

Winter Term 2026/27

Selected Topics in Deep Learning for Audio, Speech, and Music Processing

Lecture

Introduction to Music Processing

Stefan Balke, Andreas Brendel, Ben Maman, Meinard Müller

Music

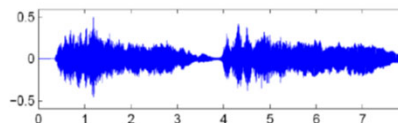


Music Information Retrieval (MIR)

Sheet Music (Image)



CD / MP3 (Audio)



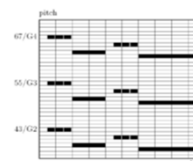
MusicXML (Text)

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

Dance / Motion (Mocap)



MIDI



Singing / Voice (Audio)



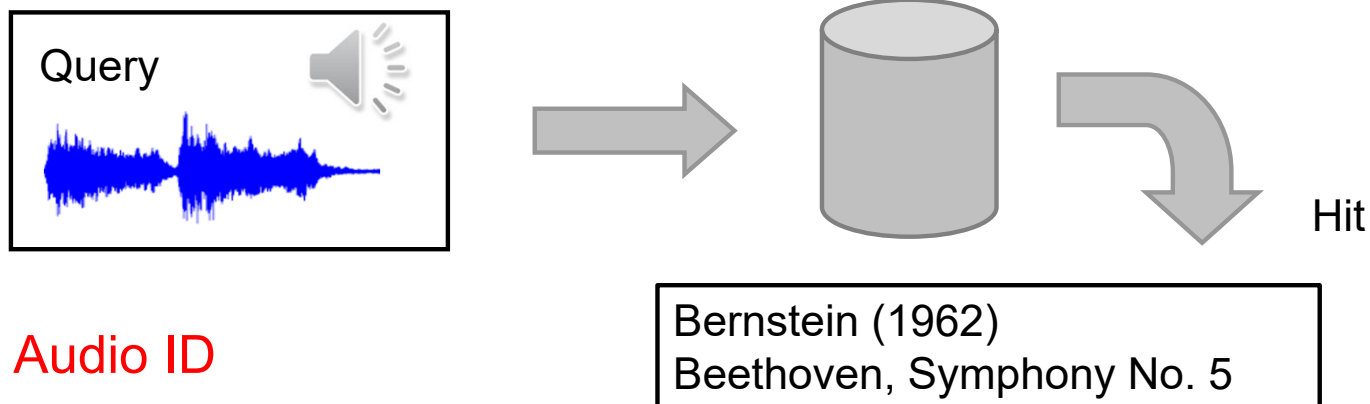
Music Film (Video)



Music Literature (Text)

