UNIVERSITEIT GENT

FACULTY OF **ARTS AND PHILOSOPHY**

Music Information Retrieval

Opportunities for digital musicology

Joren Six

IPEM, University Ghent

October 30, 2015

Overview I

Overview II

Introduction

Goal

Give an overview of the Music Information Retrieval research field while focusing on the opportunities for digital musicology. More detail about two MIR projects will be given:
(i) Tarsos: tone scale extraction and analysis.
(ii) Panako: acoustic fingerprinting.

MIR introduction

Definition

Music Information Retrieval (MIR) is the **interdisciplinary** *science of extracting and processing* **information** *from music.*

MIR combines insights from musicology, computer science, library sciences, psychology, machine learning and cognitive sciences.

MIR introduction

MIR tasks process Musical information. Musical information can be categorized into signals and symbols.

Definition

Signals are representations of analog manifestations and replicate perception. Symbols are discretized, limited and replicate content.

Example: The task of transcribing a lecture is a conversion of a signal into the symbolic domain. An audio recording serves as input, a text is the output. The symbolic representation is easy to index but lacks nuance.

Tasks - Transcription

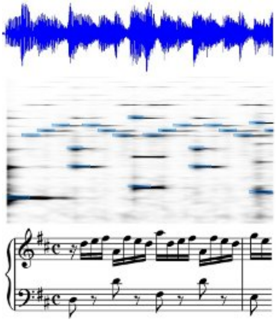


Fig: Music transcription

Transcription

- Source separation
- Instrument recognition
- Polyphonic pitch estimation and chord detection
- Tempo and Rhythm extraction

Signal $\rightarrow$ symbolic

Tasks - Structure analysis

Signal $\rightarrow$ symbolic

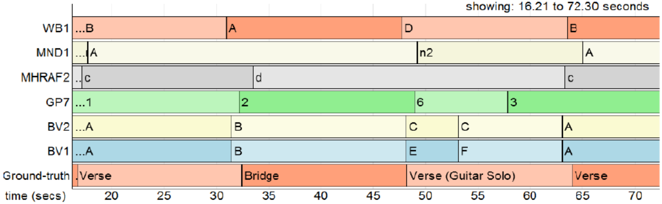


Fig: Structural analysis

Tasks - Music recommendation

Fig: Spotify automatically generates playlists based on listening behavior.

Music recommendation and automatic play-list generation.

- Content based: Signal $\rightarrow$ symbolic.
- Based on (listening) behavior: Symbolic $\rightarrow$ symbolic.

Tasks - Other Tasks

- Score following: automatic score page turning or trigger effects based on musical content.
- Emotion recognition: label audio according to emotional content.
- Automatic Cover song identification.
- Optical music recognition: convert images of scores to digital scores.
- Symbolic music retrieval.
- Automatic genre recognition.

MIR Tasks

Most tasks enable to browse, categorize, query, discover music in large databases.

Musical Information

Signals

- Recorded musical performances
  - Video
  - Audio
  - MIDI
  - Motion capture
- Scans of scores

Symbols

- Meta-data
  - Artist
  - Title
  - Album-name
  - Label
  - Composer
  - Instrumentation
- Lyrics
- Tags, reviews, ratings
- Digitized scores

Musical Information - Examples

Digital representations of Liszt's Liebestraum No.3.

Fig: Scanned score of Liszt's Liebestraum No.3.

- Scanned score
- MusicXML score
- MIDI synthesis
- MIDI performance
- Audio recording of a performance
  - Arthur Rubinstein
  - Daniel Barenboim

Musical Information

Scores can be seen as a model of a performance.

Quote

Essentially, all models are wrong, but some are useful.
- George E. P. Box

Models aim to reduce dimensions, complexity and improve understanding and readability.

'Solved' MIR Tasks

- Monophonic pitch estimation [4, 9, 12]
- Content based audio search [18]
- Automatic Genre classification

Challenging Tasks

Un-mix the mix

Decomposing a mixed audio signal is very hard. Masking, overlapping partials make e.g. polyphonic pitch detection hard.

