Music and Data

Mathieu Lagrange

October 23, 2018

Course website: https://github.com/mathieulagrange/musnu

Outline

1. Music
2. Data
3. Recommendation
4. Content based Analysis

Outline

1. Music
2. Data
3. Recommendation
4. Content based Analysis

Outline

Outline

1. Music
2. Data
3. Recommendation
4. Content based Analysis

Definition

Definition

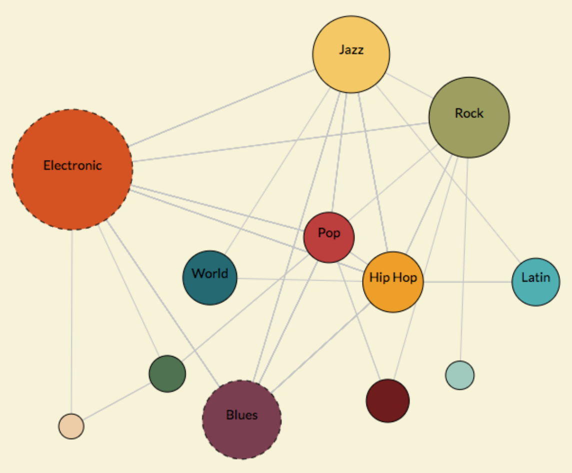
- Music is an art form consisting of sound and silence, expressed through time.
- Aesthetic: "the harmony of the spheres" and "it is music to my ears" vs. "There is no noise, only sound."
- can be divided into genres, which can be divided in sub genres

Genre, you said genre ?

A music genre is a conventional category

- that identifies some pieces of music as belonging to a shared tradition or set of conventions.
- It is to be distinguished from musical form and musical style, although in practice these terms are sometimes used interchangeably.
- Recently, academics have argued that categorizing music by genre is inaccurate and outdated.

Genre relations



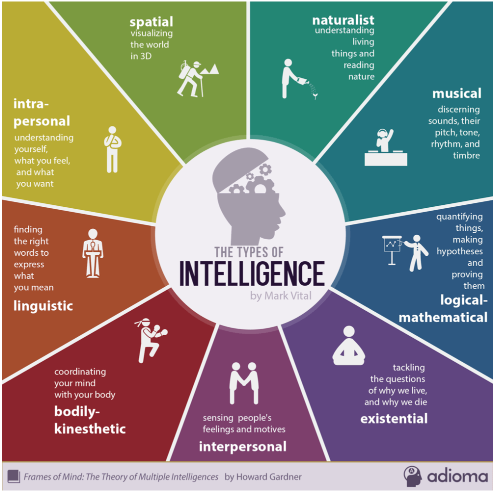
Atracktion: http://www.bsc.es/atracktion/rivers.html

Data

The dataset is the Whitburn Project list:

- it is a collective effort to gather historical weekly music sales rankings published by the Billboard company (the project is now maintained at the Bullfrogs Pond.
- The list goes back up to 1890, but data is more complete after 1954.
- The subset used contains 33560 songs.

Howard Gardner



https://blog.adioma.com/9-types-of-intelligence-infographic Adioma:

Several kinds of intelligence

- Logical-mathematical (number/reasoning smart)
- Bodily-kinesthetic (body smart)
- Linguistic (word smart)

- Naturalist (nature smart)
- Existential (life smart)
- Intra-personal (self smart)
- Interpersonal (people smart)

Music is about total engagement of those intelligence

Music and Emotion

Music is about emotion

- Most of us absolutely love music.
- We are compelled by it.
- We are provoked by it.
- We are moved by it. We are inspired by it.
- We feel connected to it.
- It reflects something profound about who we are and our experience of the world.

https://www.psychologytoday.com/us/blog/naked-truth/201410/music-is-what-feelings-sound

Another tentative definition

As a conclusion, music can be considered

- as an artifact
- but the important thing is what this artifact generates in terms of human responses.

Music is thus "**what feeling sound is like**".

