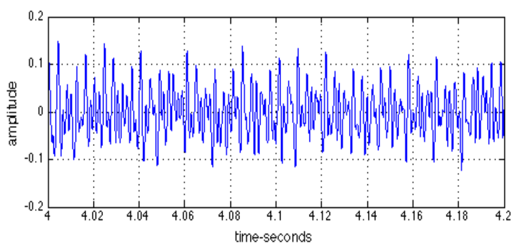
CTP 431 Music and Audio Computing

Music Information Retrieval

Graduate School of Culture Technology (GSCT)
Juhan Nam

Introduction



- ✓ Instrument: Piano
- ✓ Composer: Chopin
- ✓ Key: E-minor
- ✓ Melody
  - ELO “After all”
  - Radiohead “Exit Music”
- ✓ Transcription – Music notation
- ✓ Genre: Classical
- ✓ Mood: Melancholy, Sad, …

Music Information Retrieval (MIR)

- Information in Music
  - Factual: track, artist, years
  - Acoustic: loudness, pitch, timbre
  - Symbolic: Instrument, melody, rhythm, chords, structure
  - Semantic: genre, mood, user preference

- Area of research that aims to infer various types of information from music data
  - Make computer understand music as human does
  - Provide intelligent solutions to enhance human musical activities

MIR Tasks

- Audio fingerprinting
- Cover song detection
- Music transcription: melody, notes, tempo, chords
- Segmentation, structure, alignment
- Similarity-based retrieval, playlists, recommendation
- Classification: genre, mood, tags, …
- Query by humming
- Source separation: vocal removal
- Symbolic MIR: score retrieval or harmony analysis
- Optical Music Recognition (OMR)

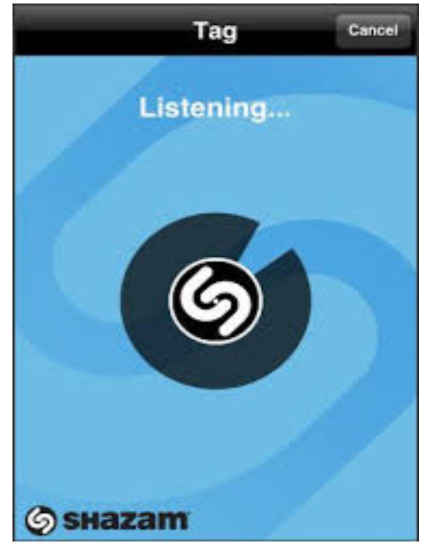
MIREX: http://www.music-ir.org/mirex/wiki/MIREX_HOME

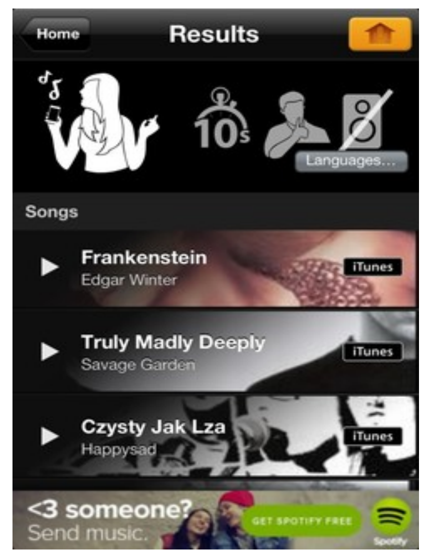
MIR Research Disciplines

- Digital Signal Processing
- Acoustics
- Music theory
- Machine Learning
- Natural language processing / Computer vision
- Psychology
- Human-Computer Interaction

Application: Music Search

- Query by music
  – Search a single unique song identified by the query
  – Audio fingerprint
  – Applied to movies, TV and ads, too
- Query by humming
  – Sing with humming and find closest matches
  – Melody match

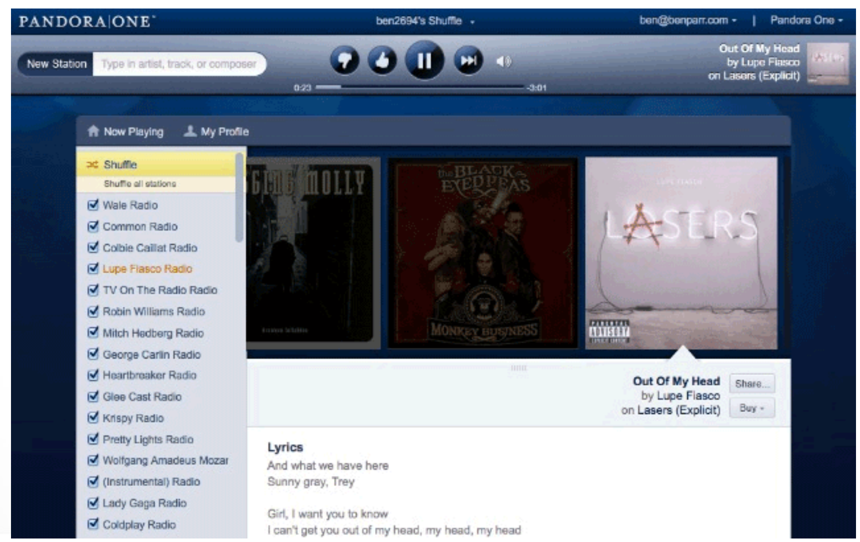




Application: Music Recommendation

- Personalized Radio
  - Generate Playlist
  - Based on user data, similarity and context

