2023 September 24

Deep Learning for Music Analysis and Generation

Pitch

Music transcription, melody extraction & chord recognition (audio → score)

Yi-Hsuan Yang Ph.D.
yhyangtw@ntu.edu.tw

FMP Notebook

https://www.audiolabs-erlangen.de/resources/MIR/FMP/C1/C1.html

Part	Title	Notions, Techniques & Algorithms	HTML	IPYNB
B	Basics	Basic information on Python, Jupyter notebooks, Anaconda package management system, Python environments, visualizations, and other topics	[html]	[ipynb]
0	Overview	Overview of the notebooks (https://www.audiolabs-erlangen.de/FMP)	[html]	[ipynb]
1	Music Representations	Music notation, MIDI, audio signal, waveform, pitch, loudness, timbre	[html]	[ipynb]
2	Fourier Analysis of Signals	Discrete/analog signal, sinusoid, exponential, Fourier transform, Fourier representation, DFT, FFT, STFT	[html]	[ipynb]
3	Music Synchronization	Chroma feature, dynamic programming, dynamic time warping (DTW), alignment, user interface	[html]	[ipynb]

Part	Title	Notions, Techniques & Algorithms	HTML	IPYNB
4	Music Structure Analysis	Similarity matrix, repetition, thumbnail, homogeneity, novelty, evaluation, precision, recall, F-measure, visualization, scape plot	[html]	[ipynb]
5	Chord Recognition	Harmony, music theory, chords, scales, templates, hidden Markov model (HMM), evaluation	[html]	[ipynb]
6	Tempo and Beat Tracking	Onset, novelty, tempo, tempogram, beat, periodicity, Fourier analysis, autocorrelation	[html]	[ipynb]
7	Content-Based Audio Retrieval	Identification, fingerprint, indexing, inverted list, matching, version, cover song	[html]	[ipynb]
8	Musically Informed Audio Decomposition	[illegible] signal reconstruction, instantaneous frequency, fundamental frequency (F0), trajectory, nonnegative matrix factorization (NMF)	[html]	[ipynb]

ISMIR 2018& 2021 Tutorials

https://rachelbittner.weebly.com/tutorials-and-courses.html

Tutorials

Programming MIR Baselines from Scratch: Three Case Studies 2021

International Society for Music InformationRetrieval (ISMIR) conference

- Part 1: Transcription with NMF (Ethan Manilow)
- Part 2: Pitch Tracking with pytorch (**Rachel Bittner**)
- Part 3: Instrument Classification with OpenL3 & Tensorflow (Mark Cartwright)

📺 See the recording here.

Fundamental Frequency Estimation in Music 2018

International Society for Music Information Retrieval (ISMIR) conference

- Part 1: Pitch (Alain de Cheveigné)
- Part 2: Polyphonic fundamental frequency estimation (**Rachel Bittner**)
- Part 3: Applications (Johana Devaney)

Outline

- **Melody extraction**
- Chord recognition
- Multi-pitch estimation
- Transcription

Melody Extraction vs. Note Transcription

- **Melody extraction**: F0 (can reflect *overshoot*, *vibrato*, *glissando*, etc)
- **Note transcription**: Note pitch (quantized in frequency)