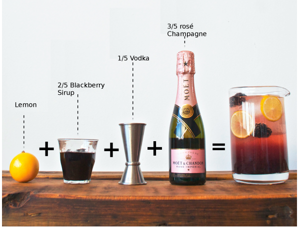


Fig: How to unmix the mix?

Tools - Sonic Visualizer

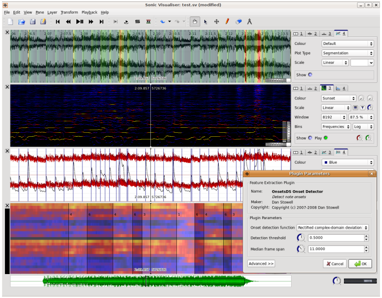


Fig: Sonic Visualizer, an application for viewing and analysing the contents of music audio files.

Sonic Visualizer offers a plugin-system with:

- Beat tracking
- Onset deteciton
- Pitch tracking
- Melody detection
- Chord estimations

sonicvisualiser.org

Tools - Tartini

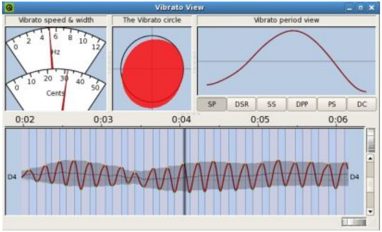


Fig: Tartini an application for pitch analysis.

Specialized tool for pitch analysis

- Vibrato analysis
- Pitch contour
- Transcription

http://miracle.otago.ac.nz/tartini

Tools - Music21

Fig: music21: programming environment for symbolic music analysis

Symbolic music queries:

- Query rhythmic features
- Melodic contours
- Chord progressions,...

http://web.mit.edu/music21/

Tools - Tarsos

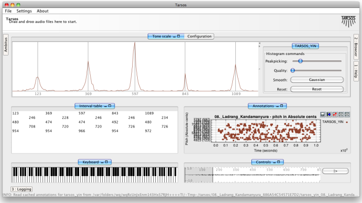


Fig: Tarsos: tone scale extraction and analysis

Extracting and analysing tone scales from music.

- Tone scale extraction
- Tone scale analysis
- Transcription of ethnic music

http://0110.be/Software

MIR Methods

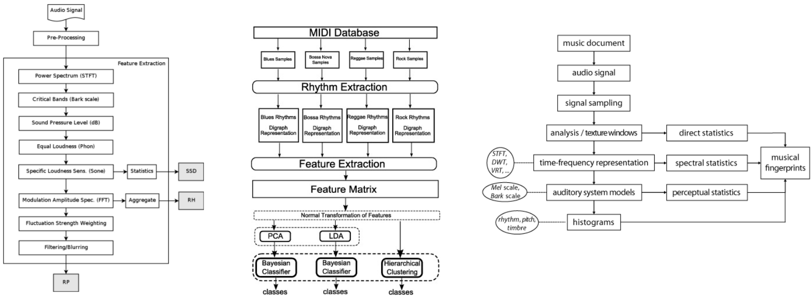


Fig: Input → feature(s) → feature processing → output.

MIR Methods

Bag of features approach to represent e.g. a musical genre. Sometimes more than 100 features are used[8].

- MFCC, timbral characteristic
- Spectral centroid
- Spectral moment
- Zero crossing rate
- Number of low energy frames
- Autocorrelation lag
- Frequency
- ...

Methodological problems

MIR research is often limited by (over?) simplification:

- It focuses mainly on classical western art music or popular music with ethnocentric terminology like scores, chords, tone scale, chromagrams, instrumentation, rhythmical structures.
- It is mainly goal oriented and pragmatic (MIREX) without explaining processes[1]. More engineering than science?
- Unclear which features correlate with which cognitive processes.
- It is mainly concerned with a limited, disembodied view on music: disregarding social interaction, movement, dance, the body, individual or cultural preferences.