iTunes Radio



Pandora

Application: Score Following

- Listen to performance and track the notes
  - Example: JKU, Tonara

Application: Score Following

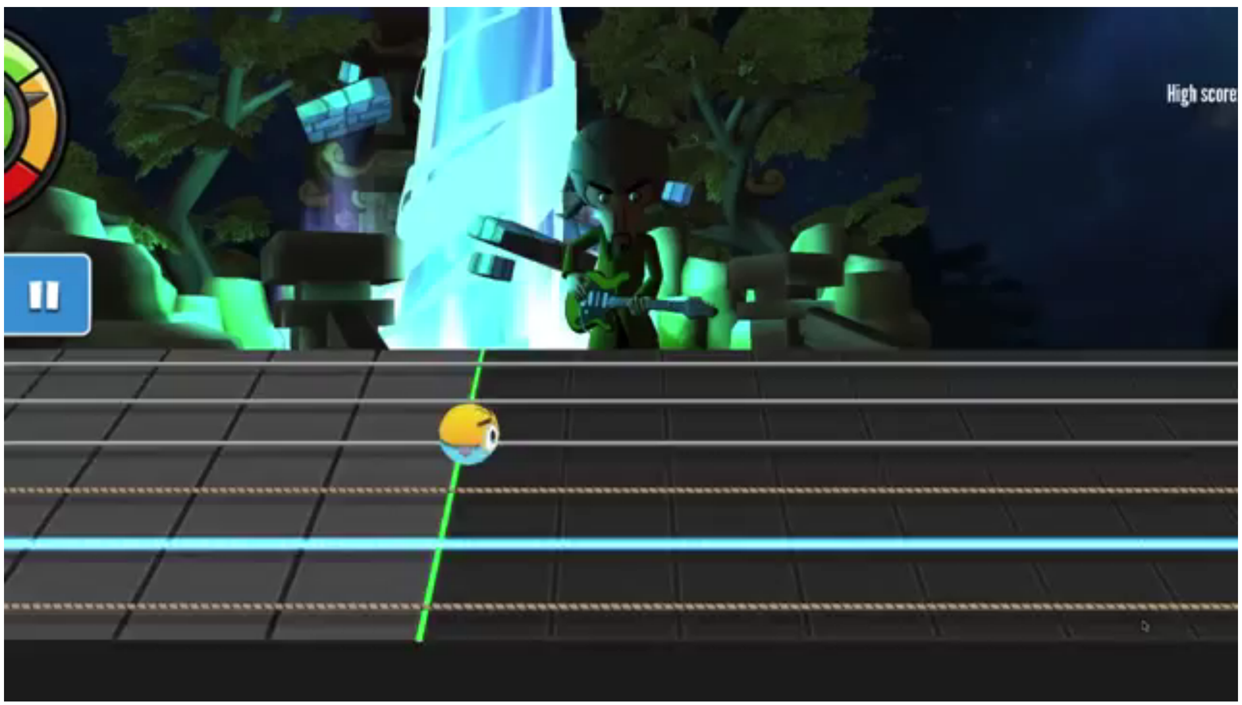
- The Piano Music Companion (2013)
  - Along with song identification

Application: Automatic Accompaniment

- Score following + Interactive Performance
  - Examples: IRCAM's Antefesco, Sonation's Cadenza

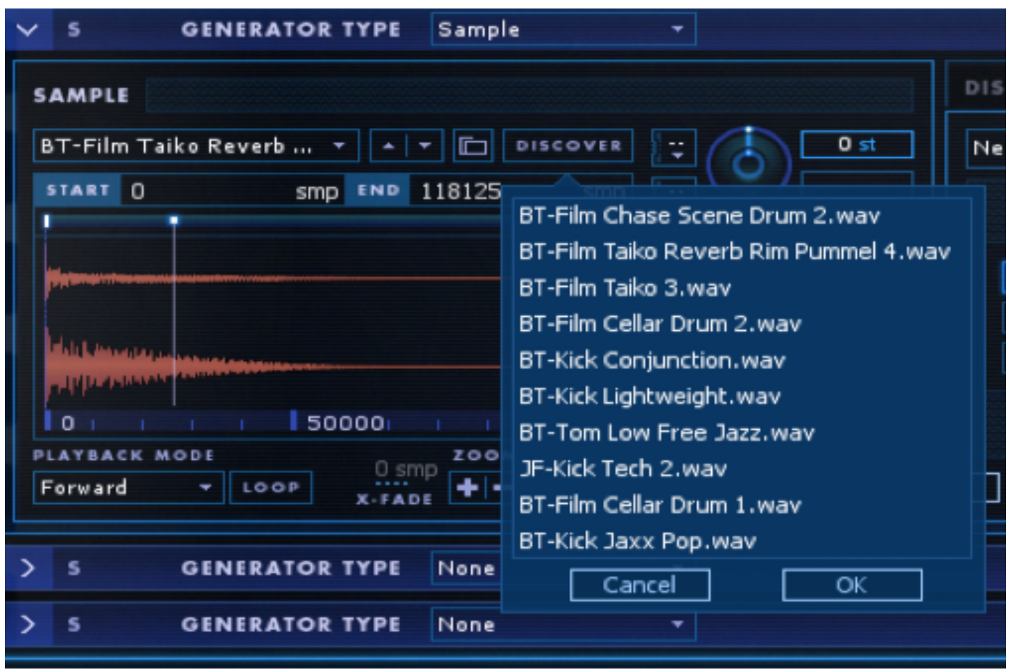
Application: Entertainment / Education

- Focus on performance evaluation
  - Learning musical instrument
  - Examples: Ovelin's Yousician, MakeMusic's Smartmusic, Ubisoft's RockSmith, RockProdigy



