Music Information Retrieval 2.0

Markus Schedl
Peter Knees

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

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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Why 2.0?

Music Information Retrieval (MIR) from a very traditional IR perspective is to have a short piece of music (symbolic, audio) and to find relevant pieces in a repository (e.g., that exact piece, "similar sounding" pieces, cover versions)

MIR spans more than that:

- MIR is not only query-by-example retrieval
- MIR is about music retrieval in general, i.e., also (textual) meta-data and not just musical representations alone
- Moreover, "MIR 2.0" deals with vast amounts of data + utilizes the power-of-the-crowds, e.g., for "semantic" song annotations

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Who are we ?

Markus Schedl

Assistant Professor at the
Department of Computational Perception **/ Johannes Kepler University 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 at the
Department of Computational Perception **/ Johannes Kepler University 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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

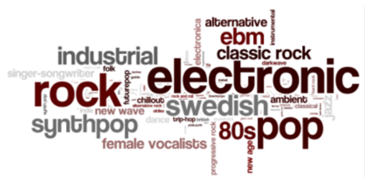
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Overview

Introduction to Music Information Retrieval

Content-based Feature Extraction

Context- and Web-based Methods



Future Directions

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Introduction to MIR

1. Definitions
2. Applications
3. Typical Tasks and Challenges
4. Types of Computational Features

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Introduction to MIR: Definitions

"MIR is a **multidisciplinary** research endeavor that strives to develop innovative **content-based searching schemes**, novel **interfaces**, and evolving **networked delivery** mechanisms in an effort to make the world's vast store of music accessible to all."

[Downie, 2004]

"...actions, methods and procedures for **recovering stored data** to provide information on music."

[Fingerhut, 2004]

"MIR is concerned with the **extraction**, **analysis**, and **usage** of information about **any kind of music entity** (for example, a song or a music artist) on **any representation level** (for example, audio signal, symbolic MIDI representation of a piece of music, or name of a music artist).

[Schedl, 2008]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Automatic Playlist Generation

Pandora.com

- continously plays similar music
- based on the *Music Genome Project*
- manually annotated tracks
- songs can be rated

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Browsing Music Collections

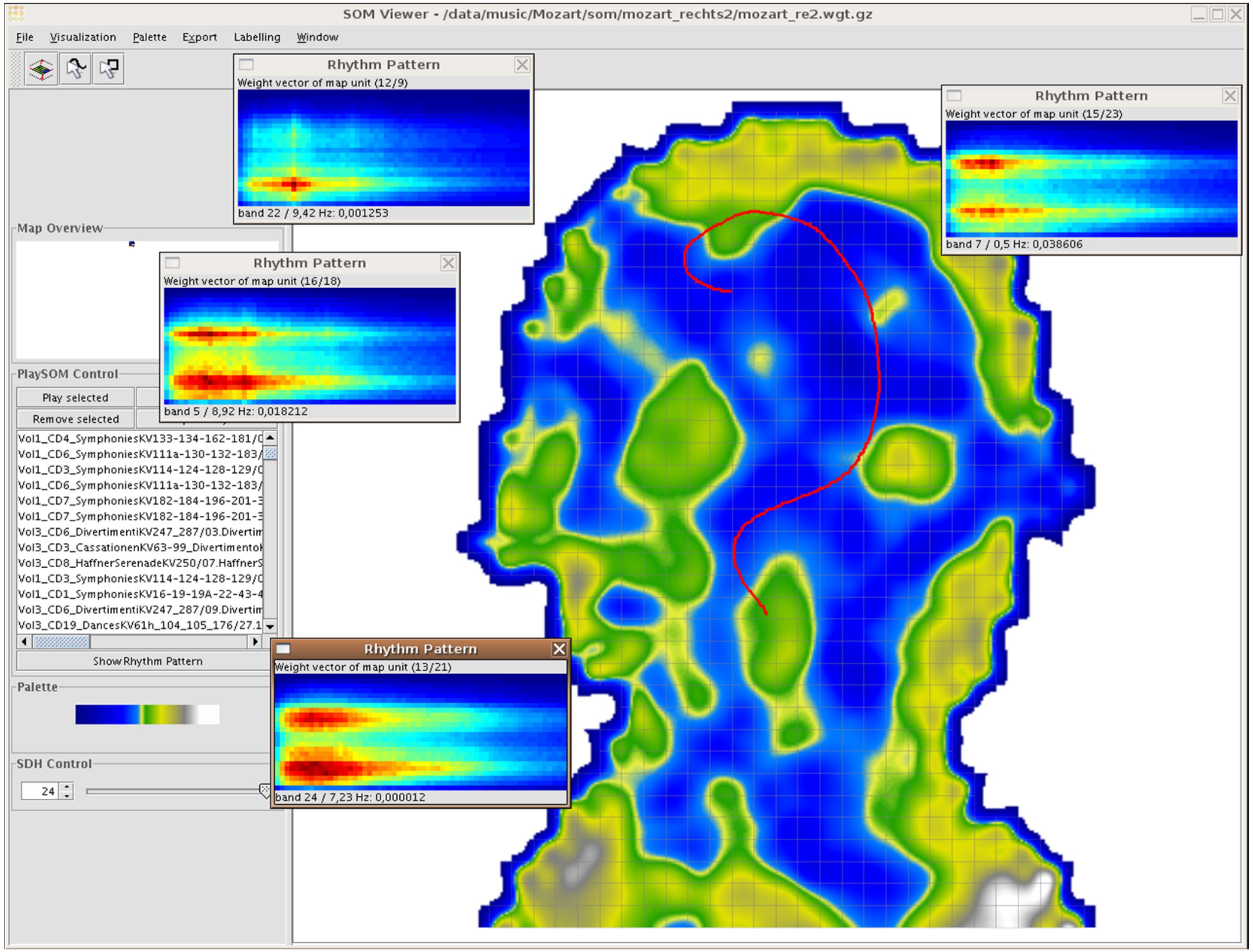
- music collections become larger and larger (on PCs as well as on mobile players)
- most UIs of music players still only allow organization and searching by textual properties accoding to scheme (*genre-)artist-album-track*

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

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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at



Browsing Music Collections

IFS @
TU Wien, 2006

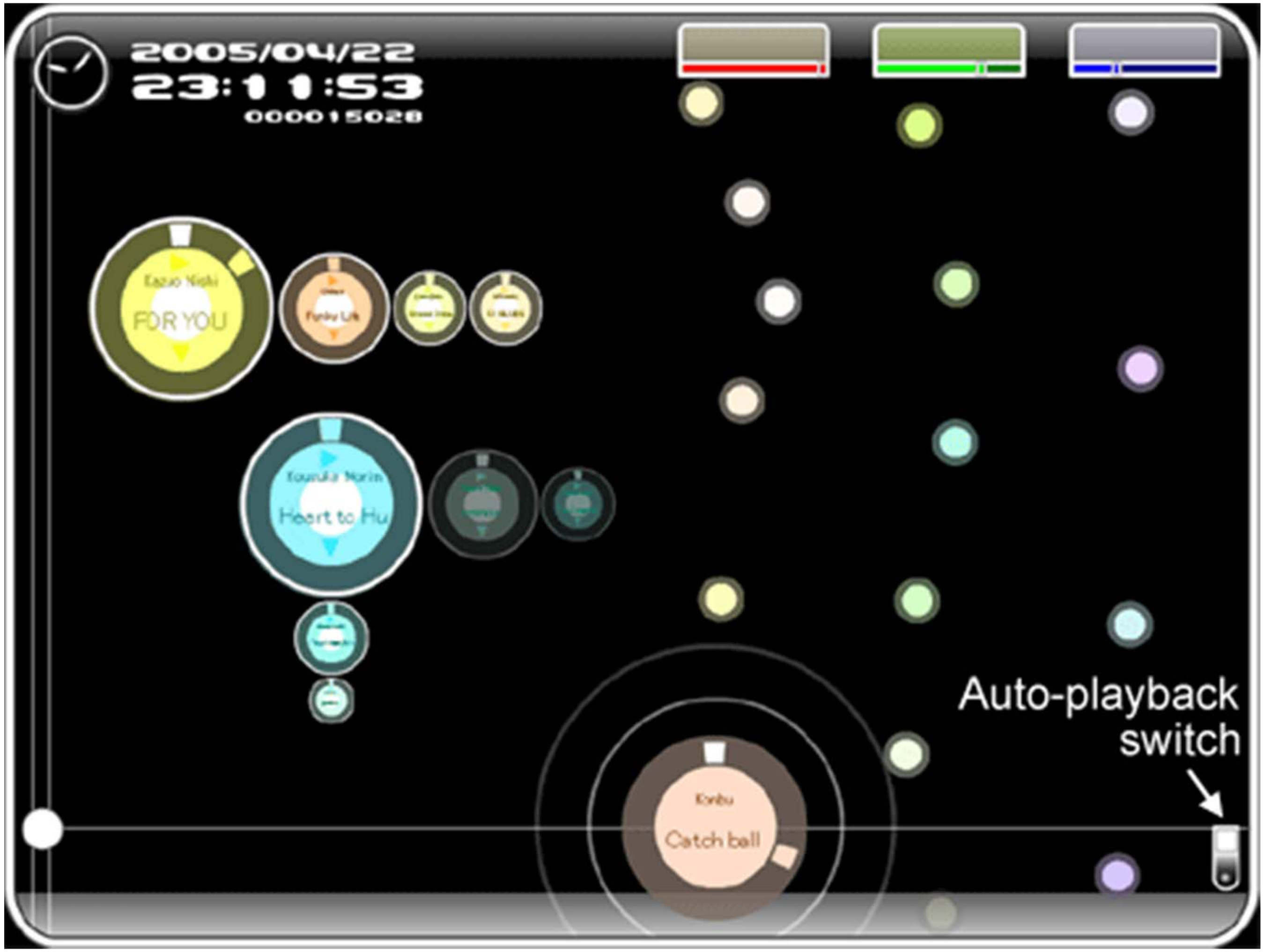
[Mayer et al., ISMIR 2006]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications



„Musicream“, [Goto and Goto, ISMIR 2005]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Audio Fingerprinting

Query-by-example:
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 hum by the user

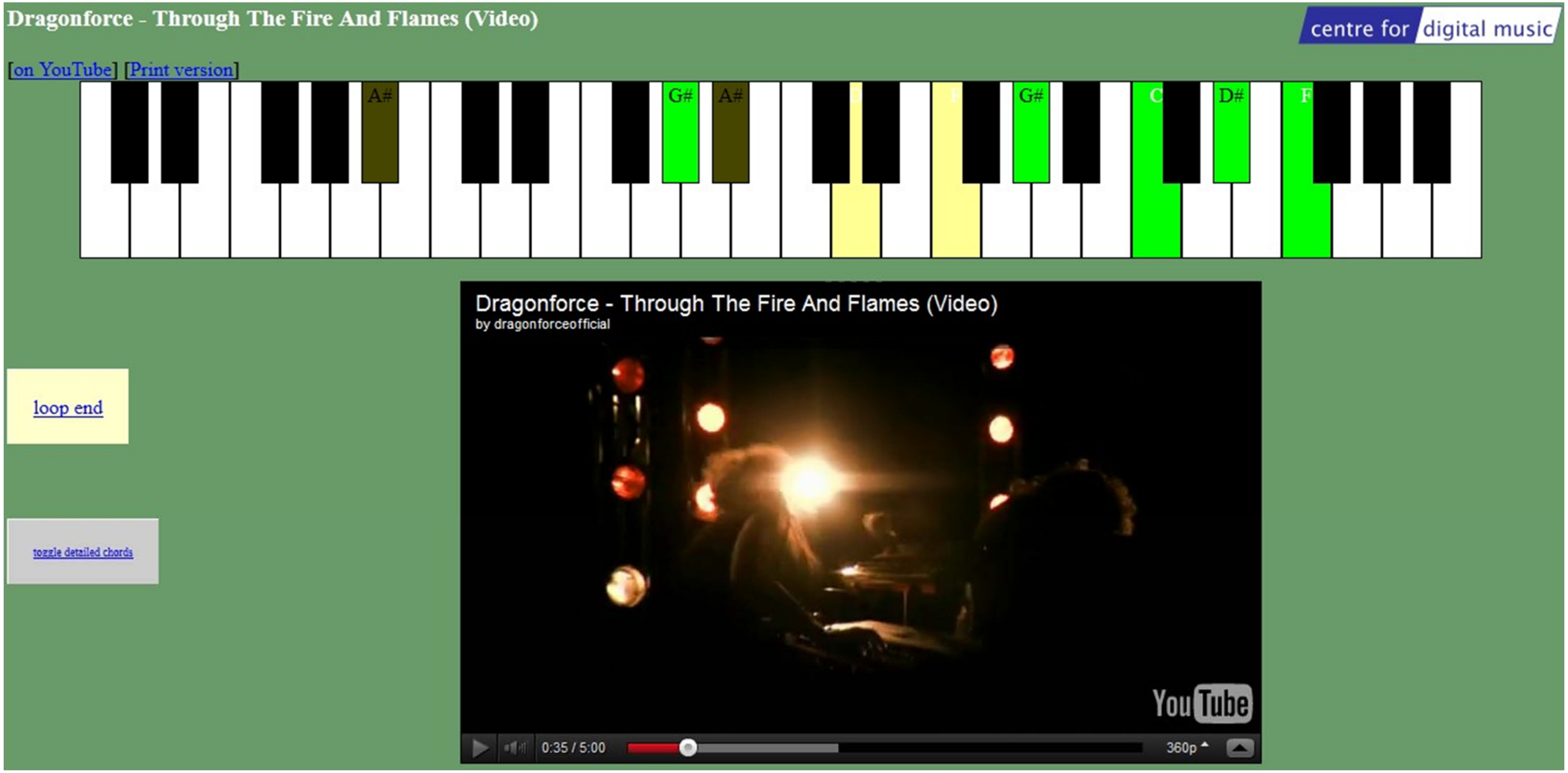
www.shazam.com
www.musicline.de/de/melodiesuche
www.soundhound.com

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

„Yanno" – Chord Detection in Youtube videos, C4DM@QMUL, 2012



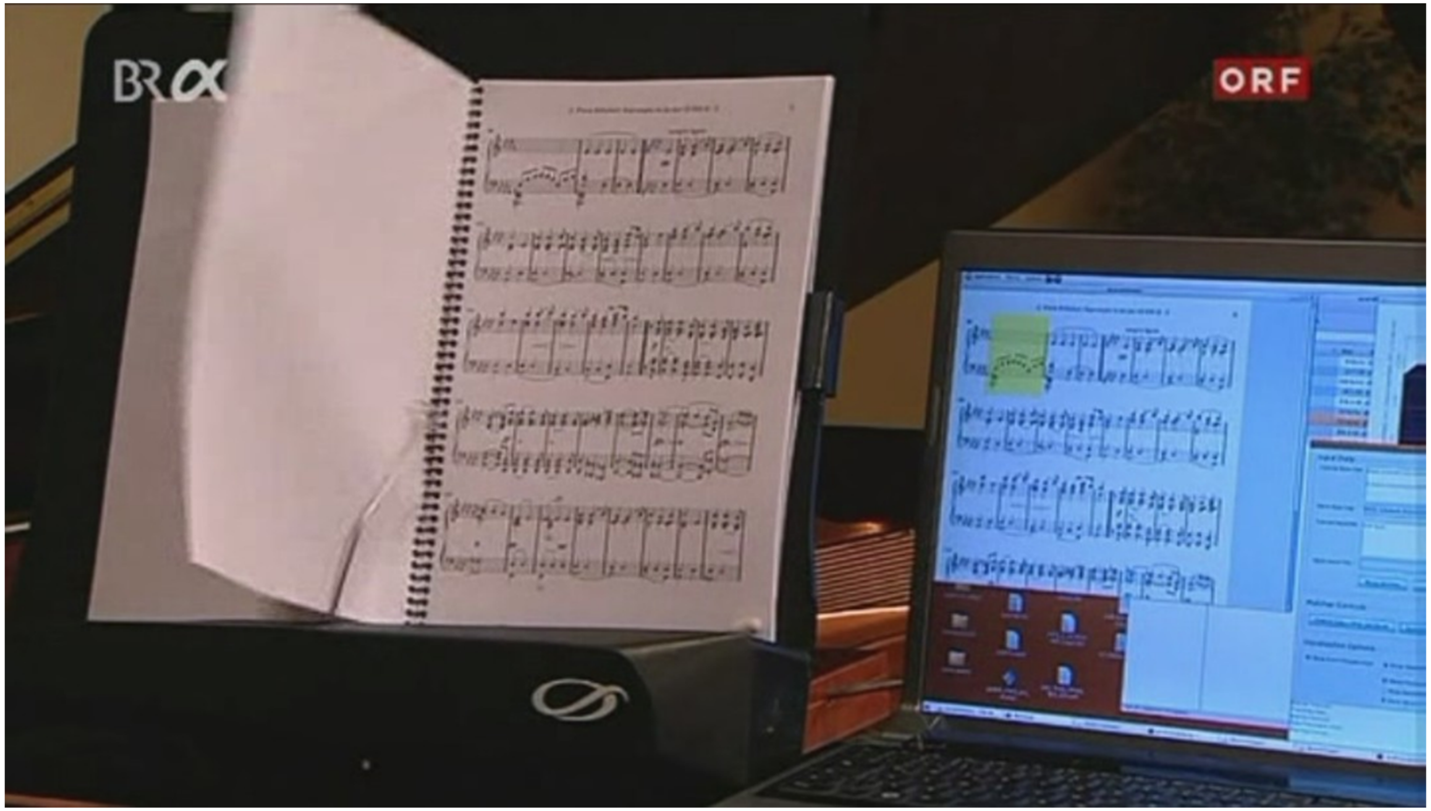
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Real-time Music Following



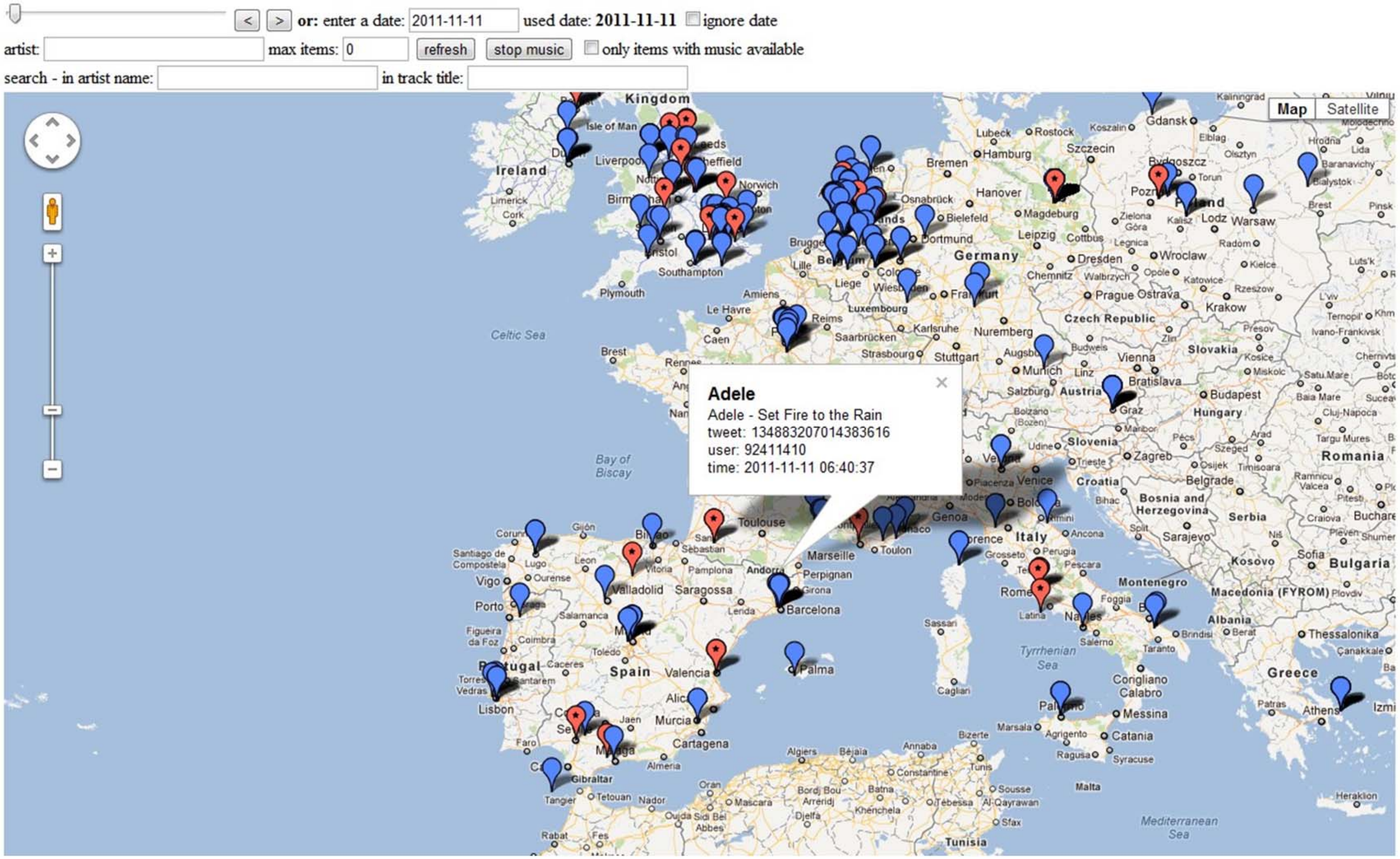
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Geospatial Popularity Estimation

