

Music Retrieval and Recommendation

Full Day Tutorial

**Peter Knees
Markus Schedl**

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

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

Overview

Goals of this class

- Introduction to the field of music similarity estimation
- Approaches to music retrieval and recommendation
- Deepening the understanding of the Music IR domain

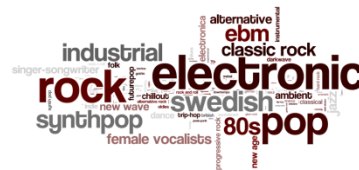
Schedule for today:

9:00 – 10:30: Introduction to Music IR

11:00 – 12:30: Music Content Analysis and Similarity

14:30 – 16:00: Music Context-Based Similarity and Indexing

16:30 – 18:00: Listener-centric and Collaborative Similarity



Who we are



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



Markus Schedl

*Associate Professor of the **Department of Computational Perception, JKU Linz***

M.Sc. in Computer Science from Vienna University of Technology

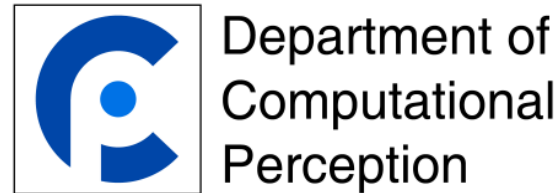
Ph.D. in Computational Perception from Johannes Kepler University Linz

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

Acknowledgements

Support from our institution



Support from two EU-funded projects



giant...steps...



00:15

SYMPHONY NO. 3 IN E FLAT MAJOR, OP. 55 'EROICA'

Ludwig van Beethoven



PHENICX

Performances as Highly Enriched
aNd Interactive Concerts eXperiences

Symphony No. 3

Allegro con brio. (♩. = 60)

Flauti 1, 2

Oboi 1, 2

Clarineti 1, 2
in B \flat

Fagotti 1, 2

Corni 1, 2
in E \flat



FISCHER

SCORE

BAR TO BAR



00:15

- Performances as Highly Enriched aNd Interactive Concert eXperiences
- Aims at making classical concerts appealing to new audiences, in particular, the younger generation
- Social media as a means to create user profiles and elaborate personalized music information and recommendation systems (pre-, during-, post-concert experiences)
- Motivate fans of classical music to use social media

00:15



-54:55



Giant Steps

STRIDING FORWARD IN
ELECTRONIC MUSICAL
CREATIVITY, EXPERIMENTATION
AND PERFORMANCE

- New products for professionals and amateurs in music creation
- Multimodal Music Information Retrieval
- User-centric design

- Features from audio signal and the web
- Beat detection, drum transcription, rhythm
- Extracting expert knowledge from DJ forums

