Tampere University

# Content-based music retrieval

SGN 14007

Lecture 12

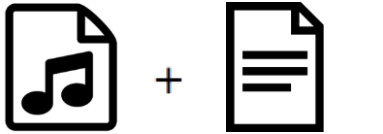
Annamaria Mesaros

# Music information retrieval (MIR)

- The amount of music we have access to has increased several orders of magnitude since early 2000's.
  - This requires a new mechanism for selecting and identifying music
- MIR is a multidisciplinary science of retrieving information from music
  - Aims to develop strategies to quickly access music collections through content analysis.

# MIR

- Subtasks of MIR include:
  - Beat tracking: Automatic analysis of temporal structure of music (beat, tempo, rhythm, meter)
  - Chord and key recognition
  - Melody estimation: Melody is a strong indicator of the identity of a musical piece.
  - Music fingerprinting (for exact identification of a musical piece)
  - Instrument recognition
  - Score alignment (synchronize audio and score)
  - Structure analysis
  - Genre recognition
  - Autotagging
  - …



# MIR

- **Metadata** describes the content of the musical piece
  - Can be analyzed by humans (domain experts) or computers
  - Factual metadata
    - performer, year, name of the song, album, etc.
  - Cultural metadata
    - attributes like mood, emotion, genre, style, etc.. (Referred also as “tags”)

# Finding music

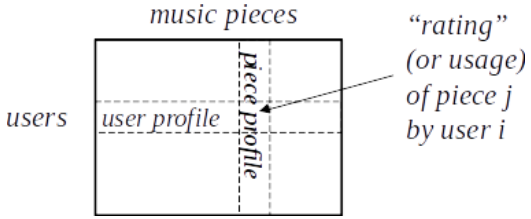
- Traditional ways of finding music are no longer sufficient
  - We cannot browse through all the music we would potentially like
  - Record companies and radio stations are no longer critical gatekeepers in music distribution
- Relying just on popularity statistics is not effective
  - Music tastes are so different that averaging opinions does not produce precise information for an individual
  - ”UK Singles Chart” etc. sales statistics work badly as a guide for the consumer

- Finding music:
  - Spotify – Discover weekly, daily mix
  - YouTube – Recommended
  - LastFM – Recommended artists, songs, tags

# Searching music

Two complementary approaches:

- **Collaborative filtering**
  - Based on (users x items) matrix ”music likings metadata”
  - Recommend music by comparing user profiles and predicting likings for new pieces
  - Measure similarity of music pieces (acoustics, usage, etc.) based on piece profiles




- **Content-based retrieval**
  - Either based on automatic signal analysis or collaborative tagging by users

- Old ways of discovering music are still relevant too (though ineffective)
  - Talking to friends, relying on experts (e.g. listening to FM radio you like)

# Audio-based music retrieval

- Collaborative filtering (CF) does not solve it all
  - CF does not allow separating the various dimensions of music similarity, but these are all mixed in the piece profile
  - CF alone is not able to deal with items that are new or do not have many listeners
- Audio-based MIR addresses the above problems
  - Enables “truly musical queries” with specific musical criteria, such as requesting pieces with certain vocal characteristics or slow tempo
  - Can be employed even on media libraries that do not have any audience of listeners
- On the other hand, audio-based MIR alone cannot measure aspects like quality, usage, or culture
  - The two approaches are complementary

# Manual tagging?

- Music annotation by human experts is costly and limits the coverage
  - Pandora.com is audio-based MIR service (US only) based on expert tagging

- Collaborative tagging by music service users (for example last.fm) is effective for items that are sufficiently popular
- Tagging games can achieve better coverage, but (currently) less users

MajorMiner

Music Labeling Game

ww.majorminer.com

listen game

www.listengame.com

# Content-based audio retrieval

- Content based retrieval make use or the raw music data, rather than rely on manually generated information

- Content based music description allows:
  - Identify what the user is searching, even if the user does not specifically know herself
  - Identify music captured from loudspeakers in a noisy space
  - Identify song name based on user’s humming (query by humming)

- Content based retrieval
  - Retrieval process starts with a query (text, audio)
  - Retrieval system returns all items that are somehow related to the query