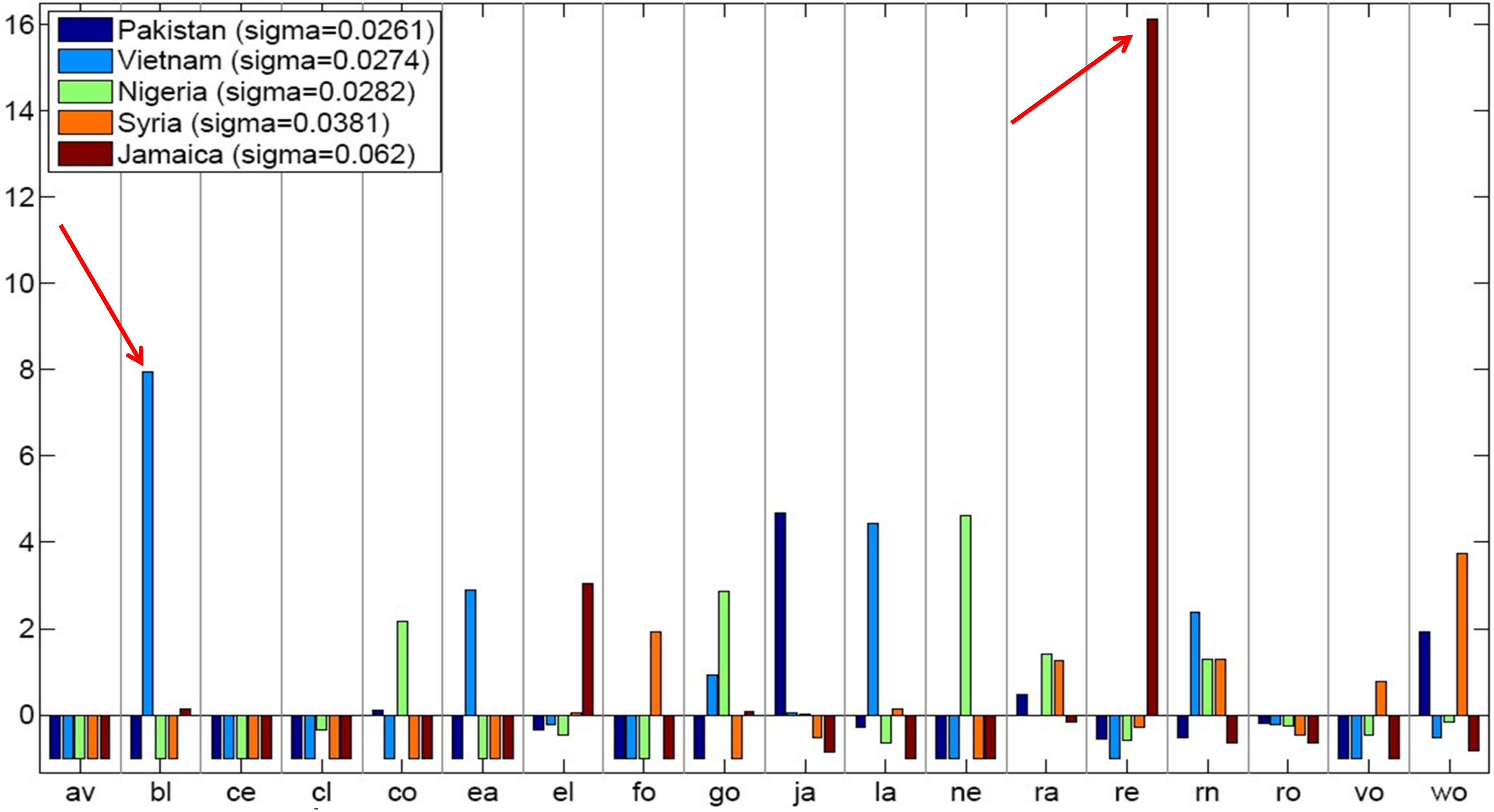


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Geospatial Listening Patterns

Applications

[Schedl and Hauger, AdMIRe 2012]

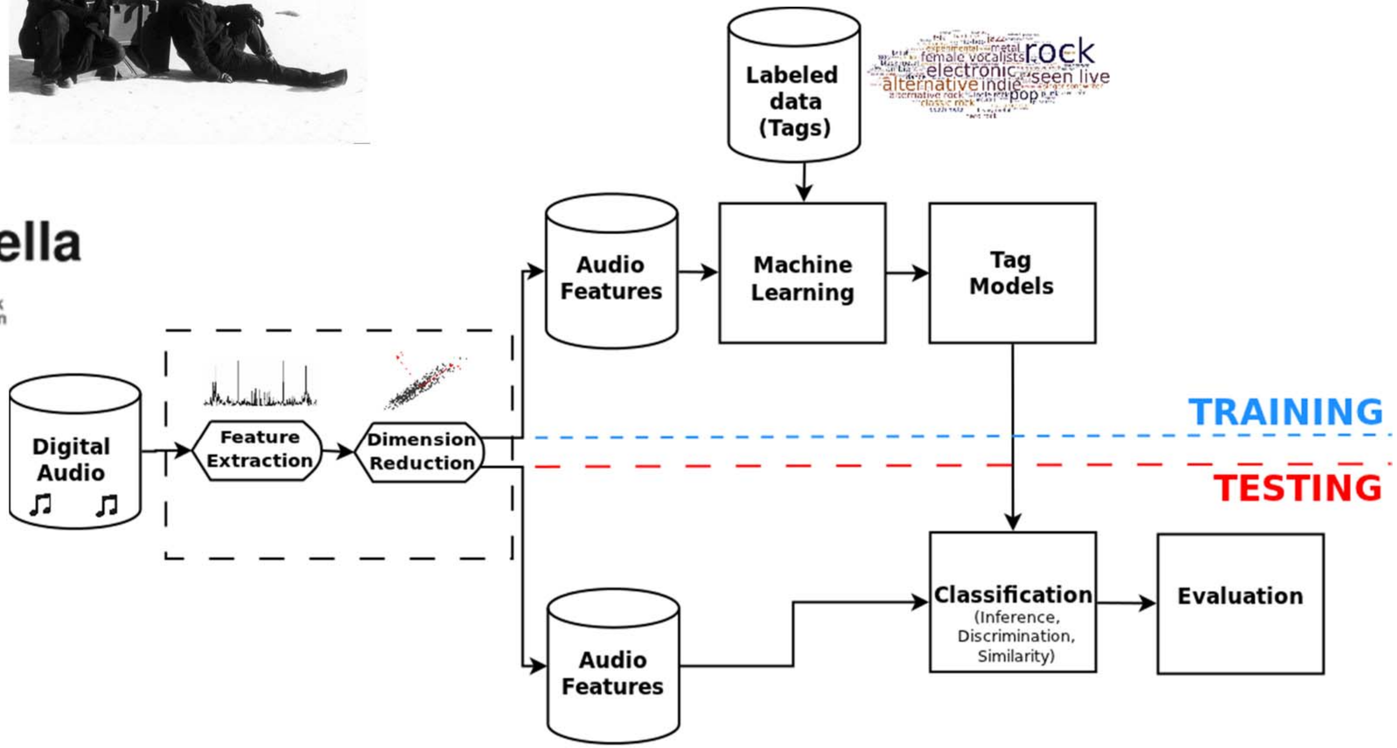


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Applications

Auto-tagging / Retrieval by Tag

[Sordo, PhD 2012]



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Typical Tasks and Challenges

- feature extraction (audio-based vs. context-based approaches)
- similarity measurement, recommendation, automated playlist generation (last.fm, Pandora, Echo Nest, ...)
- user interfaces, visualization, and interaction
- audio fingerprinting (copyright infringement detection, music identification services like shazam.com or musicbrainz.org)
- voice and instrument recognition, speech/music discrimination
- structural analysis, alignment, and transcription (segmentation, self-similarities, music summarization, audio synthesis, audio and lyrics alignment, audio to score alignment aka score following, and audio to score transcription)
- optical music recognition (OMR)
- classification and evaluation (ground truth definitions, quality measurement, e.g. for feature extraction algorithms, genre classification)

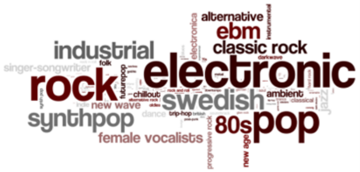
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

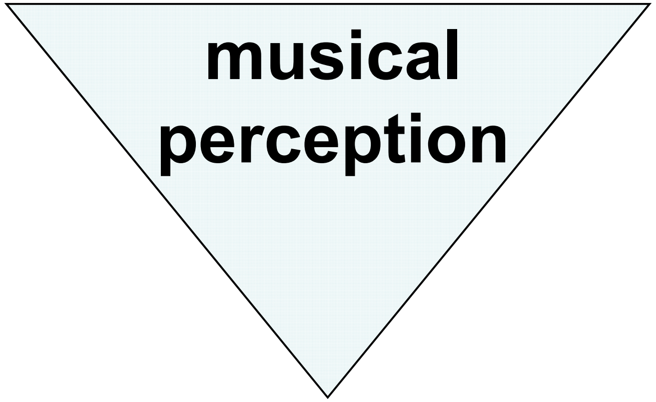
The Feature Extraction Triangle

music content



music context

Examples:
- rhythm patterns
- MFCC models
- melodiousness
- percussiveness
- loudness



Examples:
- collaborative tags
- song lyrics
- album cover artwork
- artist's background
- playlist co-occurrences

user context

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

[Schedl and Knees, **Personalization in Multimodal Music Retrieval**, Proc. AMR 2011]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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

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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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
- Artist names that equal common speech terms

Challenge for both Content and Context Analysis

- Extraction of relevant features from *noisy signal*

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Content-based Feature Extraction

1. Categorization of Content-based Features
2. Scheme of a Content-based Feature Extractor
3. Different Feature Extractors (low-level, mid-level, high-level)

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Categorization of Content-based Features

Level of abstraction:

low-level
closest to audio signal (e.g., energy, zero-crossing-rate);
app: audio identifcation

mid-level
aggregate low-level, apply psycho-acoustic models (e.g., MFCC, FP);
app: similarity estimation, query-by-humming

high-level
directly meaningful to end user (e.g., semantic tags learned from audio features);
app: autotagging, genre classification

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Categorization of Content-based Features

Temporal scope:

instantaneous
feature is valid for a "point in time", time resolution of ear is several msec!

segment
feature is valid for a segment, e.g., phrase, chorus (on a high level), or a chunk of X consecutive seconds in the audio signal

global
feature is valid for whole audio excerpt or piece of music

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Categorization of Content-based Features

♪

Acoustic property to describe:

timbre

rhythm

pitch and melody

chords and harmony

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Categorization of Content-based Features

Domain:

time domain
consider signal in *time/amplitude* representation

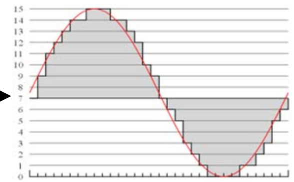
frequency domain
consider signal in *frequency/magnitude* representation, Fast Fourier Transform

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Scheme of a Content-based Feature Extractor

analog signal

sampling
quantization

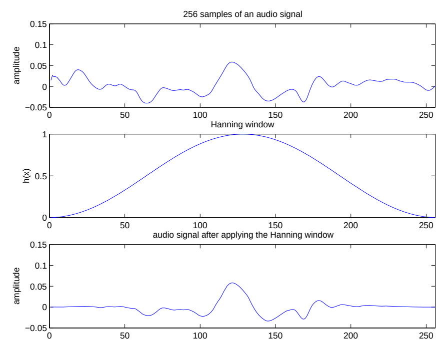


Pulse Code Modulation (PCM)

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

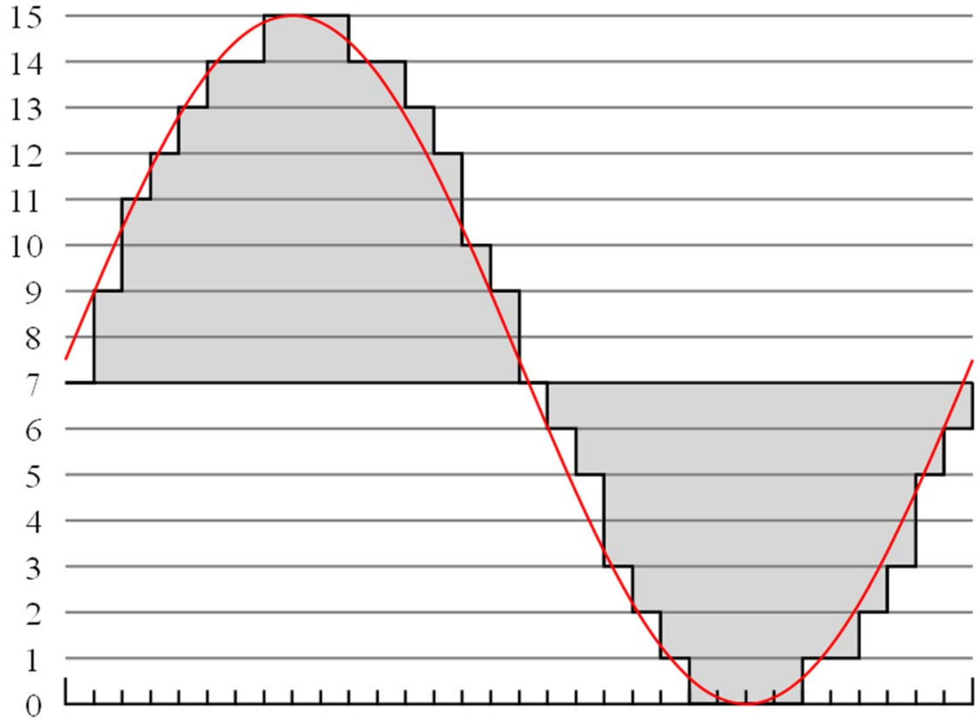
aggregation, model building (mean, median, sum, GMM, HMM)

feature value, vector, or matrix

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Analog-Digital-Conversion (ADC)



PCM: analog signal is sampled at equidistant intervals and quantized in order to store it in digital form (here with 4 bits)

Problems that may occur in ADC:

quantization error: difference between the actual analog value and quantized digital value

solution: finer resolution (use more bits for encoding), common choice in music encoding: 16 bits per channel

due to **Nyquist–Shannon Sampling Theorem**, frequencies above ½ of sampling frequency (Nyquist frequency) are discarded or heavily distorted

solution: choose a sampling frequency that is high enough (e.g. 44,100 Hz for Audio CDs)

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Frame

In short-time signal processing, pieces of music are cut into segments of fixed length, called frames, which are processed one at a time; typically, a frame comprises 256 - 4096 samples.

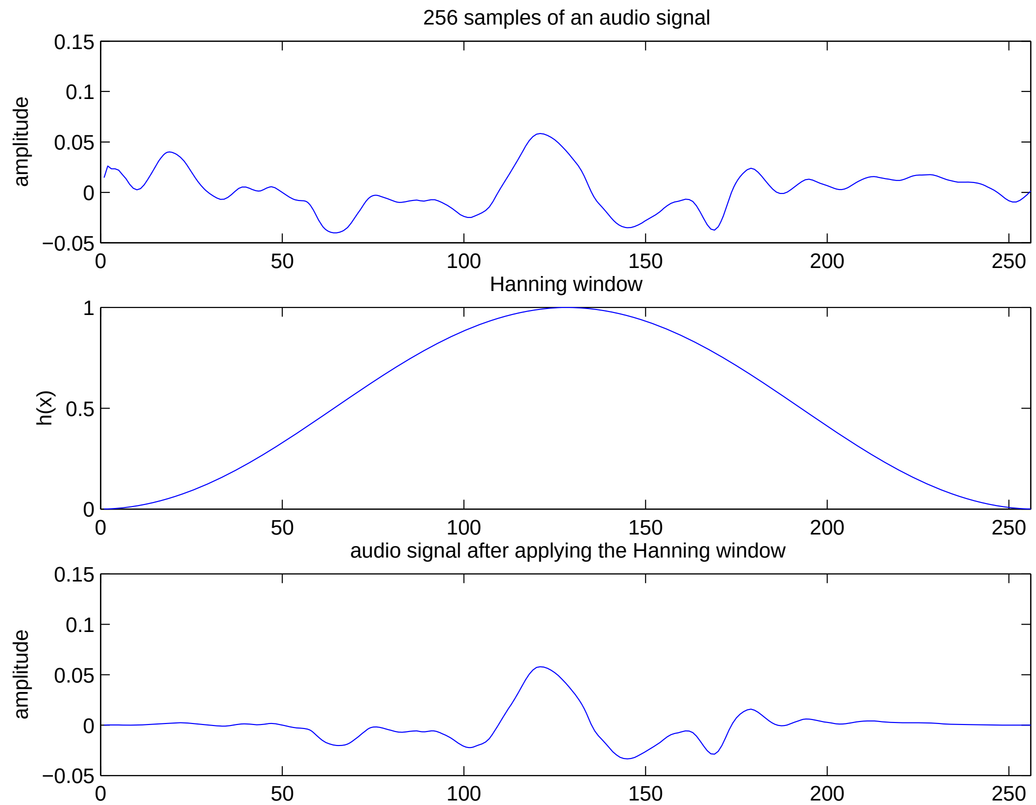
Windowing

To calculate the FFT, we need a periodic signal; thus, the PCM magnitude values of each frame are multiplied point-by-point with a suited function

e.g. Hanning window (Julius von Hann, Austrian meteorologist)

$$h(x) = \frac{1}{2} \cdot \left[1 + \cos\left(\frac{\pi + x}{n} \right) \right]$$

n...number of samples in frame

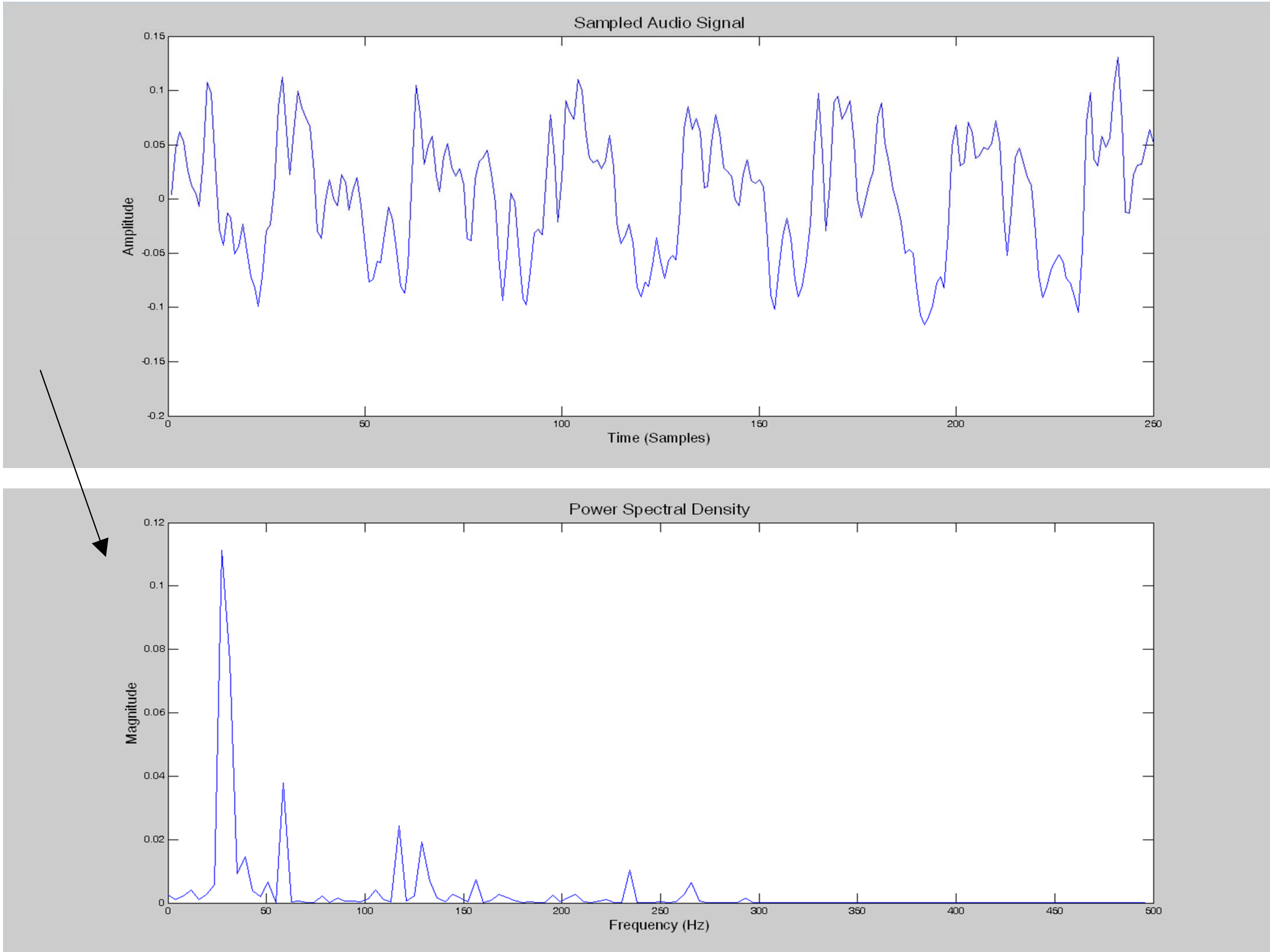


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Fourier Transform
(after "Jean Baptiste Joseph Fourier")

