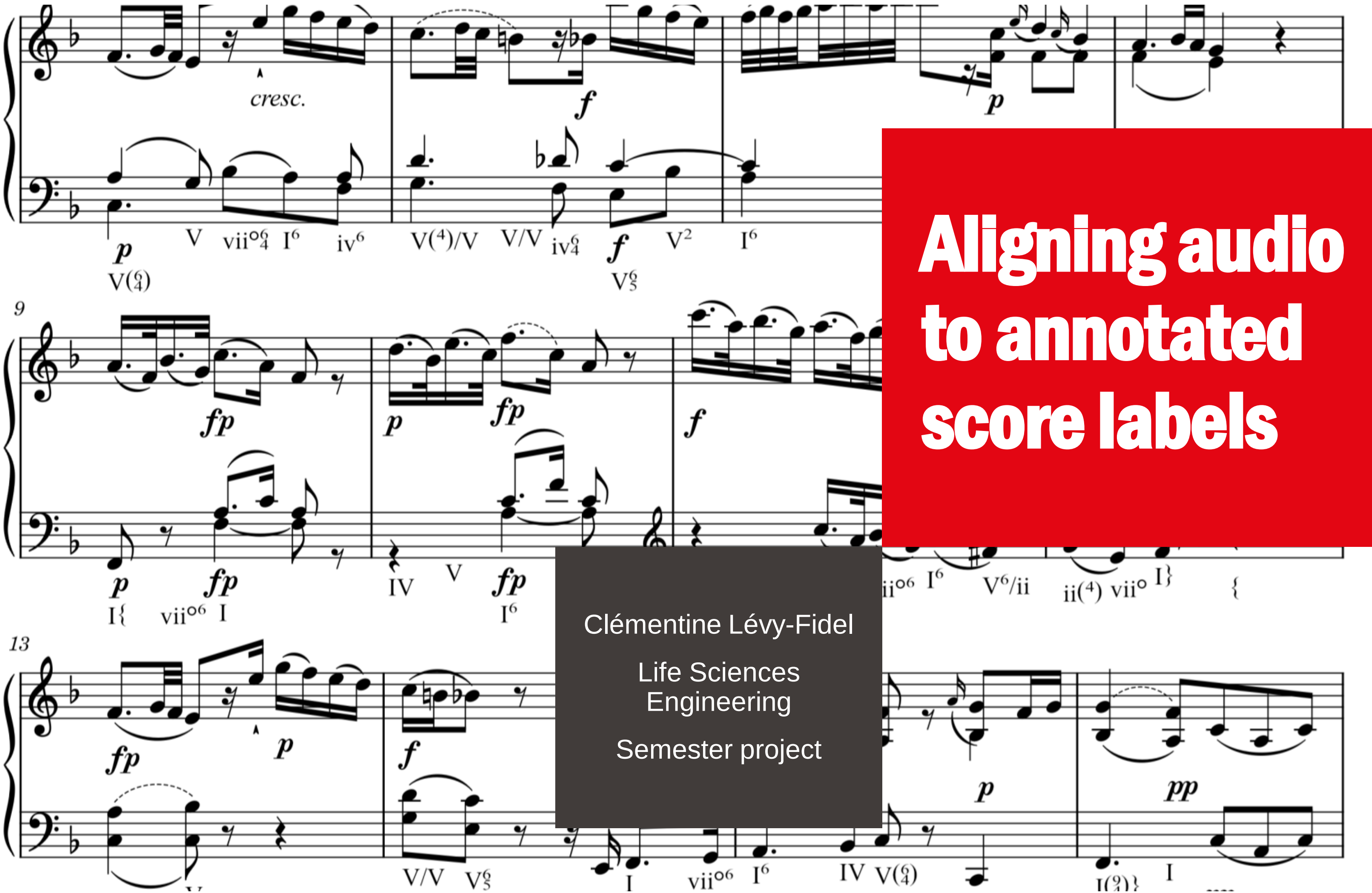
EPFL



Aligning audio to annotated score labels

Clémentine Lévy-Fidel

Life Sciences Engineering

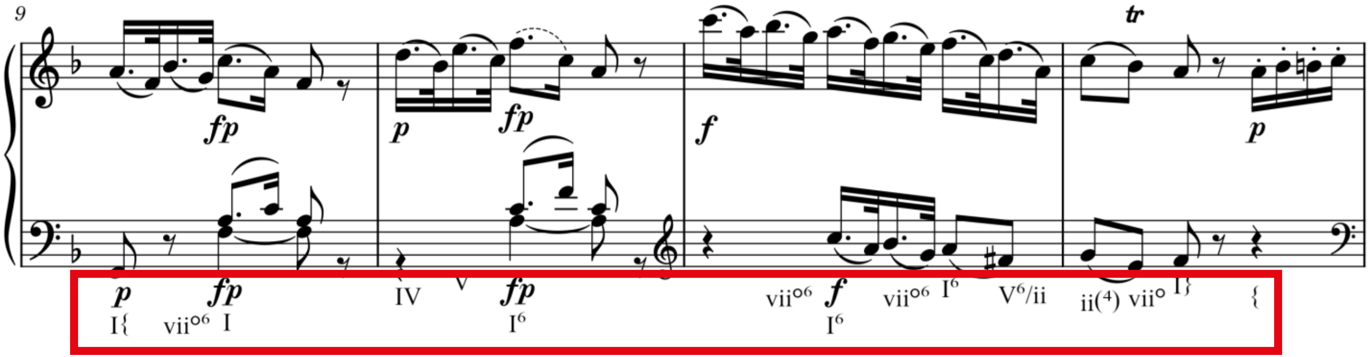
Semester project

Agenda

- Motivation
- Applications
- Related work
- Approaches
  - Beat tracking and meter reconstruction
  - Chord detection
  - Score following
  - Dynamic Time Warping
- The Aligner tool
- Next steps
- Project highlights
- Acknowledgments

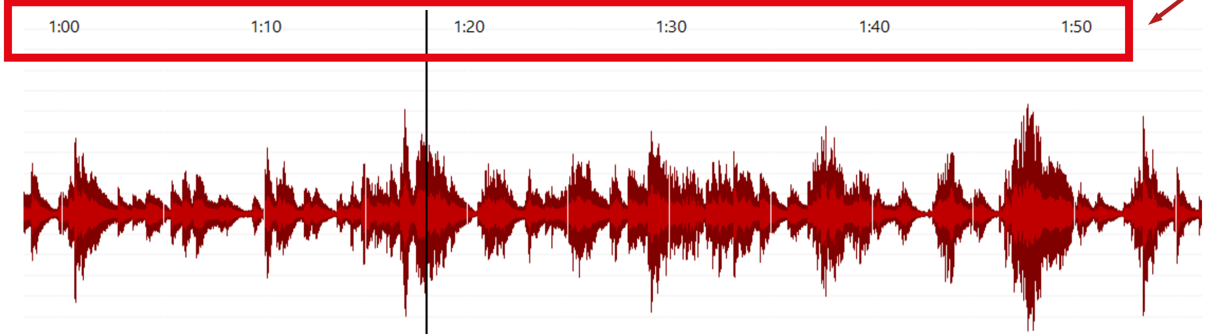
Motivation – score labels to audio alignment