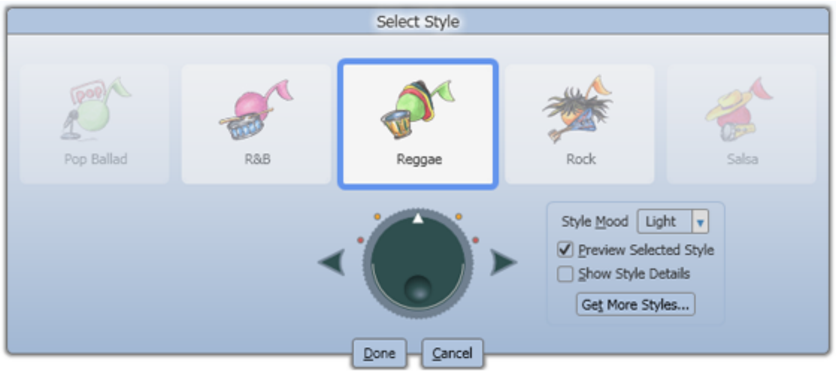
Application: Music Production

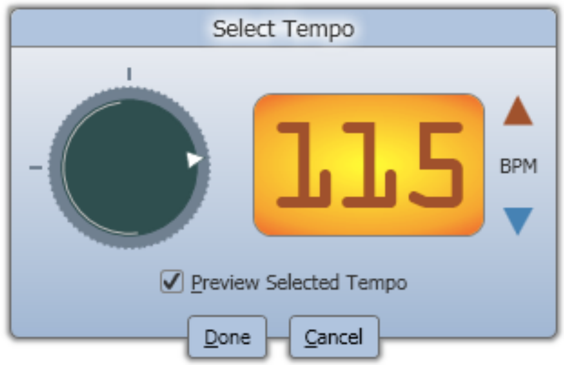
- Sound Sample search
  - Imagine Research's MediaMind: search sound effect sample for media production (e.g. film, drama)
  - Izotope's Breaktweaker: search similar timbre of drum sounds

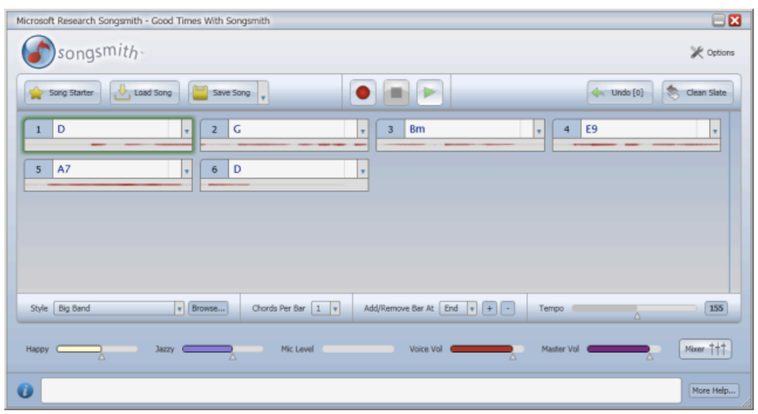


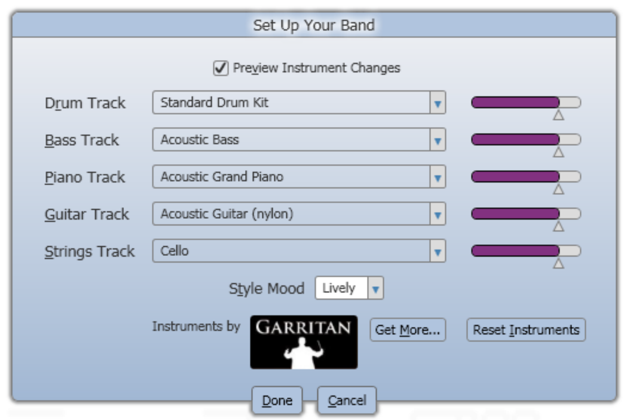
Application: Music Composition

- Automatic Song writing
  - Automatic arrangement
  - Example: MSR's Songsmith









CASE STUDY: Music Recommendation

Backgrounds

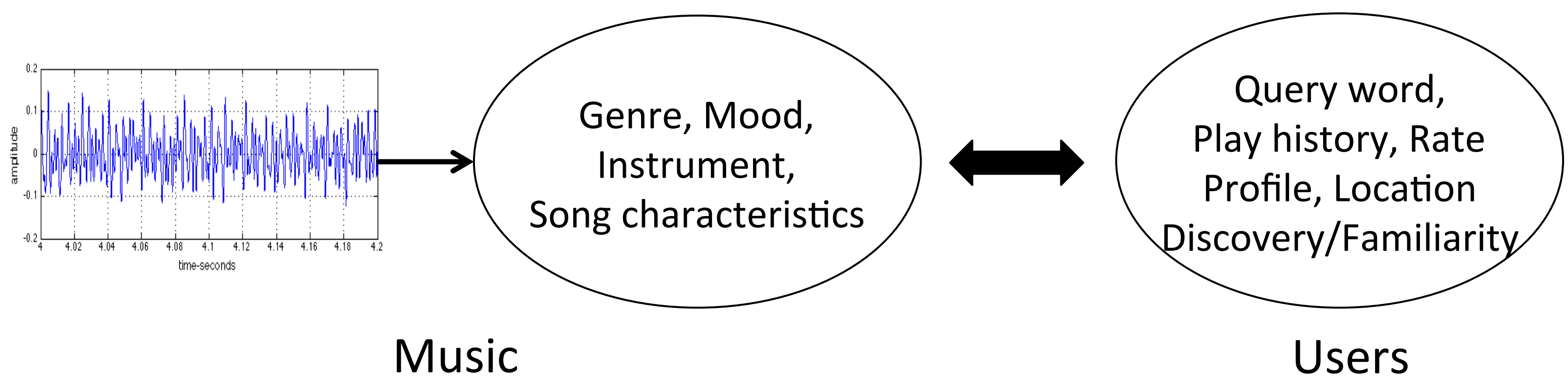
- Music record market
  - Offline → Online music services
  - CD → MP3 → Streaming audio
- Scale and diversity of music contents
  - Commercial music tracks
    - Spotify: 30M+ songs (2015)
    - Bugs music: 4.1M+ songs (2015)
  - User contents
    - YouTube: 300h+ video uploaded per min (2015)
    - SoundCloud: 12h+ audio uploaded per minute (2014)
  - TV, cables and online media
    - Music program, concert, music videos, audition, …