Dr. Lagrange :)

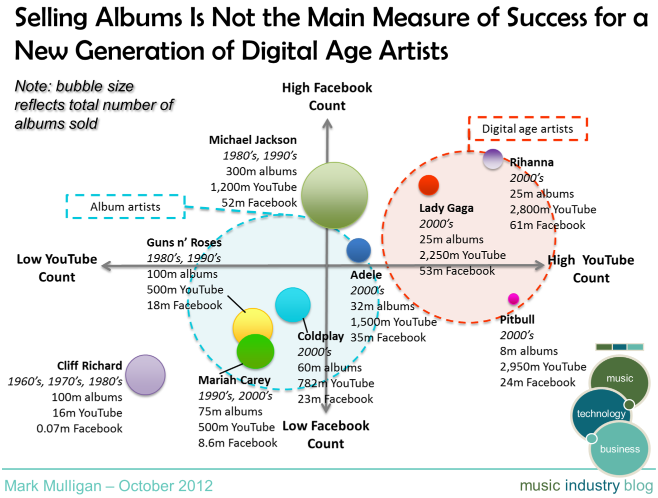
1. Music
2. Data
3. Recommendation
4. Content based Analysis

Music IS Data

That being said, music nowadays is:

- part of the digital revolution
- its artifact definition is present in many digital forms
- that follows a production / consumption pipeline that has drastically evolved

Success ??


Selling Albums Is Not the Main Measure of Success for a New Generation of Digital Age Artists
Note: bubble size reflects total number of albums sold
High Facebook Count
Low Facebook Count
Low YouTube Count
High YouTube Count
Digital age artists
Album artists
Michael Jackson
1980's, 1990's
300m albums
1,200m YouTube
52m Facebook
Rihanna
2000's
25m albums
2,800m YouTube
61m Facebook
Lady Gaga
2000's
25m albums
2,250m YouTube
53m Facebook
Guns n' Roses
1980's, 1990's
100m albums
500m YouTube
18m Facebook
Adele
2000's
32m albums
1,500m YouTube
35m Facebook
Pitbull
2000's
8m albums
2,950m YouTube
24m Facebook
Coldplay
2000's
60m albums
782m YouTube
23m Facebook
Cliff Richard
1960's, 1970's, 1980's
100m albums
16m YouTube
0.07m Facebook
Mariah Carey
1990's, 2000's
75m albums
500m YouTube
8.6m Facebook
music
technology
business
Mark Mulligan – October 2012
music industry blog

Data

The materials involved can be divided into 3 types of data that can be organized in an **emission, transmission, reception** sequence.

Data

- Creation: Score, Lyrics, Track metadata
- Production: Audio (live or studio)
- Consumption: Reviews, Ratings, Download patterns, Micro-blogging, ...

Scale ?

Major music streaming providers

- itunes music
- spotify
- google play
- deezer
- youtube

Spotify

- 75M users
- 30M songs
- 1.5B playlists
- 1TB of user data every day
- number of songs added each day > 20 000

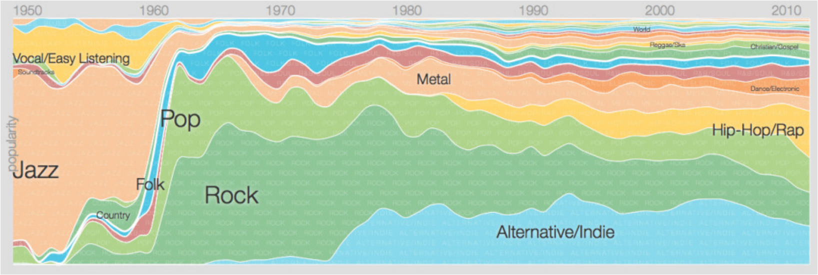
IT infrastructure:

- 42 PB storage,
- 200 TB data generated / day,
- 1300 servers

Youtube

- 100 hours of video are uploaded to YouTube every minute
- Content ID scans over 250 years of video every day
- among 15 million references