- Score labels



- Audio recording

Temporal axis



Motivation – score labels to audio alignment

Clémentine Lévy-Fidel

- ***Goal***: automate alignment
- ***Desired outcome***: temporal positions of labels

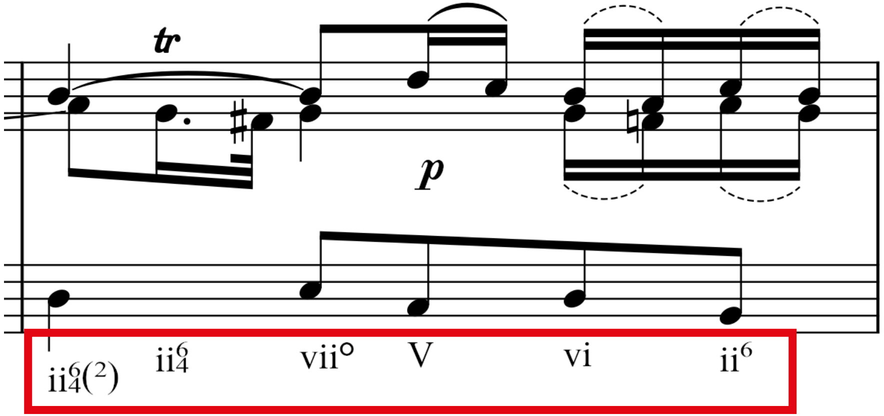
Timestamp	***Label***
0.90	F.I{
1.48	viio6
2.22	I
4.96	IV
5.66	V
…	…

Motivation – score labels to audio alignment

Clémentine Lévy-Fidel

Challenges:

- Extensive count of labels per piece:
  - Annotated Mozart Sonatas corpus: 104 to 756 labels
  - Time-consuming manual alignment

