Audio-Symbolic Alignment of Popular Music with application to Automatic Chord Estimation

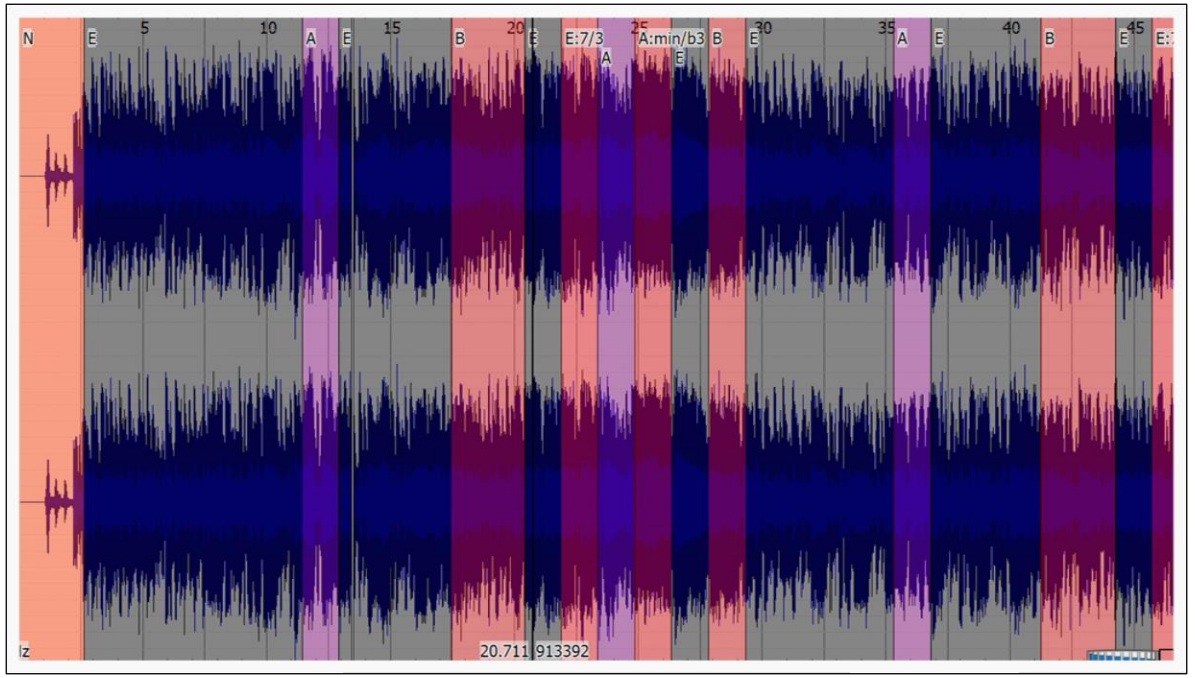
DAPHNE ODEKERKEN

Automatic Chord Estimation (ACE)

Estimate chord sequence in a music piece

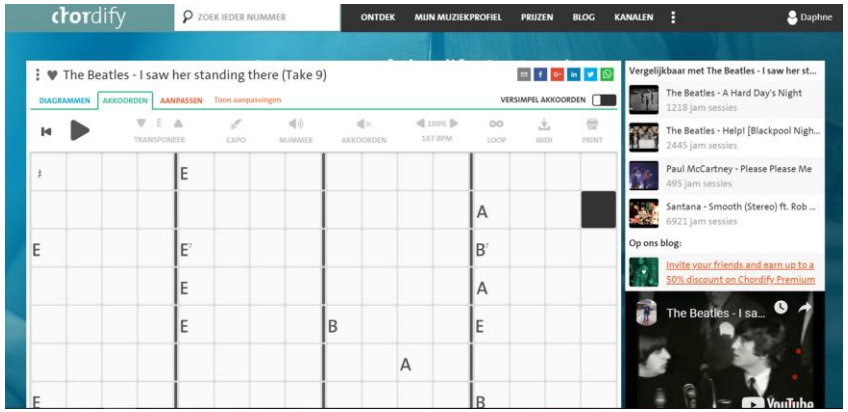
Example:

```
0.000000 2.612267 N
2.612267 11.459070 E
11.459070 12.921927 A
12.921927 17.443474 E
17.443474 20.410362 B
20.410362 21.908049 E
21.908049 23.370907 E:7/3
23.370907 24.856984 A
24.856984 26.343061 A:min/b3
26.343061 27.840748 E
27.840748 29.350045 B
```



Applications of ACE

Music Performance



Research

Subtasks in other MIR research topics:

- Cover song identification
- Key detection
- Genre classification
- Lyric interpretation
- Audio-to-lyric alignment

Evaluation: Chord Symbol Recall

Compare to Ground Truth data set
E.g. *Isophonics Data Set* with 200 songs by The Beatles and Queen

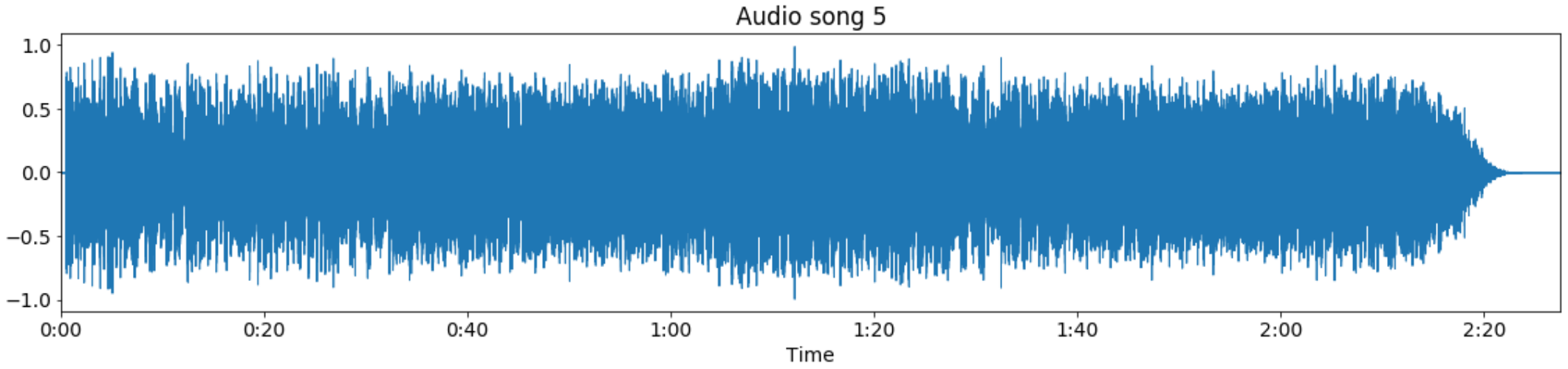
Chord Symbol Recall (CSR):

Ground Truth	C	C	C	G	G	G	G	C	C	C	
Estimation A	C	C	C	E	E	G	G	C	C	C	CSR: 80%
Estimation B	C	G	G	G	G	G	C	C	C	C	CSR: 70%

