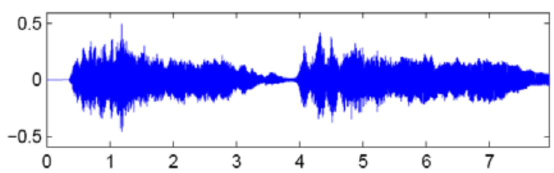
Lecture

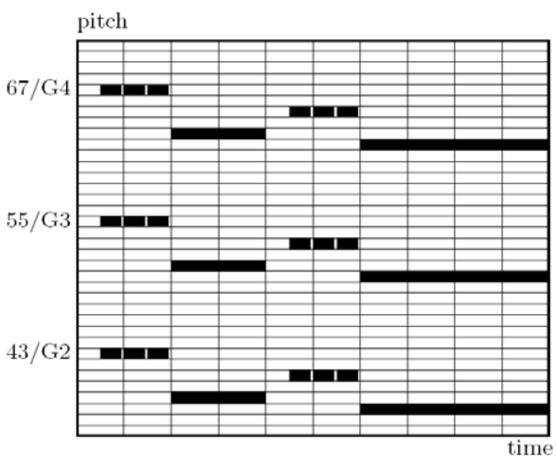
Music Processing

Music Synchronization

Meinard Müller
International Audio Laboratories Erlangen
meinard.mueller@audiolabs-erlangen.de

Music Data





Music Data

Various interpretations – Beethoven‘s Fifth

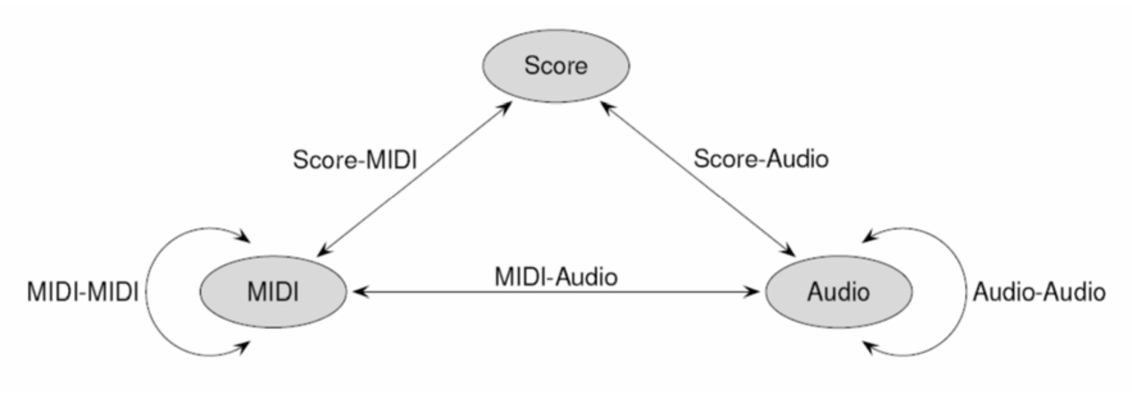
Bernstein	▶
Karajan	▶
Scherbakov (piano)	▶
MIDI (piano)	▶

General Goals

- Automated organization of complex and inhomogeneous music collections
- Generation of annotations and cross-links
- Tools and methods for multimodal search, navigation and interaction

Music Information Retrieval (MIR)

Music Synchronization



Schematic view of various synchronization tasks

Music Synchronization (Audio Alignment)

- Turetsky/Ellis (ISMIR 2003)
- Soulez/Rodet/Schwarz (ISMIR 2003)
- Arifi/Clausen/Kurth/Müller (ISMIR 2003)
- Hu/Dannenberg/Tzanetakis (WASPAA 2003)
- Müller/Kurth/Röder (ISMIR 2004)
- Raphael (ISMIR 2004)
- Dixon/Widmer (ISMIR 2005)
- Müller/Mattes/Kurth (ISMIR 2006)
- Dannenberg /Raphael (Special Issue ACM 2006)
- Kurth/Müller/Fremerey/Chang/Clausen (ISMIR 2007)
- Fujihara/Goto (ICASSP 2008)
- Wang/Iskandar/New/Shenoy (IEEE-TASLP 2008)
- Ewert/Müller/Grosche (ICASSP 2009)

Music Synchronization: Audio-Audio

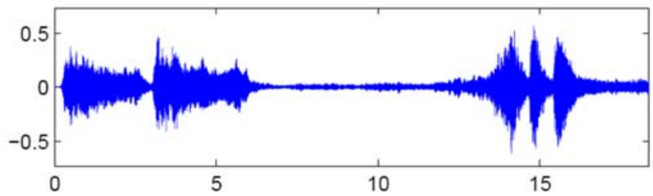
Given: Two different audio recordings of the same underlying piece of music.

Goal: Find for each position in one audio recording the musically corresponding position in the other audio recording.

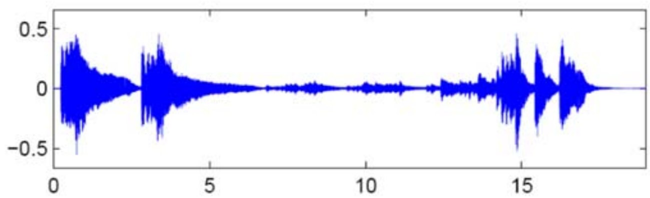
Music Synchronization: Audio-Audio

Beethoven‘s Fifth

Karajan



Scherbakov

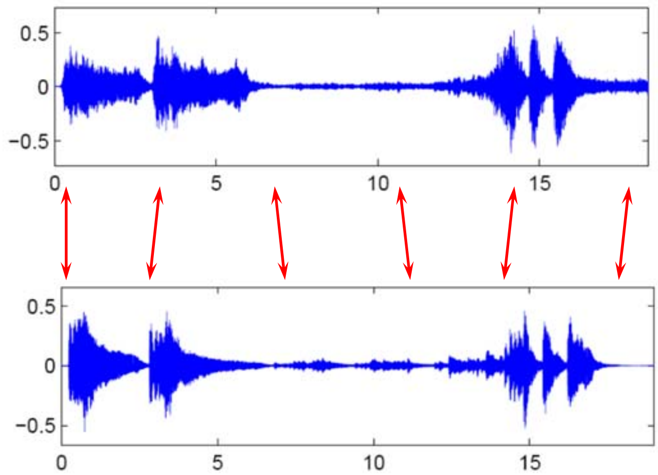


Music Synchronization: Audio-Audio

Beethoven's Fifth

Karajan

Scherbakov

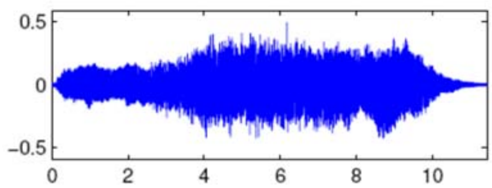


Synchronization: Karajan → Scherbakov

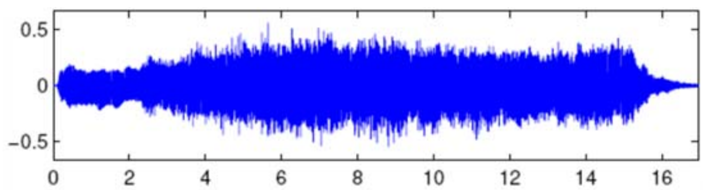
Music Synchronization: Audio-Audio

Bach Toccata

Koopman



Ruebsam

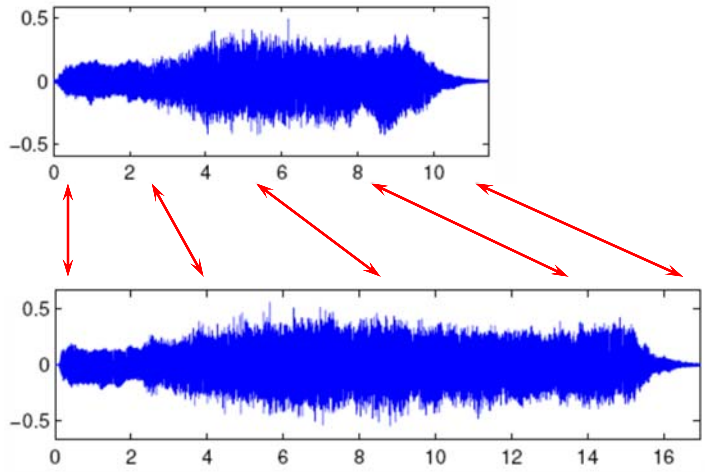


Music Synchronization: Audio-Audio

Bach Toccata

Koopman

Ruebsam



Synchronization: Koopman → Ruebsam

Music Synchronization: Audio-Audio

- Transformation of audio recordings into sequences of feature vectors

  $\rightsquigarrow \quad V := (v^1, v^2, \ldots, v^N)$

  $\rightsquigarrow \quad W := (w^1, w^2, \ldots, w^M)$

- Fix cost measure c on the feature space

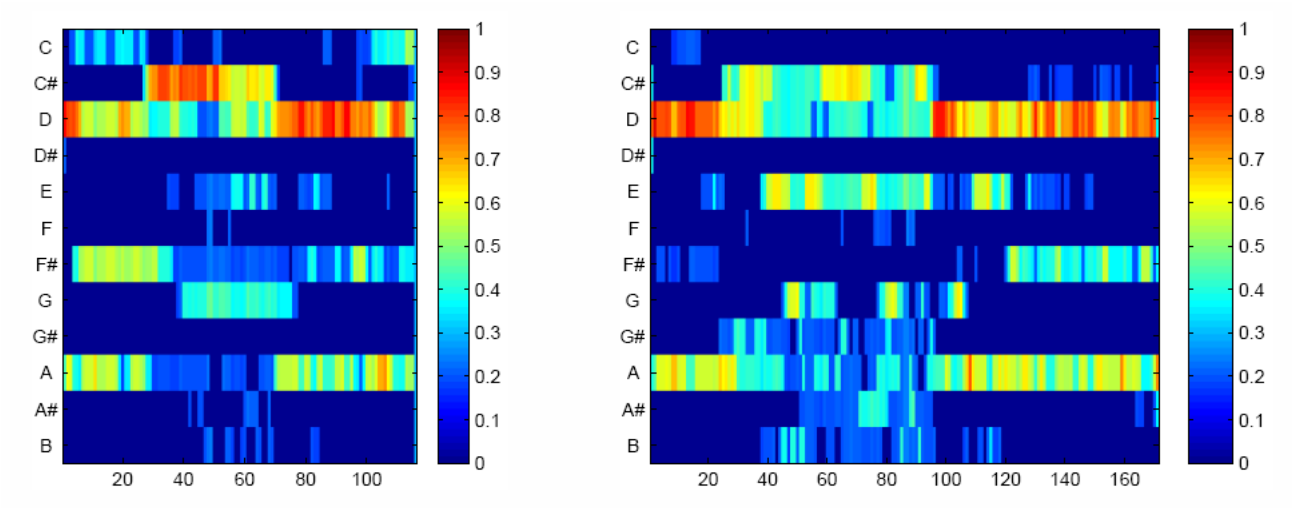
- Compute $N \times M$ cost matrix $C(n,m) := c(v^n, w^m)$