Methodological problems

Quote

Essentially, all MIR-research is wrong, but some is useful.
- Me

What follows are two examples of what aims to be useful MIR-research.

Introduction

Tarsos

Tarsos[14, 15] is a tool to extract, analyze and document tone scales and tone scale diversity.

It is mainly useful for analyzing music with an undocumented tone-scale. This is the case for a lot of ethinic music.

Introduction

Tarsos was developed to analyze the dataset of the museum for Central Africa, Tervuren

- 30000 digitized sound recordings
- 3000 hours of music
- Meta-data database with contextual data

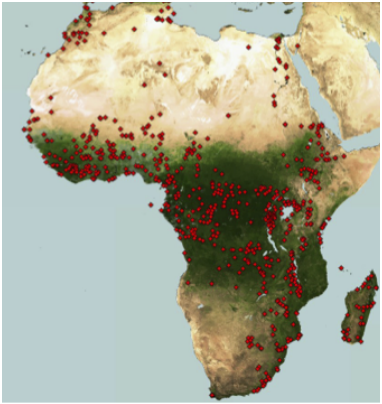


Fig: Locations of recordings

Demo

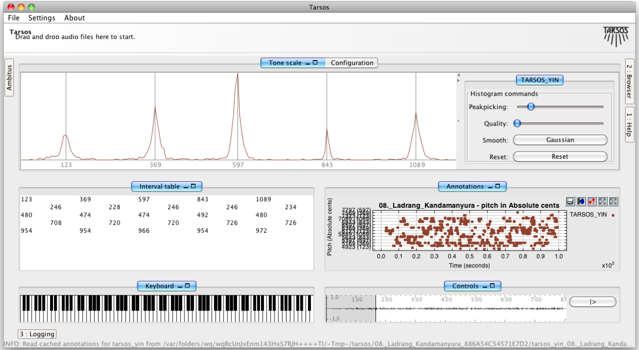


Fig: Tarsos live demonstration

Demo

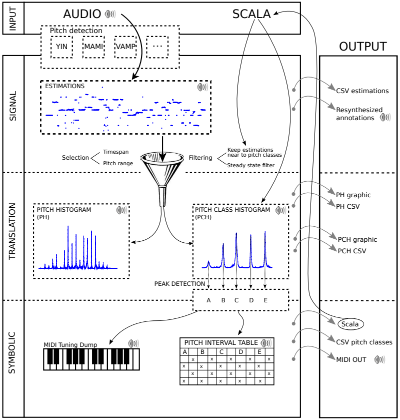


Fig: Tarsos block diagram.

Pitch Class Histogram construction

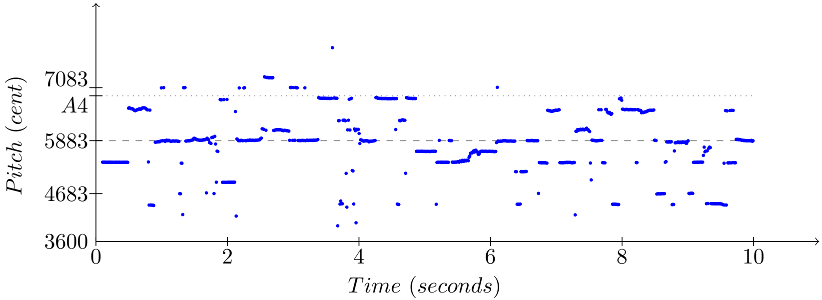


Fig: Step 1, pitch estimation.

Pitch Class Histogram construction

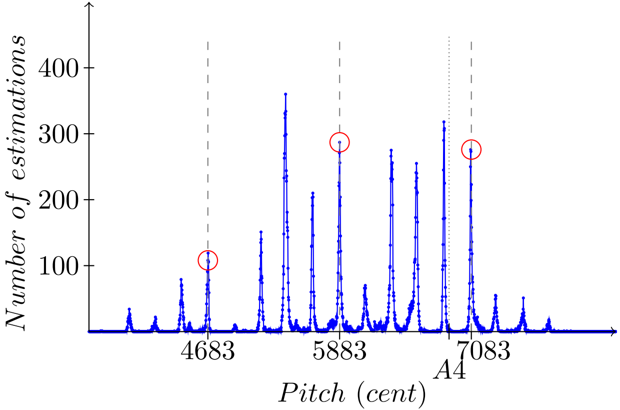


Fig: Step 2, pitch histogram creation.

