Department of Computational Perception JKU JOHANNES KEPLER UNIVERSITY LINZ

Content- and Context-based Music Similarity and Retrieval

Markus Schedl
Peter Knees

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

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

Overview

Goals of this tutorial

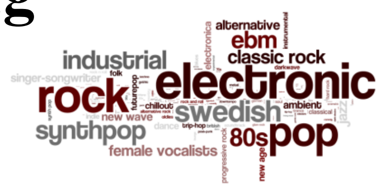
– Introduction to the field of music similarity estimation
– Approaches to music retrieval

Parts:

I. About Music Similarity

II. Music Content Analysis and Similarity

III. Music Context-Based Similarity and Indexing



IV. Personalization and User Adaptation

Department of
Computational
Perception

Schedule

Monday (today!)

Introduction to MIR, About music similarity, Evaluation of MIR systems, Basics in audio signal processing

Tuesday

Music content based methods, MFCCs, FPs, PCPs, Similarity calculation

Wednesday

Music context based methods, Text based methods, Co-occurrences, Collaborative filtering

Thursday

User context, Personalization, Hybrid Methods

Friday

Practical Exercise: Hybrid Music Recommender

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

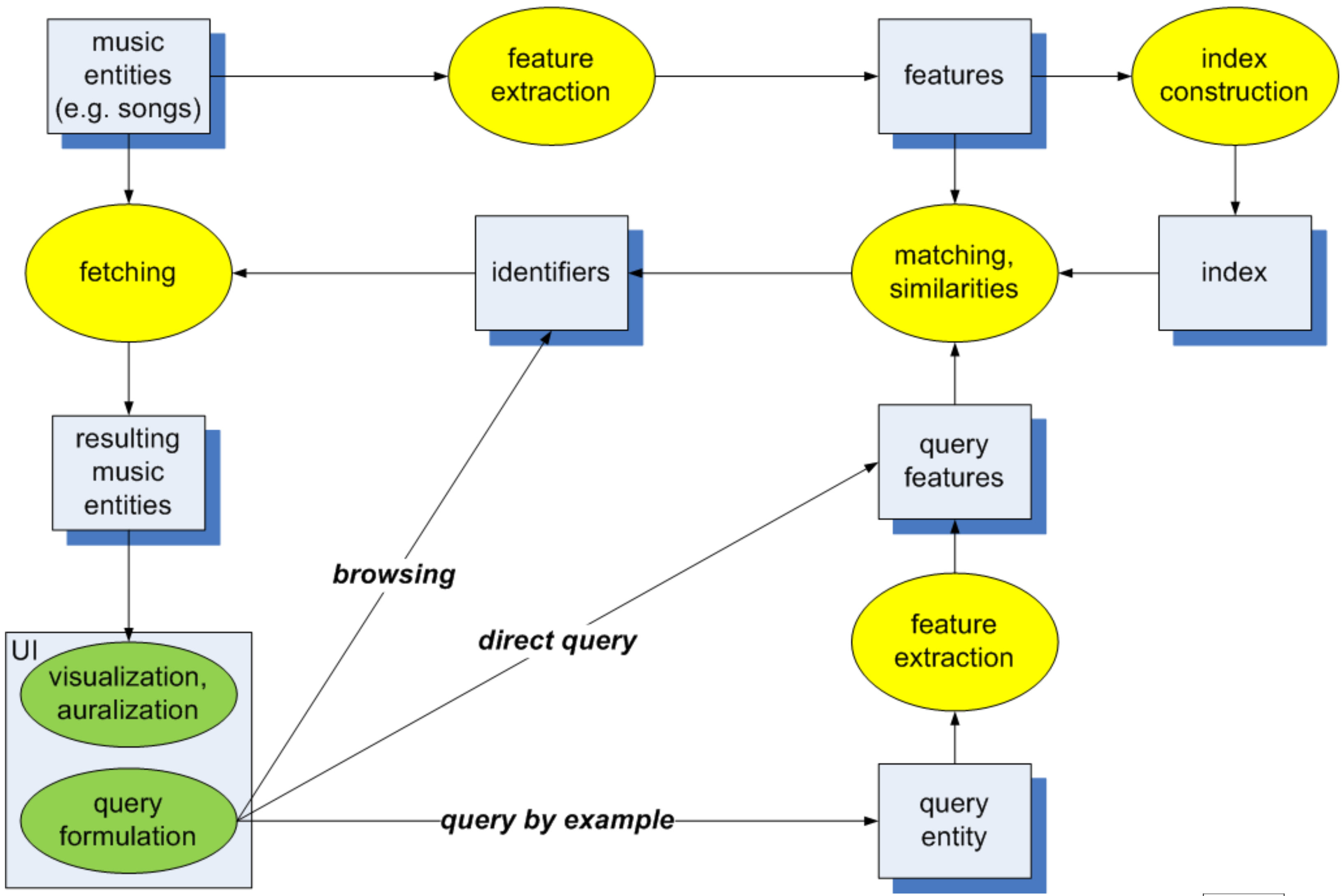
Research interests: social media mining, music and multimedia information retrieval, recommender systems, information visualization, and intelligent/personalized user interfaces