Genre through time



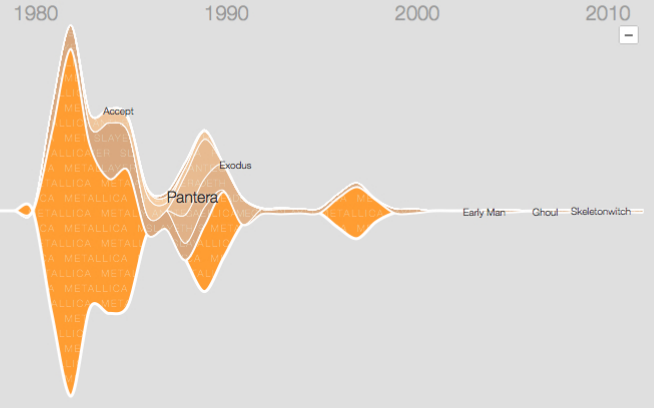
Google display: http://research.google.com/bigpicture/music

Data

The Music Timeline is based on album and artist statistics aggregated from Google Play Music

- **Popularity** : defined by how many users have an artist or album in their music library.
- **Normalization**: overview data is normalized by the total number of albums from that year.
-

Genre through time



Google display: http://research.google.com/bigpicture/music

Recommendation

A recommendation system

- is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item.
- Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags,
- and products in general.

Recommendation works !!

"[Amazon.com] recommendations generated a couple orders of magnitude more sales than just showing top sellers."

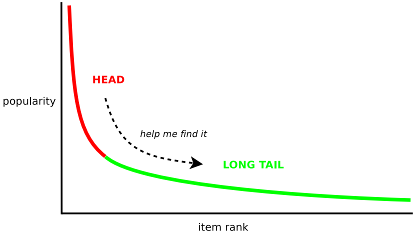
Greg Linden, who implemented the first recommendation engine for Amazon

Popularity

The popularity curve is composed by a small number of popular items (the hits), and the rest are located in the tail of the curve.

- the huge shift from physical media to digital media, and the fall in production costs
- make everything available, in contrast to the limitations of the brick-and-mortar stores.
- But, personalized recommendations and filters are needed to help users find the right content in this humongous digital space

The long tail problem



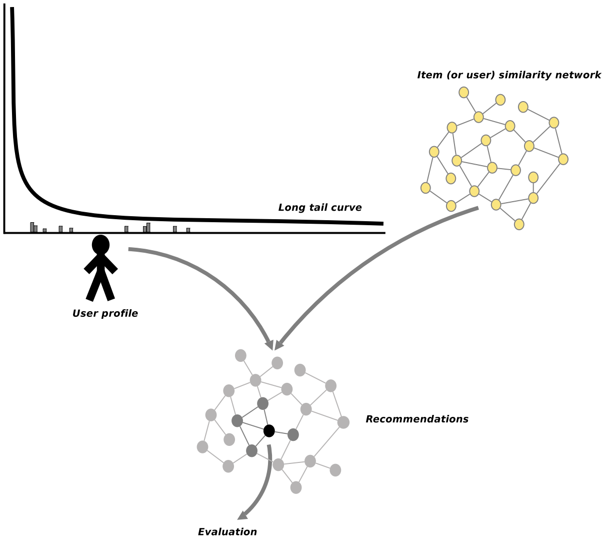
The Long Tail of items in a recommender system.
An important role of a recommender is to drive the user from the head region (popular items) to the long tail of the curve (Anderson, 2006)

Music consumption

Music consumption based on sales is biased towards a few popular artists.

- Ideally, by providing personalised filters and discovery tools to users, music consumption would diversify.
- There is a need to assist people to discover, recommend, personalise and filter the huge amount of music content.

Assumption about user behaviors

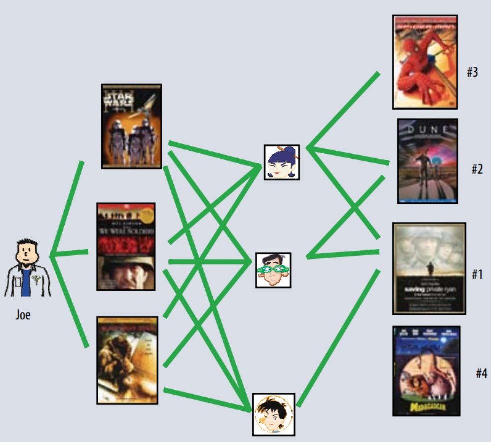


Netflix

Since October 2006, this field enjoyed an increase of interest thanks to the Netflix competition.

