INTERNATIONAL AUDIO LABORATORIES ERLANGEN

AUDIO LABS

Cross-Modal Music Retrieval

Meinard Müller

The Jazzomat Project

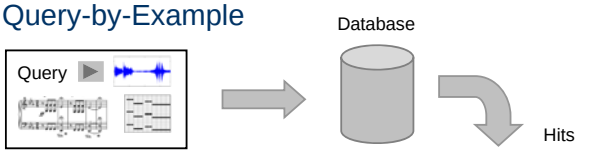
FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS

Music Retrieval

- Textual metadata
  - Traditional retrieval
  - Searching for artist, title, …
- Rich and expressive metadata
  - Generated by experts
  - Crowd tagging, social networks
- Content-based retrieval
  - Automatic generation of tags
  - Query-by-example

Query-by-Example



Retrieval tasks:

- Audio identification
- Audio matching
- Version identification
- Category-based music retrieval

Bernstein (1962)
Beethoven, Symphony No. 5

Beethoven, Symphony No. 5:
- Bernstein (1962)
- Karajan (1982)
- Gould (1992)

- Beethoven, Symphony No. 9
- Beethoven, Symphony No. 3
- Haydn Symphony No. 94

Query-by-Example

Taxonomy

Retrieval tasks:

- Audio identification
- Audio matching
- Version identification
- Category-based music retrieval

	Specificity level	**Granularity level**
	High specificity	Fragment-based retrieval
	Low specificity	Document-based retrieval

Overview

- Audio Identification
- Audio Matching
- Audio Analysis

Thanks:

- Frank Kurth
- Sebastian Ewert
- Michael Clausen
- Joan Serrà
- Peter Grosche
- Jonathan Driedger
- Stefan Balke

Overview

- Audio Identification
- Audio Matching
- Audio Analysis

Literature

- Wang “Shazam” (ISMIR 2003)
- Allamanche et al. (AES 2001)
- Cano et al. (AES 2002)
- Haitsma/Kalker (ISMIR 2002)
- Kurth/Clausen/Ribbrock (AES 2002)
- Dupraz/Richard (ICASSP 2010)
- Ramona/Peeters (ICASSP 2011)
- …

Audio Identification

Database: Huge collection consisting of all audio recordings (feature representations) to be potentially identified.

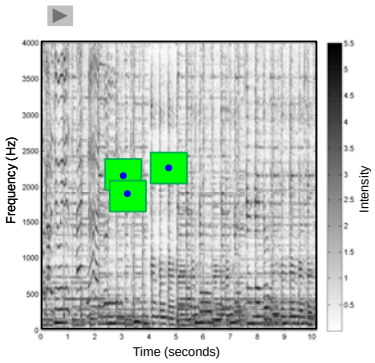
Goal: Given a short query audio fragment, identify the original audio recording the query is taken from.

Notes:
- Fragment-based retrieval
- High specificity

Application Scenario

- User hears music playing in the environment
- User records music fragment (5-15 seconds) with mobile phone
- Audio fingerprints are extracted from the recording and sent to an audio identification service
- Service identifies audio recording based on fingerprints
- Service sends back metadata (track title, artist) to user

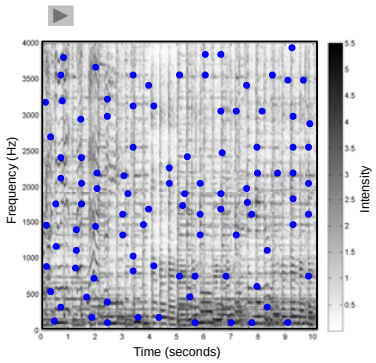
Fingerprints (Shazam)



Steps:
1. Spectrogram
2. Peaks (local maxima)

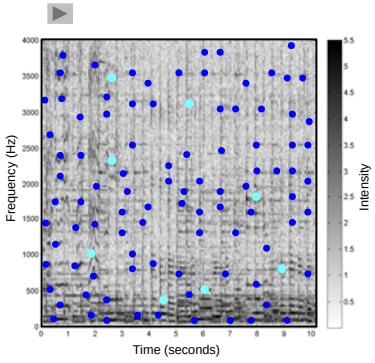
- *Efficiently computable*
- *Standard transform*
- *Robust*

Fingerprints (Shazam)



Steps:
1. Spectrogram
2. Peaks

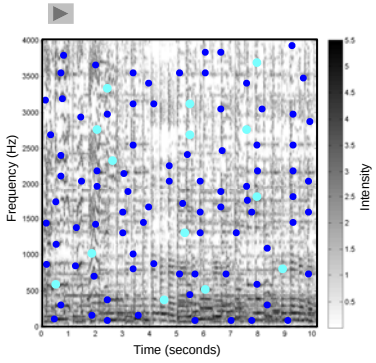
Fingerprints (Shazam)



Steps:
1. Spectrogram
2. Peaks / differing peaks

Robustness:
- Noise, reverb, room acoustics, equalization

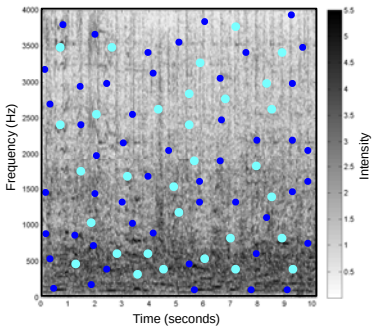
Fingerprints (Shazam)



Steps:
1. Spectrogram
2. Peaks / differing peaks

Robustness:
- Noise, reverb, room acoustics, equalization
- Audio codec

Fingerprints (Shazam)



Steps:

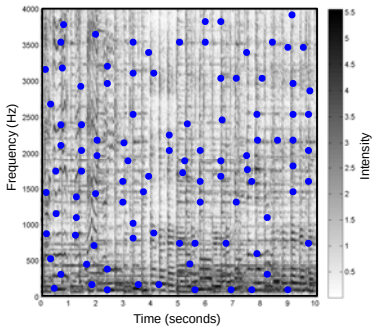
1. Spectrogram
2. Peaks / differing peaks

Robustness:

- Noise, reverb, room acoustics, equalization
- Audio codec
- Superposition of other audio sources

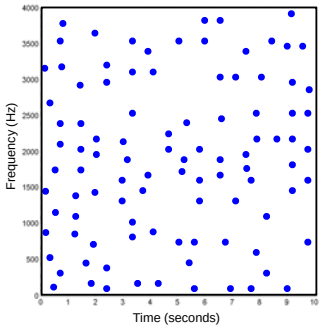
Matching Fingerprints (Shazam)

Database document



Matching Fingerprints (Shazam)

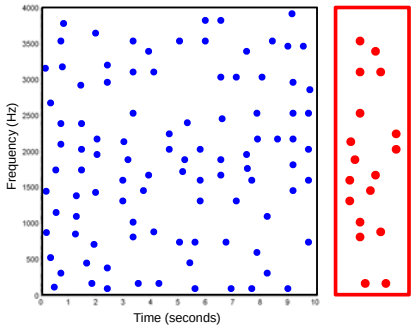
Database document (constellation map)



Matching Fingerprints (Shazam)

Database document (constellation map)

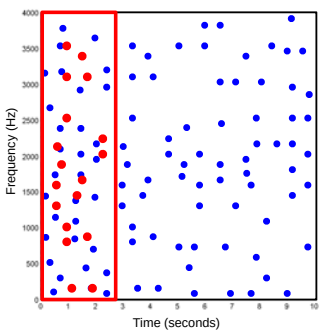
Query document (constellation map)



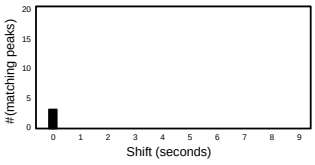
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



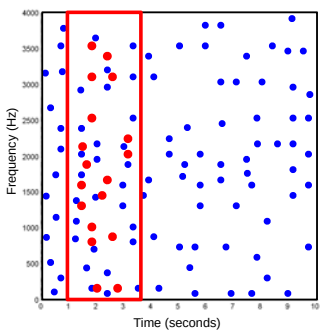
1. Shift query across database document
2. Count matching peaks



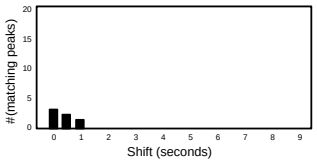
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



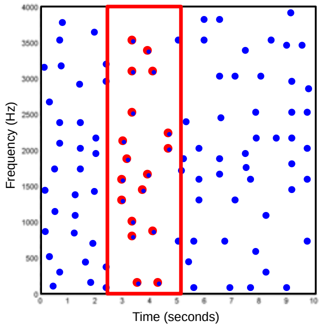
1. Shift query across database document
2. Count matching peaks



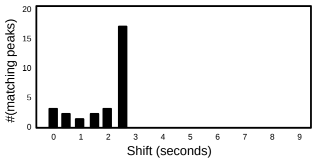
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



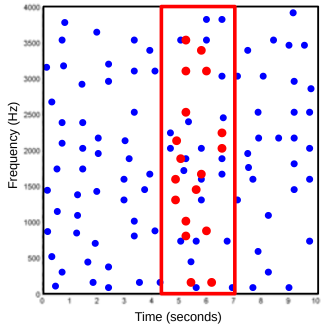
1. Shift query across database document
2. Count matching peaks



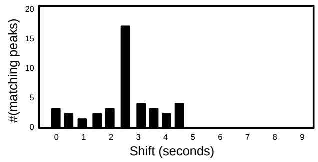
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



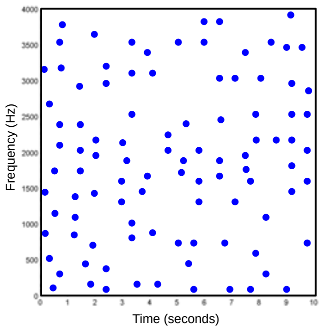
1. Shift query across database document
2. Count matching peaks



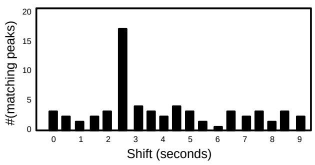
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



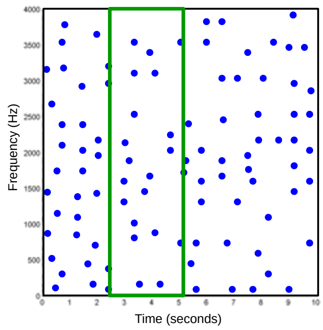
1. Shift query across database document
2. Count matching peaks



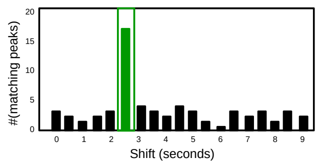
Matching Fingerprints (Shazam)

Database document (constellation map)

Query document (constellation map)



1. Shift query across database document
2. Count matching peaks
3. High count indicates a hit (document ID & position)



Summary (Audio Identification)

- Indexing crucial
- Delicate trade-off between specificity, robustness, and efficiency
- Fingerprint database
- Audio recording is identified (**not** a piece of music)
- Does not generalize to identify different interpretations or versions of the same piece of music