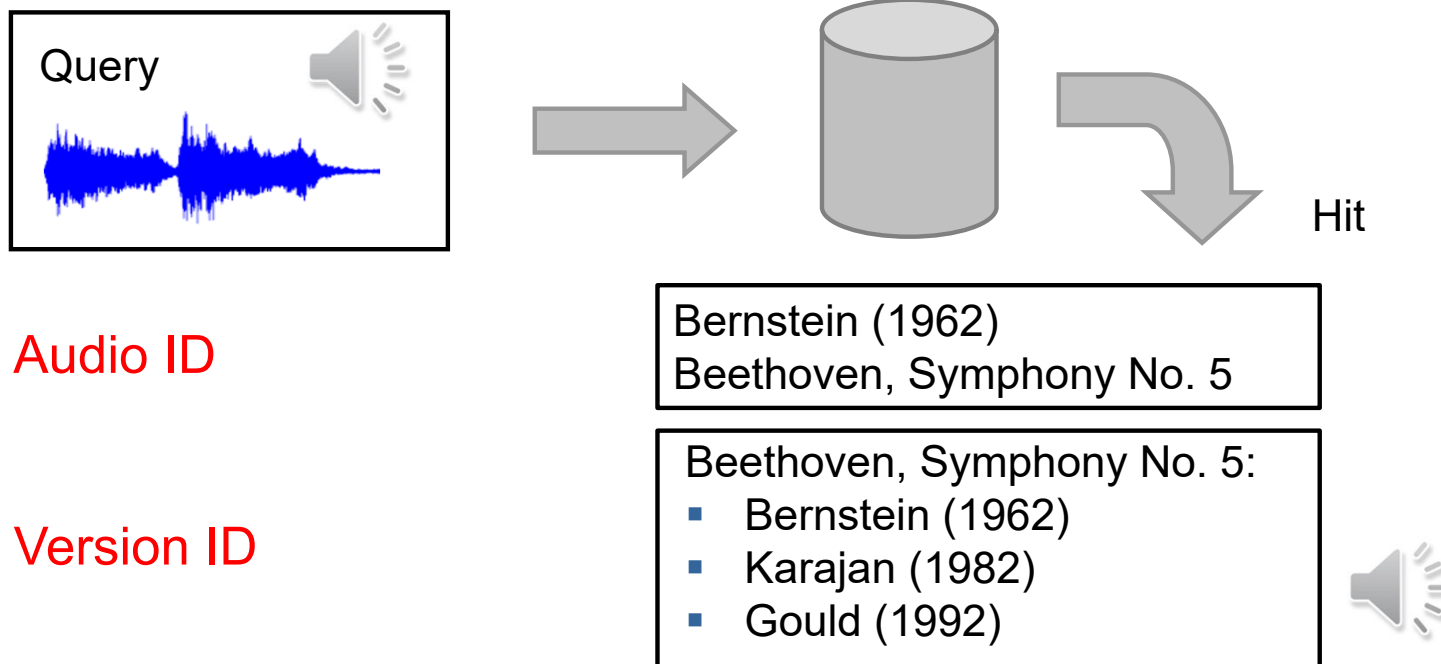


Music Retrieval

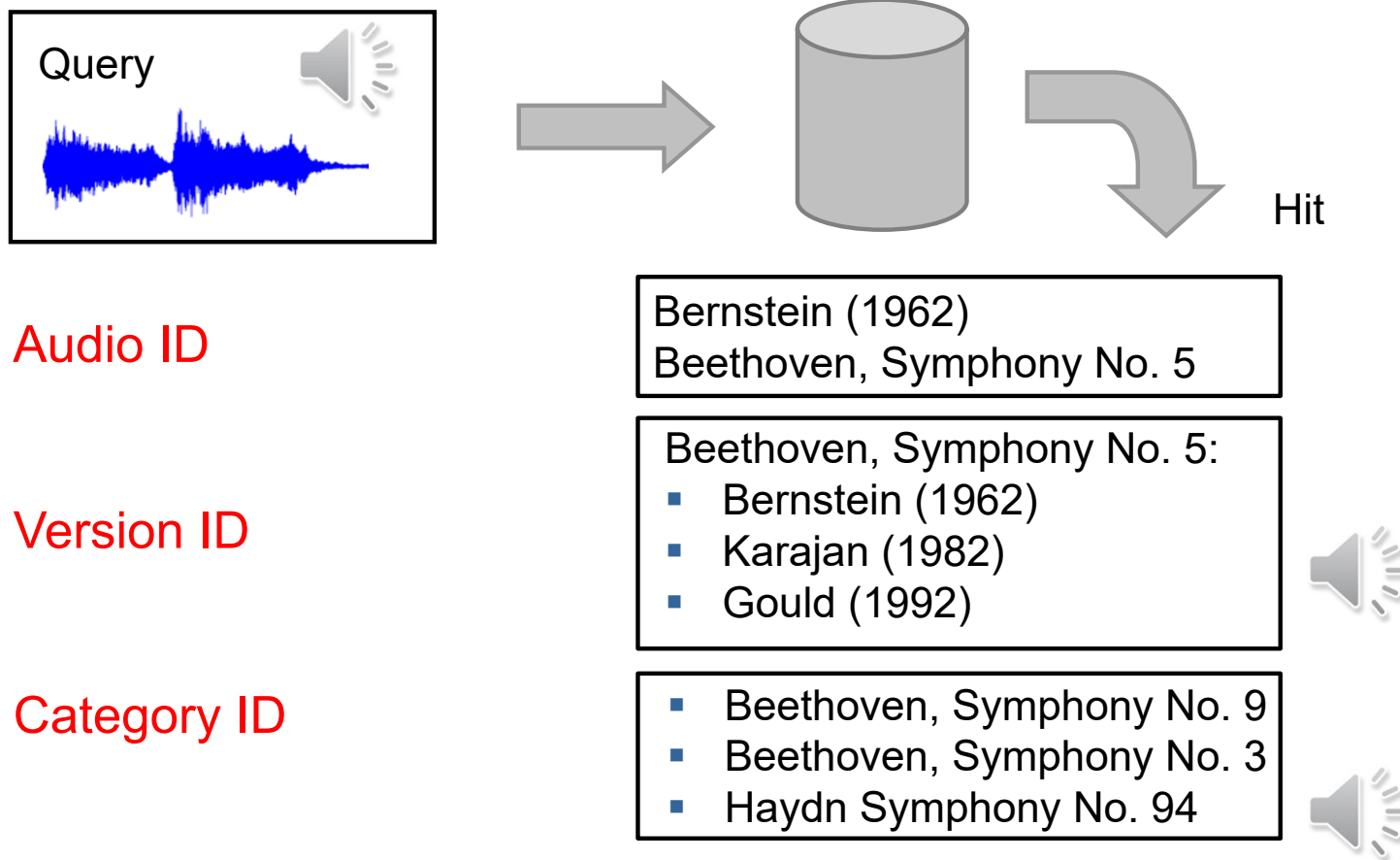


Audio ID

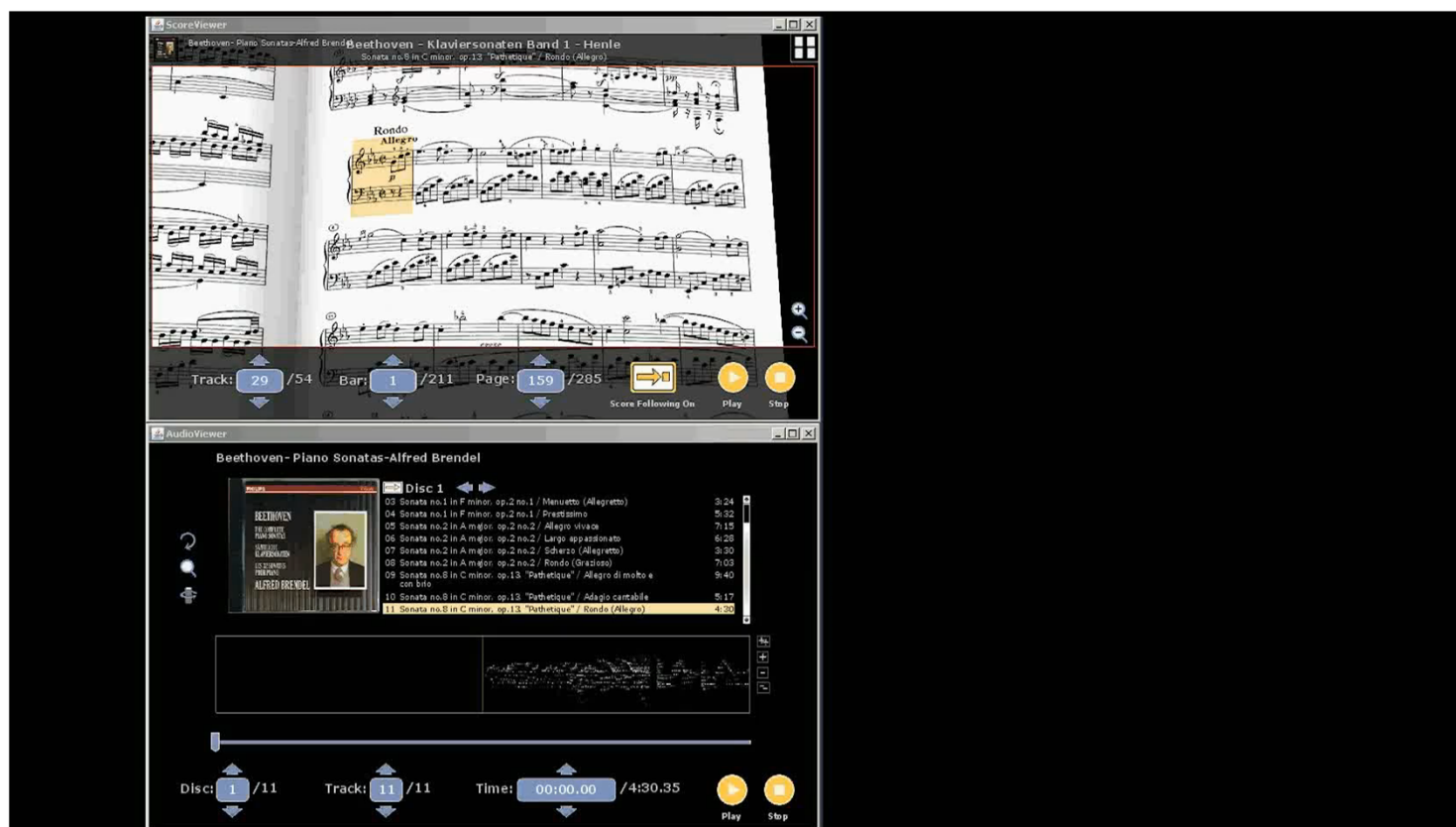
Music Retrieval



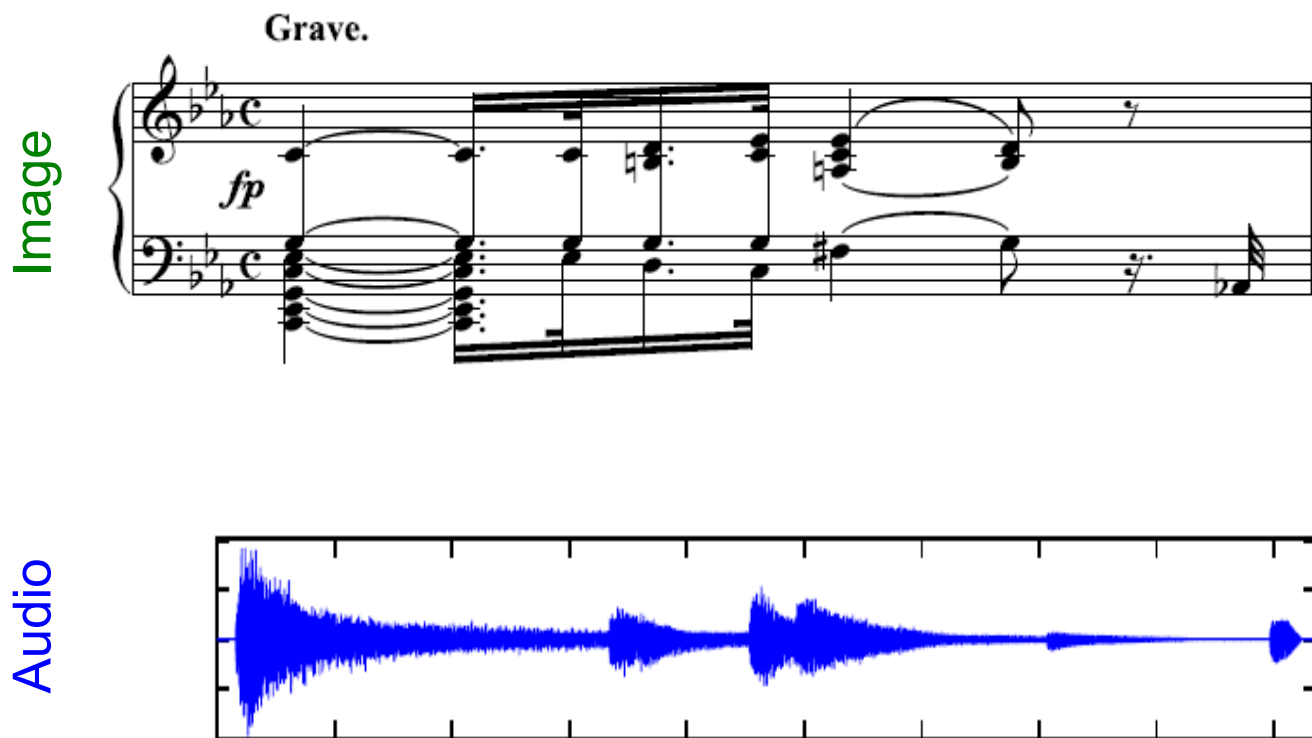
Music Retrieval



Music Synchronization



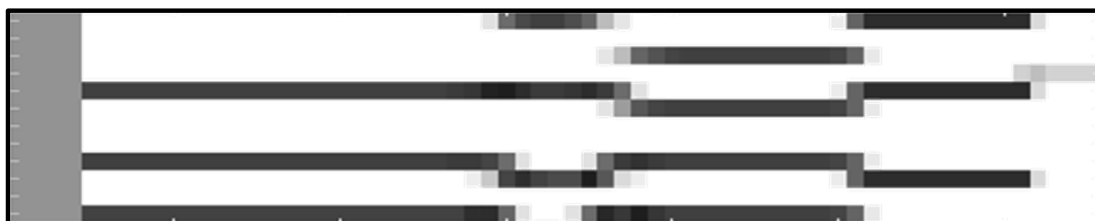
Music Synchronization: Image-Audio



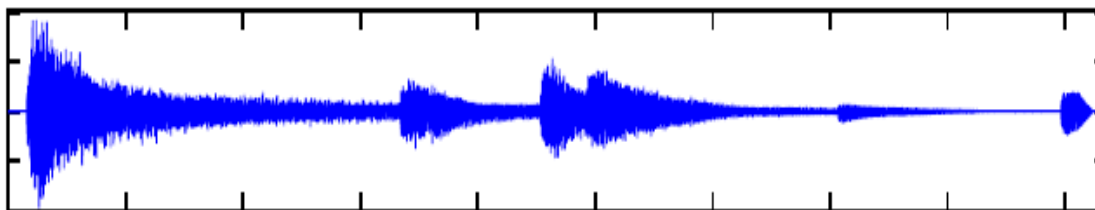
Music Synchronization: Image-Audio

Image Processing: Optical Music Recognition

Image



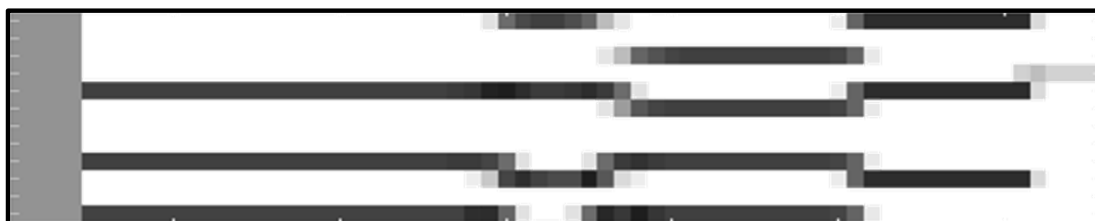
Audio



Music Synchronization: Image-Audio

Image Processing: Optical Music Recognition

Image



Audio

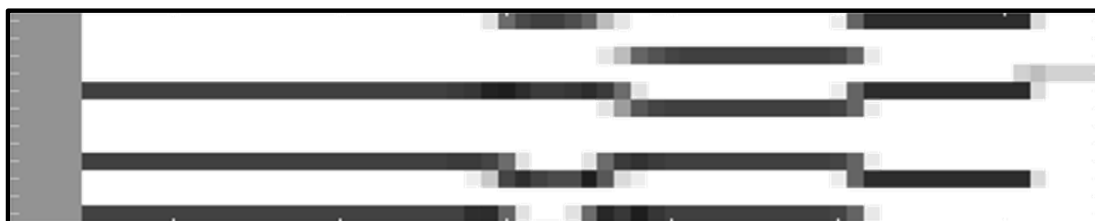


Audio Processing: Fourier Analysis

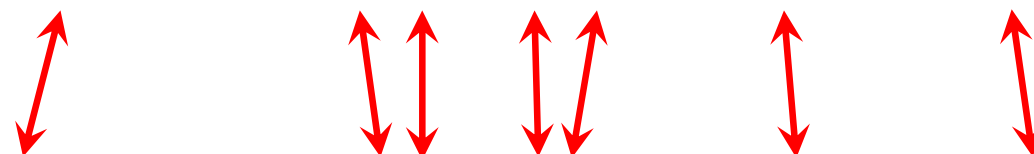
Music Synchronization: Image-Audio

Image Processing: Optical Music Recognition

Image



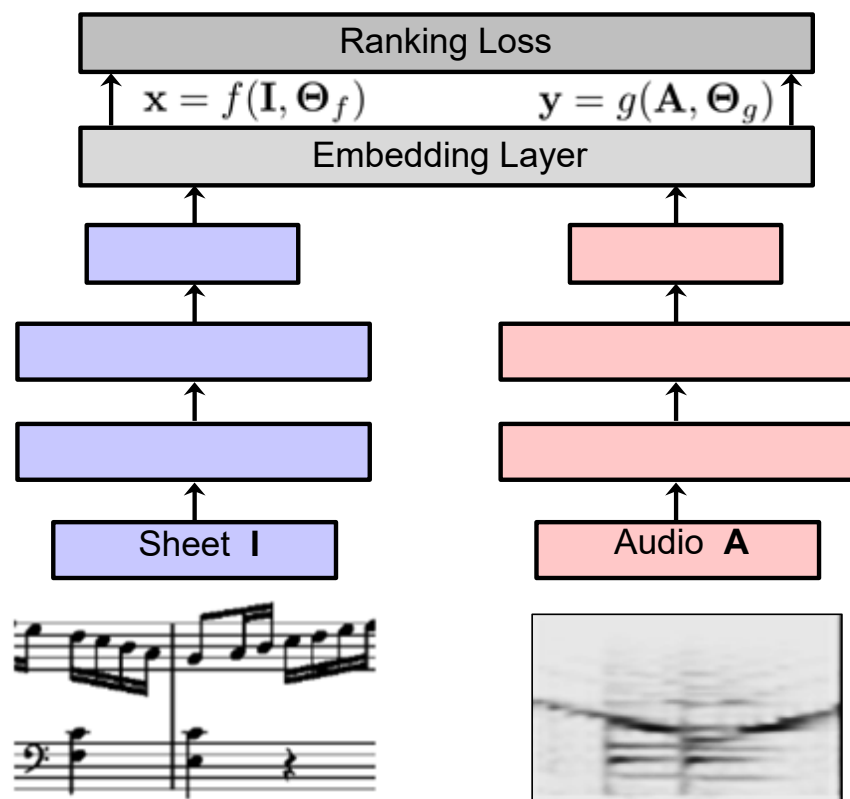
Audio



Audio Processing: Fourier Analysis

Music Synchronization: Image-Audio

Müller, Arzt, Balke, Dorfer, Widmer: **Cross-Modal Music Retrieval and Applications.**
IEEE Signal Process. Mag. 36(1): 52–62, 2019



- Deep learning
- Embedding techniques
- Weak annotations
- Loss functions
- ...

Concert Band



Concert Band



Four-Voice Choir (SATB)

Drese: Jesu geh voran

Soprano (S)



Alto (A)



Tenor (T)



Bass (B)