Backgrounds

- Connection with human data
  - Number of users
    - Spotify: +24M active users (as of Jan, 2014)
    - YouTube: +1B unique users' visit each month (as of Dec, 2014)
  - Personal data
    - Play history, rate, personal music library
    - Profile: age, occupation, …
  - Social data
    - The majority of online services can be logged in via SNS
    - Friends, followers
    - Daily posting, blog (reviews), comments

Challenges

- There are too many choices of music contents
- How can we find music more easily or in a human-friendly way?
  - Searching music with various queries (e.g. text, humming, audio tracks)
  - Recommendation based on user data (e.g. play history, rating, location)
- We need to extract semantic or musical information from audio tracks, and match them to the query or user data



Current Approaches

- Manual Curation
- Human Expert Analysis
- Collaborative Filtering
- Content-based Analysis (by computers)

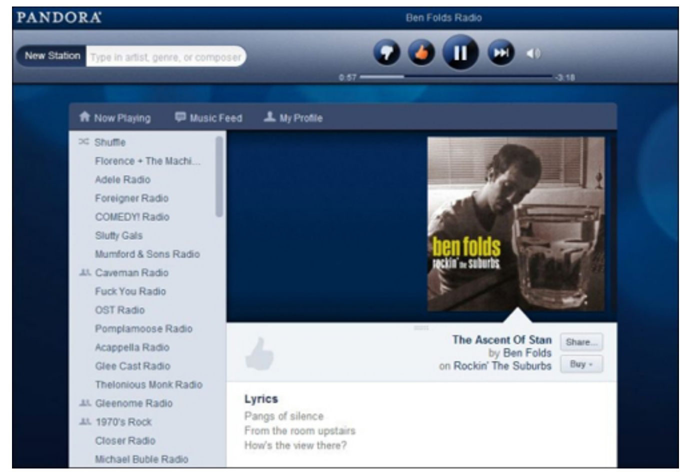
Manual Curation

- Playlist generation by music experts (or users)
  - Traditional: AM/FM radio
  - The majority of current music services are based on this approach
- Advantages
  - Effective for usage-based music services (workout, study, driving or prenatal education)
  - Good for music discovery
  - Often with story-telling
- Limitations
  - No personalization
  - Not scalable

[www.soribada.com]

Human Expert Analysis

- Pandora: music genome project (1999)
  – Musicologists analyze a song for about 450 musical attributes in various categories
  – Big success as a music service



- Advantages
  – High-quality analysis
  – Good for music discovery

- Limitations
  – Expensive: take 20-30 minutes for a song to be analyzed
  – Not scalable : only for commercial tracks ?

Collaborative Filtering (CF)

- Basic idea

Person A: I like songs A, B, C and D. Person B: I like songs A, B, C and E. Person A: Really? You should check out song D. Person B: Wow, you also should check out song E.

- Formation
  - Matrix factorization (or matrix completion) problem

$$\begin{array}{r c}
 & \\
 & \\
 & \\
 & \\
 & \\
\text{Juhan} &
\end{array}
\begin{pmatrix}
1&0&0&0&1&0&0&1\\
0&0&1&0&0&1&0&0\\
1&0&1&0&0&0&1&1\\
0&1&0&0&0&1&0&0\\
0&0&1&0&0&1&0&0\\
1&0&0&0&1&0&0&0
\end{pmatrix}
\approx
\underbrace{\begin{pmatrix} \\ & X & \\ \\ \end{pmatrix}}_{f}
x_u
\left(\quad Y \quad y_s \right)\Big\} f$$

Gangnam Style

Juhan's latent vector

Gangnam Style's latent vector

Song Preference

$$p_{us} = x_u^T y_s$$

User Similarity

$$q_{u1u2} = x_{u1}^T x_{u2}$$

Song Similarity

$$r_{s1s2} = y_{s1}^T y_{s2}$$

Collaborative Filtering

- Advantages
  - Capture semantics of music in the aspect of human
  - Enable personalized recommendation (by nature)

- Limitations
  - The cold start problem: what if a song was never played by anyone?
  - Popularity bias: likely to recommend (already) well-known songs or songs from the same musician or album

Collaborative Filtering

- Bad examples

Can you find songs similar to this musician?

Query	Most similar tracks (WMF)
Jonas Brothers - Hold On	Jonas Brothers - Games Miley Cyrus - G.N.O. (Girl's Night Out) Miley Cyrus - Girls Just Wanna Have Fun Jonas Brothers - Year 3000 Jonas Brothers - BB Good
Beyoncé - Speechless	Beyoncé - Gift From Virgo Beyonce - Daddy Rihanna / J-Status - Crazy Little Thing Called Love Beyoncé - Dangerously In Love Rihanna - Haunted
Coldplay - I Ran Away	Coldplay - Careful Where You Stand Coldplay - The Goldrush Coldplay - X & Y Coldplay - Square One Jonas Brothers - BB Good
Daft Punk - Rock'n Roll	Daft Punk - Short Circuit Daft Punk - Nightvision Daft Punk - Too Long (Gonzales Version) Daft Punk - Aerodynamite Daft Punk - One More Time / Aerodynamic

These songs are already what I know well !

[Oord et. al, 2013]

Content-Based Analysis: Music Auto-tagging

- Google has music service as part of Google play
  - Their main features "Instant mix", which automatically generates a playlist based on user's music collections or play history
- They do CF but also make use of audio content. How?

Inside Google's Infinite Music Intelligence Machine

In May, Google launched a music service that will challenge Spotify and Pandora for radio domination. We asked Google research scientist Doug Eck how it works.