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



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]

Department of
Computational
Perception

Typical MIR Tasks

- 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)
- 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”

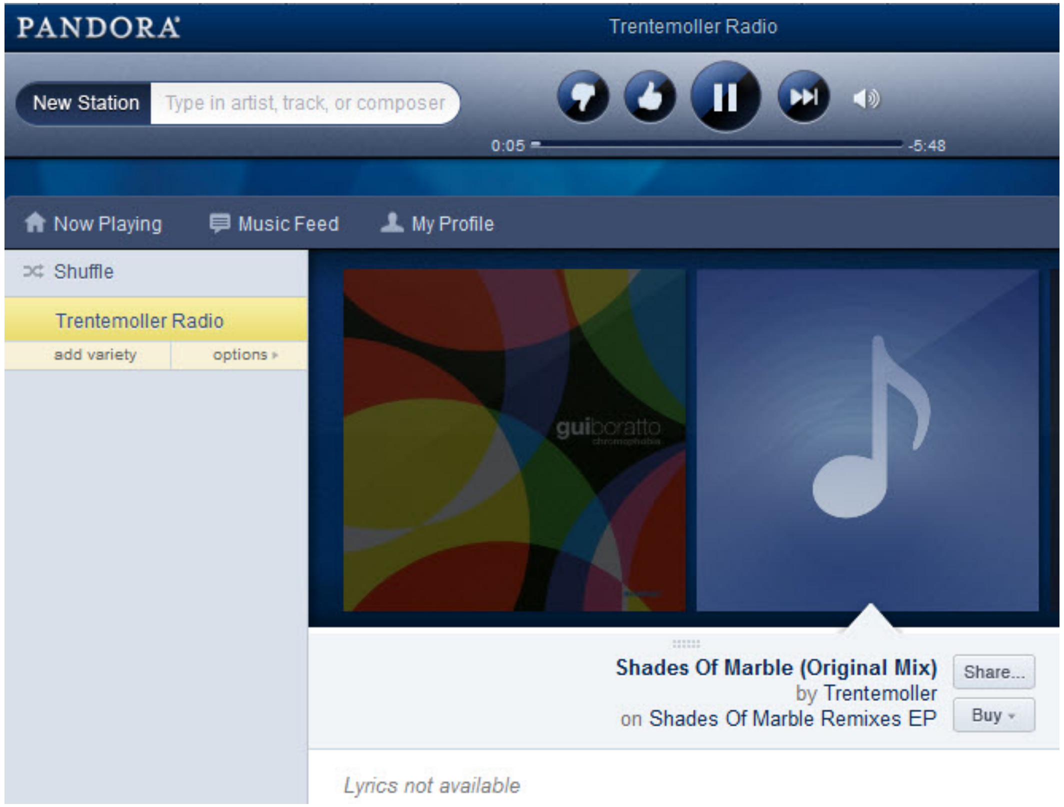
e.g.