- transformation of the signal from **time domain** (time vs. amplitude) to **frequency domain** (frequency vs. magnitude)
- theorem: **any continuous periodic function with a period of 2π can be represented as the sum of sine and/or cosine waves**
- implication: any (analog) audio signal can be decomposed into an infinite number of overlapping waves (of different frequencies) given that it is periodic (windowing!)
- if the domain of the input function (audio signal) is discrete as in our case: → **Discrete Fourier Transform (DFT)**
- in practice: DFT is efficiently calculated via **Fast Fourier Transform (FFT)** by [Cooley and 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)

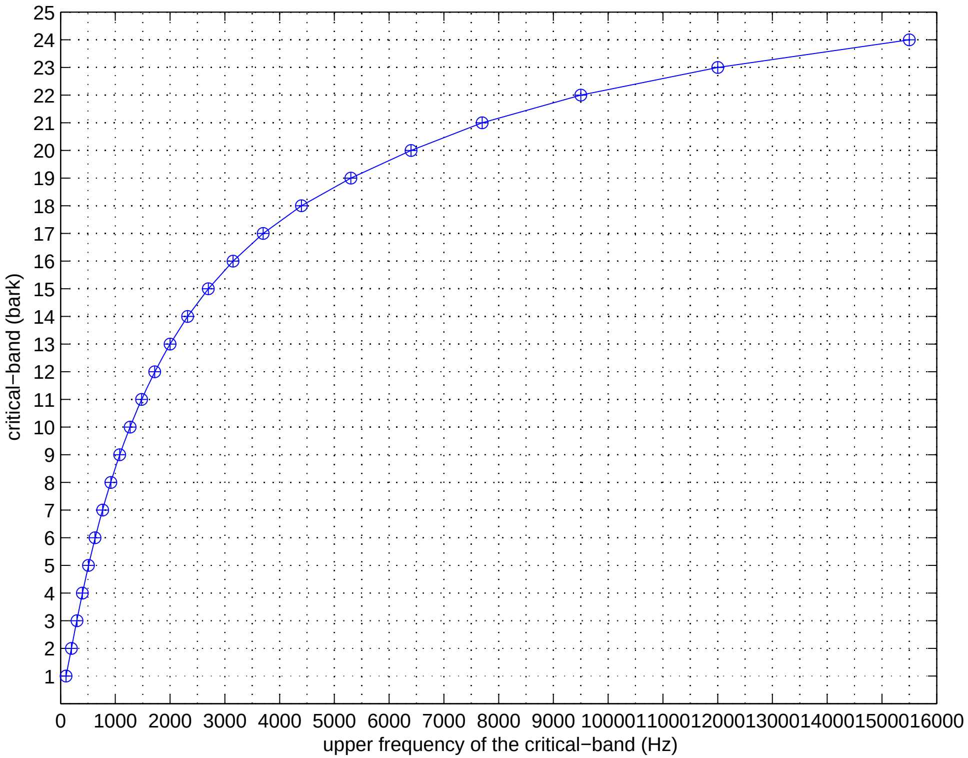
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Critical Bands (Bark Scale)

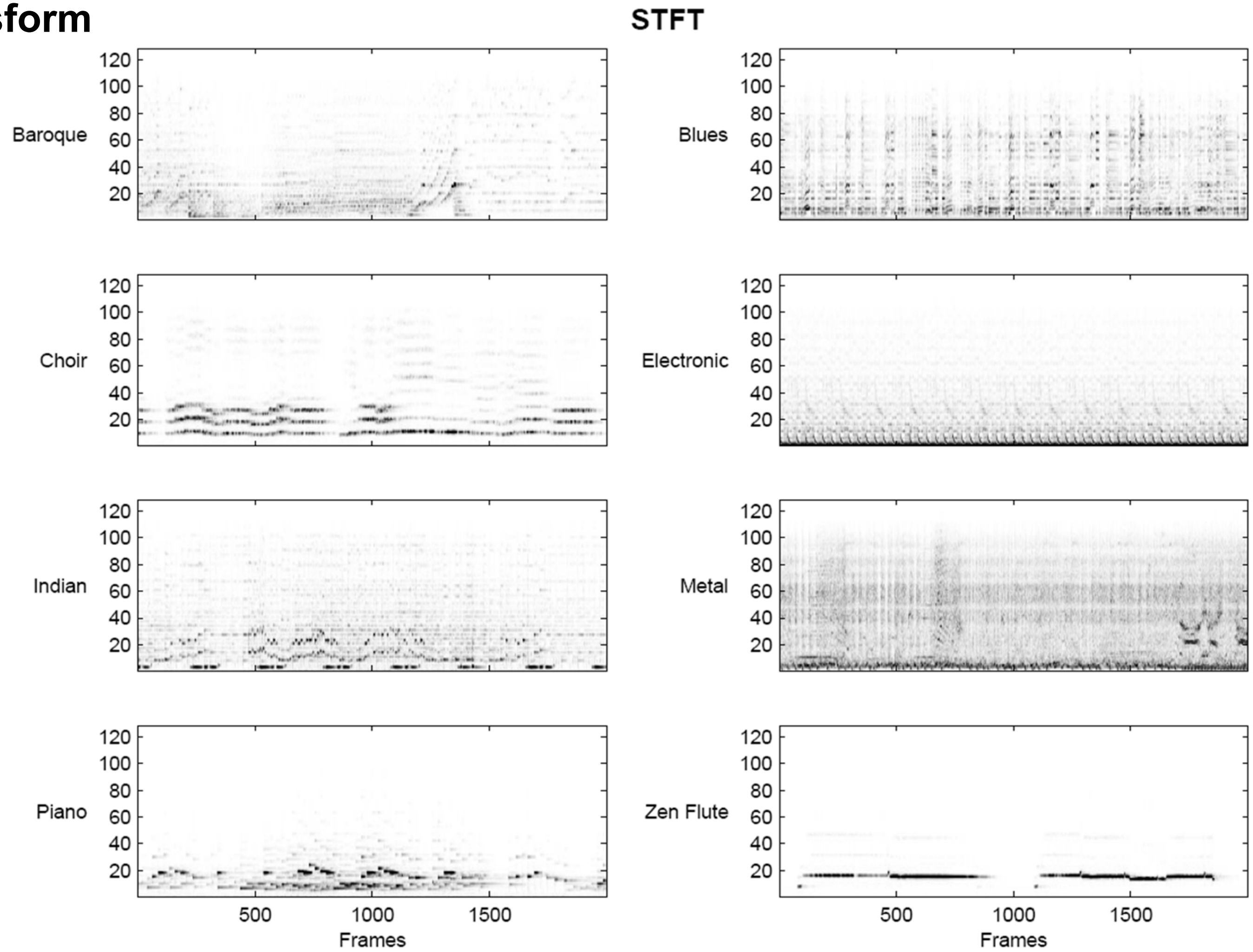
- perceptually uniform measure of frequency
- reflects characteristics of human auditory system (our auditory perception is logarithmic!)
- measured according to **bark scale**
- used for psychoacoustic postprocessing



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Representation as Short-time Fourier Transform



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Low-Level Feature: Root-Mean-Square (RMS) Energy

(aka RMS power, RMS level, RMS amplitude)

scope: time domain

calculation:

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

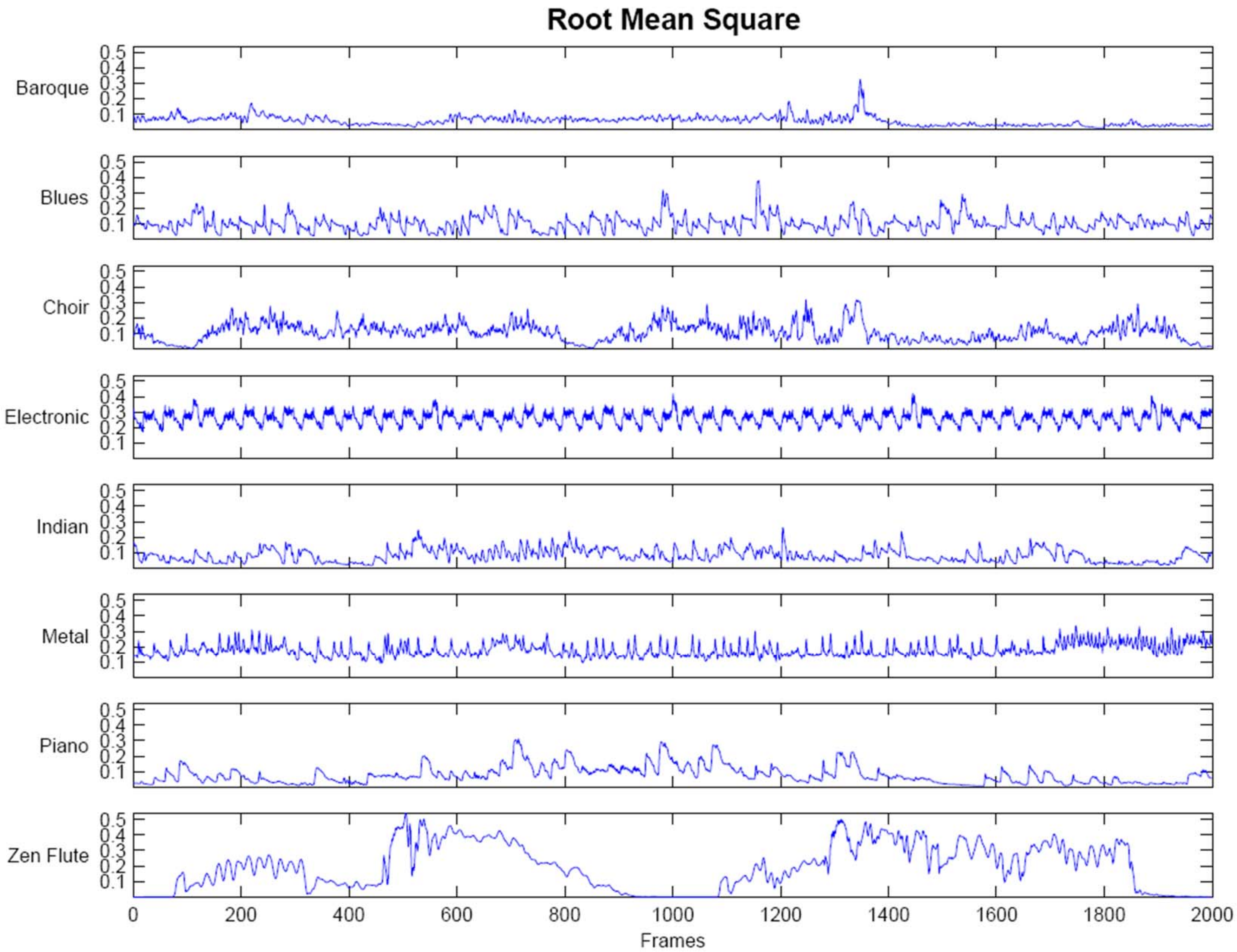
remarks:
+ beat-related feature, can be used for beat detection
+ related to perceived intensity
+ good loudness estimation
– discriminative power not clear

$M_t(n)$...magnitude in frequency domain at frame t and frequency bin n
s(k)...amplitude of k^{th} sample in time domain
K...frame size (number of samples in each frame)
N...number of highest frequency band

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

RMS Energy: Illustration

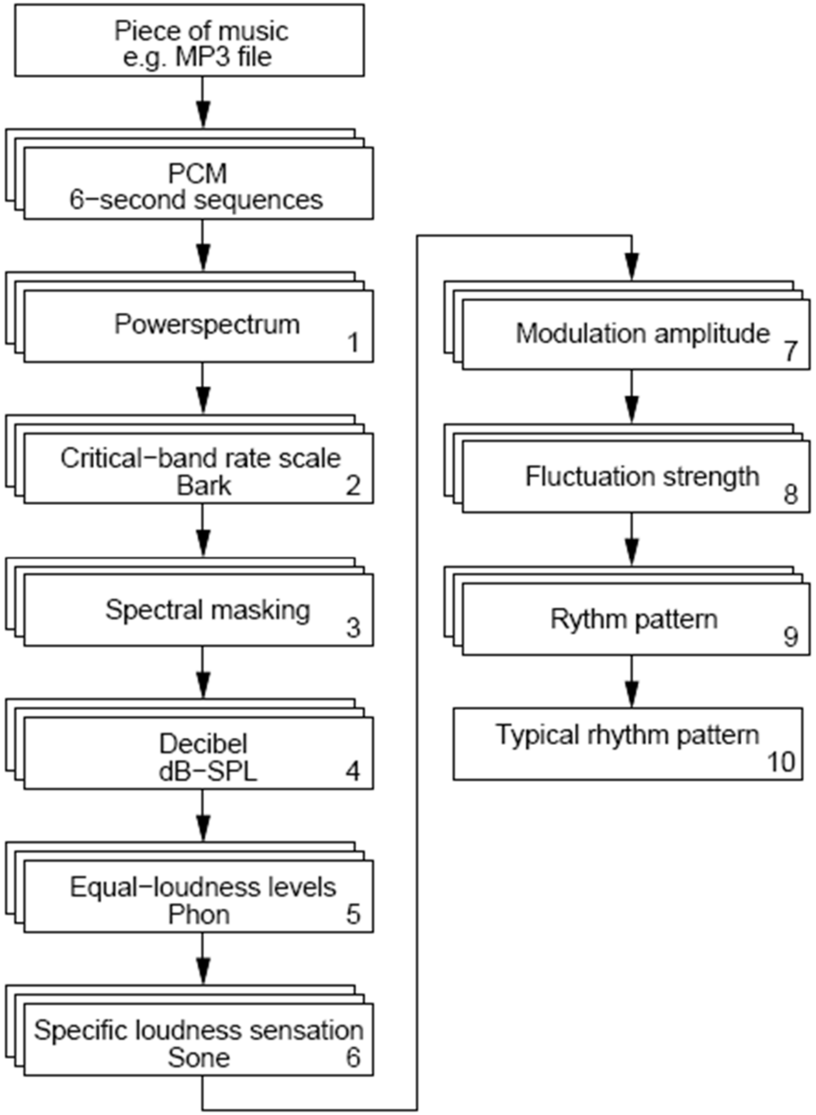


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Mid-Level Feature Extractor: Fluctuation Patterns

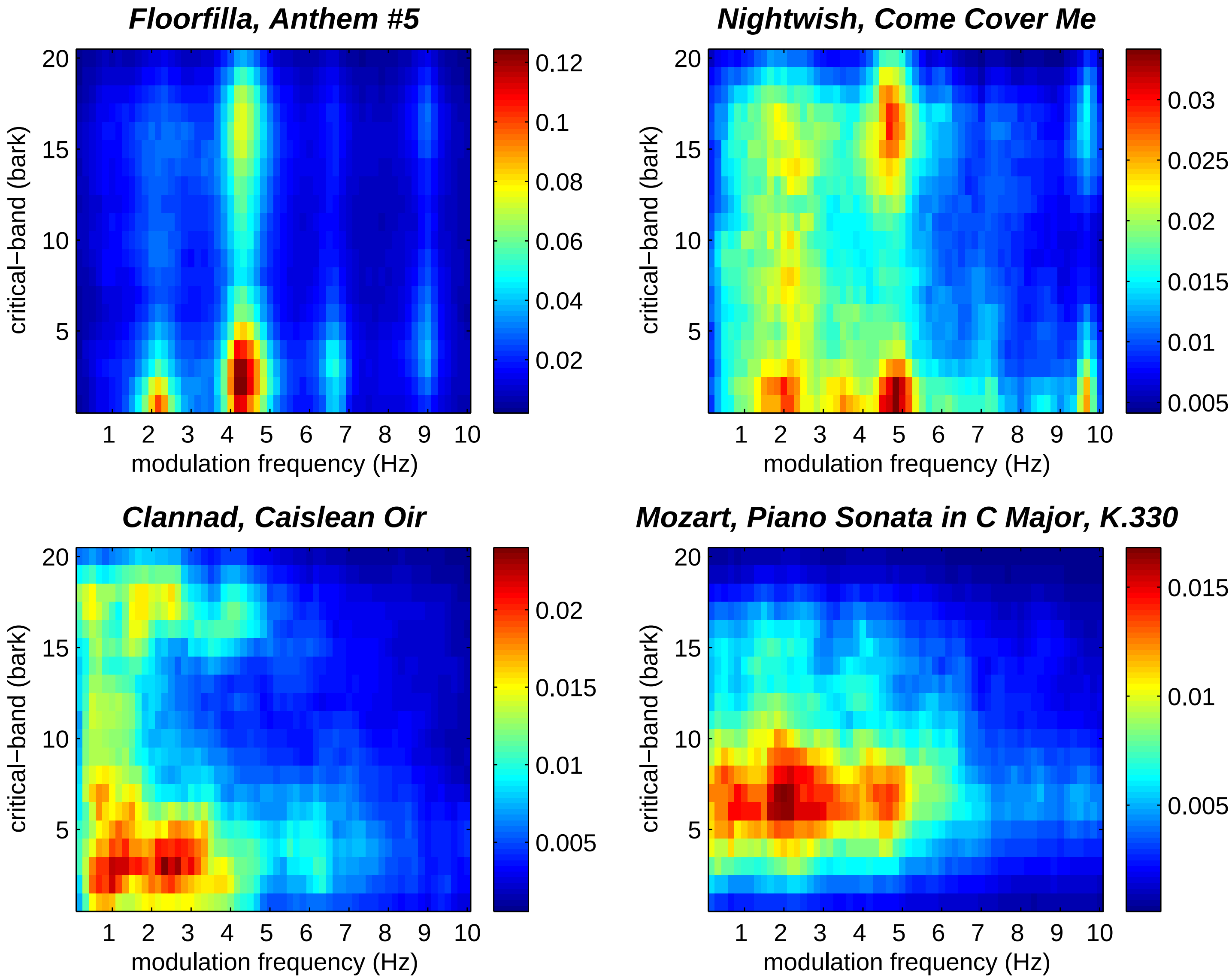
Idea:
measure how strong and fast beats are played within certain perceptually adjusted frequency bands



[Pampalk et al., ACM Multimedia 2002]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Fluctuation Patterns: Illustration



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

[Seyerlehner et al., DAFx 2010]

Most low-level and mid-level feature extractors apply the **Bag-of-Frames** (BoF) model, i.e., features calculated for each frame are treated in an unordered fashion when aggregating.

→ **Loss of Temporal Information**

However, for most non-trivial music relations, temporal information is vital.

Idea of BLF:
Process **blocks of frames** instead of single frames.