- Compute cost-minimizing warping path from C

Chroma Features

Example: Bach Toccata

Koopman Ruebsam

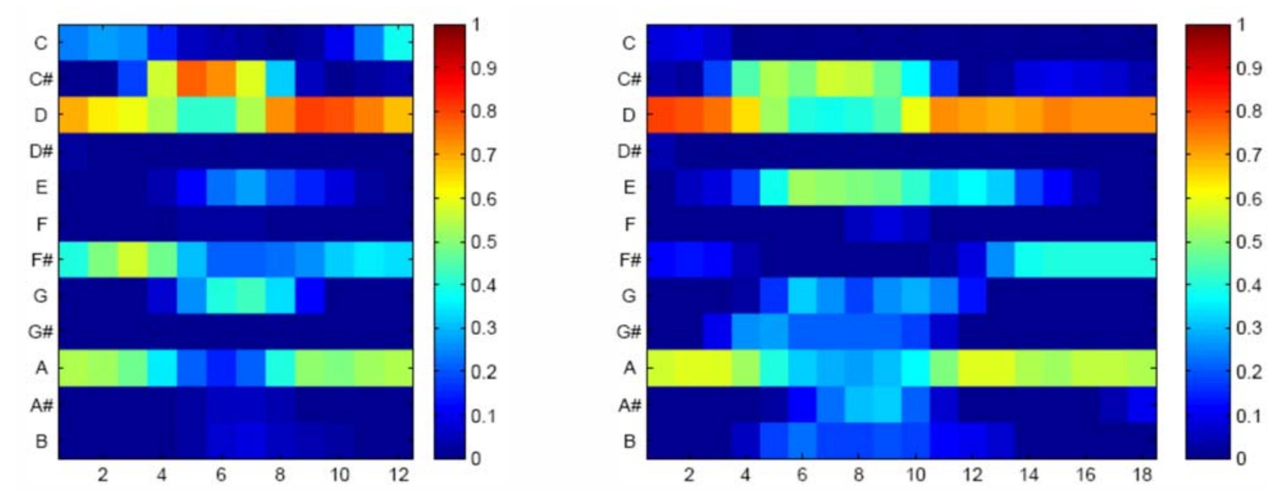


Feature resolution: 10 Hz

Chroma Features

Example: Bach Toccata

Koopman Ruebsam



Feature resolution: 1 Hz

Music Synchronization: Audio-Audio

- Koopman $\rightsquigarrow$ $V := (v^1, v^2, \ldots, v^N)$ $N = 12$

  Ruebsam $\rightsquigarrow$ $W := (w^1, w^2, \ldots, w^M)$ $M = 18$

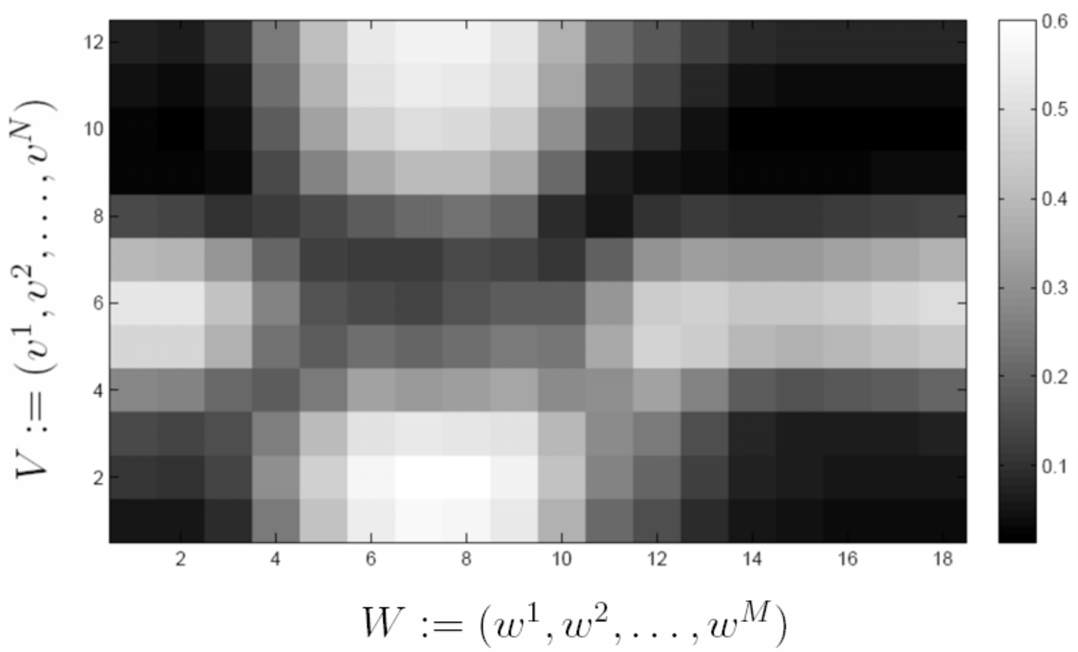
- v^n, w^m = 12-dimensional normalized chroma vectors

- Local cost measure $c : \mathbb{R}^{12} \times \mathbb{R}^{12} \to \mathbb{R}$

  $$c(v^n, w^m) := 1 - \langle v^n, w^m \rangle$$

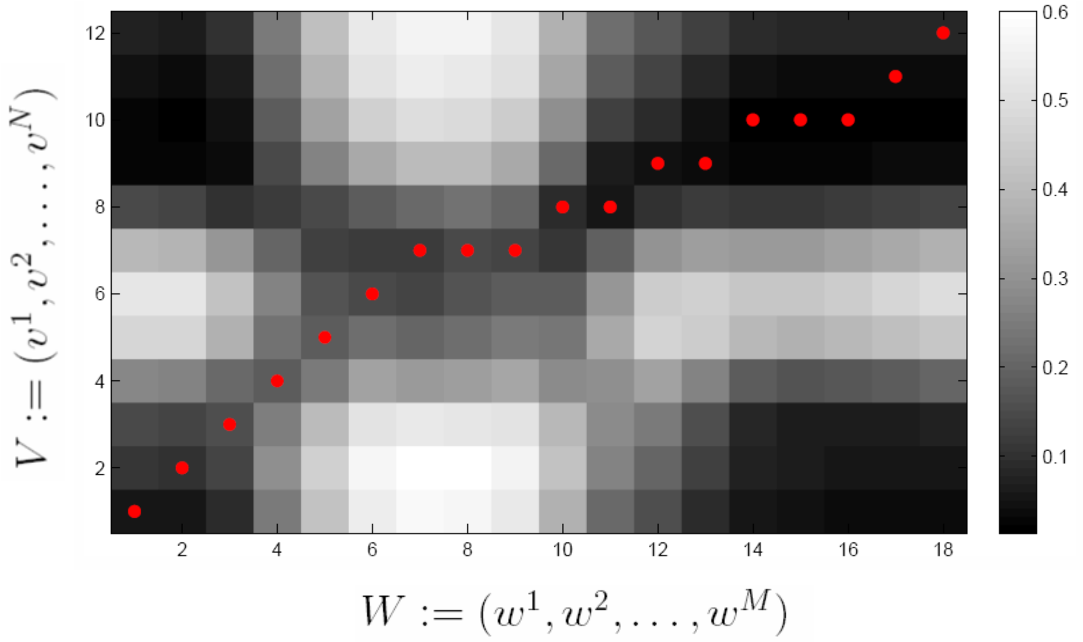
- $N \times M$ cost matrix $C(n,m) := c(v^n, w^m)$

Music Synchronization: Audio-Audio



Music Synchronization: Audio-Audio

Cost-minimizing warping path



Cost-Minimizing Warping Path

- Computation via dynamic programming

  $\rightsquigarrow$ Dynamic Time Warping (DTW)

- Memory requirements and running time: $O(NM)$
- Problem: Infeasible for large N and M
- Example: Feature resolution 10 Hz, pieces 15 min

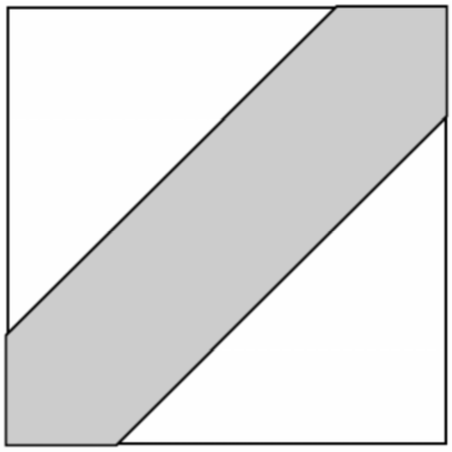
  $\Rightarrow$ $N, M \sim$ 10,000

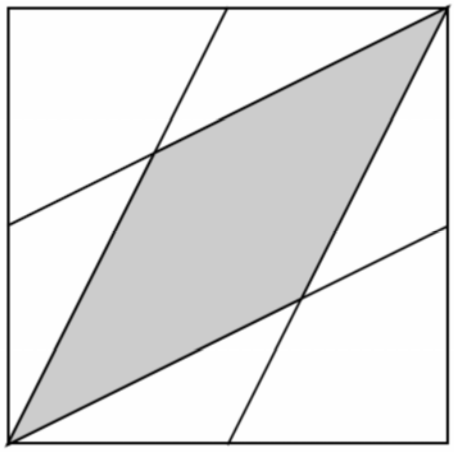
  $\Rightarrow$ $N \cdot M \sim 100{,}000{,}000$

Strategy: Global Constraints

Sakoe-Chiba band

Itakura parallelogram

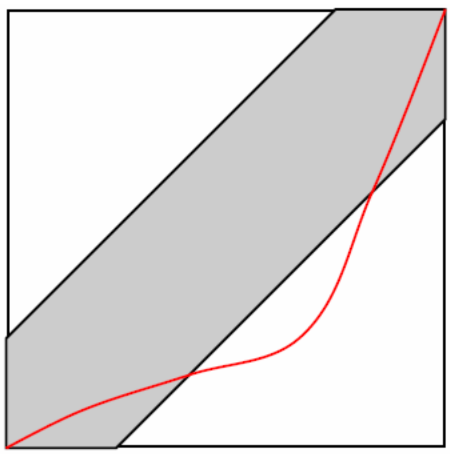


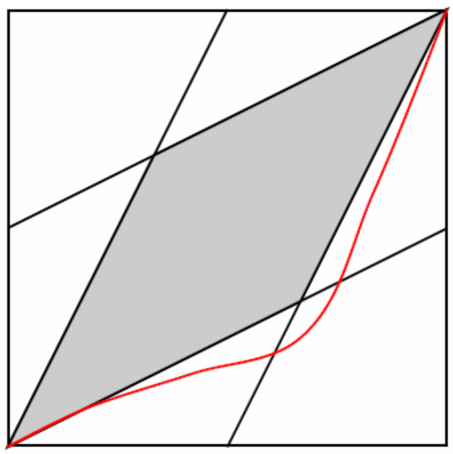


Strategy: Global Constraints

Sakoe-Chiba band

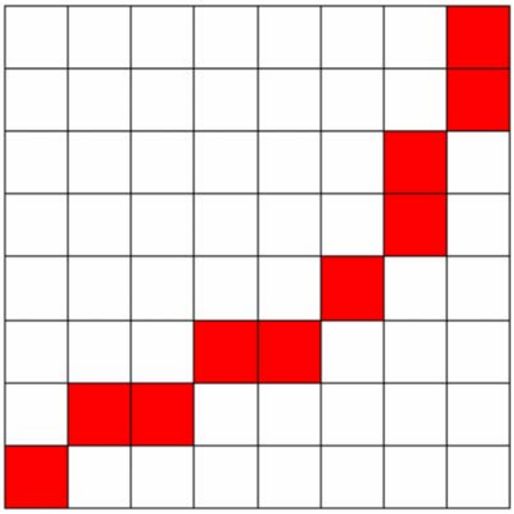
Itakura parallelogram





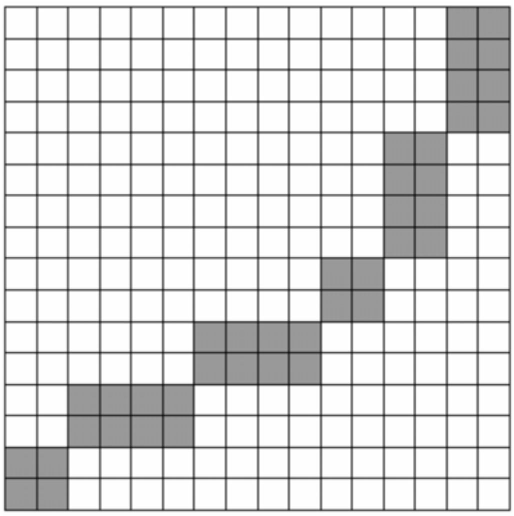
Problem: Optimal warping path not in constraint region

Strategy: Multiscale Approach



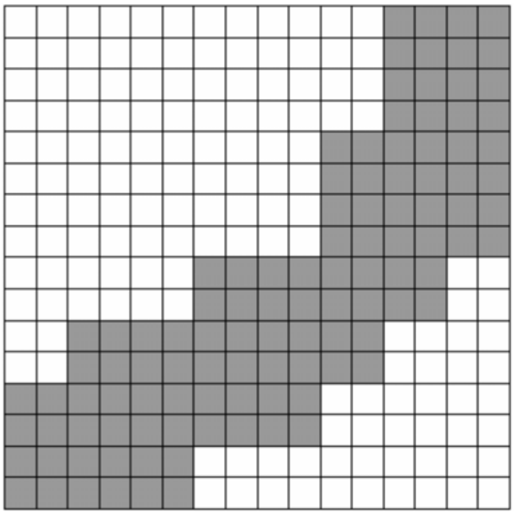
Compute optimal warping path on coarse level