Pitch Class Histogram construction

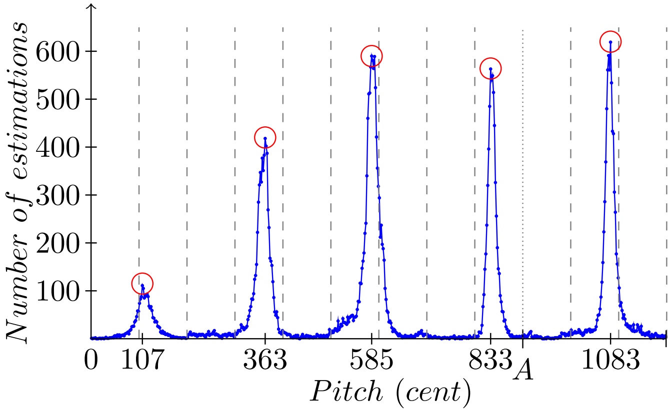


Fig: Step 3, pitch class histogram creation.

Examples

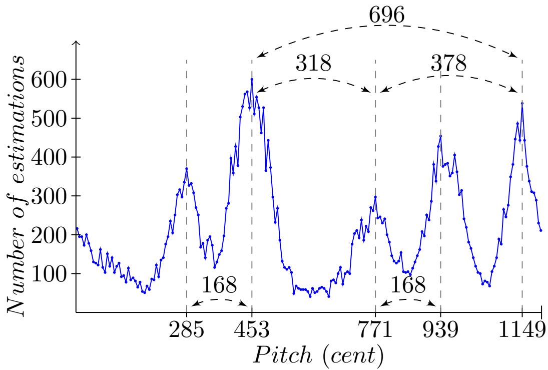


Fig: A unequally divided pentatonic tone scale with a near perfect fifth consisting of a pure minor and pure major third.

Concept of tone scale

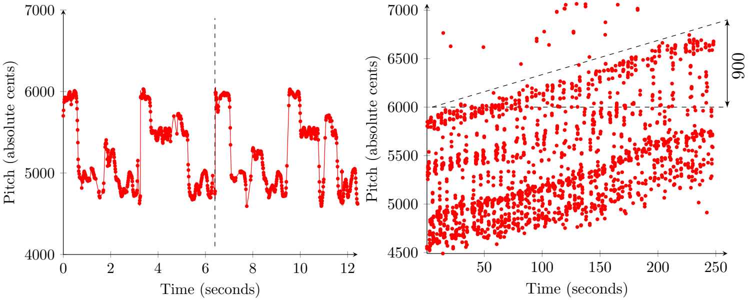


Fig: Pitch steps shift upwards during a Finnish joik.

Concept of Tone

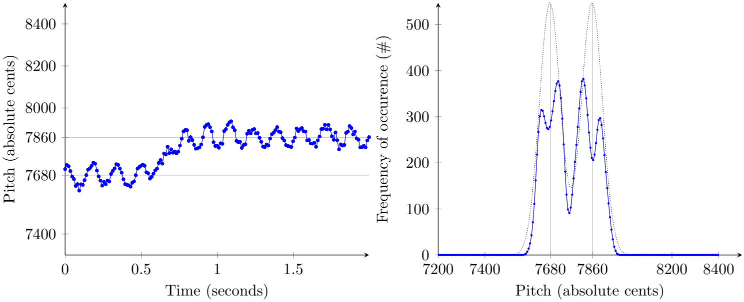


Fig: Tonal center of Western vibrato.