Sheet music for the Four-Voice Choir (SATB) of the piece 'Drese: Jesu geh voran'. The music is written for Soprano (S), Alto (A), Tenor (T), and Bass (B) voices. The Soprano part is highlighted in orange, the Alto part in red, the Tenor part in green, and the Bass part in blue. The music is in 3/4 time and features a mix of whole, half, and quarter notes, with some rests.

Four-Voice Choir (SATB)

Drese: Jesu geh voran

Soprano (S)



Alto (A)



Tenor (T)



Bass (B)



Sheet music for the Four-Voice Choir (SATB) of the chorale 'Jesu geh voran'. The music is written for Soprano (S), Alto (A), Tenor (T), and Bass (B) voices. The key signature is one sharp (F#) and the time signature is 4/4. The Soprano part is highlighted in orange, the Alto part in red, the Tenor part in green, and the Bass part in blue.



ChoraleBricks

Multitrack Dataset for Wind Music Research

- Dataset featuring 10 SATB chorales
- 193 isolated tracks with 13 instruments
- Recorded along conducting video for synchronization

Balke, Berndt, Müller: **ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research**. TISMIR, 8(1): 39–54, 2025



ChoraleBricks

Multitrack Dataset for Wind Music Research

- Dataset featuring 10 SATB chorales
- 193 isolated tracks with 13 instruments
- Recorded along conducting video for synchronization

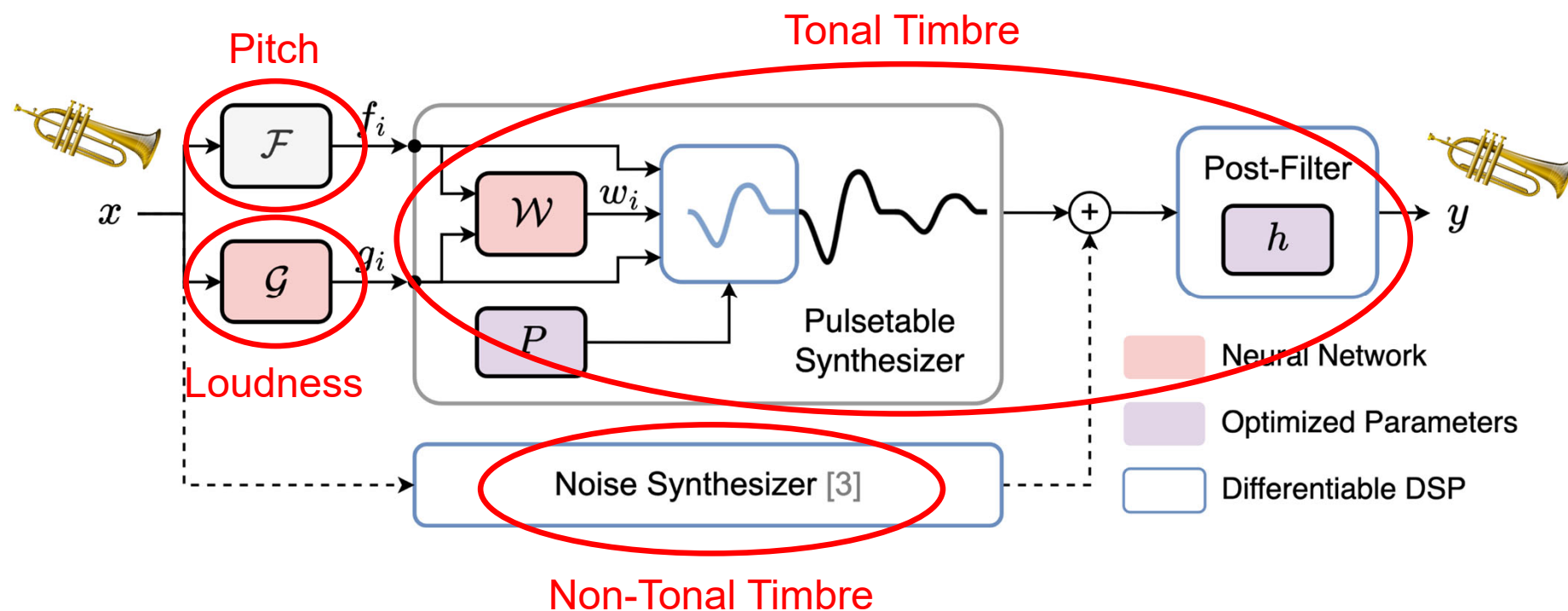
Balke, Berndt, Müller: **ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research**. TISMIR, 8(1): 39–54, 2025