Overview

- Audio Identification
- Audio Matching
- Audio Analysis

Literature

- Casey et al. (IEEE TASLP 2008)
- Ellis/Polliner (ICASSP 2007)
- Kurth/Müller (IEEE TASLP 2008)
- Marolt (IEEE-TMM, 2008)
- Müller/Kurth/Clausen (ISMIR 2005)
- Pickens et al. (ISMIR 2002)
- Serrà et al. (IEEE TASLP 2008)
- Serrà (PhD 2011)
- Suyoto et al. (IEEE TASLP 2008)
- Yu et al. (ACM MM 2010)

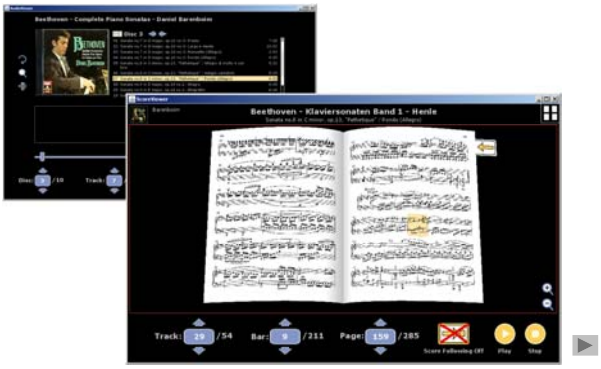
Audio Matching

Database: Audio collection containing:
- Several recordings of the same piece of music
- Different interpretations by various musicians
- Arrangements in different instrumentations

Goal: Given a short query audio fragment, find all corresponding audio fragments of similar musical content.

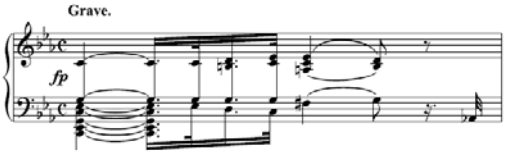
Notes:
- Fragment-based retrieval
- Medium specificity

Application Scenario

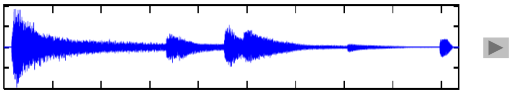


How to make the data comparable?

Image



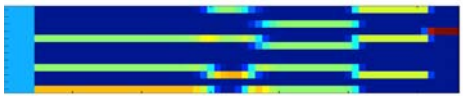
Audio



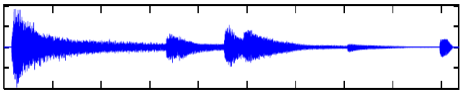
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



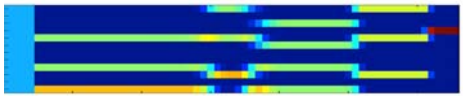
Audio



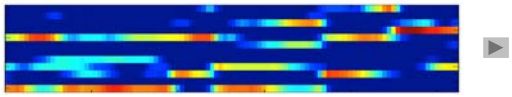
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio

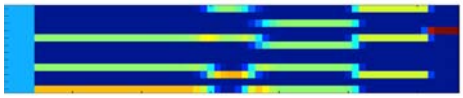


Audio Processing: Fourier Analyse

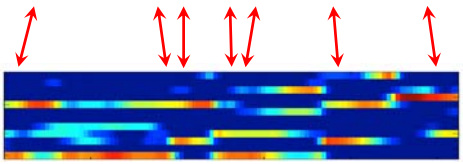
How to make the data comparable?

Image Processing: Optical Music Recognition

Image



Audio



Audio Processing: Fourier Analyse

Feature Representation

General goal: Convert an audio recording into a mid-level representation that captures certain musical properties while supressing other properties.

- Timbre / Instrumentation
- Tempo / Rhythm
- Pitch / Harmony

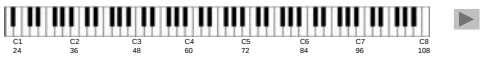
Feature Representation

General goal: Convert an audio recording into a mid-level representation that captures certain musical properties while supressing other properties.

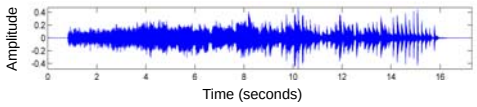
- Timbre / Instrumentation
- Tempo / Rhythm
- Pitch / Harmony