Concept of Tone II

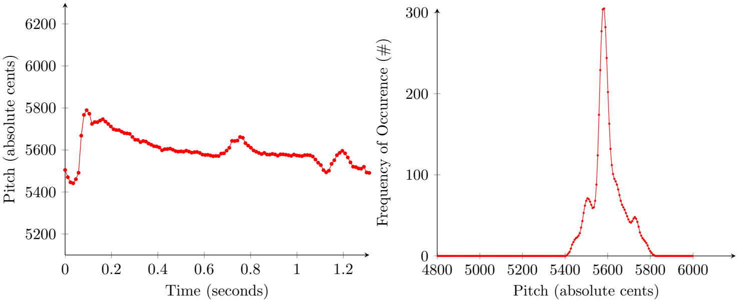


Fig: Pitch gesture in an Indian raga.

Concept of Tuning

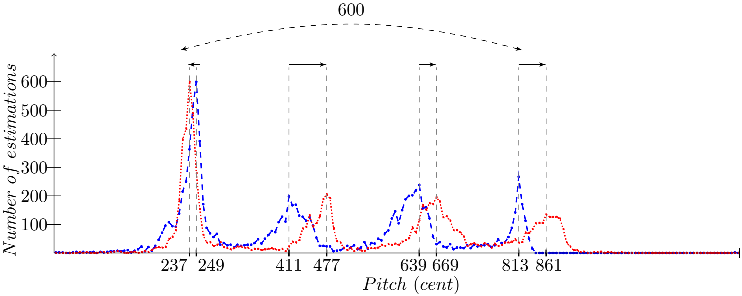


Fig: Detuning of a mono-chord during performance.

Relating Timbre and Scale

Question

Why are some tones scales or pitch intervals much more popular than others? Why are instruments tuned the way they are?

There is a theory[13, 10] that relates scale and timbre. The theory identifies points of maximum consonance that can be used to construct an optimal[1] scale.

[1] In terms of consonance

Relating Timbre and Scale

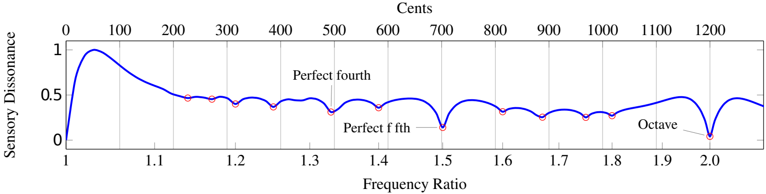


Fig: Dissonance curve for idealized harmonic instrument.

Relating Timbre and Scale

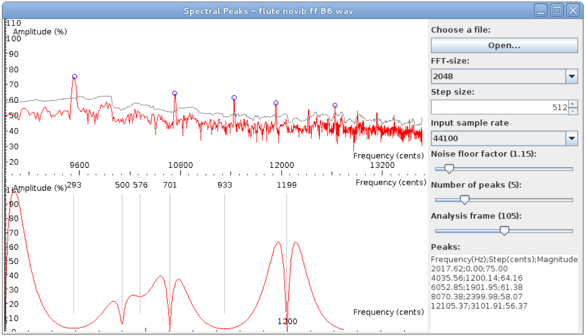


Fig: Screenshot of automatic timbre-scale mapping.

Relating Timbre and Scale

The Theory is currently not well supported by a lot of data. The dataset with African music has a large diversity in instrumentation and tone scales and offers an opportunity to support the theory.

Conclusion

Question

Tarsos offers opportunities to answer basic musicological questions:

- Is there a change in tone scale use over time? Is the 100 cents interval used more in recent years? Is there an acculturation effect?
- Is there a systematic relation between timbre and scale?

What is Acoustic Fingerprinting

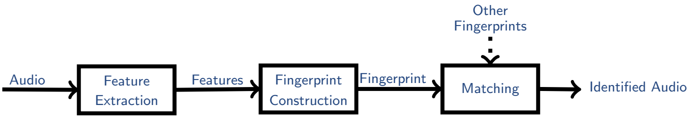


Figure: A generalized audio fingerprinter scheme.