Differentiable Digital Signal Processing

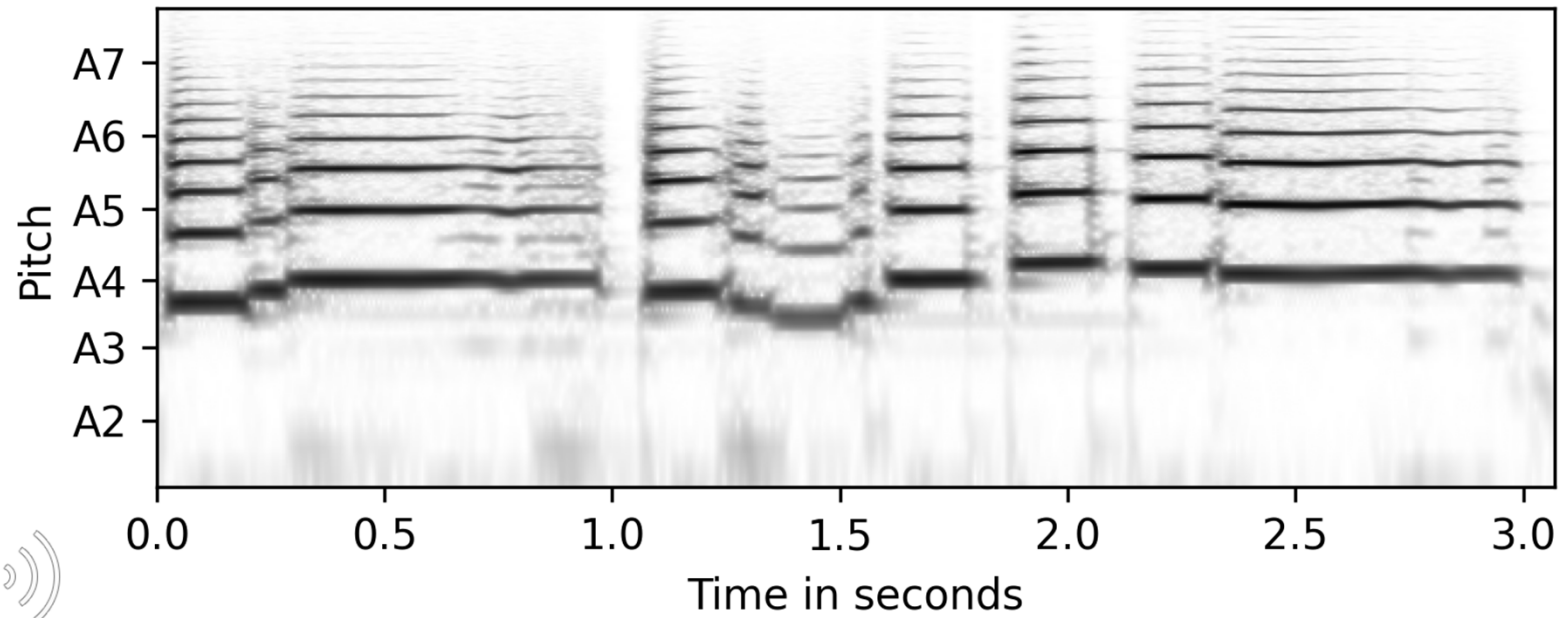
Pulsetable Synthesis

Dittmar, Zeitler, Balke, Schwär, Müller: **PULSE-IT: Lightweight and Expressive Synthesis of Wind Instrument Playing**. LBD ISMIR, 2025