Weighted CSR (WCSR): weigh the CSR by the length of the song

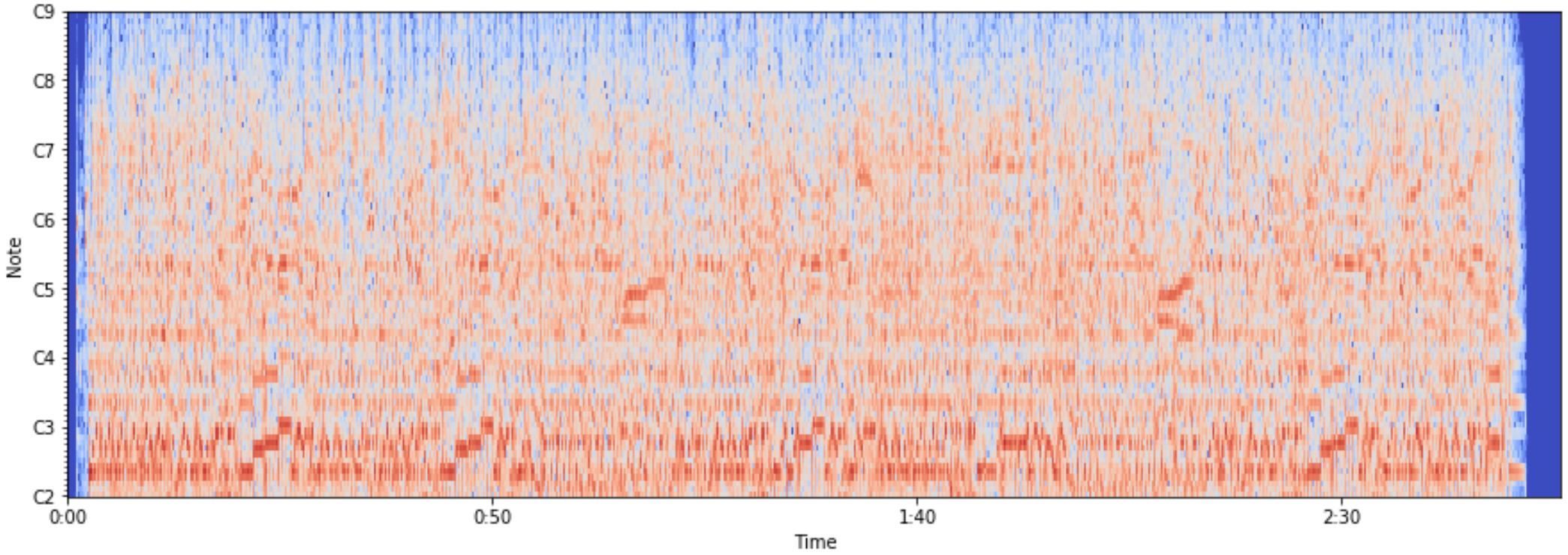
Audio representation

Digitization of the waveform

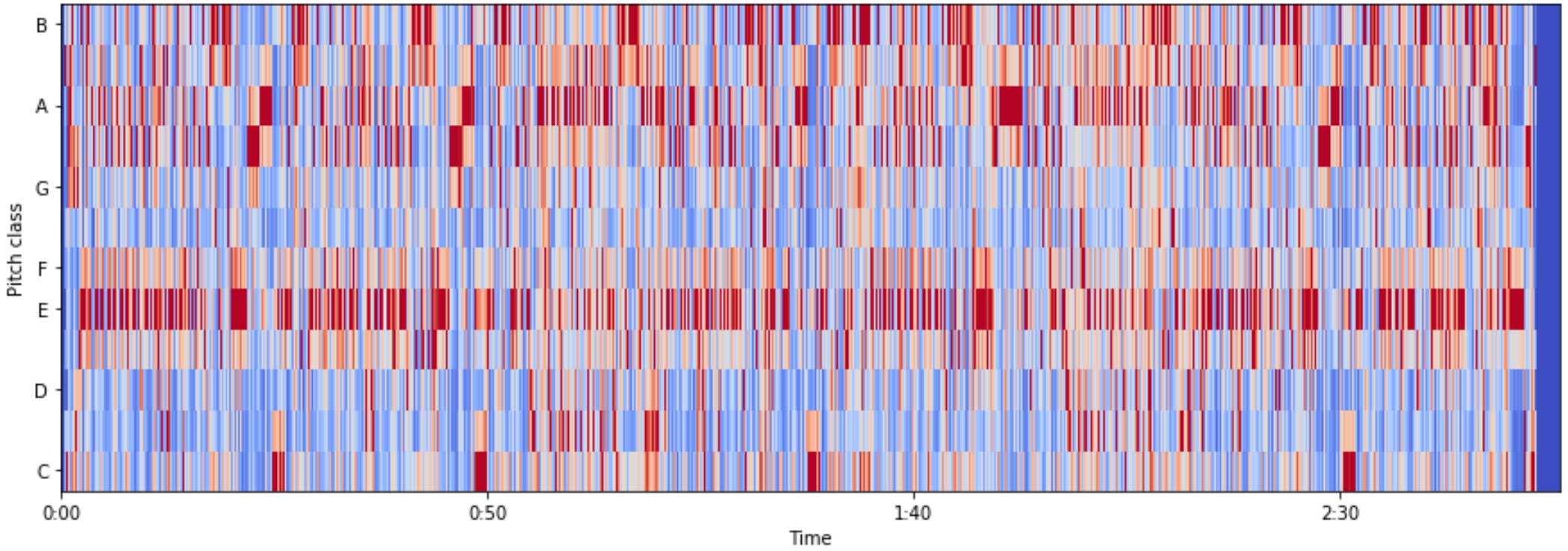


We cannot read the pitches directly

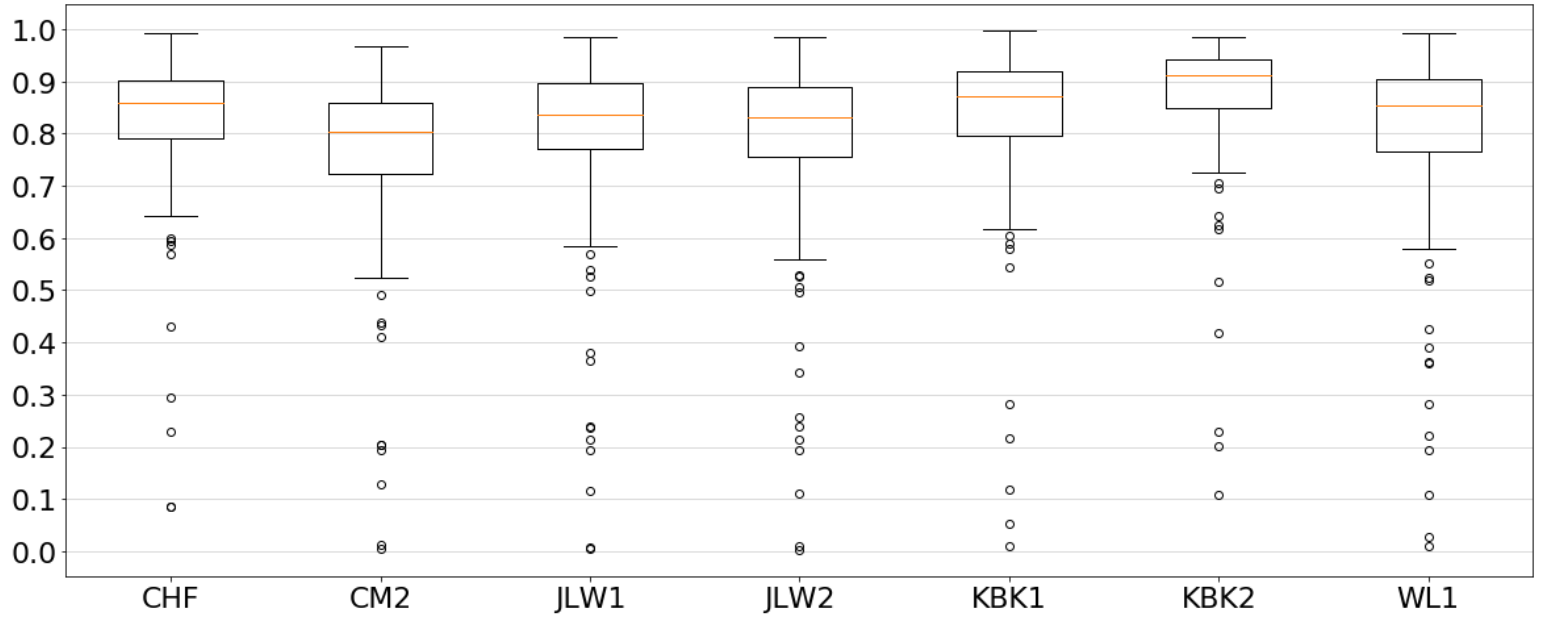
CQT Spectrogram



Chroma



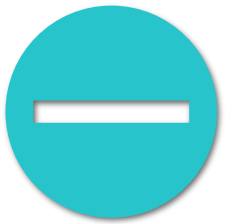
State-of-the-art Audio ACE Systems – Evaluation (CSR)



CHF	Chordify algorithm (Industry)	Industry
CM2	Chris Cannam, Matthias Mauch	MIREX 2017
JLW1, JLW2	Junyan Jiang, Wei Li, Yiming Wu	MIREX 2017
KBK1, KBK2	Filip Korzeniowski, Sebastian Böck, Florian Krebs	MIREX 2017
WL1	Yiming Wu, Wei Li	MIREX 2017

How to improve?

