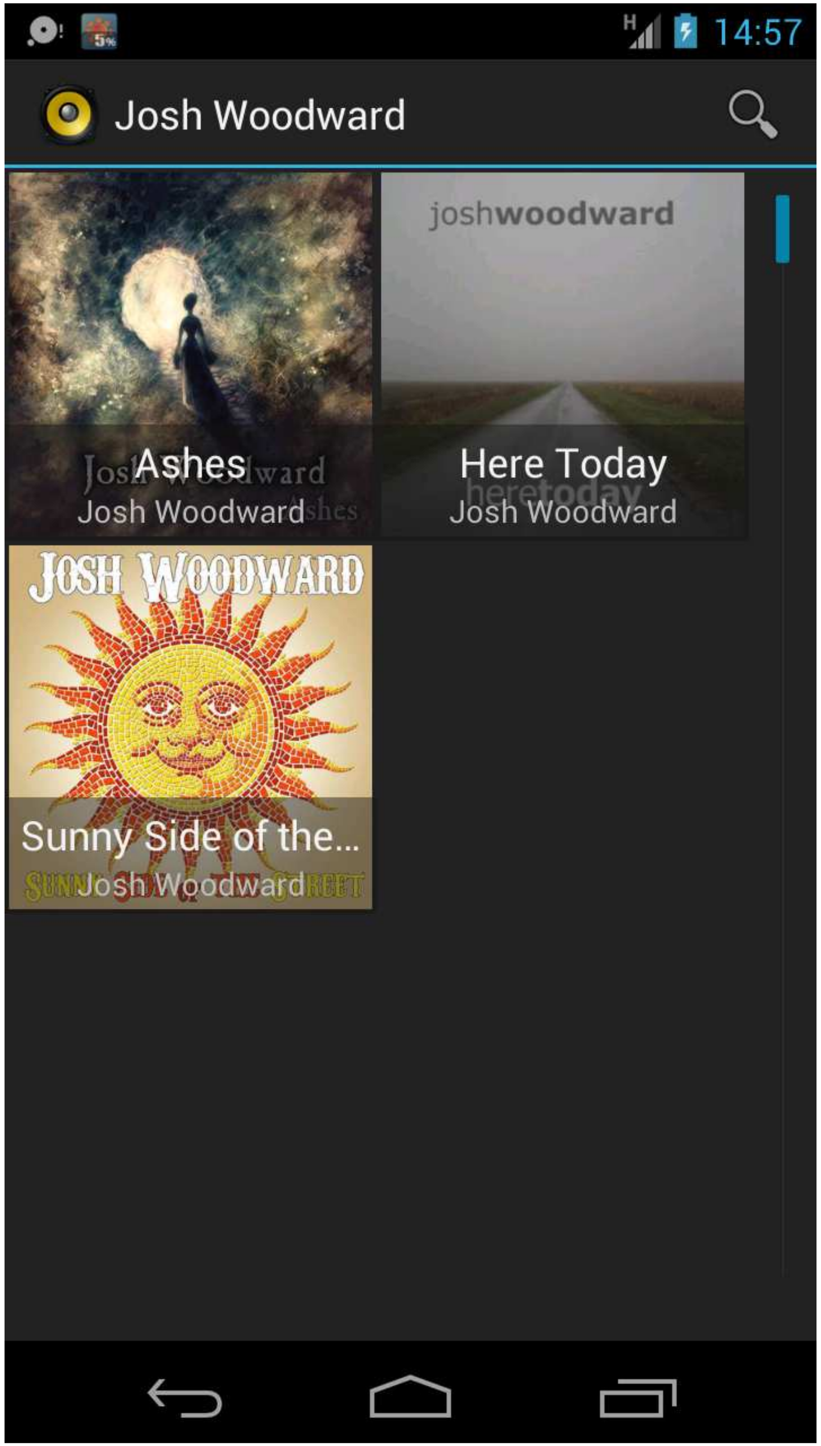
Department of Computational Perception



Mobile Music Genius

Music player in adaptive playlist generation mode

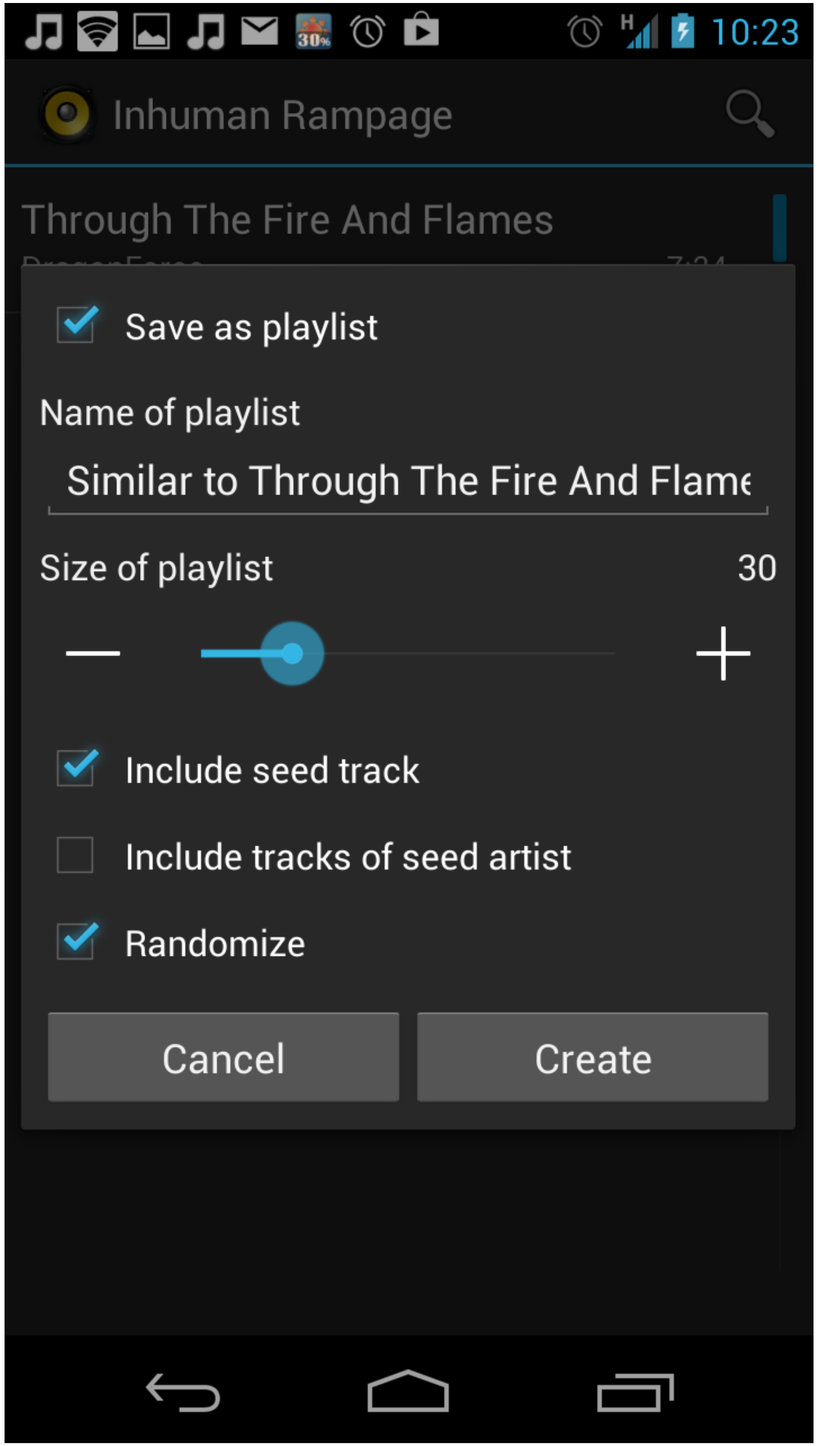
Department of Computational Perception



Mobile Music Genius

Album browser
in cover view

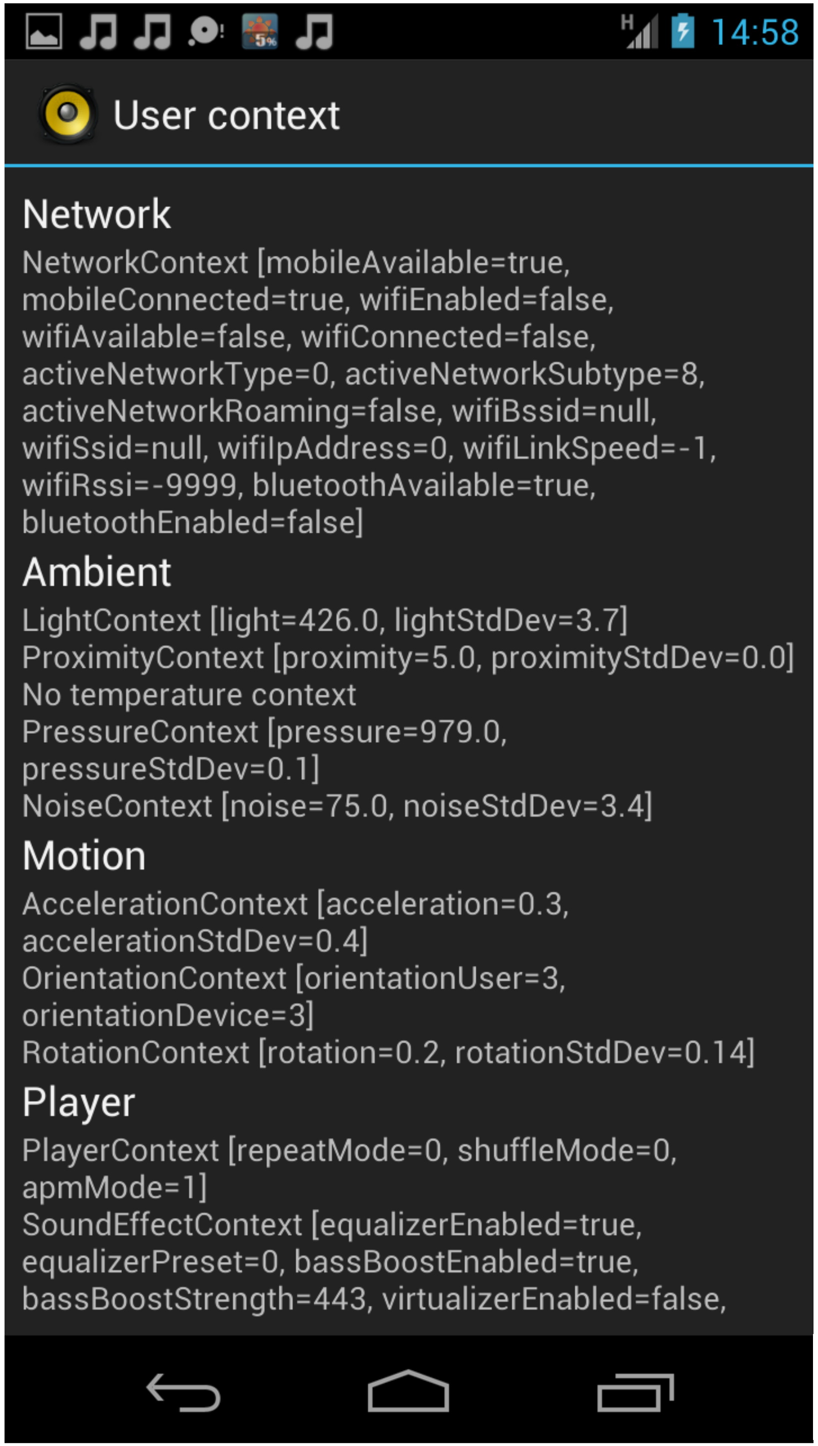
Department of
Computational
Perception



Mobile Music Genius

Automatic playlist generation based on music context (features and similarity computed based on Last.fm tags)

Department of Computational Perception



Mobile Music Genius

Some user context features gathered while playing

Department of Computational Perception

User Context Features from Android Phones

Time: timestamp, time zone

Personal: userID/eMail, gender, birthdate

Device: devideID (IMEI), sw version, manufacturer, model, phone state, connectivity, storage, battery, various volume settings (media, music, ringer, system, voice)