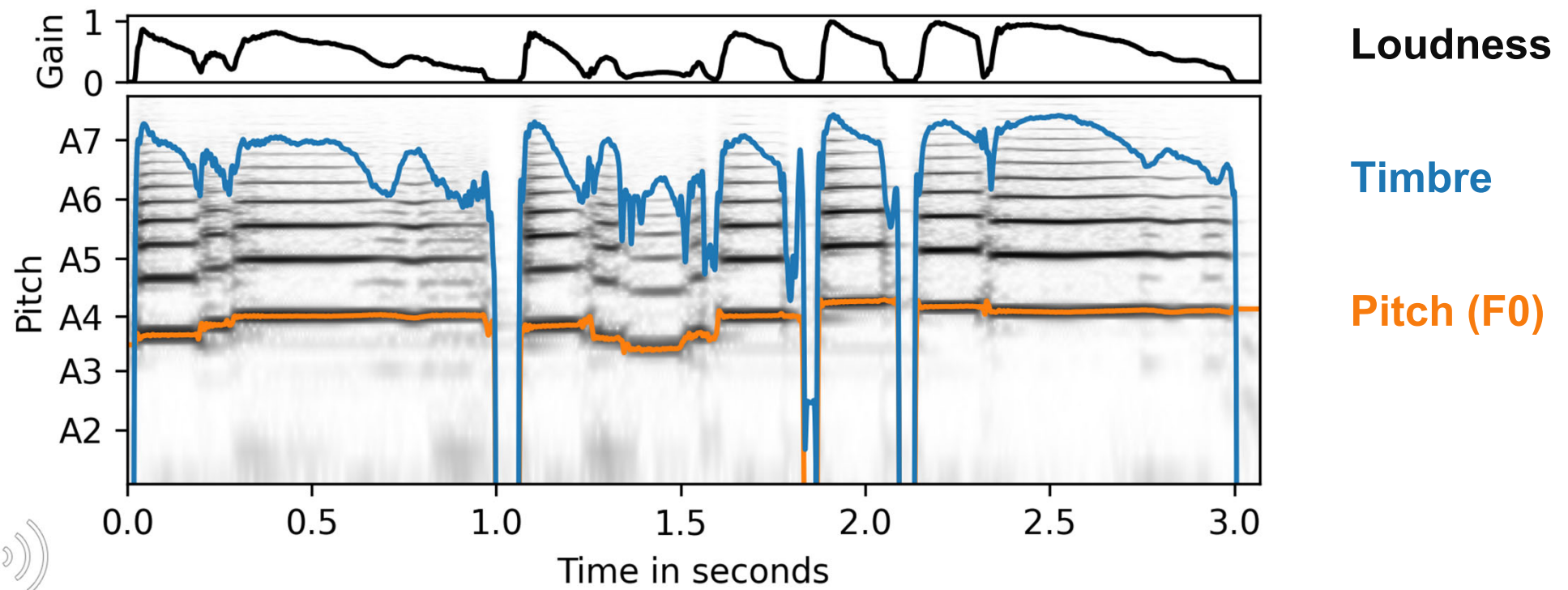
Wind Instrument Synthesis

Input Signal



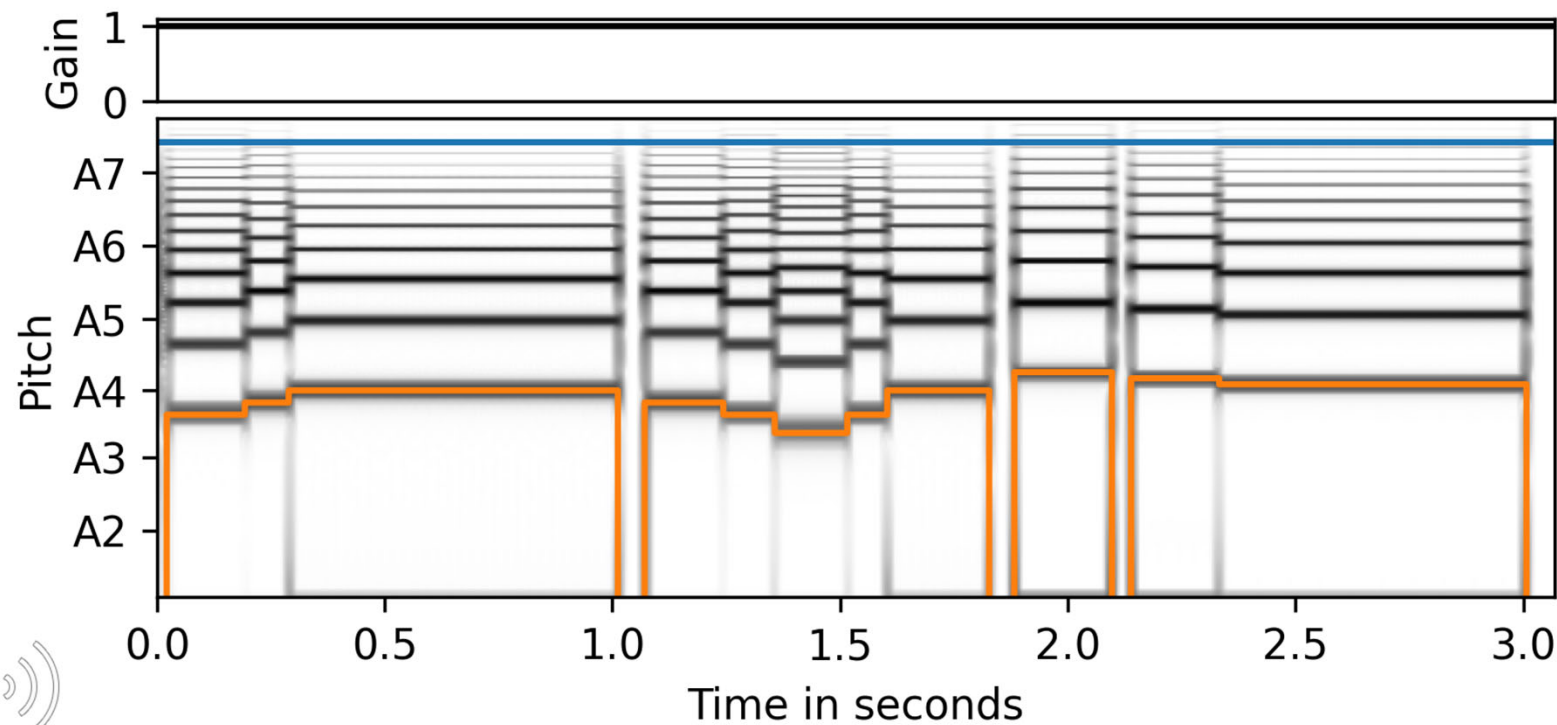
Wind Instrument Synthesis

Extract Control Signals



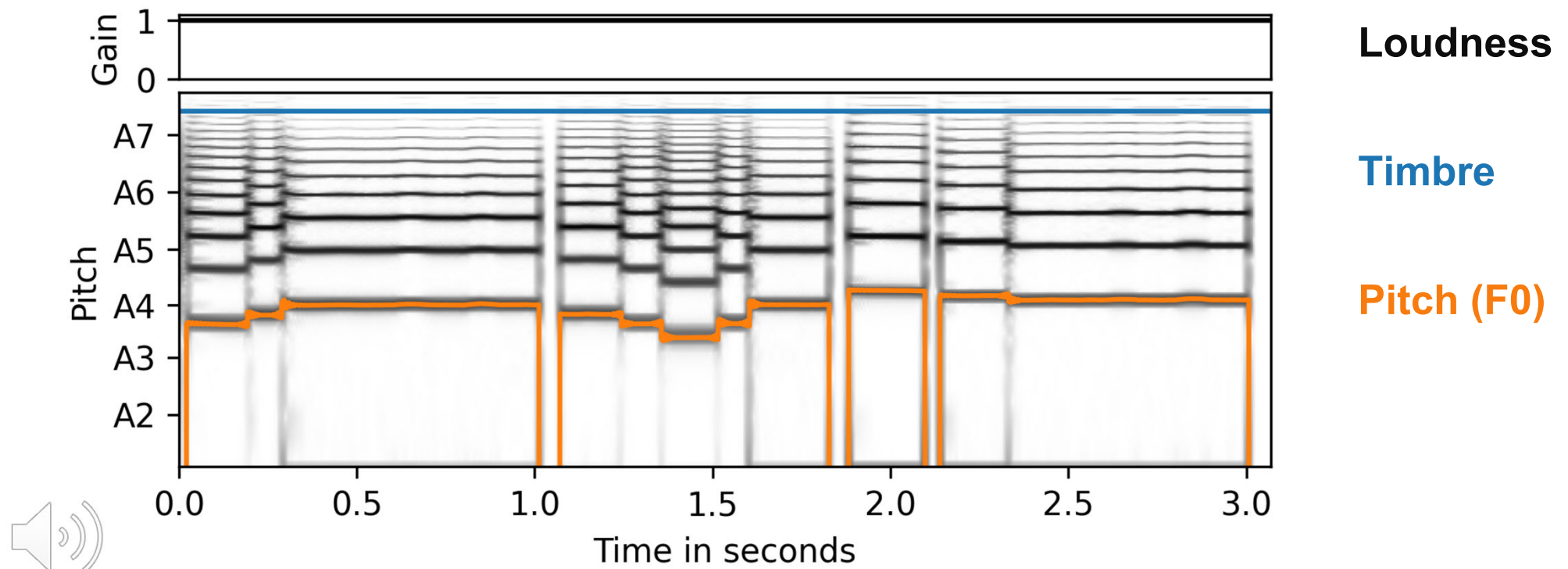
Wind Instrument Synthesis

Generated Control Signals



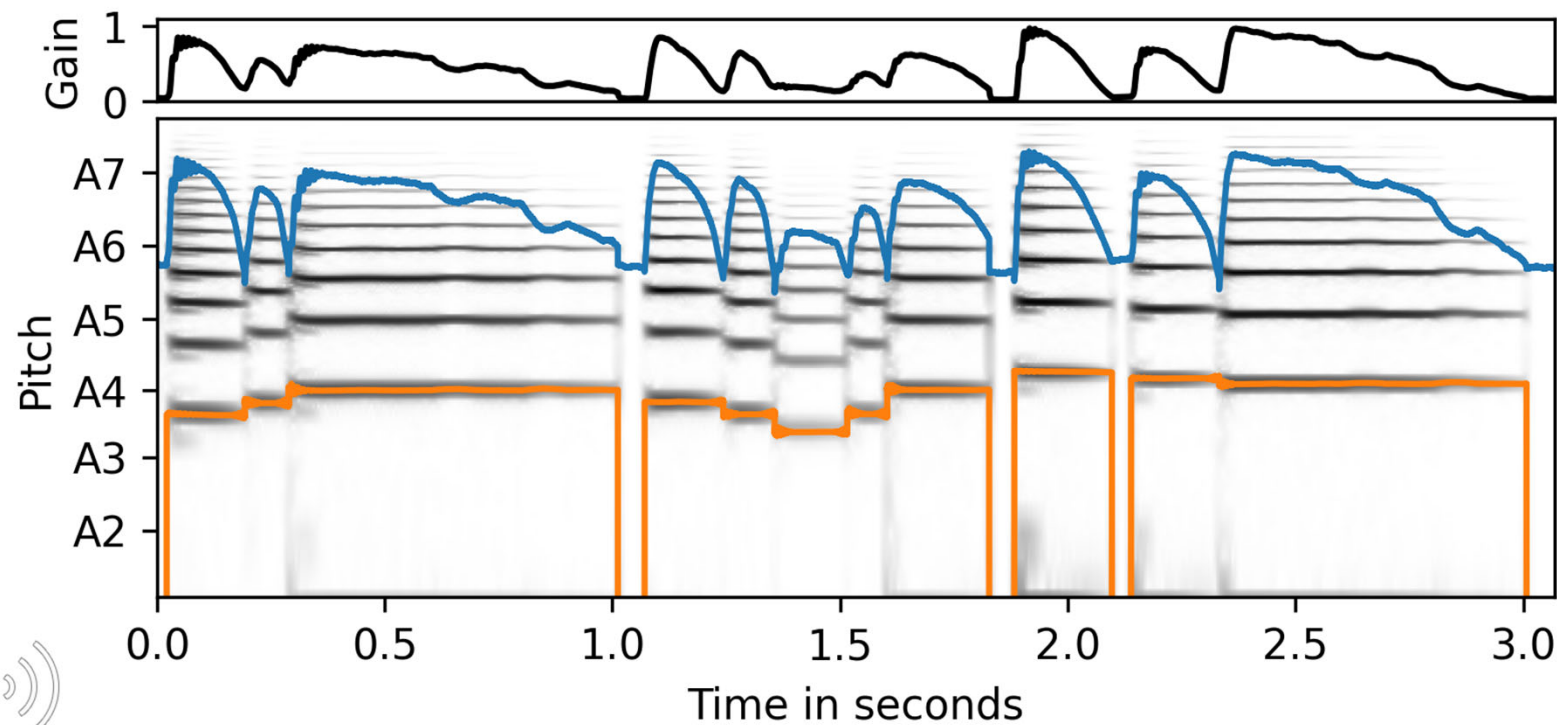
Wind Instrument Synthesis

Generated Control Signals with Pitch Fluctuations



Wind Instrument Synthesis