Feature Representation

Example: Chromatic scale

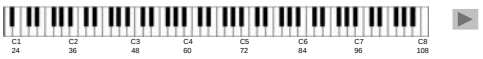


Waveform

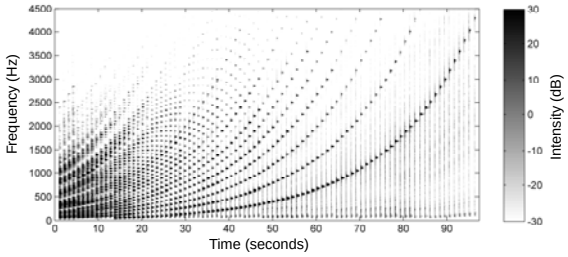


Feature Representation

Example: Chromatic scale

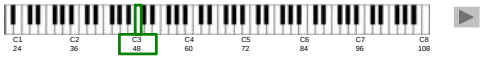


Spectrogram

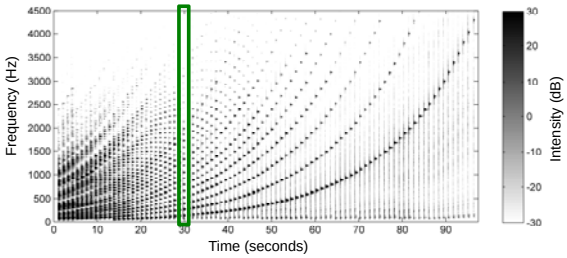


Feature Representation

Example: Chromatic scale

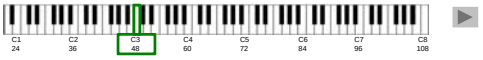


Spectrogram

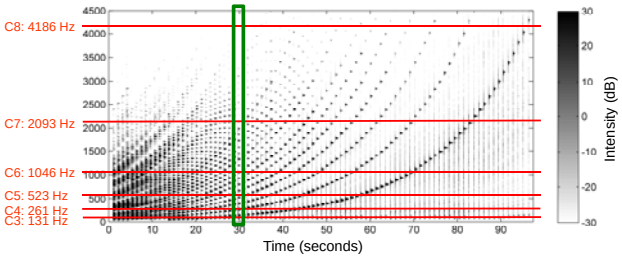


Feature Representation

Example: Chromatic scale

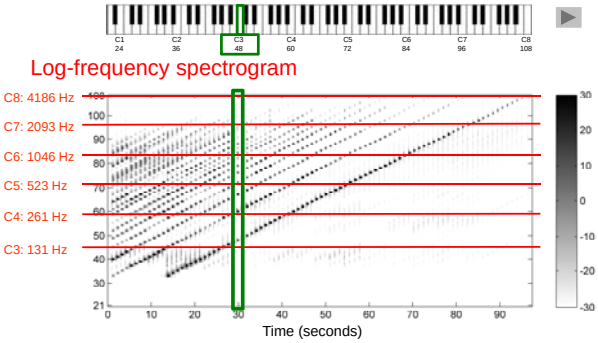


Spectrogram



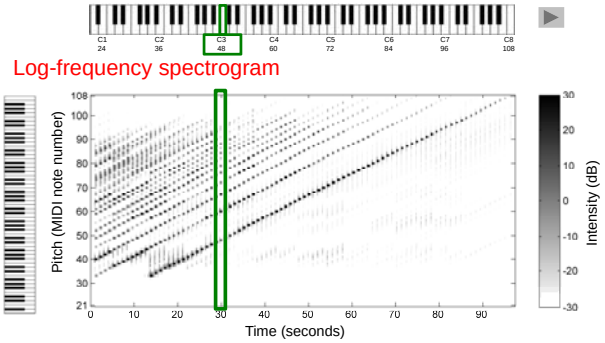
Feature Representation

Example: Chromatic scale



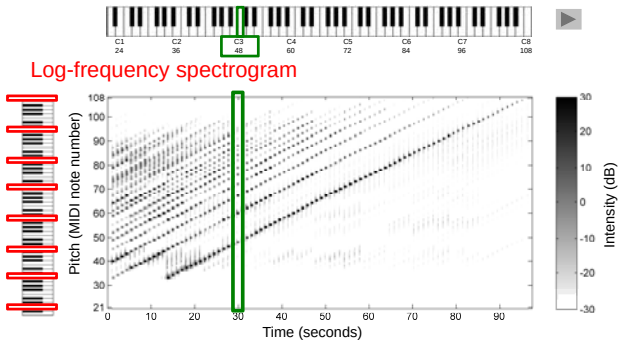
Feature Representation

Example: Chromatic scale



Feature Representation

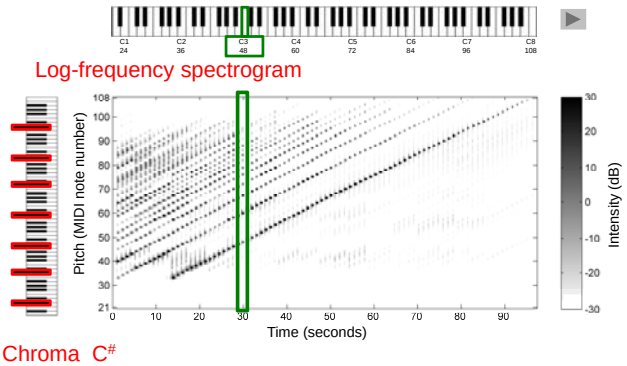
Example: Chromatic scale



Chroma C

Feature Representation

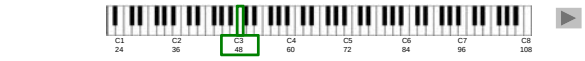
Example: Chromatic scale



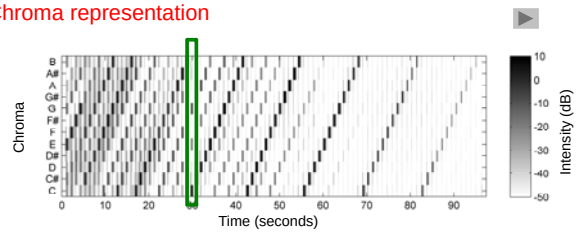
Chroma C#

Feature Representation

Example: Chromatic scale



Chroma representation



Overview

- Audio Identification
- Audio Matching
- Audio Analysis

Literature

- Müller/Ewert (IEEE TASLP 2010)

Audio Analysis

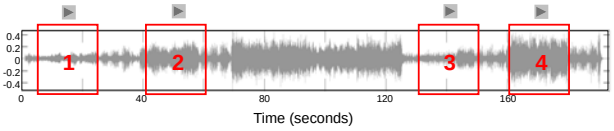
Idea:

Use “Audio Matching” for analyzing and understanding audio & feature properties:

- Relative comparison
- Compact
- Intuitive
- Quantitative evaluation

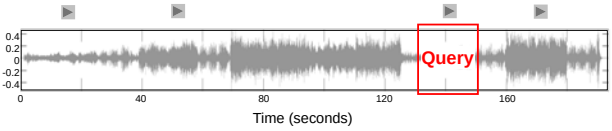
Audio Analysis

Example: Shostakovich, Waltz (Yablonsky)