Location: longitude/latitude, accuracy, speed, altitude

Place: nearby place name (populated), most relevant city

Weather: wind direction, speed, clouds, temperature, dew point, humidity, air pressure

Ambient: light, proximity, temperature, pressure, noise, digital environment (WiFi and BT network information)

Activity: acceleration, user and device orientation, screen on/off, running apps

Player: artist, album, track name, track id, track length, genre, plackback position, playlist name, playlist type, player state (repeat, shuffle mode), audio output (headset plugged)

mood and activity (direct user feedback)

Department of Computational Perception

Evaluation: ongoing

- collected user context data from JKU students over a period of 2 months
- about 8,000 listening data items and corresponding user context gathered

To be analyzed:

(i) Which granularity/abstraction level to choose for representation/learning?

(ii) Which user context features are the most discriminative to predict music preference?

First results for predicting class "artist":

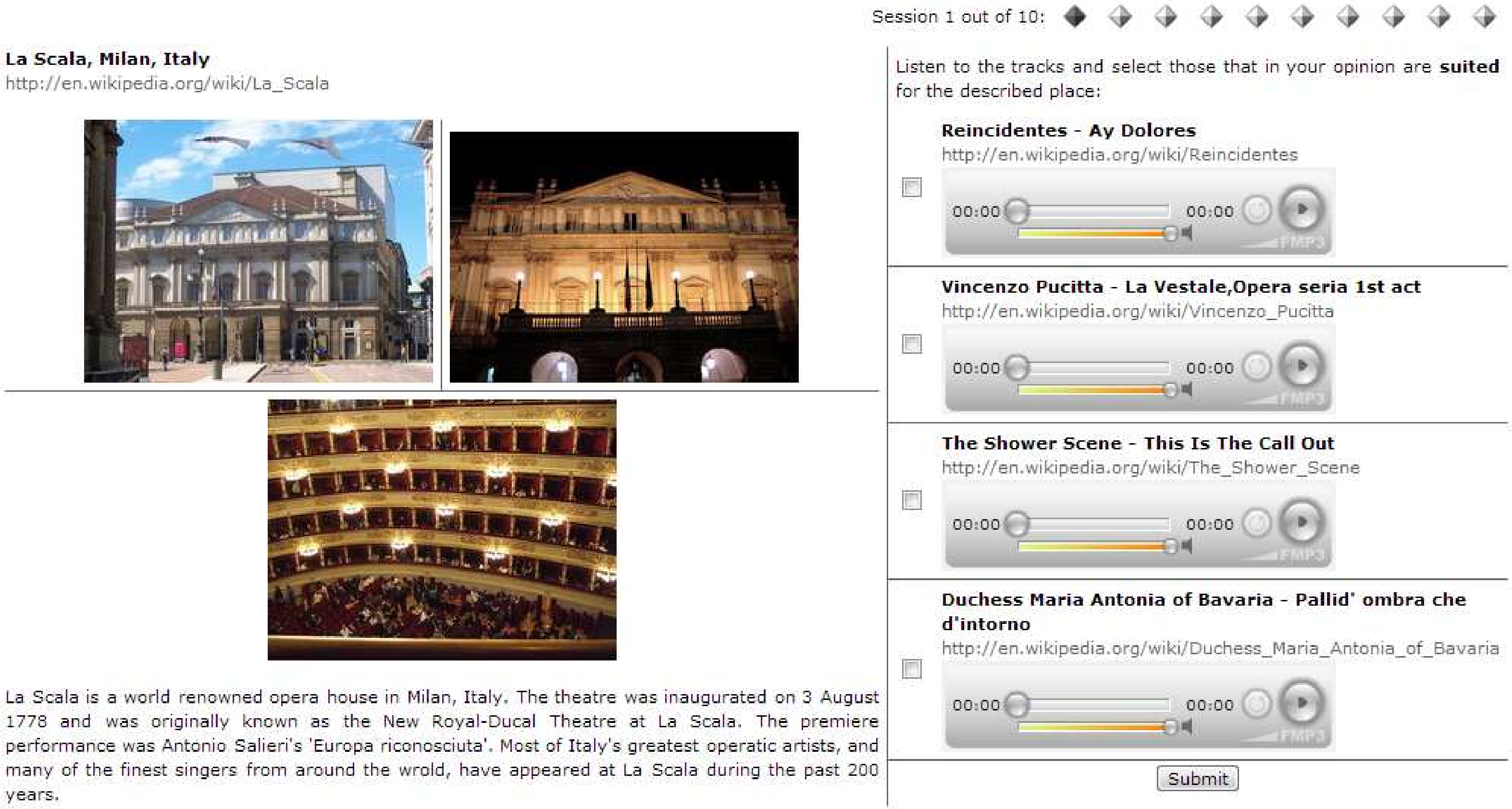
ZeroR (baseline) classifier	15% accuracy
k-nearest neighbors	42% accuracy
JRip rule learner	51% accuracy
J48 decision tree	55% accuracy

Department of Computational Perception

Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

recommend music that is suited to a place of interest (POI) of the user (context-aware)



Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Approaches:

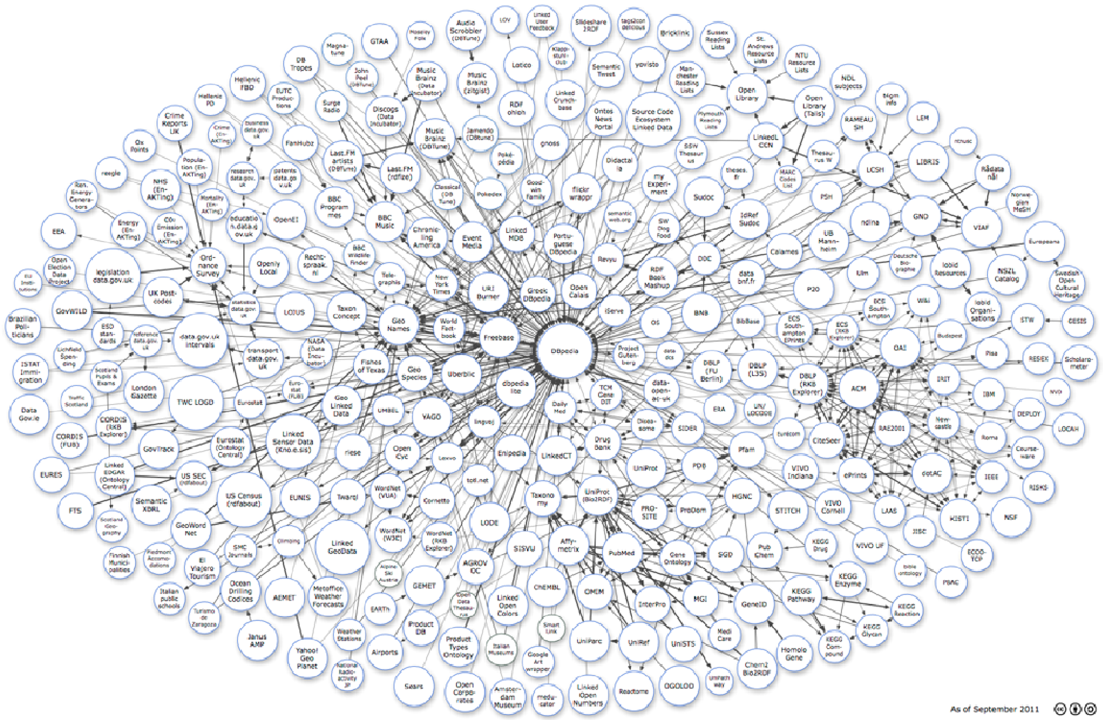
- *genre-based*: only play music belonging to the user's preferred genres (baseline)

Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Approaches:

- *knowledge-based*: use the DBpedia knowledge base (relations between POIs and musicians)

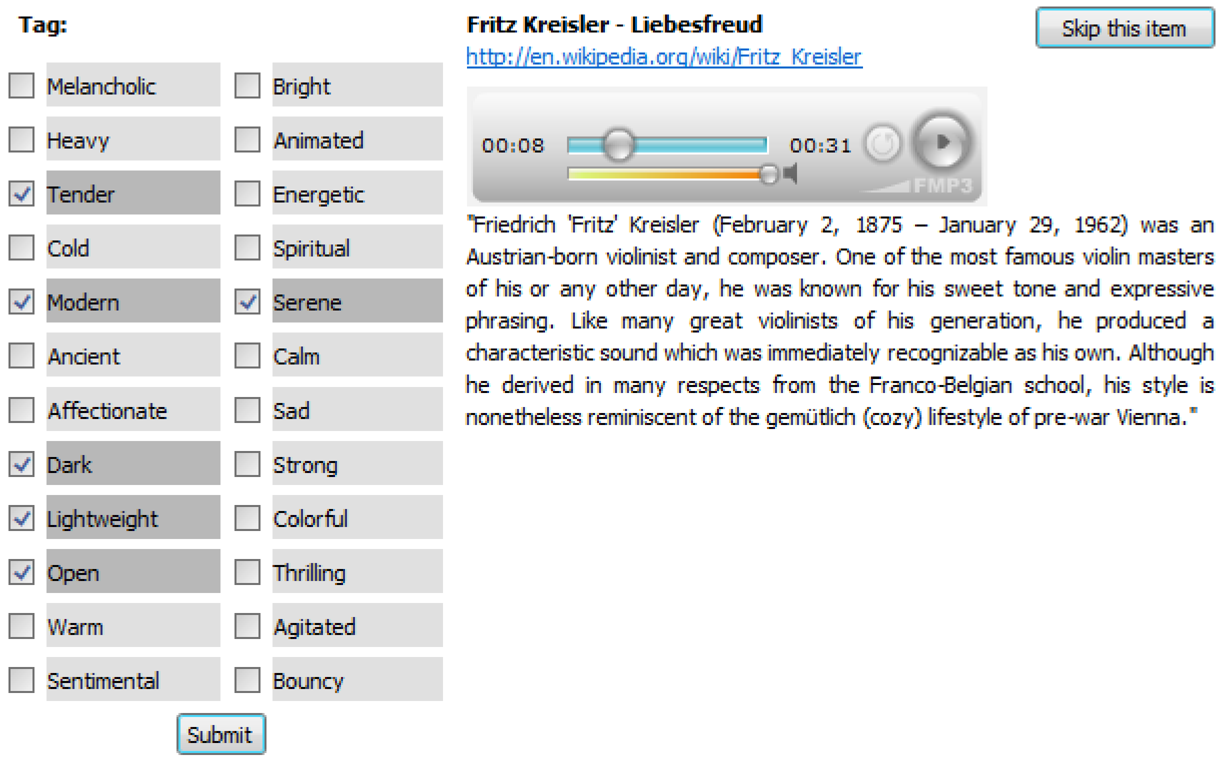


Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Approaches:

- *tag-based*: user-assigned emotion tags describing images of POIs and music, Jaccard similarity between music-tag-vectors and POI-tag-vectors