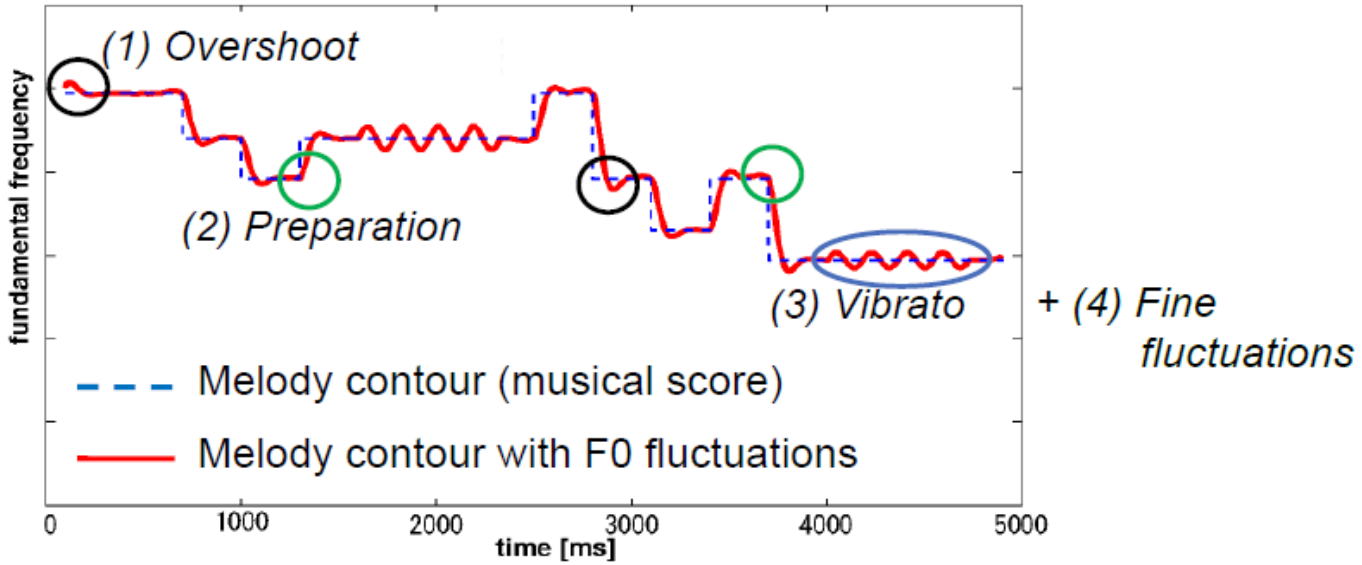
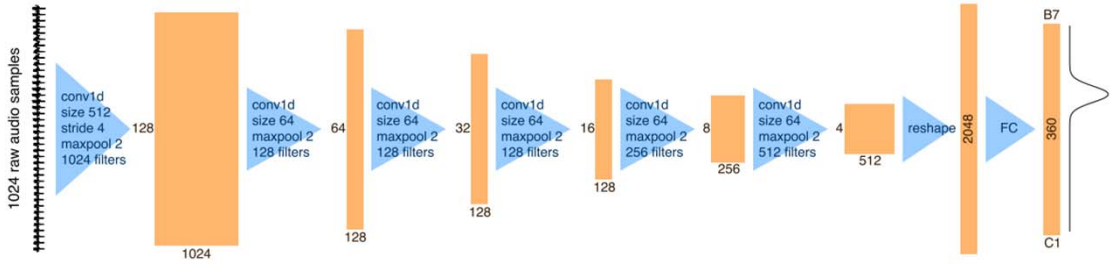
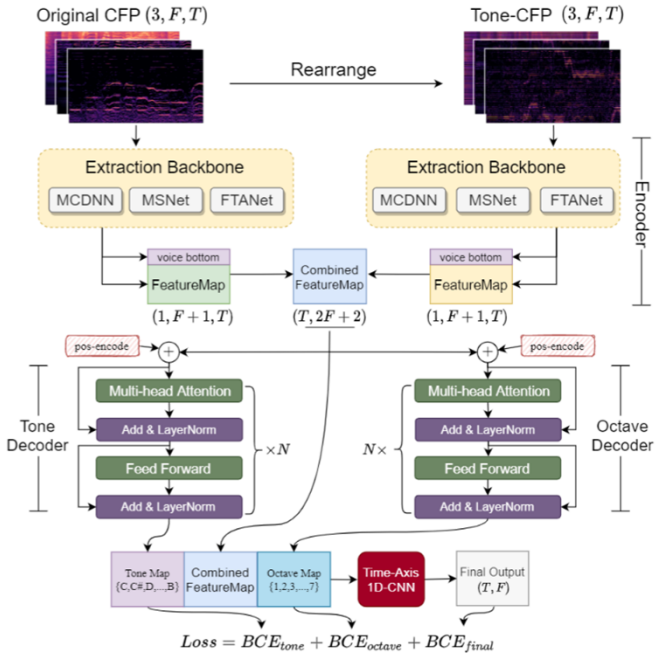


Figure from: Saitou et al, "Speech-to-singing synthesis: converting speaking voices to singing voices by controlling acoustic features unique to singing voices," WASPAA 2007

Melody Extraction Tools

- DSP-based
  - **YIN**: https://librosa.org/doc/main/generated/librosa.pyin.html
  - **WORLD**: https://github.com/JeremyCCHsu/Python-Wrapper-for-World-Vocoder
- DL-based
  - **CREPE** (ICASSP'18): https://github.com/marl/crepe
    - Also used in DDSP (ICLR'20)
  - **TONet** (ICASSP'22): https://github.com/RetroCirce/TONet