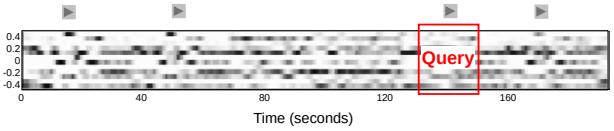
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



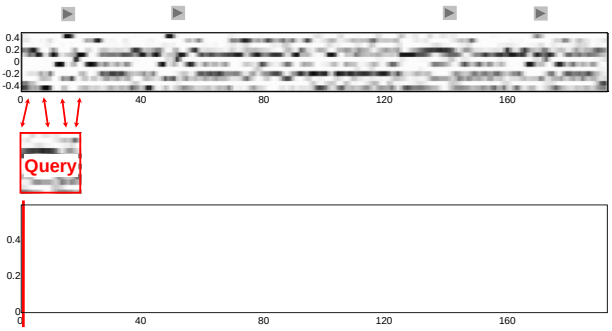
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



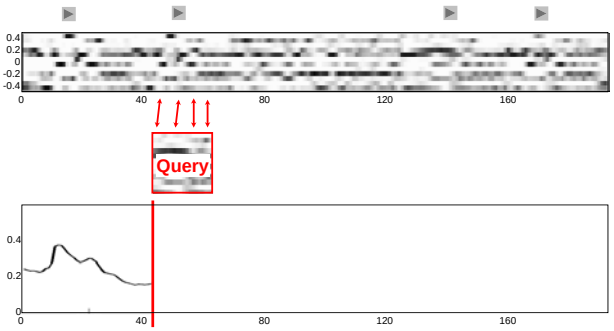
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



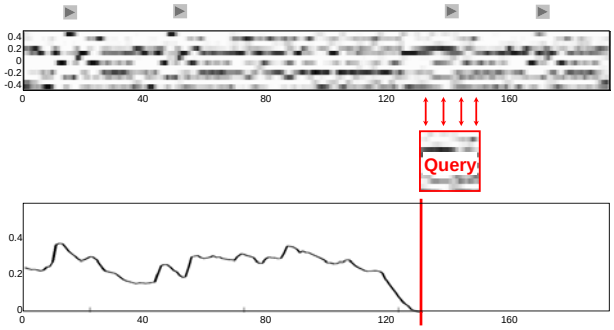
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



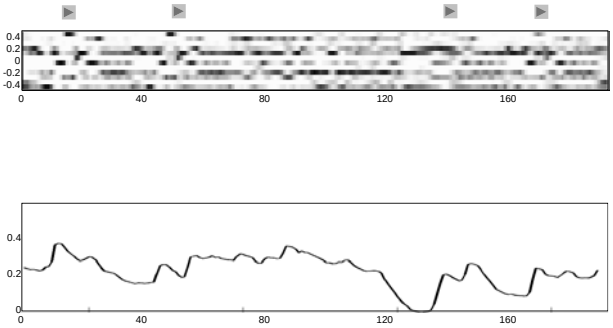
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



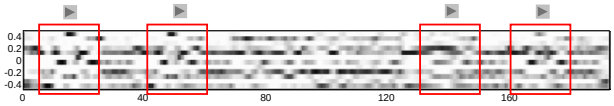
Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)

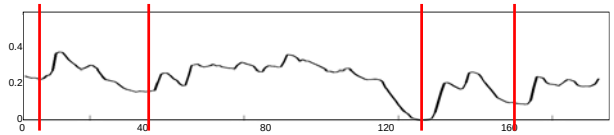


Audio Analysis

Query: Shostakovich, Waltz (Yablonsky)



Expected matching positions (should have local minima)

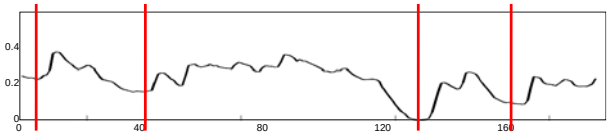


Audio Analysis

Idea:

- Use matching curve for analyzing feature properties

Expected matching positions (should have local minima)

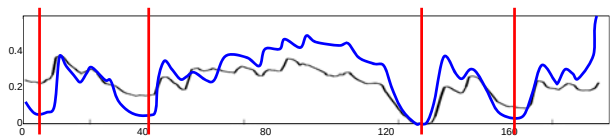


Audio Analysis

Idea:

- Use matching curve for analyzing feature properties
- Example: Chroma feature of higher timbre invariance

Expected matching positions (should have local minima)



Jazzomat

Task: Matching of music data of various types and formats

- Queries
  - Symbolic format (transcript)
  - Monophonic (solo)
- Database
  - Audio format
  - Polyphonic

Idea: Use the audio matching framework for designing musically relevant feature representations.