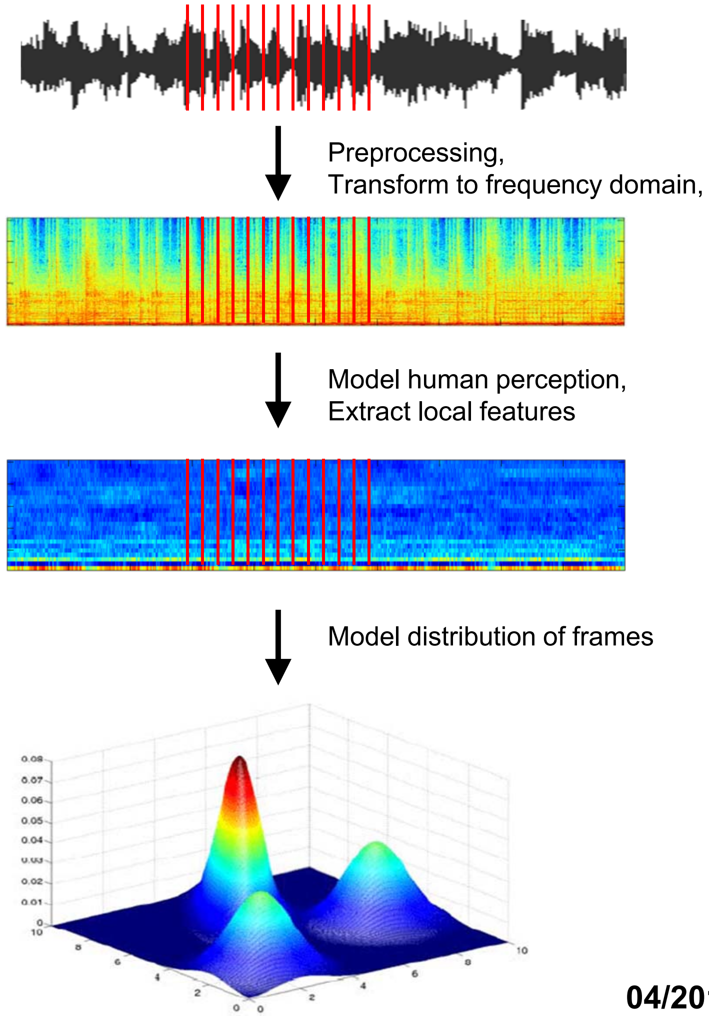
Department of Computational Perception
JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

A Typical Frame-Level Similarity Algorithm

1) digitized audio signal (PCM)
2) transform to frequency domain (FFT)
3) extract frame-level features
4) model each song as distribution of its feature vectors → **Bag of Frames (BoF)**



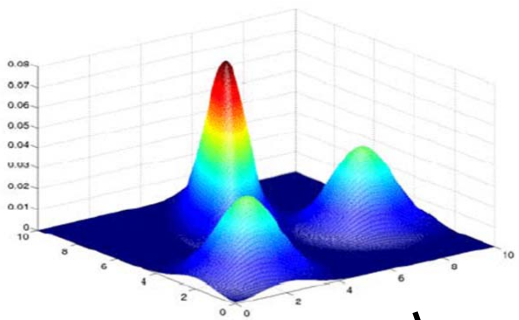
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

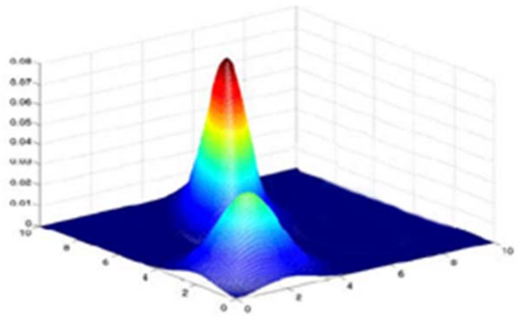
A Typical Frame-Level Similarity Algorithm: *How to measure similarity?*

As distance function the **KL-**Divergence (Kullback–Leibler), a measure of the relative entropy of two probability distributions, is frequently used.

S1



S2



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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

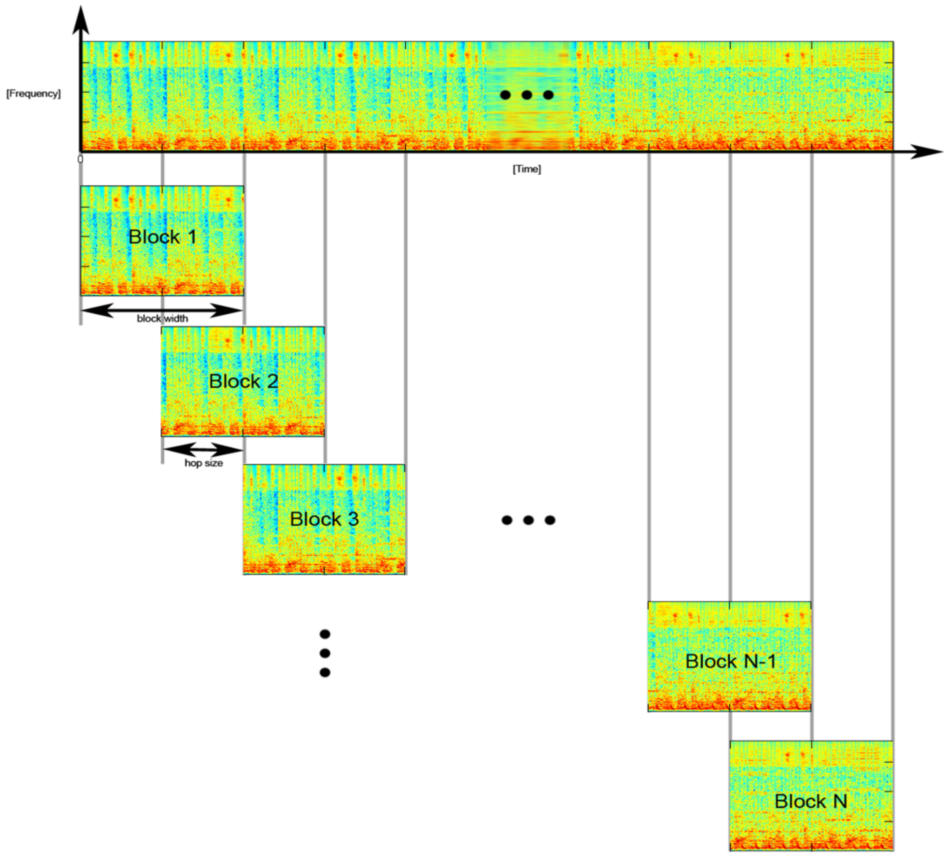
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

The Block-Processing Framework

Idea:
process audio signal in blocks instead of single frames

- better capability to capture temporal information
- entire feature extraction process is defined on block-level



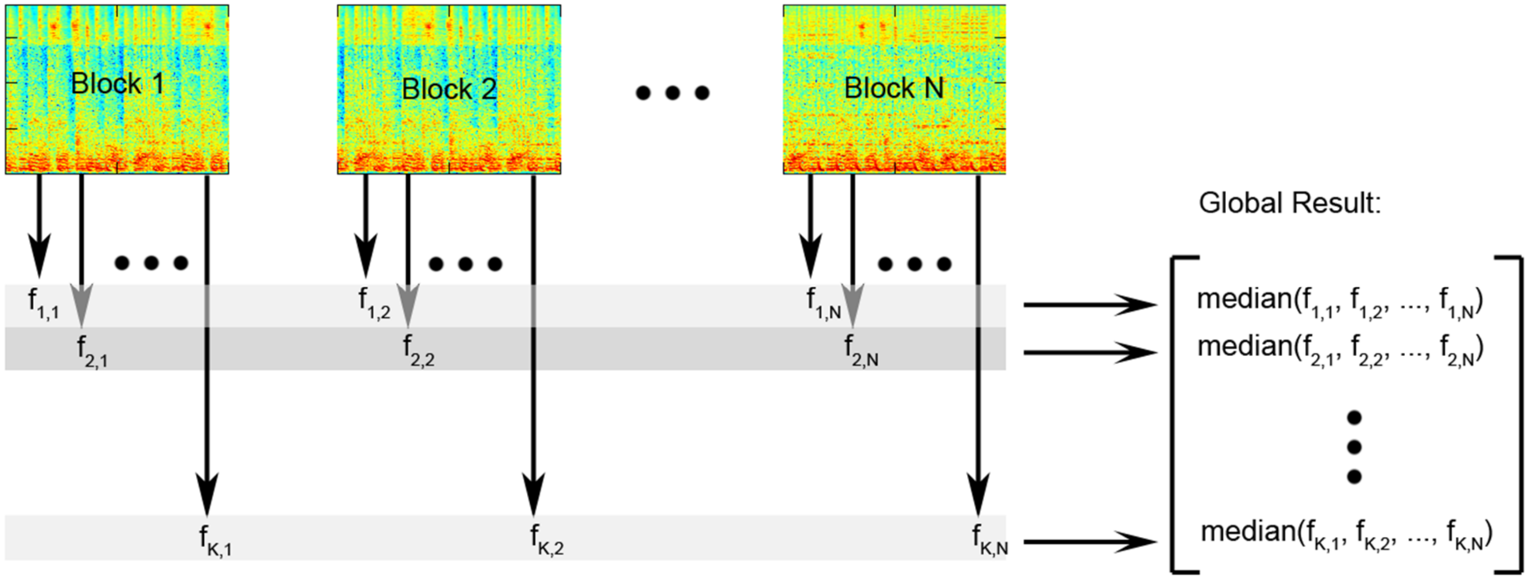
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor:
Block-Level Framework (BLF)

Aggregating Block-Level to Song-Level Features

To obtain one feature vector for the entire song, a generalization step is performed, which is realized by an ***aggregation function***.



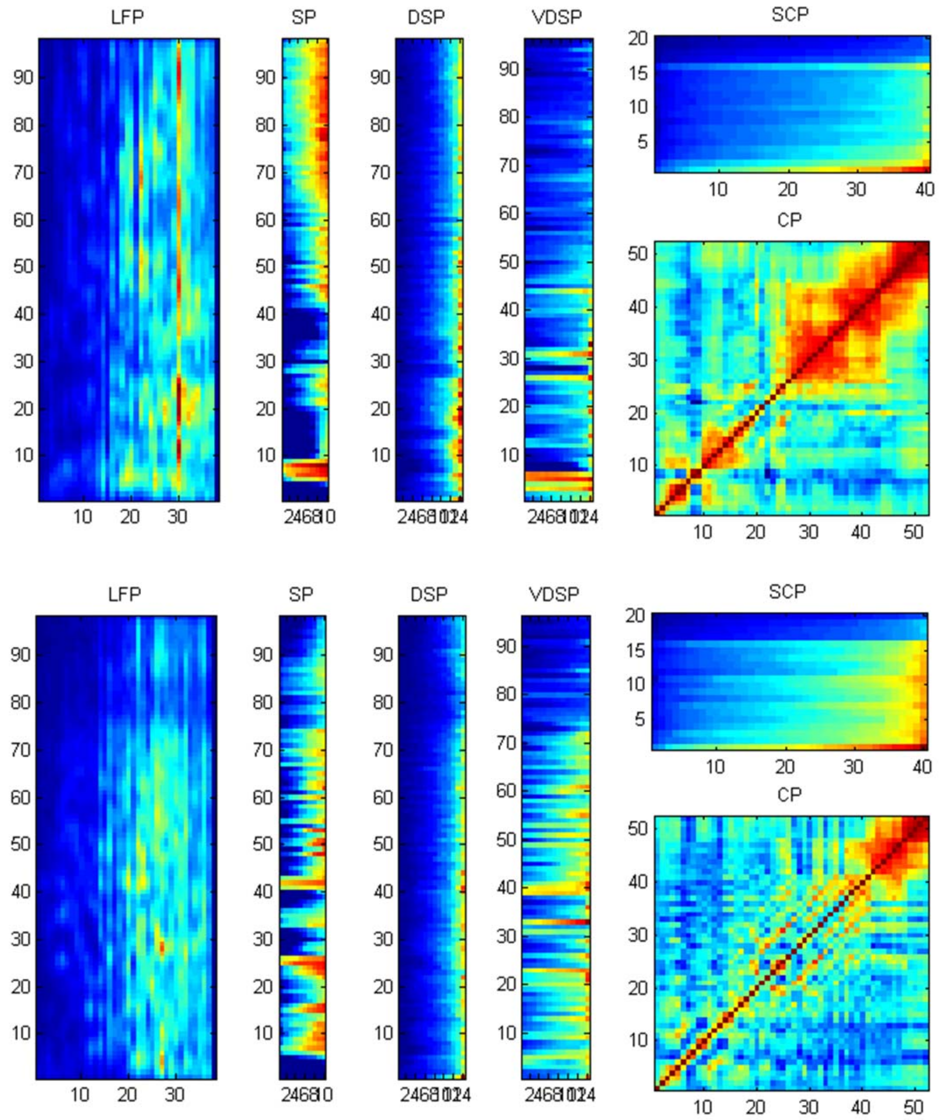
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

BLF differ significantly from other features.

- **Advantages:**
  - + Vector Space Model: The whole mathematical toolbox of vector spaces is available.
  - + easy to use in classification
  - + song models can be visualized
- **Disadvantages:**
  - – high dimensional feature space



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Aggregating Block-Level to Song-Level Features

Several Block-Level Features have been defined in the Block-Processing Framework:

1. **Spectral Pattern (SP)**
2. **Delta-Spectral Pattern (DSP)**
3. **Variance Delta-Spectral Pattern (VDSP)**
4. **Logarithmic Fluctuation Pattern (LFP)**
5. **Correlation Pattern (CP)**
6. **Spectral Contrast Pattern (SCP)**

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

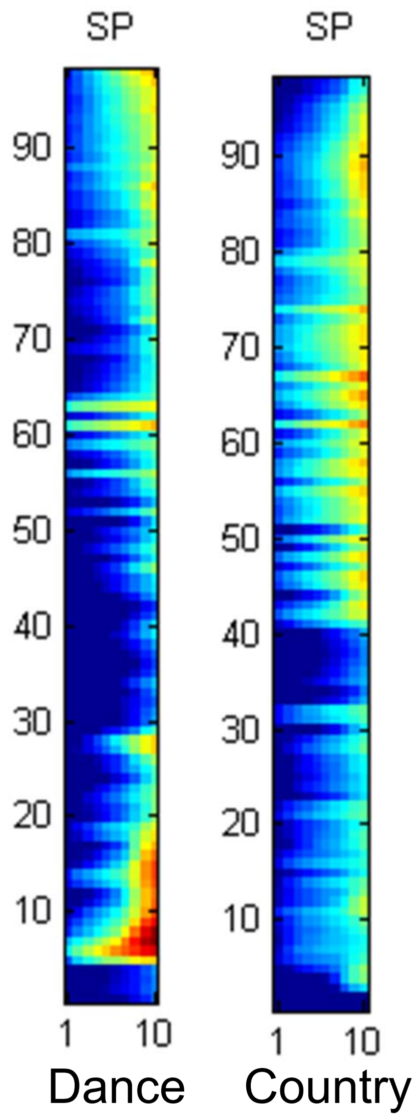
High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Pattern (SP)

- An audio descriptor that captures the frequency content of an audio file
- Each frequency band of an audio block is simply sorted along the time axis.

$$\mathrm{SP} = \begin{bmatrix} \mathrm{sort}(b_{H,1} & \cdots & b_{H,W}) \\ \vdots & \ddots & \vdots \\ \mathrm{sort}(b_{1,1} & \cdots & b_{1,W}) \end{bmatrix}$$

- The 0.9-percentile is used as aggregation function.

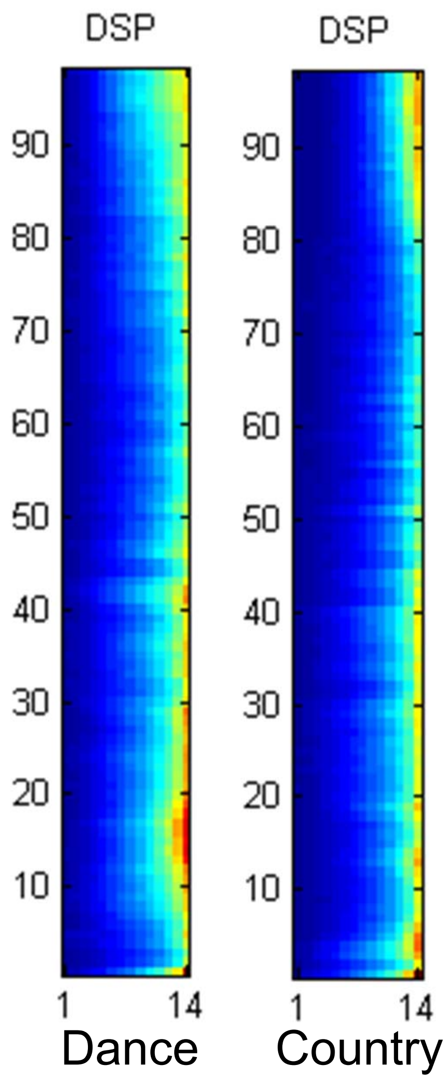


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Delta-Spectral Pattern (DSP)

- First the delta spectrum, i.e., the difference of the original and a delayed version (by 3 frames) is computed to emphasize onsets.
- The same strategy as for the SP is followed – all frequency bands are sorted.
- The 0.9-percentile is used as aggregation function.

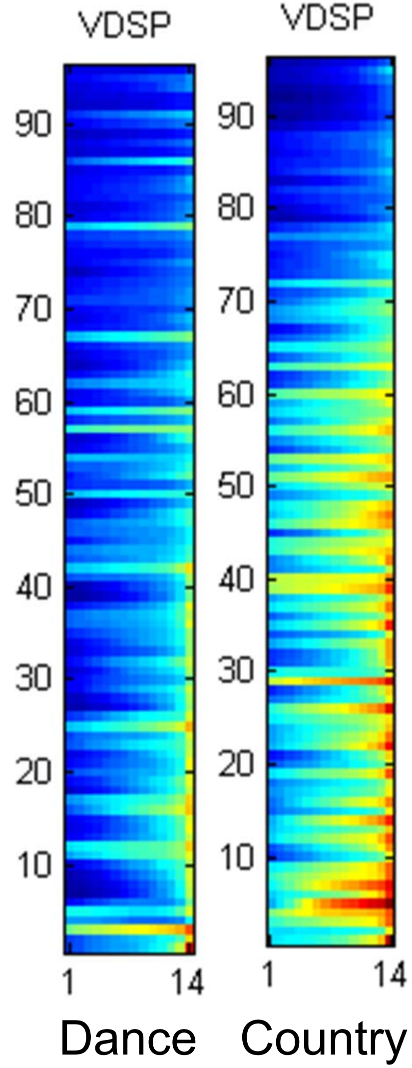


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Variance Delta-Spectral Pattern (VDSP)

- Similar to the DSP
- The variance is used instead of the 0.9-percentile as aggregation function.
- Captures temporal variations of the onset strength



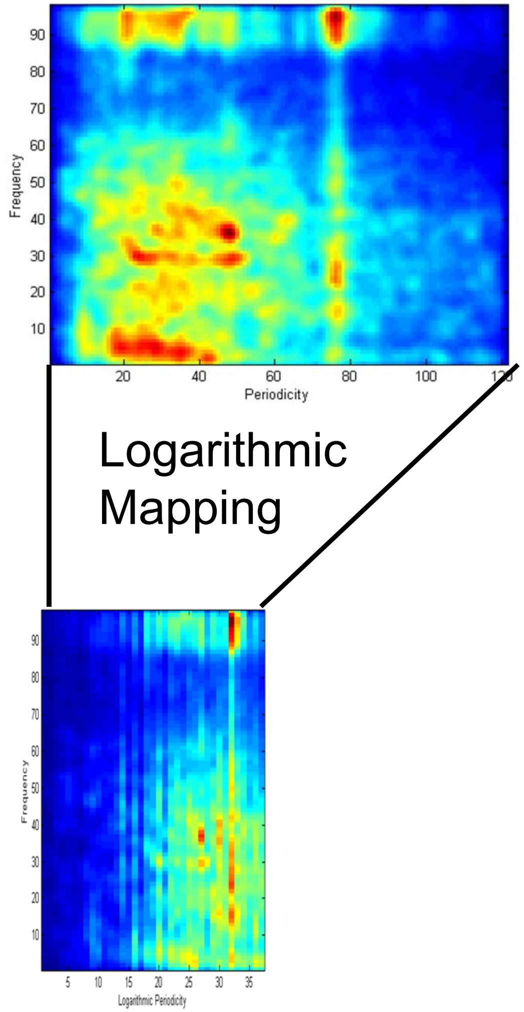
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Logarithmic Fluctuation Pattern (LFP)

- Modified version of the Fluctuation Patterns
- The Cent scale is used instead of the Sone representation.
- The periodicity dimension of the pattern is logarithmically scaled to yield a more tempo invariant representation.



