

PAT 463/563 (Fall 2026)

Music & AI

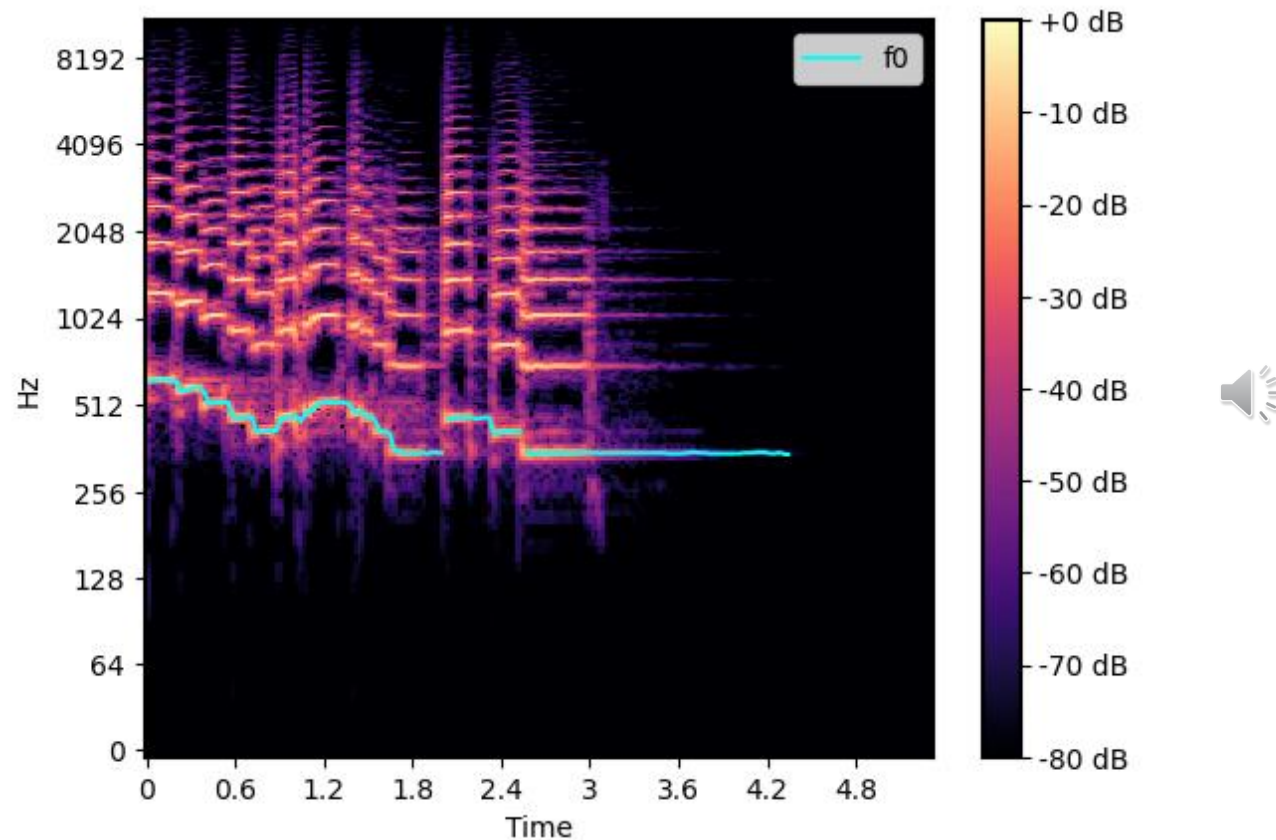
Lecture 11: Music Transcription & Analysis

Instructor: Hao-Wen (Herman) Dong

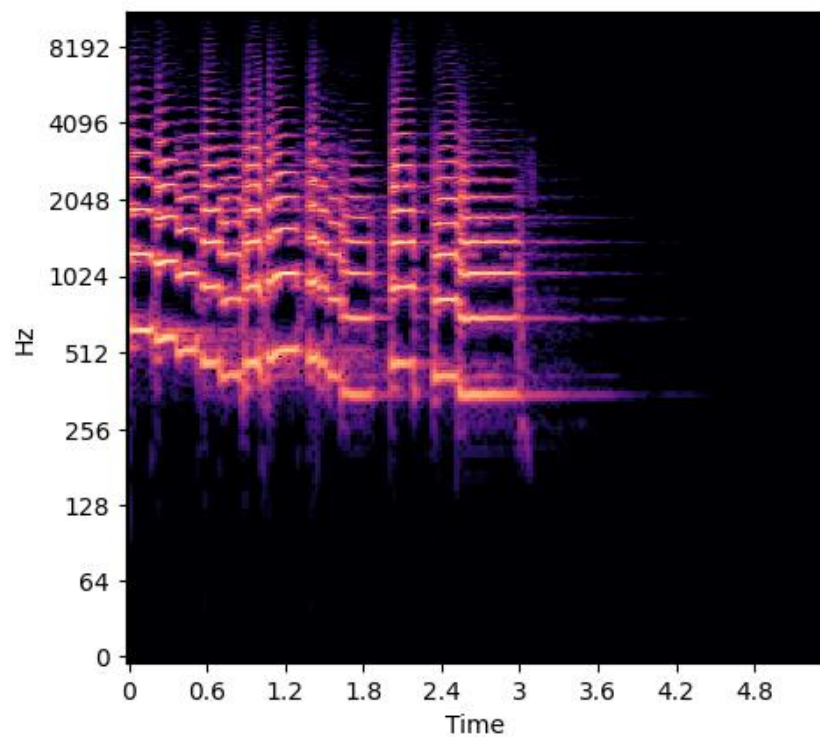
Pitch Estimation

Fundamental Frequency (F0)

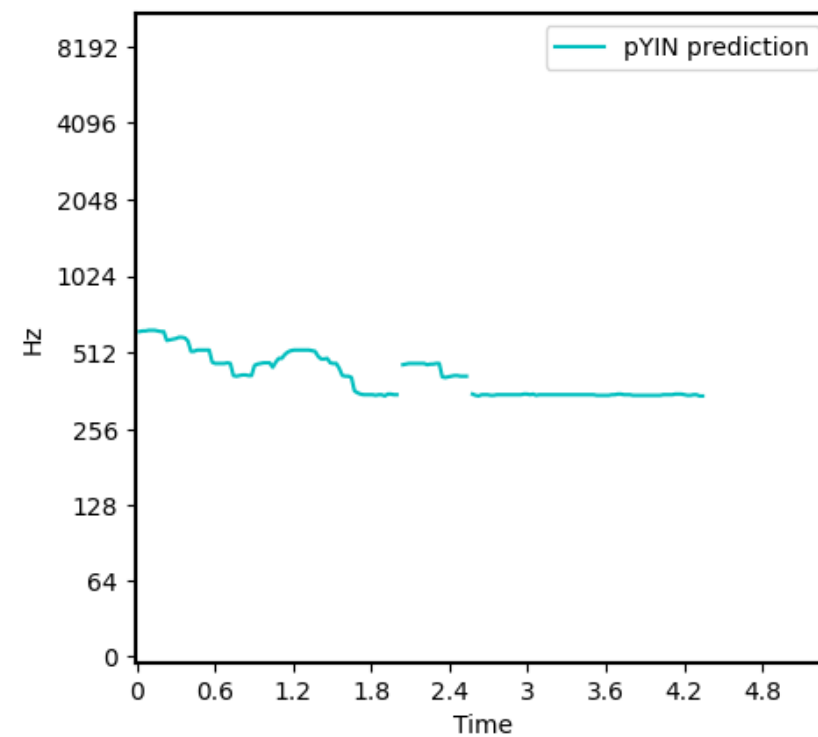
- **Definition:** The **lowest frequency component** of a waveform



Fundamental Frequency (F0) Estimation



→ **F0 Estimation** →



Auto-tune Artist



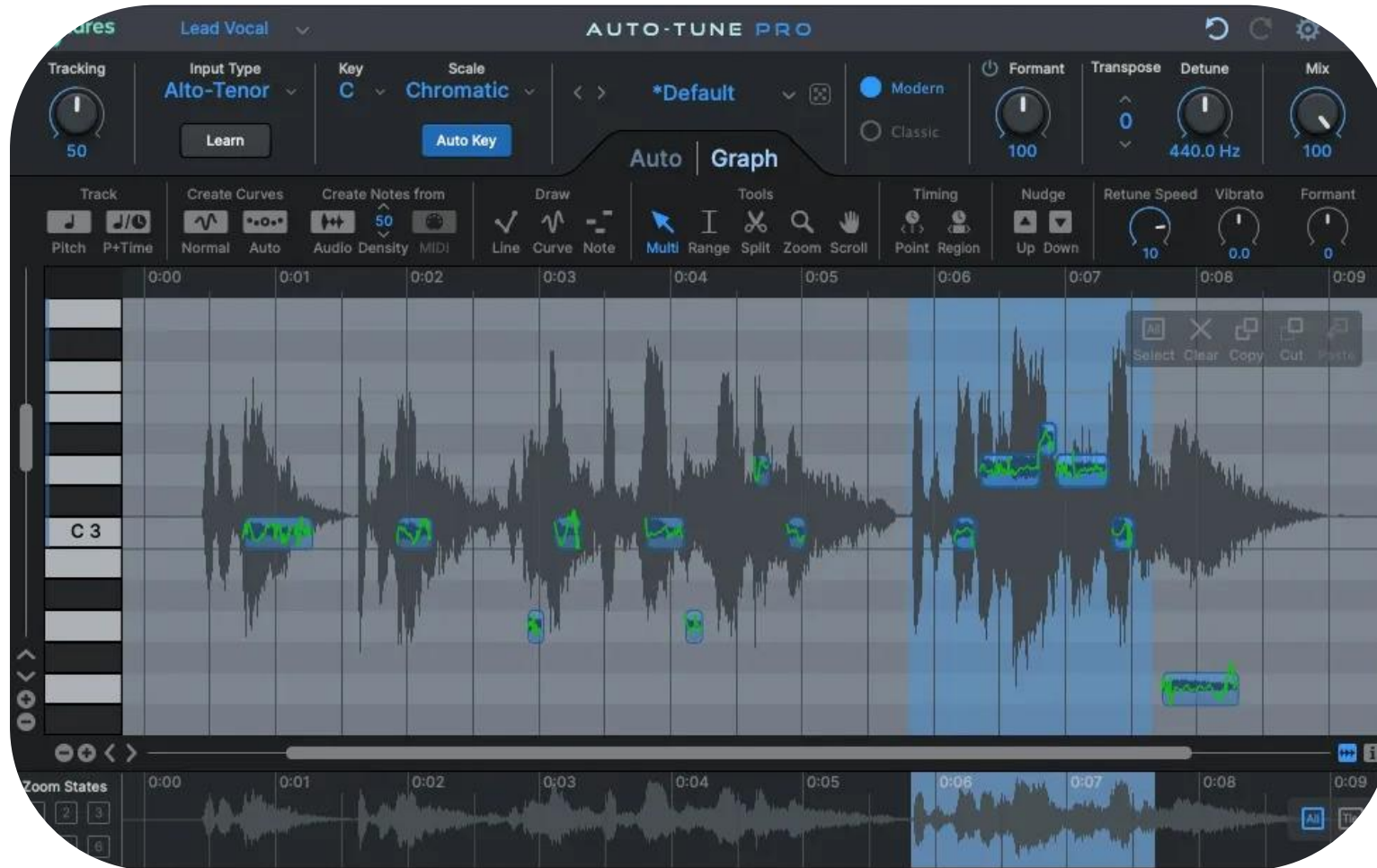
(Source: Antares Audio Technologies)

| Auto-tune



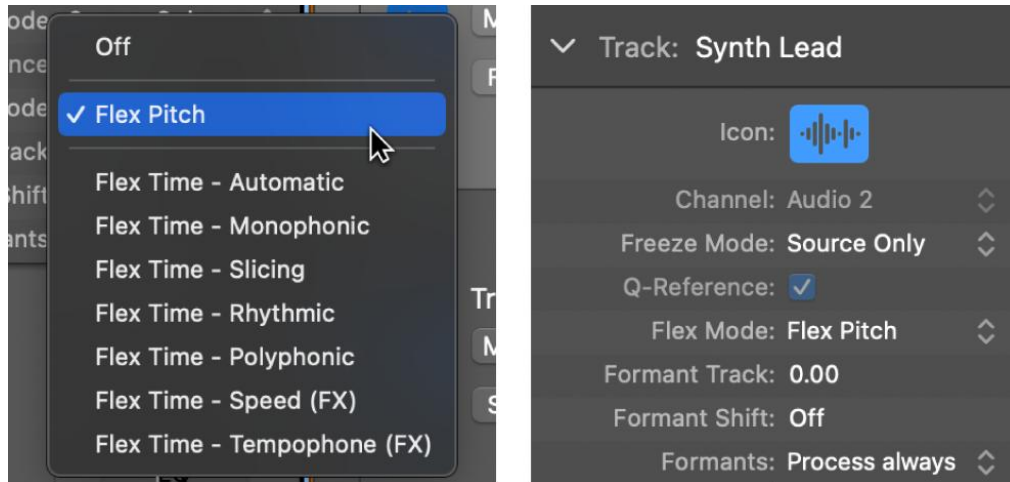
[youtube.com/shorts/
Kg8OSbKRETA](https://youtube.com/shorts/Kg8OSbKRETA)

Auto-tune Pro

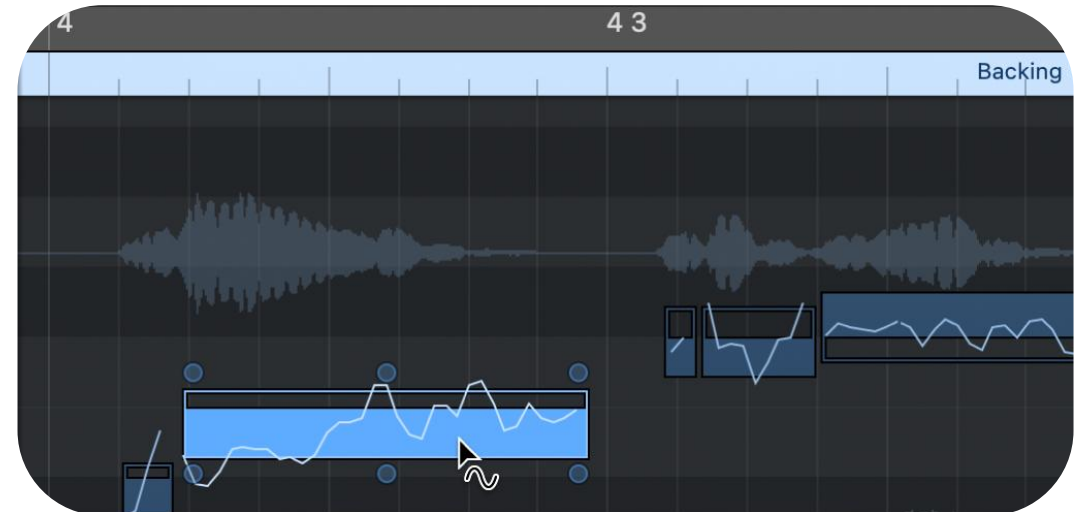


(Source: Antares Audio Technologies)

Pitch Correction in Logic Pro



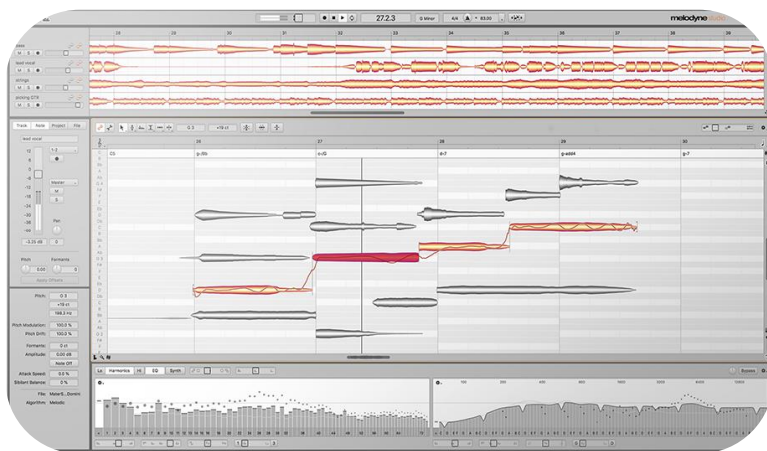
(Source: Logic Pro User Guide)



(Source: Logic Pro User Guide)

Other Auto-tune & Pitch Correction Tools

Melodyne



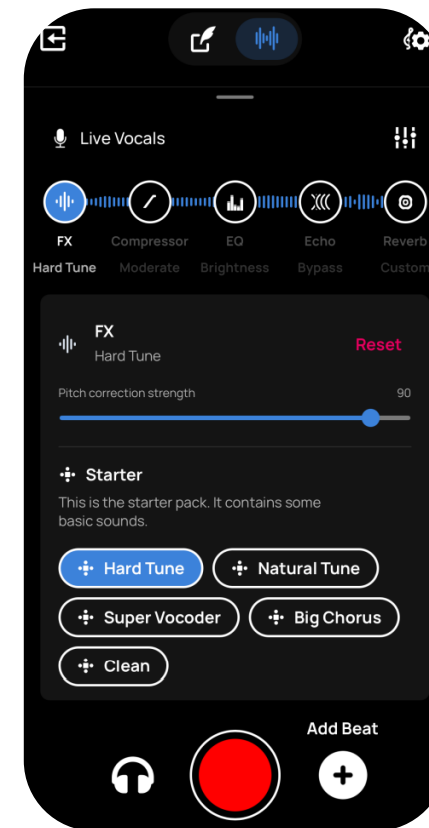
(Source: Celemony)

iZotope Nectar



(Source: iZotope Team)

Voloco



(Source: Voloco User Manual)