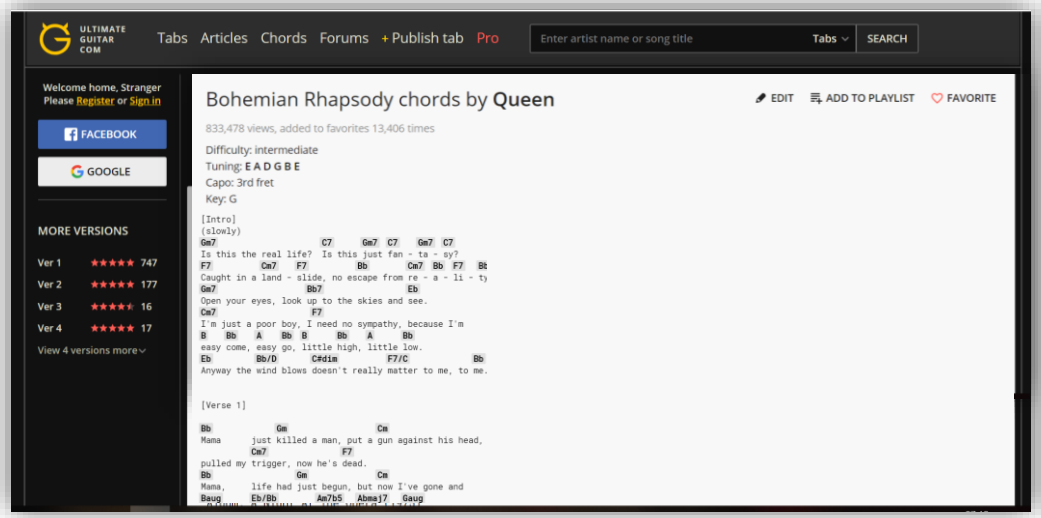
Better results

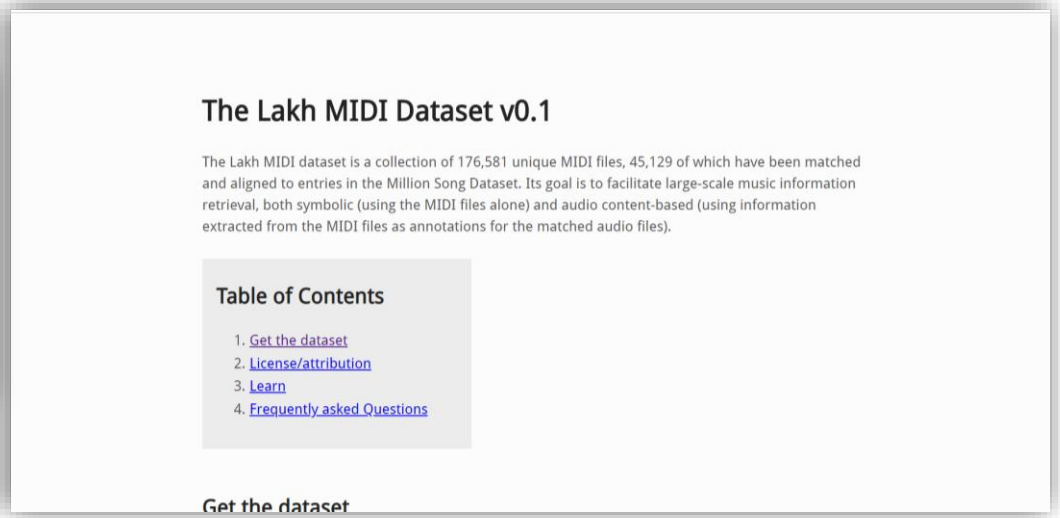


Without overfitting

Incorporating musical knowledge

Crowd-sourced musical knowledge

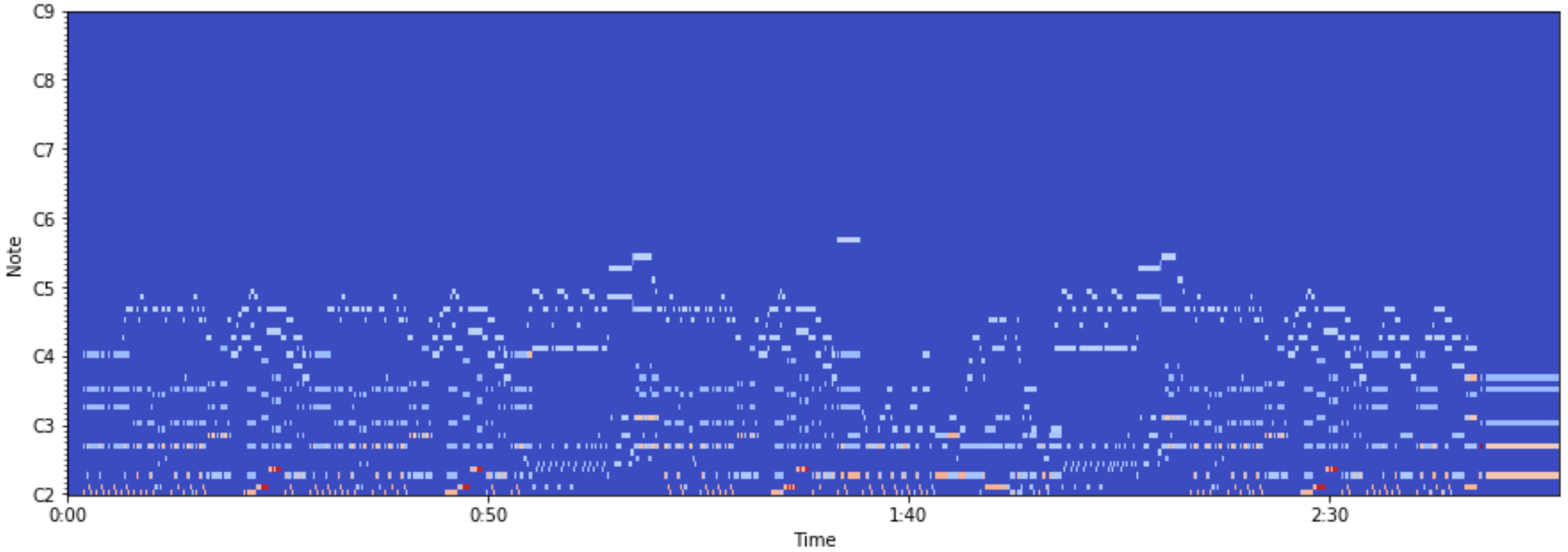




UltimateGuitar: 1,100,000 tabs

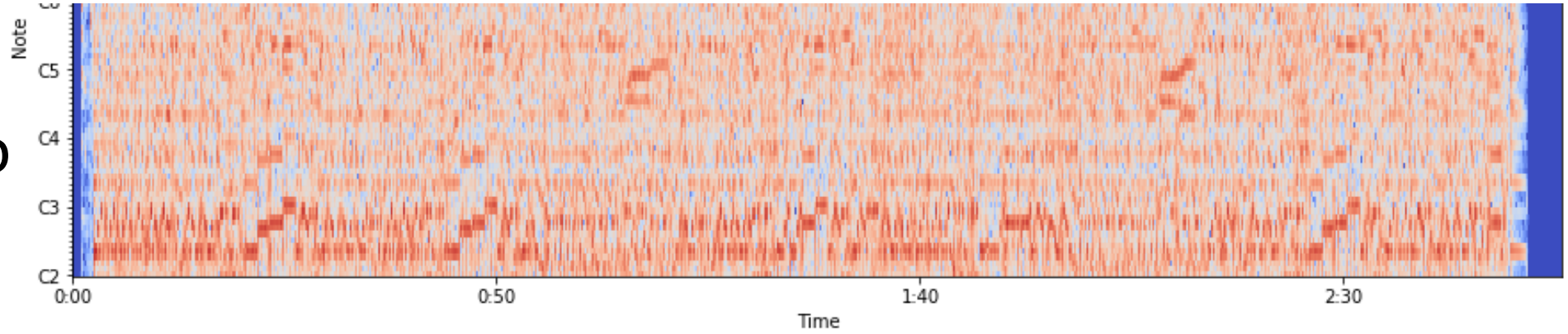
Lakh: 176,581 unique MIDIs

MIDI representation: piano roll

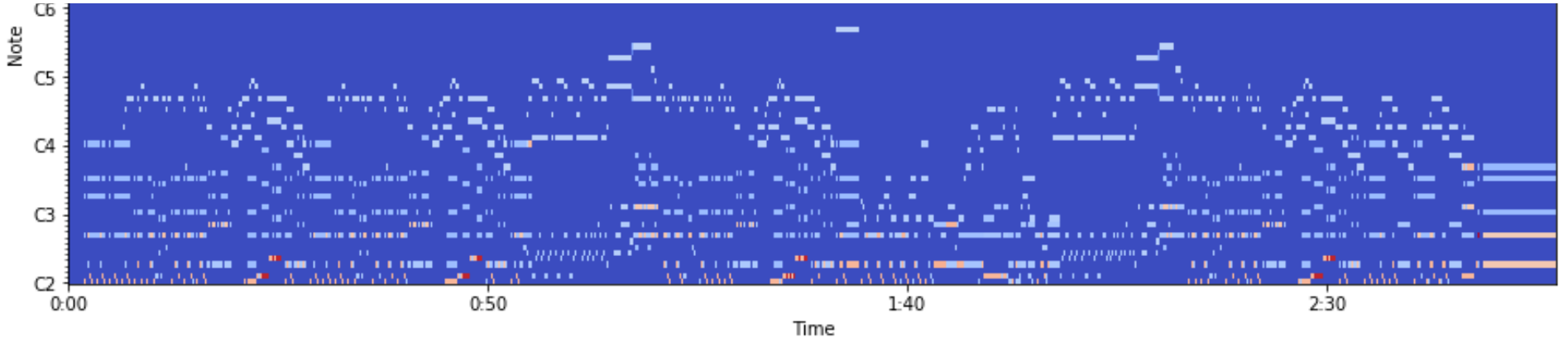


MIDI vs CQT Spectrogram

Audio



MIDI



Tab representation

GUITAR TABLATURE

```
e|-------------------------------|-------------------------------|
B|-------------------------------|-------------------------------|
G|-------------------------------|-------------------------------|
D|-------------------------------|-------------------------------|
A|2---2---2-------2---2-------2--|2---2---2---2-------2---2------|
E|0---0---0-------0---0-------0--|0---0---0---0-------0---0------|
                                        Well, she was   just

         E7
e|-------------------------------|-------------------------------|
B|--------3-------3---3----------|--------3---3-------3---3------|
G|--------1-------1---1----------|--------1---1-------1---1------|
D|--------2-------2---2----------|--------2---2-------2---2------|
A|--------2-------2---2----------|--------2---2-------2---2------|
E|--------0-------0---0----------|--------0---0-------0---0------|
                 seventeen,                             you know

         A7                            E7
e|--------3-------3---3----------|--------0-------0---0----------|
B|--------2-------2---2----------|--------3-------3---3----------|
G|--------2-------2---2----------|--------1-------1---1----------|
D|--------2-------2---2----------|--------2-------2---2----------|
A|--------0-------0---0----------|--------2-------2---2----------|
E|-------------------------------|--------0-------0---0----------|
                 what I      mean                       and the
```

CHORD SHEET

```
I SAW HER STANDING THERE
THE BEATLES

[Verse]
            E7                          A7           E7
Well she was just seventeen and you know what I mean
                                          B7
And the way she looked was way beyond compare
   E            E7            A7     Am/C
So how could I dance with another oh,
        E7       B7        E7
when I saw her standing there


[Verse]
     E7                          A7       E7
Well she looked at me and I, I could see
                                                 B7
That before too long I'd fall in love with her
E             E7            A7
She wouldn't dance with another
Am/C           E7       B7        E7
Oh,  when I saw her standing there
```

Audio vs MIDI vs Tabs

AUDIO

+ Input data

- Difficult to extract chords

MIDI

+ Easy to extract notes

+ Available on the Internet

+ Musical knowledge

- Timing may differ

- Various qualities

TAB

+ Easy to extract chords

+ Available on the Internet

+ Musical knowledge

- No timing

- Various qualities

Research Question

Can we exploit symbolic representations to improve audio ACE?

DECIBEL Framework

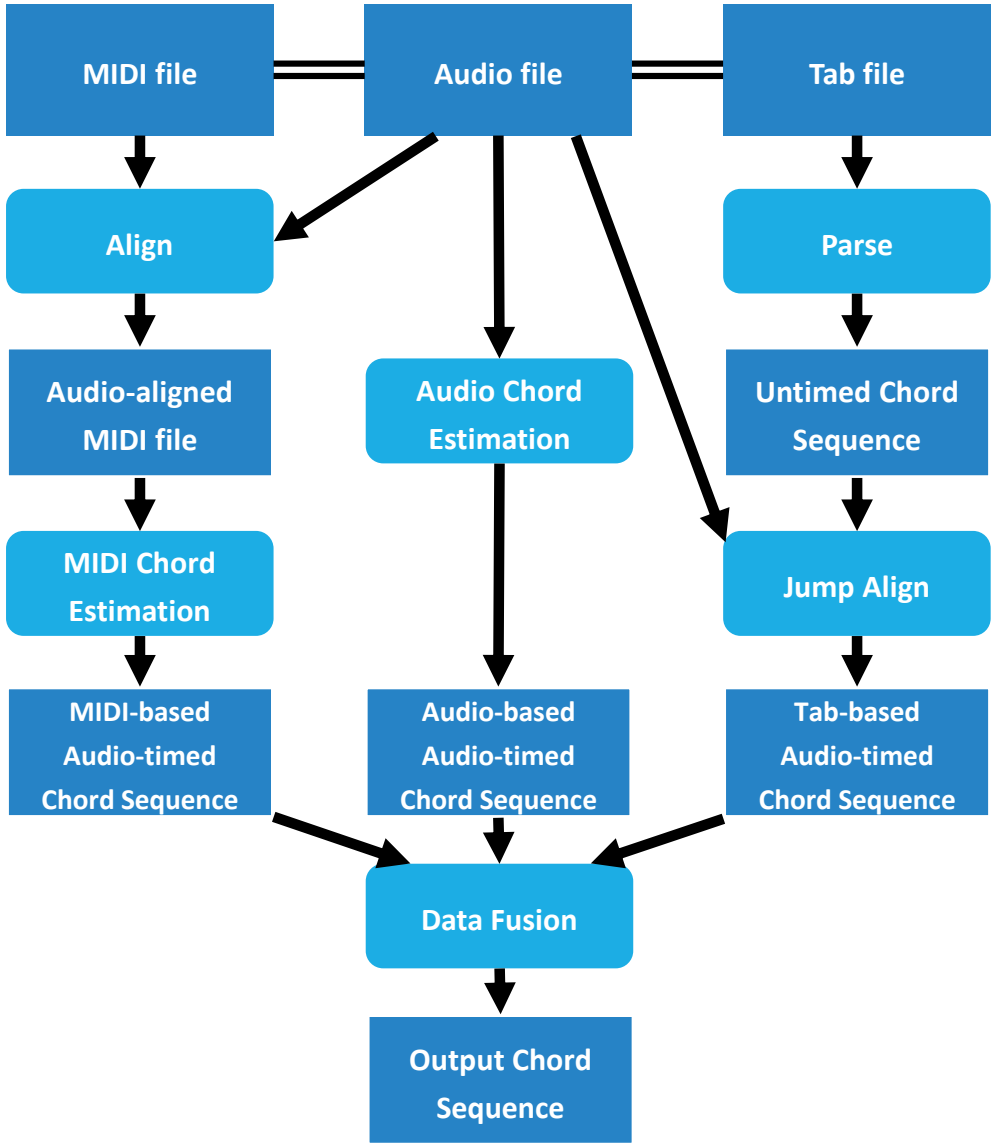
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

Evaluate: is DECIBEL better than the original audio ACE system?