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

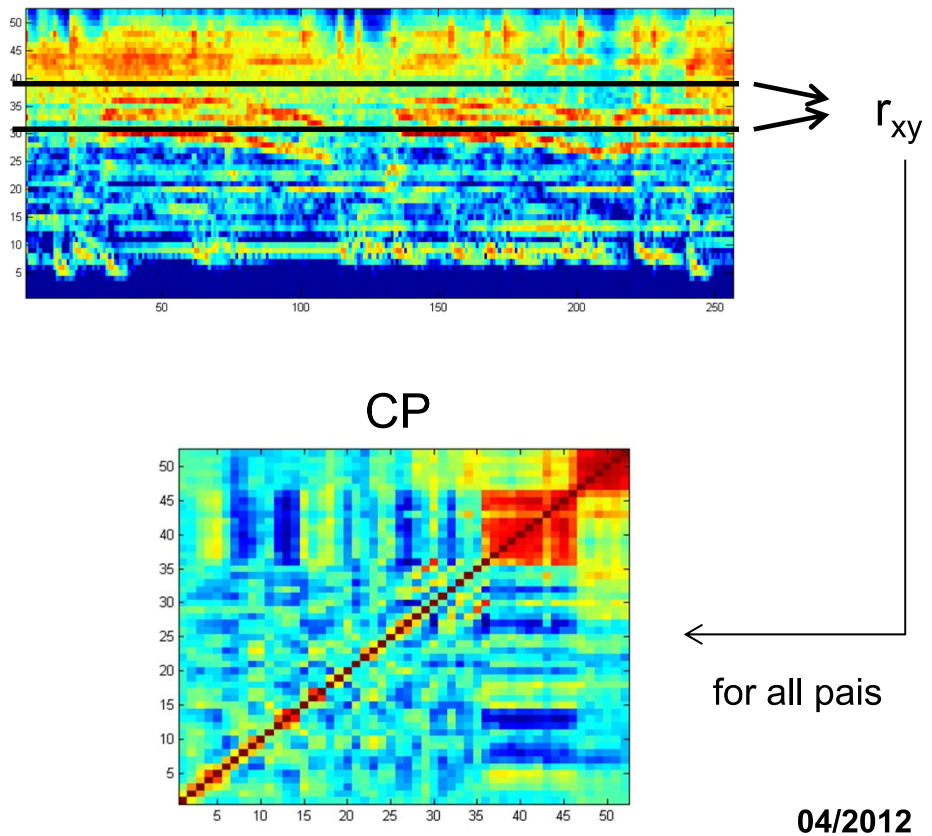
High-Level Feature Extractor: Block-Level Framework (BLF)

Correlation Pattern (CP)

- Reduce the Cent spectrum to 52 frequency bands
- Captures the temporal relation of the frequency bands
- Compute the pairwise linear correlation between each frequency band.

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1) s_x s_y}$$

- The 0.5-percentile is used as aggregation function.

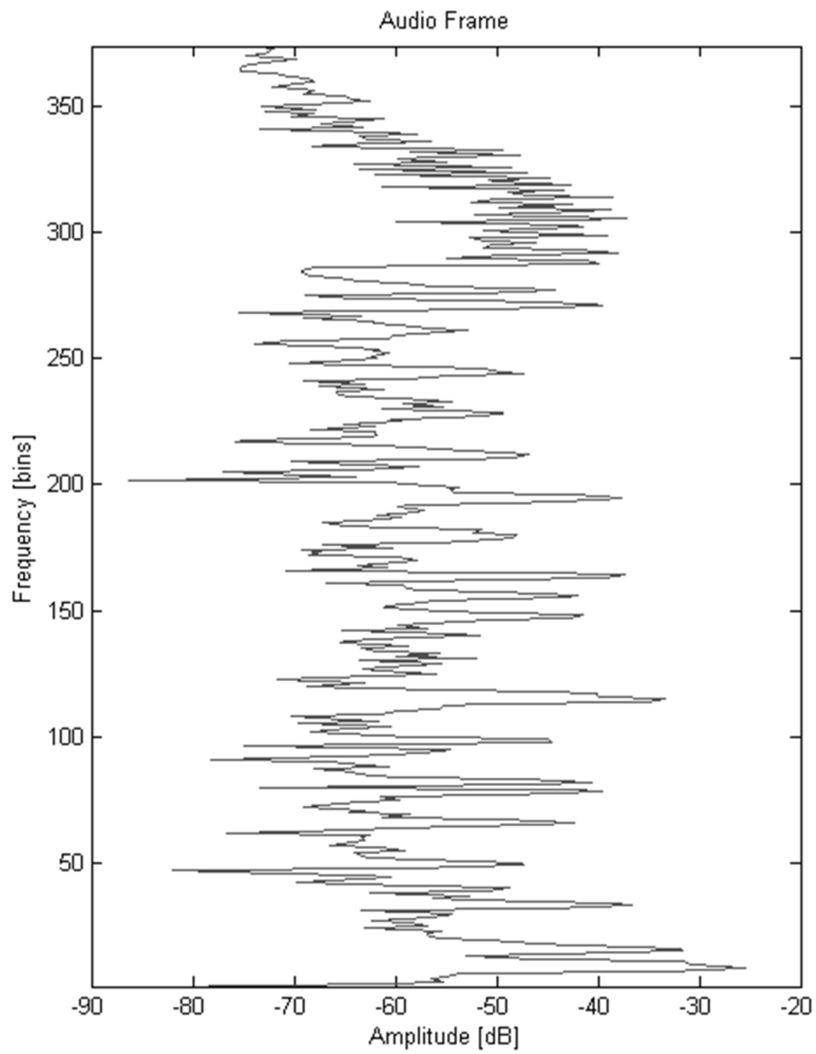


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.



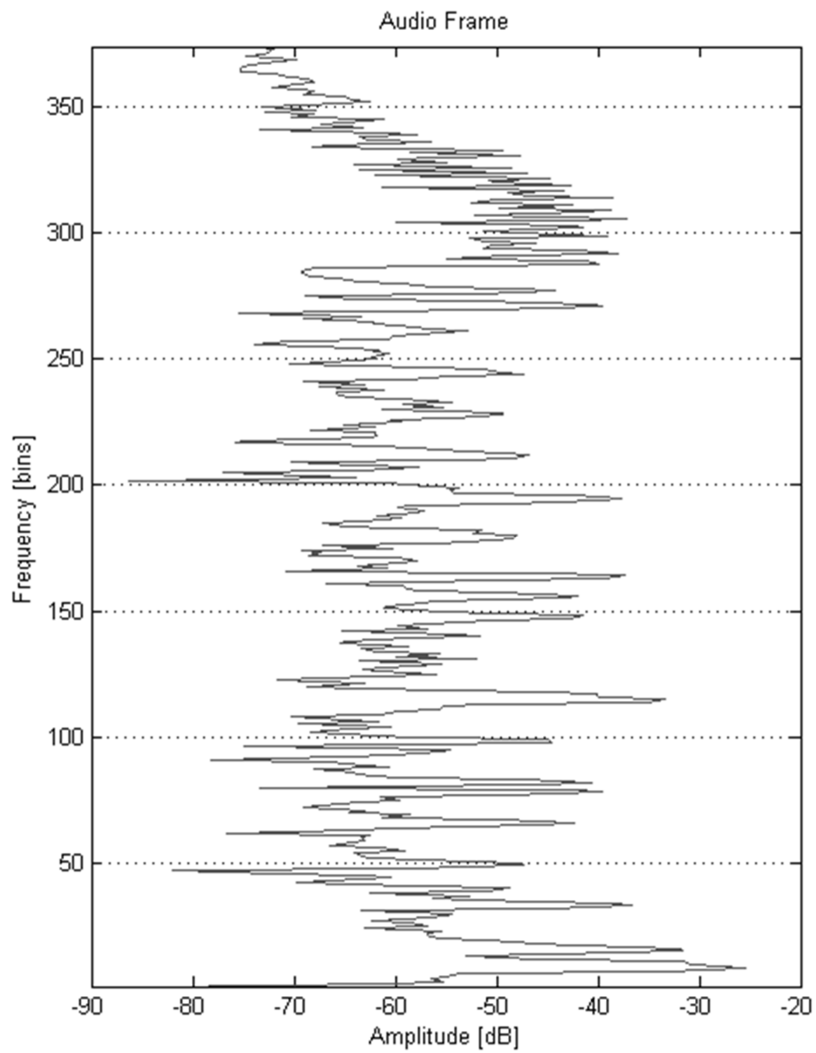
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.



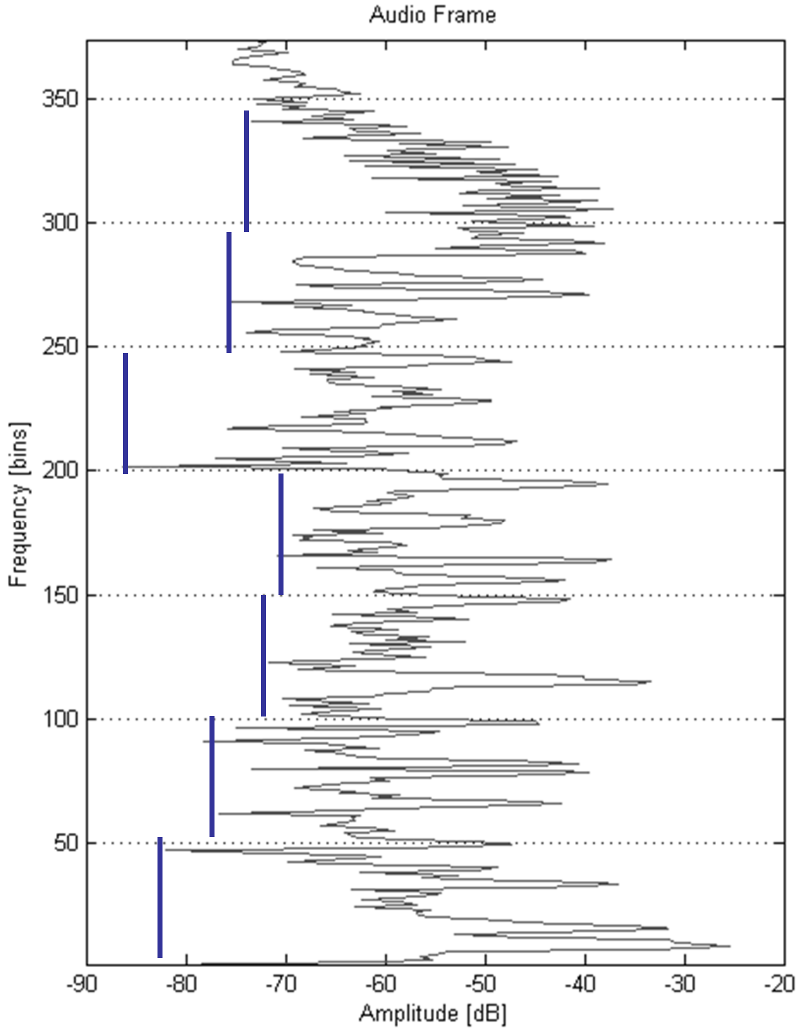
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.

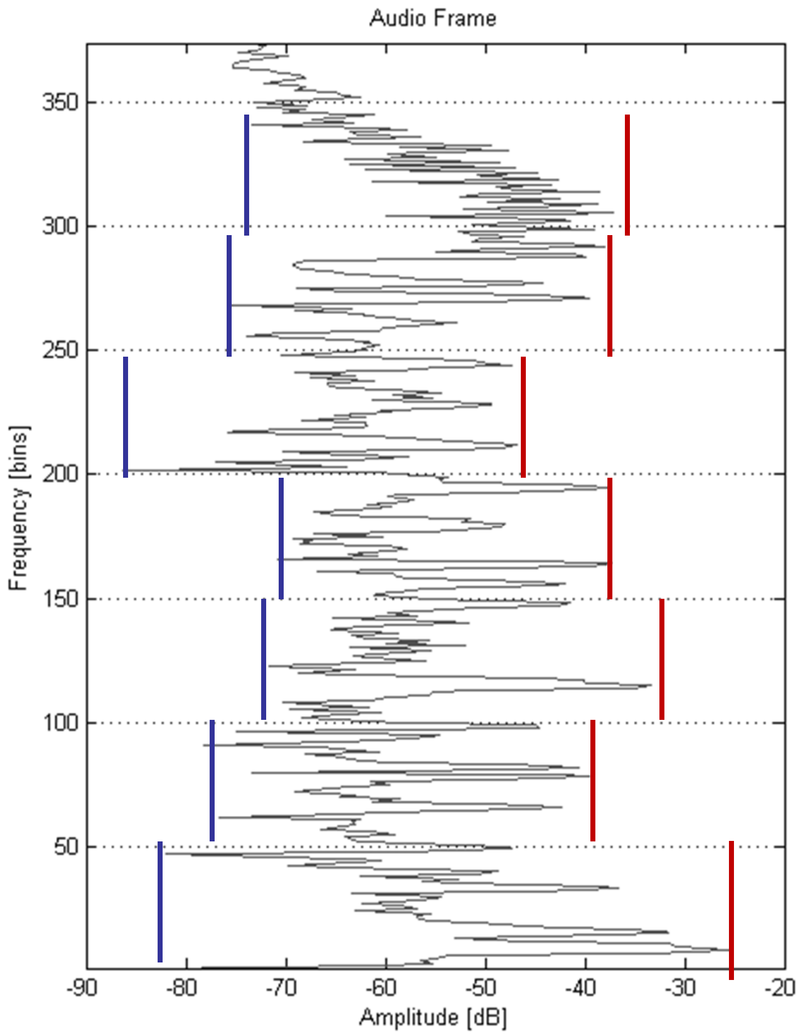


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.

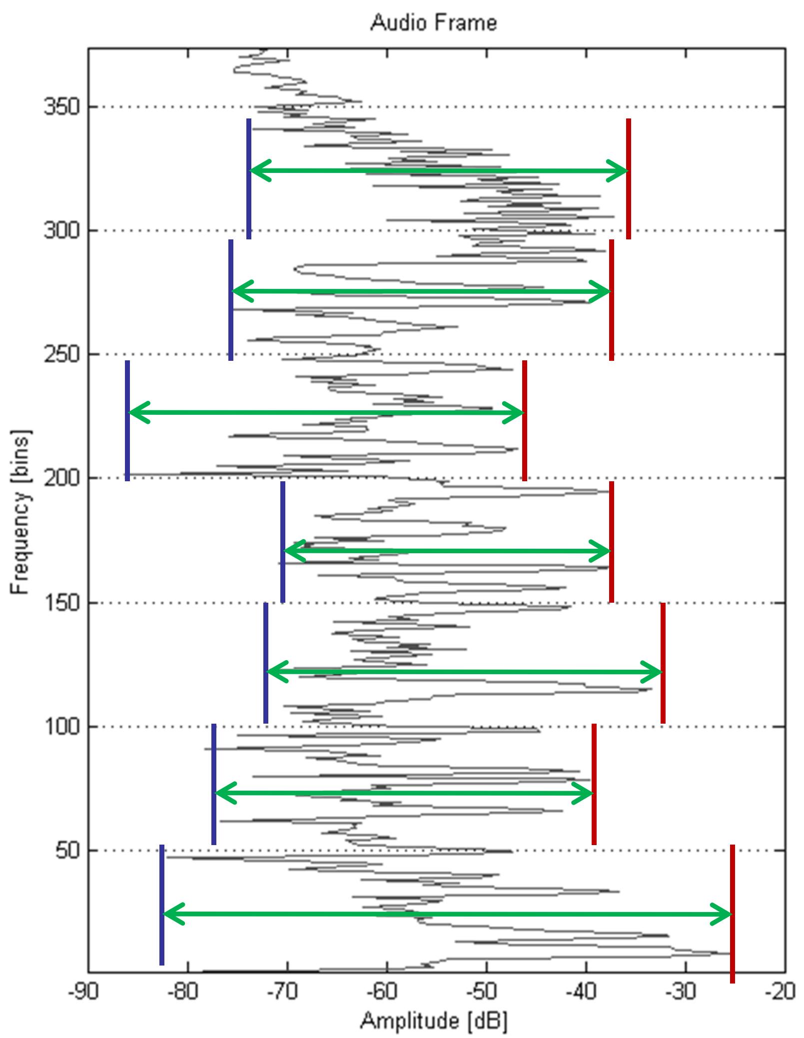


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.

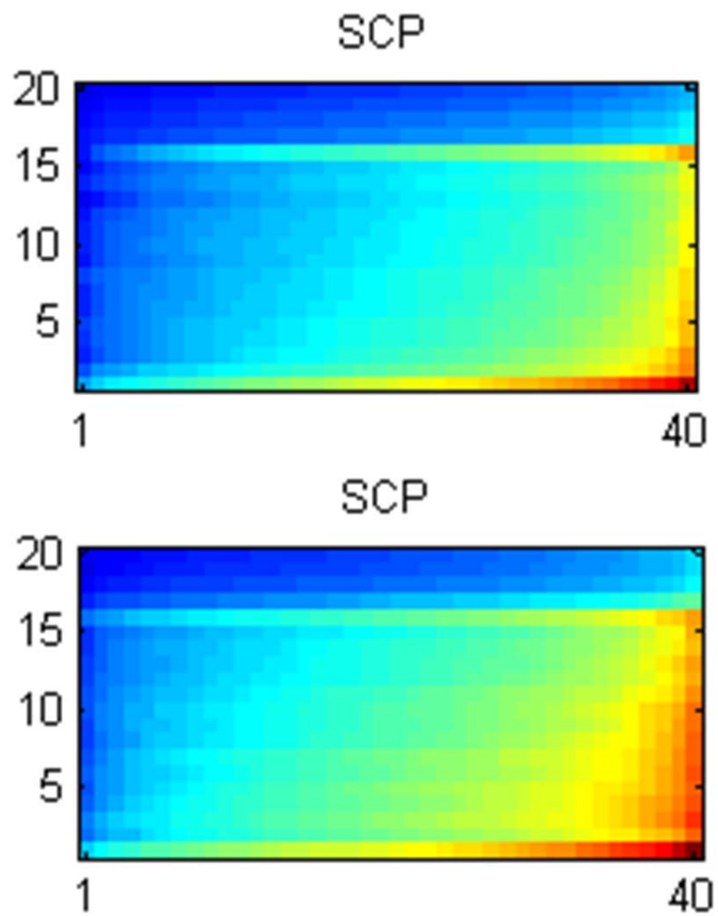


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

Spectral Contrast Pattern (SCP)

- Compute the **spectral contrast** per frame to estimate the ``*tone-ness*´´.
- This is performed separately for 20 frequency bands of the Cent spectrum.
- Sort the spectral contrast values of each frequency band along the whole block.
- The aggregation function is the 0.1-percentile.



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Defining Similarities in the BLF

How to estimate song similarities for multiple block-level features?

D(S1,S2) =**?**

Two music similarity algorithms possible:

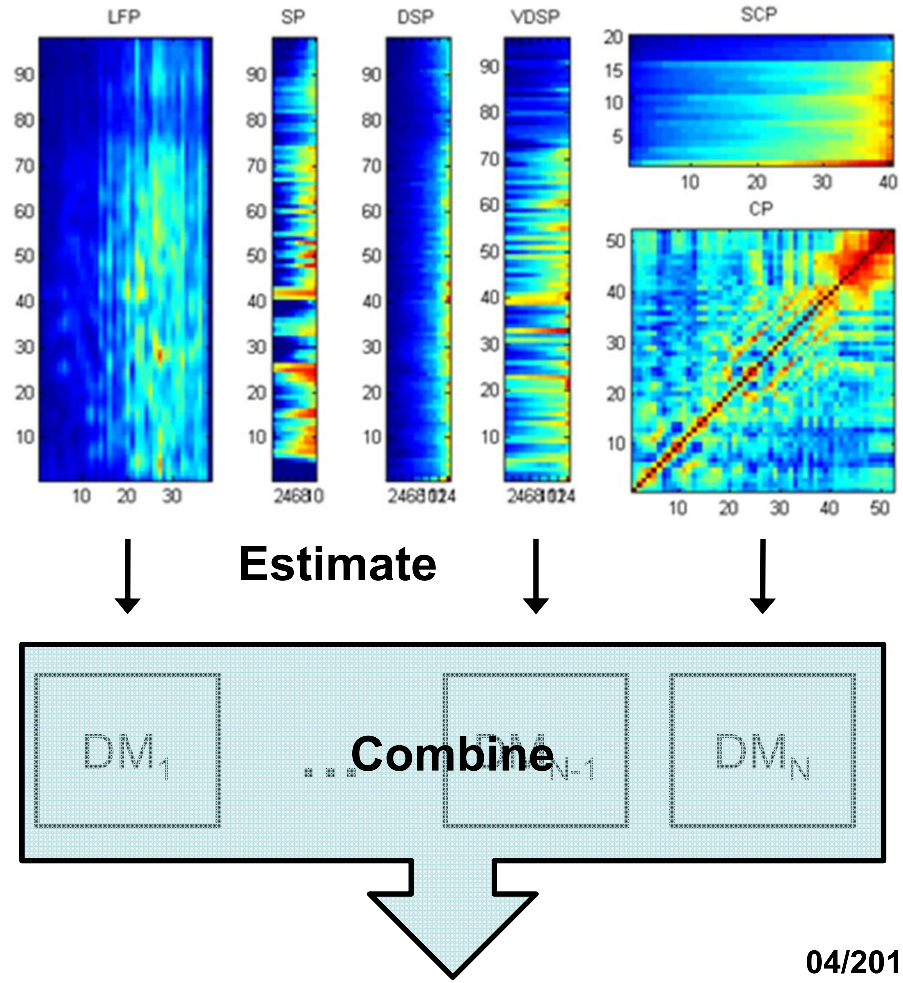
- The first one is based on directly comparing and combining the block-level features (**BLS**).
- The second similarity measure performs a mapping over a semantic tag space (**TAG**).