Fast Company, July, 2013

How Google Music Intelligence Works

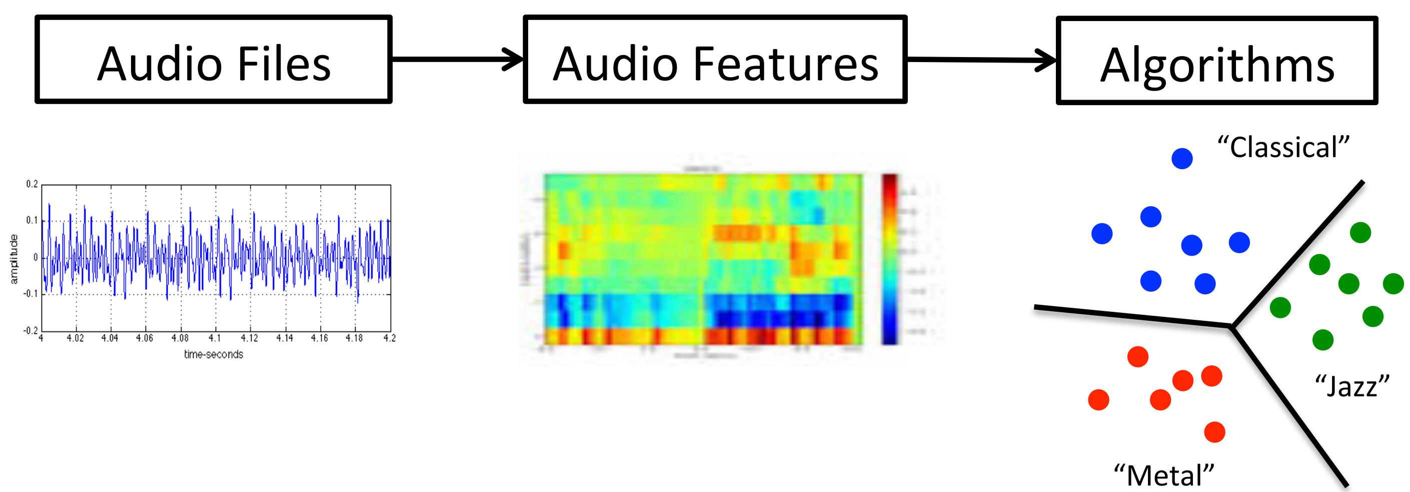
Eck's team is focused on the technical side of this equation, relying on a dual-sided machine learning methodology. One component of that is collaborative filtering of the variety employed by Netflix and Amazon to recommend horror flicks and toasters. The other involves machine listening. That is, computers "listen" to the audio and try to pick out specific qualities and details within each song.

"We use a kind of neural network backend to do this," says Eck.

Content-Based Analysis: Music Auto-tagging

- An intelligent approach that makes computers listen to music and predict descriptive words (i.e. tags) from audio tracks
  - Features: MFCC, Chroma,…
  - Algorithms: GMM, SVM, Neural Networks
  - Tags: genre, mood, instrument, voice quality, usage

- Basic Framework



Example of Auto-tagging

This is a [ ] song that is [ ], [ ] and [ ]. It features [ ] and [ ] vocal. It is a song with [ ] and [ ] that you might like to listen to while [ ].

This is a [**very danceable**] song that is [**arousing/awakening**], [**exciting/thrilling**] and [**happy**]. It features [**strong**] and [**fast tempo**] vocal. It is a song with [**high energy**] and [**high beat**] that you might like to listen to while [**at a party**].

James Brown – Give it up or turn it a loose

This is a [**pop**] song that is [**happy**], [**carefree/lighthearted**] and [**light/playful**]. It features [**high-pitched**] vocal and [**altered with effects**] vocal. It is a song with [**positive feeling**] that you might like to listen to while [**at a party**].

Cardigans - Lovefool

Text-based Music Retrieval by Auto-tagging

- Sort the probability of the query tag and choose top-N songs
  - Like text-based Google search

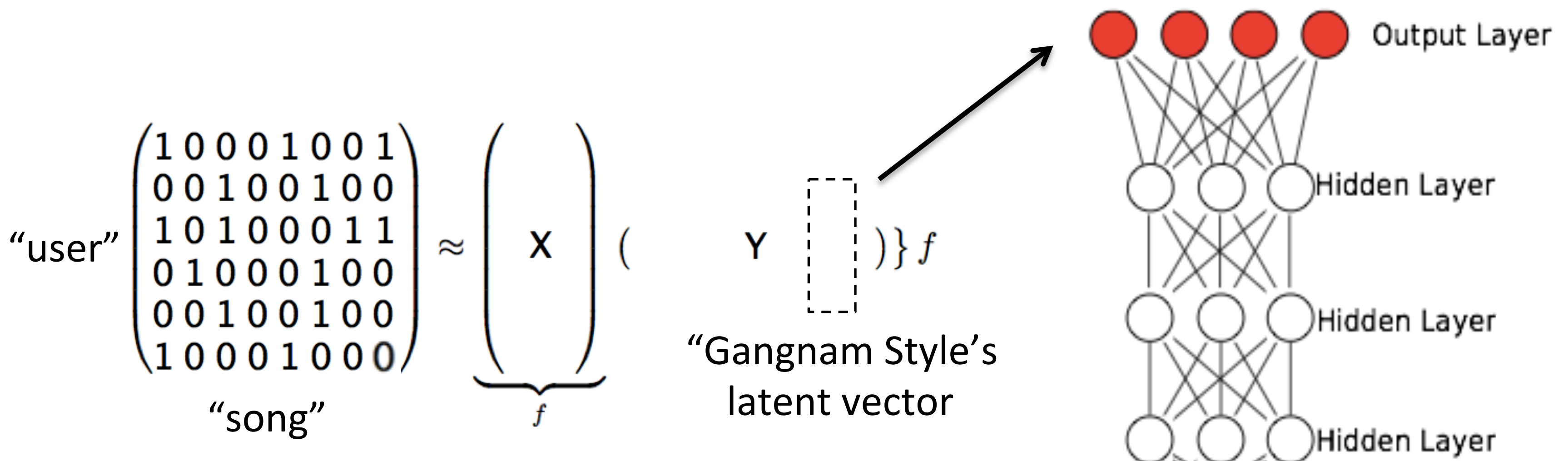
Query word: "Female Lead Vocals" →

Top 5 ranked songs
Norah Jones – Don't know why
Dido – Here with me
Sheryl Crow – I shall believe
No doubt – Simple kind of like
Carpenters – Rainy days and Mondays