DECIBEL Framework

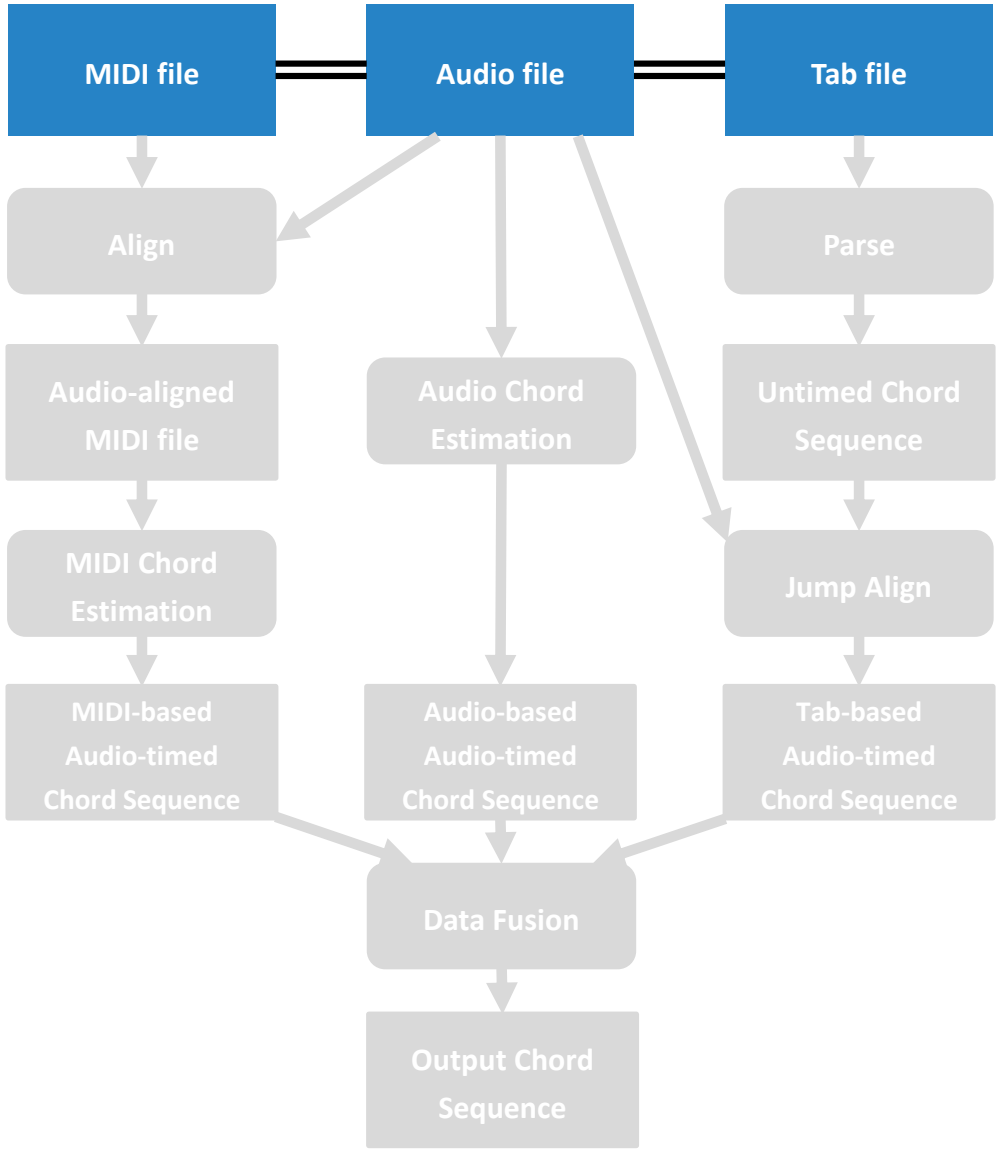
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

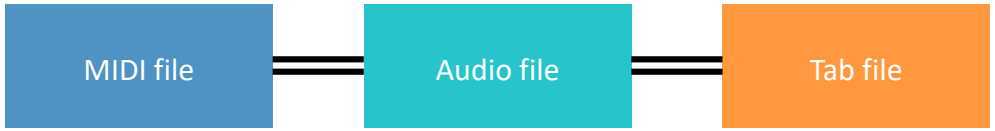
- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

Evaluate: is DECIBEL better than the original audio ACE system?



Data Set

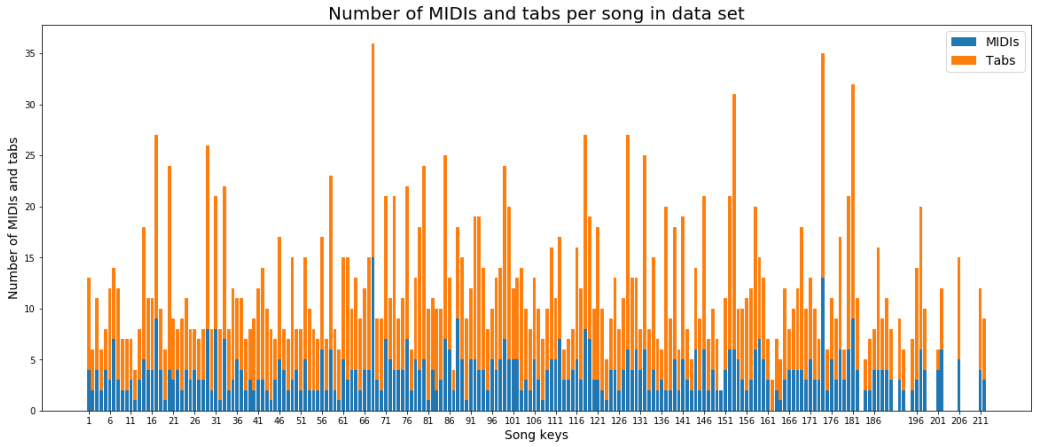


Subset of Isophonics Reference Annotations (Beatles & Queen)

200 annotated audio files

770 MIDI files, manually matched to audio files

1668 tabs, manually matched to audio files



DECIBEL Framework

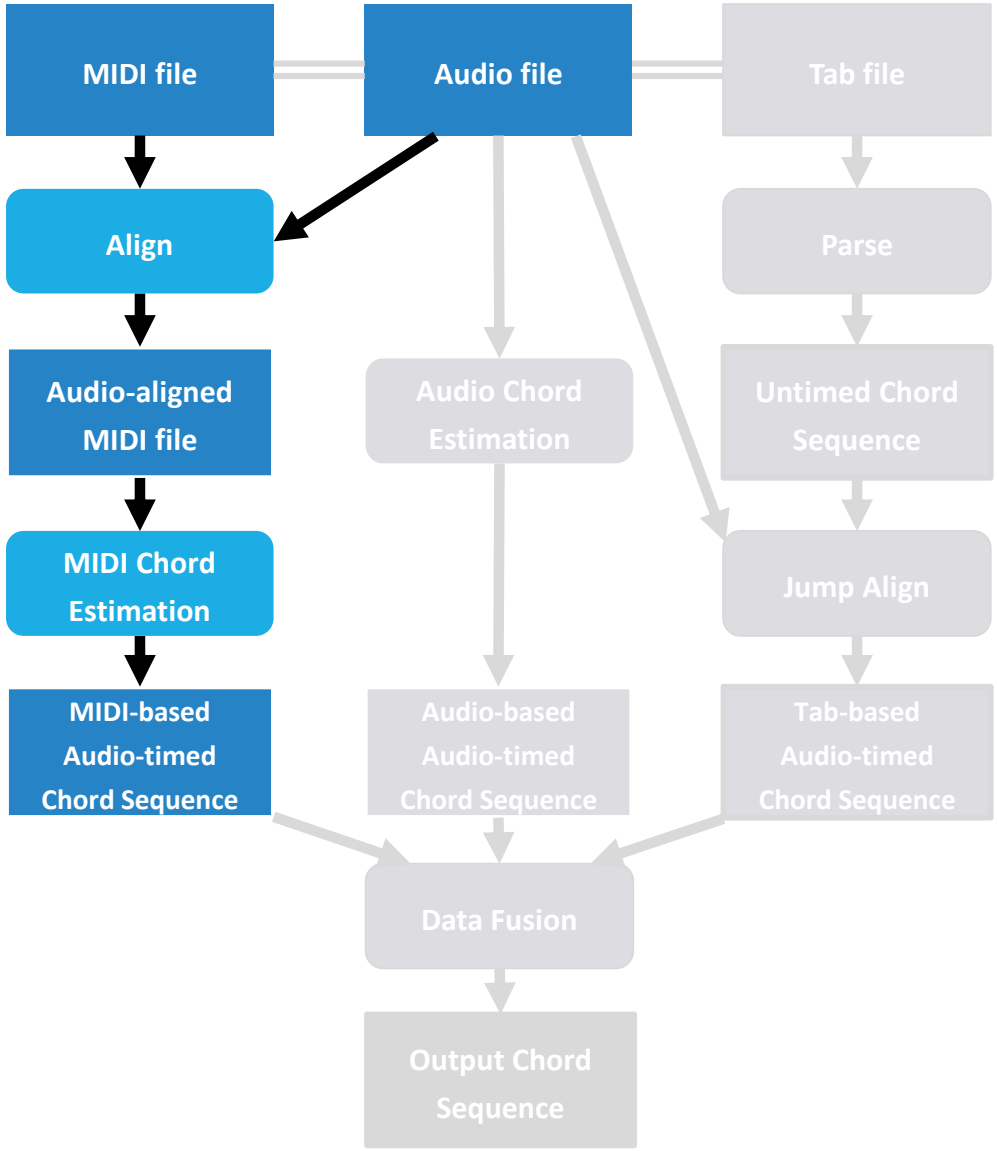
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

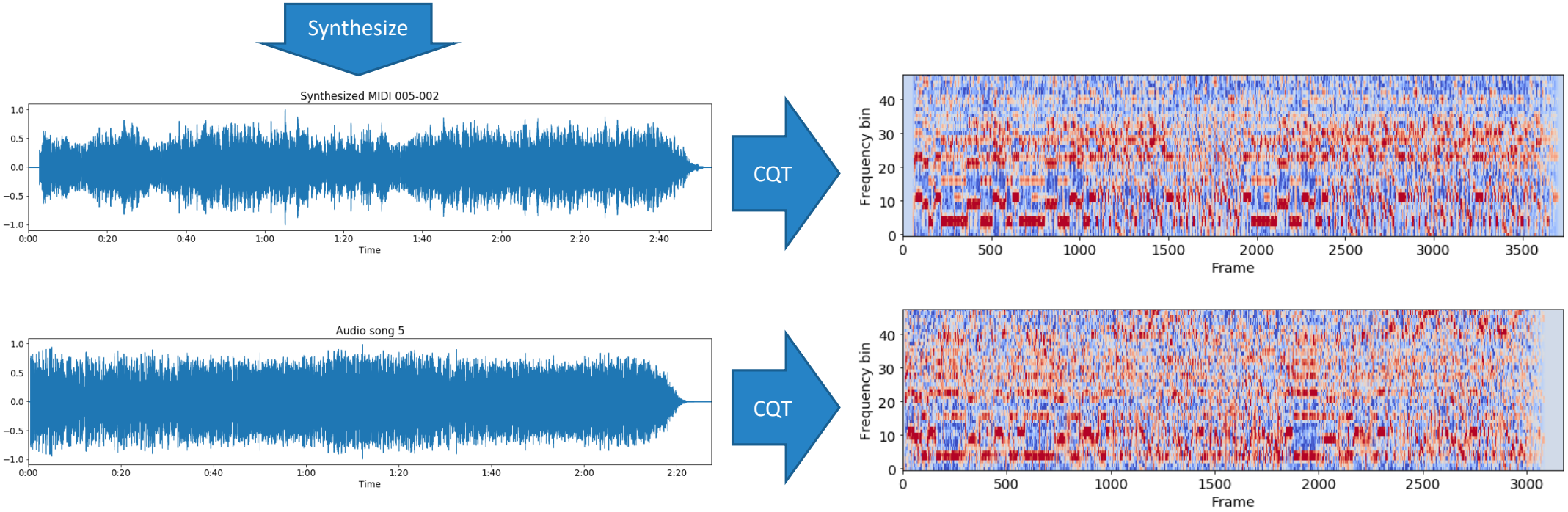
- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