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

BLS: Fusing Features

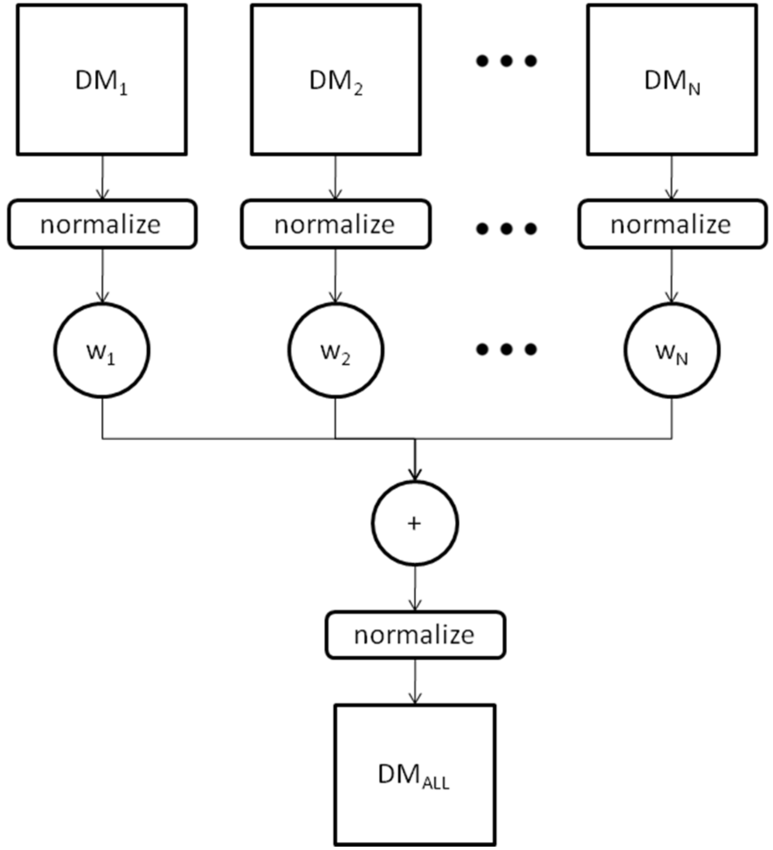


- **Estimate** song similarities separately for each pattern (by computing **Manhattan distance**).
- **Combine** the similarity estimates of the individual patterns into a single result.

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

BLS: Combining Estimates



Naïve approach: linearly weighted combination of BLFs

Problem: Similarity estimates of the different patterns (block-level features) have different scales.

→ special normalization strategy is used:
Distance Space Normalization

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

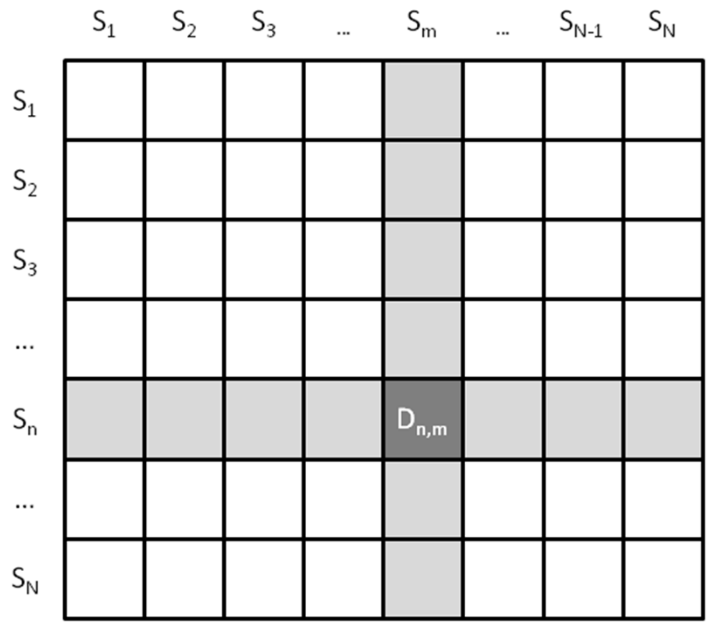
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

BLS: Distance Space Normalization (DSN)

- Operates on the distance matrix
- Each distance $D_{n,m}$ is normalized using Gaussian normalization.
- Mean and standard deviation are computed over both **column** and **row** of the distance matrix.
- Each distance has its own normalization parameters.

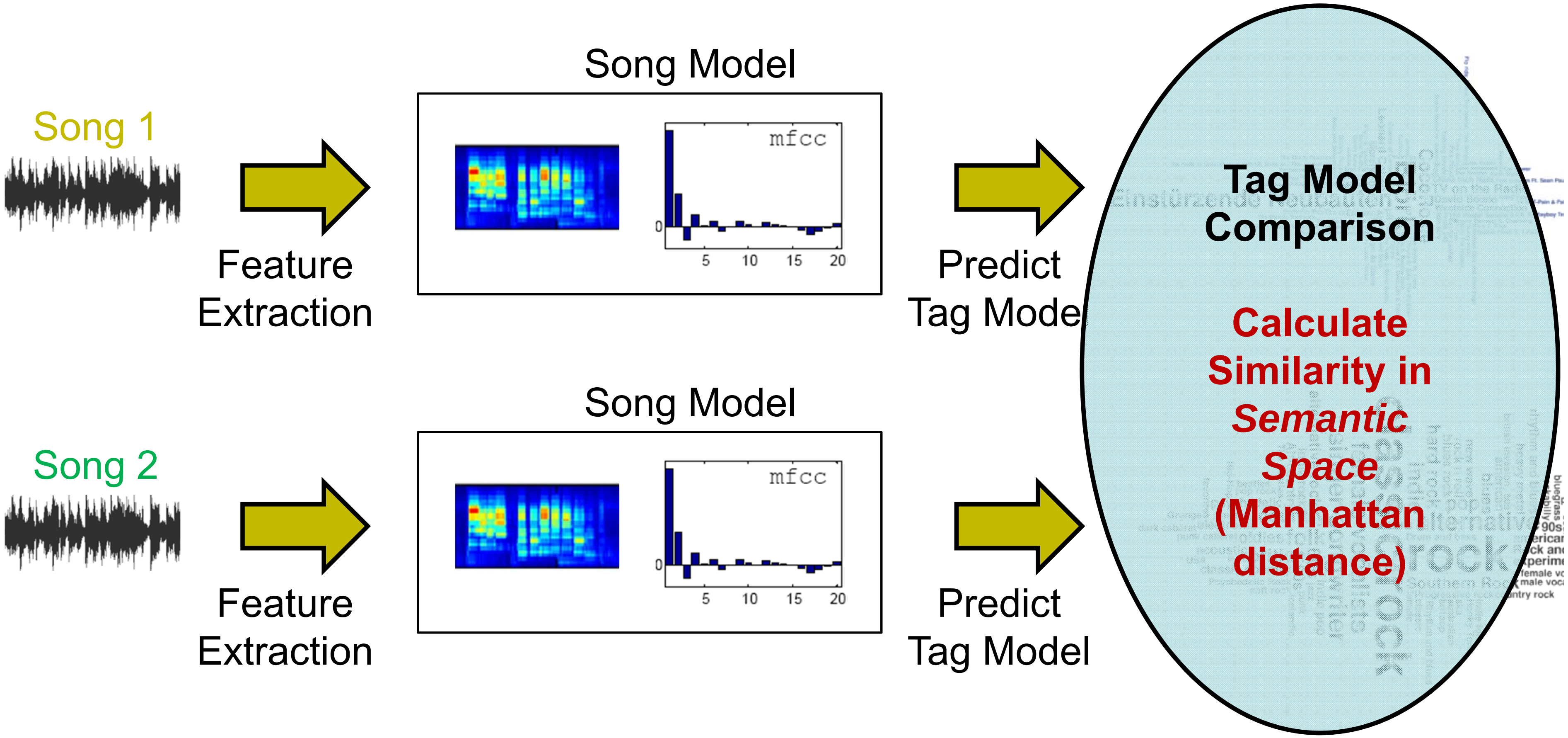
- **Observation:** The operation itself can improve the nearest neighbor classification accuracy.



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

High-Level Feature Extractor: Block-Level Framework (BLF)

TAG: The Semantic Space Approach



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

nepTune

Browsing Music Collections...

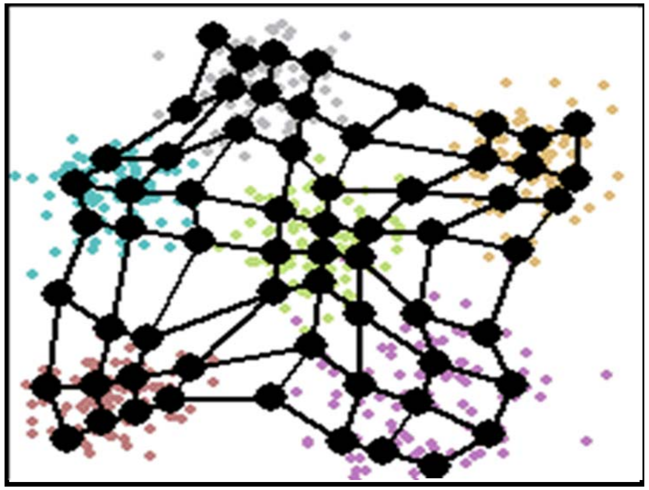
...can be fun

CP@JKU, 2006

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

nepTune

Clustering of music pieces

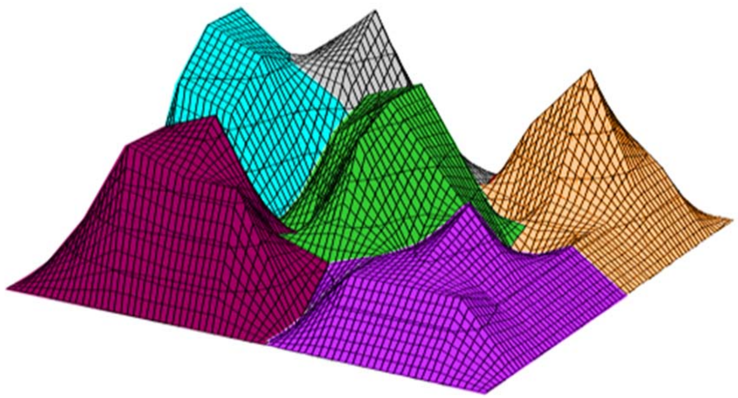


Each song corresponds to point in feature (similarity) space

Self-organizing Map

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

Number of pieces per cluster → landscape height profile

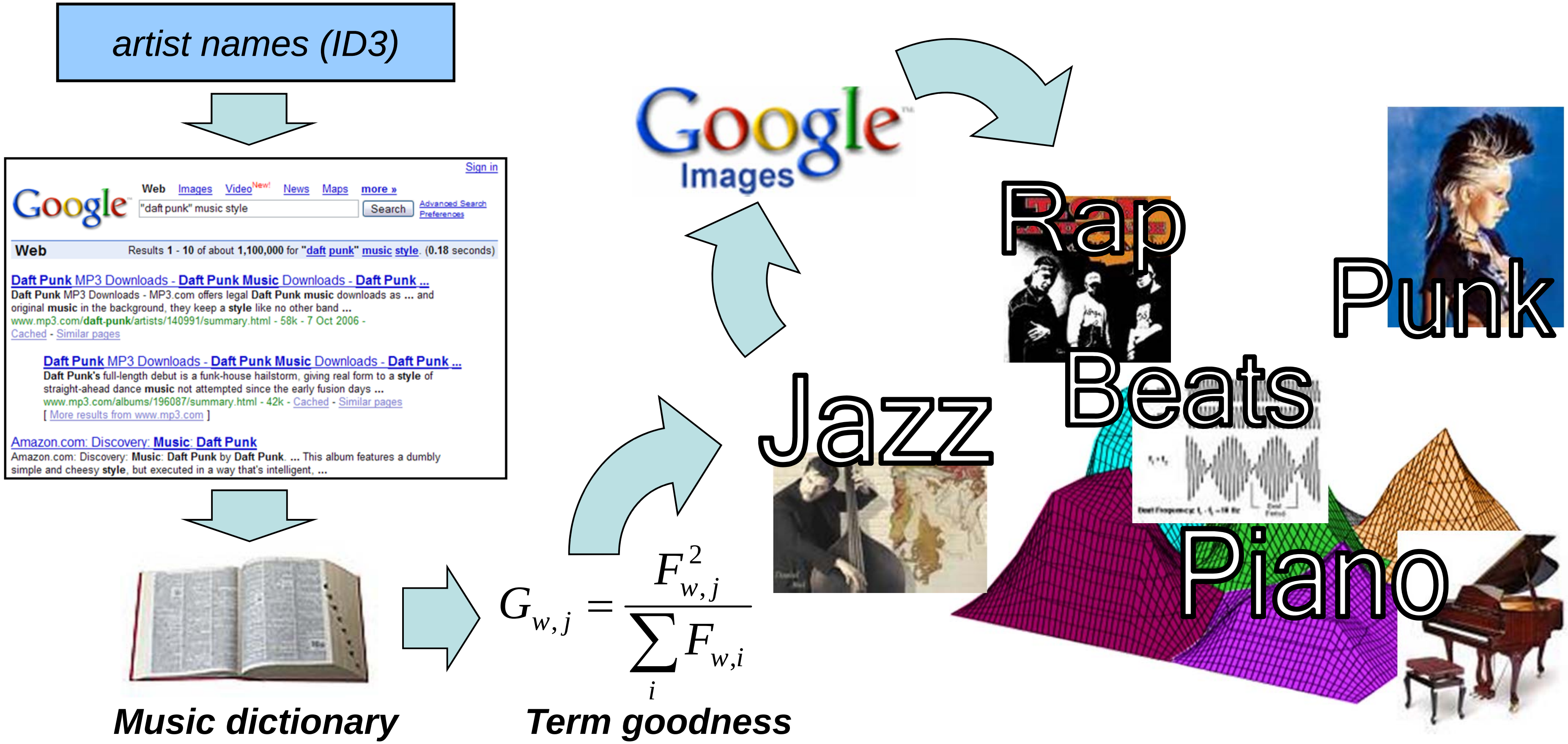


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

nepTune

Automatic description of landscape via Web term extraction



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

nepTune: Demo Video

Music Information Retrieval 2.0

Part II – Context-based MIR

Markus Schedl
Peter Knees

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

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Context- and Web-based Methods

"Music is more than just the signal"

"Similarity" as perceived by humans is not exclusively determined by the audio but also by the outside world, i.e., the context

Influenced by, e.g., marketing strategies, political views of artists, language of songs, decade of activity, etc.
(also listening context, mood, peers – see later)

Analysis of audio alone can't capture these aspects

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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 a central concept

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Text-based Approaches

Data sources:

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

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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

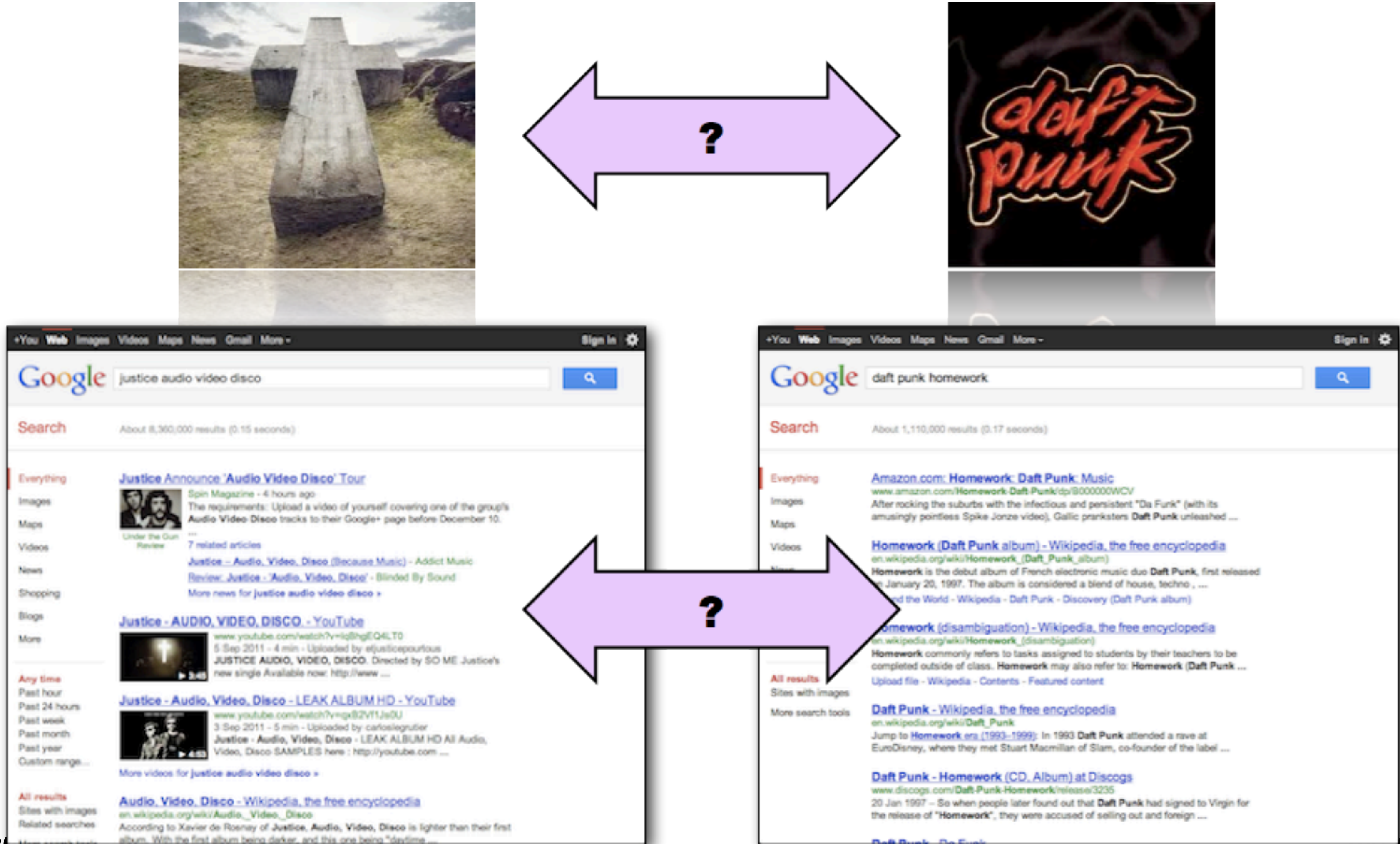
Large range of possible similarity measures

- Overlap, Manhattan, Euclidean, Cosine, etc.