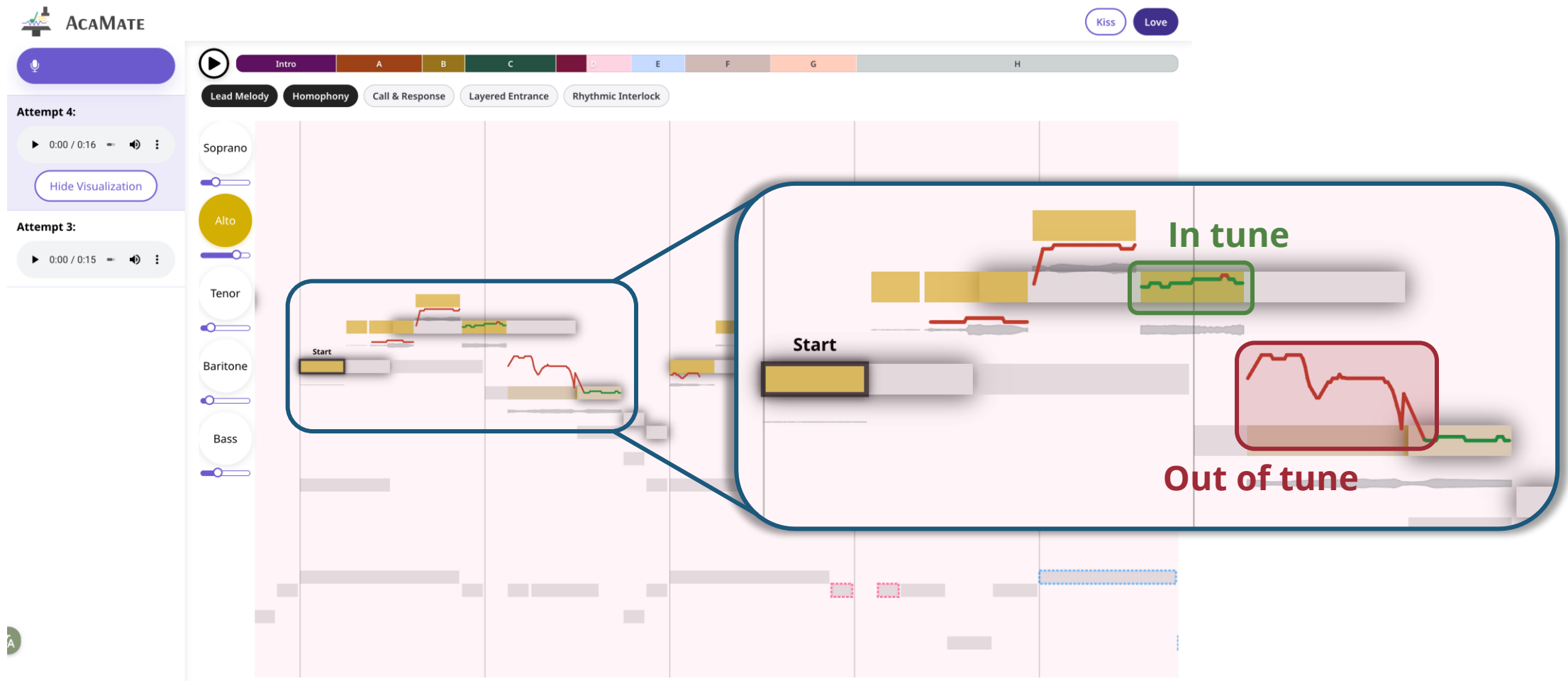
celemony.com/en/melodyne/what-is-melodyne
izotope.com/en/learn/why-upgrade-to-nectar-4.html
resonantcavity.com/assets/docs/voloco-user-manual.pdf

F0 Estimation Models

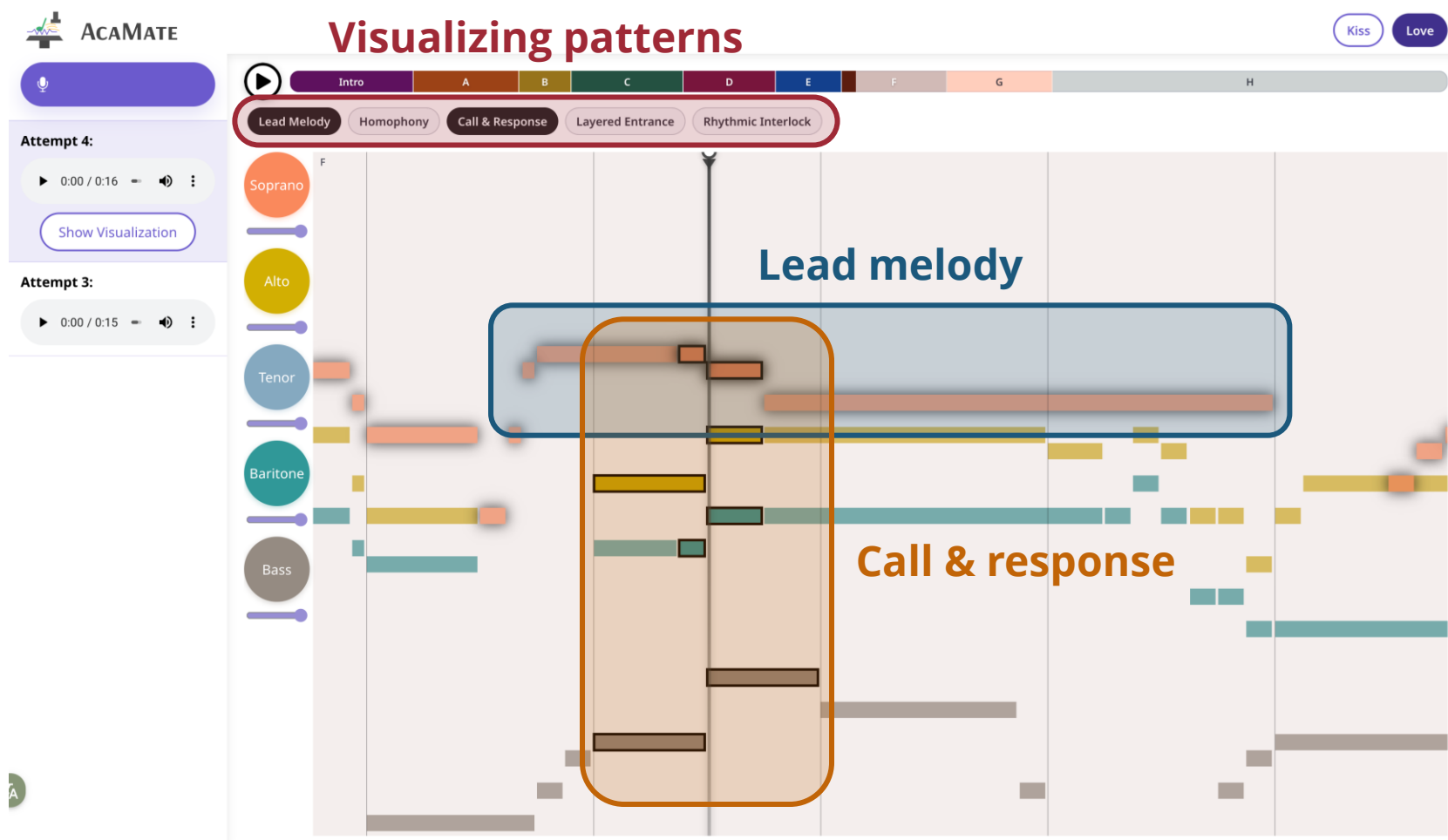
- **pYIN** (de Cheveigné et al., 2002)
 - `librosa.pyin`
- **CREPE** (Kim et al., 2018)
 - github.com/marl/crepe
- **PESTO** (Riou et al., 2023)
 - github.com/SonyCSLParis/pesto

Alain de Cheveigné and Hideki Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, 111(4):1917-1930, 2002.
Matthias Mauch and Simon Dixon, "pYIN: A fundamental frequency estimator using probabilistic threshold distributions," *ICASSP*, 2014.
Jong Wook Kim, Justin Salamon, Peter Li, and Juan Pablo Bello, "CREPE: A Convolutional Representation for Pitch Estimation," *ICASSP*, 2018.
Alain Riou, Stefan Lattner, Gaëtan Hadjeres, and Geoffroy Peeters, "PESTO: Pitch Estimation with Self-supervised Transposition-equivariant Objective," *ISMIR*, 2023.

AcaMate: Vocal Coach for A Cappella Singers (Ju et al., 2025)

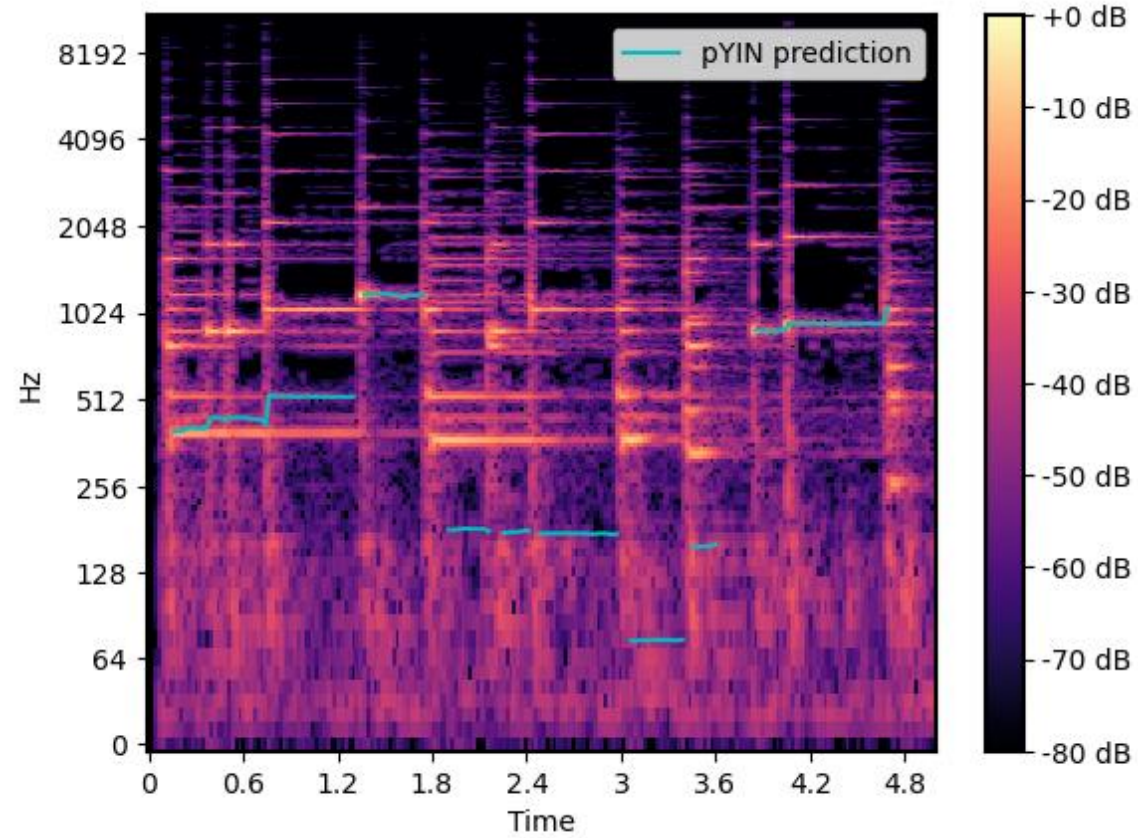


AcaMate: Vocal Coach for A Cappella Singers (Ju et al., 2025)

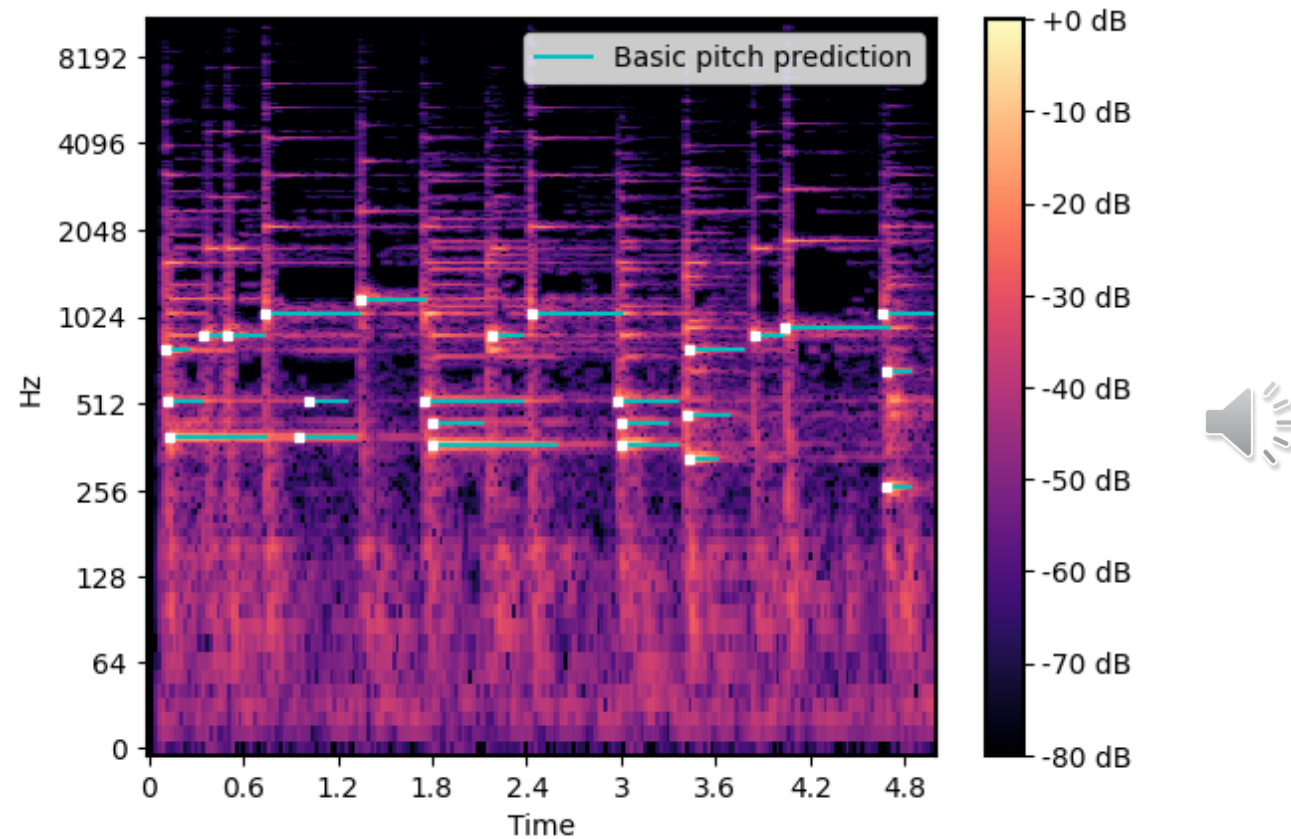


Music Transcription

Polyphonic F0 Estimation



Polyphonic F0 Estimation



basicpitch.spotify.com

Polyphonic F0 Estimation Models

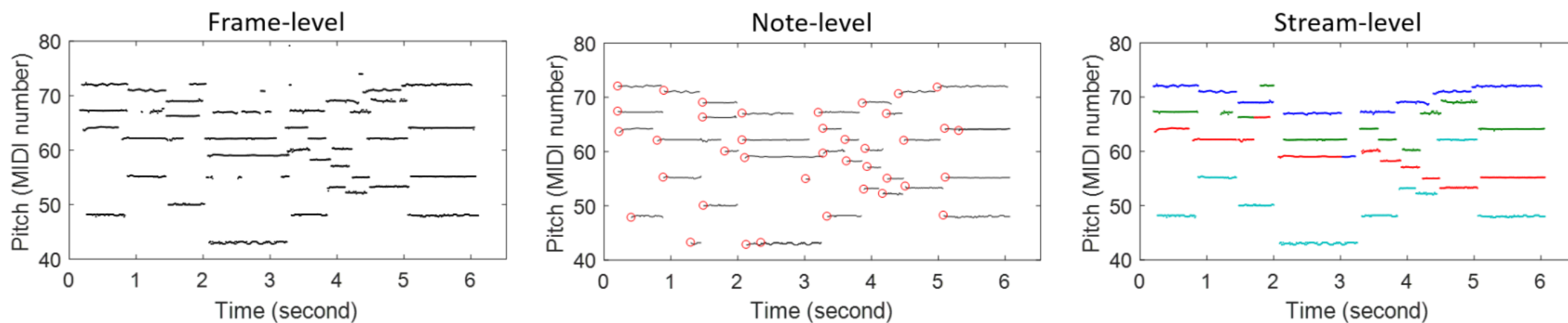
- **Deep Salience** (Bittner et al., 2017)
 - github.com/rabitt/ismir2017-deepsalience
- **Onset and Frames** (Hawthorne et al., 2018)
 - Piano only
 - github.com/jongwook/onsets-and-frames
- **Basic Pitch** (Bittner et al., 2022)
 - github.com/spotify/basic-pitch
 - basicpitch.spotify.com

Rachel M. Bittner, Brian McFee, Justin Salamon, Peter Li, and Juan P. Bello, "Deep Salience Representations for F0 Estimation in Polyphonic Music," *ISMIR*, 2017.

Curtis Hawthorne, Erich Elsen, Jialin Song, Adam Roberts, Ian Simon, Colin Raffel, Jesse Engel, Sageev Oore, and Douglas Eck, "Onsets and Frames: Dual-Objective Piano Transcription," *ISMIR*, 2018.

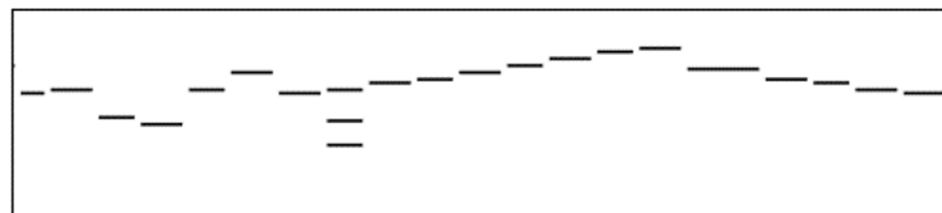
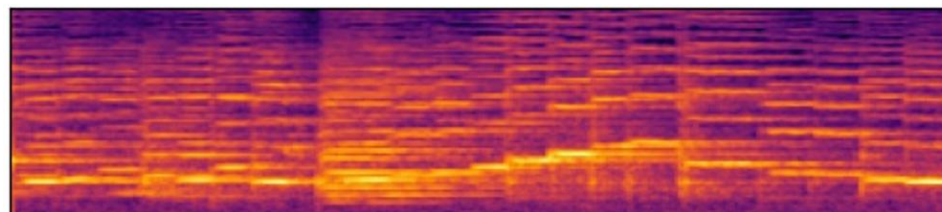
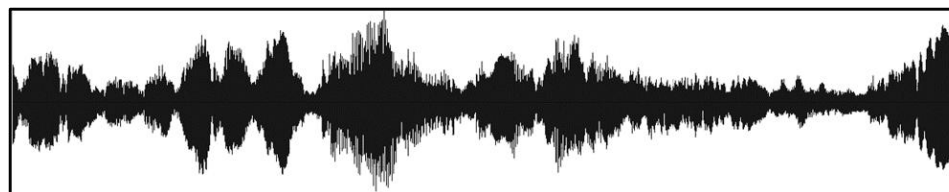
Rachel M. Bittner, Juan José Bosch, David Rubinstein, Gabriel Meseguer-Brocal, and Sebastian Ewert, "A Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and Multipitch Estimation," *ICASSP*, 2022.