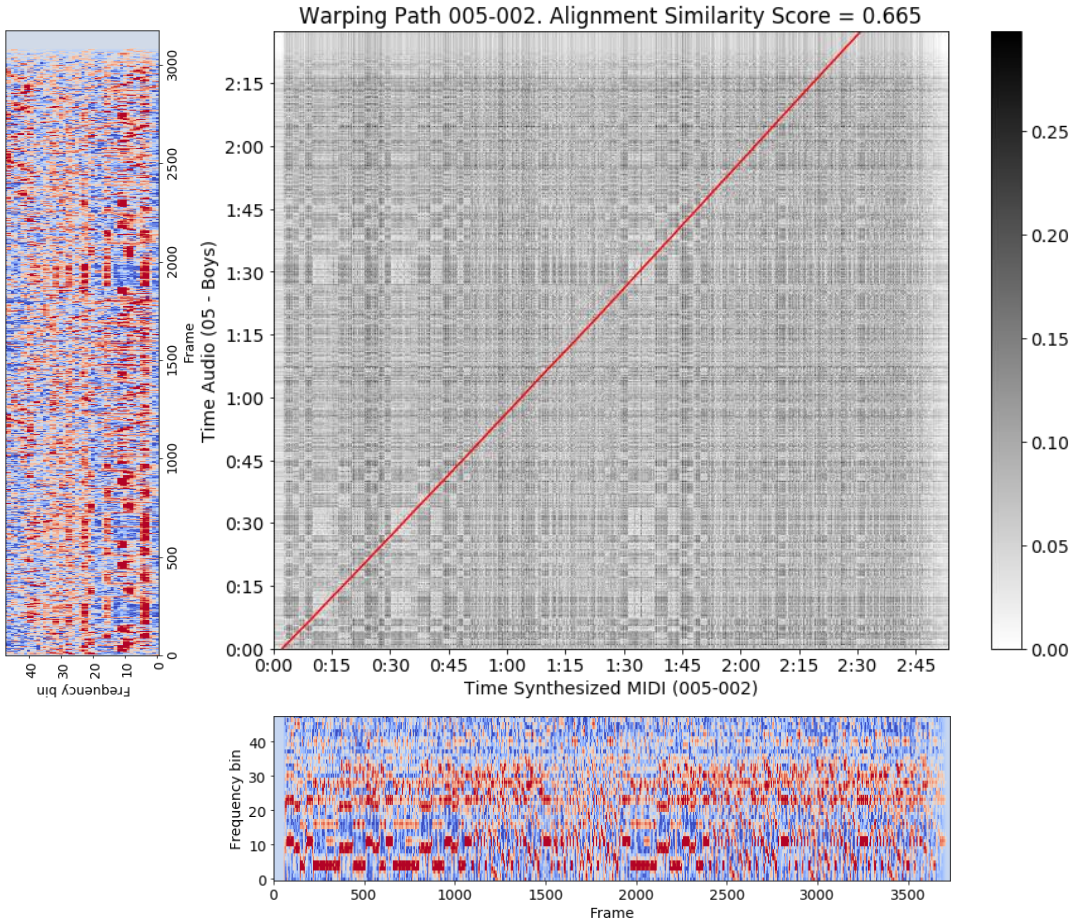
Evaluate: is DECIBEL better than the original audio ACE system?



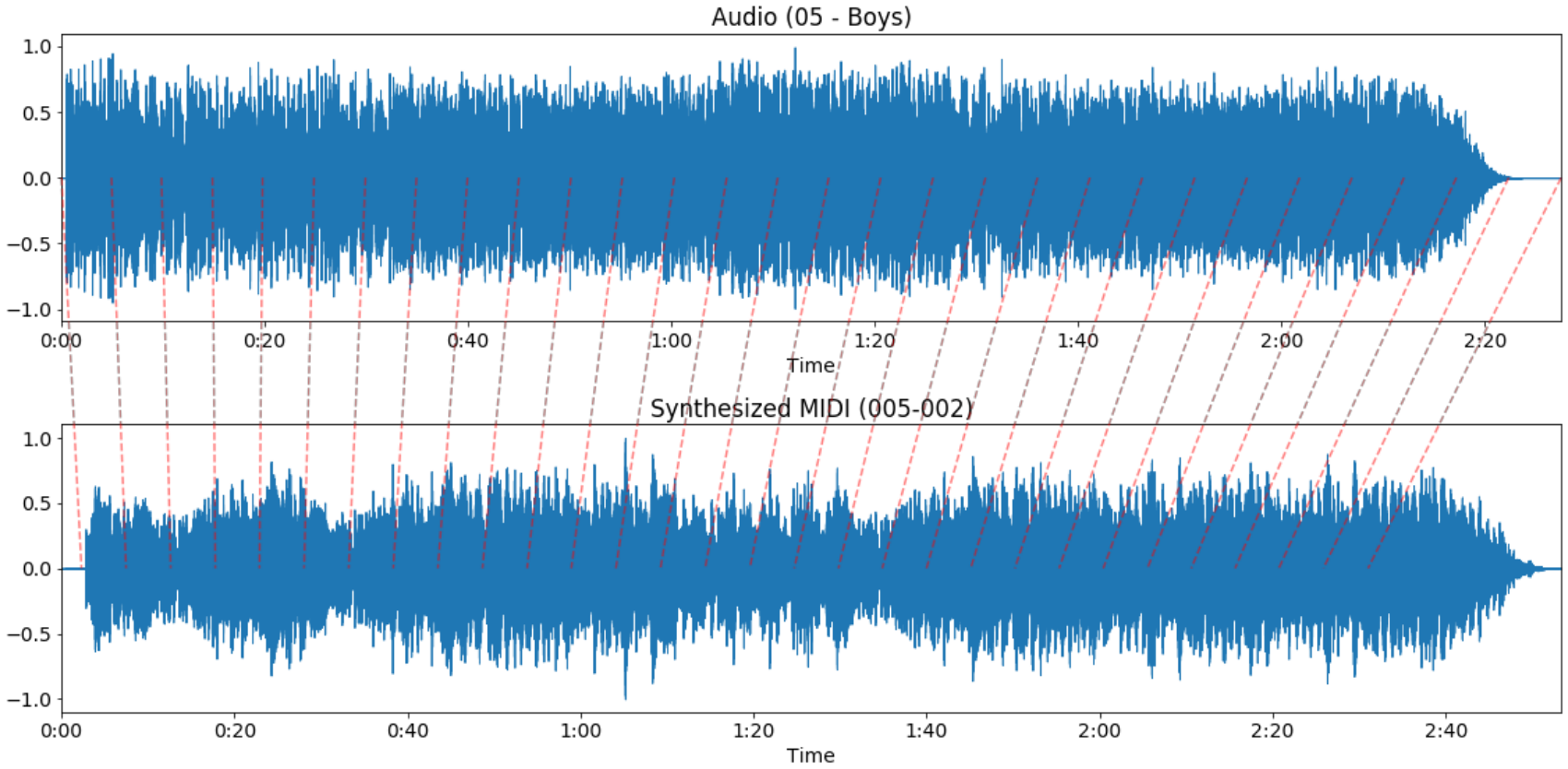
Synthesize MIDI & CQT



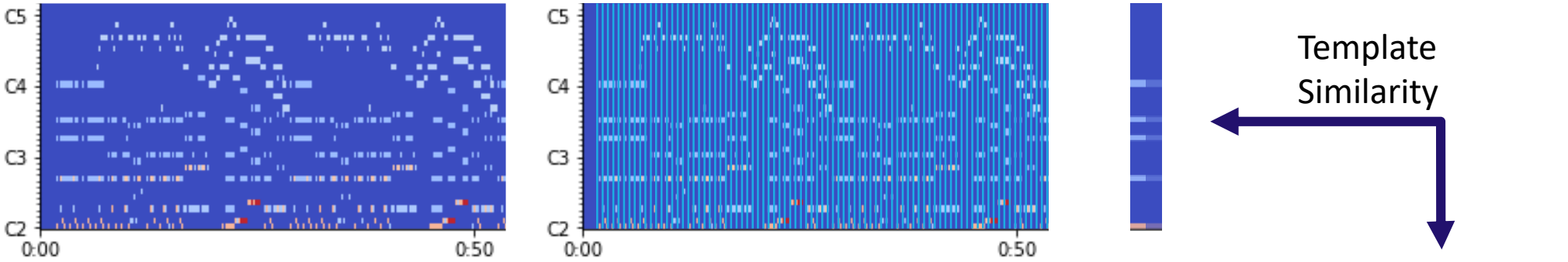
DTW-based Alignment



Audio-Audio Alignment



MIDI chord recognition: CASSETTE



Chord estimation **A**pplied to **S**ymbolic music by **S**egmentation, **E**xtraction and **T**iebreaking **TE**mplate matching

MIDI selection and WCSR

	Beat Segmentation WCSR	Bar Segmentation WCSR
All MIDIs	72.5%	69.2%
Best-ATS MIDI	75.7%	72.9%
Best-CSR MIDI	79.5%	75.5%