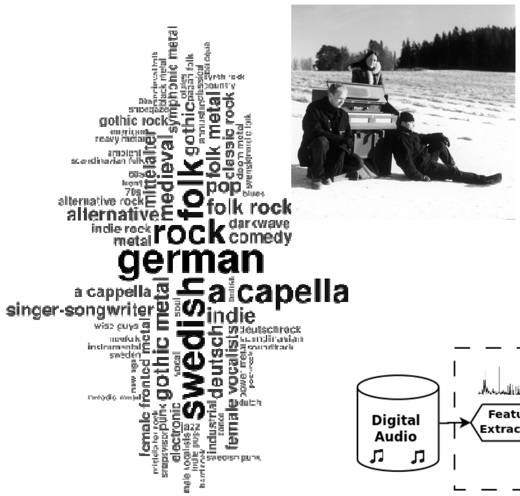


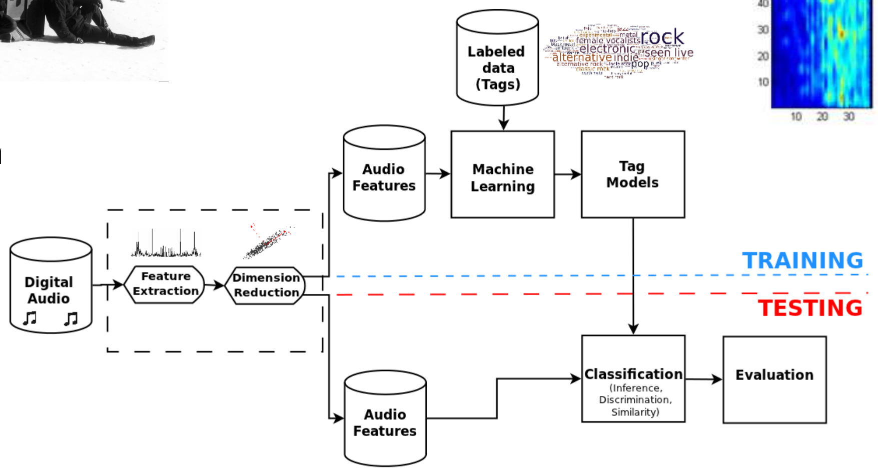
Matching Places of Interest and Music

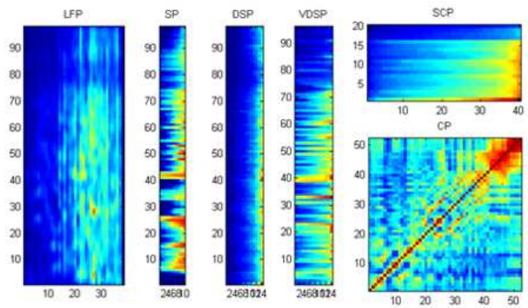
(Kaminskas et al.; RecSys 2013)

Approaches:

- *auto-tag-based*: use state-of-the-art music auto-tagger based on the Block-level Feature framework to automatically label music pieces; then again compute Jaccard similarity between music-tag-vectors and POI-tag-vectors







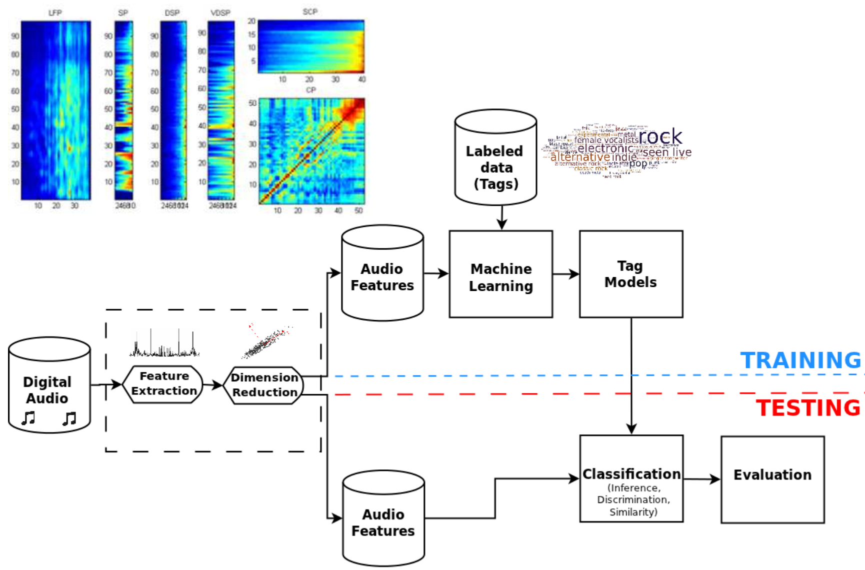
Department of Computational Perception

Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Approaches:

- *combined*: aggregate music recommendations w.r.t. ranks given by knowledge-based and auto-tag-based approaches



Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Approaches:

- *genre-based*: only play music belonging to the user's preferred genres (baseline)
- *knowledge-based*: using the DBpedia knowledge base (relations between POIs and musicians)
- *tag-based*: user-assigned emotion tags describing images of POIs and music, Jaccard similarity between music-tag-vectors and POI-tag-vectors
- *auto-tag-based*: using state-of-the-art music auto-tagger based on the Block-level Feature Framework to automatically label music pieces; then again use Jaccard similarity between music-tag-vectors and POI-tag-vectors
- *combined*: aggregate music recommendations w.r.t. ranks given by knowledge-based and auto-tag-based approaches