1. Audio is fed into the system,
2. Features are extracted and fingerprints constructed
3. The fingerprints are compared with a database containing fingerprints of reference audio.
4. The audio is either identified or, if no match is found, labeled as unknown.

Why Audio Fingerprinting?

- Identifying short audio fragments
- Duplicate detection in large digital music archives
- Digital rights management applications (SABAM)
- Music structure analysis
- Analysis of techniques and repertoire in DJ-sets
- Synchronization of audio (and video) streams
- Alignment of extracted features with audio[17]

Fig: Shazam music recognition service

Demo Panako

Panako[16]

System Design

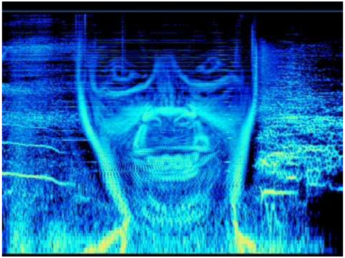


Fig: Spectrogram in Aphex Twin's *Windowlicker*

Current audio fingerprinting systems use fingerprints based on:

- **Spectral Peaks [18, 16, 6]**
- Onsets in spectral bands [5]
- Other features [2, 7, 11, 3]

System Design

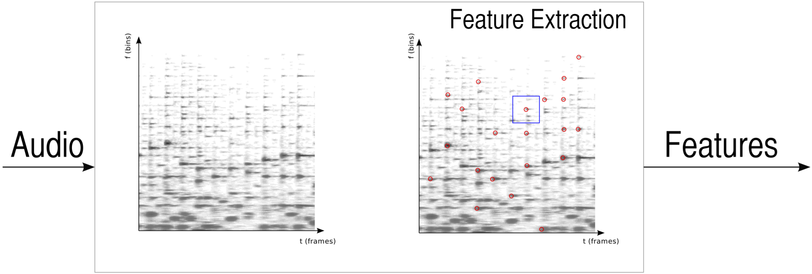


Fig: Step 1, extracting spectral peaks.

System Design

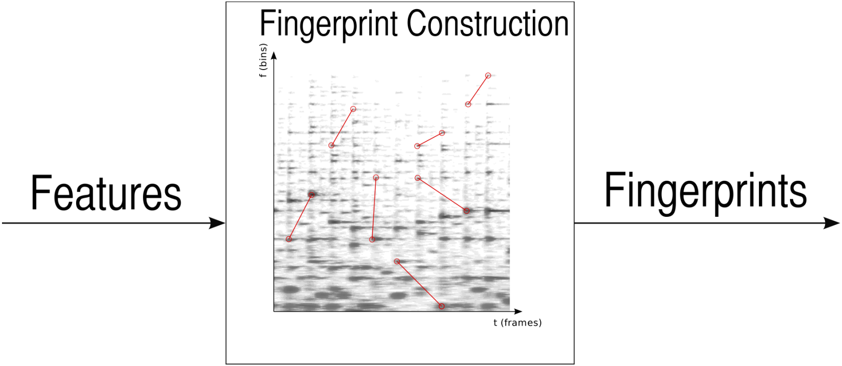
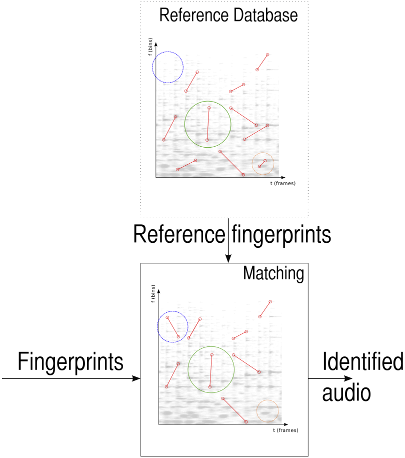


Fig: Step 2, creating fingerprints by combining spectral peaks.