Strategy: Multiscale Approach



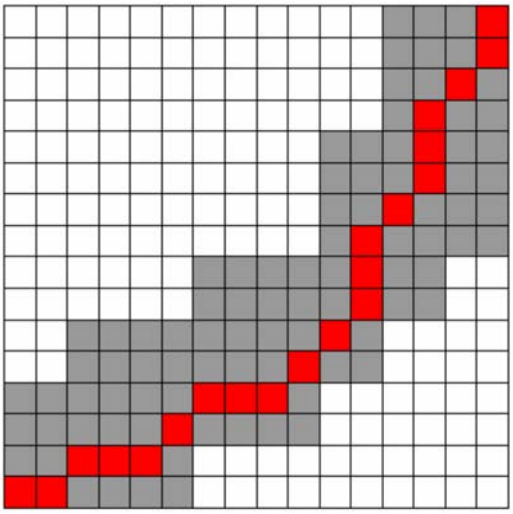
Project on fine level

Strategy: Multiscale Approach



Specify constraint region

Strategy: Multiscale Approach



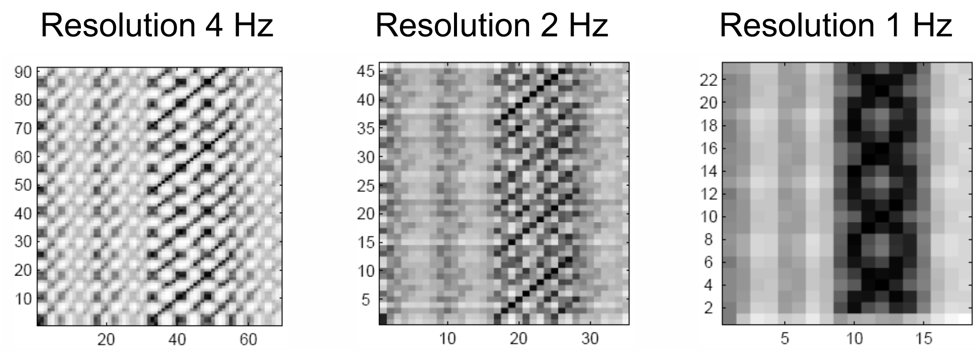
Compute *constrained* optimal warping path

Strategy: Multiscale Approach

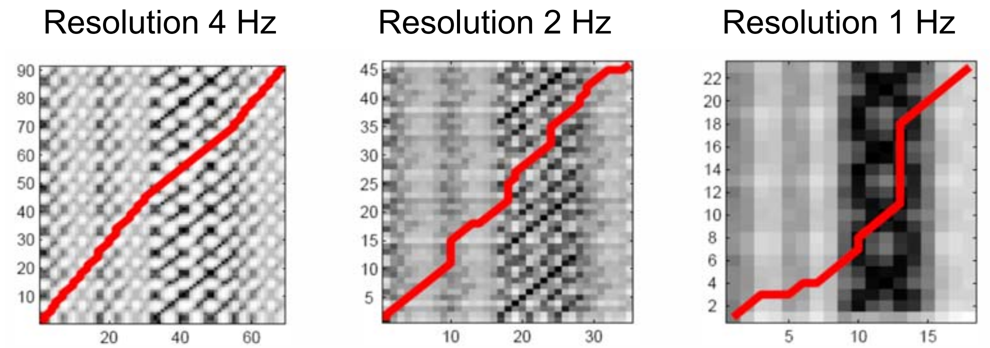
- Suitable features?
- Suitable resolution levels?
- Size of constraint regions?

Good trade-off between efficiency and robustness?

Strategy: Multiscale Approach



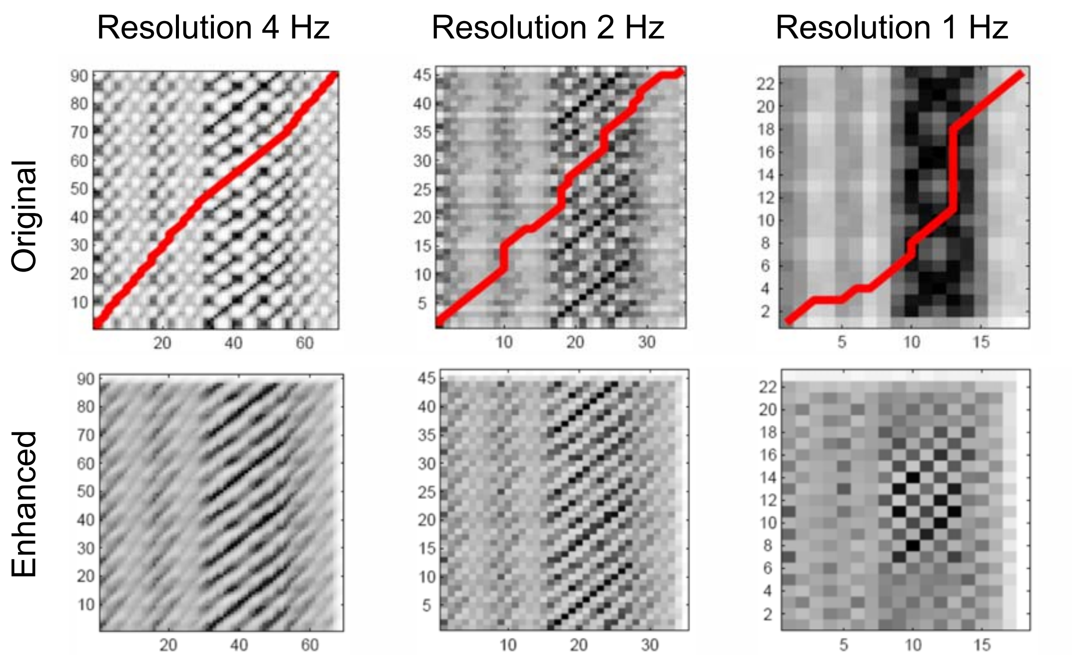
Strategy: Multiscale Approach



Problem: Cost matrix may degenerate
⇝ useless warping path

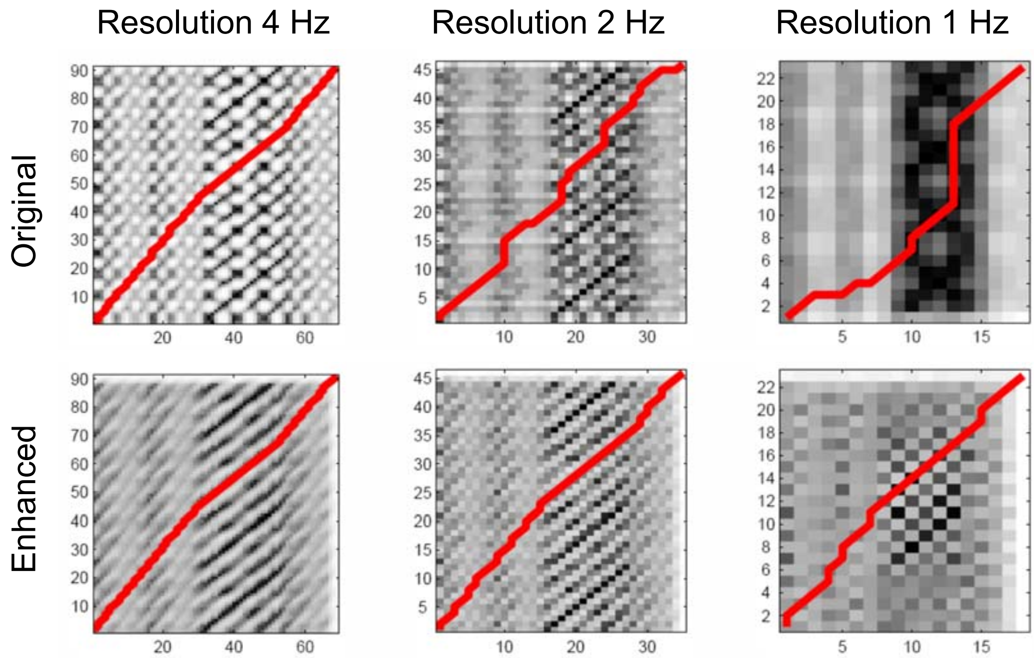
Strategy: Multiscale Approach

Improve robustness by enhancing cost matrix



Strategy: Multiscale Approach

Improve robustness by enhancing cost matrix



Strategy: Multiscale Approach

Chroma features at three levels: 0.33 Hz / 1 Hz / 10 Hz

Recording 1	length [sec]	Recording 2	length [sec]	t_{DTW} [sec]	t_{MsDTW} [sec]	[%]
Beet9Bern	1144.9	Beet9Kar	1054.8	31.18	1.08	3.46

Strategy: Multiscale Approach

Chroma features at three levels: 0.33 Hz / 1 Hz / 10 Hz

Recording 1	length [sec]	Recording 2	length [sec]	t_{DTW} [sec]	t_{MsDTW} [sec]	[%]
Beet9Bern	1144.9	Beet9Kar	1054.8	31.18	1.08	3.46

Number of matrix entries needed for DTW and MsDTW:

	DTW	MsDTW	%
Level 1	120,808,050	2,117,929	1.75
Level 2	1,209,030	17,657	1.46
Level 3	134,464	134,464	100

Music Synchronization: Audio-Audio

Conclusions

- Chroma features

  $\rightsquigarrow$ suited for harmony-based music

- Relatively coarse but good global alignments

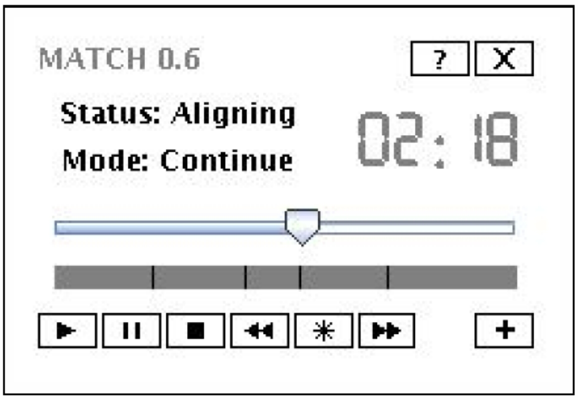
- Multiscale approach: simple, robust, fast

Music Synchronization: Audio-Audio

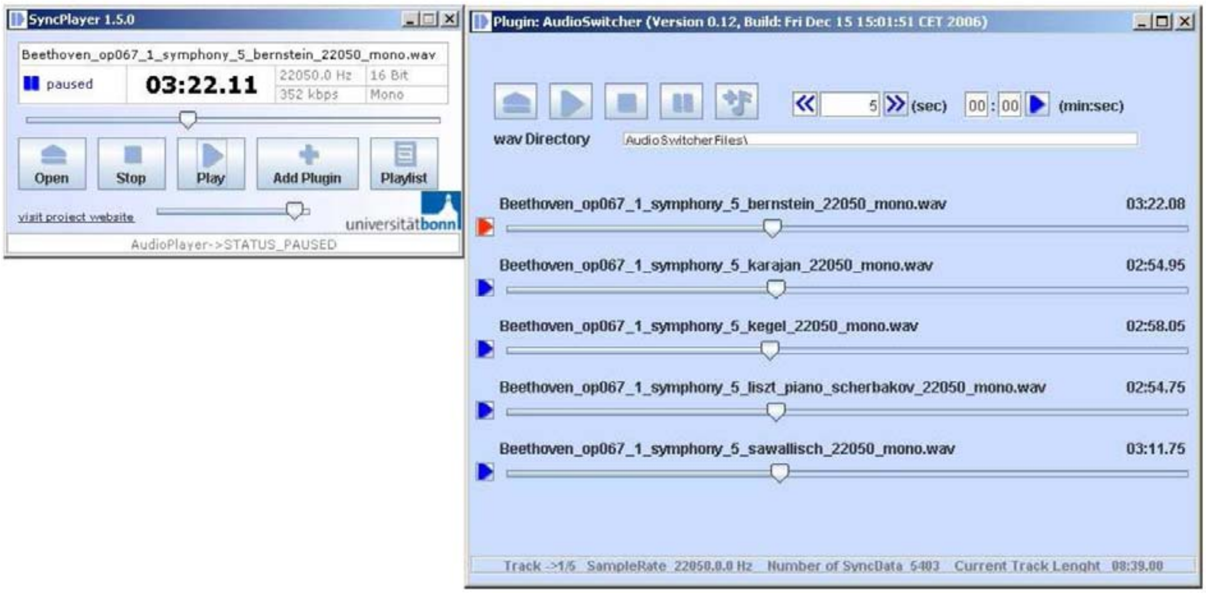
Applications

- Efficient music browsing
- Blending from one interpretation to another one
- Mixing and morphing different interpretations
- Tempo studies