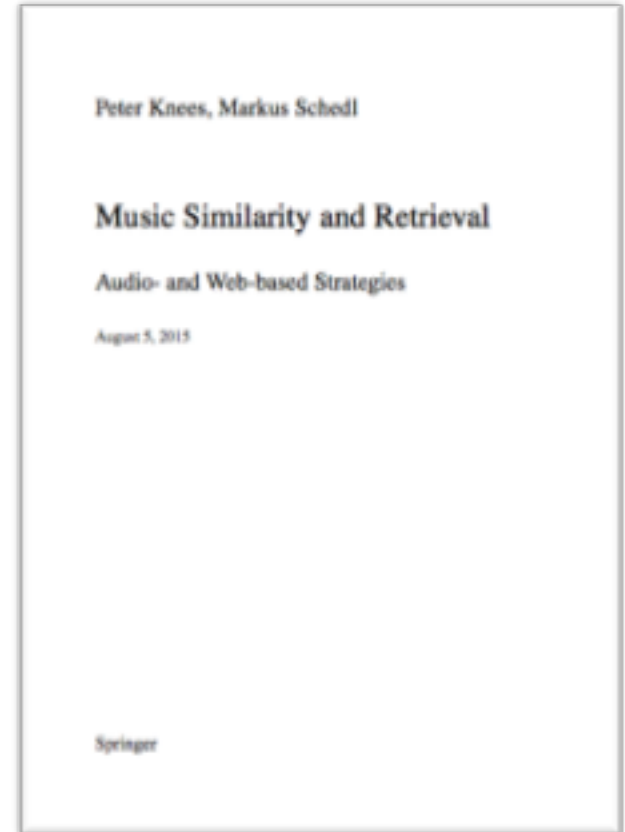


Book accompanying the tutorial

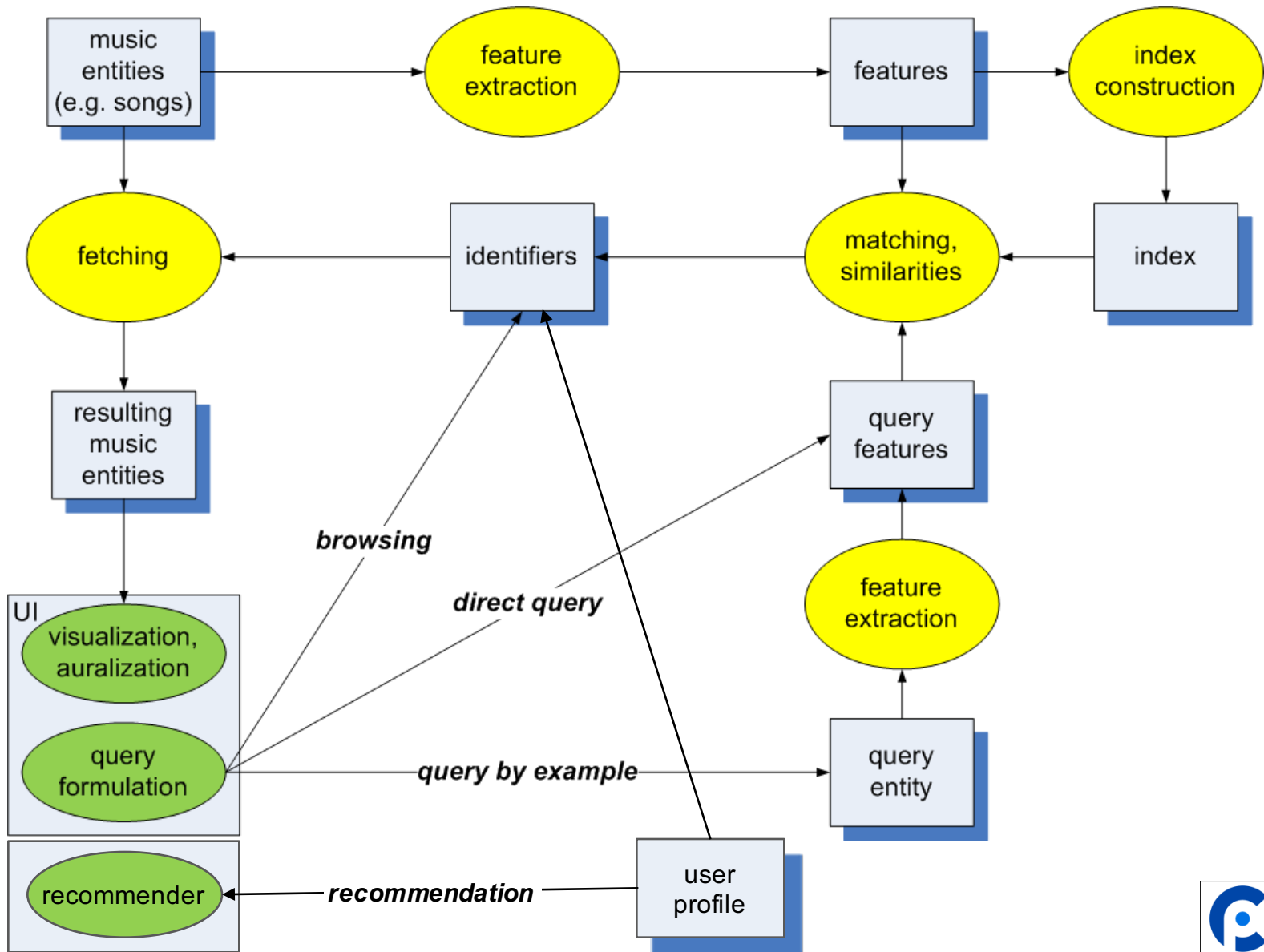
Music Similarity and Retrieval Audio- and Web-based Strategies

Peter Knees and Markus Schedl

To be published by the end of 2015 in
Springer's Information Retrieval series



What is MIR? An Information Retrieval View



Some Definitions of Music IR

“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]

Typical MIR Tasks

- Feature extraction (audio-based vs. context-based approaches)
- Similarity measurement, recommendation, automated playlist generation (last.fm, Pandora, Echo Nest, ...)
- Detection of musical events (onsets, beats, downbeats, key changes, etc.)
- User interfaces, visualization, and interaction
- Audio fingerprinting (copyright infringement detection, music identification services like shazam.com or musicbrainz.org, track identification in music sets)
- Cover song detection
- Voice and instrument recognition and extraction, 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)
- Classification and evaluation (ground truth definitions, quality measurement, e.g. for feature extraction algorithms, genre classification)
- Optical music recognition (OMR)



Applications: Automatic Playlist Generation

“Personalized Radio Stations”

e.g.