- The competition offers a prize of $1,000,000 to those that improve their movie recommendation system
- the Netflix competition provides the largest open dataset, containing more than 100 million movie ratings from anonymous users.
- The research community was challenged in developing algorithms to improve the accuracy of the current Netflix recommendation system.

Joe using Netflix



Collaborative Filtering ?

Insight: Personal preferences are correlated

- If Jack loves A and B, and Jill loves A, B, and C, then Jack is more likely to love C

Collaborative Filtering Task

- Discover patterns in observed preference behavior (e.g. purchase history, item ratings, click counts) across community of users
- Predict new preferences based on those patterns

Does not rely on item or user attributes (e.g. demographic info, author, genre)

- Content-based filtering: complementary approach

Collaborative Filtering ?

Insight: Personal preferences are correlated

- If Jack loves A and B, and Jill loves A, B, and C, then Jack is more likely to love C

Collaborative Filtering Task

- Discover patterns in observed preference behavior (e.g. purchase history, item ratings, click counts) across community of users
- Predict new preferences based on those patterns

Does not rely on item or user attributes (e.g. demographic info, author, genre)

- Content-based filtering: complementary approach

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \ldots, U\}$
- Items $i \in \{1, \ldots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \ldots & 4 \\ 3 & ? & ? & \ldots & ? \\ ? & 5 & ? & \ldots & 5 \end{bmatrix}\right\} U \text{ users}}_{M \text{ items}}$$

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \ldots, U\}$
- Items $i \in \{1, \ldots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \ldots & 4 \\ 3 & ? & ? & \ldots & ? \\ ? & 5 & ? & \ldots & 5 \end{bmatrix}\right\}}_{M \text{ items}} U \text{ users}$$

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \ldots, U\}$
- Items $i \in \{1, \ldots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \ldots & 4 \\ 3 & ? & ? & \ldots & ? \\ ? & 5 & ? & \ldots & 5 \end{bmatrix}\right\} U \text{ users}}_{M \text{ items}}$$

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \ldots, U\}$
- Items $i \in \{1, \ldots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \ldots & 4 \\ 3 & ? & ? & \ldots & ? \\ ? & 5 & ? & \ldots & 5 \end{bmatrix}\right\} U \text{ users}}_{M \text{ items}}$$

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \ldots, U\}$
- Items $i \in \{1, \ldots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \ldots & 4 \\ 3 & ? & ? & \ldots & ? \\ ? & 5 & ? & \ldots & 5 \end{bmatrix}\right\} U \text{ users}}_{M \text{ items}}$$

What is Collaborative Filtering?

Given:

- Users $u \in \{1, \dots, U\}$
- Items $i \in \{1, \dots, M\}$
- Training set *train* with observed, real-valued preferences r_{ui} for some user-item pairs (u, i)
  - r_{ui} = e.g. purchase indicator, item rating, click count ...

Goal: Predict unobserved preferences

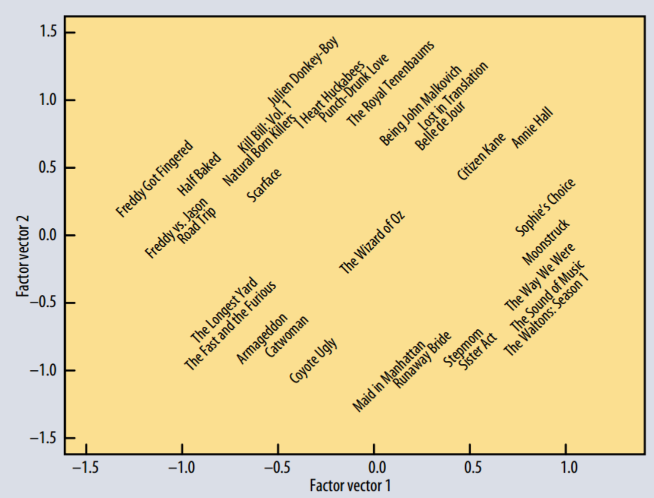
- Test set *test* with pairs (u, i) not in *train*