- We also can compute similarity between songs using the estimated tag probabilities
  - E.g. cosine distance between two tag probability vectors
  - Applicable to query by audio

Content-based Music Recommendation

- Blending audio and user data
  - Replace the text-based tags with the latent vector of a song



Matrix factorization from collaborative filtering

[Oord et. al, 2013]

Audio Track of "Gangnam Style"

Music Retrieval Results

Query	Most similar tracks (WMF)	Most similar tracks (predicted)
Jonas Brothers - Hold On	Jonas Brothers - Games Miley Cyrus - G.N.O. (Girl's Night Out) Miley Cyrus - Girls Just Wanna Have Fun Jonas Brothers - Year 3000 Jonas Brothers - BB Good	Jonas Brothers - Video Girl Jonas Brothers - Games New Found Glory - My Friends Over You My Chemical Romance - Thank You For The Venom My Chemical Romance - Teenagers
Beyoncé - Speechless	Beyoncé - Gift From Virgo Beyonce - Daddy Rihanna / J-Status - Crazy Little Thing Called Love Beyoncé - Dangerously In Love Rihanna - Haunted	Daniel Bedingfield - If You're Not The One Rihanna - Haunted Alejandro Sanz - Siempre Es De Noche Madonna - Miles Away Lil Wayne / Shanell - American Star
Coldplay - I Ran Away	Coldplay - Careful Where You Stand Coldplay - The Goldrush Coldplay - X & Y Coldplay - Square One Jonas Brothers - BB Good	Arcade Fire - Keep The Car Running M83 - You Appearing Angus & Julia Stone - Hollywood Bon Iver - Creature Fear Coldplay - The Goldrush
Daft Punk - Rock'n Roll	Daft Punk - Short Circuit Daft Punk - Nightvision Daft Punk - Too Long (Gonzales Version) Daft Punk - Aerodynamite Daft Punk - One More Time / Aerodynamic	Boys Noize - Shine Shine Boys Noize - Lava Lava Flying Lotus - Pet Monster Shotglass LCD Soundsystem - One Touch Justice - One Minute To Midnight

Collaborative Filtering only

Collaborative Filtering
+ Audio Content

[Oord et. al, 2013]

Content-Based Analysis: Music Auto-tagging

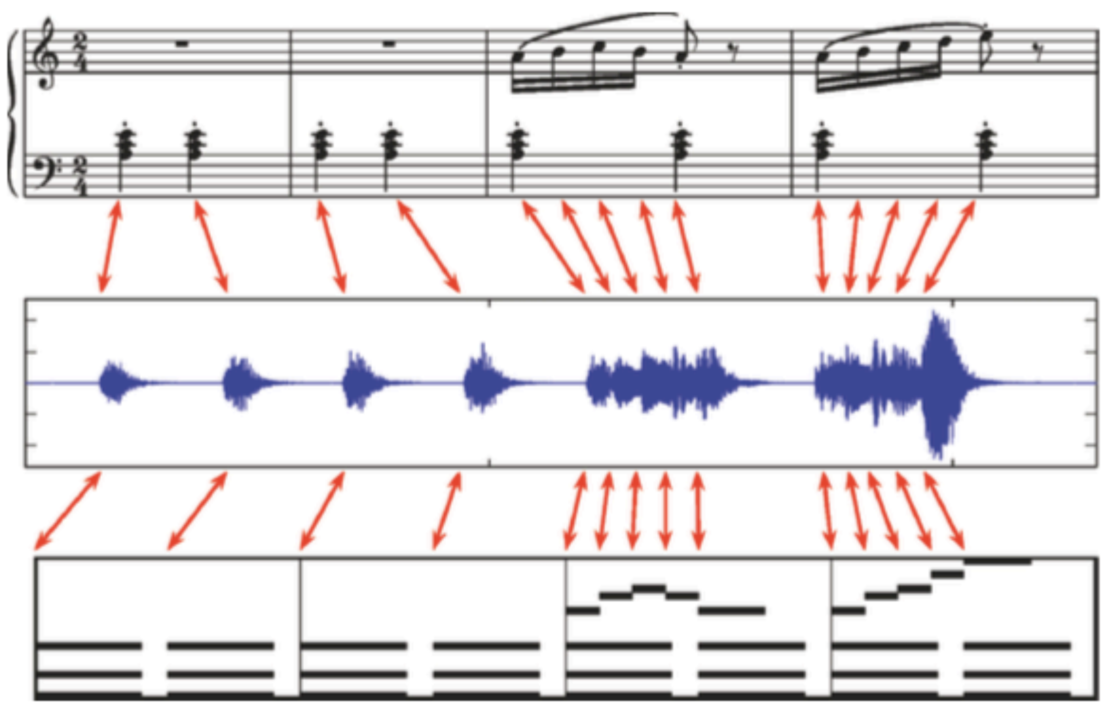
- Advantages
  - Free of cold-start and popularity bias
  - Highly scalable: using high-performance computing
  - Works for music in other media or user content as well
  - Can be combined with other approaches

- Limitations
  - Some tags are unpredictable: indy, idol, …
  - Hard to measure music quality (or level of performance), especially for user contents