Generated Control Signals with Pitch & Loudness Fluctuations



Loudness

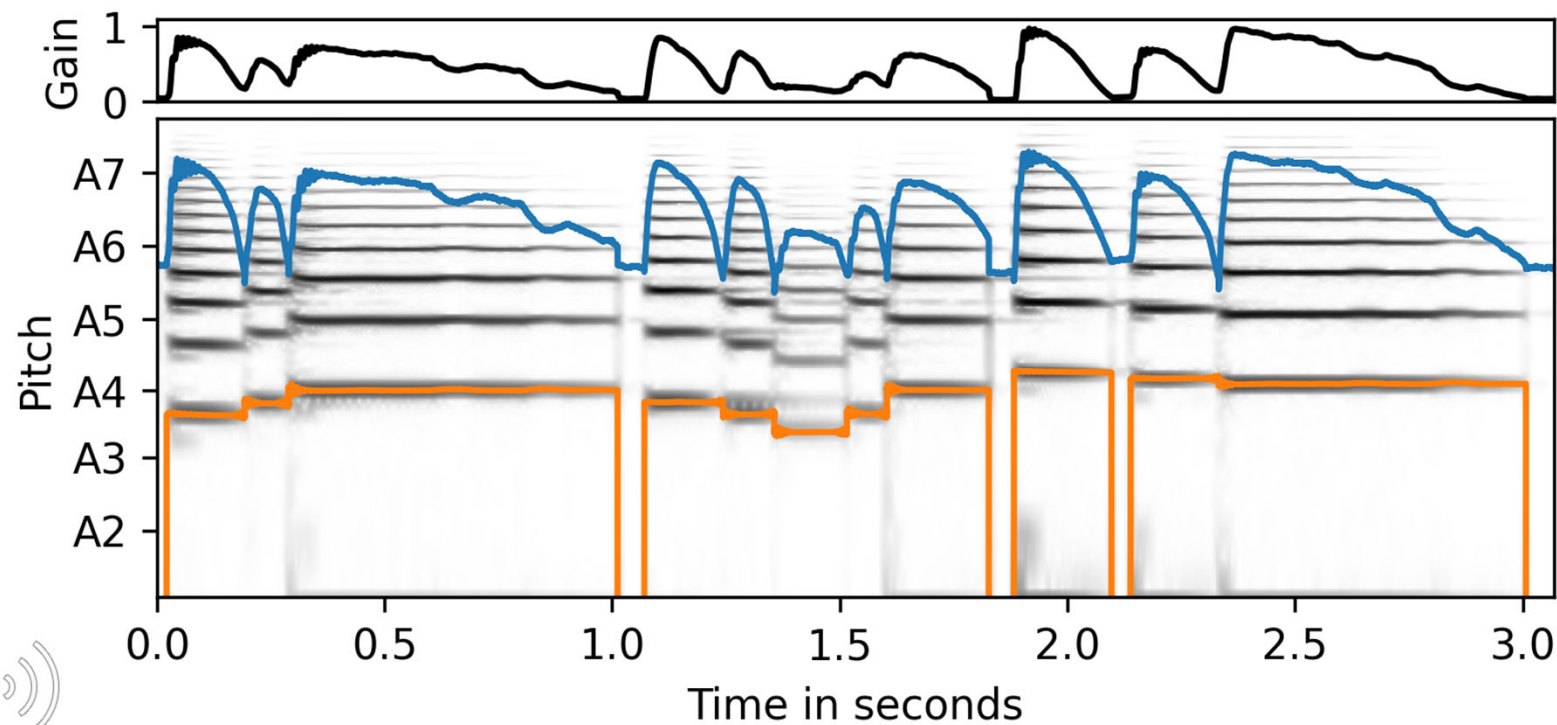
Timbre

Pitch (F0)



Wind Instrument Synthesis

Generated Control Signals with Fluctuations and Room Acoustics



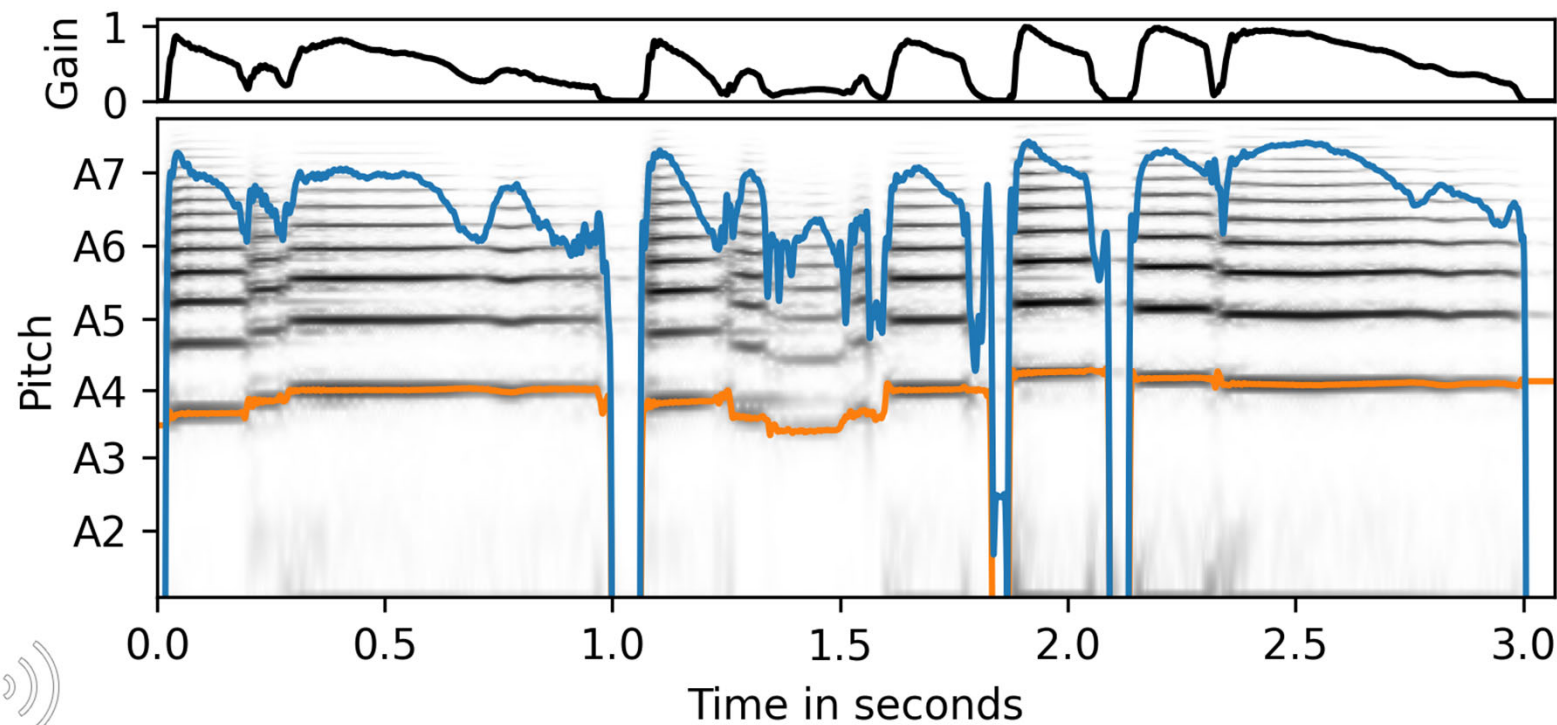
Loudness

Timbre

Pitch (F0)

Wind Instrument Synthesis

Resynthesized Trumpet



Loudness

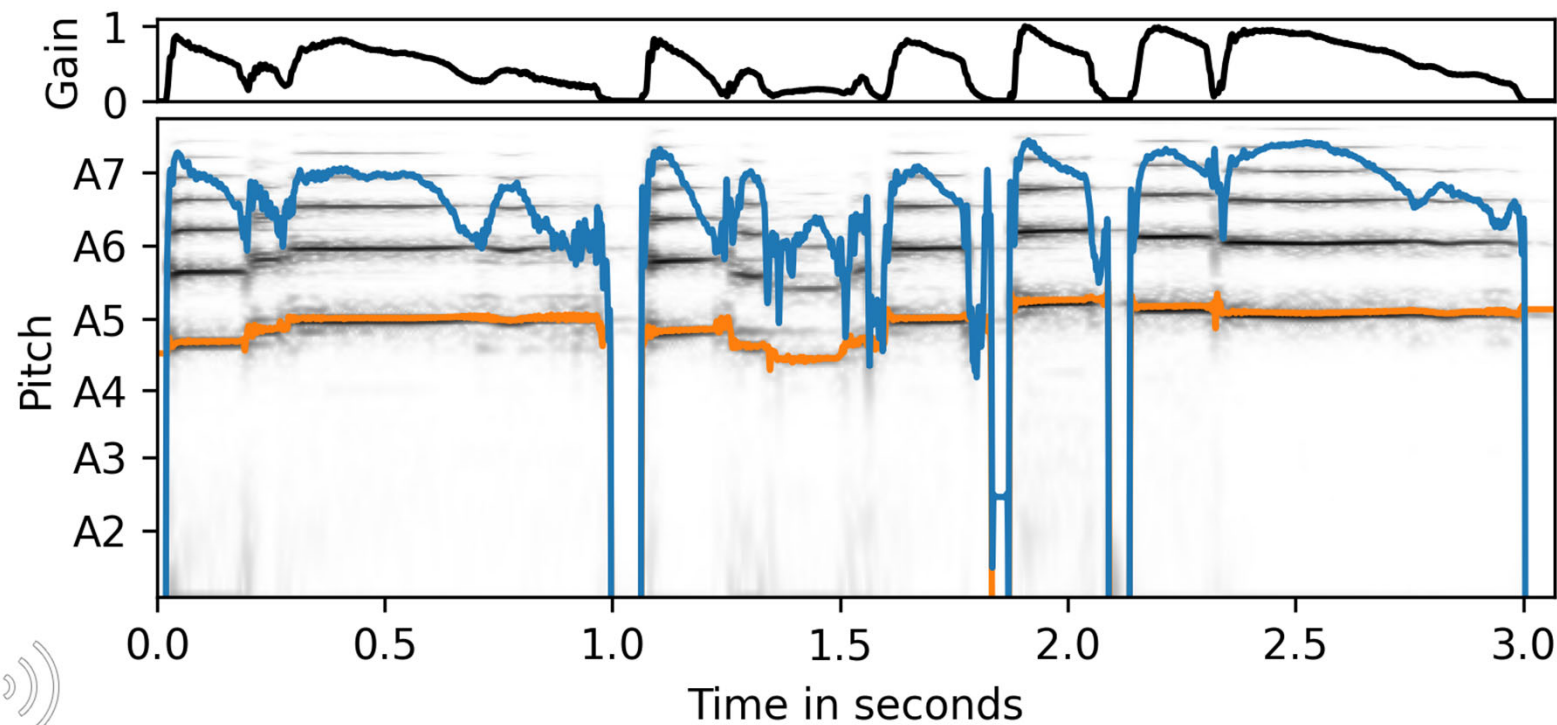
Timbre

Pitch (F0)