F0 Estimation vs. Music Transcription



(Source: Benetos et al., 2019)

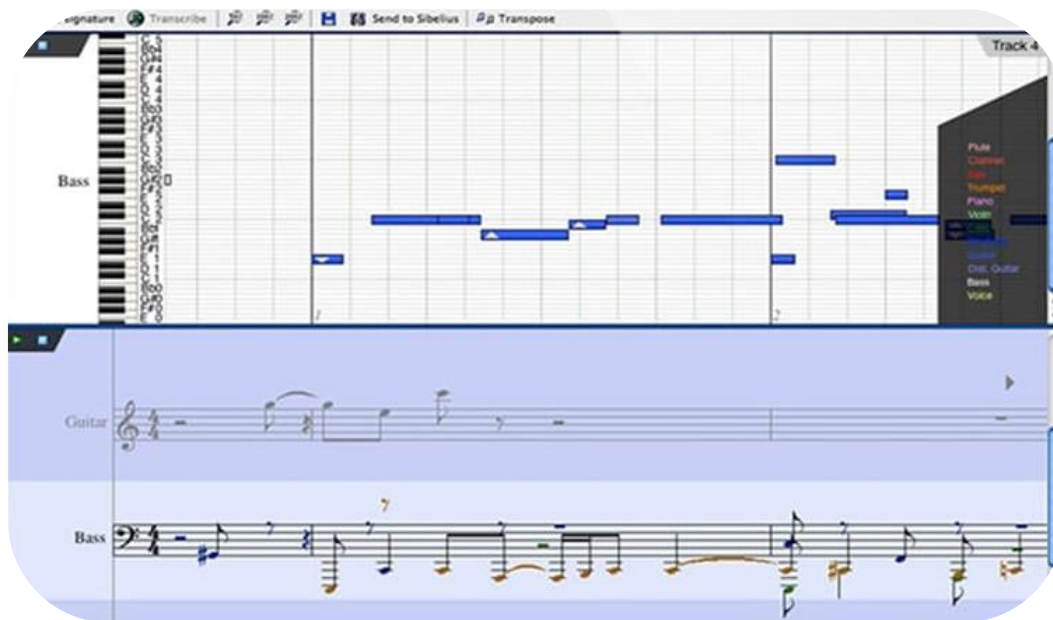
Music Transcription



(Source: Dong et al., 2022)

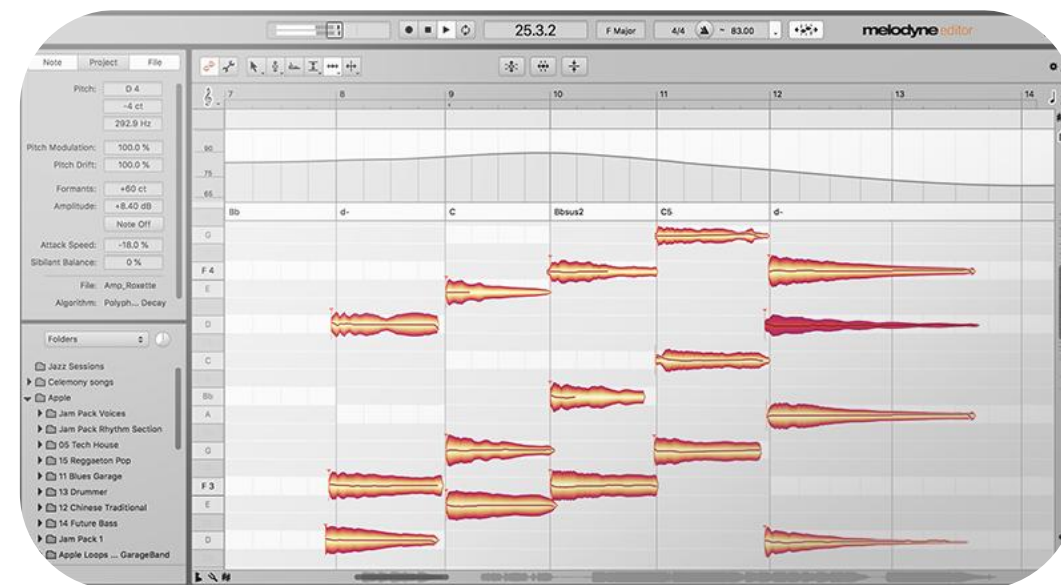
Commercial Music Transcription Software

AudioScore in Sibelius



(Source: Avid)

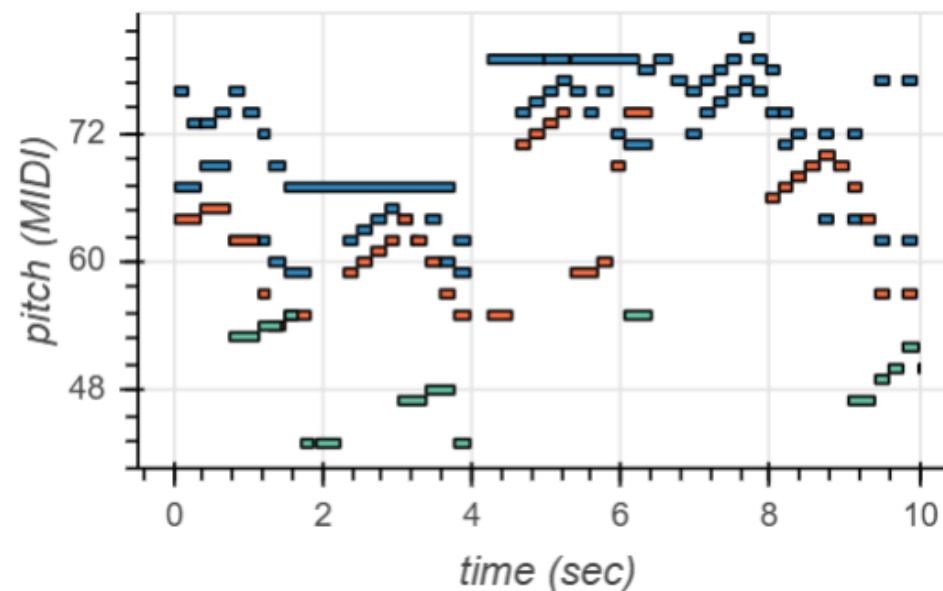
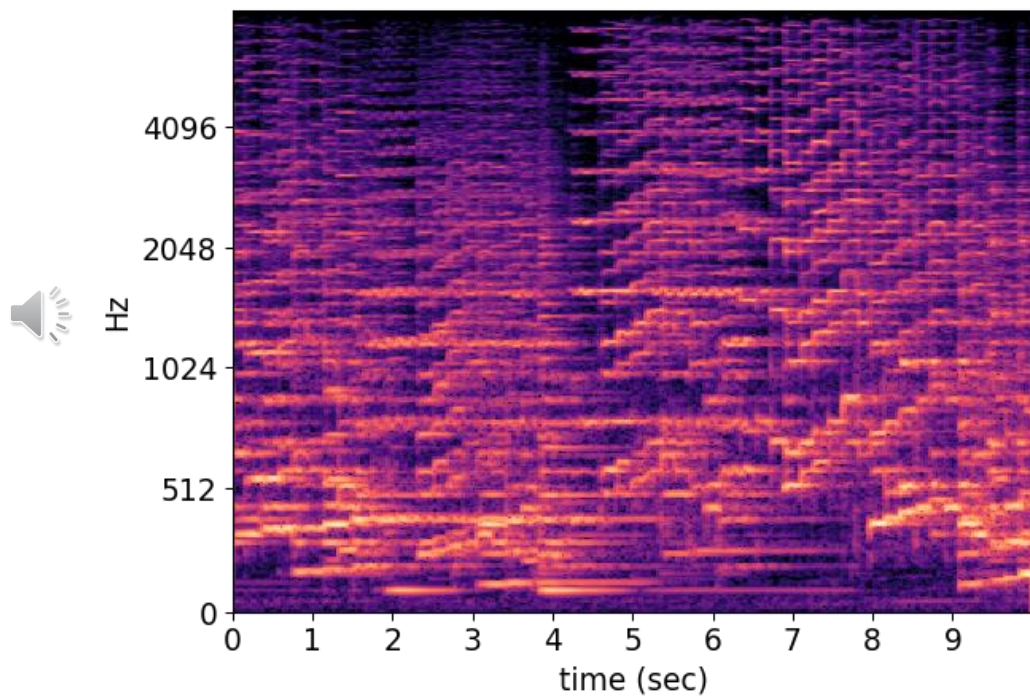
Melodyne Editor



(Source: Celemony)

Multitrack Transcription Models

- **MT3** (Gardner et al., 2022)
 - github.com/magenta/mt3

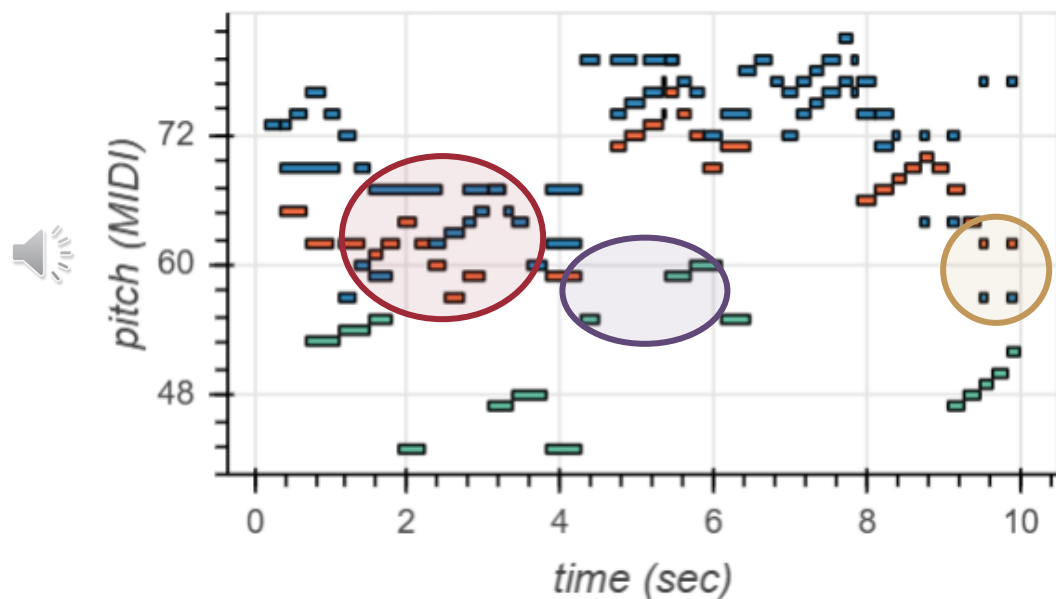


(Source: Gardner et al., 2022)

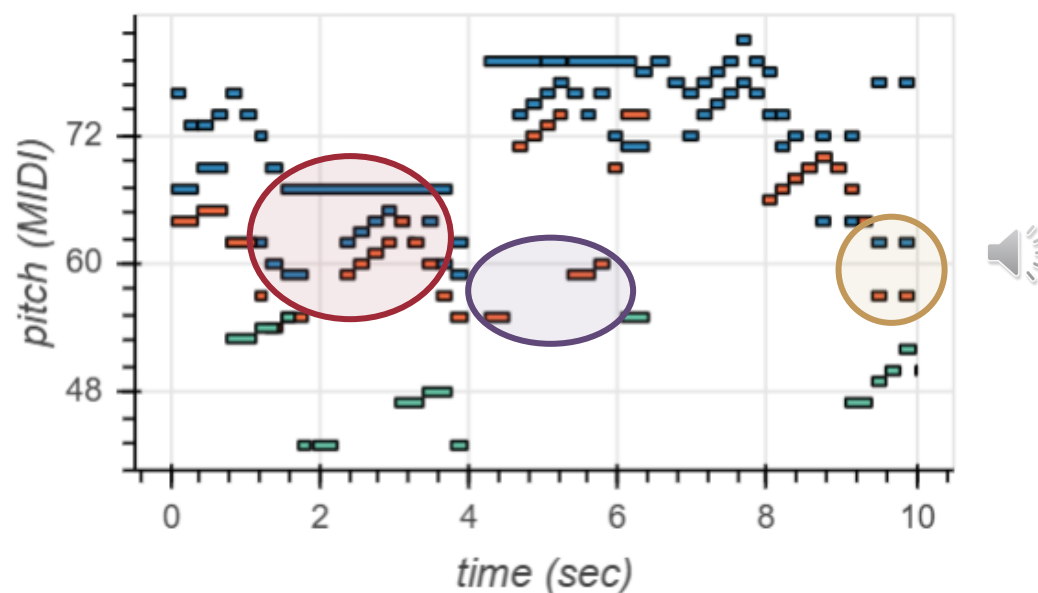
Multitrack Transcription Models

- **MT3** (Gardner et al., 2022)
 - github.com/magenta/mt3

Ground truth



MT3 prediction



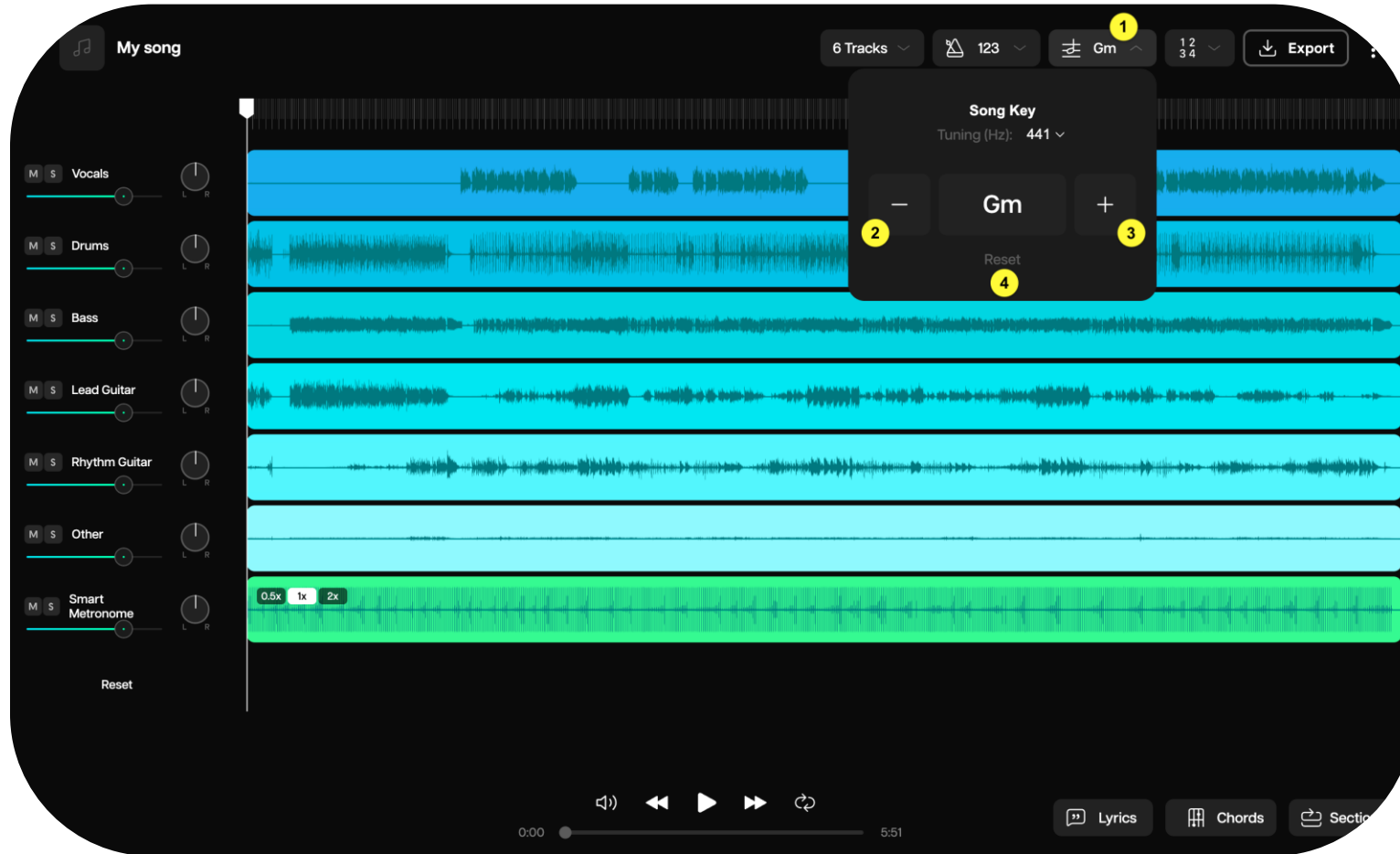
(Source: Gardner et al., 2022)

Resources

- Rachel Bittner, Mark Cartwright, and Ethan Manilow, “Programming MIR Baselines from Scratch: Three Case Studies,” *Tutorials of ISMIR*, 2021.
 - Part 1: [Transcription with NMF](#) (Ethan Manilow)
 - Part 2: [Pitch Tracking with pytorch](#) (Rachel Bittner)
 - Part 3: [Instrument Classification with OpenL3 & Tensorflow](#) (Mark Cartwright)
- Rachel Bittner, Alain de Cheveigné, and Johana Devaney, “Fundamental Frequency Estimation in Music,” *Tutorials of ISMIR*, 2018.
 - Part 1: [Pitch](#) (Alain de Cheveigné)
 - Part 2: [Polyphonic fundamental frequency estimation](#) (Rachel Bittner)
 - Part 3: [Applications](#) (Johana Devaney)

Harmony & Rhythm Analysis

Key Detection in Moises



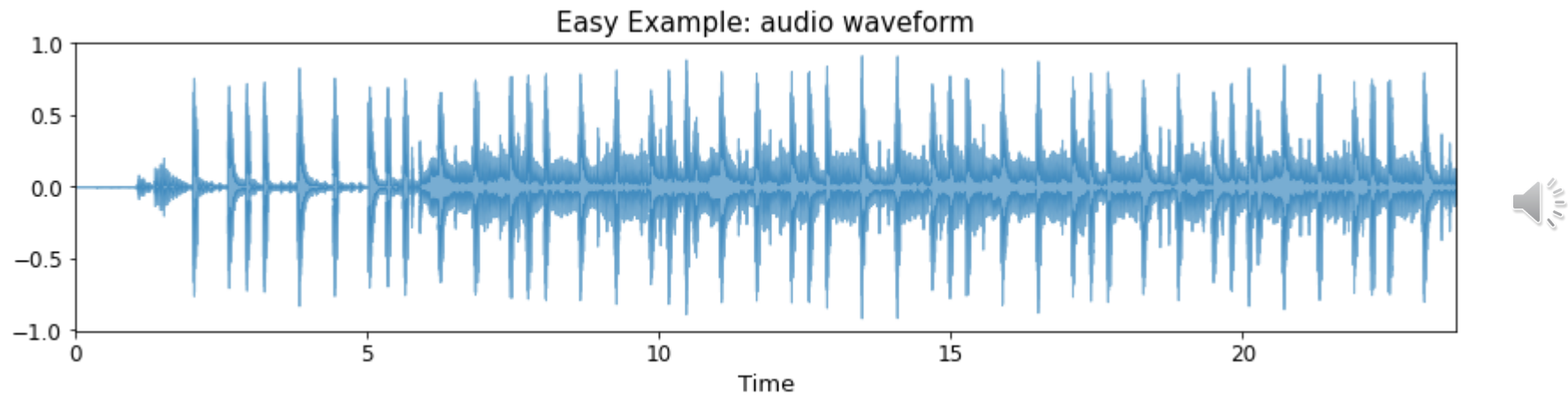
(Source: Moises)

| Chord Detection



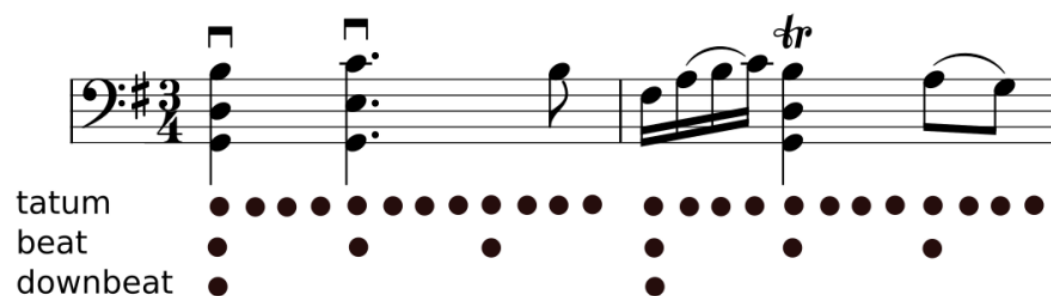
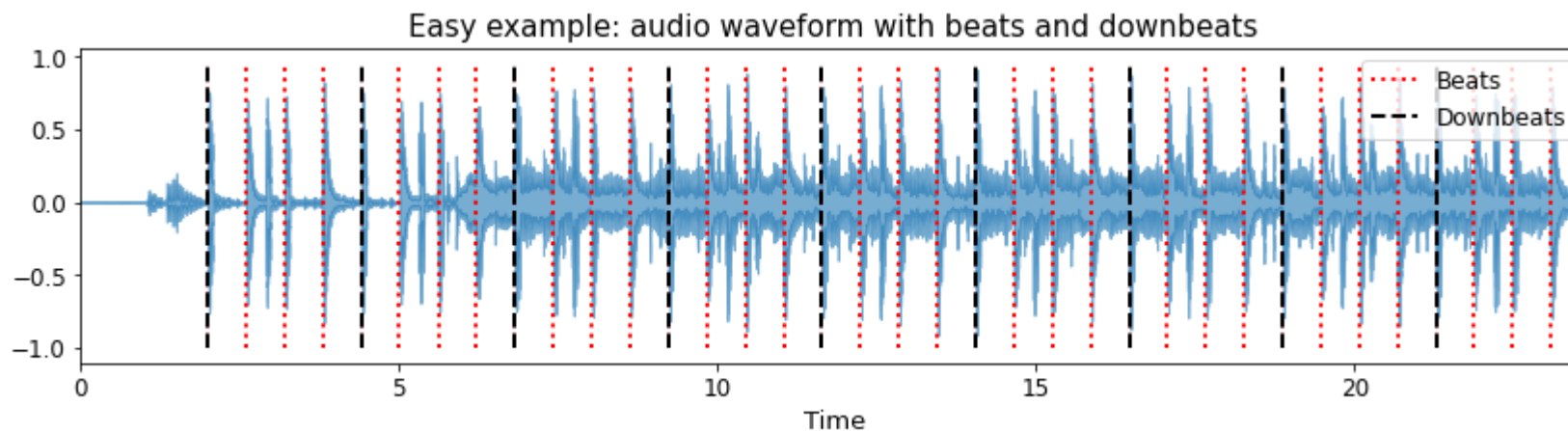
youtube.com/shorts/_N3b_GARMfA