DECIBEL Framework

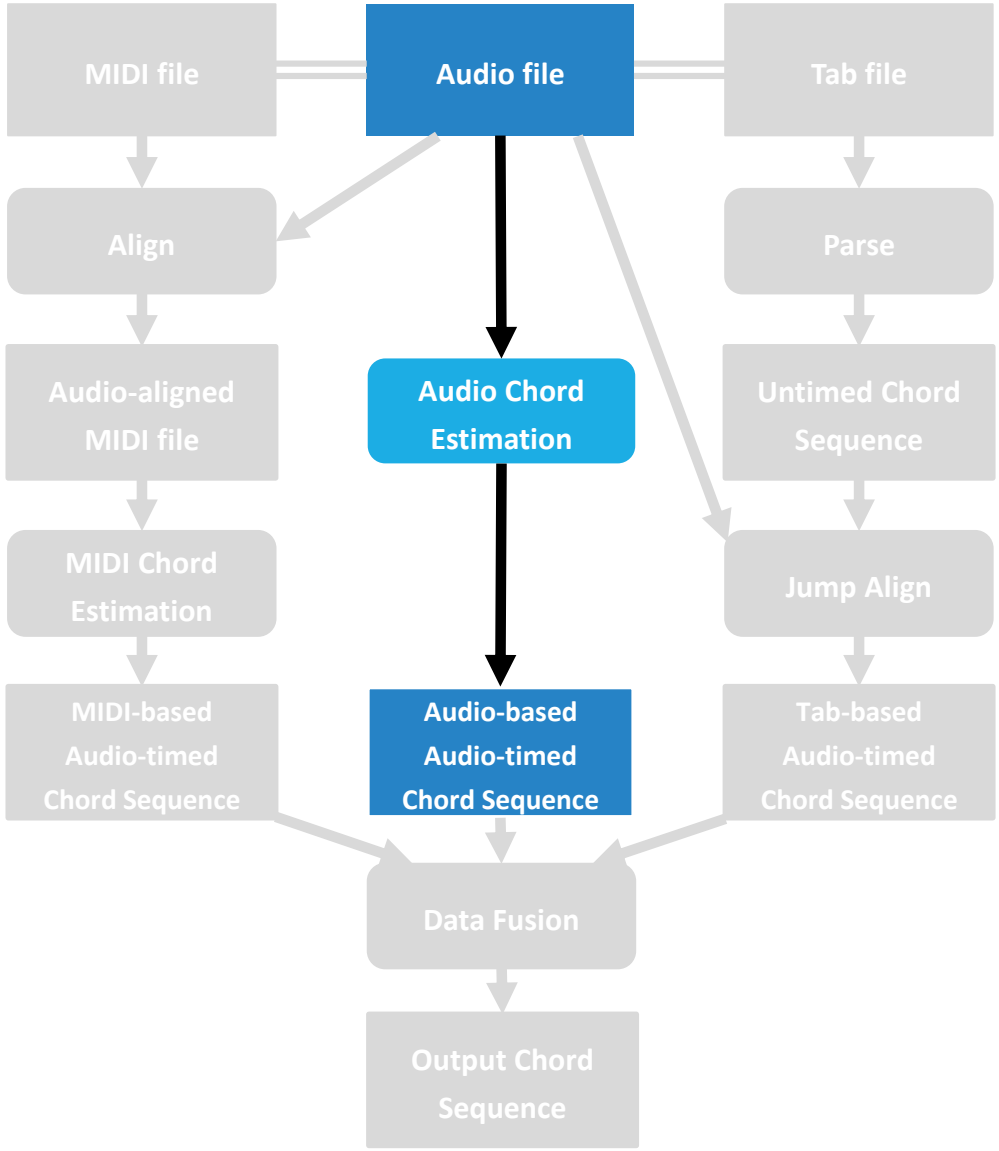
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

Evaluate: is DECIBEL better than the original audio ACE system?



DECIBEL Framework

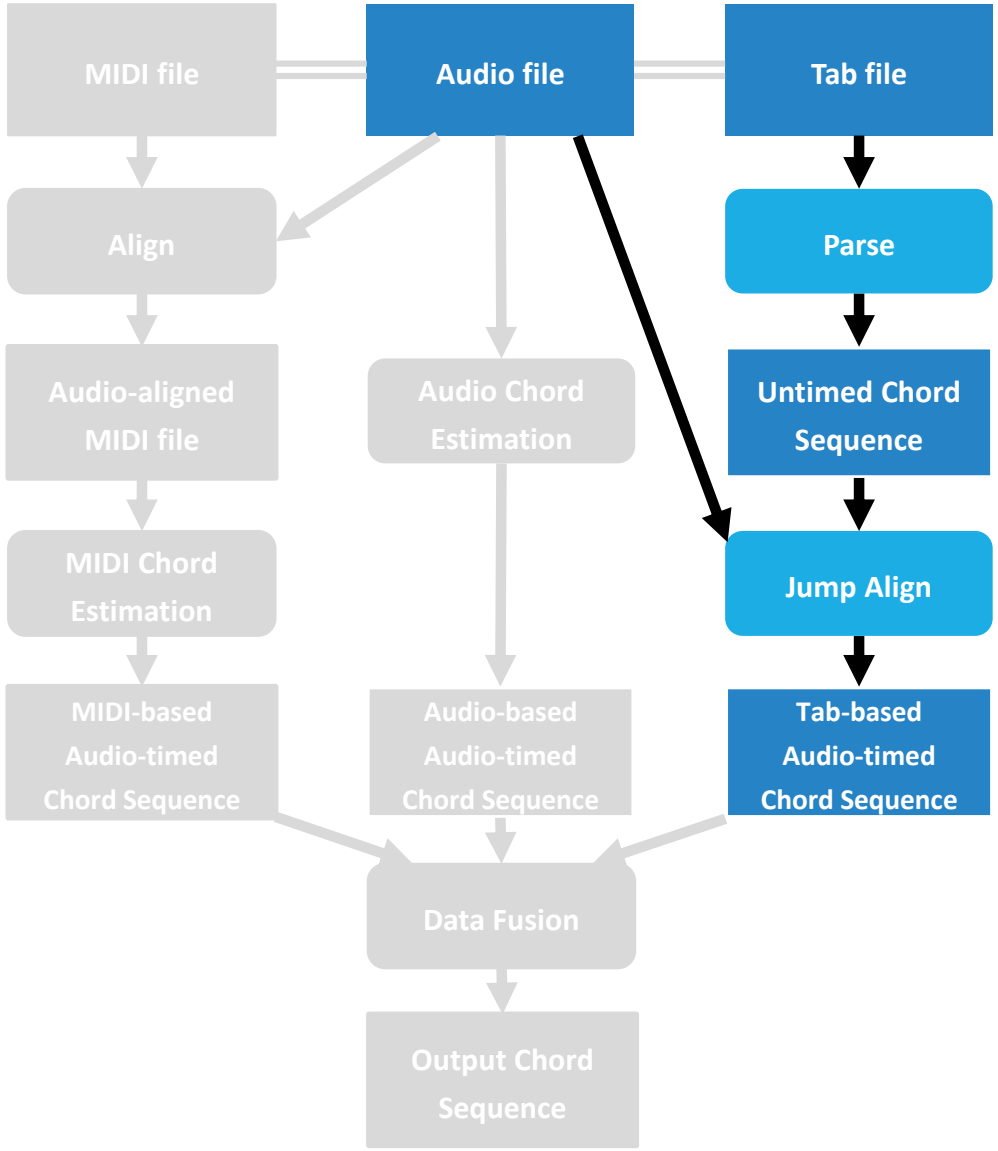
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

Evaluate: is DECIBEL better than the original audio ACE system?



Tab Parsing

```
I SAW HER STANDING THERE
THE BEATLES

[Verse]
            E7                      A7          E7
Well she was just seventeen and you know what I mean
                                      B7
And the way she looked was way beyond compare
   E          E7           A7    Am/C
So how could I dance with another oh,
       E7      B7       E7
when I saw her standing there


[Verse]
     E7                      A7      E7
Well she looked at me and I, I could see
                                          B7
That before too long I'd fall in love with her
E            E7           A7
She wouldn't dance with another
Am/C          E7      B7       E7
Oh,  when I saw her standing there
```

Parse

[E7 A7 E7]

[B7]

[E E7 A7 Am/C]

[E7 B7 E7]

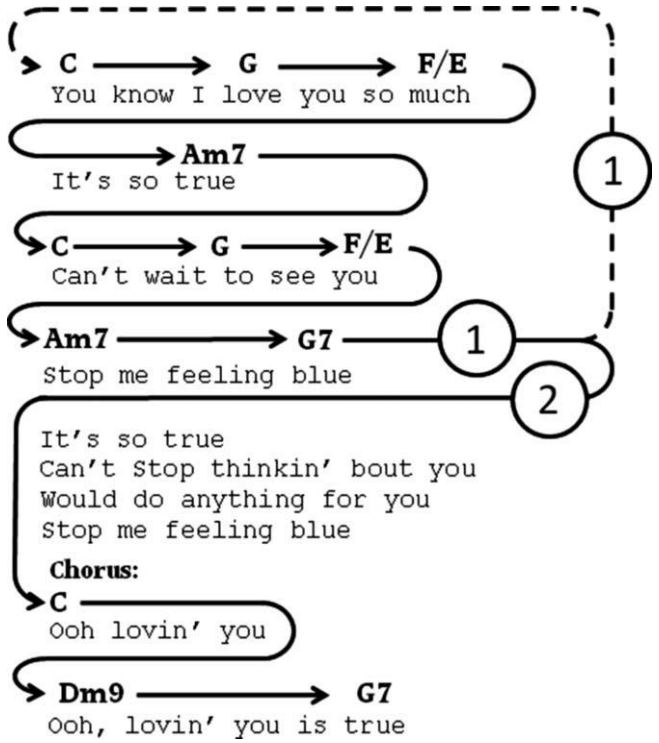
[E7 A7 E7]

[B7]

[E E7 A7]

[Am/C E7 B7 E7]

HMM with altered Transition Probabilities



Matt McVicar, Yizhao Ni, Raul Santos-Rodriguez, and Tijl De Bie. Using online chord databases to enhance chord recognition. *Journal of New Music Research*, 40(2):139–152, 2011.