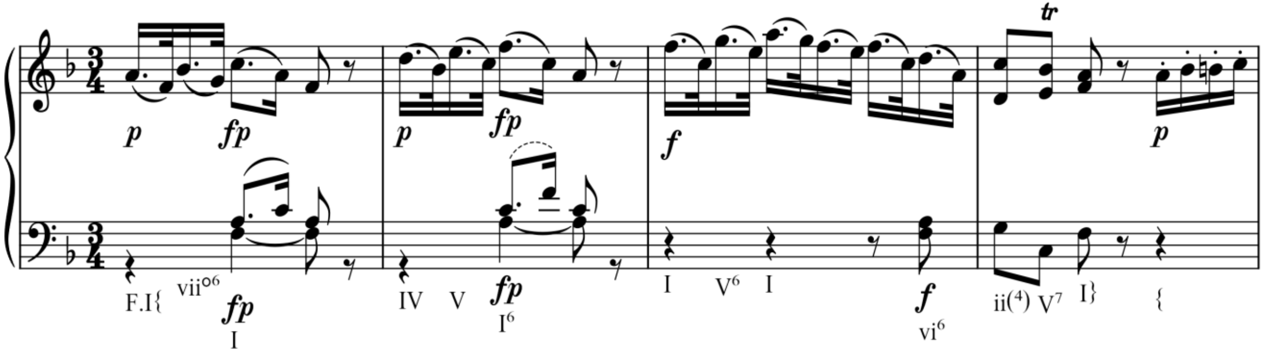


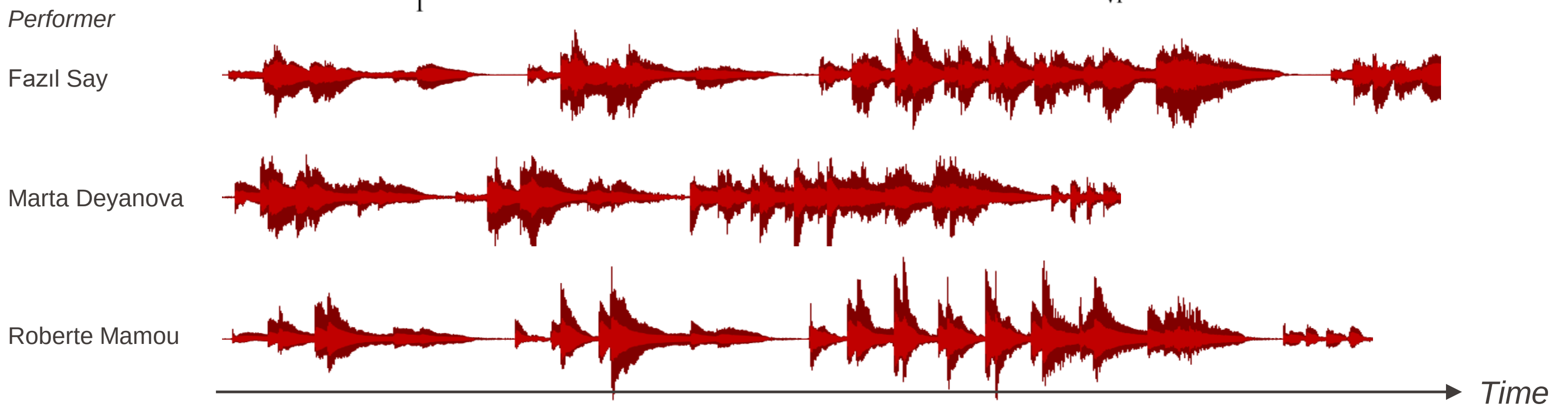
Motivation – score labels to audio alignment

Clémentine Lévy-Fidel

Challenges:

- Varying annotation or audio profiles





Applications

- Practical usability:
  - Transfer annotations from one recording to another
  - Switch between music formats: audio, score, MIDI…
- Behavioural experiments:
  - Align with EEG measurements
  - Align with assessments of perceived musical tension
- Support music theory:
  - Understand experts' annotations live
  - Live visualisation in context of presentations, teaching…
- Extend the use of the *Annotated Mozart Sonatas* corpus
  - 18 sonatas, 56 scores
- Apply to other databases and research, e.g. *Montreux Jazz Festival*

Related work

- **Meter inference –** via stochastic models

  Beat and downbeat tracking systems:

  - Bar pointer model (Whiteley et. al, 2006)

- **Music tracking –** via stochastic models

  Score following for live performance accompaniment:

  - *AnteScofo* (IRCAM, 2009)
  - *MusicPlusOne* (Christopher Raphael, 2001)
  - Deep-learning based approaches on music sheet images

- **Music synchronization** – via Dynamic Time Warping

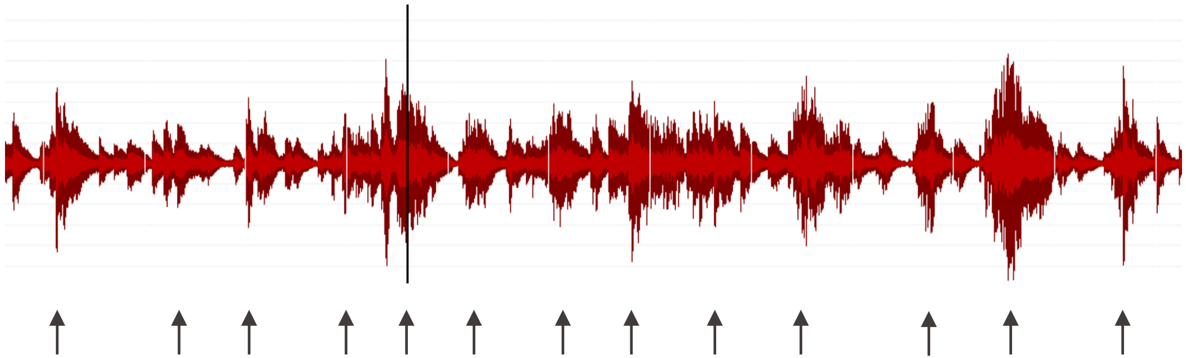
  International Audio Laboratories Erlangen:

  - Fundamentals of Music Processing (FMP) (Müller et. al, 2015)
  - *LibFMP* (Müller et. al, 2020)
  - *SyncToolBox* (Müller et. al, 2021)

Approaches – 1. Beat tracking

Clémentine Lévy-Fidel

Detection by beat **[1]** ***and downbeat*** **[2]** ***tracking algorithms from*** **madmom's** ***models:***

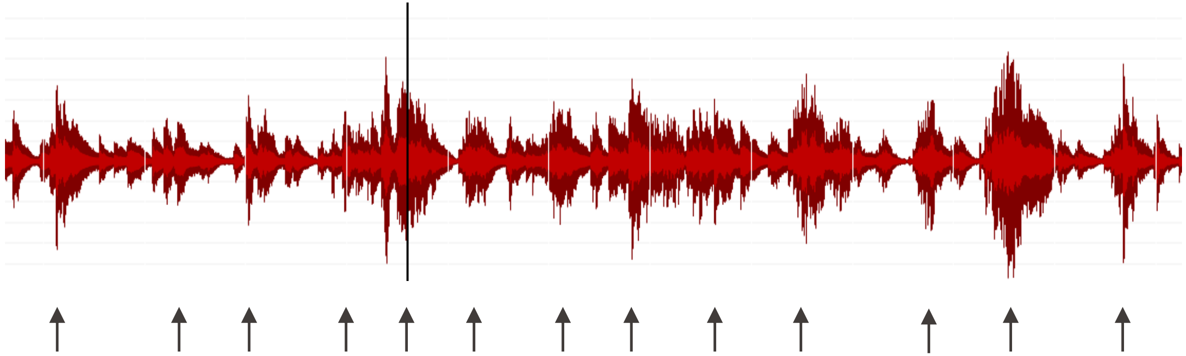


Detected **beat** sequence	1	2	3	1	2	3	1	2	3	1	2	3	1	…
Corresp. **bar** number	1			2			3			4			5	…

[1] Krebs et. al, 2015
[2] Böck et. al, 2016

Approaches – 1. Beat tracking

Alignment:



Detected beat sequence	1	2	3	1	2	3	1	2	3	1	2	3	1	…
Corresp. bar number	1			2			3			4			5	…

+

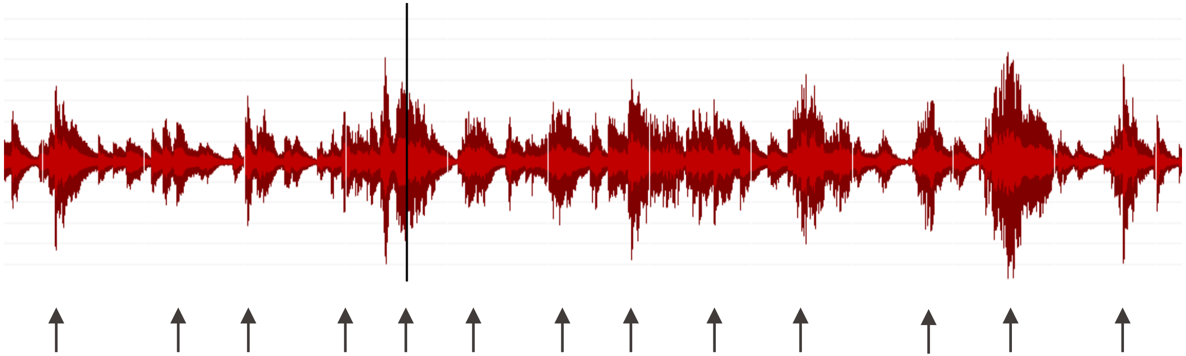
Beat	1	1	2	1	2	…
Bar	1	2	2	3	4	…
Label	C	Dm7	G7	C	Am	…

Clémentine Lévy-Fidel

Approaches – 1. Beat tracking

Pitfalls:

- Misalignment



Detected beat sequence	1	2	3	1	2	3	1	2	3	1	2	3	1	…
Corresp. bar number	1			2			3			4			5	…
Real beat sequence	1	2	3	1		2	3	1	2	3	1	2	3	…
Real bar number	1			2				3			4			…

Misalignment

Approaches – 1. Beat tracking

Clémentine Lévy-Fidel

Pitfalls:

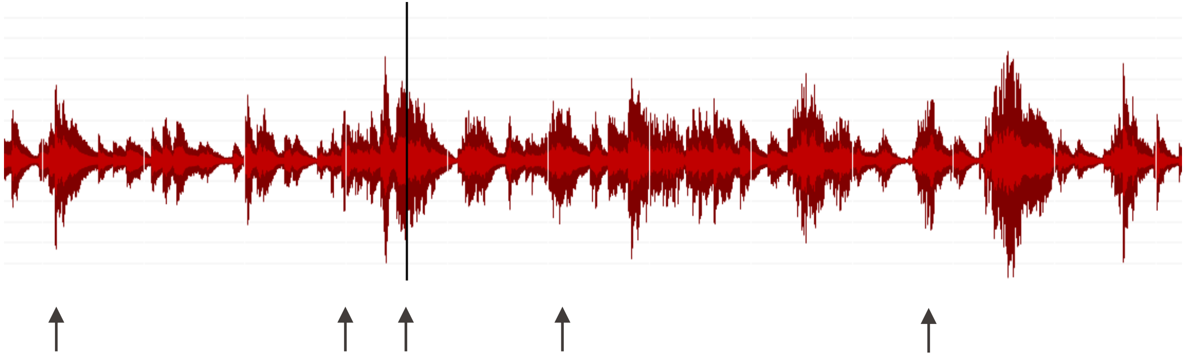
- Misalignment
- Phase locks with a correct metrical grid but not necessarily at the intended beat level:

→ Interprets half beats as beats

→ In general, happens to detect a multiple of the number of beats (“octave error”)

Approaches – 2. Chord detection

Chord detection by chroma detection algorithms from* madmom's *models* [3]*:



Expected chord sequence C Dm7 C7 C Am …

[3] Korzeniowski et. al, 2016

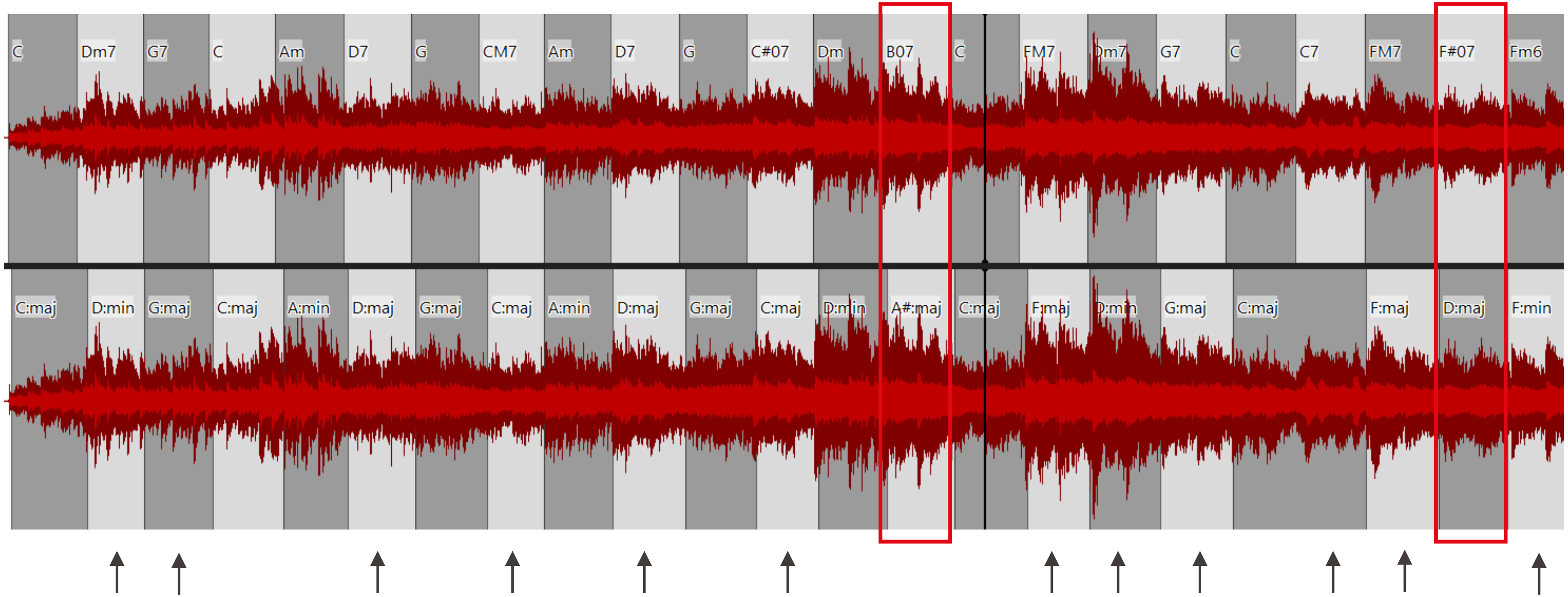
Approaches – 2. Chord detection

Clémentine Lévy-Fidel

Bach, Prelude in C

Aligned annotations

Detected chords

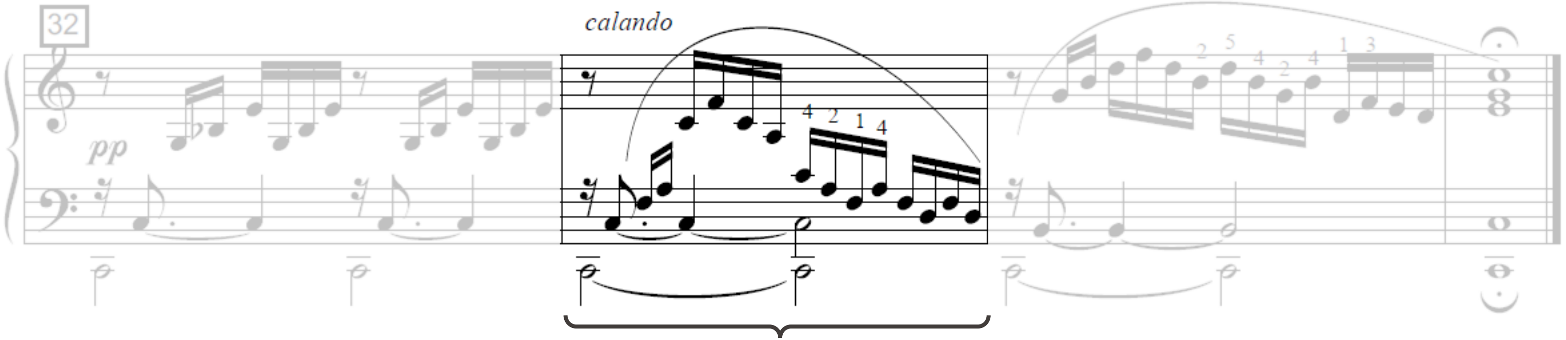


Approaches – 2. Chord detection

Clémentine Lévy-Fidel

Pitfalls:

- Not robust enough:
  - Simplified chords (only major or minor)
  - Frequent false alarms / misses
- Correctness can vary upon annotator



Annotator: **C64**
Detected chord: **F**

Approaches – 3. Score following

Deep learning approach reading score images *[4]*

Pitfalls:

- More demanding to use score images than XML or MIDI files if they are provided
- Method difficult to adapt
- Accuracy not entirely guaranteed

[4] Henkel et. al, 2021

Break – solutions?

Clémentine Lévy-Fidel

Possible workarounds for misalignment pitfalls:

- Use state-of-the-art sequence alignment algorithms
- Train *madmom*'s models or score following models for our type of music

  → higher correct detection rate

  → more robust alignment

Remaining problems:

→ Need of manual verification

→ Lack of aligned data for comparison with ground truth

Approaches – 4. Dynamic Time Warping

Synchronization approach:

For *every* event in the score sequence (i.e., note), match with a corresponding event in the audio sequence

→ Dynamic Time Warping:

Find alignment path between two time series by minimizing distance

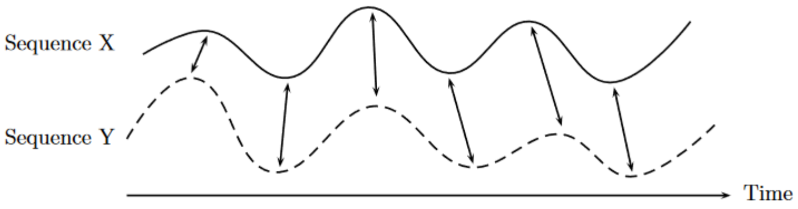


Fig. 4.1. Time alignment of two time-dependent sequences. Aligned points are indicated by the *arrows*

Image source: Müller et al., 2007.

Approaches – 4. Dynamic Time Warping (1/2)

Following Ewert et al. (2009):

- Audio feature extraction
  - Chroma onset (CO) features combining:
    - Chroma filtering
    - Onset energy peak picking
  - Locally adaptive normalization

→ normalize with local maximum of sequence to make CO features invariant to dynamic variations