Melody Extraction in the Symbolic Domain

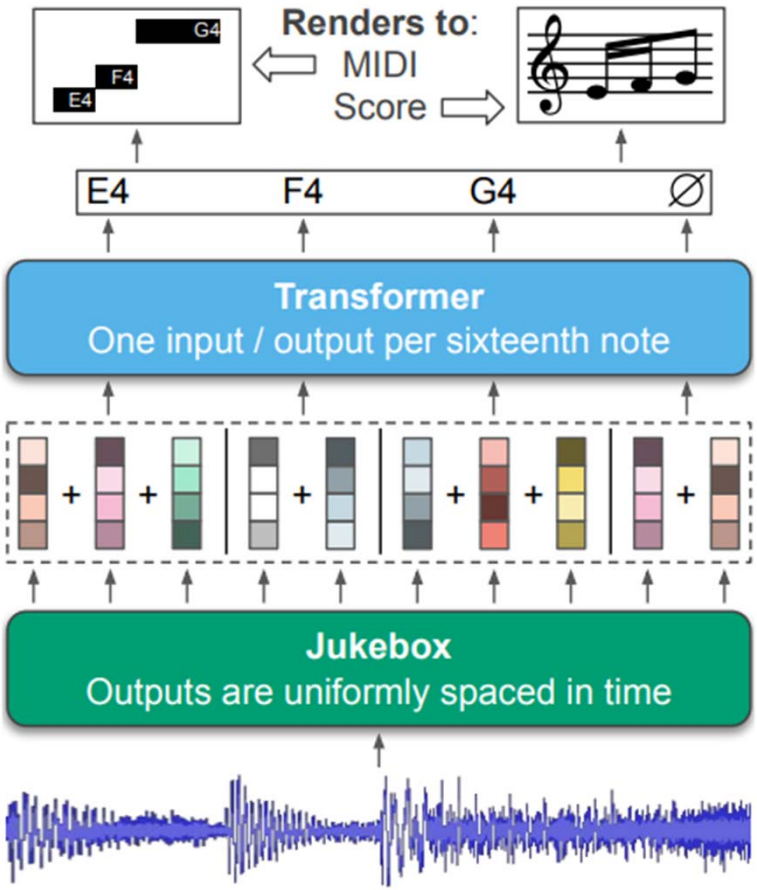
- **Symbolic-domain melody identification**: identify the melody track from multi-track MIDI
  - **Lyrics-informed method**: https://github.com/gulnazaki/lyrics-melody
    - Used in Compose & Embellish (ICASSP'22)
  - **Random forest**: https://github.com/wayne391/midi-track-identification
- **Symbolic-domain melody extraction** from a single track of polyphonic music (e.g., piano)
  - **Skyline algorithm**: https://github.com/wazenmai/MIDI-BERT/tree/CP/melody_extraction/skyline
    - Used in Compose & Embellish (ICASSP'22) and MidiBERT (arXiv'21)
  - **CNN**: https://github.com/sophia1488/symbolic-melody-identification

Outline

- Melody extraction
- **Chord recognition**
- Multi-pitch estimation
- Transcription

Chord Recognition in the Audio Domain

- **Sheet Sage** (ISMIR'22)**:** https://github.com/chrisdonahue/sheetsage
  - For both melody and chord
  - Computationally heavy but pretty accurate



Chord Recognition in the Symbolic Domain

- Rule-based
  - **Chorder**: https://github.com/joshuachang2311/chorder

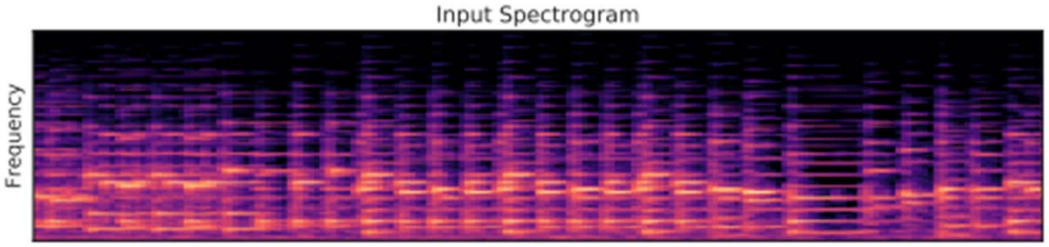
- DL-based

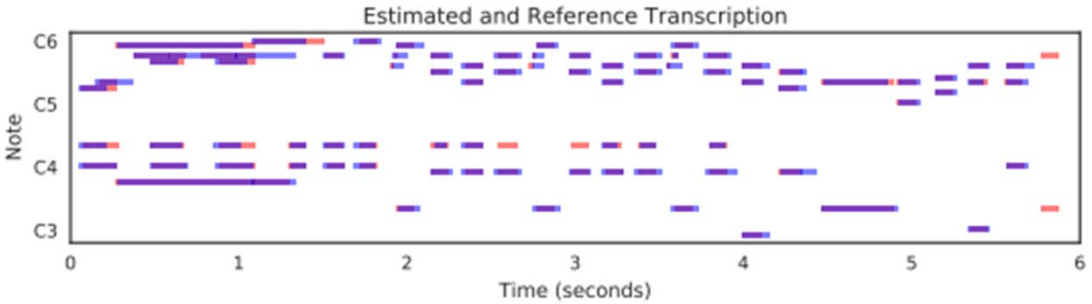
Outline

- Melody extraction
- Chord recognition
- **Multi-pitch estimation**
- Transcription