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

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]

Department of Computational Perception

Applications: Audio Identification

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

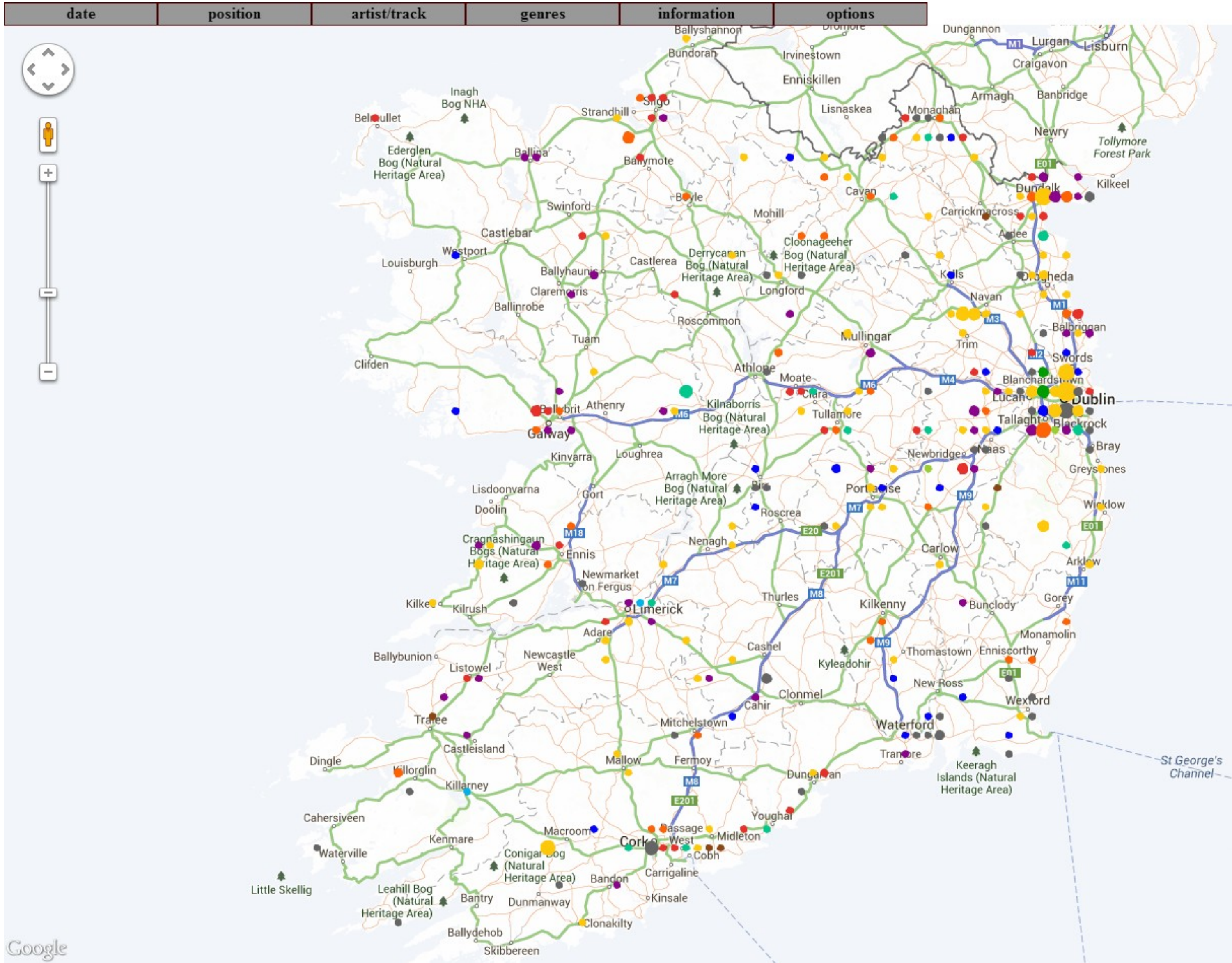
Query-by-humming:
input is not excerpt of a song, but melody hummed by the user

Examples:

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

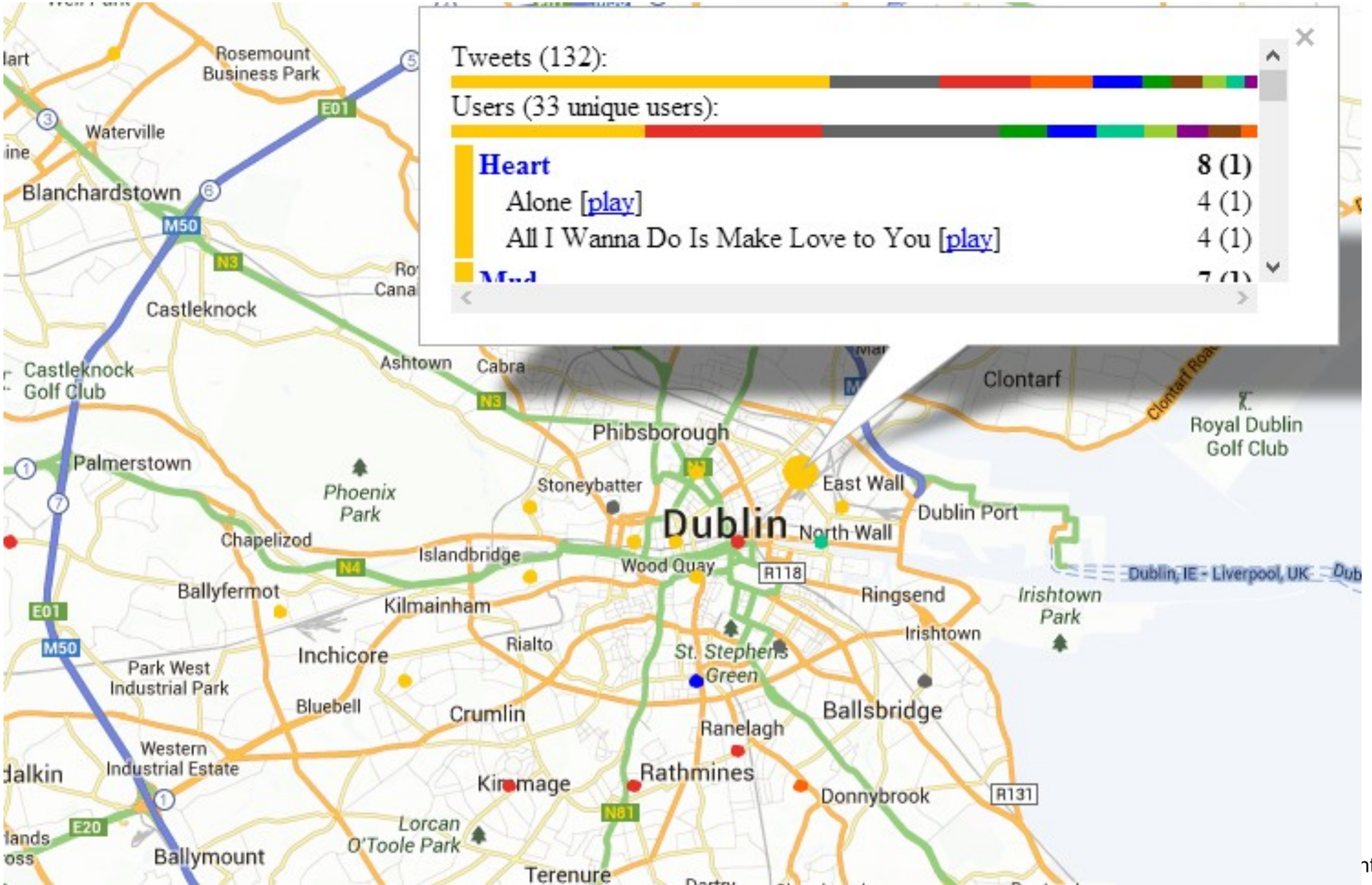
Melodiesuche
Query by Humming

Applications: Music Tweet Map



Department of Computational Perception

Applications: Music Tweet Map



nt of
Computational
Perception

(Raphael; 2003)

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

Department of Computational Perception

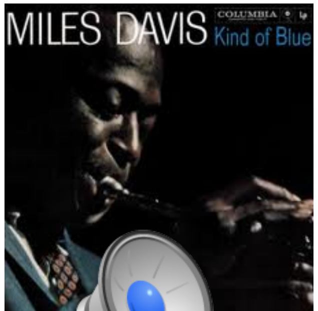
Music Similarity Examples

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!)

Computational Factors Influencing Music Perception and Similarity

(Schedl et al.; JIIS 2013)

music content

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

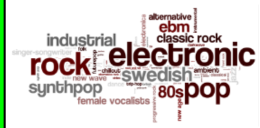
Examples:
- mood
- activities
- social context
- spatio-temporal context
- physiological aspects

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

Other Evaluation Campaigns

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

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

The MusiClef 2012 Data Set