  - Add temporal decay

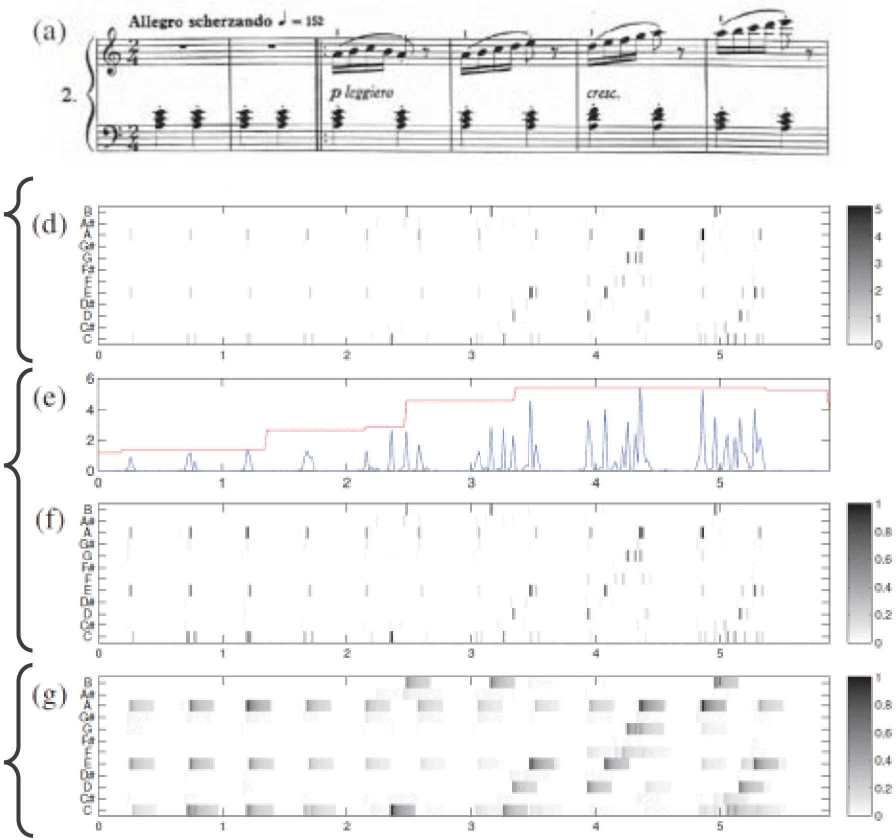


Fig. 1. **(a)** First six measures of Burgmüller, Op. 100, Etude No. 2. **(b)** - **(g)** feature representations of a corresponding audio recording (see Sect. 2 for a description).

Approaches – 4. Dynamic Time Warping (2/2)

Following Ewert et al. (2009)

- Synchronization algorithm
    - Similarity (or cost) matrix: evaluates local similarity cost measure for each pair of features between both sequences
    - Determine optimum-cost alignment path

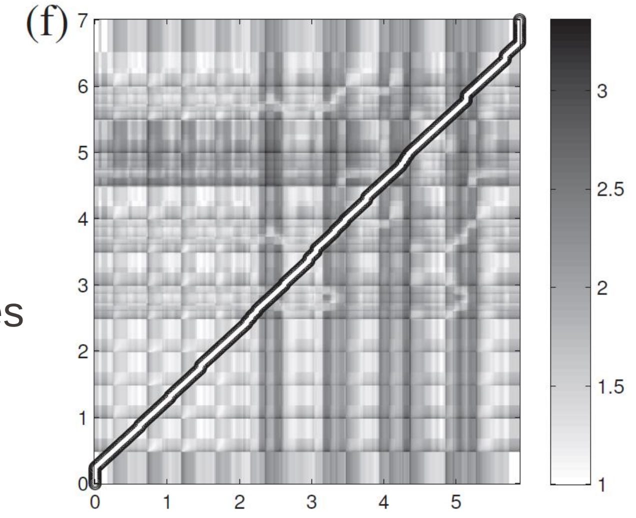


Fig. 2. (a)-(c) Illustration of the effect of the decay operation on the cost matrix level. (d) $C_{\mathbf{chroma}}$, (e), $C_{\mathbf{DLNCO}}$ (f) $C_{\mathbf{chroma}} + C_{\mathbf{DLNCO}}$ for **Burg2**.

The Aligner tool

Applying DTW to the Annotated Mozart Sonatas *corpus*

Dataset content:

- *Notes*

quarterbeats	duration_qb	mc	mn	mc_onset	mn_onset	timesig	staff	voice	duration	gracenote	nominal_duration	scalar	tied	tpc	midi	chord_id
0	0.375	1	1	0	0	3/4	1	1	3/32		1/16	3/2		3	69	0
3/8	0.125	1	1	3/32	3/32	3/4	1	1	1/32		1/32	1		-1	65	1
1/2	0.375	1	1	1/8	1/8	3/4	1	1	3/32		1/16	3/2		-2	70	2

- *Harmonies*

quarterbeats	duration_qb	mc	mn	mc_onset	mn_onset	timesig	staff	voice	label	globalkey	localkey	pedal	chord	numeral	form	figbass
0	0.5	1	1	0	0	3/4	2	1	F.I{	F	I		I	I		
1/2	0.5	1	1	1/8	1/8	3/4	2	1	viio6	F	I		viio6	vii	o	6
1	2.0	1	1	1/4	1/4	3/4	2	1	I	F	I		I	I		

Clémentine Lévy-Fidel

The Aligner tool

Input: notes, labels, audio

Steps:

- Work on **notes** events: convert **symbolic annotations** to **synthesized audio** representation