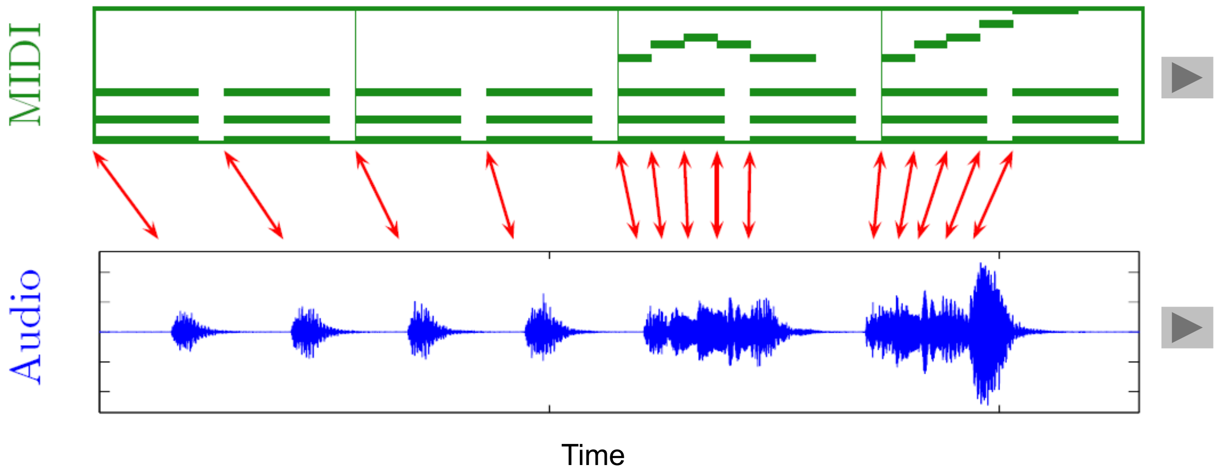
System: Match (Dixon)



System: SyncPlayer/AudioSwitcher



Music Synchronization: MIDI-Audio



Music Synchronization: MIDI-Audio

MIDI = meta data

Automated annotation

Audio recording

Sonification of annotations ▶ ▶

Music Synchronization: MIDI-Audio

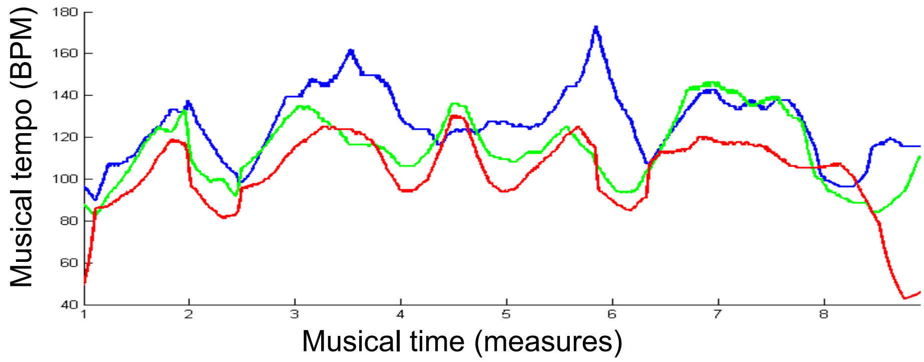
MIDI = reference (score)

Tempo information

Audio recording

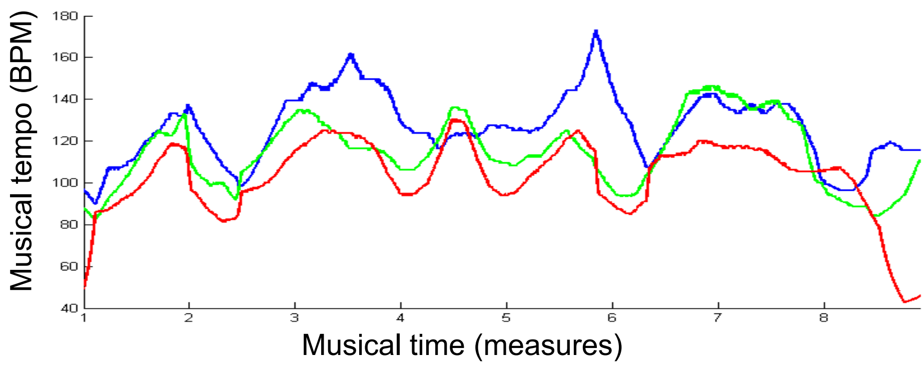
Performance Analysis: Tempo Curves

Schumann:
Träumerei



Performance Analysis: Tempo Curves

What can be done if no reference is available?

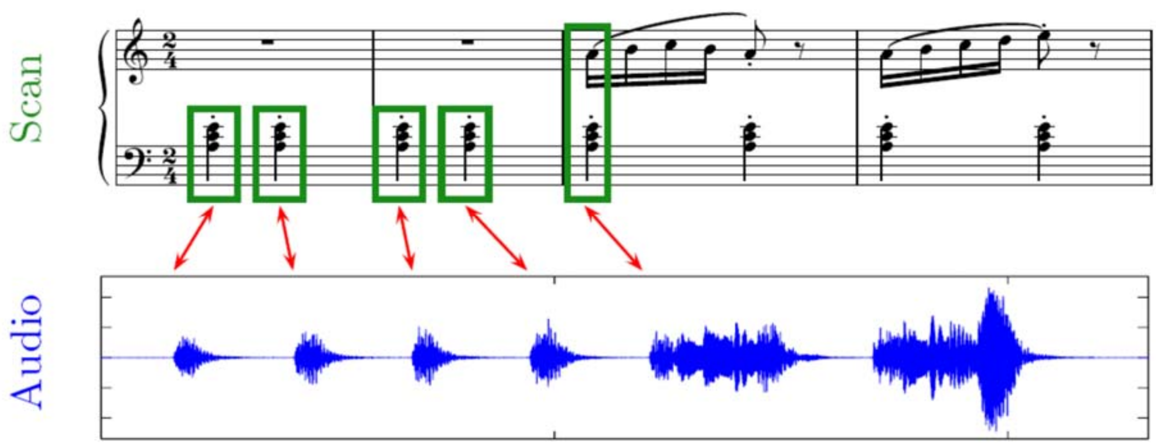


Music Synchronization: MIDI-Audio

Applications

- Automated audio annotation
- Accurate audio access after MIDI-based retrieval
- Automated tracking of MIDI note parameters during audio playback
- Performance Analysis