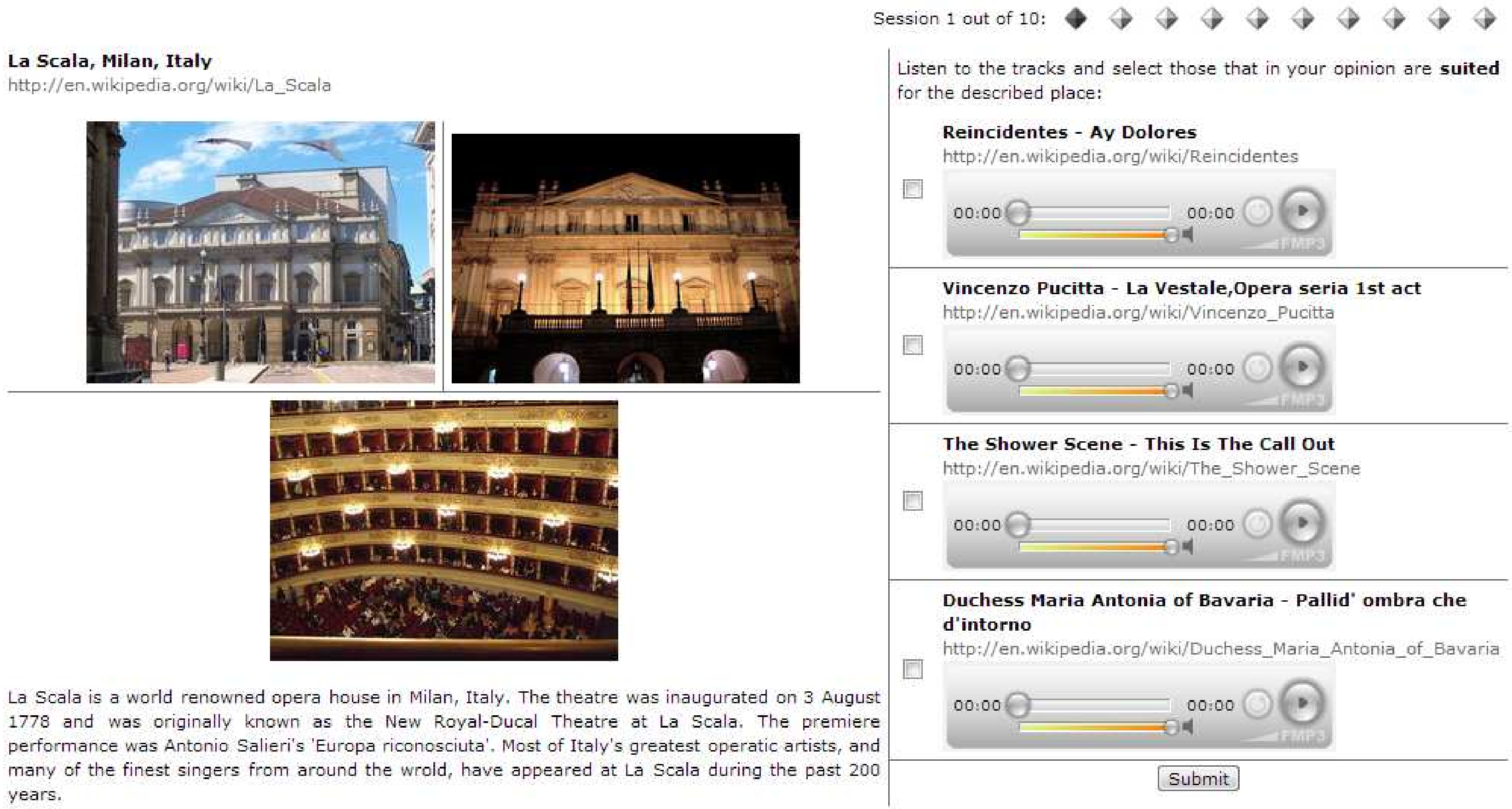
Department of Computational Perception

Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Evaluation:

- user study via web interface (58 users, 564 sessions)

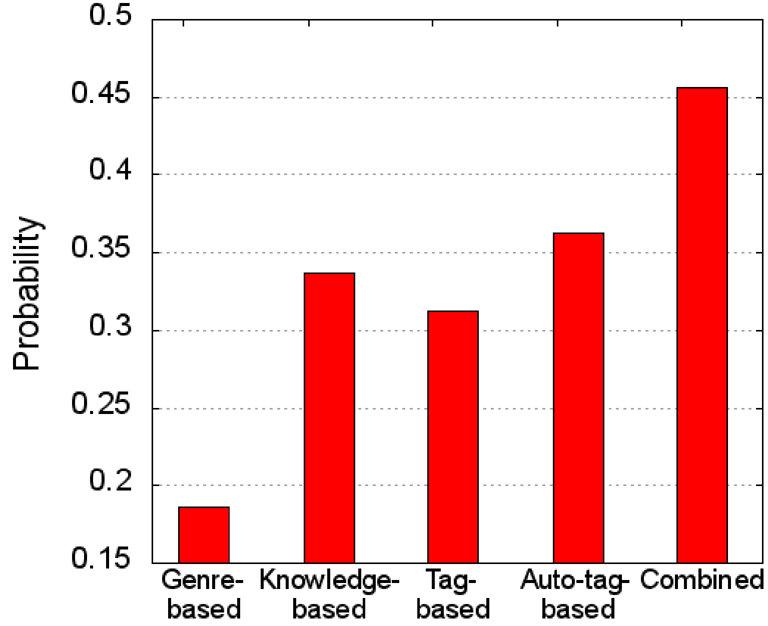


Matching Places of Interest and Music

(Kaminskas et al.; RecSys 2013)

Evaluation:

- Performance measure: number of times a track produced by each approach was considered as well-suited in relation to total number of evaluation sessions, i.e. probability that a track marked as well-suited by a user was recommended by each approach



Department of Computational Perception

SUMMARY

Music Information Retrieval is a great field

Various approaches to extract information from the audio signal

Various sources and approaches to extract contextual data and similarity information from the Web

Multi-modal modeling and retrieval is important and allows for exciting applications

Next big challenges:

- modeling user properties and context
- personalization
- situation-based retrieval
- new and better suited evaluation strategies

REFERENCES

(Aucouturier, Pachet; 2004) **Improving Timbre Similarity: How High's the Sky?**, Journal of Negative Results in Speech and Audio Sciences, 1(1), 2004.

(Baccigalupo, Plaza, Donaldson; 2008) **Uncovering Affinity of Artists to Multiple Genres from Social Behaviour Data**, Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008), Philadelphia, September 2008.

(Baumann, Hummel; 2003) **Using Cultural Metadata for Artist Recommendation**, Proceedings of the 3rd International Conference on Web Delivering of Music (WEDELMUSIC 2003), Leeds, UK, September 2003

(Bertin-Mahieux, Eck, Maillet, Lamere; 2008) **Autotagger: A Model for Predicting Social Tags From Acoustic Features on Large Music Databases**, Journal of New Music Research, 37(2):115-135, 2008.

(Cano, Koppenberger; 2004) **The Emergence of Complex Network Patterns in Music Artist Networks**, Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004), Barcelona, Spain, October 2004.

(Casey, Veltkamp, Goto, Leman, Rhodes, Slaney; 2008) **Content-based music information retrieval: Current directions and future challenges**, Proceedings of the IEEE 96 (4), 668-696, 2008.

(Celino, Turati, Della Valle, Cerizza; 2006) **Squiggle: a Semantic Search Engine at Work**, Proceedings of the 1st International Conference on Semantics And digital Media Technology (SAMT 2006), Athens, Greece, December 2006

(Celma; 2010) **Music Recommendation and Discovery in the Long Tail – The Long Tail, Long Fail, and Long Play in the Digital Music Space**, Springer, 2010.