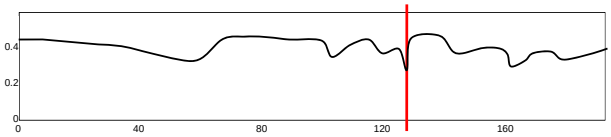
Jazzomat

Example: Solo by Wayne Shorter on “Down Under”

Query: (monophonic solo transcript)

Database: (real audio)

1. Idea: Use standard chroma features



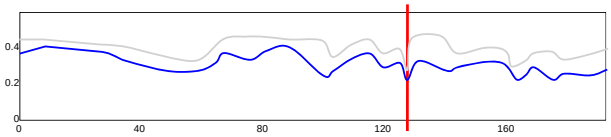
Jazzomat

Example: Solo by Wayne Shorter on “Down Under”

Query: (monophonic solo transcript)

Database: (real audio)

1. Idea: Use standard chroma features
2. Idea: Use only dominant chroma entry



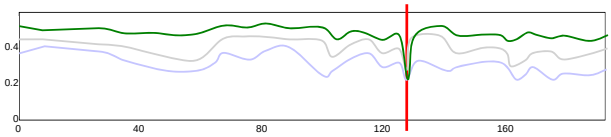
Jazzomat

Example: Solo by Wayne Shorter on “Down Under”

Query: (monophonic solo transcript)

Database: (real audio)

1. Idea: Use standard chroma features
2. Idea: Use only dominant chroma entry
3. Idea: Use chroma from salience spectrogram



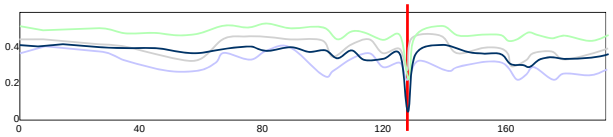
Jazzomat

Example: Solo by Wayne Shorter on “Down Under”

Query: (monophonic solo transcript)

Database: (real audio)

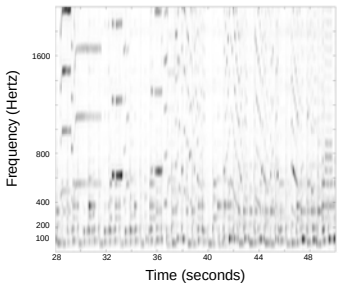
1. Idea: Use standard chroma features
2. Idea: Use only dominant chroma entry
3. Idea: Use chroma from salience spectrogram
4. Idea: Combine ideas from 2. and 3.



Jazzomat

Fundamental frequency (F0) estimation

Spectrogram

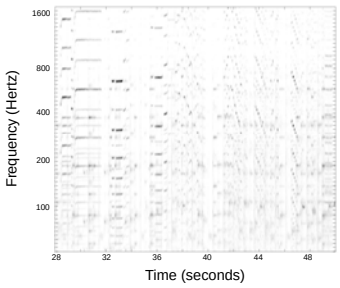


Original audio

Jazzomat

Fundamental frequency (F0) estimation

Salience log-spectrogram

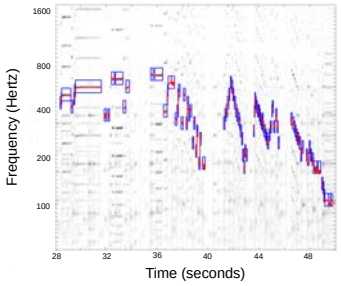


Original audio

Jazzomat

Fundamental frequency (F0) estimation

Score-informed F0

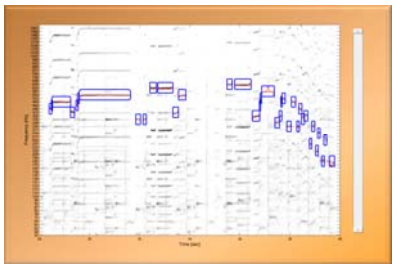


Original audio

F0 estimation

Jazzomat

Solo separation and editing

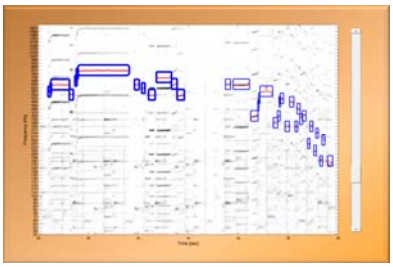


Original audio

Separated solo

Jazzomat

Solo separation and editing



Original audio

Separated solo

Pitch shift

Jazzomat

Fantastic dataset useful for various music processing tasks:

- Cross-modal music retrieval
- Melody (predominant F0) tracking
- Music transcription
- Beat tracking
- Performance analysis

Dataset only useful when including the audio material!

Book Project

A First Course on Music Processing

Textbook (approx. 500 pages)

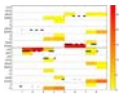
1. Music Representations
2. Fourier Analysis of Signals
3. Music Synchronization
4. Music Structure Analysis
5. Chord Recogntion
6. Temo and Beat Tracking
7. Content-based Audio Retrieval
8. Music Transcription

Springer

To appear (plan):
End of 2015

Projects with Musicology

DFG



Computergestützte Analyse harmonischer Strukturen

Kooperationspartner:
Prof. Rainer Kleinertz
Universität des Saarlandes
Institut für Musikwissenschaft

Bundesministerium für Bildung und Forschung

FREISCHÜTZ DIGITAL

Freischütz Digital

Kooperationspartner:
Prof. Joachim Veit, Universität Paderborn / Detmold
Prof. Thomas Betzwieser, Universität Frankfurt
Prof. Gerd Szwillus, Universität Paderborn

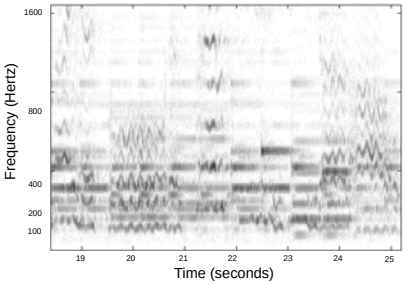
Chroma Toolbox

- http://www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/
- MATLAB implementations for various chroma variants