# Query mechanisms

- Query by example
  - Given music representation or fragment, retrieve music with similar parts or aspects.
- Browse by similarity
  - Find music similar to the user taste/choice
- Query by humming or tapping
  - Recognize the song the user wants based on humming/tapping
- Tempo
  - Find music with desired tempo
- Lyrics
  - Find songs containing given keywords
- Music categories
  - Genre, mood, tags

# Music similarity

- Music similarity estimation enables query by example and browsing by artist similarity

- Widely used acoustic features
  - Mel-frequency cepstral coefficients (MFCCs)  timbre/instrumentation
  - Chroma: collapse spectral content into one octave and use 12 bins for the total spectral energy on each pitch class (c, c#, d,...,b)  harmonic content
  - Rhythmogram (or, fluctuation patterns): cosine transform in blocks that extend in time direction  rhythm

# Features define “similarity”

- ”Similarity” as such is not well-defined

Is Bohemian rhapsody by Queen more similar to:

a. Bohemian rhapsody by London Symphony Orchestra, or

b. Killer Queen by Queen?

- The riddle is solved by choosing the acoustic features
  - chroma  a) is more similar (composition)
  - MFCCs  b) is more similar (instrumentation)
- User may wish to specify the features when doing query by example

# Audio identification

- Aims to identify a particular recording within a large archive of recordings.
- Applications:
  - User applications to identify songs (Shazam)
  - Copyright detection from user uploaded videos (youtube: www.youtube.com/t/contentid )

- Audio fingerprinting based on spectral peak pairs:

  for **each song** in the database

  - Take STFT of song to get time-frequency representation X(t,f), where t denotes time, and f is frequency.
  - Find peaks (landmarks), i.e., time–frequency points $\{t_i,f_i\}$
  - Form pairs of landmarks, keys: $<f_i,f_j,t_j-t_i>$
  - Create hash values from key f(key)=value, i.e., a [binary string]
  - Add hash value with song ID and time of landmark $t_i$ to database

  end for

Tampere University

# Audio landmarks

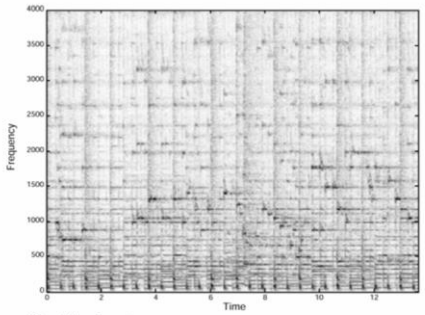



Fig. 1A - Spectrogram

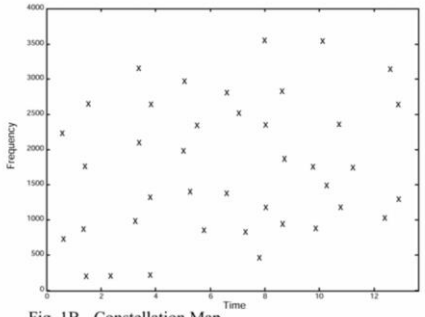



Fig. 1B - Constellation Map

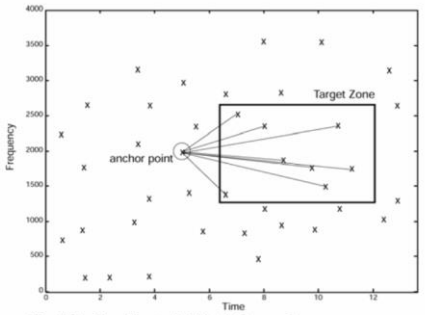



Fig. 1C - Combinatorial Hash Generation

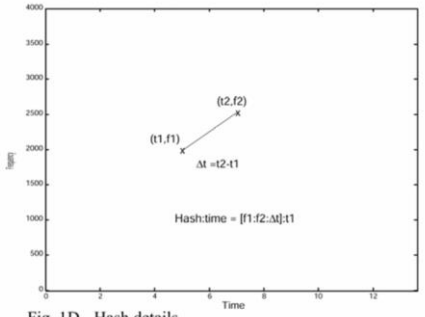



Fig. 1D - Hash details

# Audio identification

- Query part:
  - Take STFT of query audio signal
  - Extract peaks from the query audio
  - Form pairs of landmarks (*keys*)
  - For each pair:
    - Extract hash value from key (and time)
    - Retrieve list of song IDs with same hash value
    - Calculate offset time between landmark time in query audio and landmark time in database song.
  - Find most frequently occurring songs in the retrieved lists
    - Offset time should be similar in the same song