Beat & Downbeat Estimation



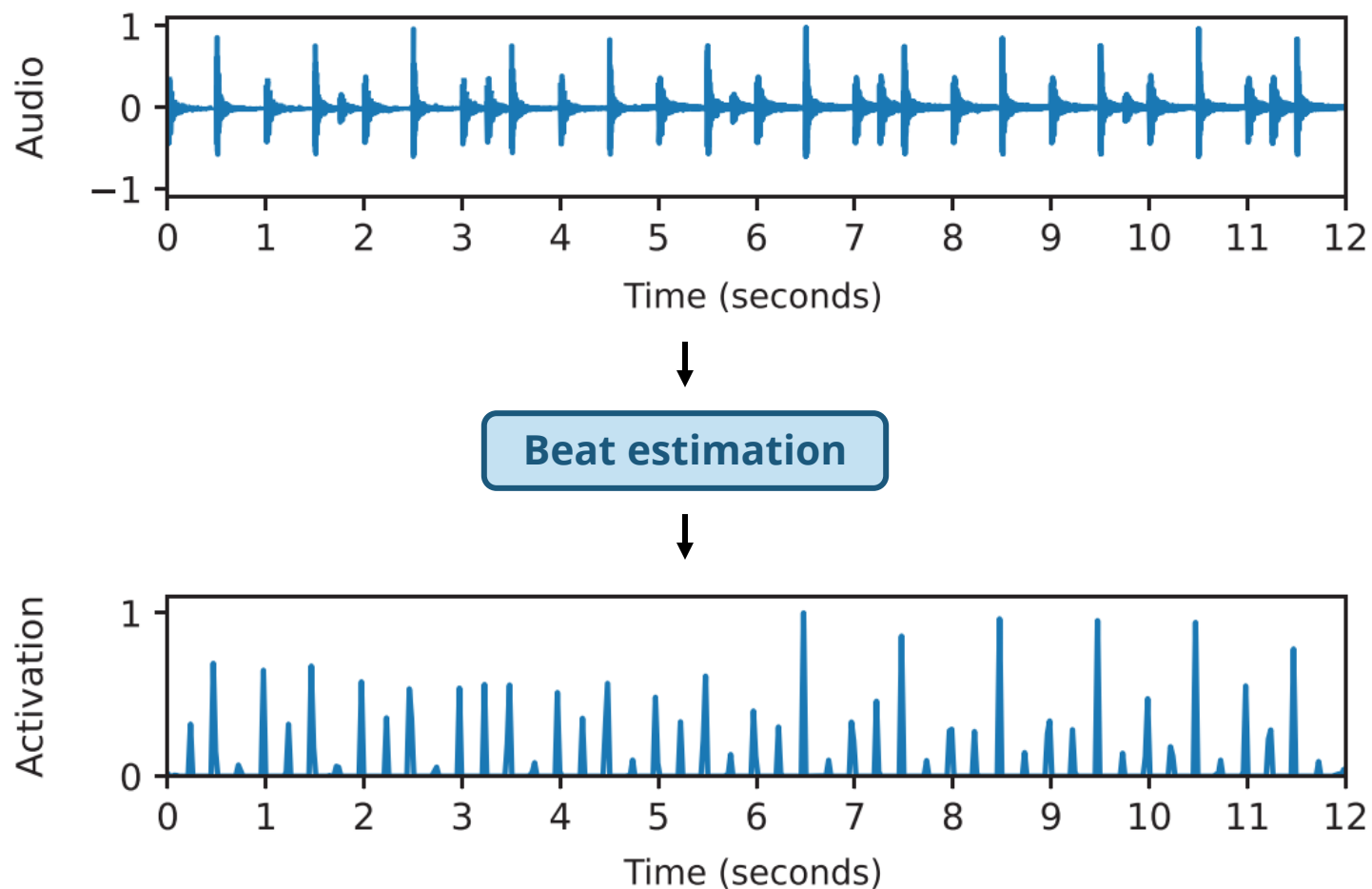
(Source: Davies et al., 2021)

Beat & Downbeat Estimation



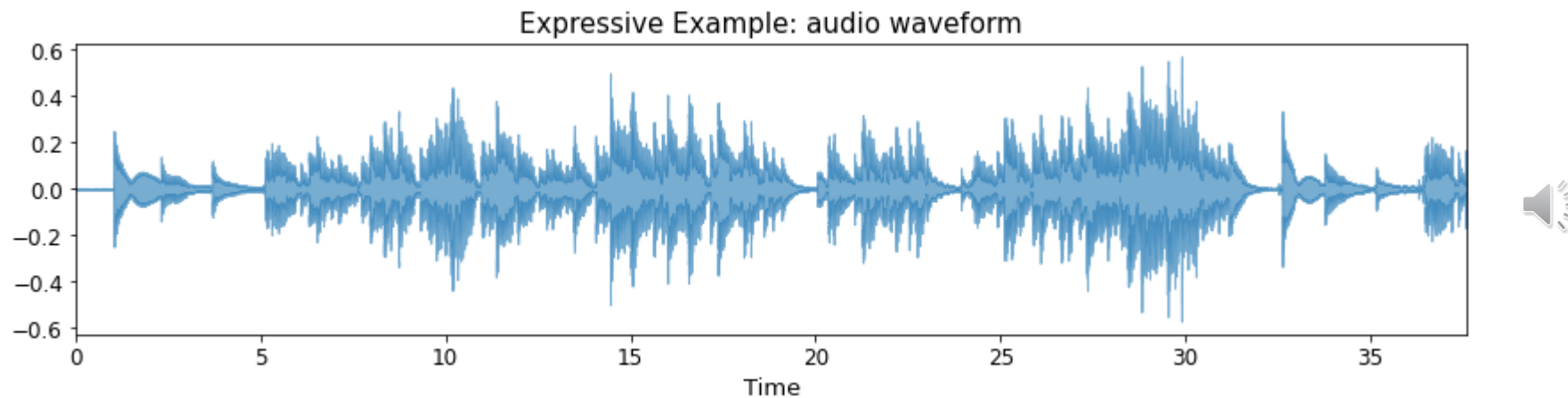
(Source: Davies et al., 2021)

Beat & Downbeat Estimation



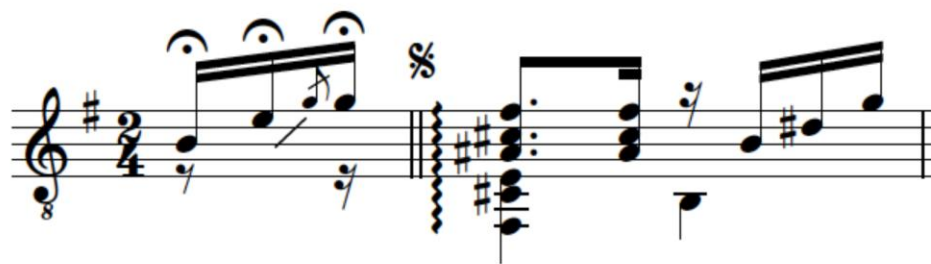
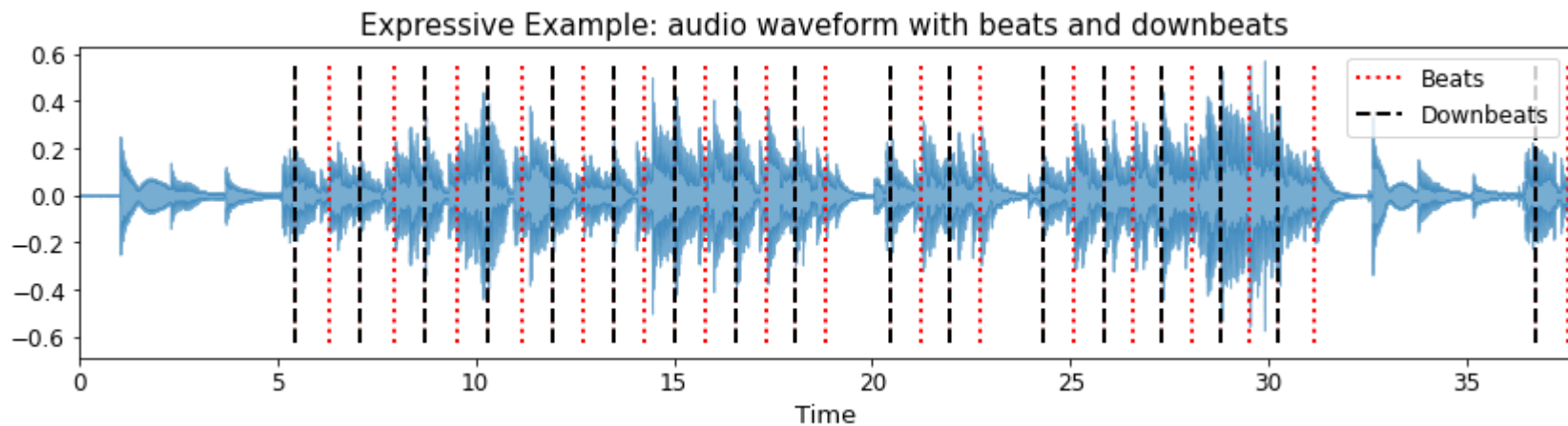
(Source: Meier et al., 2024)

Beat & Downbeat Estimation



(Source: Davies et al., 2021)

Beat & Downbeat Estimation



(Source: Davies et al., 2021)

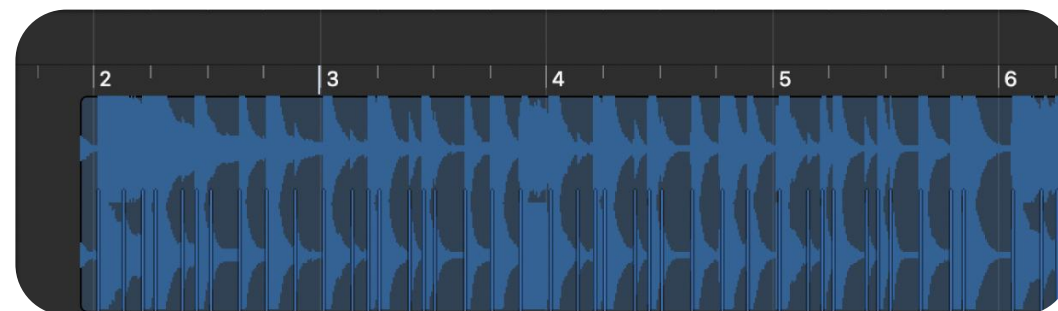
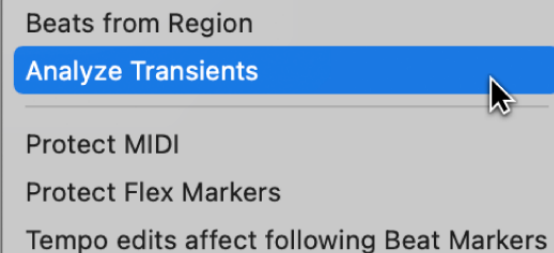
Beat Tracking in Pro Tools & Logic Pro

Beat Detective in Pro Tools



(Source: Logic Pro User Guide)

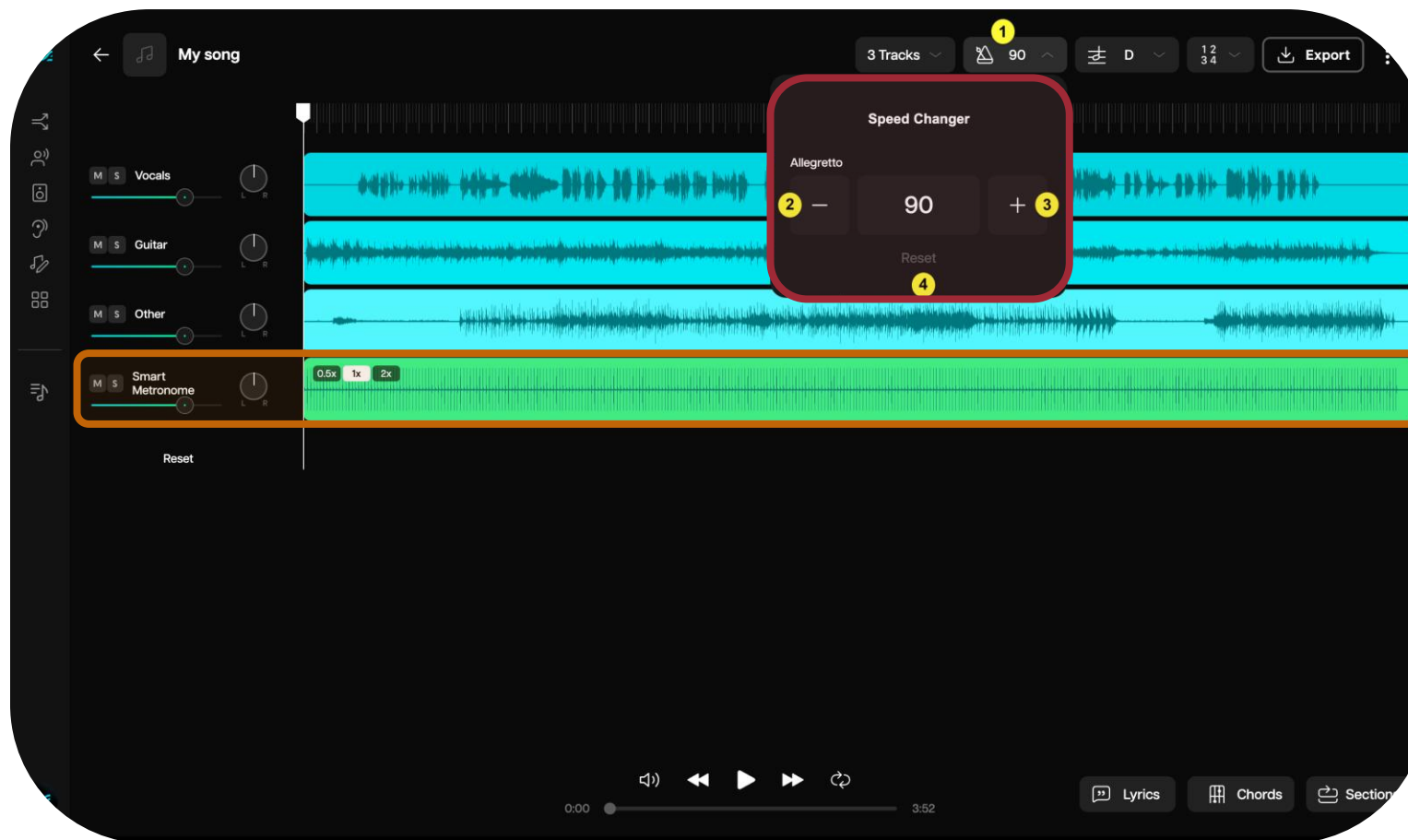
Beat Mapping in Logic Pro



(Source: Logic Pro User Guide)

Tempo Estimation & Beat Tracking in Moises

Tempo estimation



Beat tracking

(Source: Moises)

Resources on Rhythm Analysis

- Matthew E. P. Davies, Sebastian Böck, and Magdalena Fuentes, "[Tempo, Beat and Downbeat Estimation](#)," *Tutorials of ISMIR*, 2021.
- Hendrik Schreiber, Julián Urbano, and Meinard Müller, "[Music Tempo Estimation: Are We Done Yet?](#)," *TISMIR*, 3(1):111-125, 2020.

Structure Analysis

Structure Analysis

Music segmentation

The musical score is divided into sections labeled A1 through D. The sections are color-coded: A1 (pink), A2 (pink), B1 (green), B2 (green), C (light blue), A3 (pink), B3 (green), B4 (green), and D (yellow). The score includes various musical notations such as dynamics (f, ff, p, sf), tempo markings (Allegro, Vivace, in tempo, poco rit.), and articulation (accents, slurs).