- Pandora
- Last.fm
- Spotify Radio
- iTunes Radio
- Google Play Music All Access
- Groove (was: Xbox Music)

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, on mobile players, in the Cloud)
- most UIs of music players still only allow organization and searching by textual properties according 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

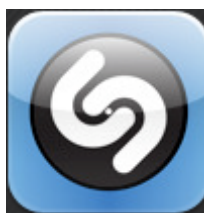
Query-by-humming:

input is not excerpt of a song, but melody hummed by the user

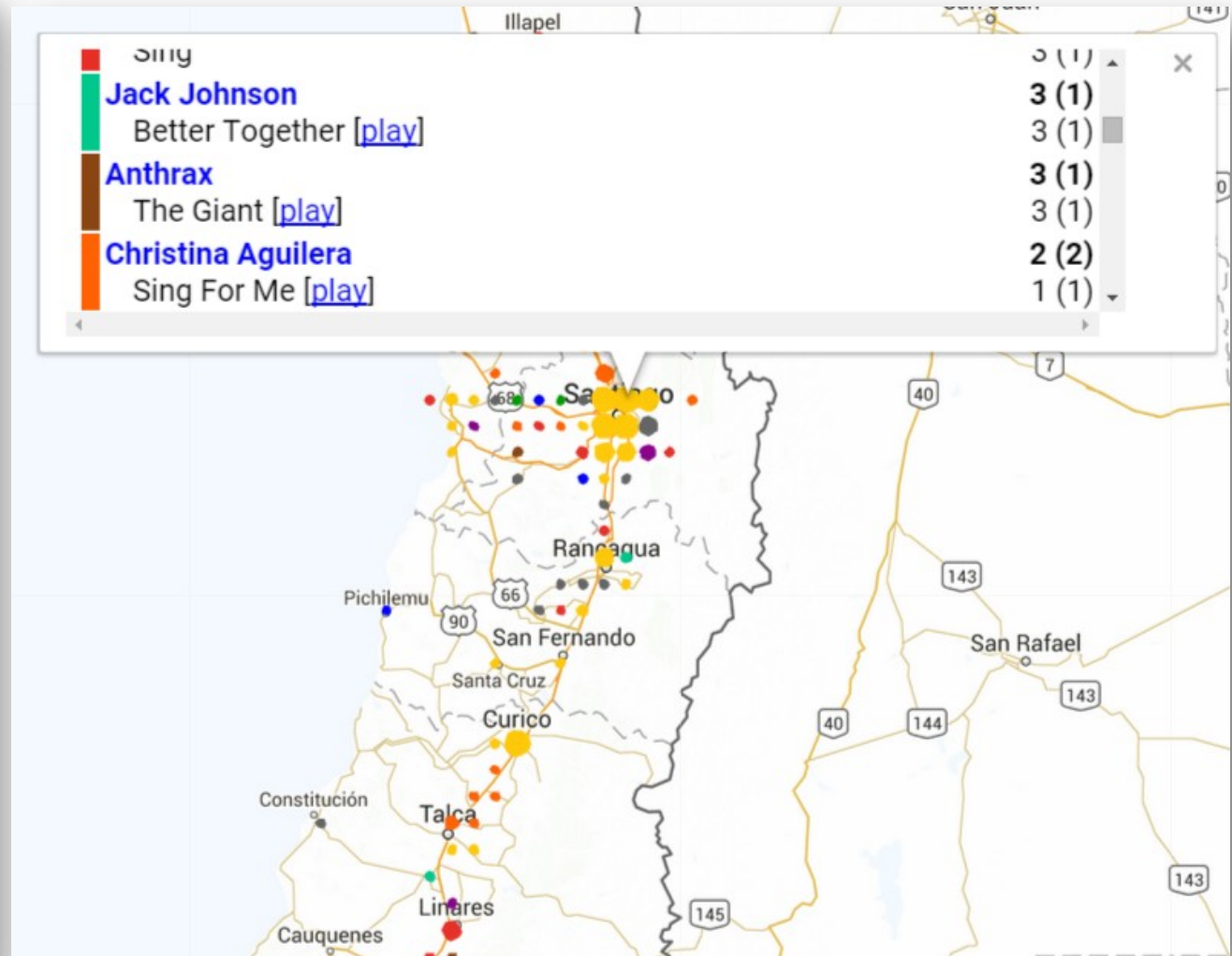
Examples:

www.shazam.com

www.soundhound.com



Applications: Music Tweet Map



Applications: Automatic Accompaniment



Part I

ABOUT MUSIC SIMILARITY

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

Music Similarity: Examples

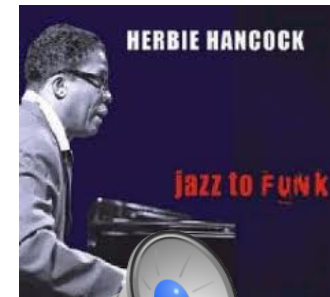
Three different genres?



Which go together?



Which are more similar?



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)

→ 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)

→ Music similarity is highly subjective

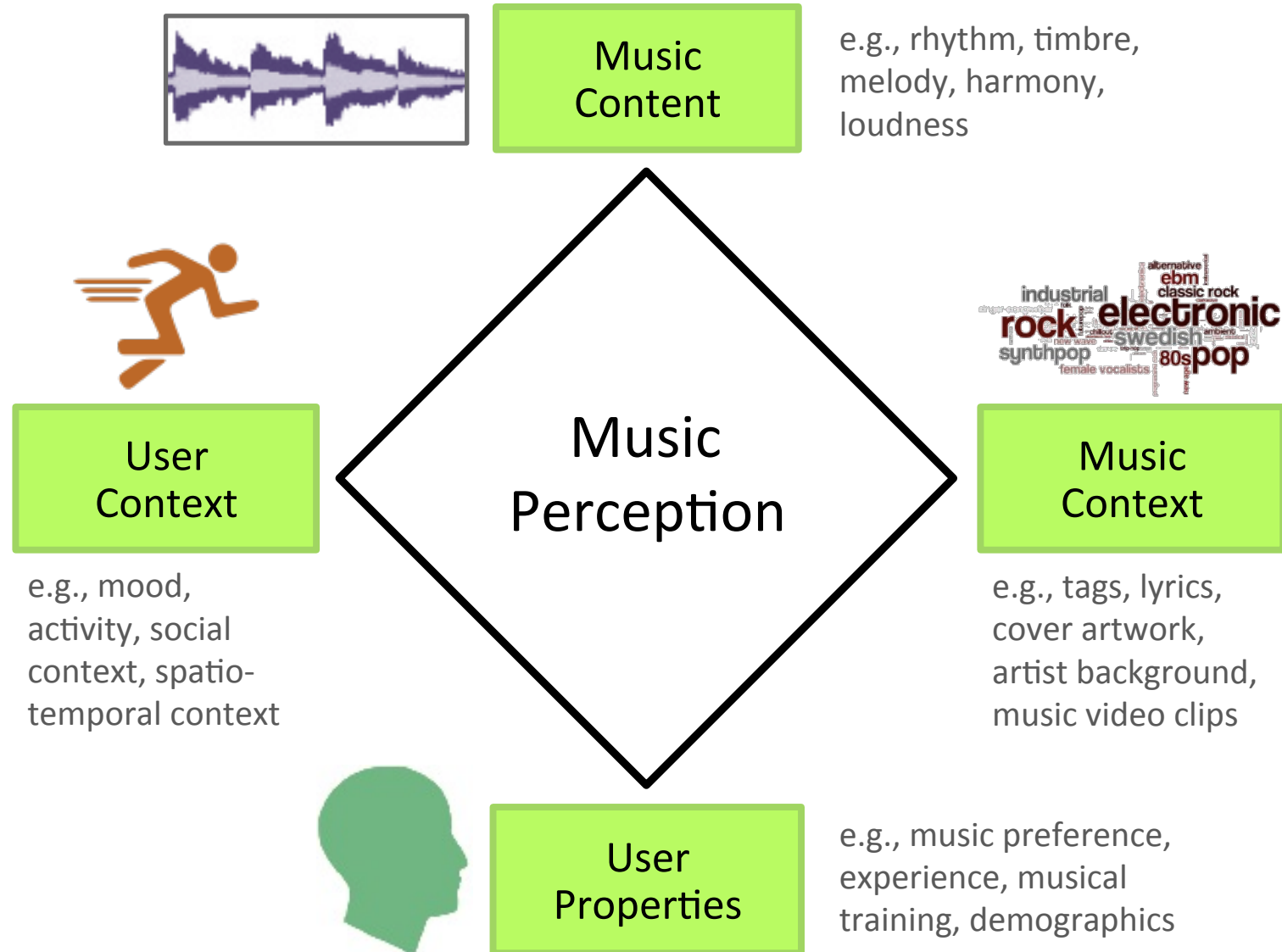
NB: Similarity definition is currently a hot and controversial topic in MIR! (see in a bit)