View as matrix completion problem

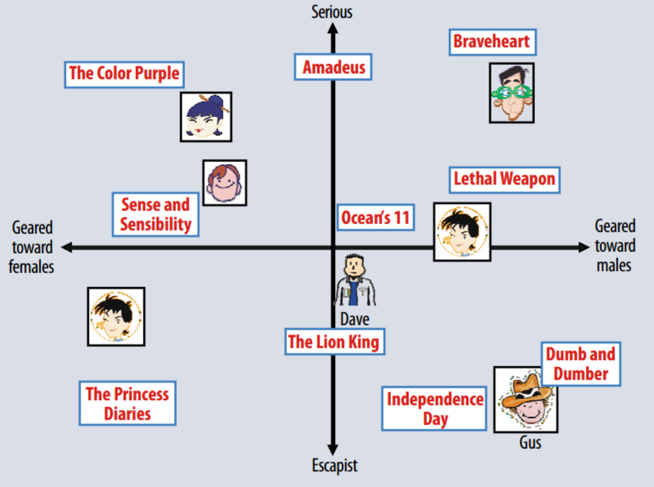
- Fill in unknown entries of sparse preference matrix

$$R = \underbrace{\left.\begin{bmatrix} ? & ? & 1 & \dots & 4 \\ 3 & ? & ? & \dots & ? \\ ? & 5 & ? & \dots & 5 \end{bmatrix}\right\} U \text{ users}}_{M \text{ items}}$$

Projection


Factor vector 2
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
Factor vector 1
Julien Donkey-Boy
I Heart Huckabees
Punch-Drunk Love
The Royal Tenenbaums
Being John Malkovich
Lost in Translation
Belle de Jour
Kill Bill: Vol. 1
Natural Born Killers
Freddy Got Fingered
Half Baked
Scarface
Citizen Kane
Annie Hall
Freddy vs. Jason
Road Trip
The Wizard of Oz
Sophie's Choice
Moonstruck
The Way We Were
The Sound of Music
The Waltons: Season 1
The Longest Yard
The Fast and the Furious
Armageddon
Catwoman
Coyote Ugly
Maid in Manhattan
Runaway Bride
Stepmom
Sister Act

Projecting movies and users



What is Collaborative Filtering?

Measuring success

- Interested in error on unseen test set *test*, not on training set
- For each (u, i) let r_{ui} = true preference, $\hat{r}_{ui}$ = predicted preference
- Root Mean Square Error
  - $\text{RMSE} = \sqrt{\frac{1}{|test|} \sum_{(u,i) \in test} (r_{ui} - \hat{r}_{ui})^2}$
- Mean Absolute Error
  - $\text{MAE} = \frac{1}{|test|} \sum_{(u,i) \in test} |r_{ui} - \hat{r}_{ui}|$
- Ranking-based objectives
  - e.g. What fraction of true top-10 preferences are in predicted top 10?

5

1. Music
2. Data
3. Recommendation
4. Content based Analysis

Music Information Retrieval

MIR is the interdisciplinary science of retrieving information from music. It has background in

- musicology, psychoacoustics, psychology
- signal processing, informatics, machine learning.

The analysis can be performed at 2 levels:

- song level analysis,
- collection level analysis.