System Design



Opportunities for digital musicology

Acoustic fingerprinting can provide opportunities for digital musicology:

1. Analysis of repetition within songs
2. Comparison of versions/edits
3. Audio and audio feature alignment to share datasets
4. DJ-set analysis

Musical structure analysis

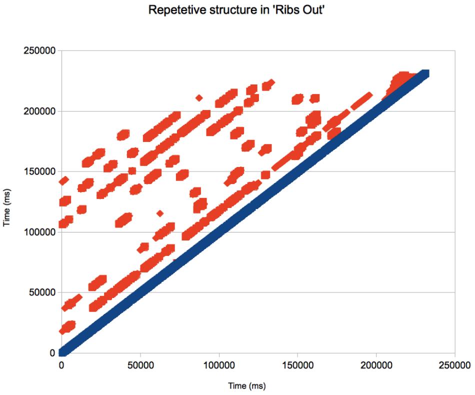


Fig: Repetition in 'Ribs Out' by Fuck Buttons[2].

[2]Unfortunately the best example I could find

Radio Edit vs. Original

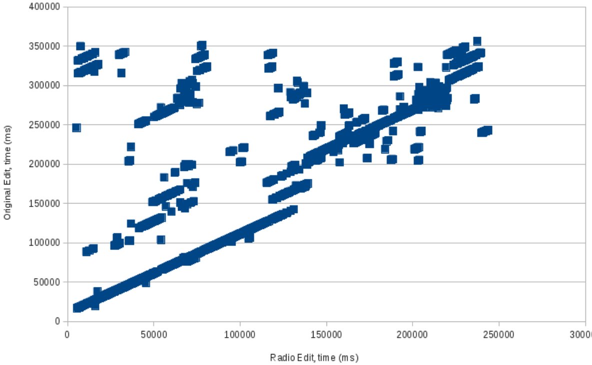


Fig: Radio edit vs. original version of Daft Punk's *Get Lucky*.

Exact Repetition Over Time

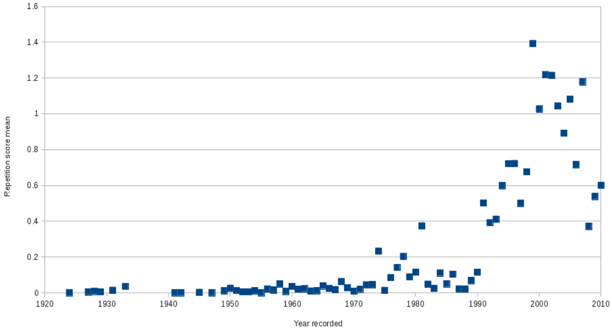


Fig: How much *cut-and-paste* is used on average for a set of 20000 recordings.

Synchronization of audio streams

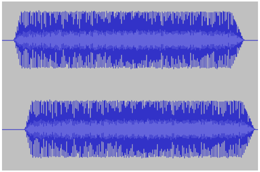


Fig: Two similar audio streams out of sync

Audio synchronization can be used for:

- Aligning unsynchronized audio streams from several microphones
- Aligning video footage by using audio
- Aligning audio and extracted features
- Aligning audio and data[17]

Synchronization of audio streams

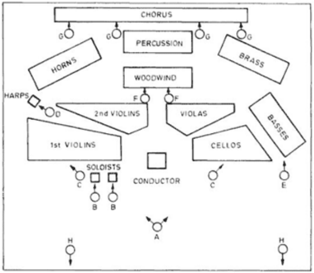


Fig: Microphone placement for symphonic orchestra and synchronization

Audio synchronization using acoustic fingerprinting is *submillisecond accurate*. If microphone placement spans several meters and with the speed of sound being 340.29m/s:

Distance (m)	Delay (ms)
1	3
2	6
3	9

Analysis of repertoire and techniques used in DJ-Sets

Fig: a DJ

An extension of the spectral peak fingerprinting method allows time-stretching, pitch-shifting and tempo change[16]. Given a DJ-set and reference audio[a] the following can be extracted automatically:

- Which parts of which songs were played and for how long
- Which modifications were applied (percentage modification of time and frequency)

[a] Tracklists of DJ-Sets can be found on `http://www.1001trackl`

Practical Audio Fingerprinting

Panako[16] was used to generate the example data[3], an open source audio fingerprinting system available on `http://panako.be`.