Figure 4.5 following [Müller, FMP, Springer 2015]

(Source: Müller & Zalkow, 2019)

Hierarchical music segmentation

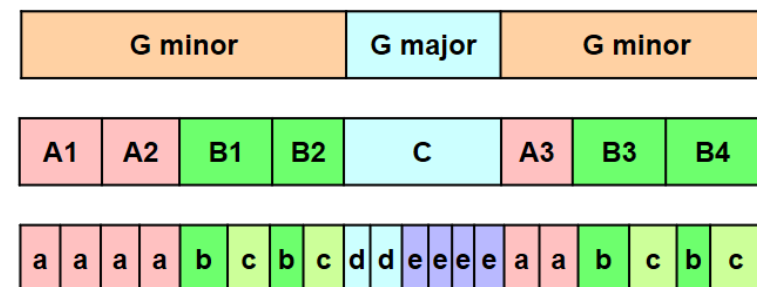
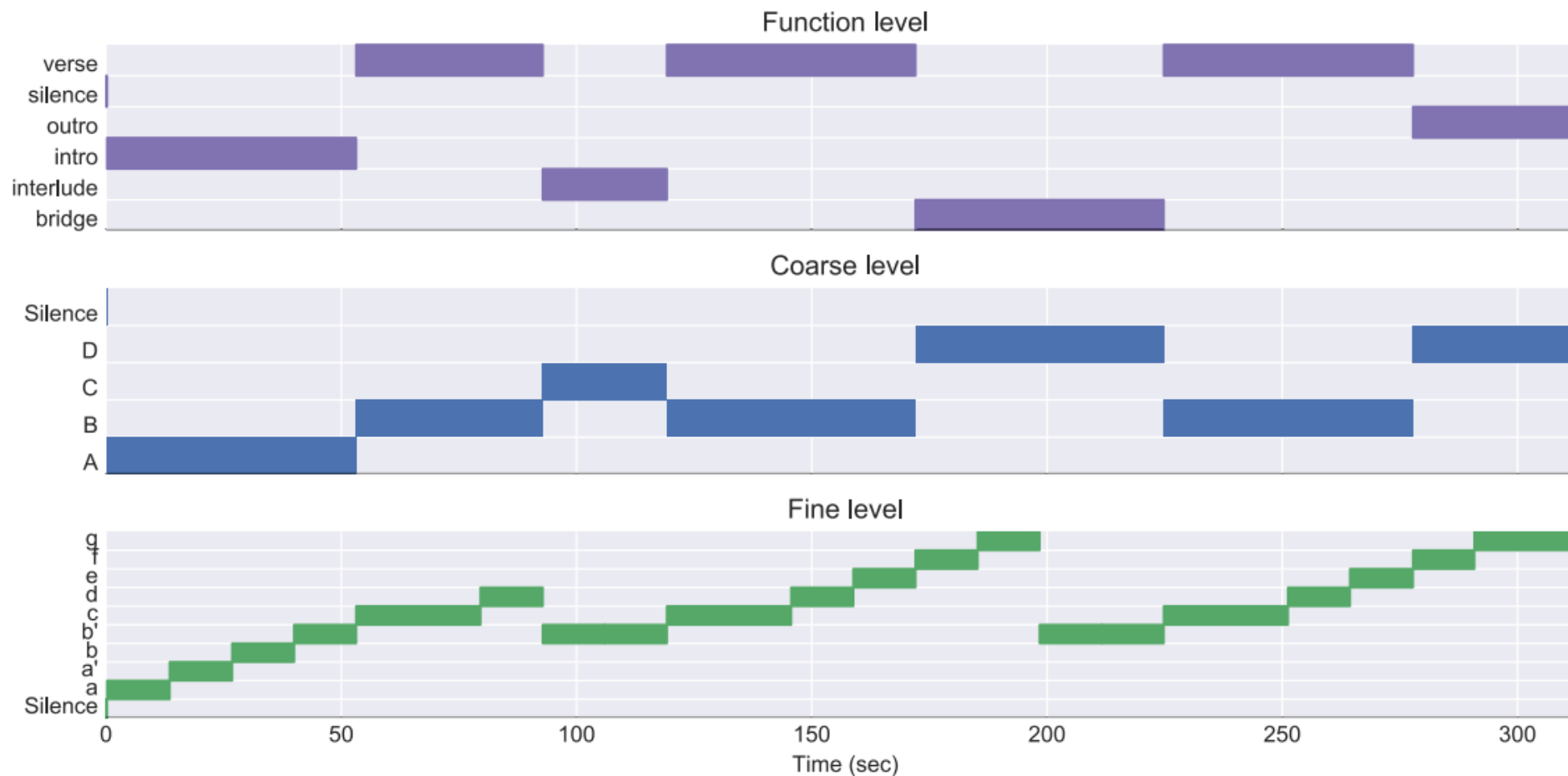


Figure 4.28 from [Müller, FMP, Springer 2015]

(Source: Müller & Zalkow, 2019)

Hierarchical Music Segmentation



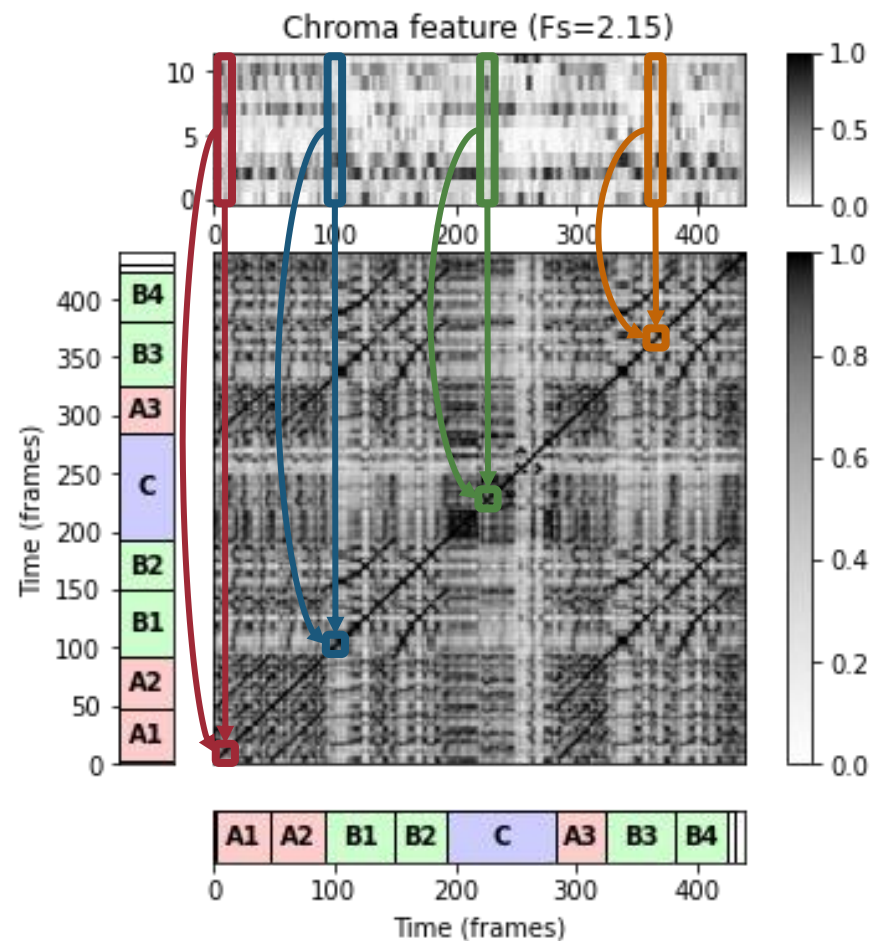
(Source: Müller & Zalkow, 2019)

Oriol Nieto, Gautham J. Mysore, Cheng-i Wang, Jordan B. L. Smith, Jan Schlüter, Thomas Grill, and Brian McFee, "Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications," *TISMIR*, 3(1):246-263, 2020.



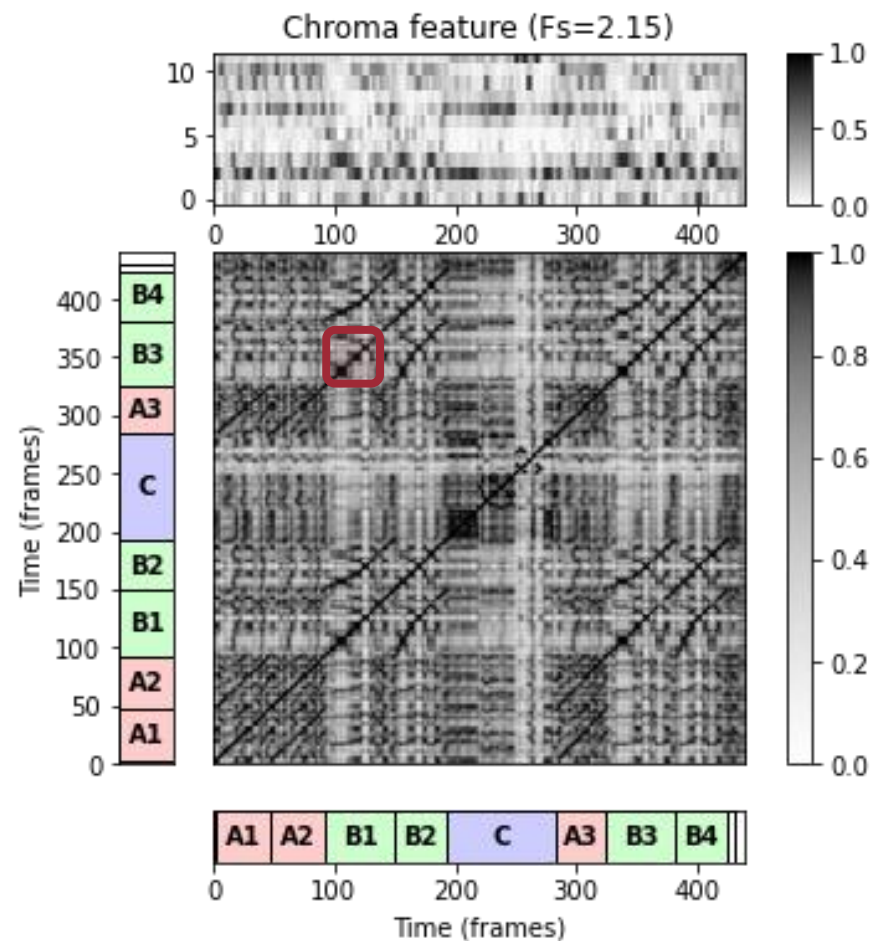
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



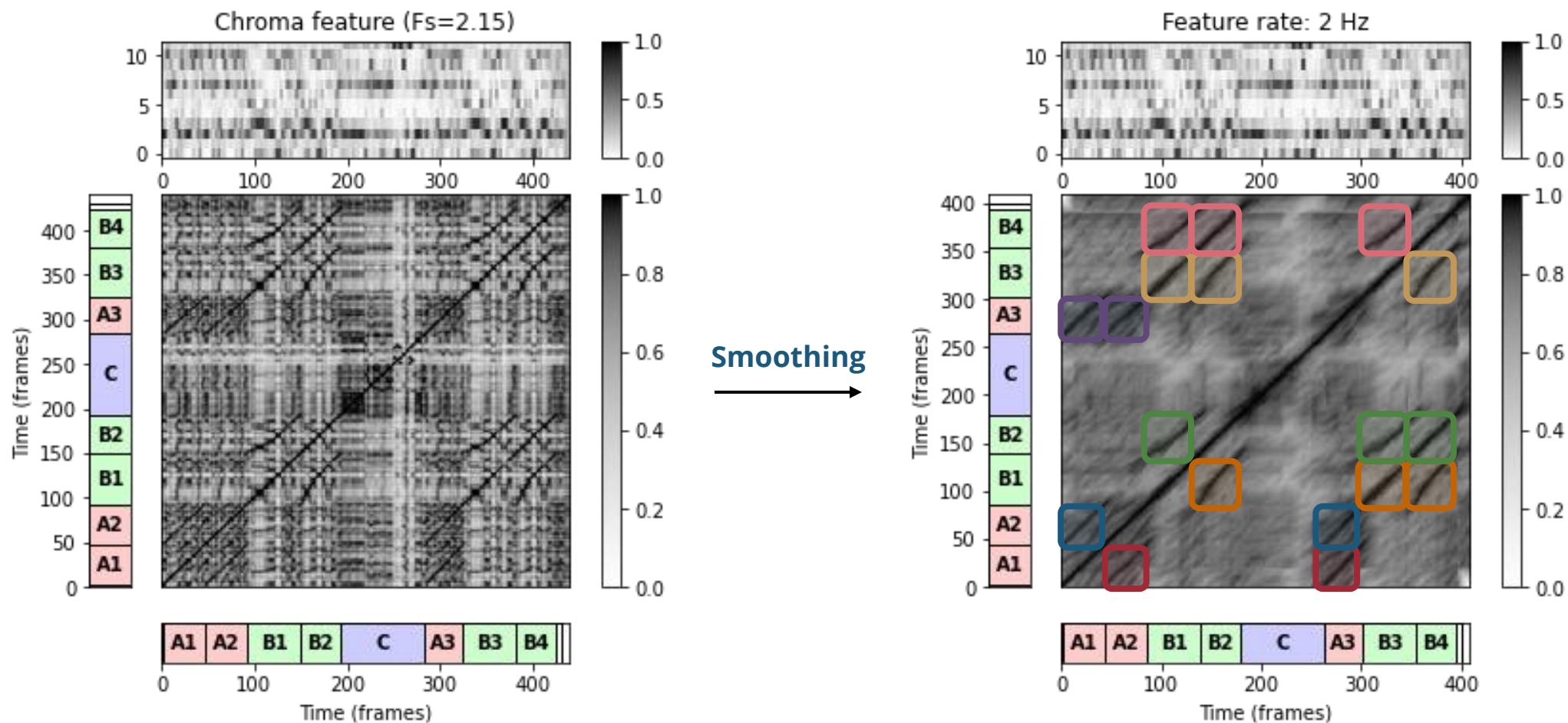
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



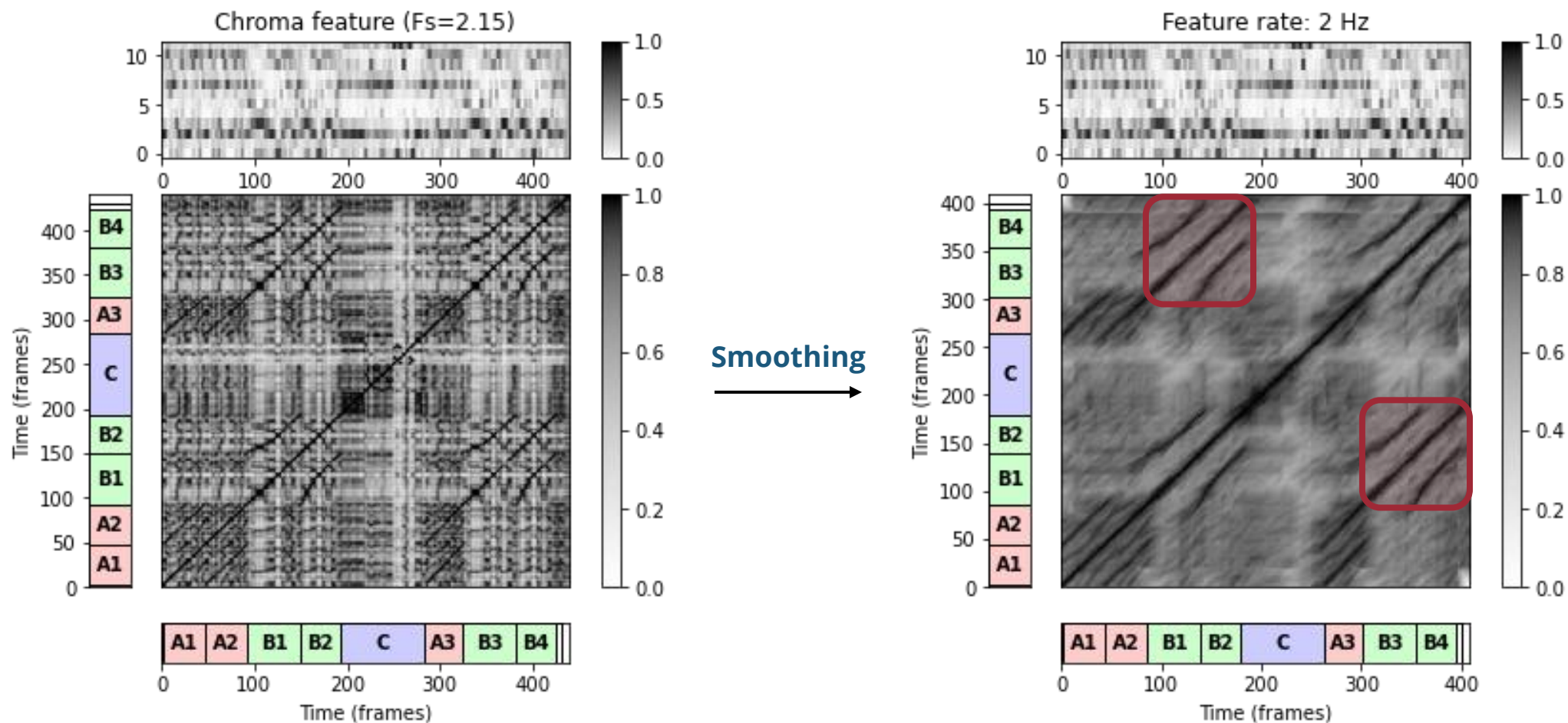
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



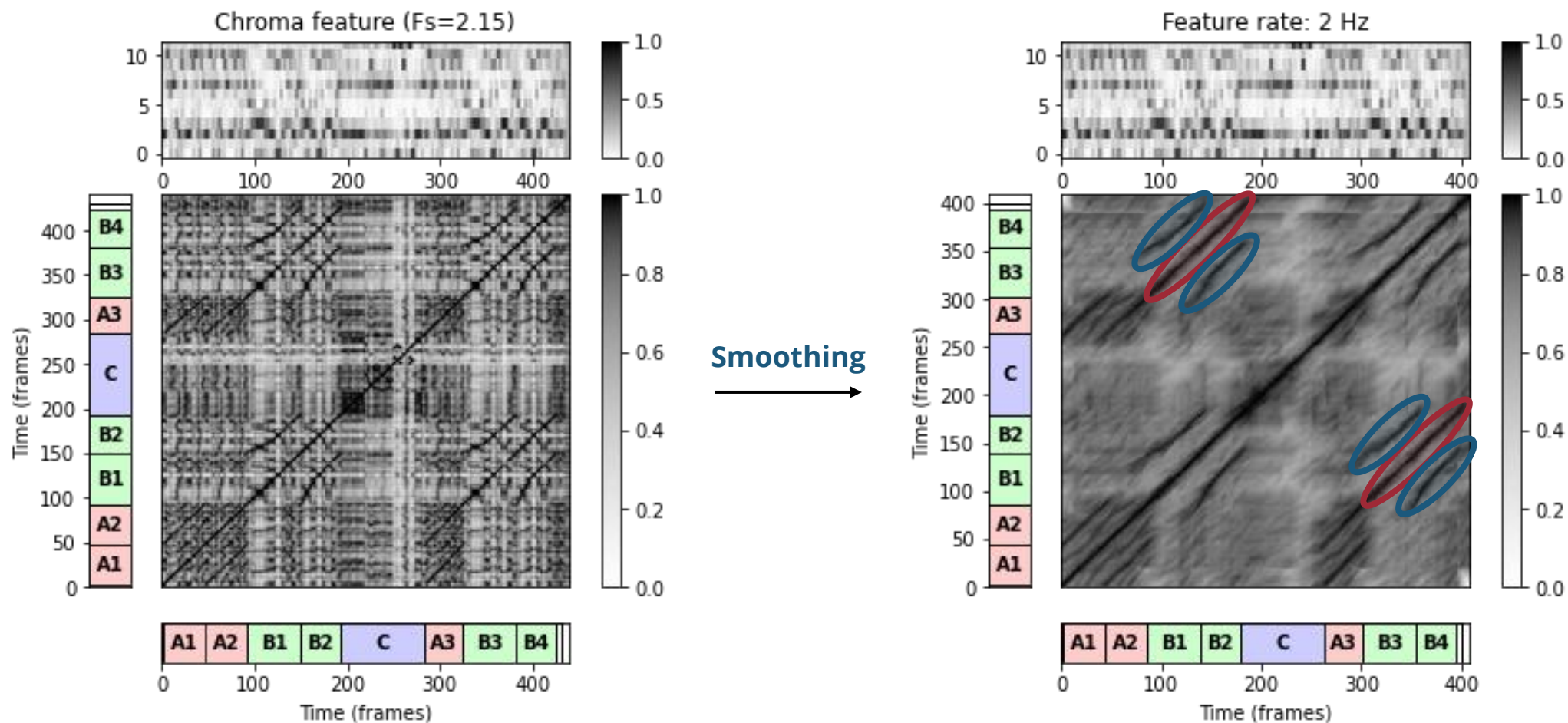
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