To the best we can do as of now, computational similarity is obtained by taking into account multiple influencing factors:

audio content — music context — user context — user properties
(the latter two being the most difficult to obtain)



Influences for Music Perception & Similarity



The Semantic Gap in Music



High-level

Musical concepts as perceived by humans



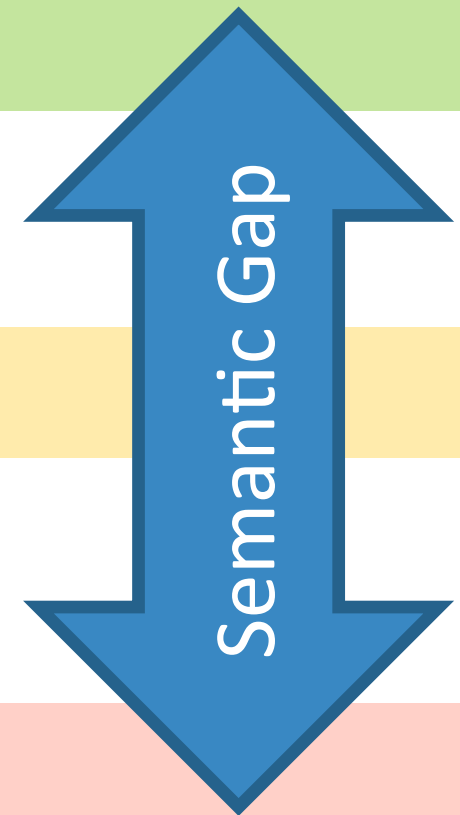
Mid-level

High-level-informed combination of low-level features



Low-level

Statistical descriptions of signal, machine-understandable data



Content vs. Music Context vs. User Context: Quick Overview

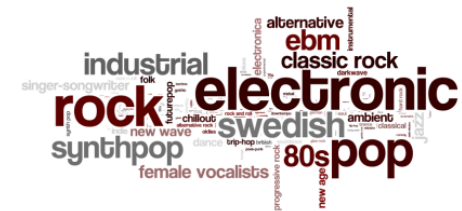
Music Content Analysis (cf. Part II)

- 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.)



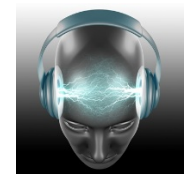
Music Context Analysis (cf. Part III)

- Captures aspects beyond pure audio signal
- No audio file necessary
- Typically textual; resemble high-level features



User Context and Interaction Analysis (cf. Part IV)

- Builds upon the ways people are “using” music
- Collected from implicit or explicit data
- Usually, user-based features are closer to “what users want”



Content vs. Context (Music + Listener)

Challenges for Context-Based Feature Extractors

- Dependence on availability of sources (Web pages, tags, playlists, ...)
- Dependence on context (e.g., different artists)

brutal death metal

Top-Künstler



- (Reliable) data often only available on artist level for music context

Challenge for both Content and Context Analysis

- Extraction of relevant features from *noisy signal*

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 audio/signal-based tasks in 2015

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

Trend towards UX challenges (3 announced for 2015)

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

What is Music Similarity?

Evaluation strategies bypass this central question

Marsden (2015) criticizes the principles of “ISMIRality”:

- Music is not just a document
- Music is not just acoustic data
- Music is a trace of human behaviour
- MIR is task-centered but the most common musical activity, *listening*, has no obvious task.

Proposes the following definition:

“Two instances of music are similar when there is a plausible musical process which produces identical or similar outcomes for the two instances.”

Examples for “plausible processes”:

- Tapping on note onsets (vs., e.g., tapping every third note)
- Identifying composer of piece (vs., e.g., selecting pieces whose composers start with ‘B’)

What is Music Similarity?

Tasks/processes which give rise to similarity should be studied:

- Variation
- Performance of jazz standards
- Cover versions
- Oral traditions
- Music that accompanies similar passages in a film/TV, similar products in ads
- Music that is close in playlists (similar listening context)
- ...

Humans perform new tasks using knowledge of old, related tasks

Overlap of tasks gives the impression of musical similarity