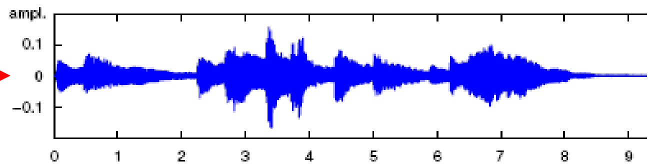
Music Synchronization: Scan-Audio



Music Synchronization: Scan-Audio

Scanned Sheet Music

Correspondence



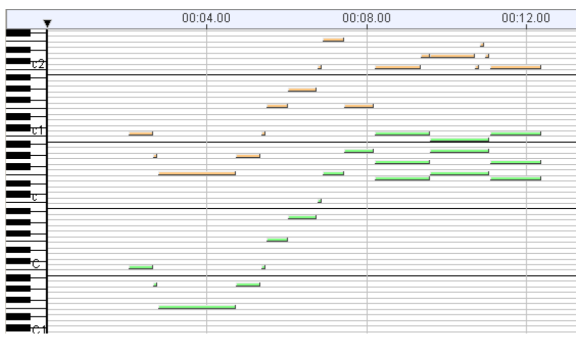
Audio Recording

Music Synchronization: Scan-Audio

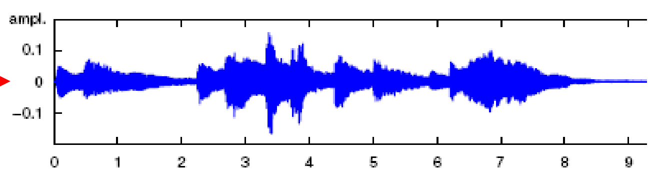
Scanned Sheet Music

Symbolic Note Events

OMR

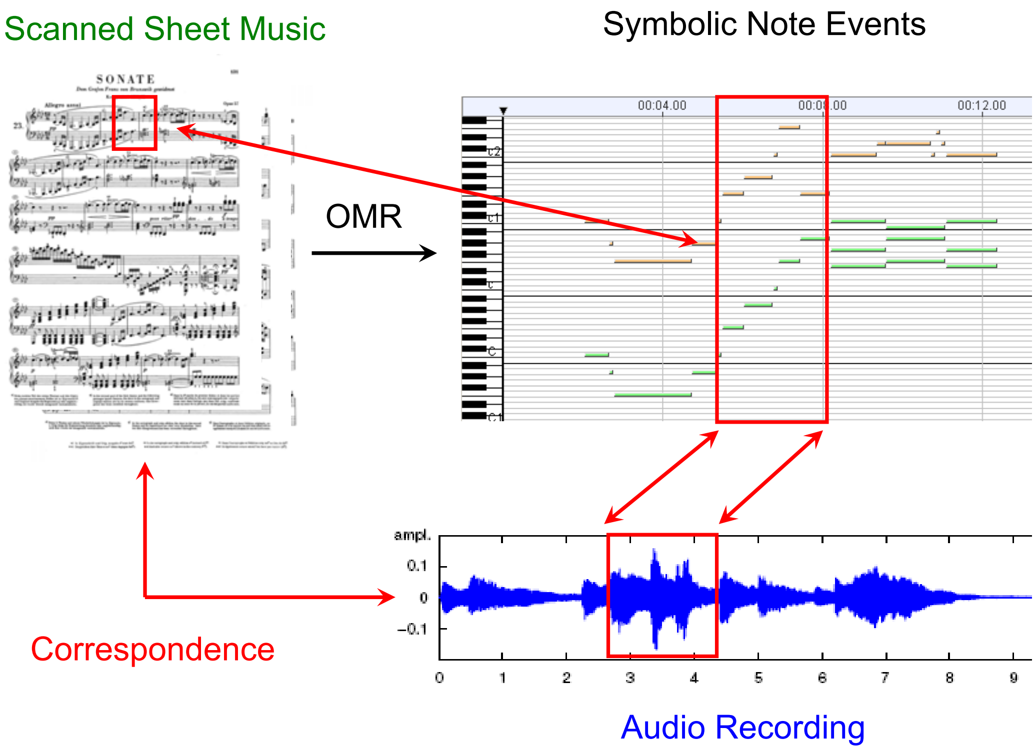


Correspondence

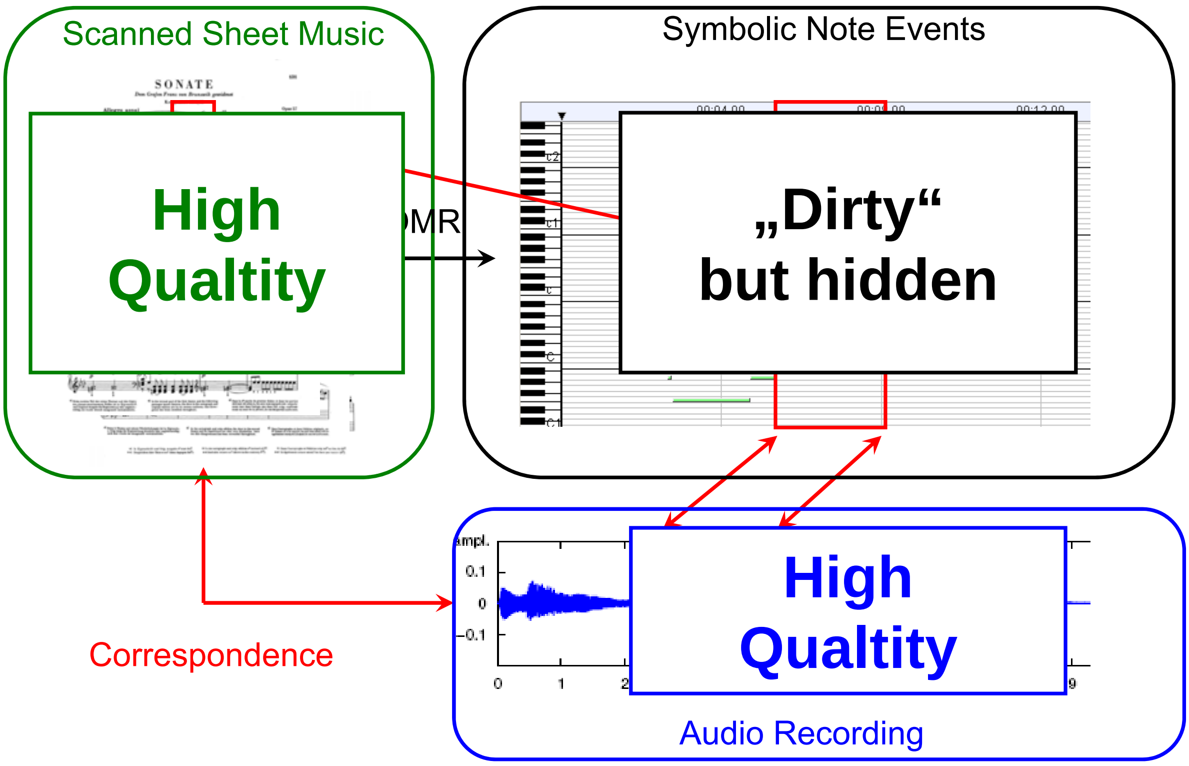


Audio Recording

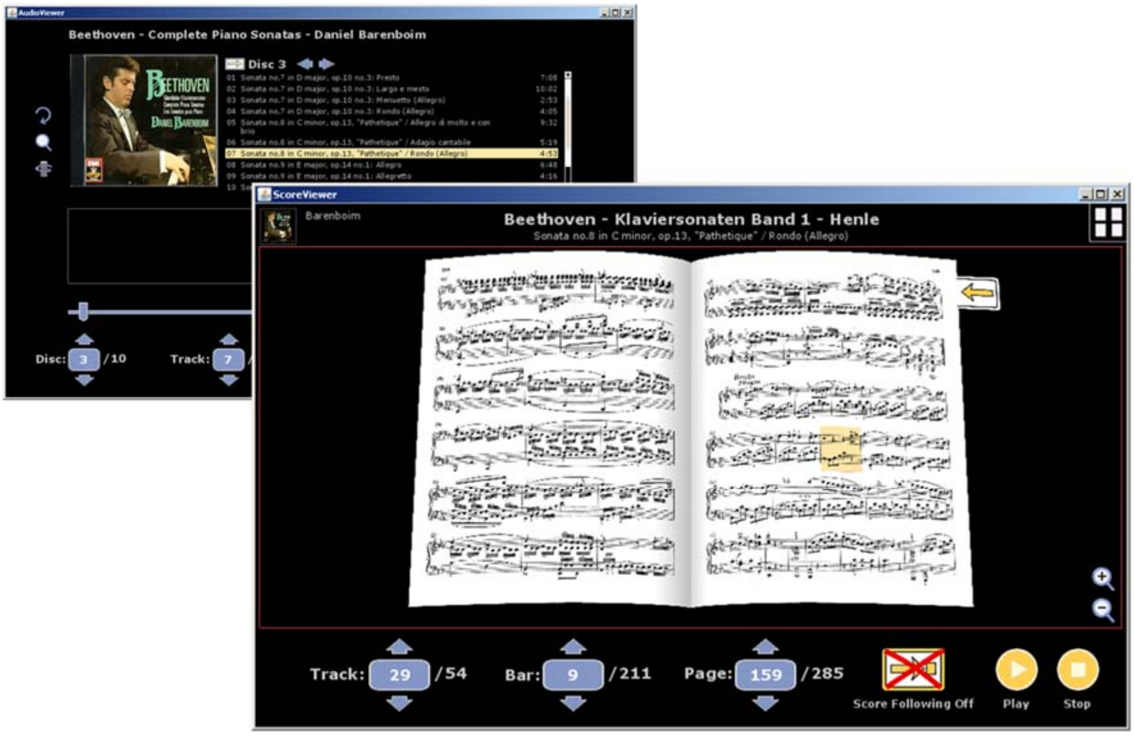
Music Synchronization: Scan-Audio



Music Synchronization: Scan-Audio

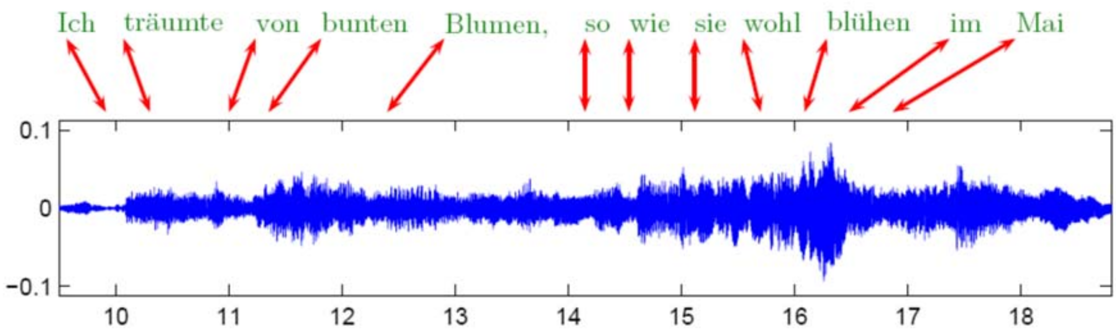


Application: Score Viewer



[ECDL 08, ICMI 08]

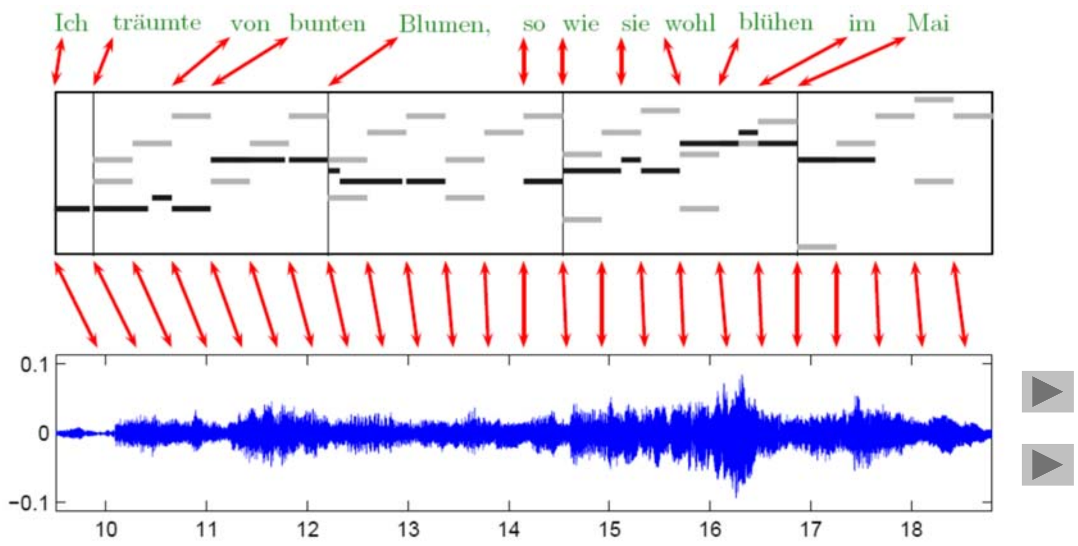
Music Synchronization: Lyrics-Audio



Difficult task!

Music Synchronization: Lyrics-Audio

Lyrics-Audio → Lyrics-MIDI + MIDI-Audio



System: SyncPlayer/LyricsSeeker

High-Resolution Music Synchronization

- Normalized chroma features
  → robust to changes in instrumentation and dynamics
  → robust synchronization of reasonable overall quality

- Drawback: low temporal alignment accuracy

- Idea: Integration of note onset information

High-Resolution Music Synchronization

- Normalized chroma features
  → robust to changes in instrumentation and dynamics
  → robust synchronization of reasonable overall quality

- Drawback: low temporal alignment accuracy

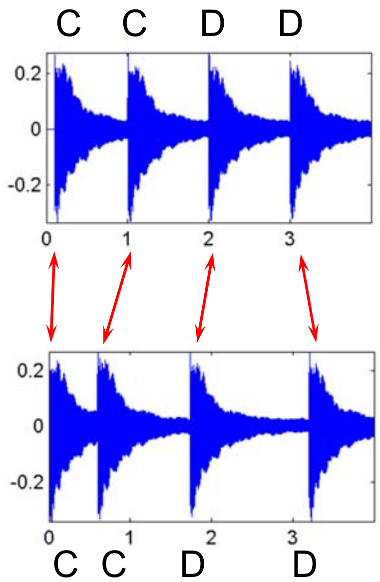
- Idea: Integration of note onset information

- Example: MIDI-Audio synchronization

  Chroma-Chroma:
  Chroma-Chroma + onset information:

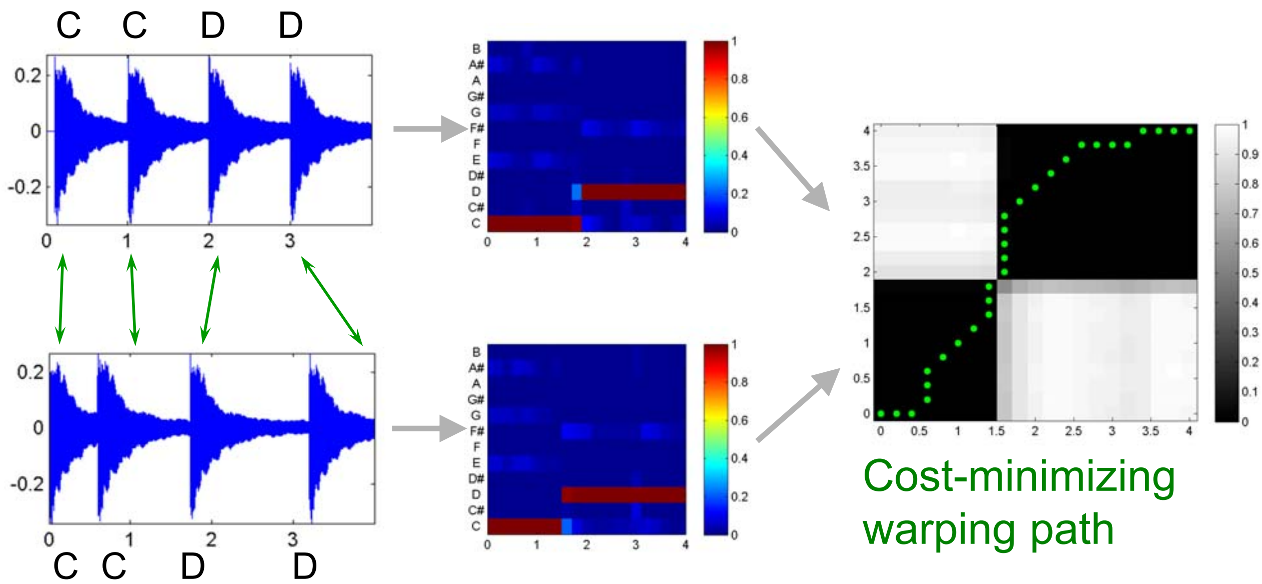
High-Resolution Music Synchronization

Example: C – C – D – D



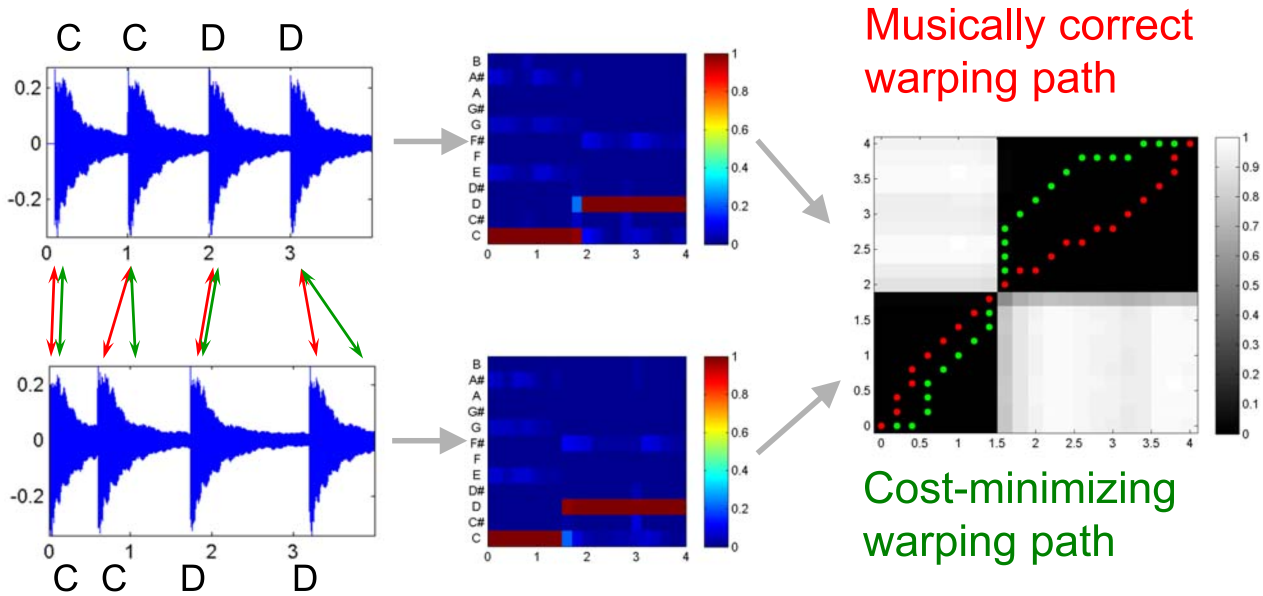
High-Resolution Music Synchronization

Example: C – C – D – D



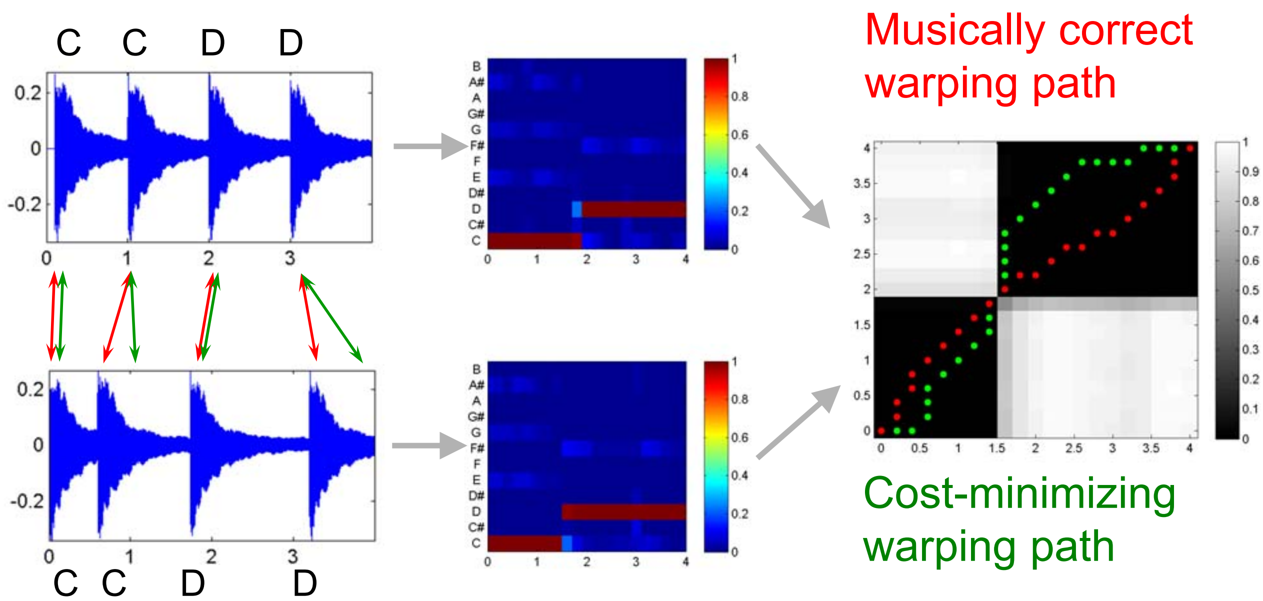
High-Resolution Music Synchronization

Example: C – C – D – D



High-Resolution Music Synchronization

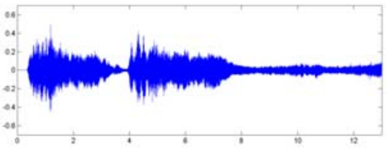
Example: C – C – D – D

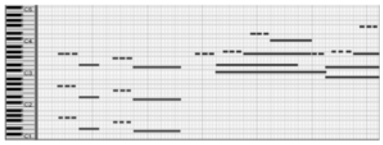


Problem: note onsets are not captured in feature representation

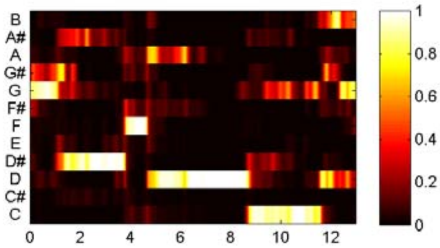
High-Resolution Music Synchronization

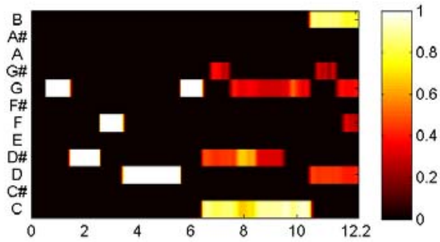
Example: Beethoven's Fifth





Chroma representations



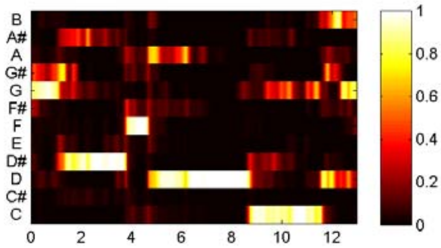


Problem: note onsets are not captured in feature representation

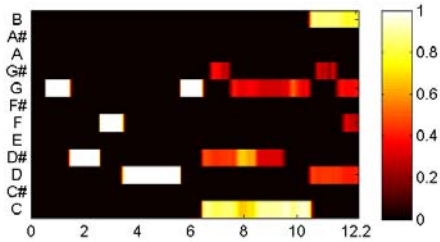
High-Resolution Music Synchronization

Example: Beethoven's Fifth

Audio

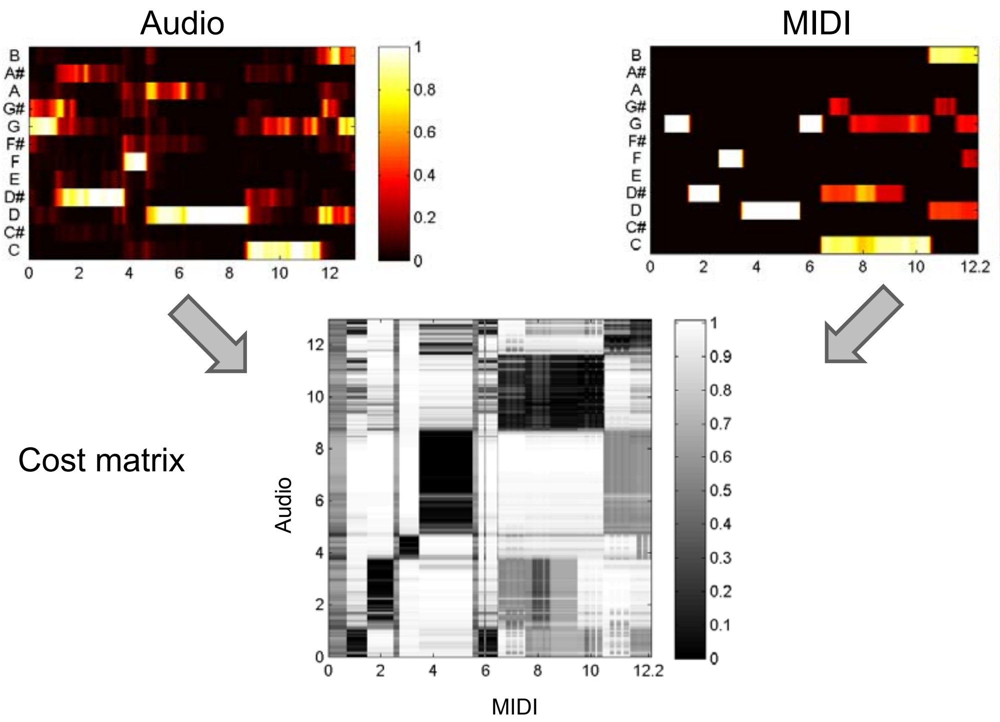


MIDI



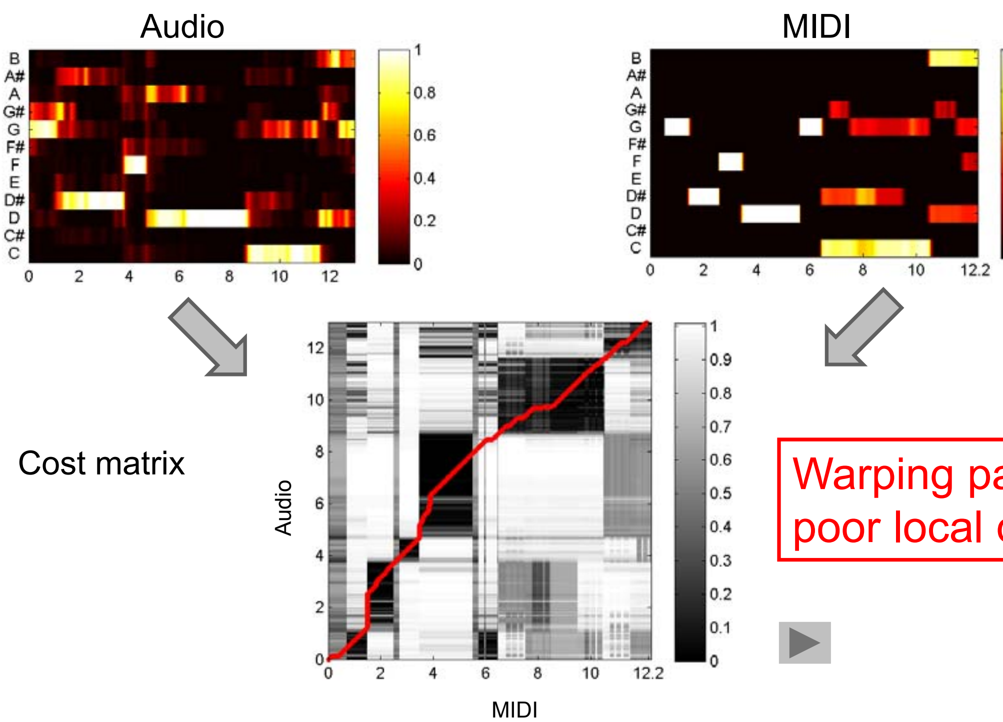
High-Resolution Music Synchronization

Example: Beethoven's Fifth



High-Resolution Music Synchronization

Example: Beethoven's Fifth



Warping path of poor local quality

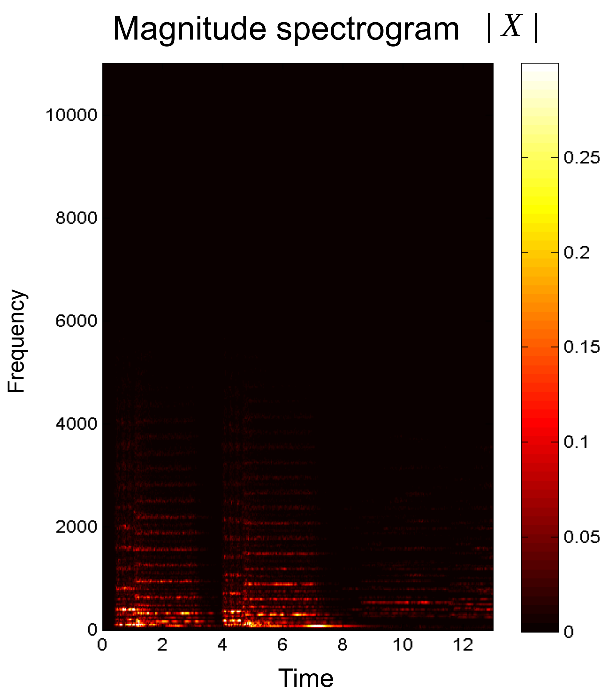
Onset Detection

- General goal: Detection of onsets of musical notes

- Typical signal properties at note onset positions:
  - increase in energy
  - change of pitch
  - change of spectral content
  - high frequency content