Example: each key (pair of landmarks) corresponds to a unique hash tag, associated with a list of songs

| Hash value | Time in query | Song id:time (s) |
|---|---|---|
| 00 FA 12 FF | 0.1 | A:15.1, G:33.1 |
| 11 EA FA 01 | 0.5 | A:15.5, D:3.1, B:1.2 |
| 59 73 A3 F1 | 0.7 | A:15.7, C:55.2 |

Table shows found query file hashes:

00 FA 12 FF
11 EA FA 01
59 73 A3 F1

And the time in the query audio (0.1, 0.5, 0.7s)

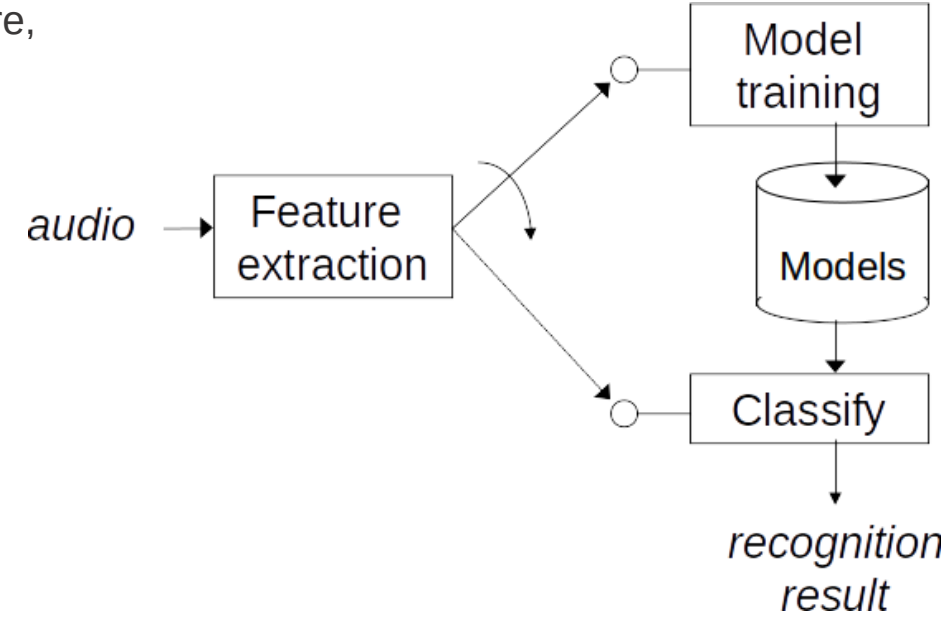
Matching hash values are found in database songs with hash time also listed.

Song A appears often. It also has a constant time-offset value of 15 seconds to the query.

# Music classification

- Music can be classified according to genre, mood, etc.
- Classical train/test supervised classification scenario (figure)

- MIREX: Classify music into categories
  - genre: rock, hip hop, jazz, classical,...
  - mood: aggressive, passionate, humorous, cheerful,..
  - artist identification
  - classical composer identification



# Query by humming

- Consists of two main steps:
  - Melody transcription of a hummed or sung query into a suitable higher-level representation
  - Matching that representation against a large database of known reference items
- Example method [Ryynänen-icassp-2008]
  - Preprocessing: extract melodies automatically from music pieces
  - Transcribe the query
- Match by Euclidean distance between the two melodic contours (allow time scaling)

  - Example A
    - query 🔊 retrieval results #1 🔊 #2 🔊 #3 🔊
  - Example B
    - query 🔊 retrieval results #1 🔊 #2 🔊 #3 🔊
  - Example C
    - query 🔊 retrieval results #1 🔊 #2 🔊 #3 🔊

# Lyrics: what is this song about?

Miranda

in the morning

takes her eggs sunny side up

# Summary

- Using audio content to search
  - Motivated by the increase in amount of music available
- Content can be queried in many forms
  - Text (factual and cultural metadata)
  - Example (audio recording, humming, etc.)
- Common structure
  - Features are extracted from the query
  - Database contains features for each song
  - Return similar songs (similar in some respect)