Evaluation of tab-based ACE

	WCSR
All tabs	73.8%
Best-likelihood tab	77.0%
Best-CSR tab	77.8%

DECIBEL Framework

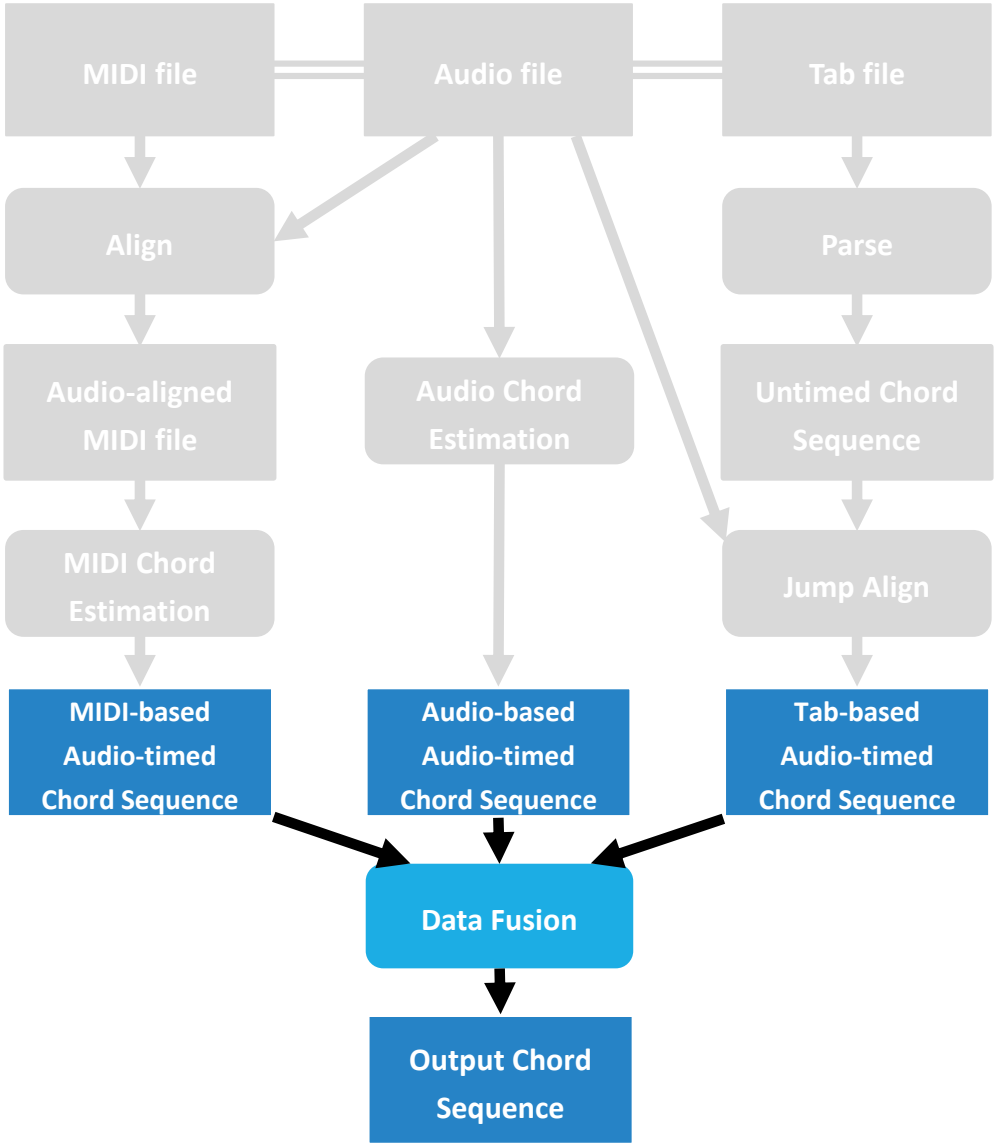
Data set (audio, MIDI, tabs, annotations)

3 representation-specific subsystems:

- **MIDI** alignment and ACE
- **Audio** ACE
- **Tab** parsing and alignment

Data fusion

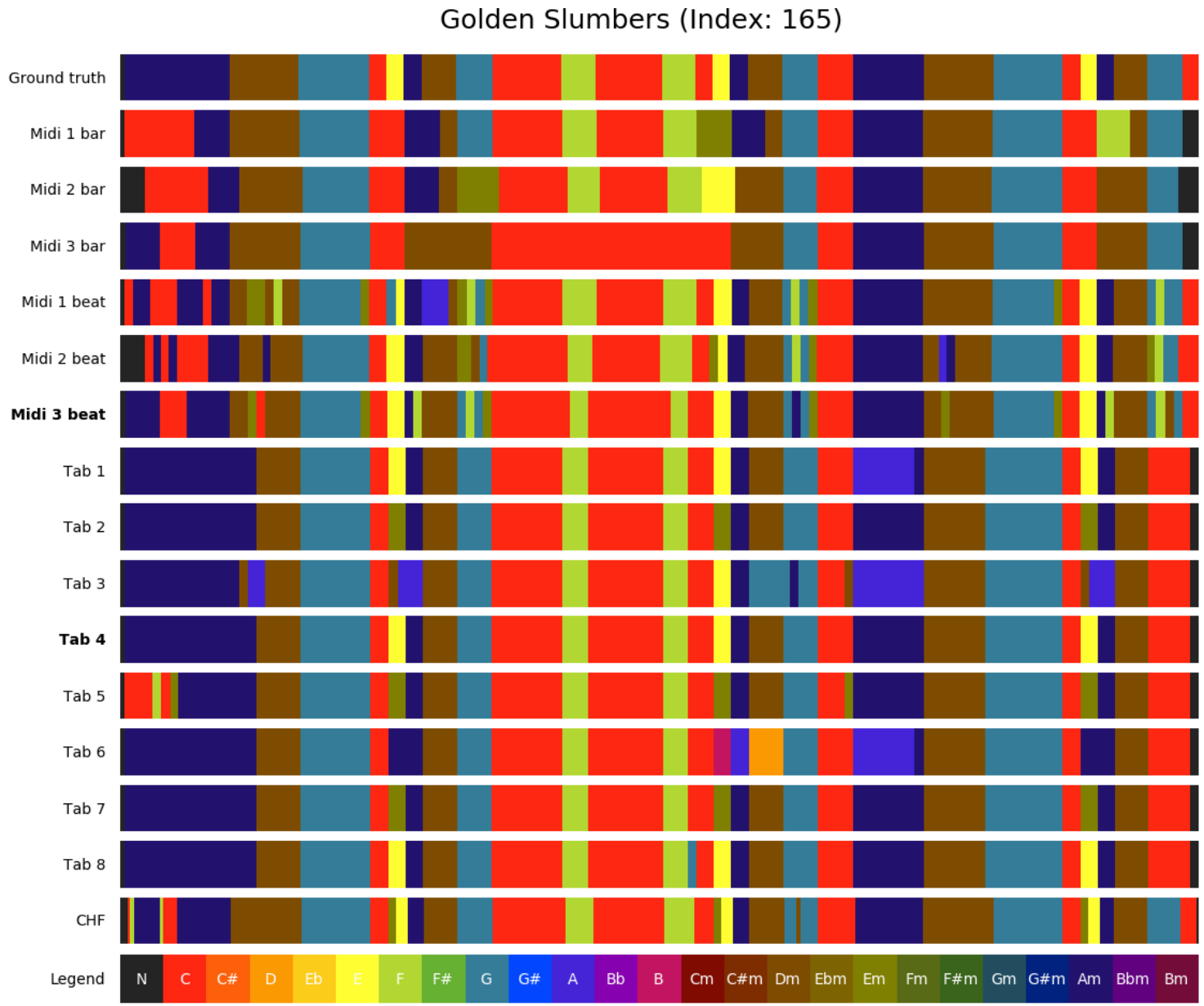
Evaluate: is DECIBEL better than the original audio ACE system?



Data Fusion

Rich harmonic representation

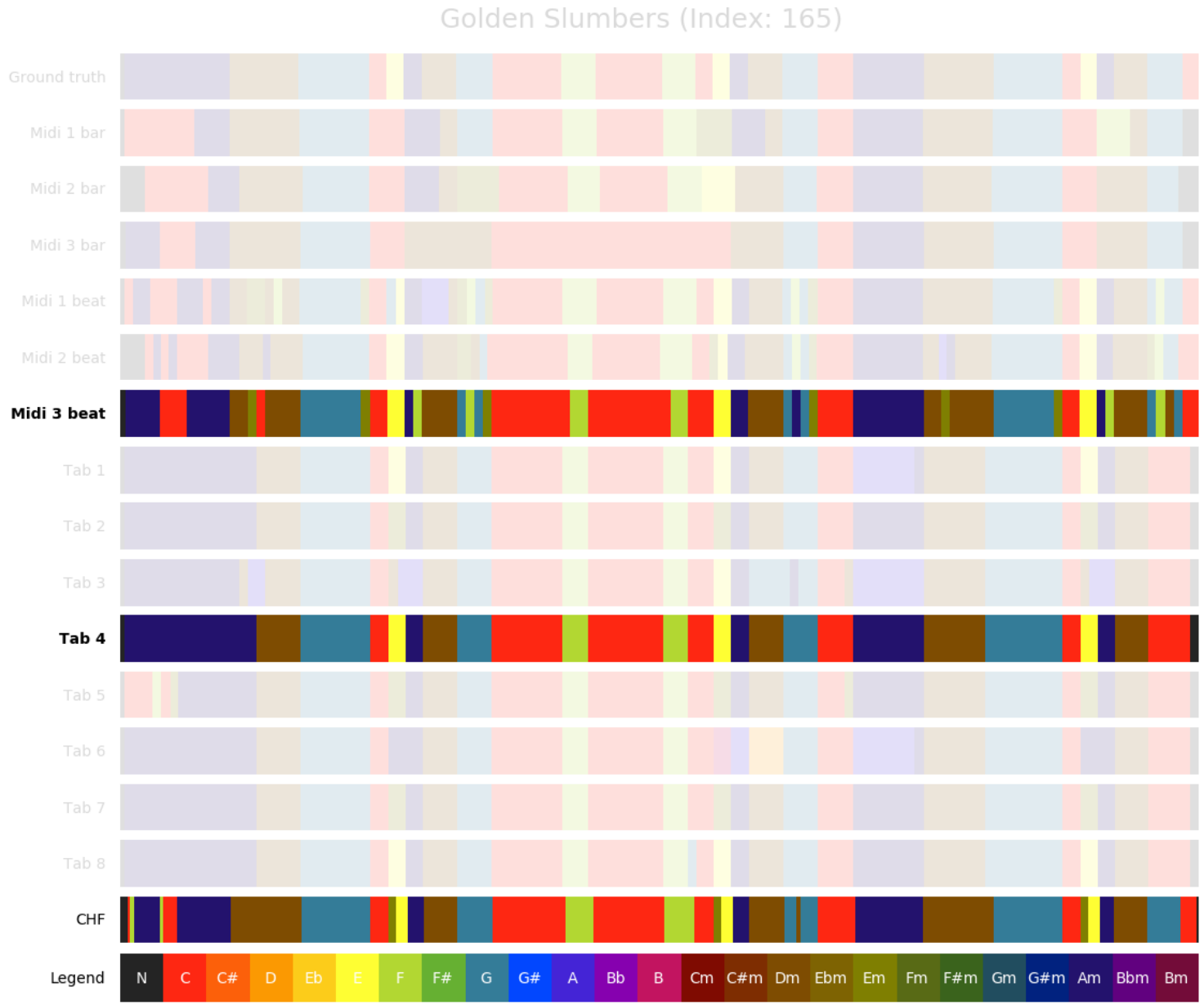
How to find one final output?