Schubert

Fundamental frequency (F0) estimation

Spectrogram

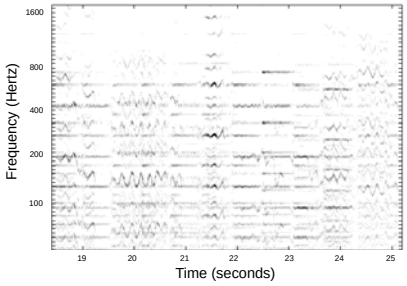


Original audio

Schubert

Fundamental frequency (F0) estimation

Salience log-spectrogram

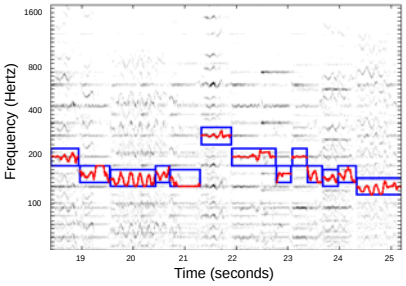


Original audio

Schubert

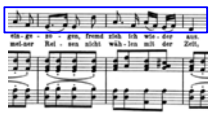
Fundamental frequency (F0) estimation

Score-informed F0



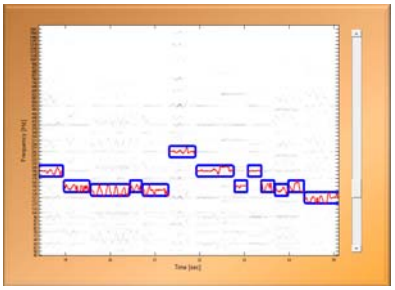
Original audio

F0 estimation



Schubert

Voice separation and editing

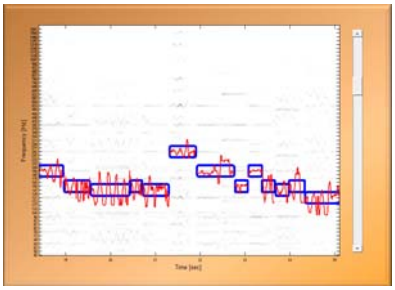


Original audio

Separated voice

Schubert

Voice separation and editing



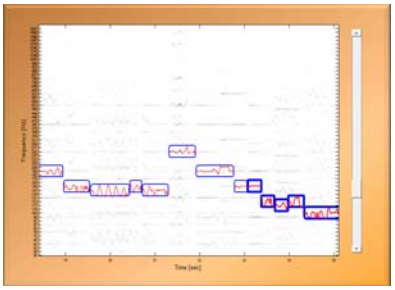
Original audio

Separated voice

Amplified vibrato

Schubert

Voice separation and editing



- Original audio
- Separated voice
- Amplified vibrato
- Pitch shift

References (Audio Identification)

- E. Allamanche, J. Herre, O. Hellmuth, B. Fröba, and M. Cremer.
  AudioID: Towards content-based identification of audio material.
  In *Proc. 110th AES Convention*, 2001.
- P. Cano, E. Batlle, T. Kalker, and J. Haitsma.
  A review of algorithms for audio fingerprinting.
  In *Proc. MMSP, pp.169–173*, 2002.
- P. Cano, E. Batlle, T. Kalker, and J. Haitsma.
  A review of audio fingerprinting.
  Journal of VLSI Signal Processing Systems, 41(3):271–284, 2005.
- E. Dupraz and G. Richard.
  Robust frequency-based audio fingerprinting.
  In *Proc. IEEE ICASSP, pp. 281-284,* 2010.
- J. Haitsma and T. Kalker.
  A highly robust audio fingerprinting system.
  In *Proc. ISMIR, pages 107–115,* 2002.
- Y. Ke, D. Hoiem, and R. Sukthankar.
  Computer Vision for Music Identification.
  In *Proc. IEEE CVPR, pages 597–604,* 2005.

References (Audio Identification)

- F. Kurth, A. Ribbrock, and M. Clausen.
  Identification of highly distorted audio material for querying large scale data bases.
  In *Proc. AES Convention, 2002.*
- M. Ramona and G. Peeters.
  Audio identification based on spectral modeling of Bark-bands energy and synchronization through onset detection.
  In *Proc. IEEE ICASSP, pp. 477-480*, 2011.
- J.S. Seo, J.Haitsma, and T. Kalker.
  Linear speed-change resilient audio fingerprinting.
  In *Proc. IEEE WMPCA*, 2002.
- A. Wang.
  An industrial strength audio search algorithm.
  In *Proc. ISMIR, pp. 7–13, 2003.*

References (Audio Matching)

- C. Fremerey, M. Müller, F. Kurth, and M. Clausen.
  Automatic mapping of scanned sheet music to audio recordings.
  In *Proc. ISMIR, pp. 413–418,* 2008.
- F. Kurth and M. Müller.
  Efficient index-based audio matching.
  IEEE Trans. on Audio, Speech, and Language Processing, 16(2):382–395, 2008.
- M. Müller.
  Information Retrieval for Music and Motion.
  Springer Verlag, 2007.
- M. Müller and S. Ewert.
  Towards timbre-invariant audio features for harmony-based music.
  IEEE Trans. on Audio, Speech, and Language Processing (TASLP), 18(3):649–662, 2010.
- M. Müller, F. Kurth, and M. Clausen.
  Audio matching via chromabased statistical features.
  In *Proc. ISMIR, pp. 288–*295, 2005.
- J. Pickens, J. P. Bello, G. Monti, M. Sandler, T. Crawford, and M. Dovey.
  Polyphonic score retrieval using polyphonic audio queries: a harmonic modeling approach.
  Journal of New Music Research, 32(2): 223-236, 2003.

References (Audio Matching)

- I. S. H. Suyoto, A. L. Uitdenbogerd, and F. Scholer.
  Searching musical audio using symbolic queries.
  IEEE Transactions on Audio, Speech & Language Processing, 16(2):372–381, 2008.
- Y. Yu, M. Crucianu, V. Oria, and L. Chen.
  Local summarization and multi-level LSH for retrieving multi-variant audio tracks.
  In *Proc. ACM Multimedia, pp. 341-350,* 2009.