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Related Web Pages as Text Source

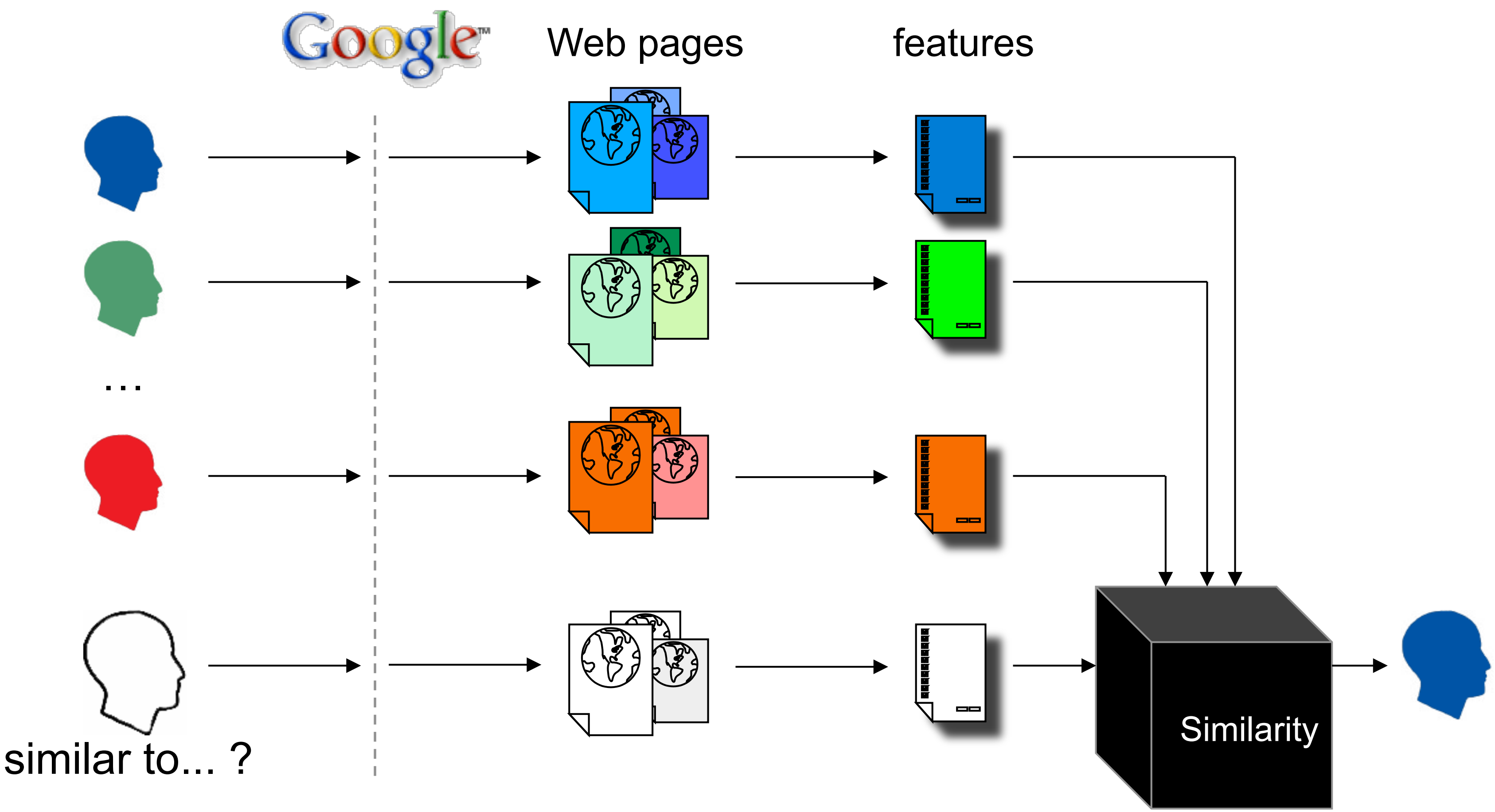


Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Related Web Pages as Text Source



Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Related Web Pages as Text Source

Using search engines and queries such as
"*artist*" +music
"*artist*" +music +review
[Whitman and Lawrence, ICMC 2002] [Baumann and Hummel, WEDELMUSIC 2003] [Knees et al., ISMIR 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)

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- term frequency

Abbr.	Description	Formulation
TF_A	Formulation used for binary match SB = b	$r_{d,t} = \begin{cases} 1 & \text{if } t \in T_d \\ 0 & \text{otherwise} \end{cases}$
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].	$r_{d,t} = \frac{(k_1+1) \cdot f_{d,t}}{f_{d,t} + k_1 \cdot \left[(1-b) + b \cdot \frac{W_d}{av_{d \in D}(W_d)}\right]}$ $k_1 = 1.2, b = 0.75$

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- **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		$w_t = \left(\max_{t' \in T} n_{t'}\right) - n_t$
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 JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- 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 JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- 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.	$\sum_{t \in T_d} r_{d,t} = 1$
NORM_MAX	Normalize maximum of each virtual document's term feature vector to 1.	$\max_{t \in T_d} r_{d,t} = 1$

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- **term frequency**
- **inverse document freq**
- **virtual document mode**
  concatenate all Web pag
- **normalization with res**
- **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	$S_{d_1,d_2} = \left(\max_{d_1',d_2'} (D_{d_1',d_2'})\right) - D_{d_1,d_2}$ $D(F,G) = \sum_i \left(f_i \log \frac{f_i}{m_i} + g_i \log \frac{g_i}{m_i}\right)$ $m_i = \frac{f_i + g_i}{2}$

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- **term frequency**
- **inverse document frequency**
- **virtual document modeling:**
  concatenate all \
- **normalization w**
- **similarity meas**
- **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).

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- **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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Large-Scale Study

[Schedl et al., TOIS 2011]

Investigating different aspects in modeling artist term profiles from Web pages (9,200 experiments):

- **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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Interesting Findings [Schedl et al., TOIS 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:

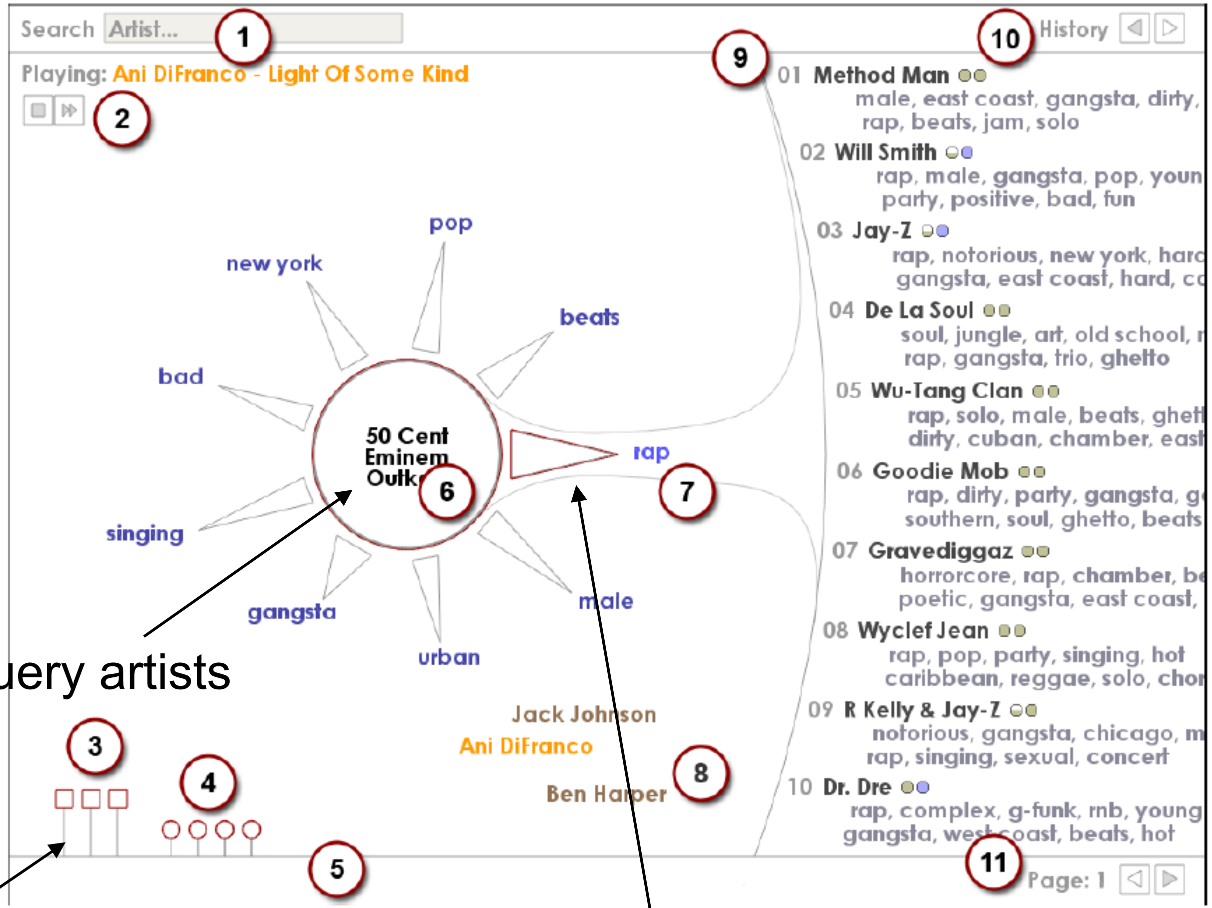
  `TF_C3.IDF_I`
  `TF_C3.IDF_H` → logarithmic formulations for TF and IDF
  `TF_C2.IDF_I`

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Web-based Descriptions for Browsing

"MusicSun"
[Pampalk and Goto, ISMIR'07]

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



Query artists

Relevant word dimension

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

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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., ISMIR 2006]:
Crawl list of RSS feeds and use Weblog entries to index pieces

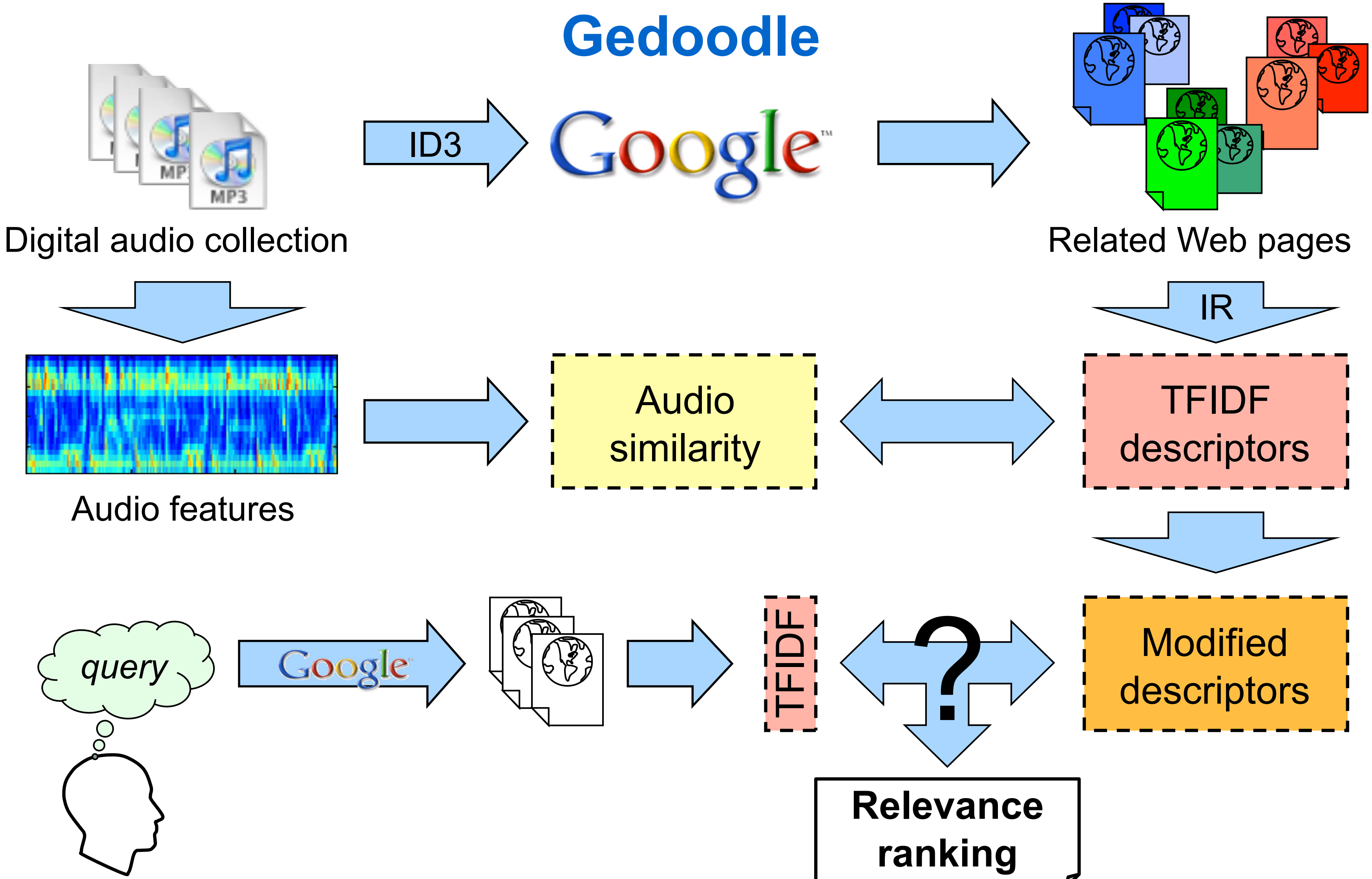
Squiggle [Celino et al, SAMT 2006]:
Combines meta-data databases (like MusicBrainz) for rich indexing

Gedoodle [Knees et al., SIGIR 2007]:
Query Google and combine Web pages to index pieces

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Gedoodle



Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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

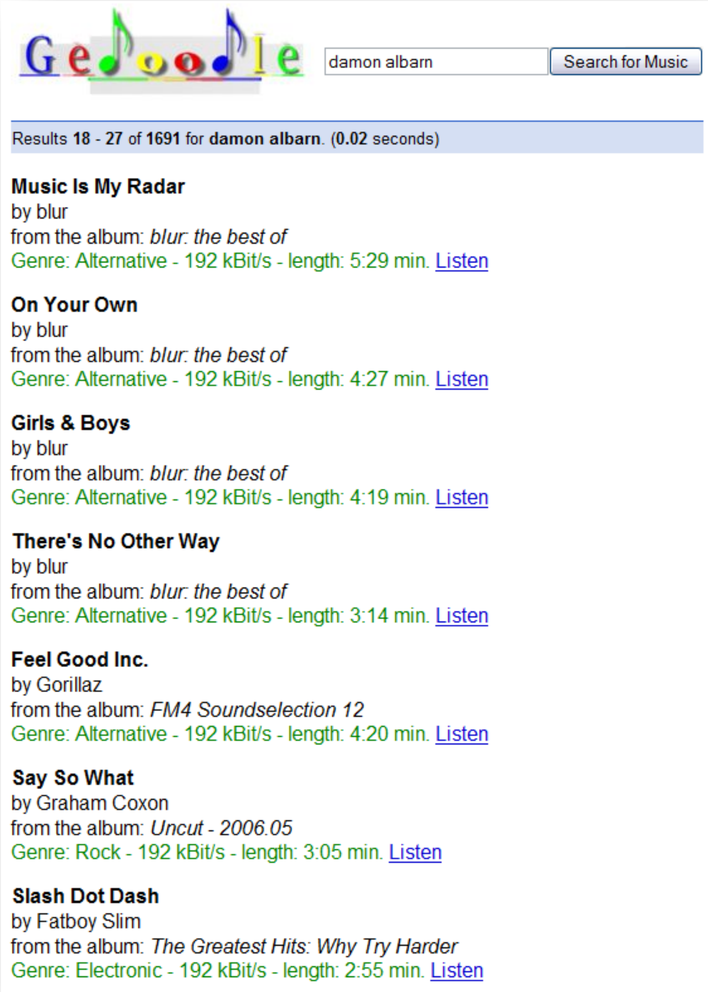
Create normalized TFIDF vector for each track

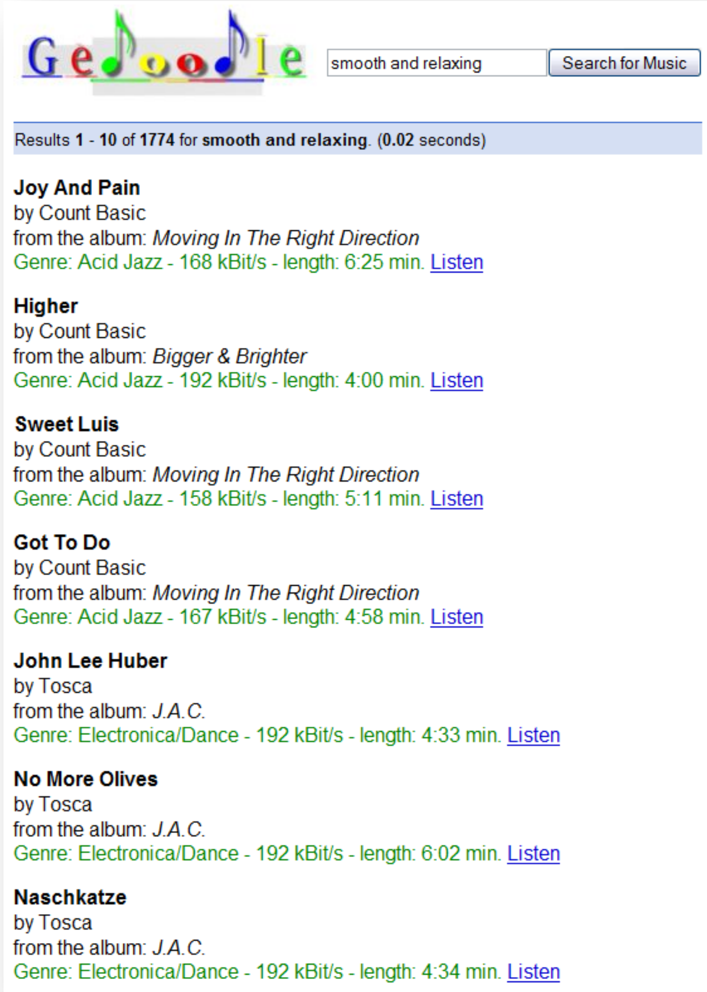
Include audio similarity for vector modification and dimensionality reduction

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Gedoodle (Example queries)





Department of Computational Perception
JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Music Information Extraction from Web Pages

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

Apply methods from NLP, IE, Named Entity Detection to extract data

- **Genres, Moods, Similarities** using Rule Patterns [Geleijnse and Korst, ISMIR'06]

- **Band Members and Line-Up** using Rule Patterns [Schedl and Widmer, AMR'07]

- **Band Members, Discography, Artist Detection** using rules in GATE [Krenmair, JKU 2010]

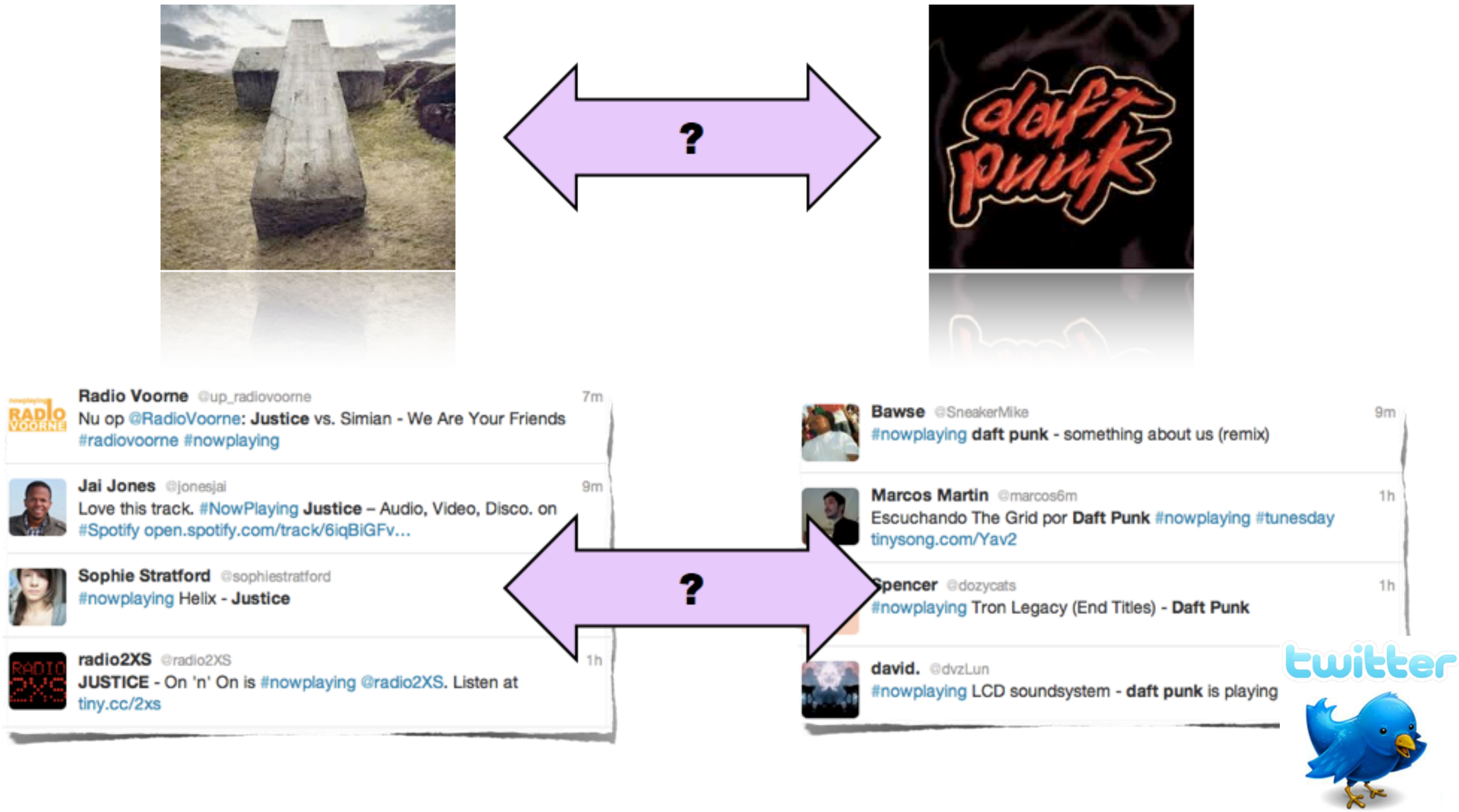
- **Band Members, Discography** using Supervised Learning (in GATE) [Knees and Schedl, WOMRAD 2011]

- **Album cover** detection and extraction [Schedl et al., ECIR 2006]