Data Fusion

Rich harmonic representation

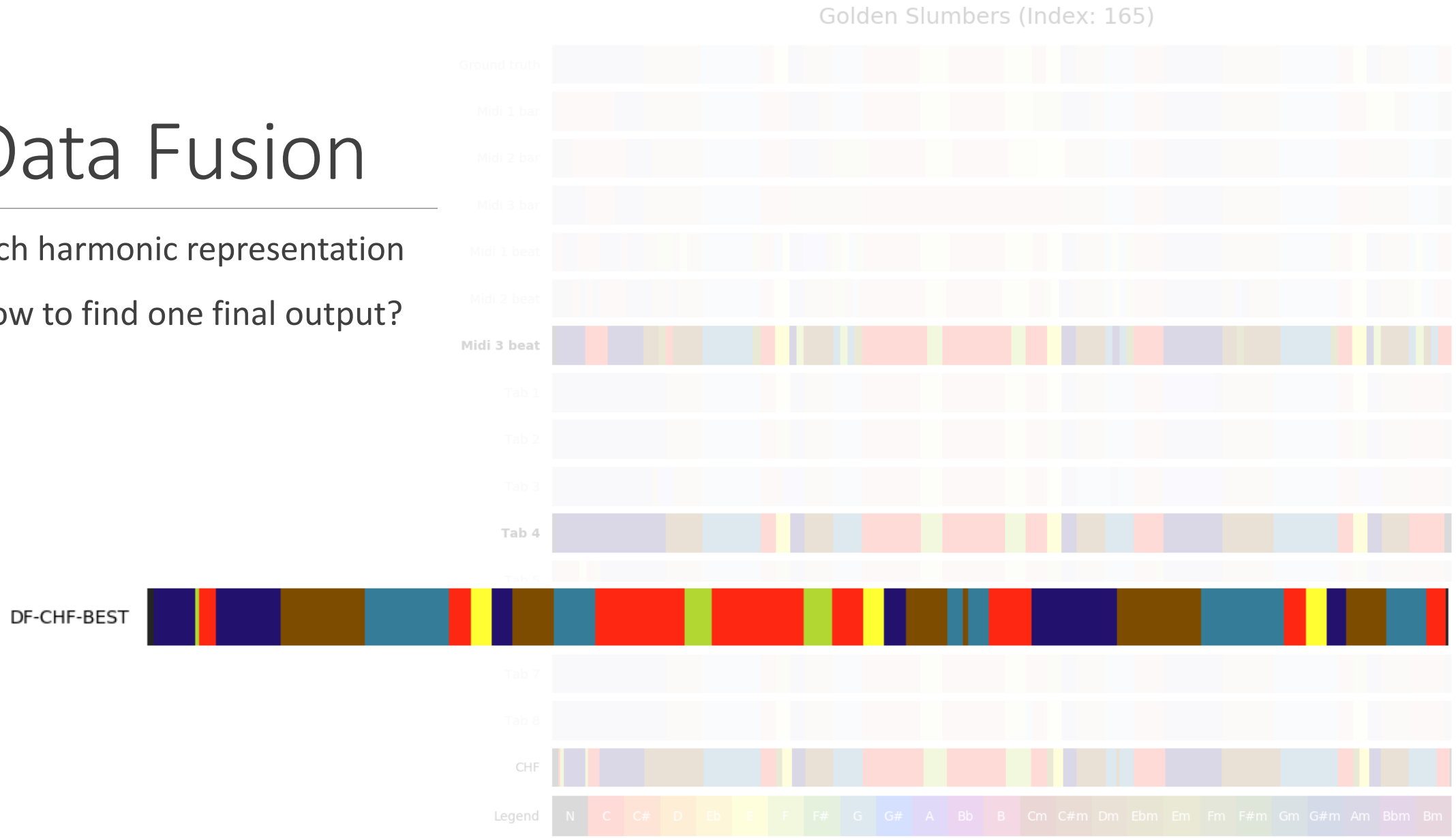
How to find one final output?



Data Fusion

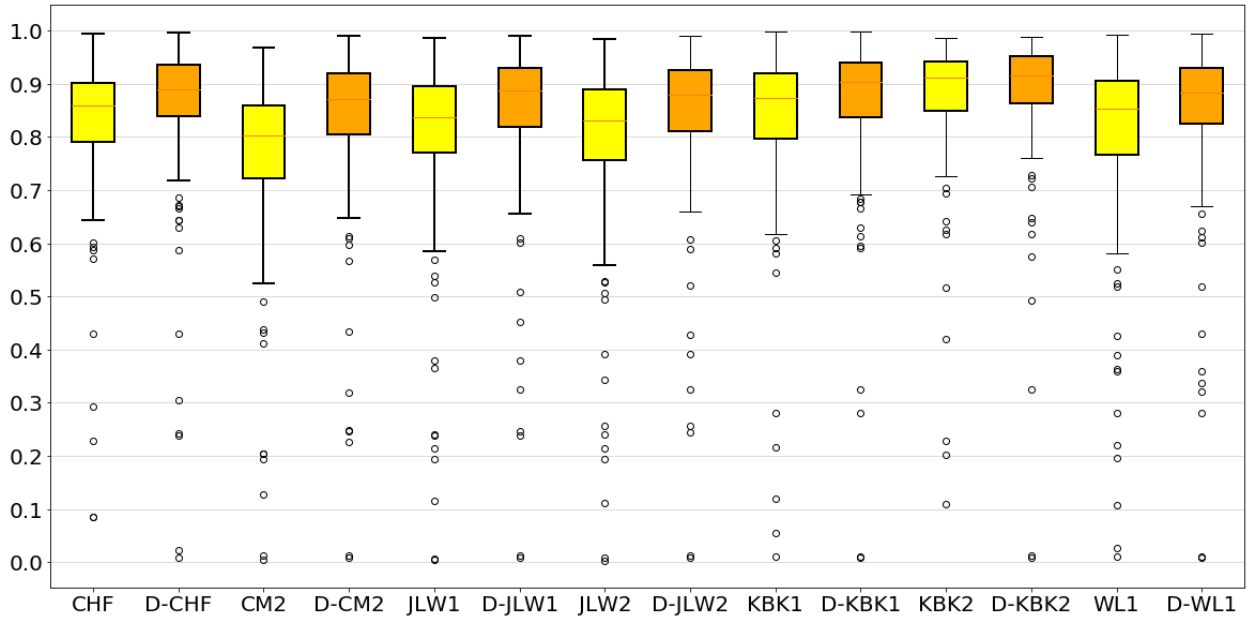
Rich harmonic representation

How to find one final output?



Audio vs. Decibel for all 7 Audio ACE systems

Audio ACE System	Audio WCSR	DECIBEL WCSR	Improve-ment
CHF	0.820	0.848	2.8%
CM2	0.752	0.817	6.5%
JLW1	0.785	0.827	4.2%
JLW2	0.779	0.826	4.7%
KBK1	0.822	0.852	3.0%
KBK2	0.867	0.875	0.8%
WL1	0.793	0.834	4.1%



Conclusion

DECIBEL significantly improves each of the tested audio ACE systems!

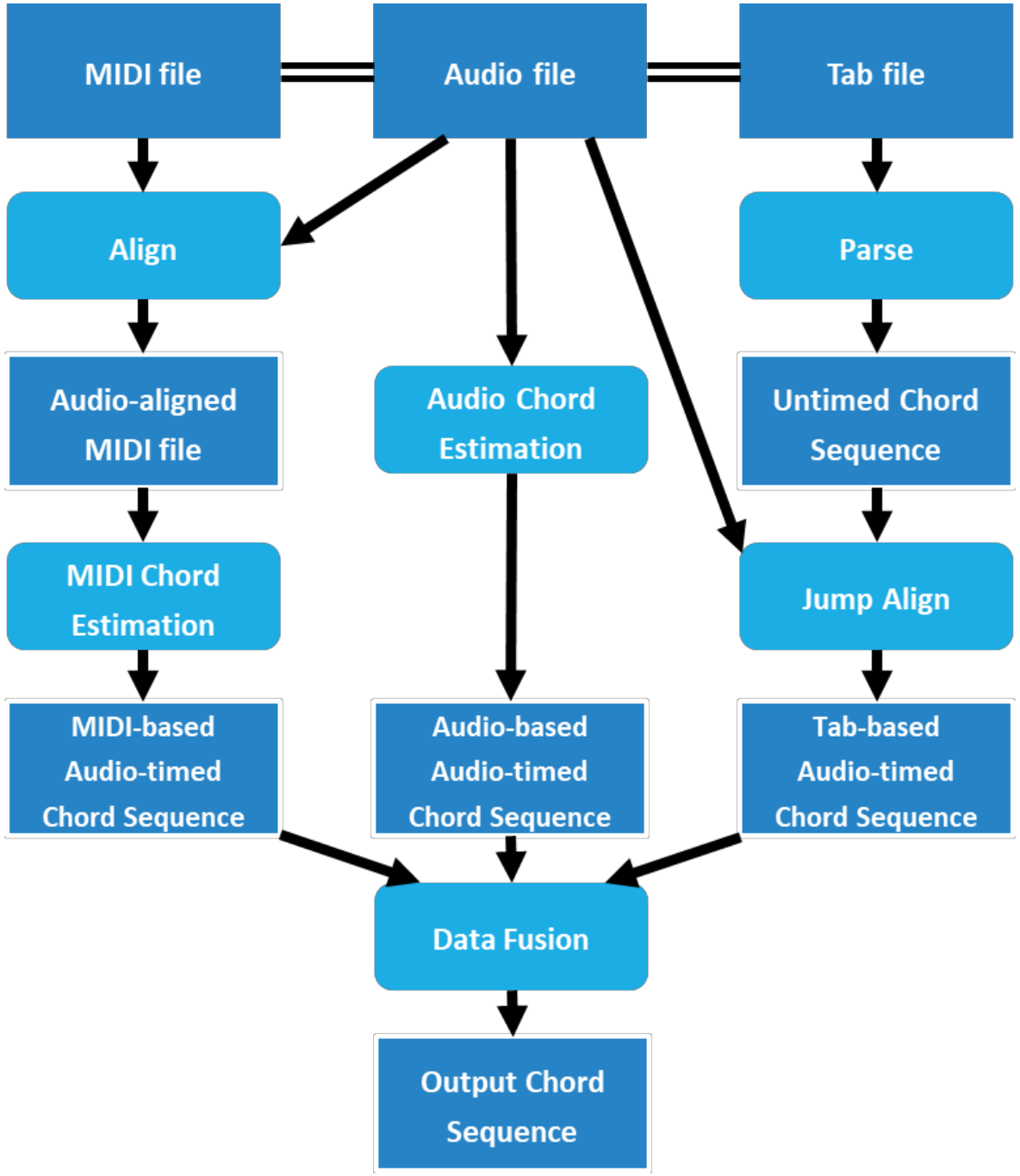
Rich harmonic representation

Full text available at: https://dspace.library.uu.nl/handle/1874/372620

Better results

Without overfitting

Incorporating musical knowledge


MIDI file
Audio file
Tab file
Align
Parse
Audio-aligned MIDI file
Audio Chord Estimation
Untimed Chord Sequence
MIDI Chord Estimation
Jump Align
MIDI-based Audio-timed Chord Sequence
Audio-based Audio-timed Chord Sequence
Tab-based Audio-timed Chord Sequence
Data Fusion
Output Chord Sequence

Questions?