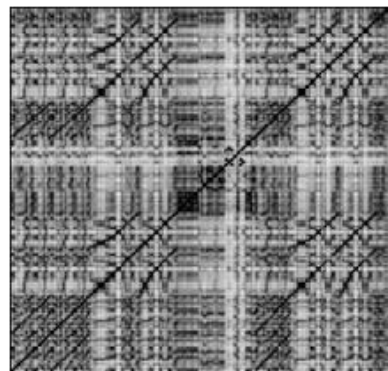
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)

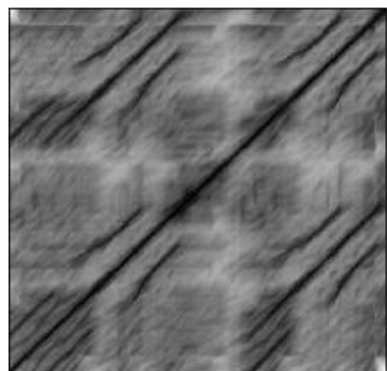


(Source: Müller & Zalkow, 2019)

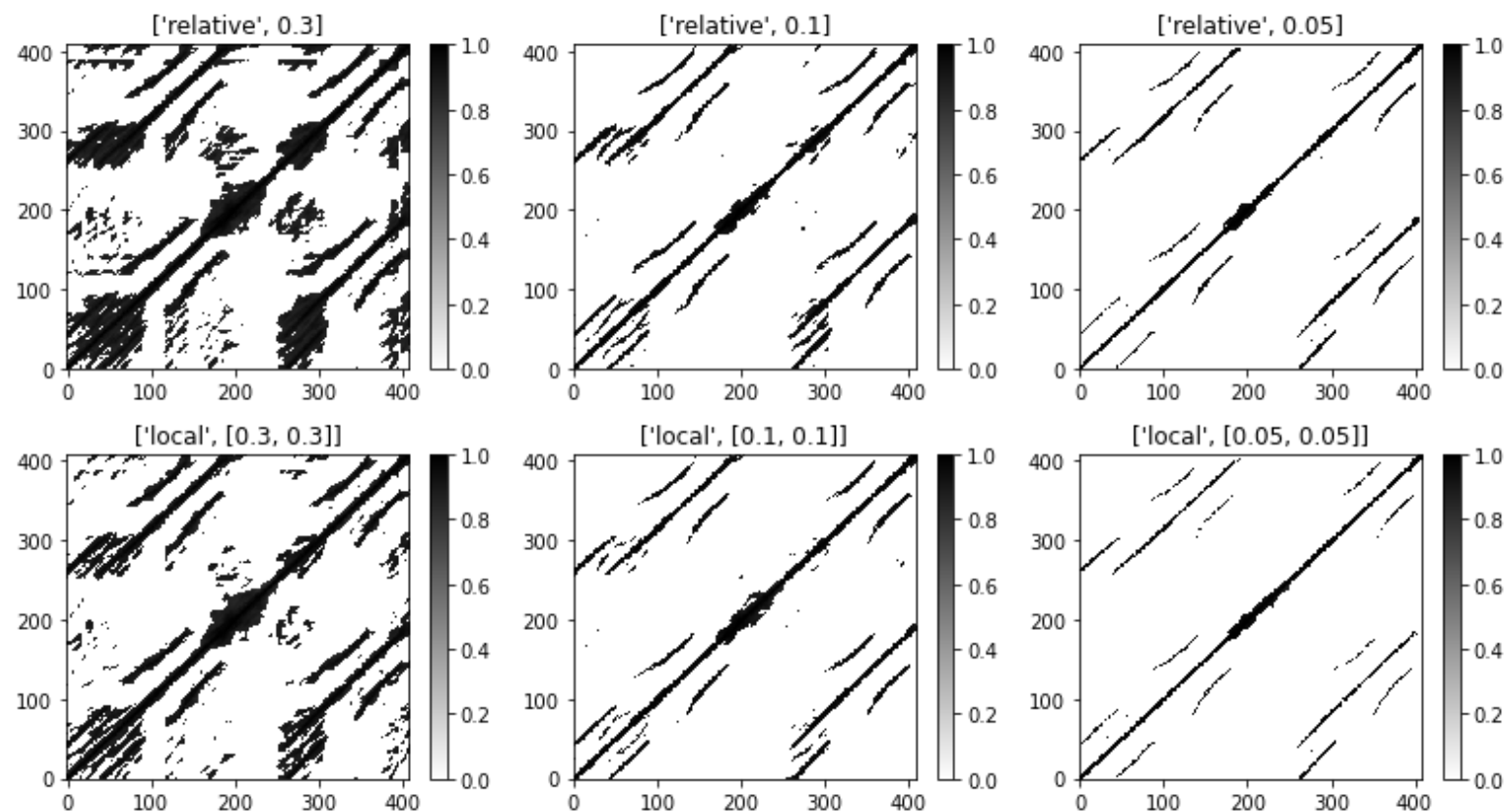
Self-Similarity Matrices (SSMs)



↓ Smoothing

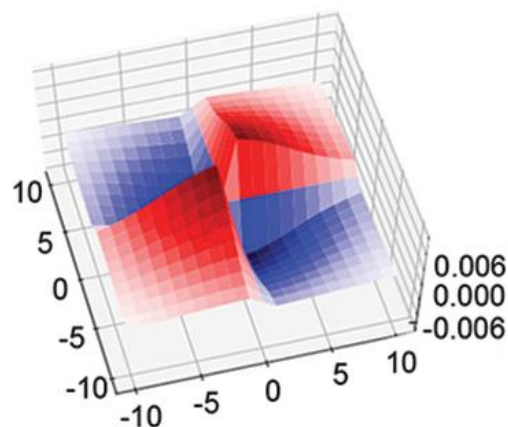
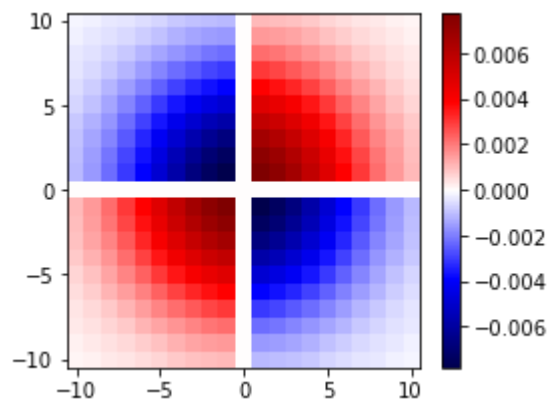


↗ Thresholding



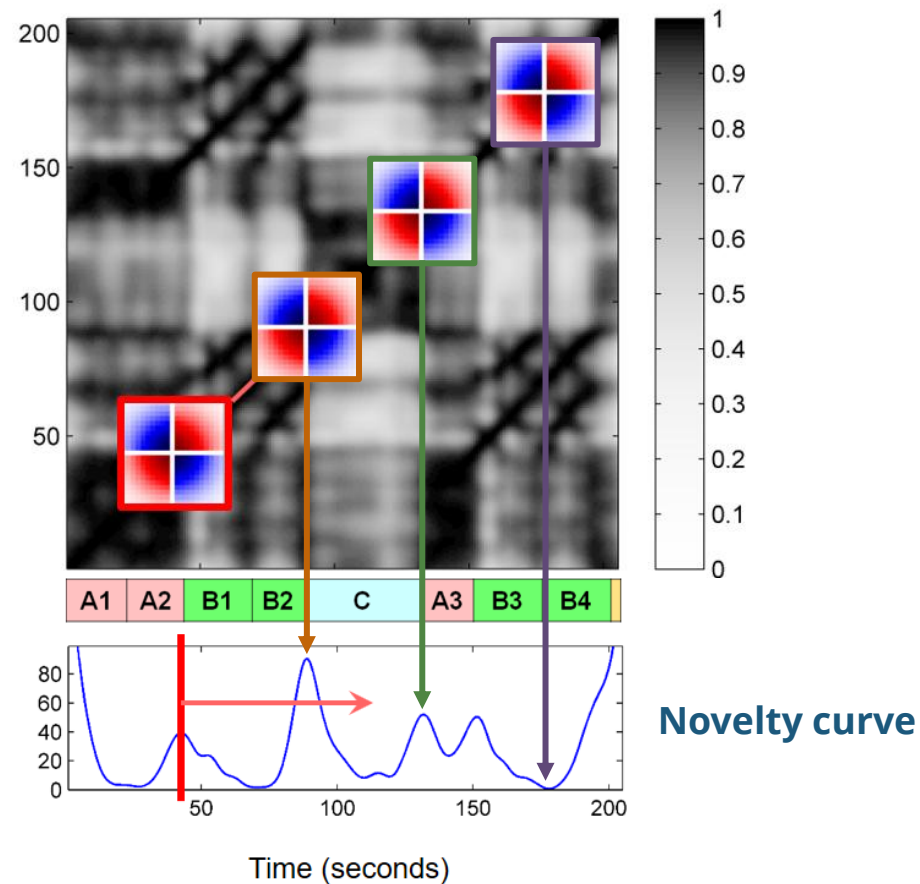
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



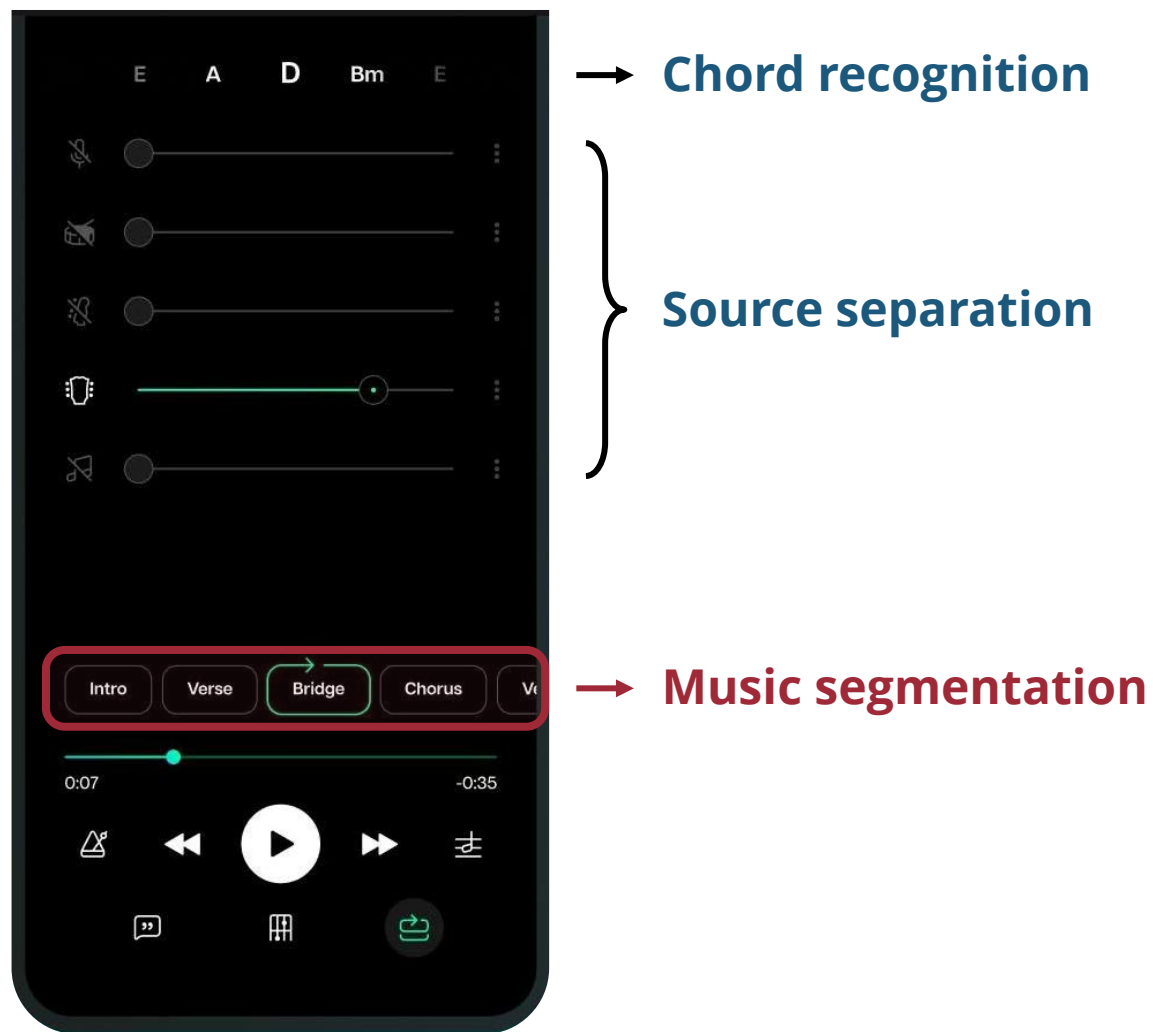
(Source: Müller & Chiu, 2024)

Figure 4.24 from [Müller, FMP, Springer 2015]



(Source: Müller & Zalkow, 2019)

Structure Analysis in Moises



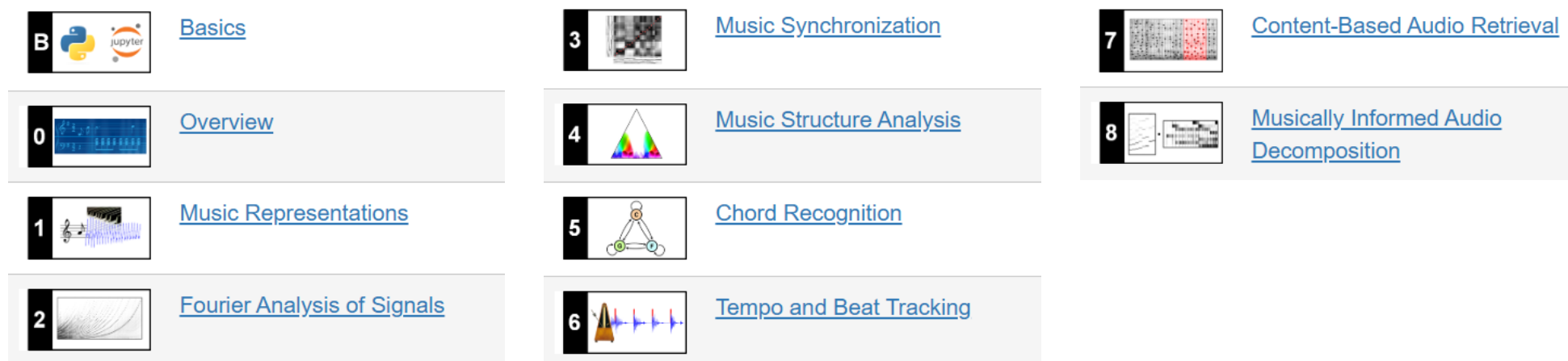
(Source: Moises)

Resources on Music Structure Analysis

- Meinard Müller and Ching-Yu Chiu, "[A Basic Tutorial on Novelty and Activation Functions for Music Signal Processing](#)," *TISMIR*, 7(1):179-194, 2024.
- Oriol Nieto, Gautham J. Mysore, Cheng-i Wang, Jordan B. L. Smith, Jan Schlüter, Thomas Grill, and Brian McFee, "[Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications](#)," *TISMIR*, 3(1):246-263, 2020.
- Meinard Müller & Jordan B. L. Smith, "Music Structure Analysis," *Tutorials of ISMIR*, 2014. ([part 1](#), [part 2](#), [part 3](#))

Resources on Music Information Research (MIR)

- Meinard Müller, “[Fundamentals of Music Processing – Using Python and Jupyter Notebooks](#),” *Springer*, 2021.
- Meinard Müller and Frank Zalkow, “[FMP Notebooks: Educational Material for Teaching and Learning Fundamentals of Music Processing](#),” *ISMIR*, 2019.
 - Jupyter notebooks available at audiolabs-erlangen.de/FMP



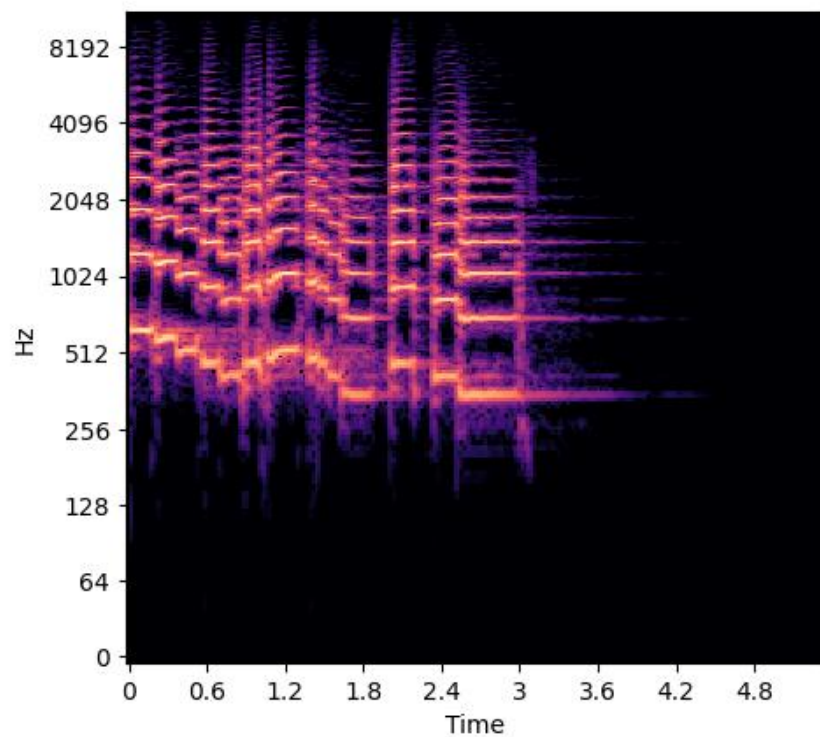
(Source: Müller & Zalkow, 2019)

Resources on Music Information Research (MIR)