Multi-pitch Estimation

- For piano
  - **Onset-and-frames** (ISMIR'18): https://github.com/magenta/magenta/tree/main/magenta/models/onsets_frames_transcription
  - **High-resolution Piano Transcription** (arXiv'20): https://github.com/bytedance/piano_transcription





Multi-pitch Estimation

- For being instrument-agnostic
  - **Basic pitch** (ICASSP'22): https://github.com/spotify/basic-pitch (it's lightweight!)

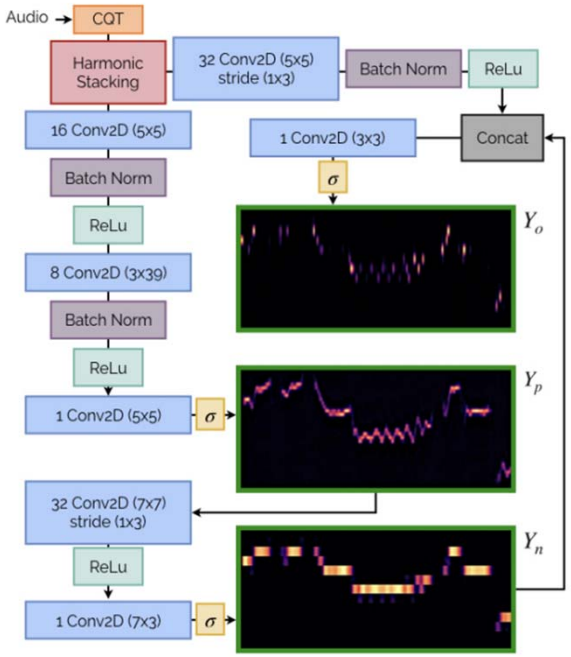
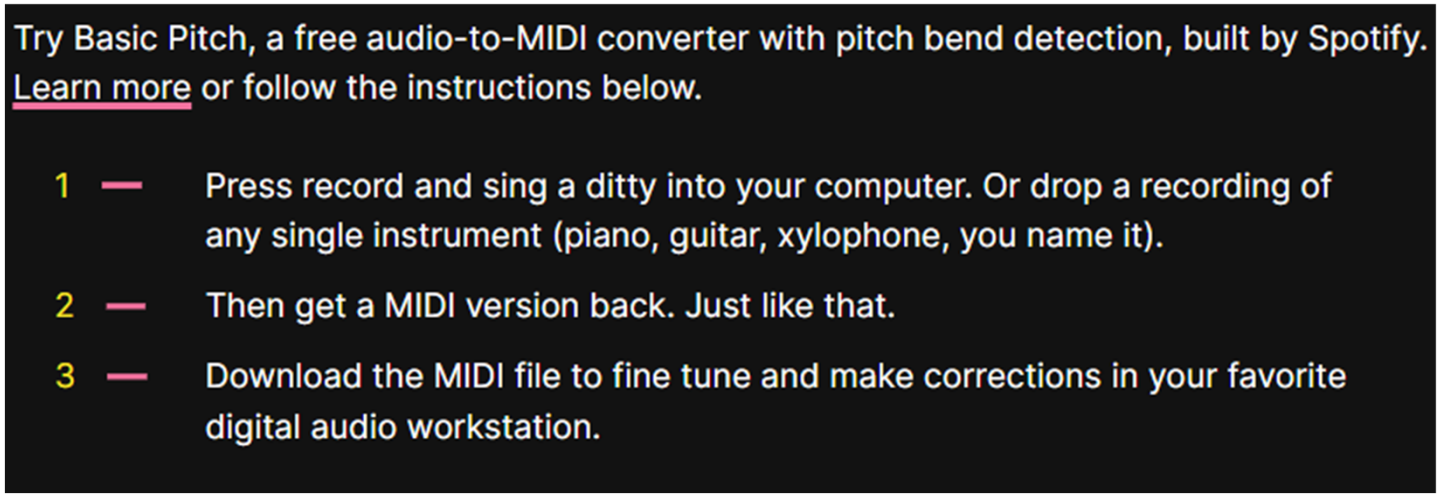


Fig. 1. The NMP architecture. The matrix posteriorgram outputs Y_o, Y_p, and Y_n are outlined in green. σ indicates a sigmoid activation.