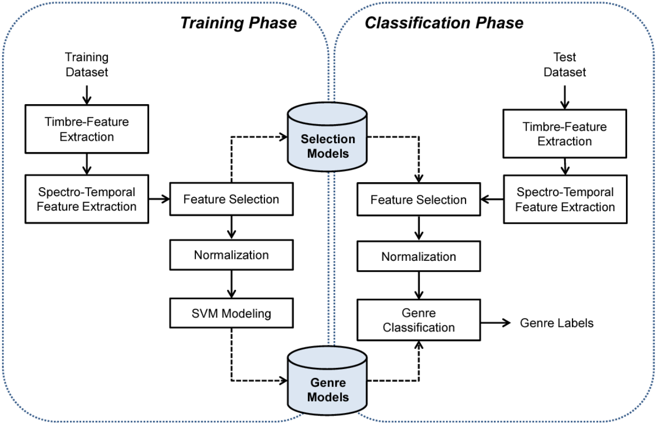
Song level analysis

- transcription
- rhythm, melody, chords
- demixing
- generation

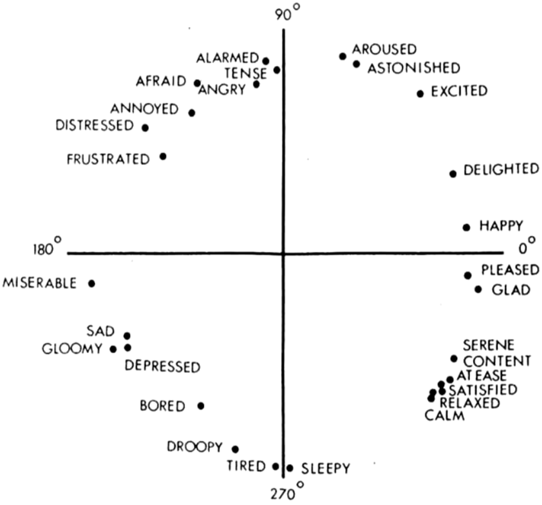
Collection level analysis

- genre classification
- mood regression
- fingerprinting
- cover song detection

Genre classification



Mood regression



Fingerprinting

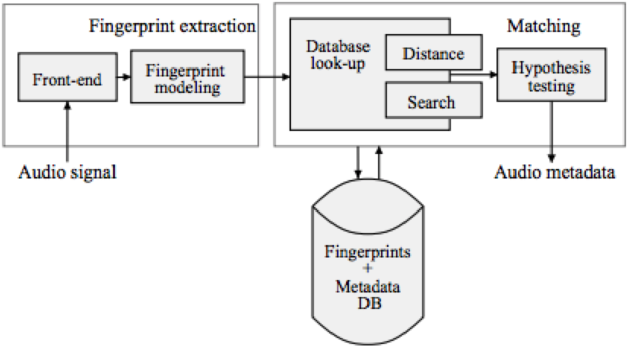
An acoustic fingerprint

- is a condensed digital summary, a fingerprint, deterministically generated from an audio signal
- that can be used to identify an audio sample or quickly locate similar items in an audio database.
-

Application scenarios:

- meta data retrieval: Shazam
- monitor the use of specific musical works and performances on radio broadcast, records, CDs, streaming media and peer-to-peer networks.
- This identification has been used in copyright compliance, licensing, and other monetization schemes.

Processing pipeline

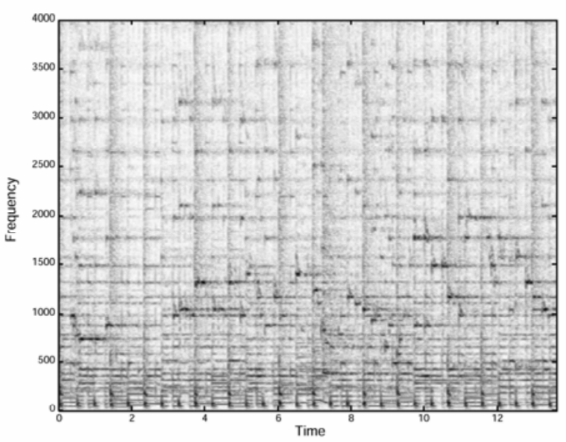


Cano & al A review of algorithms for audio fingerprinting 2002

Challenges

- Immunity to channel added noise
- Recognize fragments from anywhere in the track (the shorter, the better)
- Large corpus of reference items