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Microblogs as Text Sources



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

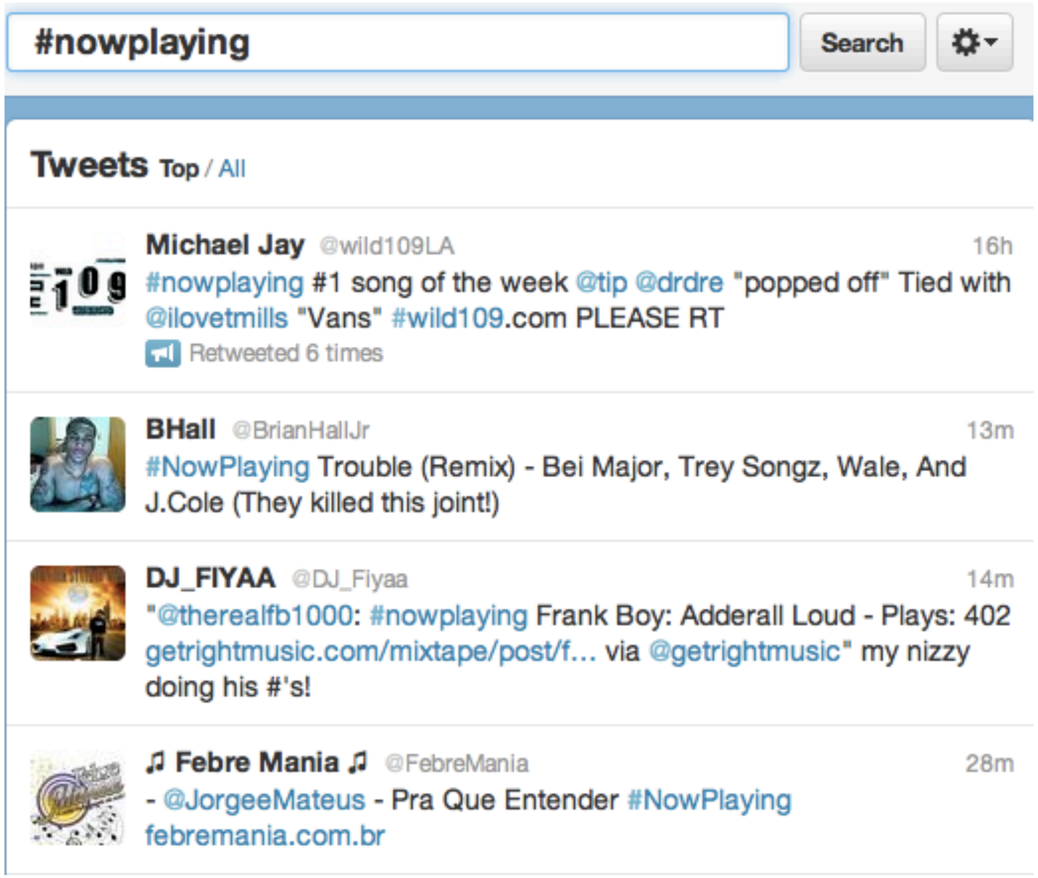
Microblogs as Text Sources

Retrieving tweets from *twitter.com* using hashtags such as
- *#nowplaying*
- *#np*
- *#itunes* [Schedl et al., ISMIR 2011]

Highly compressed, non-natural language due to 140 character restriction

Artist/track identification and matching often challenging

#nowplaying also used for movies



Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Interesting Findings on Large-Scale Evaluation

[Schedl, IR 2012]

- modeling artists as ***virtual documents*** is preferable

- using query scheme ***"artist"*** outperforms ***"artist" +music*** (sparsity!)

- ***normalization*** does not improve results! → 140-character limit

- ***term set*** has an important influence: ***TS_A*** best, but at highest cost!
  TS_N good compromise

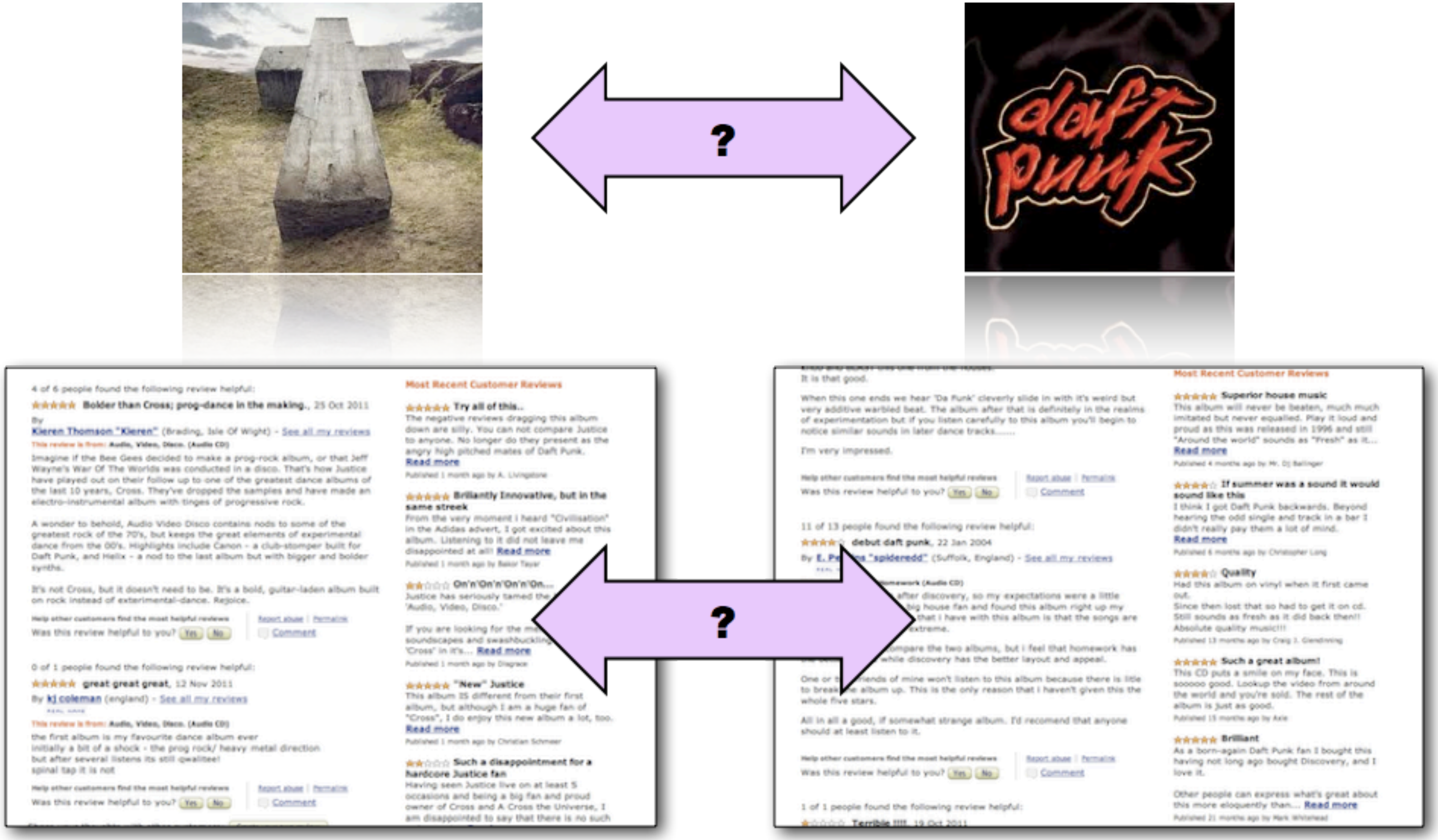
- simple ***inner product*** as similarity measure performed better than cosine
  → consistent with observation for normalization

- no clear picture for TF and IDF variants (just don't use binary match!)
  TF_C2 (log) and TF_E performed best, on average
  IDF_B2 (log), IDF_J (Okapi), IDF_E (log) performed best, on average

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Product Reviews as Text Sources



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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

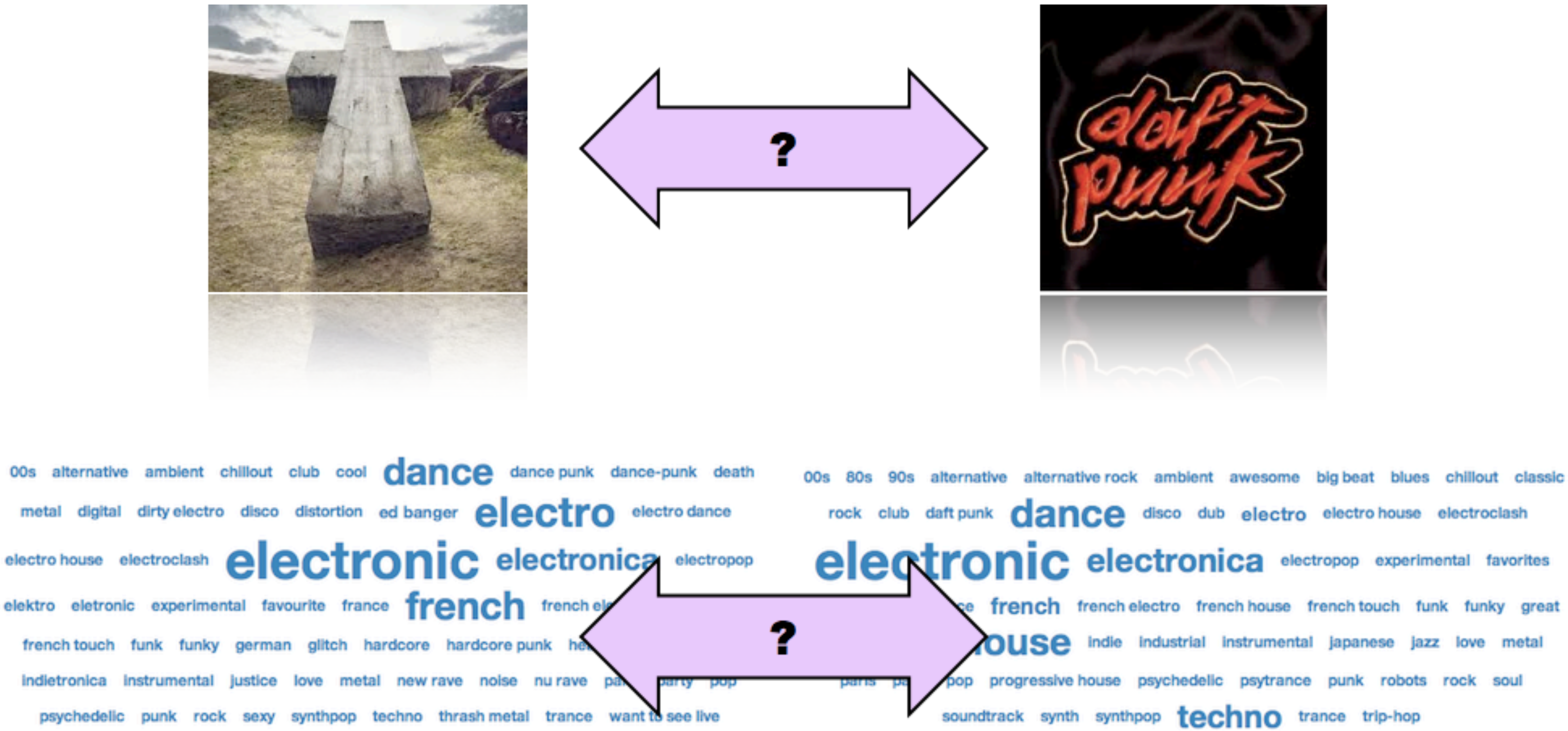
Very prone to attacks (remedy: consider “helpfulness” ratings)

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Community Tags as Text Sources

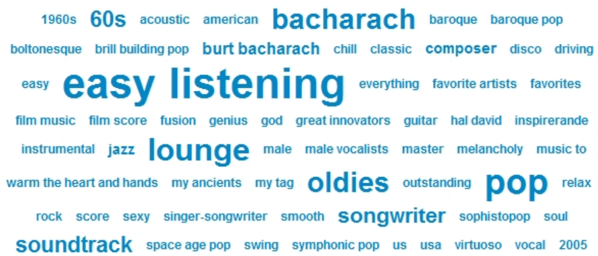


Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

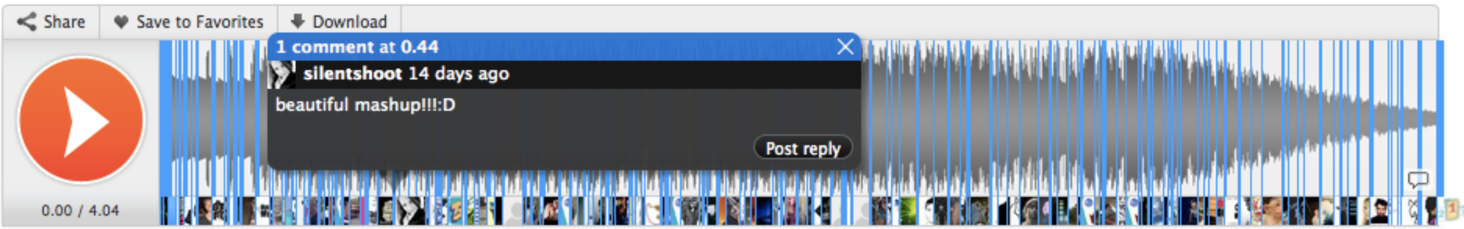
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Tag Sources

- Community
  e.g., last.fm



- Soundcloud (annotations along timeline)



- Games with a purpose
  e.g., Tag-a-Tune
  [Law and von Ahn, CHI'09]

- Autotags (see before)

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Community Tags as Text Sources

Treating collections of tags (e.g., from last.fm) as documents
[Pohle et al., ISMIR 2007] [Levy and Sandler, JNMR 2008] [Hu et al., ISMIR 2009]

- Retrieve tags for artist or track from last.fm
- Cleaning of noisy and redundant tags
  - manually
  - automatically [Geleijnse et al., ISMIR 2007]
- List of collected terms is treated as text document and TF·IDF'd [Levy & Sandler, ISMIR 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 (13,500 tags vs. >200,000 Web terms [Levy & Sandler, ISMIR 2007])
- Depends on community, needs annotators
- Hacking and Attacks!

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Lyrics as Text Source

Topic Features

[Logan et al., ICME 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., ISMIR 2008], [Hu et al., ISMIR 2009]

Rhyme Features [Mayer et al., ACM MM 2008] [Hirjee and Brown, ISMIR 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., ACM MM 2005] [Hirjee and Brown, ISMIR 2009]

- Language, structure

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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 JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Similarity from Co-Occurrences

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

Data sources:

- 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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Search Engine Page Count Estimates

[Schedl et al., CBMI 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 compl.)

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

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Shared Folders in a P2P Network

Make use of meta-data transmitted as files names or ID3 tags in P2P network OpenNap [Whitman and Lawrence, ICMC 2002] [Ellis et al., ISMIR 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 and Weinsberg, AdMIRe 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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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., CSCW'94]

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 and yet unknown tracks

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Playlist Co-Occurrences

Analysis of co-occurrences of artists and songs on radio station playlists (French radio station Fip) and compilation CD databases (CDDB) [Pachet et al., WEDELMUSIC 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 29,000 playlists from “Art of the Mix” [Cano and Koppenberger, ISMIR‘04]: artists similar if they co-occur in playlist (highly sparse)

Analysis of > 1 million playlists from “MusicStrands” [Baccigalupo et al., ISMIR 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
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

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
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Current and Future Challenges

Improving content- and context-based methods (scale and speed)

Most urgent challenge: Modelling **User Context**

Perception of music is highly subjective

Influenced by personal experiences, preferences, taste, etc.

Analysis of audio and Web data alone can't capture these aspects

Essential for truly personalized recommendations
(in contrast to e.g., collaborative filtering)

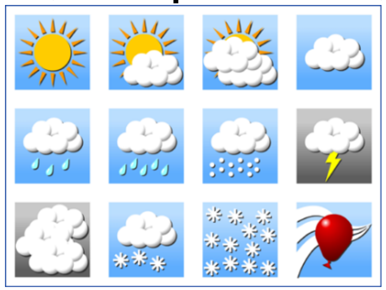
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Dimensions of User Context

adapted from [Göker and Myrhaug, 2002]

Environmental Context: people & things surrounding user, temperature, noise, light, humidity



Personal Context:

- *physiological context*: weight, blood pressure, pulse, …
- *mental context*: stress level, mood, expertise, …

Task Context: current activities, running apps

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

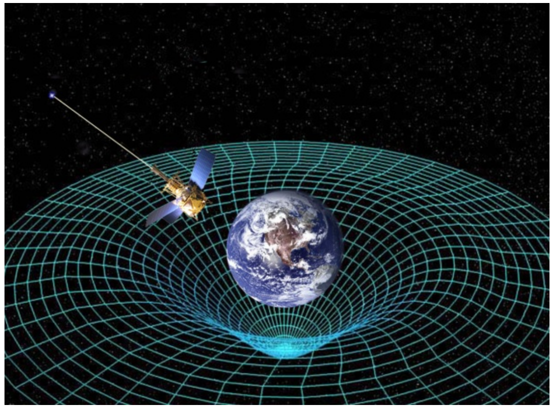
Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Dimensions of User Context

adapted from [Göker and Myrhaug, 2002]

Social Context: relatives, friends, enemies, collaborators



Spatio-temporal Context: location, place, direction, speed, and time

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ
Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

User-Aware Music Recommendation on Android Phones

(Ongoing Master's Thesis G. Breitschopf)

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

- collecting user context data while playing (next slides)
- adaptive system that learns user taste/preferences from implicit feedback (player interaction)
- two basic research questions:
  + Can we solve a binary classification problem (like/dislike) from context features and feedback? (assuming skipping indicates <dislike>, actively selecting a song indicates <like>)
  + Can we relate user context features to music the user wants to listen to?
- ultimate aim: dynamically and seamlessly update the user's playlist according to his/her current context

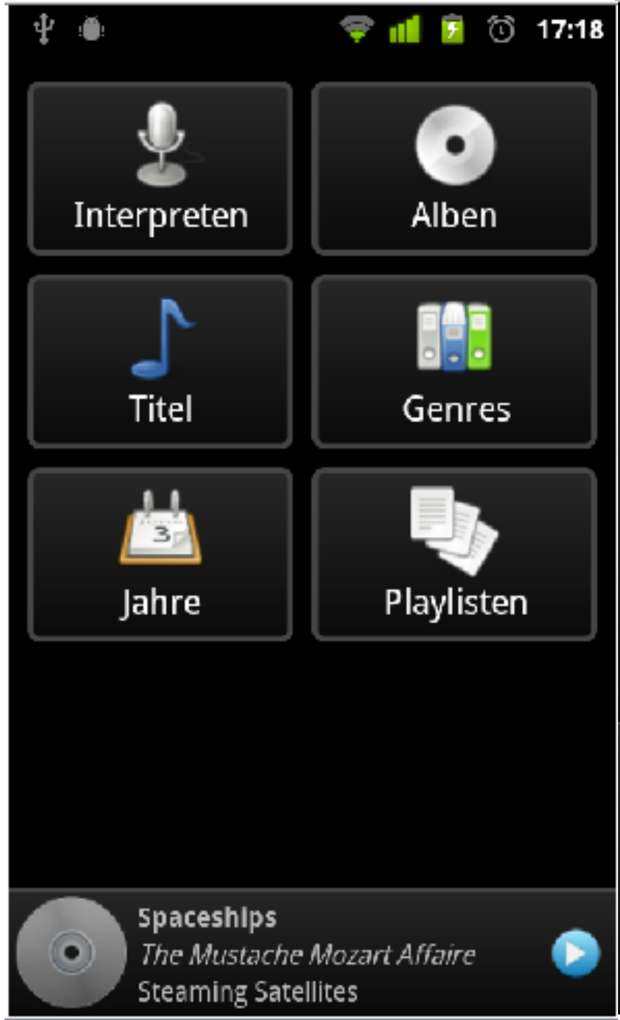
Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

User-Aware Music Recommendation on Android Phones

(Ongoing Master's Thesis G. Breitschopf)



Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees
{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

User Context Features from Android Phones

Time:	timestamp, time zone
Personal:	userID/eMail, gender, birthdate
Device:	deviceID (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, UI mode (undocked, car, desk), 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)

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Currently in Evaluation…

Which granularity/abstraction level to choose for representation/learning?

user context data:

e.g., position as GPS coordinates or nearest agglomeration

→ data transformation/aggregation

musical metadata:

artist/performer, album, track, genre, mood, other tags?

Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Conclusions

Music Information Retrieval is a broad and diverse field

Various approaches to extract information directly from the audio signal

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

Multi-modal retrieval promising and allows for exciting applications

Next big challenges:

- modeling user context
- personalization
- situation-based retrieval

Department of Computational Perception
JKU JOHANNES KEPLER UNIVERSITY LINZ

Markus Schedl, Peter Knees

{markus.schedl, peter.knees}@jku.at | http://www.cp.jku.at

Thank you !

Latest version of the slides will be available at
http://www.cp.jku.at/tutorials/ecir2012_mir20.html