(Celma, Cano, Herrera; 2006) **SearchSounds: An Audio Crawler Focused on Weblogs**, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006), Victoria, Canada, October 2006.

(Cooley, Tukey; 1965) **An Algorithm for the Machine Calculation of Complex Fourier Series**, Mathematics of Computation, 19(90): 297-301, 1965.

(Dror, Koenigstein, Koren, Weimer; 2012) **The Yahoo! Music Dataset and KDD-Cup’11**, Journal of Machine Learning Research, 18:3-18, 2012.

(Ellis, Whitman, Berenzweig, Lawrence; 2002) **The Quest For Ground Truth in Musical Artist Similarity**, Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002), Paris, France, October 2002.

(Flexer, Schnitzer, Gasser, Pohle; 2010) **Combining Features Reduces Hubness in Audio Similarity**, Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010), Utrecht, the Netherlands , August 2010.

(Geleijnse, Korst; 2006) **Web-based Artist Categorization**, Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006), Victoria, Canada, October 2006.

(Geleijnse, Schedl, Knees; 2007) **The Quest for Ground Truth in Musical Artist Tagging in the Social Web Era**, Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 2007.

(Hauger, Schedl; 2012) **Exploring Geospatial Music Listening Patterns in Microblog Data**, Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR 2012), Copenhagen, Denmark, October 2012.

(Hirjee, Brown; 2009) **Automatic Detection of Internal and Imperfect Rhymes in Rap Lyrics**, Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), Kobe, Japan, October 2009.

(Hu, Downie, Ehmann; 2009) **Lyric Text Mining in Music Mood Classification**, Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009), Kobe, Japan, October 2009

(Hu, Downie, West, Ehmann; 2005) **Mining Music Reviews: Promising Preliminary Results**, Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), London, UK, September 2005

(Kaminskas, Ricci, Schedl; 2013) **Location-aware Music Recommendation Using Auto-Tagging and Hybrid Matching**, Proceedings of the 7th ACM Conference on Recommender Systems (RecSys 2013), Hong Kong, China, October 2013.

(Knees, Pampalk, Widmer; 2004) **Artist Classification with Web-based Data**, Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004), Barcelona, Spain, October 2004.

(Knees, Pohle, Schedl, Schnitzer, Seyerlehner; 2008) **A Document-centered Approach to a Natural Language Music Search Engine**, Proceedings of the 30th European Conference on Information Retrieval (ECIR), Glasgow, Scotland, UK, 2008.

(Knees, Pohle, Schedl, Widmer; 2007) **A Music Search Engine Built upon Audio-based and Web-based Similarity Measures**, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), Amsterdam, the Netherlands, July 2007.

(Knees, Schedl; 2011) **Towards Semantic Music Information Extraction from the Web Using Rule Patterns and Supervised Learning**, Proceedings of the 2nd Workshop on Music Recommendation and Discovery (WOMRAD), Chicago, USA, October 2011.

(Knees, Schedl; 2014) **Music Similarity and Retrieval**, expected to be published by Springer in 2014.

(Knees, Schedl, Pohle, Widmer; 2006) **An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web**, Proceedings of the 14th ACM International Conference on Multimedia (MM'06), Santa Barbara, California, USA, 2006.

(Krenmair; 2010) **Musikspezifische Informationsextraktion aus Webdokumenten**, Master‘s thesis, Johannes Kepler University Linz, Austria, 2010.

(Law, von Ahn; 2009) **Input-Agreement: A New Mechanism for Collecting Data Using Human Computation Games**, Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), Boston, MA, USA, April 2009.

(Levy, Sandler; 2007) **A Semantic Space for Music Derived from Social Tags**, Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 2007

(Levy, Sandler; 2008) **Latent Semantic Models for Music**, Journal of New Music Research, 37(2):137-150, 2008.

(Lippens, Martens, Leman, Baets, Meyer, Tzanetakis; 2004) **A Comparison of Human and Automatic Musical Genre Classification**, Proceedings of the IEEE International Conference on Audio, Speech and Signal Processing (ICASSP), Montreal, Canada, 2004

(Logan, Kositsky, Moreno; 2004) **Semantic Analysis of Song Lyrics**, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2004), Taipei, Taiwan, June 2004.

(Mahedero, Martínez, Cano, Koppenberger, Gouyon; 2005) **Natural Language Processing of Lyrics**, Proceedings of the 13th ACM International Conference on Multimedia, Singapore, November 2005.

(Mayer, Neumayer, Rauber; 2008) **Combination of Audio and Lyrics Features for Genre Classification in Digital Audio Collections**, Proceedings of the ACM Multimedia, Vancouver, Canada; 2008.

(McFee, Bertin-Mahieux, Ellis, Lanckriet; 2012) **The Million Song Dataset Challenge**, Proceedings of 4th International Workshop on Advances in Music Information Research (AdMIRe 2012), Lyon, France, April 2012.

(Pachet, Westerman, Laigre; 2001) **Musical Data Mining for Electronic Music Distribution**, Proceedings of the 1st International Conference on Web Delivering of Music (WEDELMUSIC 2001), Florence, Italy, November 2001.

(Pampalk; 2001) **Islands of Music - Analysis, Organization, and Visualization of Music Archives**, Master's Thesis, TU Vienna, 2001.

(Pampalk, Goto; 2007) **MusicSun: A New Approach to Artist Recommendation**, Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 2007

(Pohle, Knees, Schedl, Widmer; 2007) **Meaningfully Browsing Music Services**, Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 2007.

(Resnick, Iacovou, Suchak, Bergstrom, Riedl; 1994) **GroupLens: An Open Architecture for Collaborative Filtering of Netnews**. Proceedings of Conference on Computer Supported Coorperative Work (CSCW), Chapel Hill, 1994.

(Schedl; 2011a) **Analyzing the Potential of Microblogs for Spatio-Temporal Popularity Estimation of Music Artists**, IJCAI 2011: International Workshop on Social Web Mining, Barcelona, Spain, July 2011.