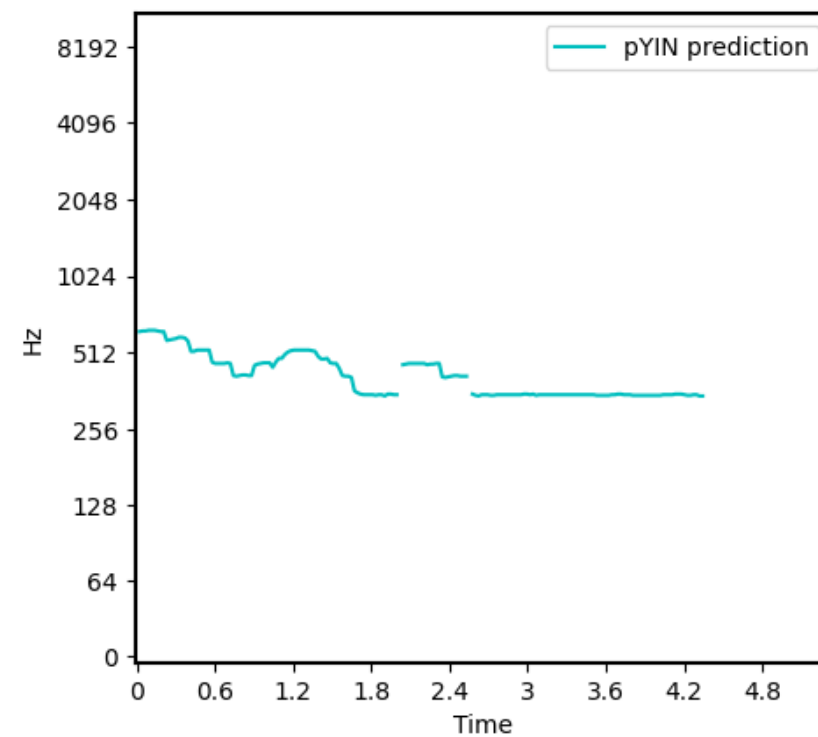
- Masataka Goto, Jin Ha Lee, and Meinard Müller, “[Exploring 25 Years of Music Information Retrieval: Perspectives and Insights](#),” *Tutorials of ISMIR*, 2024.
- Geoffroy Peeters, Gabriel Meseguer-Brocal, Alain Riou, and Stefan Lattner, “[Deep Learning 101 for Audio-based MIR](#),” *Tutorials of ISMIR*, 2024. ([book](#))
- Keunwoo Choi, György Fazekas, Kyunghyun Cho, and Mark Sandler, “[A Tutorial on Deep Learning for Music Information Retrieval](#),” *arXiv preprint arXiv:1709:04396*, 2017. ([code](#))

Recap

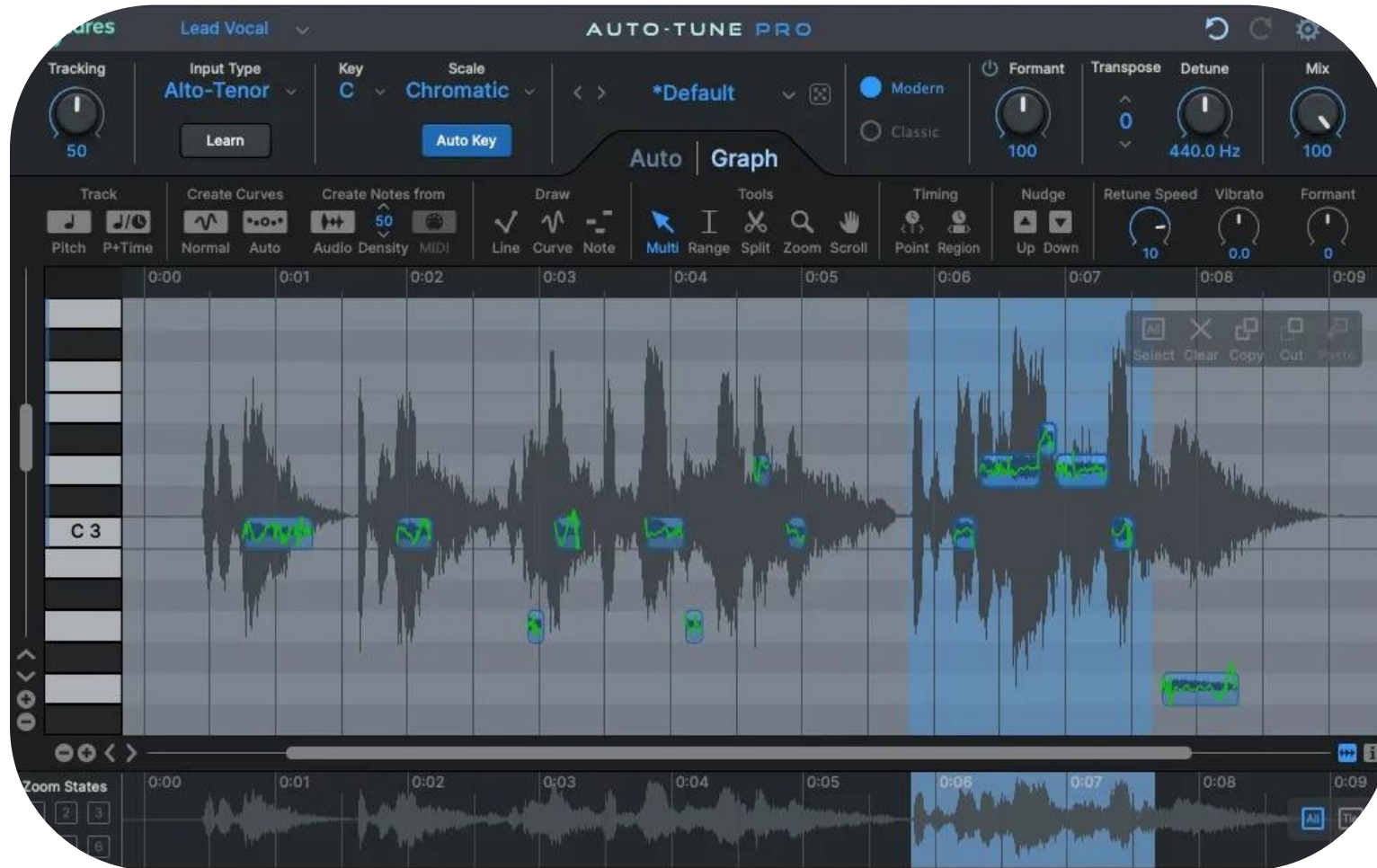
Fundamental Frequency (F0) Estimation



→ **F0 Estimation** →

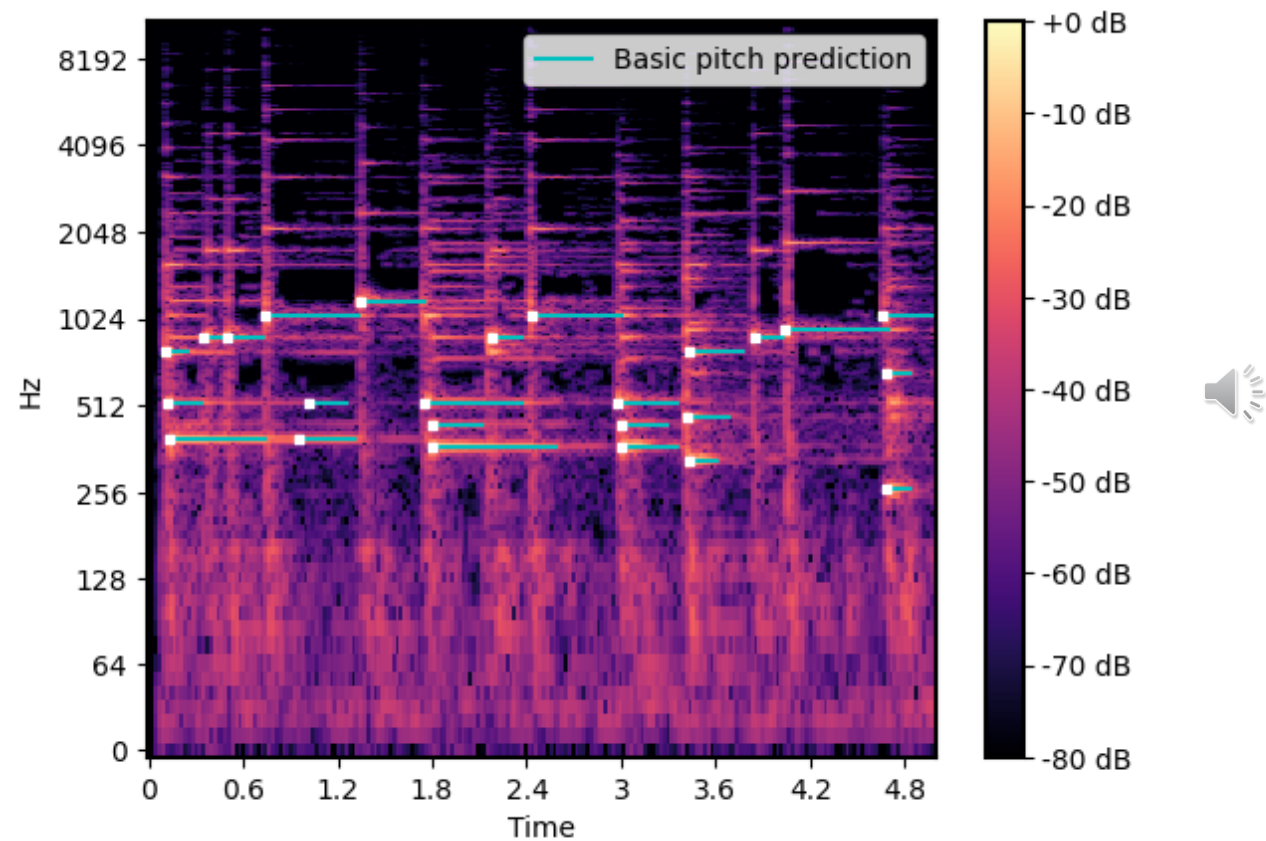


Auto-tune Pro



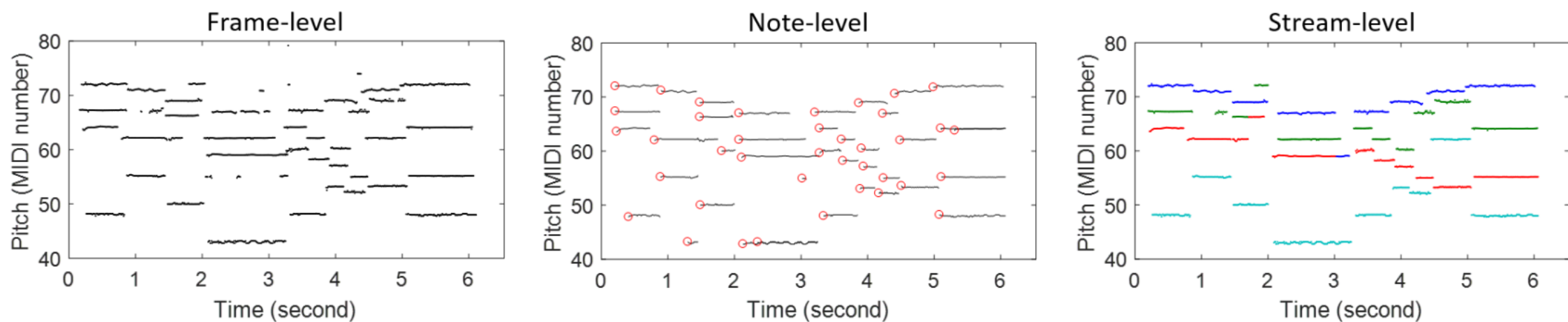
(Source: Antares Audio Technologies)

Polyphonic F0 Estimation



basicpitch.spotify.com

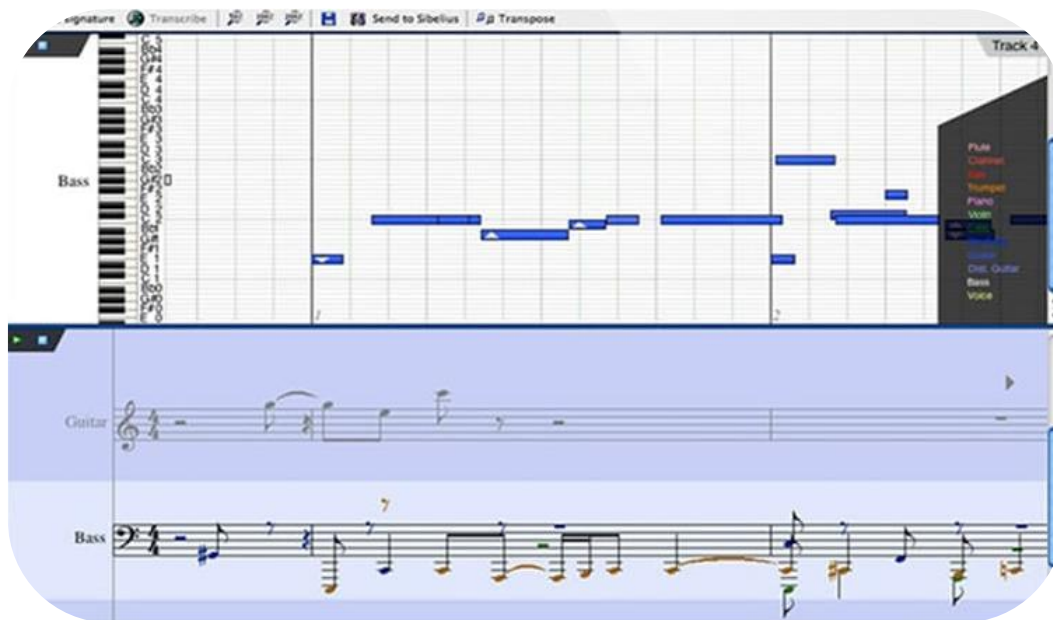
F0 Estimation vs. Music Transcription



(Source: Benetos et al., 2019)

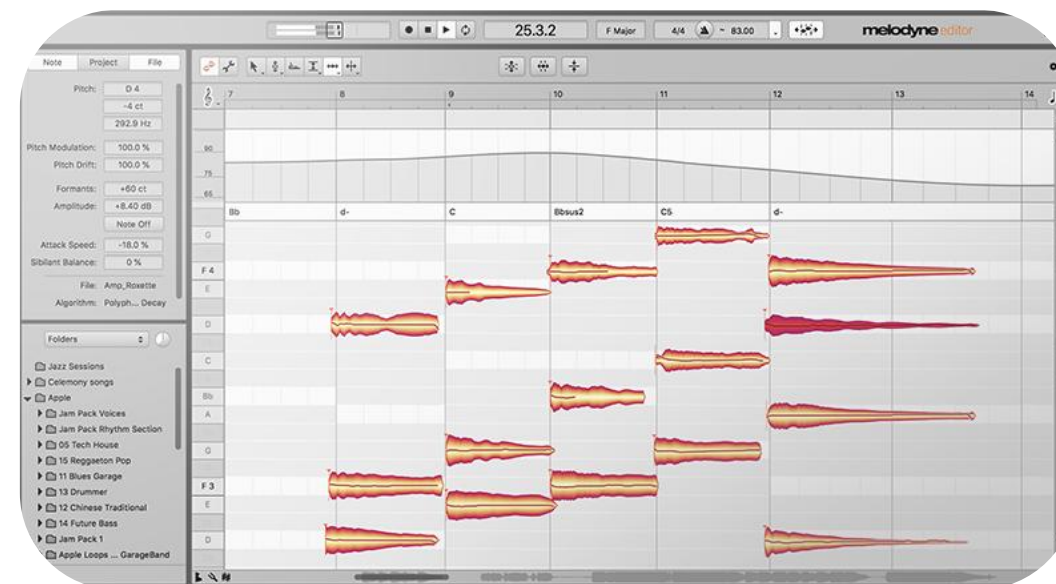
Commercial Music Transcription Software

AudioScore in Sibelius



(Source: Avid)

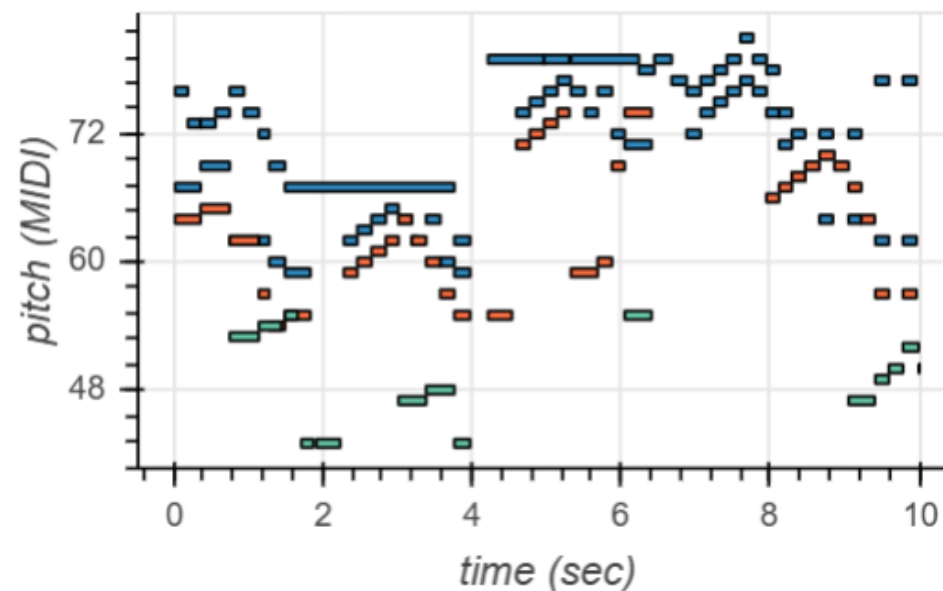
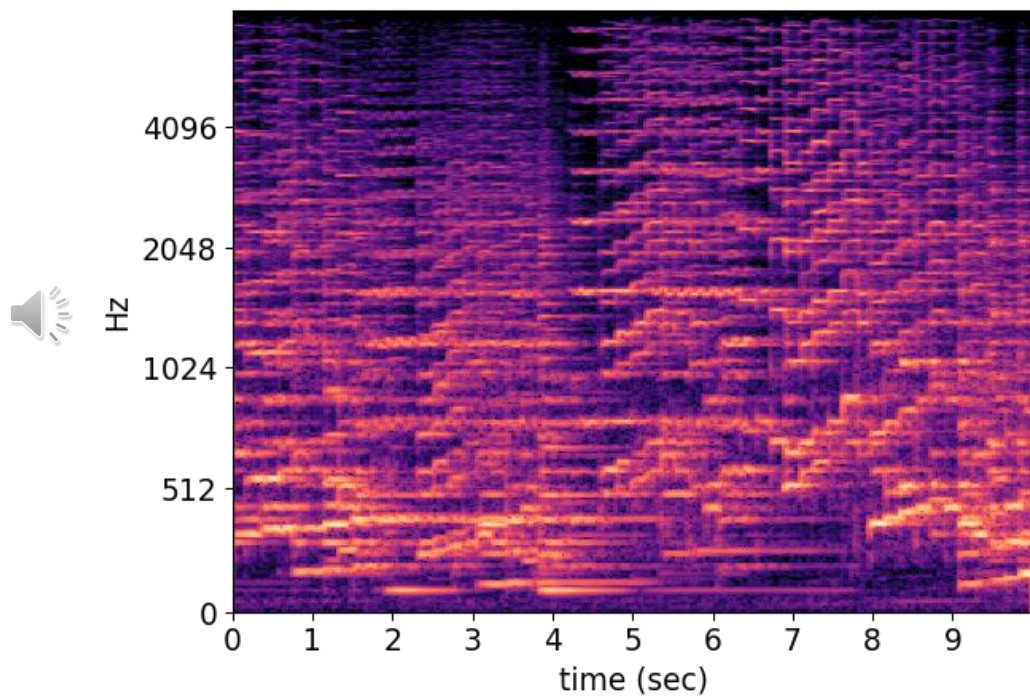
Melodyne Editor



(Source: Celemony)