- Idea: locate note onset candidates by measuring changes in spectral content

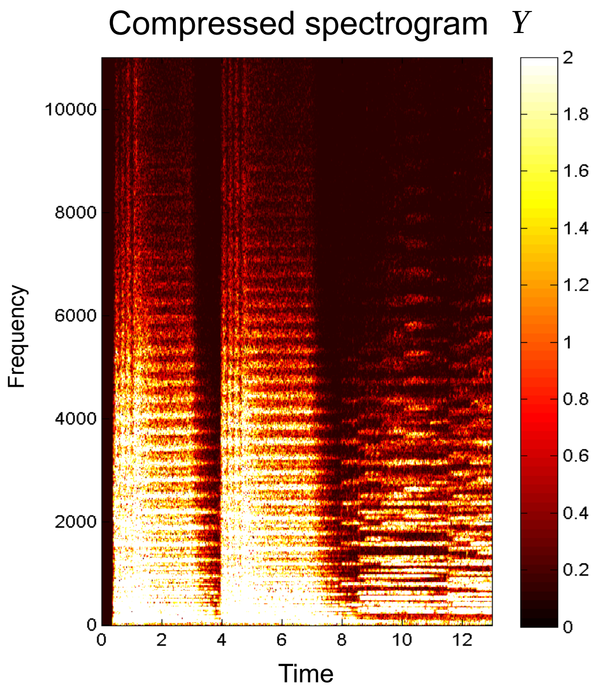
Onset Detection



Steps:

1. Spectrogram

Onset Detection



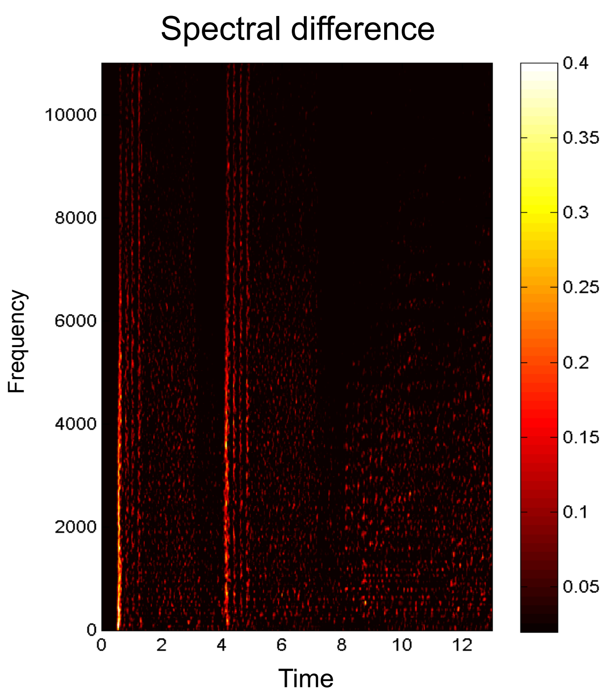
Steps:

1. Spectrogram
2. Logarithmic compression

$$Y = \log(1 + C \cdot |X|)$$

- *human sensation*
- *enhances low intensity values*
- *high frequency content*
- *reduces influence of amplitude modulation*

Onset Detection

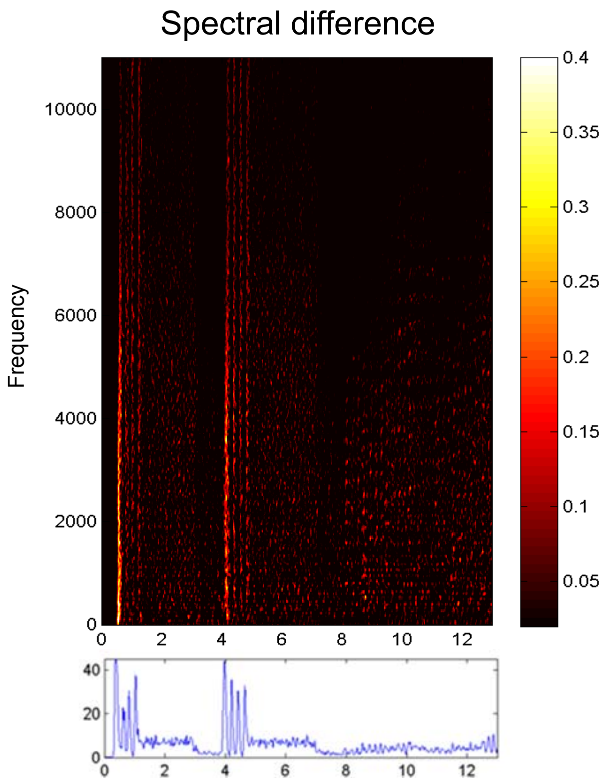


Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation

- *energy increase to be captured*
- *only positive values considered*

Onset Detection



Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation

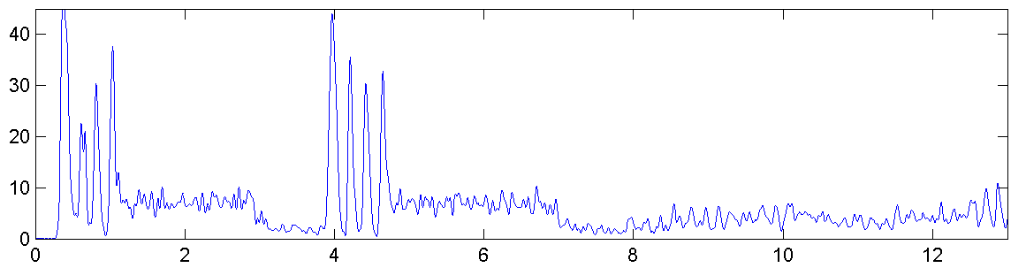
Novelty Curve

Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation

Novelty Curve



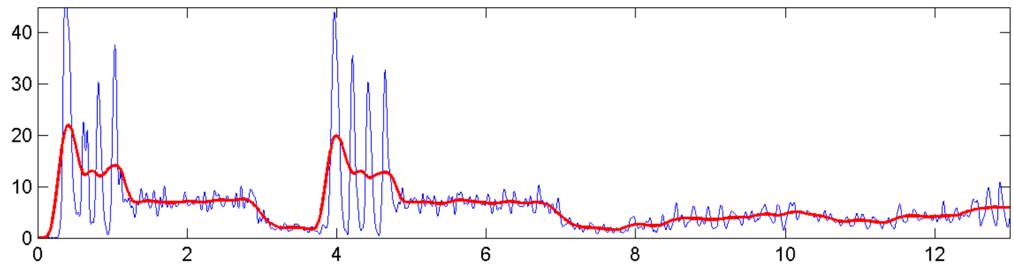
Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization

Novelty Curve

Substraction of local average

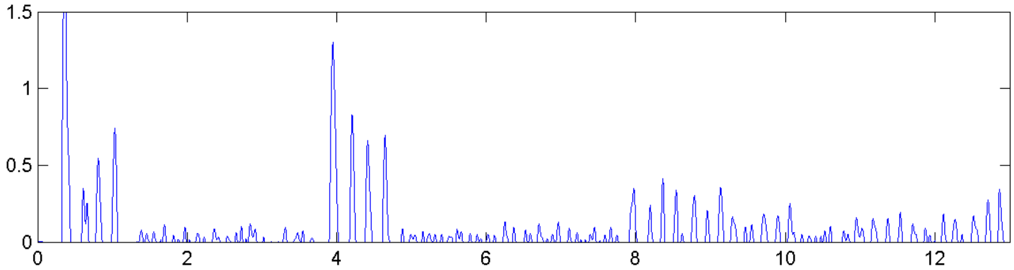


Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization

Normalized novelty curve

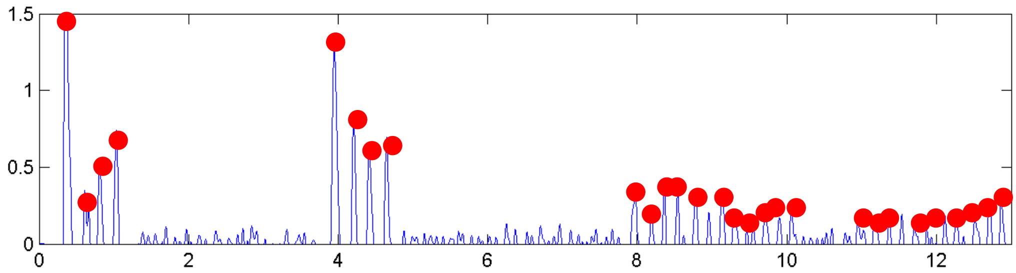


Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization
6. Peak picking

Normalized novelty curve

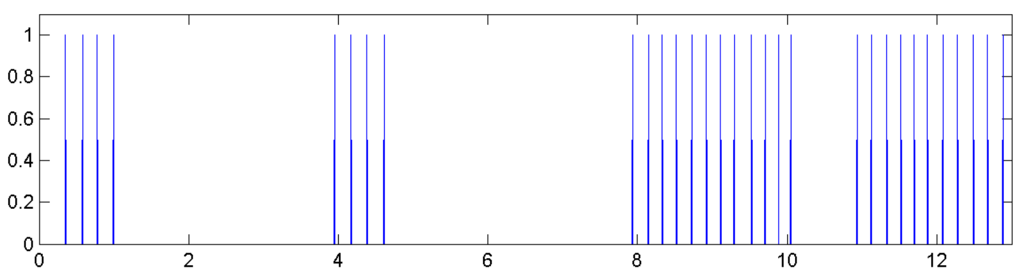


Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization
6. Peak picking

Impulses

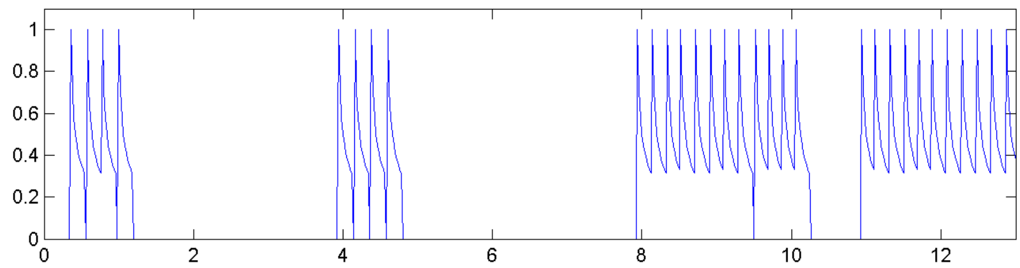


Onset Detection

Steps:

1. Spectrogram
2. Logarithmic compression
3. Differentiation
4. Accumulation
5. Normalization
6. Peak picking
7. Decay Filter