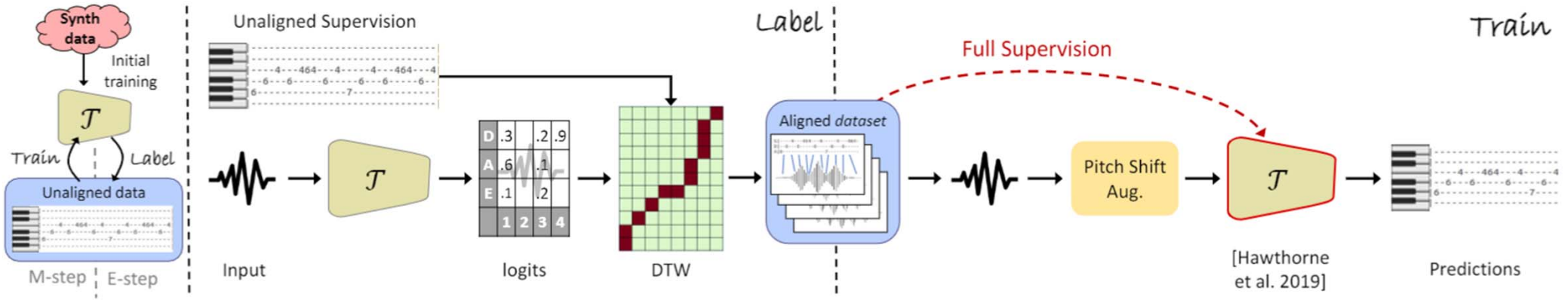


Outline

- Melody extraction
- Chord recognition
- Multi-pitch estimation
- **Transcription**

Multi-instrument Music Transcription

- **Omnizart** (JOSS'21): https://github.com/Music-and-Culture-Technology-Lab/omnizart
- **MT3** (ICLR'22): https://github.com/magenta/mt3
- **Unaligned Supervision for Automatic Music Transcription in The Wild** (ICML'22): https://github.com/benadar293/benadar293.github.io
  - Claimed to outperform MT3

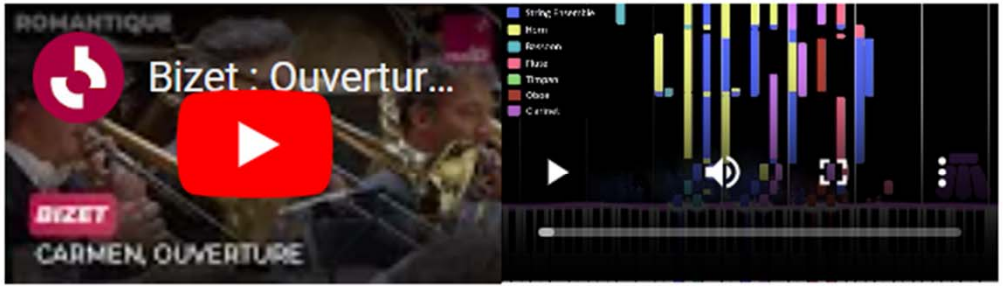


Multi-instrument Music Transcription

https://benadar293.github.io/

Carmen original

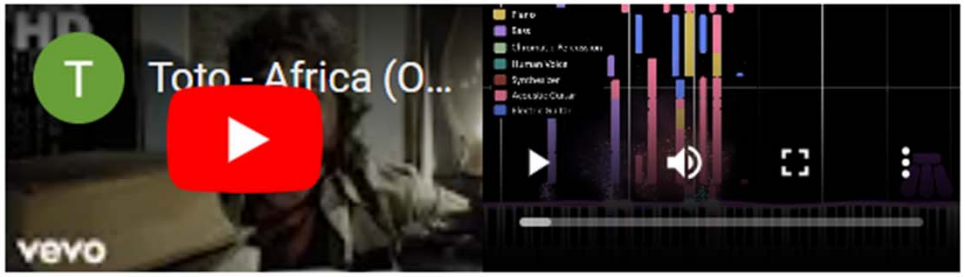
Carmen transcription



Source: https://www.youtube.com/watch?v=jL-Csf1pNCI&ab_channel=FranceMusique

Toto Africa original

Toto Africa transcription



Source: https://www.youtube.com/watch?v=FTQbiNvZqaY&ab_channel=TotoVEVO

Ref: Maman & Bermano, “Unaligned supervision for automatic music transcription in the wild,” ICML 2022