Wind Instrument Synthesis

Resynthesized Trumpet (One Octave Higher)



Loudness

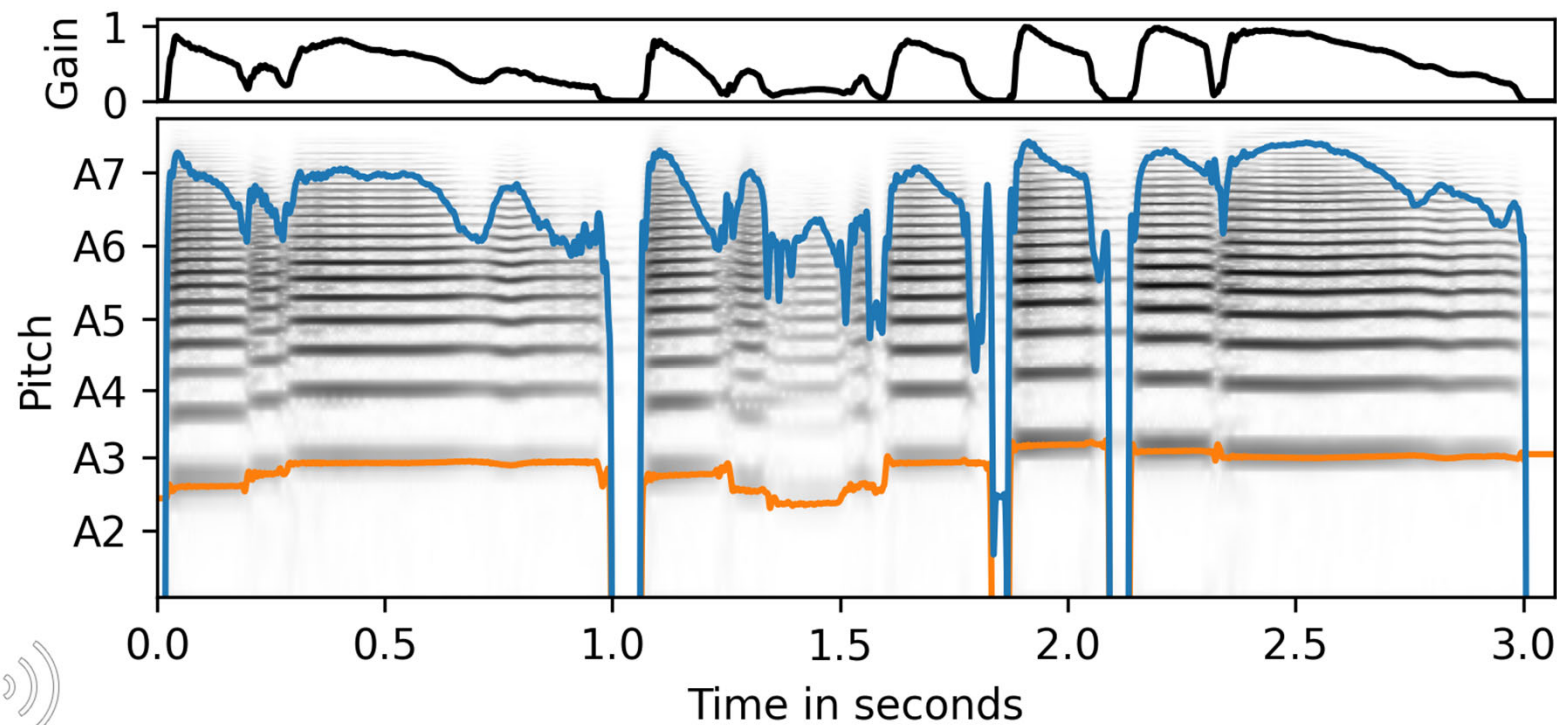
Timbre

Pitch (F0)



Wind Instrument Synthesis

Resynthesized Trumpet (One Octave Lower)



Loudness

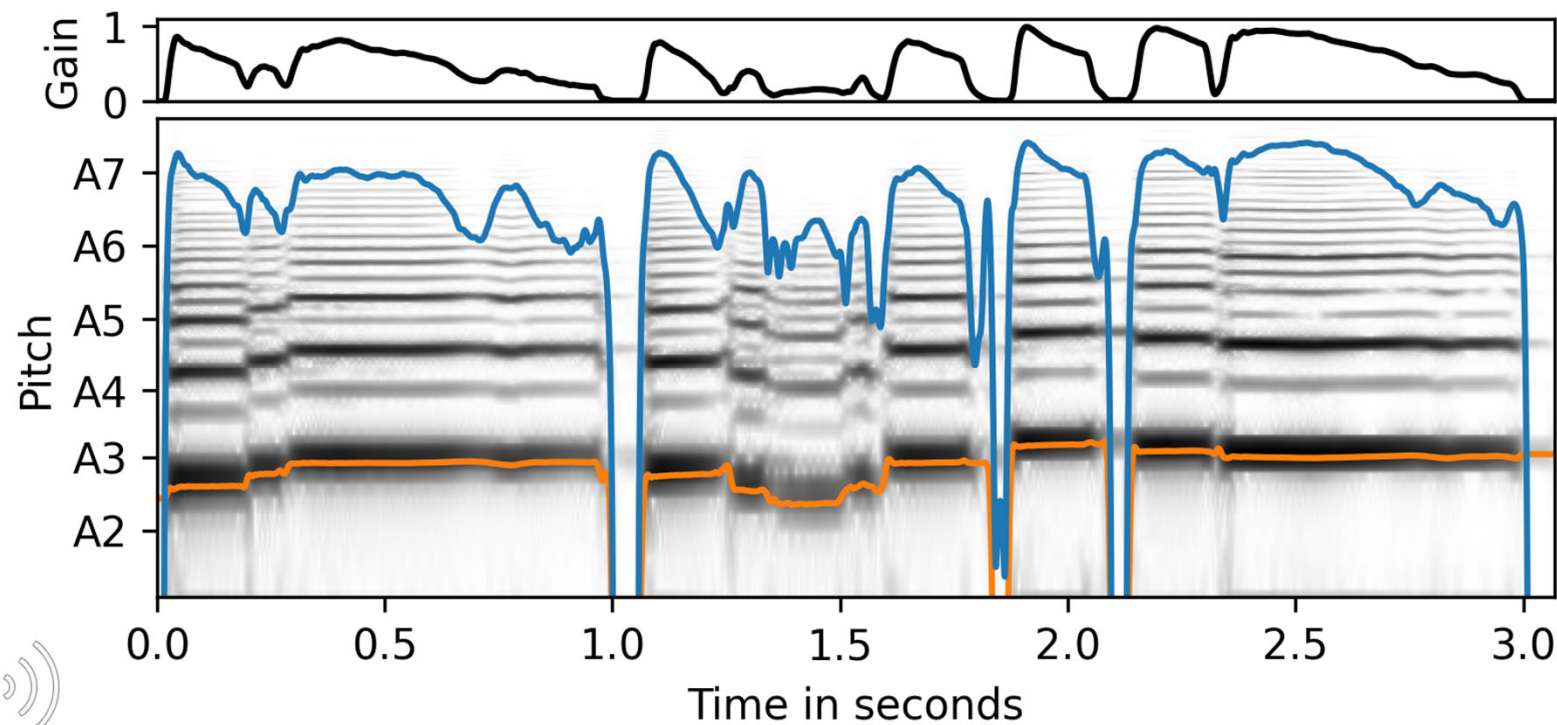
Timbre

Pitch (F0)



Wind Instrument Synthesis

Resynthesized Clarinet (One Octave Lower)