→ Temporal occurrence, duration, pitch (midi), velocity, instrument

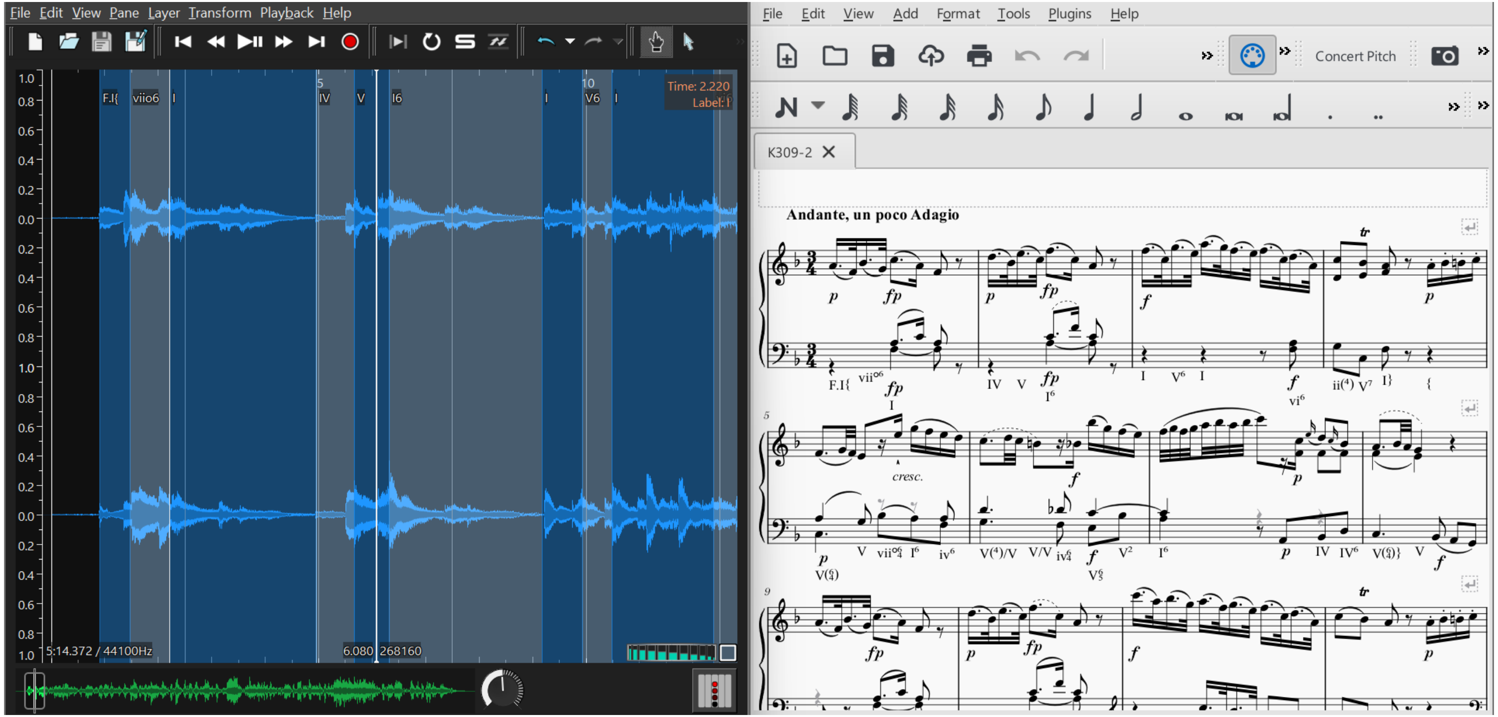
→ Extract similar features as out of audio

- Perform **synchronization to audio** by DTW

- Join **notes** and **labels** datasets on *quarterbeats*

The Aligner tool

Output



Clémentine Lévy-Fidel

The Aligner tool

Clémentine Lévy-Fidel

Github repository (https://github.com/clelf/Aligning-audio-to-annotated-score-labels) containing:

- Dataset preparation tutorial (using `ms3` parser)
- Code to adapt *SyncToolBox* to the use of the *Mozart Sonatas Annotated* dataset
- Command line interface:

```
python aligner.py -a [audio_WAV_file] -n [notes_TSV_file]
-l [labels_TSV_file] -o [CSV_file_to_write_results_to]
```

The Aligner tool

Additional features

- Output aligned notes for score following purpose
- Evaluation module to be used if ground truth data is labelled

Next steps

Clémentine Lévy-Fidel

- Generalize pipeline to data outside the *Mozart Sonatas* corpus
- Generalize algorithm to scores and/or audio recordings that contain repetitions, following Grachten et. al (2013)
- Label data manually and use ML-based approaches
- Investigate JKU's *matchmaker* library when published

EPFL

Personal highlights

- Invaluable help of Steffen, Gabriele and Johannes ☺
- Great introduction to the field of audio processing, music information retrieval, digital musicology …
- Contact with Gerhard Widmer’s lab at Johannes Kepler University (Linz)
- Creating a useful tool!

Acknowledgements

Special thanks to Steffen, Gabriele and Johannes for their support

Thanks to the whole of DCML team for the warm welcome and insightful conversations

EPFL

Clémentine Lévy-Fidel

Thank you!

Questions?