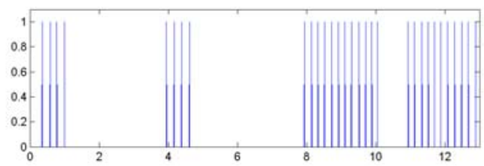
Decaying impulses



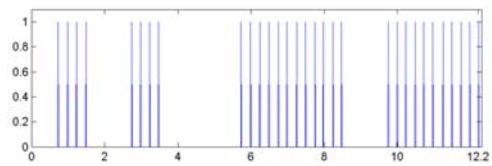
High-Resolution Music Synchronization

Cost matrix based on impulses

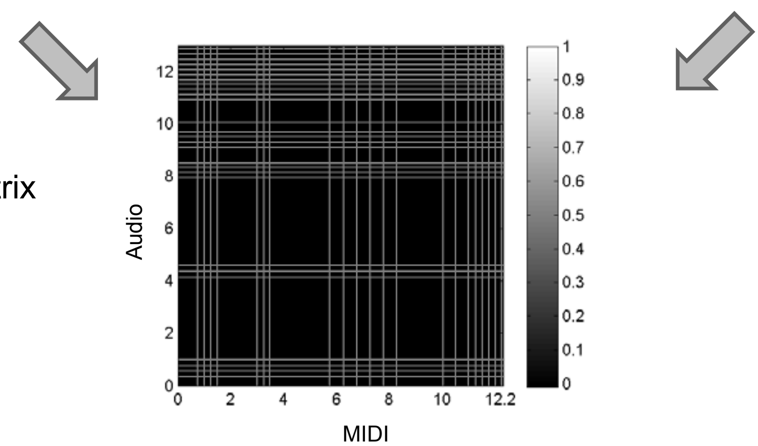
Audio



MIDI



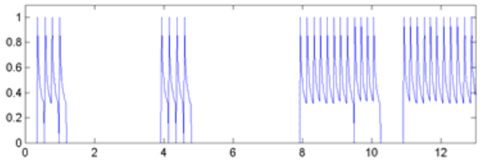
Cost matrix



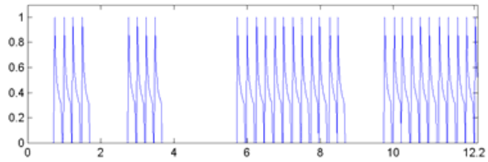
High-Resolution Music Synchronization

Cost matrix based on decaying impulses

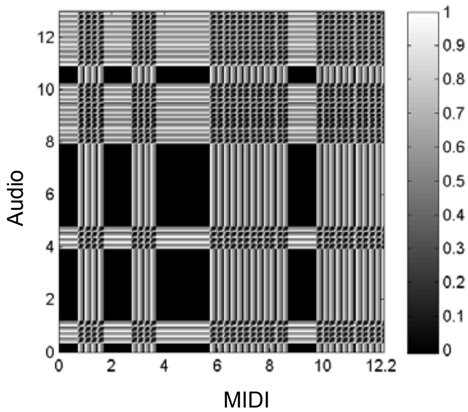
Audio



MIDI



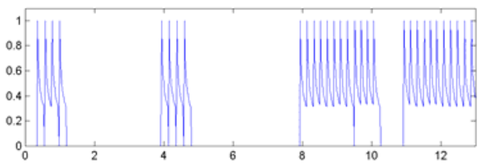
Cost matrix



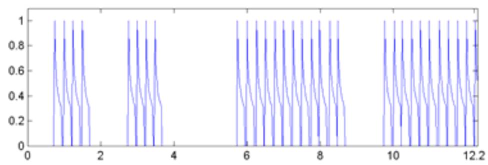
High-Resolution Music Synchronization

Cost matrix based on decaying impulses

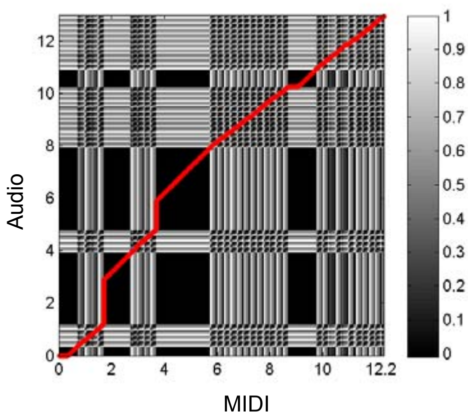
Audio



MIDI



Cost matrix



Warping path based on onset information

High-Resolution Music Synchronization

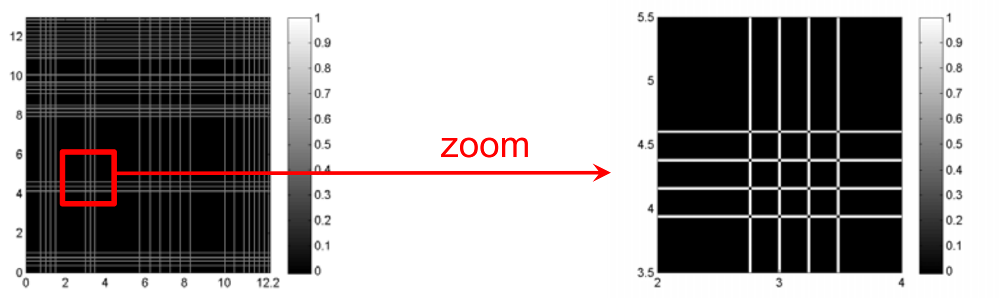
Ideas:

- Build up cost matrix with corridors of low cost
- Decaying strategy enforce corridor structure
- Each corridor corresponds to MIDI-audio pair of note onset candidates
- Warping path tends to run through corridors of low cost

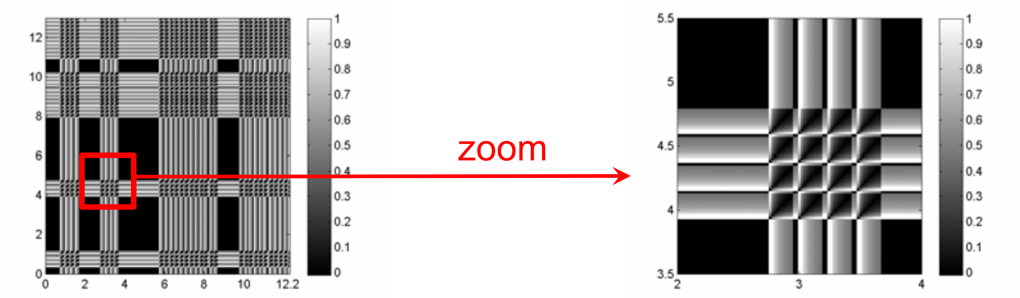
  → note onset positions are likely to be aligned

High-Resolution Music Synchronization

Impulses

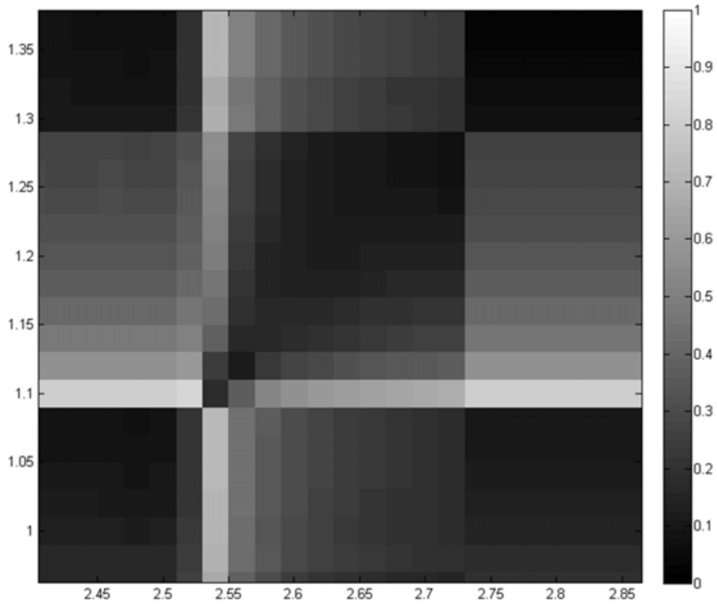


Decaying impulses



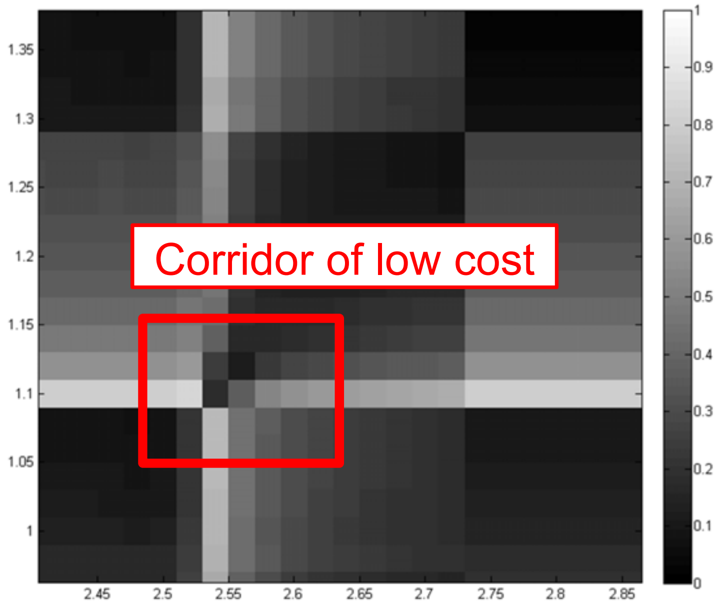
High-Resolution Music Synchronization

Cost matrix for decaying impulses



High-Resolution Music Synchronization

Cost matrix for decaying impulses



High-Resolution Music Synchronization

Combination of two different types of cost matrices:

- Cost matrix obtained from chroma features controls the global course of warping path

  → robust synchronization

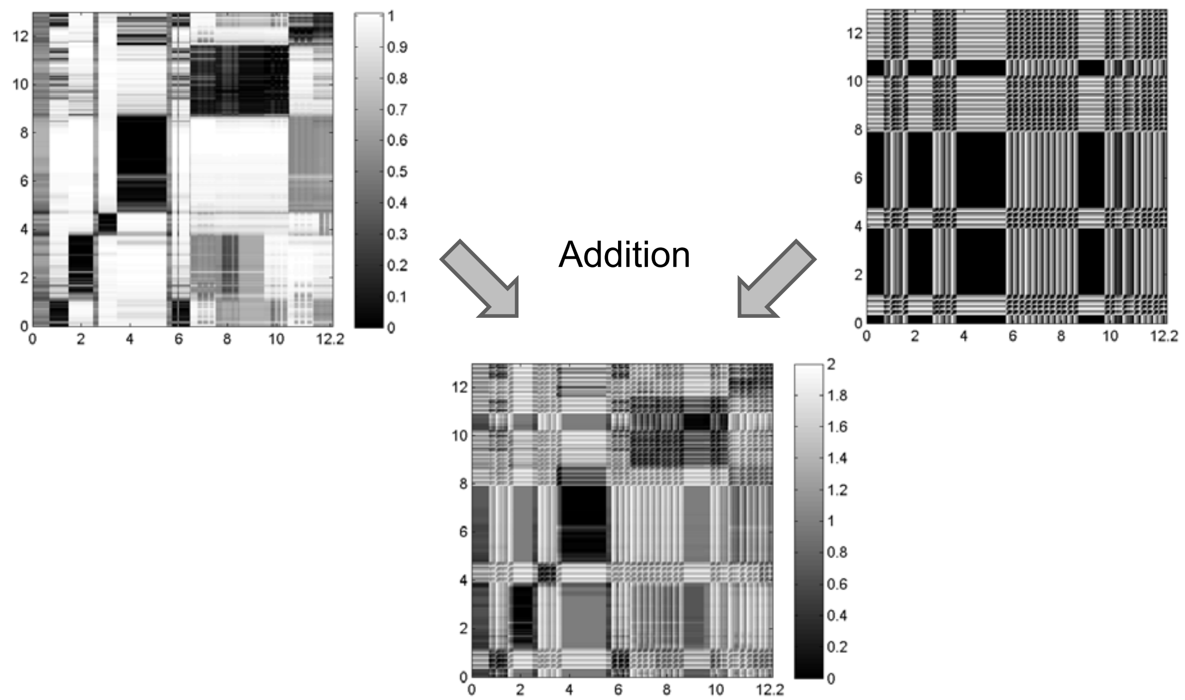
- Cost matrix obtained from onset information controls the local course of warping path

  → accurate alignment

High-Resolution Music Synchronization

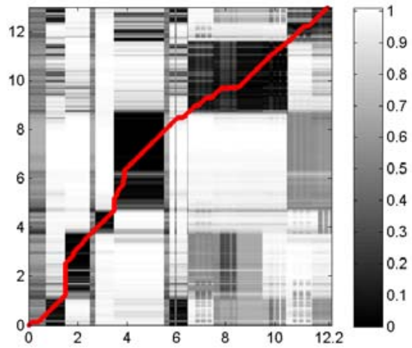
Chroma cost matrix

Onset cost matrix

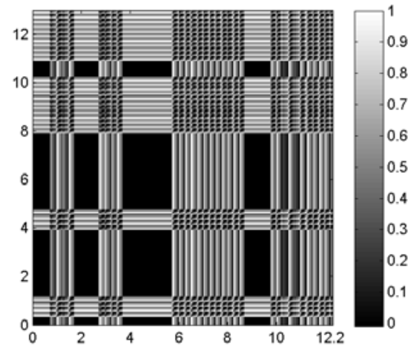


High-Resolution Music Synchronization

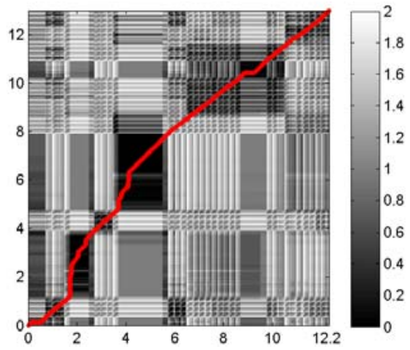
Chroma cost matrix



Onset cost matrix



Addition



Conclusions: Music Synchronization

Various requirements

- Efficiency
- Robustness
- Accuracy
- Variablity of music

Conclusions: Music Synchronization

Combination of various strategies

- Feature level
- Local cost measure level
- Global alignment level
- Evidence pooling using competing strategies

Conclusions: Music Synchronization

Offline vs. Online

- Online version: Dixon/Widmer (ISMIR 2005)
- Hidden Markov Models: Raphael (ISMIR 2004)
- Score-following
- Automatic accompaniment

Conclusions: Music Synchronization

Presence of variations

- Instrumentation
- Musical structure
- Polyphony
- Musical key
- …