Loudness

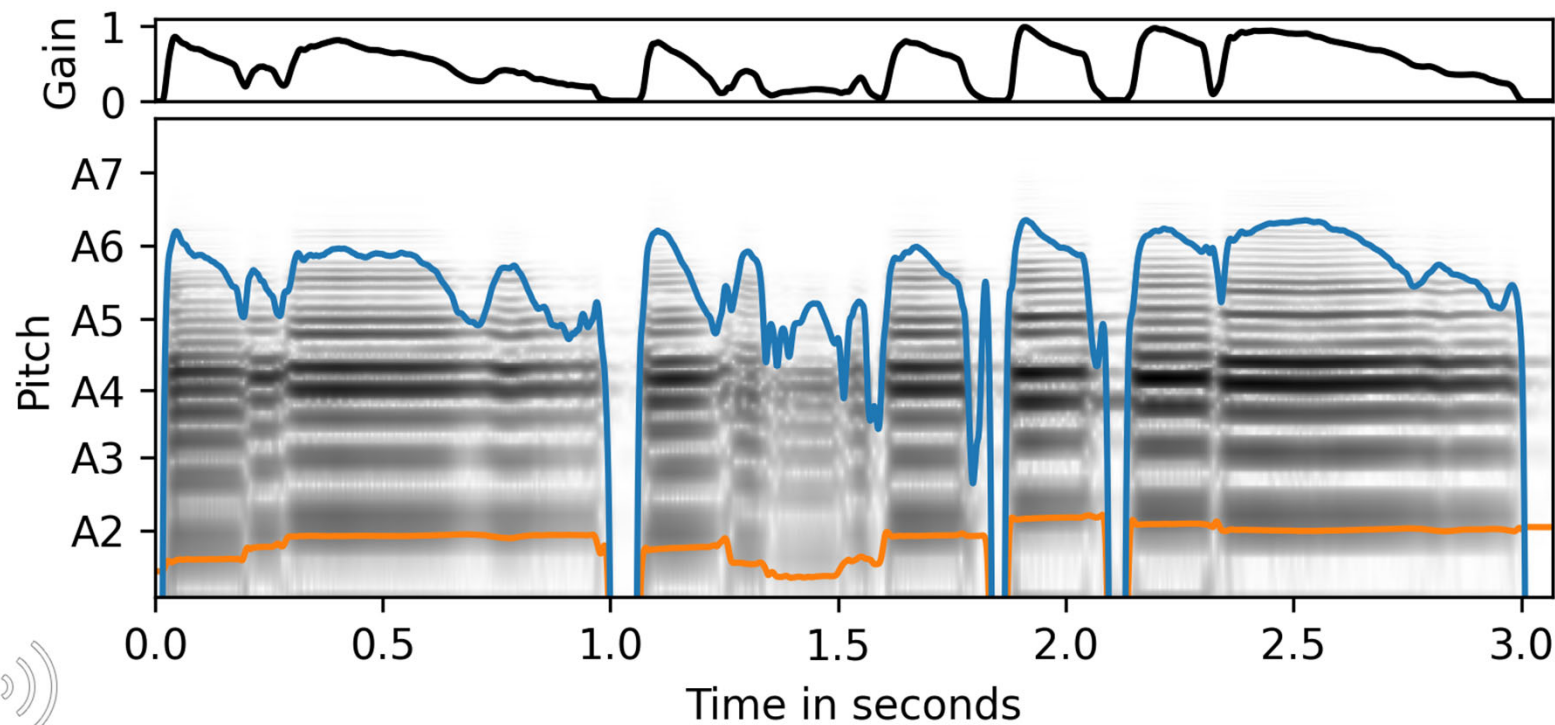
Timbre

Pitch (F0)



Wind Instrument Synthesis

Resynthesized Bassoon (Two Octaves Lower)



Loudness

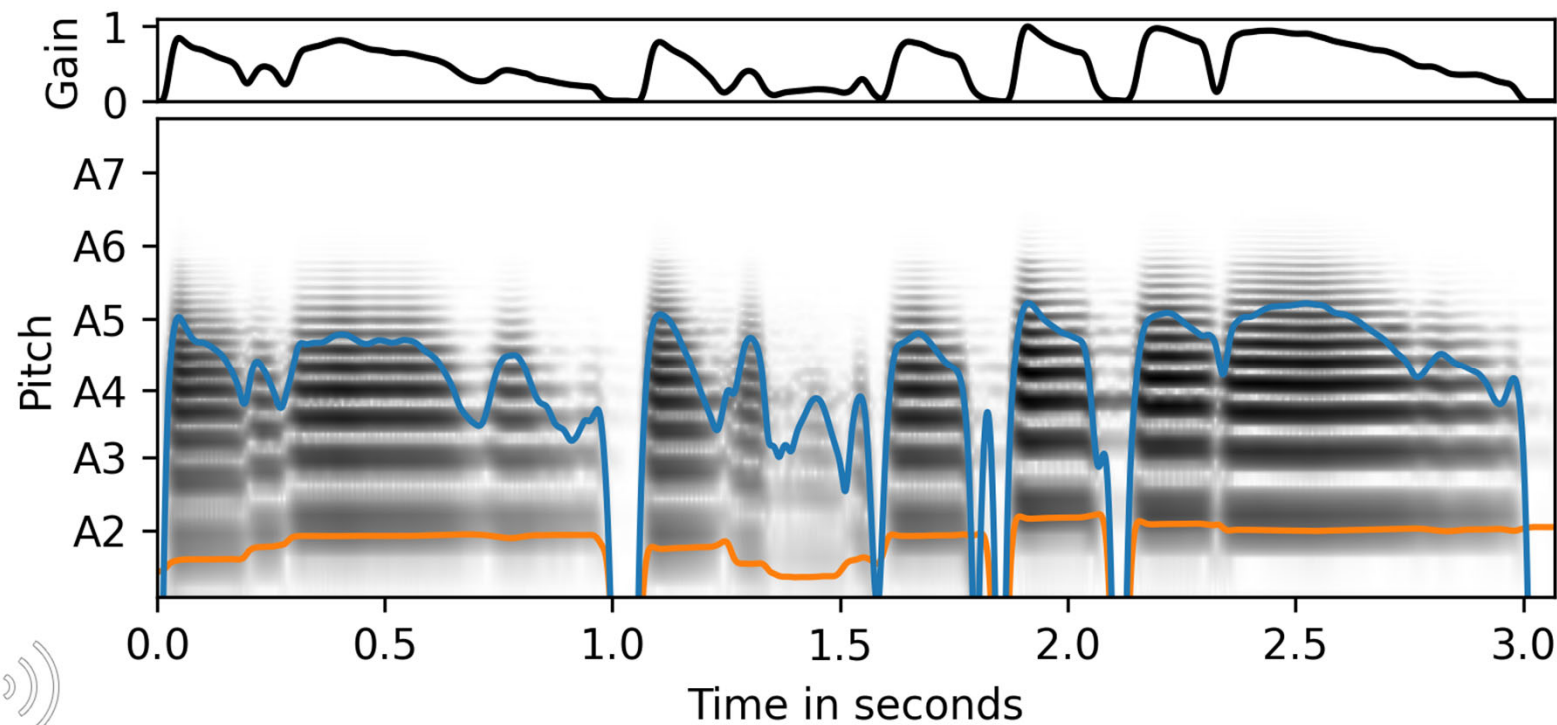
Timbre

Pitch (F0)



Wind Instrument Synthesis

Resynthesized Tuba (Two Octaves Lower)



Loudness

Timbre

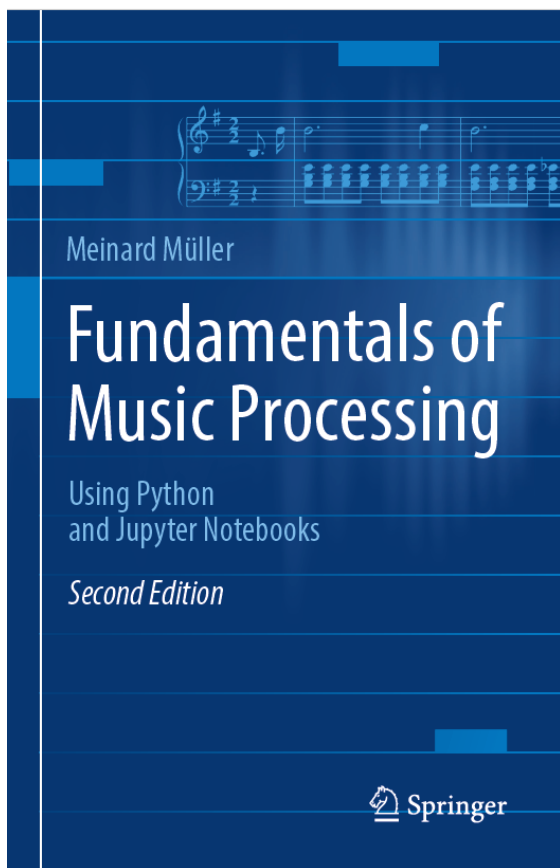
Pitch (F0)



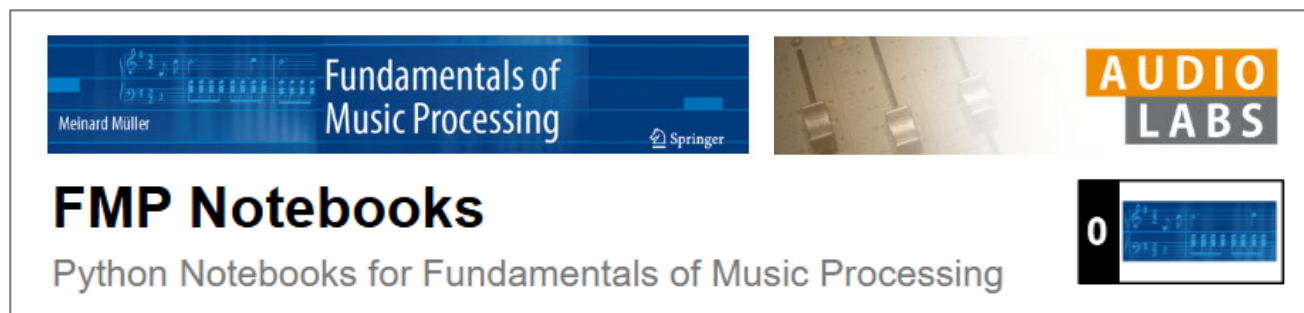
Fundamental Questions in Music Processing

- What are musical representations?
- What are waveforms and digital signals?
- What does the Fourier transform reveal?
- What does the FFT compute?
- What is a spectrogram?
- Which signal properties are important for music?
 - How are frequency and pitch related?
 - How are intensity and loudness related?
 - What are harmonics, overtones, and timbre?
 - How can playing techniques such as vibrato be described?
- What constitutes a musical interpretation?

Fundamentals of Music Processing (FMP)



Meinard Müller
Fundamentals of Music Processing
Using Python and Jupyter Notebooks
2nd edition, Springer, 2021



<https://www.audiolabs-erlangen.de/FMP>