References (Version Identification)

- M. Casey, C. Rhodes, and M. Slaney.
  Analysis of minimum distances in high-dimensional musical spaces.
  IEEE Trans. on Audio, Speech & Language Processing, 16(5), 2008.
- D. P. W. Ellis and G. E. Poliner.
  Identifying cover songs with chroma features and dynamic programming beat tracking.
  In *Proc. IEEE ICASSP, vol. 4, pp.1429–1432,* 2007.
- R. Foucard, J. L. Durrieu, M. Lagrange, and G. Richard.
  Multimodal similarity between musical streams for cover version detection.
  In *Proc. IEEE ICASSP, pp. 5514–5517,* 2010.
- H. Kantz and T. Schreiber.
  Nonlinear time series analysis, 2nd ed.
  Cambridge University Press, 2004.
- M. Marolt.
  A mid-level representation for melody-based retrieval in audio collections.
  IEEE Trans. on Multimedia, 10(8): 1617-1625, 2008.
- N. Marwan, M. C. Romano, M. Thiel, and J. Kurths.
  Recurrence plots for the analysis of complex systems.
  Physics Reports, 438(5): 237-329, 2007.

References (Version Identification)

- D. Müllensiefen and M. Pendzich.
  Court decisions on music plagiarism and the predictive value of similarity algorithms.
  Musicae Scientiae, Discussion Forum 4B, 207-238, 2009.
- S. Ravuri and D. P. W. Ellis.
  Cover song detection: from high scores to general classification.
  In *Proc. IEEE ICASSP, pp. 55–58,* 2010.
- J. Serrà.
  Identification of versions of the same musical composition by processing audio descriptions.
  PhD Thesis, Universitat Pompeu Fabra, 2011.
- J. Serrà, E. Gómez, P. Herrera, and X. Serra.
  Chroma binary similarity and local alignment applied to cover song identification.
  IEEE Trans. on Audio, Speech and Language Processing, 16(6):1138–1152, 2008.
- J. Serrà, X. Serra, and R. G. Andrzejak.
  Cross recurrence quantification for cover song identification.
  New Journal of Physics,11:093017, 2009.
- J. Serrà, H. Kantz, X. Serra, and R. G. Andrzejak.
  Predictability of music descriptor time series and its application to cover song detection.
  IEEE Trans. On Audio, Speech, and Language Processing. In Press.

References (Version Identification)

- W. H. Tsai, H. M. Yu, and H. M. Wang.
  Using the similarity of main melodies to identify cover versions of popular songs for music document retrieval.
  Journal of Information Science and Engineering, 24(6):1669–1687, 2008.

References (Category-Based Music Retrieval)

- C. M. Bishop.
  Pattern recognition and machine learning.
  Springer, 2007.
- D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra.
  Unifying low-level and high-level music similarity measures.
  IEEE Trans. on Multimedia, 13(4): 687-701, 2011.
- P. Cano and M. Koppenberger.
  Automatic sound annotation.
  IEEE Workshop on Machine Learning for Signal Processing, 2004.
- R. O. Duda, P. E. Hart, and D. G. Stork.
  Pattern classification, 2nd ed.
  John Wiley & Sons, 2000.
- Z. Fu, G. Lu, K. M. Ting, D. Zhang.
  A survey of audio-based music classification and annotation.
  IEEE Trans. on Multimedia, 13(2): 303-319, 2011.
- A. K. Jain, R. P. W. Duin, and J. Mao.
  Statistical pattern recognition: a review.
  IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(1): 4-37, 2000.

References (Category-Based Music Retrieval)

- C. Laurier, O. Meyers, J. Serrà, M. Blech, P. Herrera, and X. Serra.
  Indexing music by mood: design and implementation of an automatic content-based annotator.
  Multimedia Tools and Applications, 48(1): 161-184, 2010.
- M. Mandel and D. P. W. Ellis.
  Song-level features and support vector machines for music classification.
  In *Proc. of ISMIR, pp. 594- 599,* 2005
- N. Scaringella, G. Zoia, and D. Mlynek.
  Automatic genre classification of music content: a survey.
  IEEE Signal Processing Magazine, 23(2):133–141, 2006.
- J. Shen, M. Wang, S. Yan, H. Pang, and X. Hua.
  Effective music tagging through advanced statistical modeling.
  In *Proc. of SIGIR,* 2010.
- E. Tsunoo, T. Akase, N. Ono, and S. Sagayama.
  Musical mood classification by rhythm and bass-line unit pattern analysis.
  In *Proc. ICASSP, pp. 265-268,* 2010.
- D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet.
  Semantic annotation and retrieval of music sound effects.
  IEEE Trans. on Audio, Speech, and Language Processing, 16(2): 467-476, 2008.

References (Category-Based Music Retrieval)

- A. Tversky.
  Features of similarity.
  Psychological Reviews, 84(4): 327-352, 1977.
- G. Tzanetakis and P. Cook.
  Musical genre classification of audio signals.
  IEEE Trans. on Speech and Audio Processing, 5(10):293– 302, 2002.
- K. West.
  Novel techniques for audio music classification and search.
  PhD Thesis, University of East Anglia, 2008.
- I. A. Witten and E. Frank.
  Data mining: practical machine learning tools and techniques, 2nd ed.
  Elsevier, 2nd ed., 2005.
- L. M. Zbikowski.
  Conceptualizing music: cognitive structure, theory and analysis.
  Oxford University Press, 2002.