CASE STUDY: Score Following

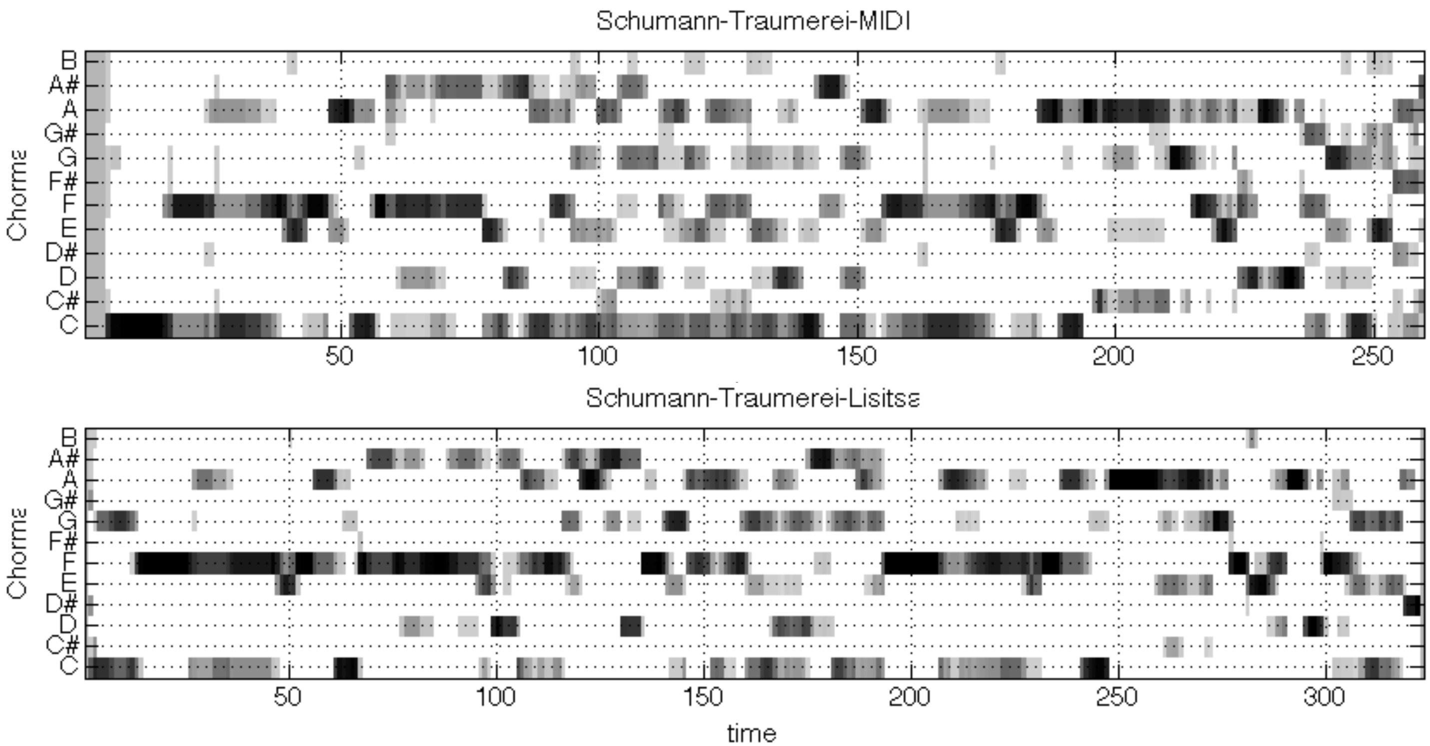
Music Score Following

- Tracking played notes while listening to the music
  - Temporally align different representations or renditions of music
  - Audio to Audio, Audio to Score (or MIDI)



Music Score Following

- Extracting Chroma Features
  - Capture harmonic (or tonal) characteristics of music