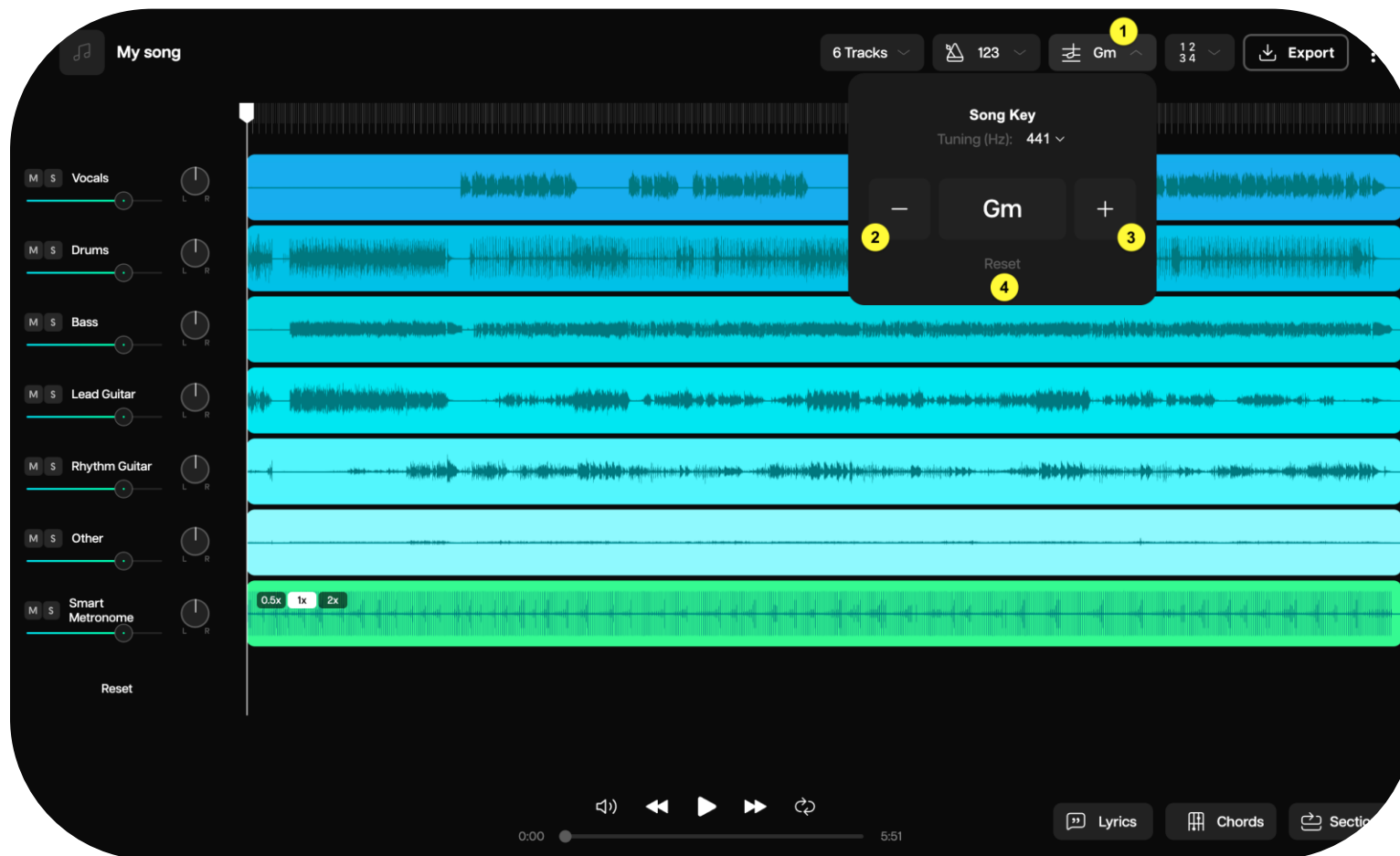
Multitrack Transcription Models

- **MT3** (Gardner et al., 2022)
 - github.com/magenta/mt3



(Source: Gardner et al., 2022)

Key Detection in Moises



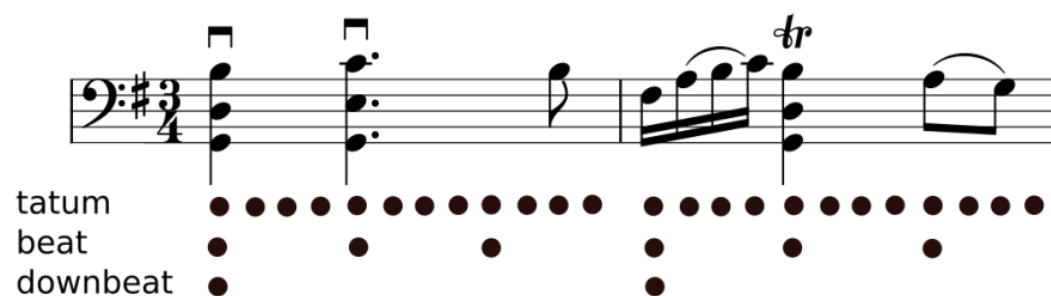
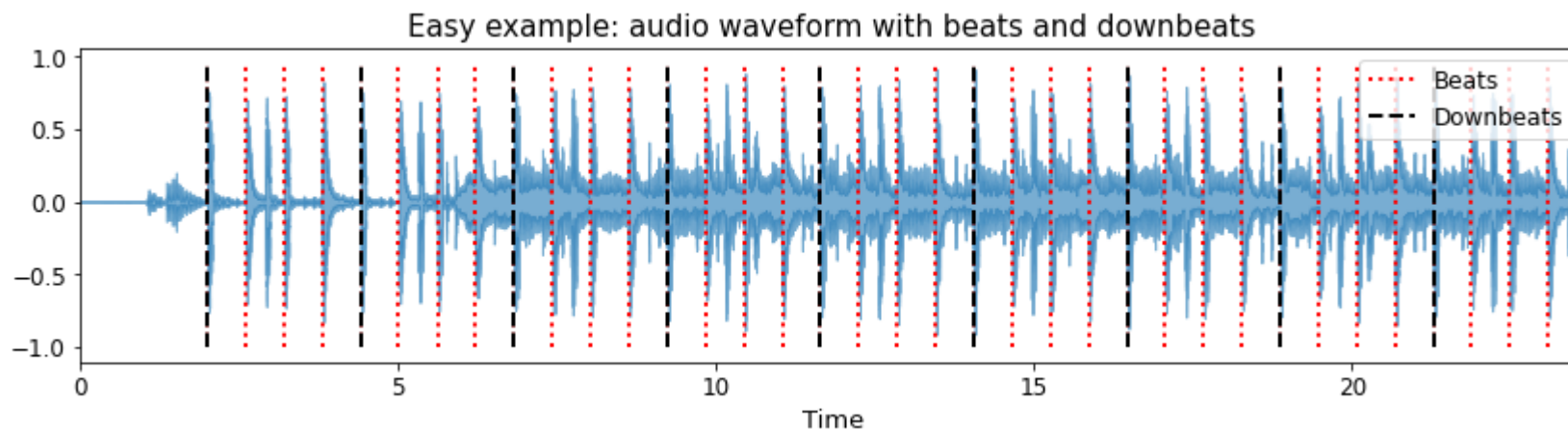
(Source: Moises)

| Chord Detection



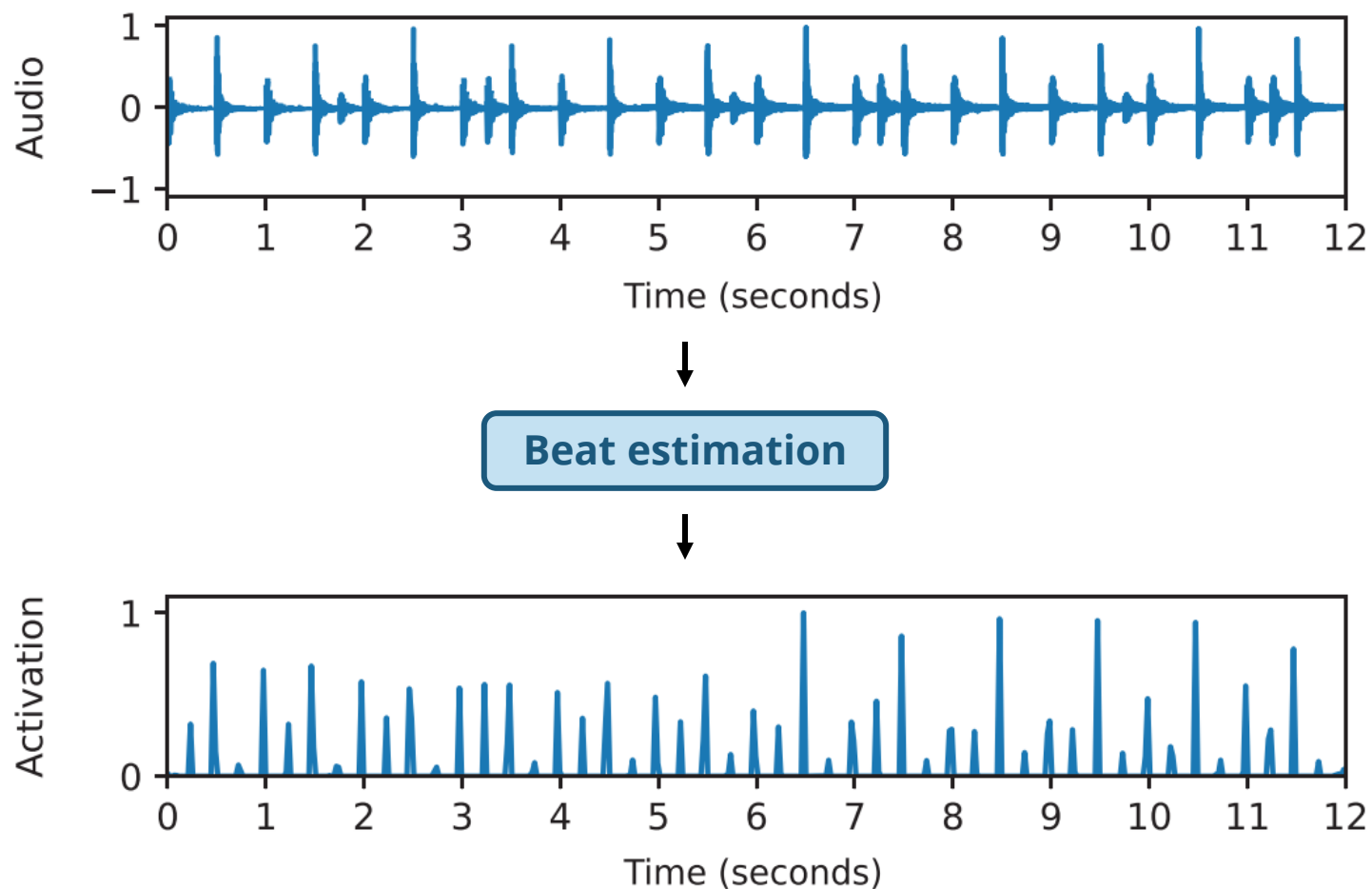
youtube.com/shorts/_N3b_GARMfA

Beat & Downbeat Estimation



(Source: Davies et al., 2021)

Beat & Downbeat Estimation



(Source: Meier et al., 2024)

Structure Analysis

Music segmentation

The musical score is divided into several sections, each with its own tempo and dynamic markings:

- Allegro.** (A1, A2): First movement, starting with *f* and *ff* dynamics.
- f marc.** (B1, B2): Second movement, marked *f marc.* and *p poco rit.*
- Vivace.** (C): Third movement, marked *sf* and *p poco rit.*
- A3, B3, B4, D**: Final section, marked *ff marcato* and *p poco ritard.*

Figure 4.5 following [Müller, FMP, Springer 2015]

(Source: Müller & Zalkow, 2019)

Hierarchical music segmentation

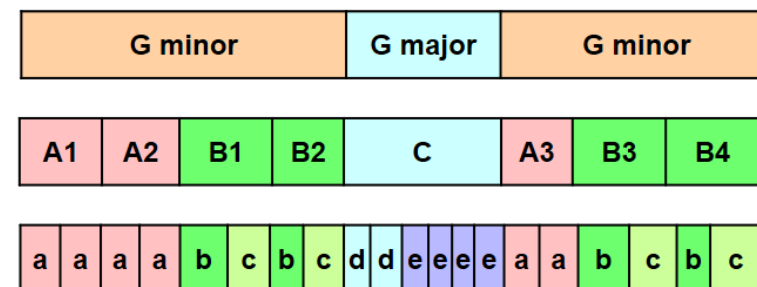
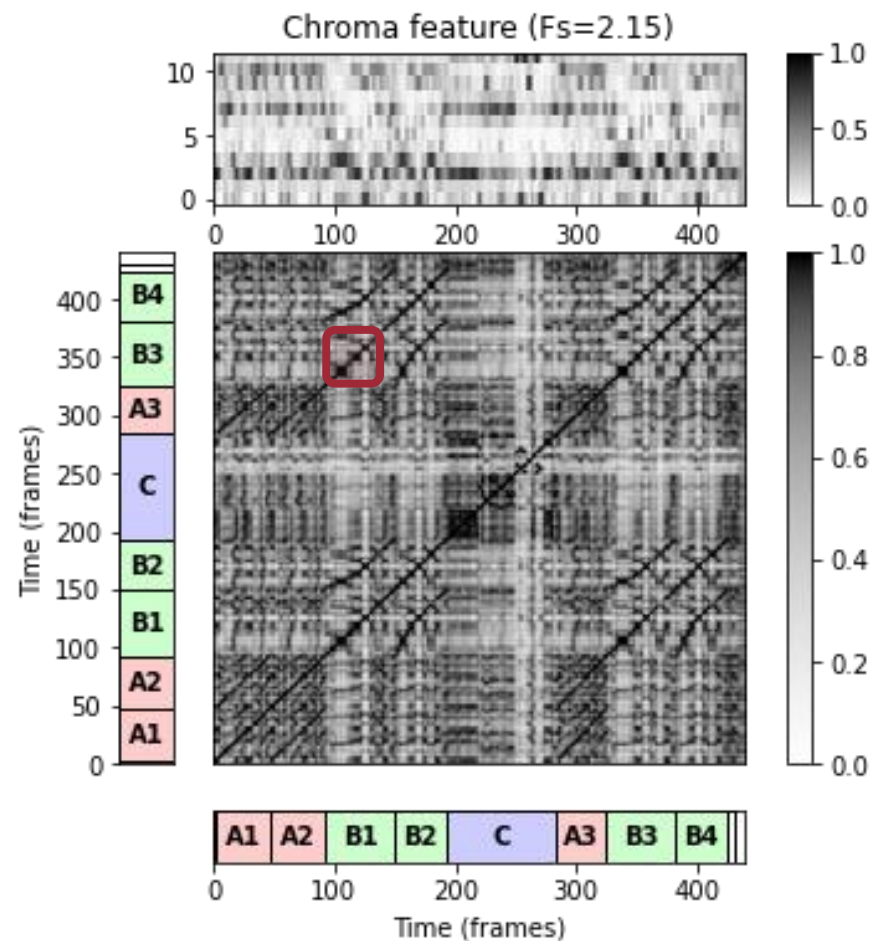


Figure 4.28 from [Müller, FMP, Springer 2015]

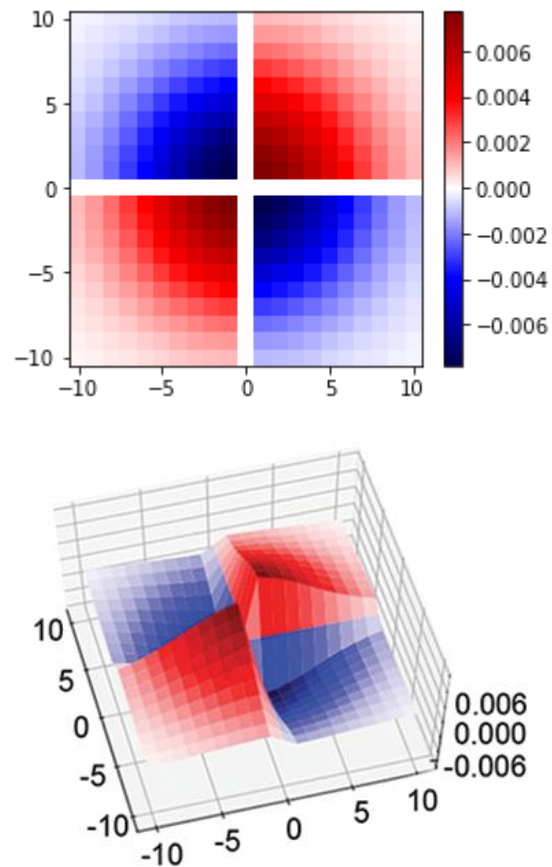
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



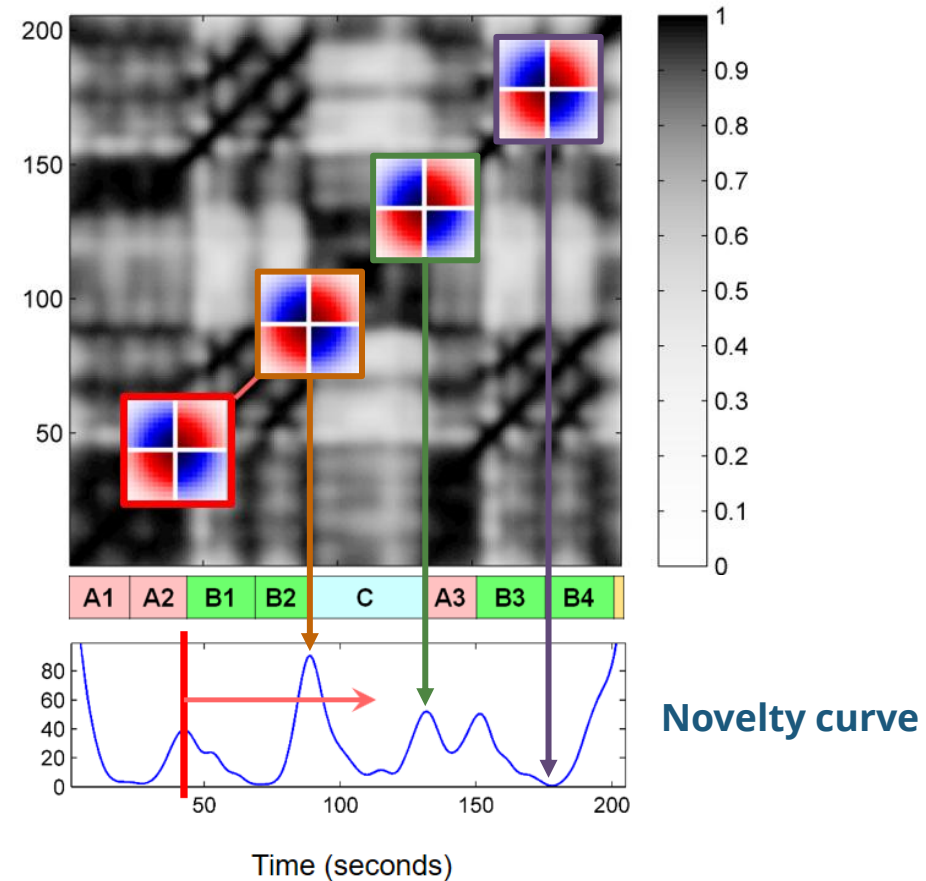
(Source: Müller & Zalkow, 2019)

Self-Similarity Matrices (SSMs)



(Source: Müller & Chiu, 2024)

Figure 4.24 from [Müller, FMP, Springer 2015]



(Source: Müller & Zalkow, 2019)

Next Lecture

Music Classification