These subapplications of Panako were used:

- `monitor` during the live demo.
- `compare` for the comparison, structure analysis.
- `monitor` can also be used for DJ-set analysis.

Other usable fingerprinters are audfprint and echoprint.

[3] Some methods implemented within Panako are patented (US6990453).

Bibliography I

Pedro Cano, Eloi Batlle, Ton Kalker, and Jaap Haitsma.
A review of audio fingerprinting.
The Journal of VLSI Signal Processing, 41:271–284, 2005.

Michele Covell and Shumeet Baluja.
Known-Audio Detection using Waveprint: Spectrogram Fingerprinting by Wavelet Hashing.
In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, 2007.

Bibliography II

Alain de Cheveigné and Kawahara Hideki.
YIN, a Fundamental Frequency Estimator for Speech and Music.
The Journal of the Acoustical Society of America, 111(4):1917–1930, 2002.

Dan Ellis, Brian Whitman, and Alastair Porter.
Echoprint - an open music identification service.
In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011)*, 2011.

Bibliography III

- Sébastien Fenet, Gaël Richard, and Yves Grenier.
  A Scalable Audio Fingerprint Method with Robustness to Pitch-Shifting.
  In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011)*, pages 121–126, 2011.

- Jaap Haitsma and Ton Kalker.
  A highly robust audio fingerprinting system.
  In *Proceedings of the 3th International Symposium on Music Information Retrieval (ISMIR 2002)*, 2002.

Bibliography IV

Marc Leman, Dirk Moelants, Matthias Varewyck, Frederik Styns, Leon van Noorden, and Jean-Pierre Martens.
Activating and relaxing music entrains the speed of beat synchronized walking.
PLoS ONE, 8(7):e67932, 07 2013.

Phillip McLeod and Geoff Wyvill.
A Smarter Way to Find Pitch.
In *Proceedings of the International Computer Music Conference (ICMC 2005)*, 2005.

Bibliography V

R. Plomp and W. J. Levelt.
Tonal consonance and critical bandwidth.
Journal of the Acoustical Society of America, 38:548–560, 1965.

M. Ramona and G. Peeters.
AudioPrint: An efficient audio fingerprint system based on a novel cost-less synchronization scheme.
In *Proceedings of the 2013 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2013)*, pages 818–822, 2013.

Bibliography VI

M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley.
Average Magnitude Difference Function Pitch Extractor.
IEEE Trans. on Acoustics, Speech, and Signal Processing, 22(5):353–362, October 1974.

William A. Sethares.
Tuning Timbre Spectrum Scale.
Springer, 2 edition, 2005.

Bibliography VII

Joren Six and Olmo Cornelis.
Tarsos - a Platform to Explore Pitch Scales in Non-Western and Western Music.
In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011)*, 2011.

Joren Six, Olmo Cornelis, and Marc Leman.
Tarsos, a Modular Platform for Precise Pitch Analysis of Western and Non-Western Music.
Journal of New Music Research, 42(2):113–129, 2013.

Bibliography VIII

Joren Six and Marc Leman.
Panako - A Scalable Acoustic Fingerprinting System Handling Time-Scale and Pitch Modification.
In *Proceedings of the 15th ISMIR Conference (ISMIR 2014)*, 2014.

Joren Six and Marc Leman.
Synchronizing Multimodal Recordings Using Audio-To-Audio Alignment.
Journal of Multimodal User Interfaces, 9(3):223–229, 2015.

Bibliography IX

Avery L. Wang.
An Industrial-Strength Audio Search Algorithm.
In *Proceedings of the 4th International Symposium on Music Information Retrieval (ISMIR 2003)*, pages 7–13, 2003.