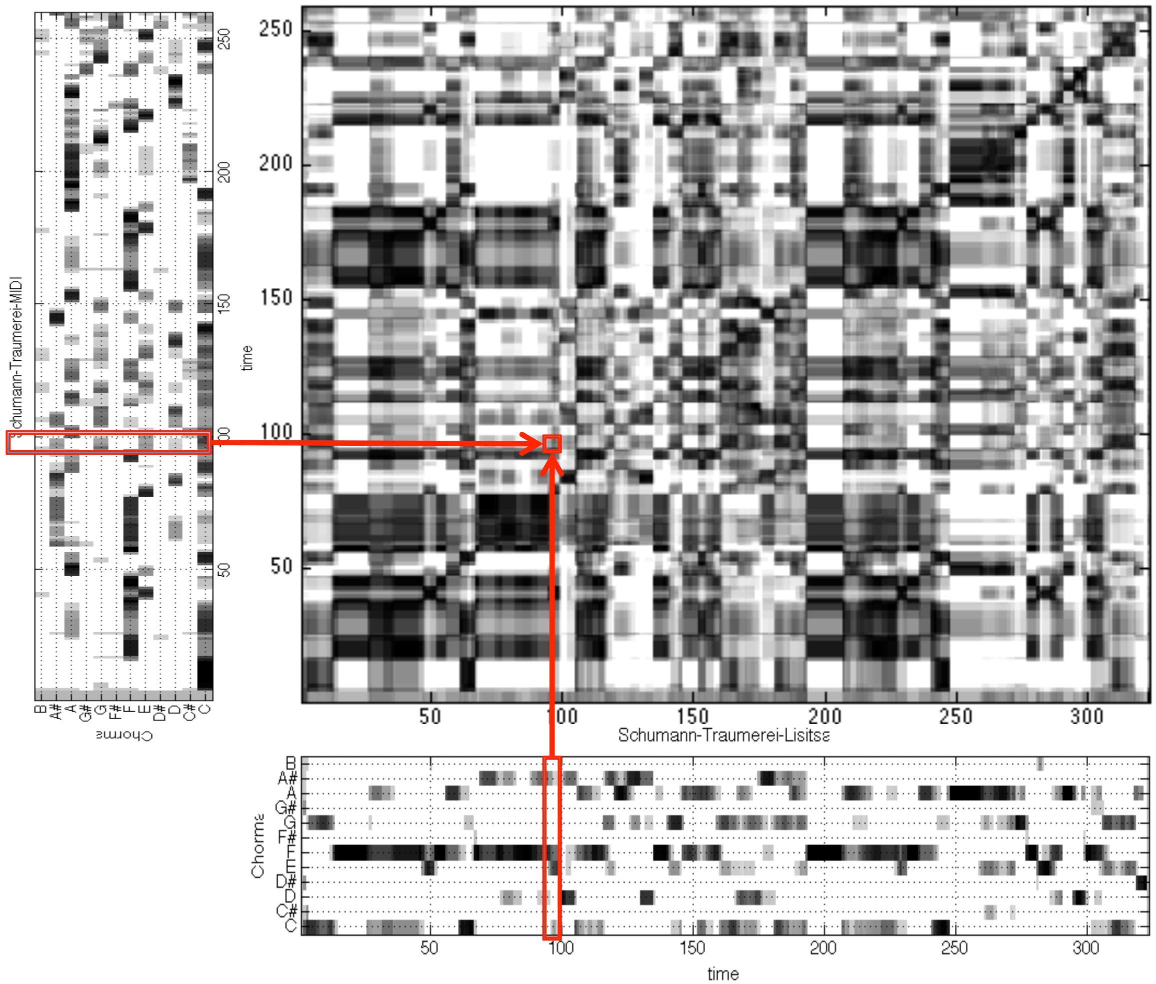
MIDI

Lisitsa

CENS : Normalized Chroma Features (Muller, 2005)

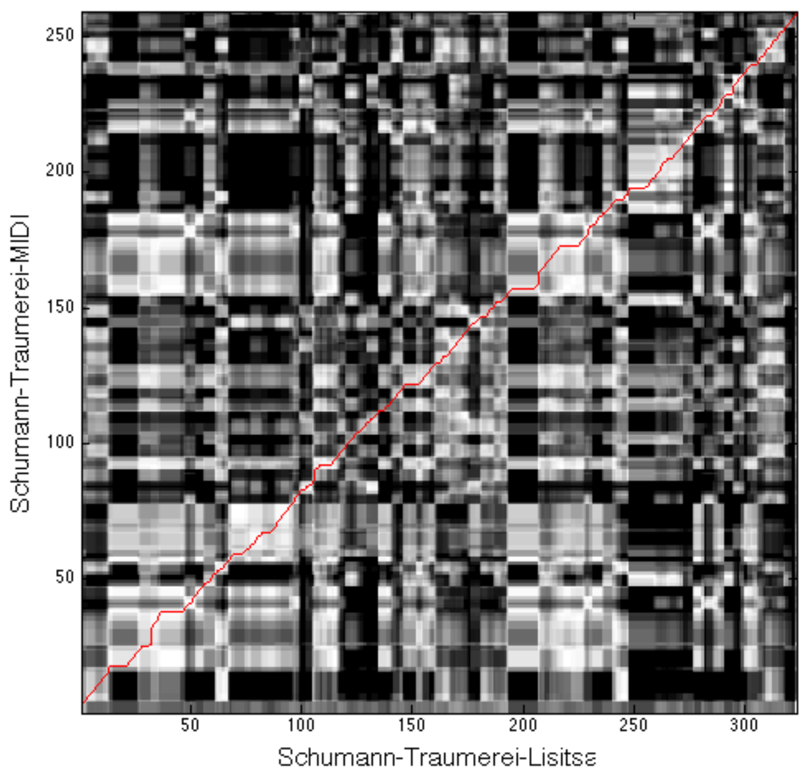
Music Score Following

- Computing (Dis)similarity Matrix

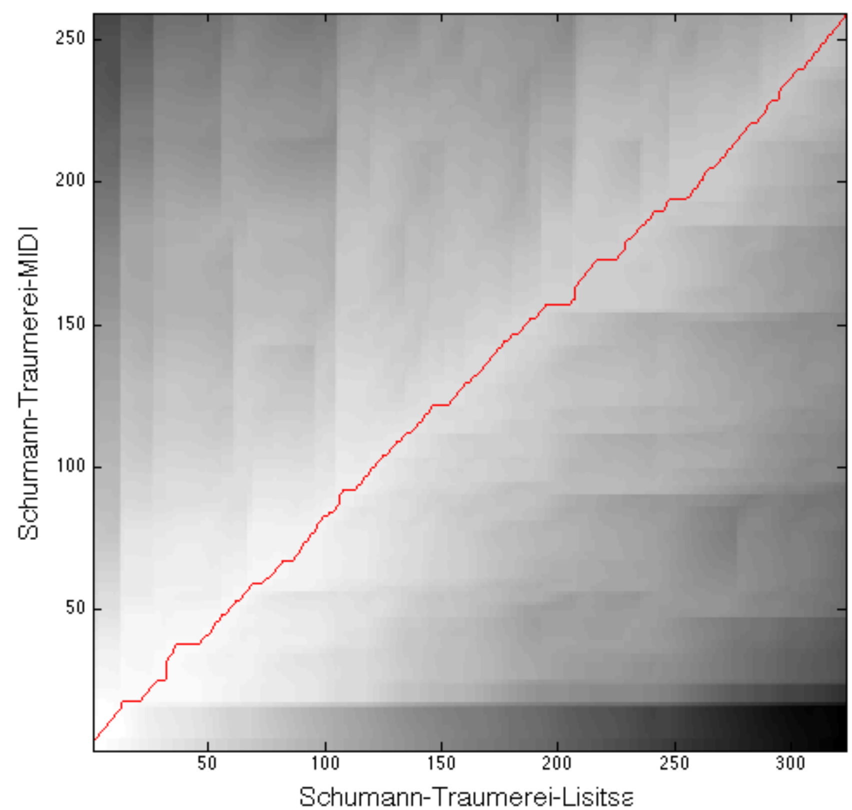


Music Score Following

- Computing the Shortest Path using Dynamic Time Warping



Local Similarity



Accumulated Similarity

Score Following Demo