(Schedl; 2011b) **Web- and Community-based Music Information Extraction**, In *Music Data Mining*, CRC Press/Chapman Hall, July 2011.

(Schedl; 2012a) **#nowplaying Madonna: A Large-Scale Evaluation on Estimating Similarities Between Music Artists and Between Movies from Microblogs**, Information Retrieval, 15(3-4), 2012.

(Schedl; 2012b) **Exploiting Social Media for Music Information Retrieval**, In Naeem Ramzan, Roelof van Zwol, Jong-Seok Lee, Kai Clüver, Xian-Sheng Hua (eds.), *Social Media Retrieval*, Springer, December 2012

(Schedl; 2013) **Leveraging Microblogs for Spatiotemporal Music Information Retrieval**, Proceedings of the 35th European Conference on Information Retrieval (ECIR 2013), Moscow, Russia, March 2013.

(Schedl, Flexer; 2012) **Putting the User in the Center of Music Information Retrieval**, Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012), Porto, Portugal, Oct 2012

(Schedl, Flexer, Urbano; 2013) **The Neglected User in Music Information Retrieval Research**, Journal of Intelligent Information Systems, 2013.

(Schedl, Hauger; 2012) **Mining Microblogs to Infer Music Artist Similarity and Cultural Listening Patterns**, Proceedings of the 21st International World Wide Web Conference (WWW 2012): 4th International Workshop on Advances in Music Information Research (AdMIRe 2012), Lyon, France, April 2012.

(Schedl, Hauger, Schnitzer; 2012) **A Model for Serendipitous Music Retrieval**, IUI 2012: 2nd International Workshop on Context-awareness in Retrieval and Recommendation (CaRR 2012), Lisbon, Portugal, 2012.

(Schedl, Knees; 2011) **Personalization in Multimodal Music Retrieval**, Proceedings of the 9th Workshop on Adaptive Multimedia Retrieval (AMR 2011), Barcelona, Spain, July 2011.

(Schedl, Knees, Pohle, Widmer; 2006) **Towards Automatic Retrieval of Album Covers**, Proceedings of the 28th European Conference on Information Retrieval (ECIR 2006), London, UK, April 2006.

(Schedl, Knees, Widmer; 2005) **A Web-Based Approach to Assessing Artist Similarity using Co-Occurrences**, Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing, Riga, Latvia, 2005.

(Schedl, Pohle, Knees, Widmer; 2011) **Exploring the Music Similarity Space on the Web**, ACM Transactions on Information Systems, 29(3):14, July 2011.

(Schedl, Pohle, Koenigstein, Knees; 2010) **What's Hot? Estimating Country-Specific Artist Popularity**, 11th International Society for Music Information Retrieval Conference (ISMIR), Utrecht, the Netherlands, 2010.

(Schedl, Schnitzer; 2013) **Hybrid Retrieval Approaches to Geospatial Music Recommendation**, Proceedings of the 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2013), Dublin, Ireland, July-August 2013.

(Schedl, Sordo, Koenigstein, Weinsberg; 2013) **Mining User-generated Data for Music Information Retrieval,** In Marie-Francine Moens, Juanzi Li, Tat-Seng Chua (eds.), *Mining of User Generated Content and Its Applications,* CRC Press, 2013.

(Schedl, Widmer; 2007) **Automatically Detecting Members and Instrumentation of Music Bands via Web Content Mining**, Proceedings of the 5th Workshop on Adaptive Multimedia Retrieval (AMR 2007), Paris, France, 2007.

(Schnitzer, Flexer, Schedl, Widmer; 2012) **Local and Global Scaling Reduce Hubs in Space**, Journal of Machine Learning Research, 13:2871-2902, October 2012.

(Schnitzer, Pohle, Knees, Widmer; 2007) **One-Touch Access to Music on Mobile Devices**, Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia (MUM 2007), Oulu, Finland, December 2007.

(Shavitt, Weinsberg; 2009) **Songs Clustering Using Peer-to-Peer Co-occurrences**, Proceedings of the IEEE International Symposium on Multimedia (ISM2009): International Workshop on Advances in Music Information Research (AdMIRe 2009), San Diego, USA, December 2009.

(Seyerlehner; 2010) **Content-Based Music Recommender Systems: Beyond Simple Frame-Level Audio Similarity**, PhD Thesis, Johannes Kepler University Linz, December, 2010.

(Seyerlehner, Widmer, Knees; 2010) **A Comparison of Human, Automatic and Collaborative Music Genre Classification and User Centric Evaluation of Genre Classification Systems**, Proceedings of the 8th International Workshop on Adaptive Multimedia Retreival (AMR), Linz, Austria, 2010.

(Seyerlehner, Widmer, Schedl, Knees; 2010) **Automatic Music Tag Classification based on Block-Level Features**, Proceedings of the 7th Sound and Music Computing Conference (SMC), Barcelona, Spain, July 2010.

(Sordo; 2012) **Semantic Annotation of Music Collections: A Computational Approach**, PhD Thesis, Universitat Pompeu Fabra, Barcelona, Spain, February 2012.

(Turnbull, Barrington, Torres, Lanckriet; 2007) **Towards Musical Query-by-Semantic-Description using the CAL500 Data Set**, Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), Amsterdam, the Netherlands, July 2007.

(Turnbull, Barrington, Yazdani, Lanckriet; 2009) **Combining Audio Content and Social Context for Semantic Music Discovery**, Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Boston, USA, July 2009.

(Urbano, Schedl; 2012) **Minimal Test Collections for Low-Cost Evaluation of Audio Music Similarity and Retrieval Systems**, International Journal of Multimedia Information Retrieval, 2012.

(Whitman, Ellis; 2004) **Automatic Record Reviews**, Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain, 2004.

(Whitman, Lawrence; 2002) **Inferring Descriptions and Similarity for Music from Community Metadata**, Proceedings of the 2002 International Computer Music Conference (ICMC), Göteborg, Sweden, 2002.

(West, Cox, Lamere; 2006) **Incorporating Machine-Learning Into Music Similarity Estimation**, Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, Santa Barbara, USA, October 2006.