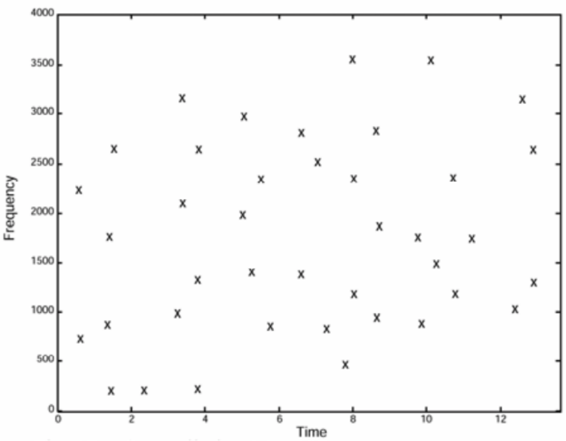
The Shazam algorithm



Spectrogram

Wang & al An industrial-strength audio search algorithm 2003

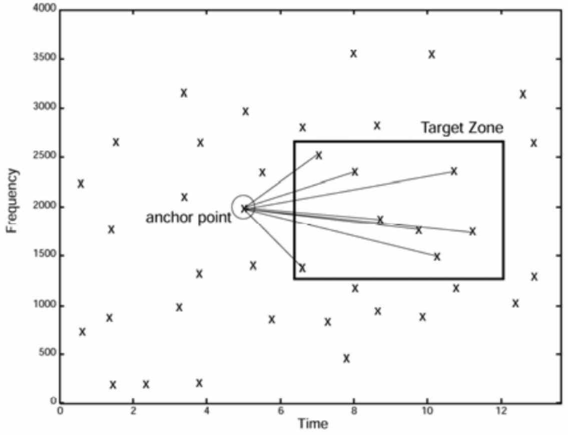
The Shazam algorithm



Peak selection

Wang & al An industrial-strength audio search algorithm 2003

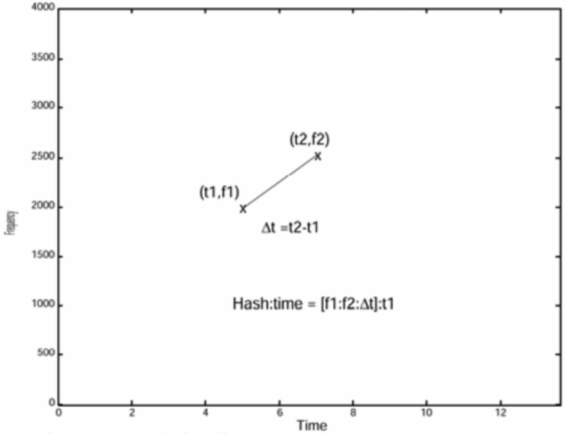
The Shazam algorithm



Pairing lookout

Wang & al An industrial-strength audio search algorithm 2003

The Shazam algorithm



Hash composition

Wang & al An industrial-strength audio search algorithm 2003

Cover song

A cover song is

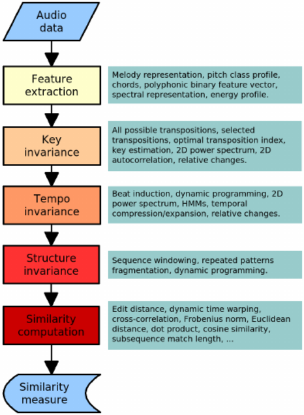
- a song initially played by a given band (say Let it Be from the Beatles) played by another band (cover from David Bowie).
- Even if the cover is easily and universally recognized by a human listeners (even non musically trained),
- the sound features that triggers the detection of covers are challenging to design.

Cover song detection

Cover songs may have variations

- in key, tempo, structure, harmony, timbre, language.
- Most of the time, melody and lyrics are largely preserved.
- Detection systems focus on melody.

Processing pipeline



Serra & al Audio cover song identification and similarity: background, approaches, evaluation, and beyond 2010

Building invariant representations

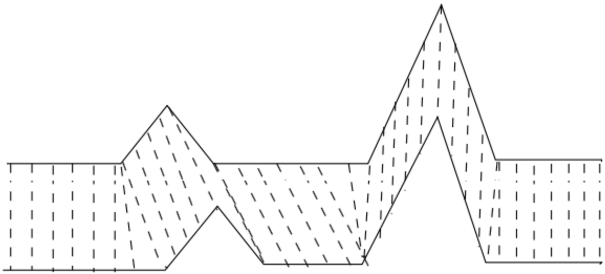
1. translation: phase shift
2. transposition: frequency shift
3. tempo change: time warping

Dynamic time warping

In time series analysis, dynamic time warping (DTW) is

- one of the algorithms for measuring similarity between two temporal sequences which may vary in speed.
- The sequences are "warped" non-linearly in the time dimension
- to determine a measure of their similarity independent of certain non-linear variations in the time dimension.

Sequence warping



DTW algorithm

```
for i := 1 to n
    for j := 1 to m
        cost := d(s[i], t[j])
        DTW[i, j] := cost + minimum(DTW[i-1, j  ],    // insertion
                                    DTW[i  , j-1],    // deletion
                                    DTW[i-1, j-1])    // match

return DTW[n, m]
```