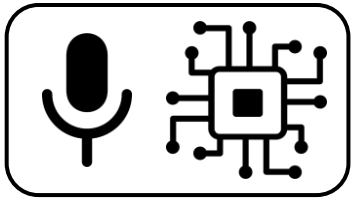
Computational Analysis of Sound and Music

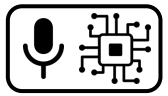


Music Information Retrieval – Music Transcription 1/2

Dr.-Ing. Jakob Abeßer

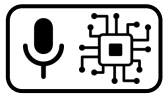
Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de



Outline

- **Pitch Tracking**

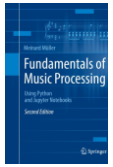


Pitch Detection

Introduction

- Pitch
  - Perceptual sound attribute
  - Allows ordering from low to high in a frequency-related scale

- Pitch detection → Two Sub-tasks



FMP Notebooks

1) Pitch detection

2) Voicing detection

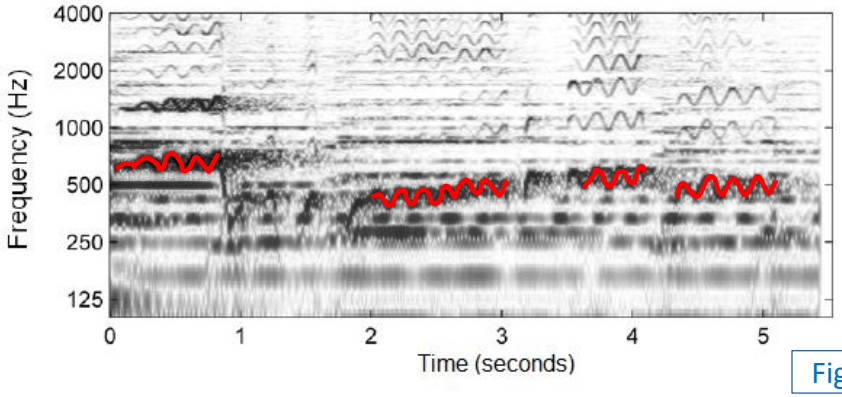


Fig-M3-1

Own

Pitch Detection

Introduction

- Application Scenarios
  - Music Instrument Tuning
  - Music Education
  - Music Transcription
  - Bioacoustics (e.g., bird call recognition)

Fig-M3-2

Fig-M3-3

Fig-M3-4

Pitch Detection

Complexity Levels

- Pitch detection of isolated monophonic instruments

Aud-M3-1

- Predominant melody extraction in polyphonic music

Aud-M3-2

- Polyphonic melody extraction

Aud-M3-3

Increasing Difficulty

Pitch Detection

Traditional Method

- MELODIA [Salamon & Gomez, 2012]
    - Melody extraction from polyphonic music
- Steps
    - Sinusoid Extraction
        - Equal loudness filter
        - STFT
        - Detection of predominant peaks
        - Frequency refinement via instantaneous frequency (IF)

Pitch Detection

Traditional Method

- Steps
  - Harmonic summation
    - Sum over possible harmonic frequencies
    - Frequencies → pitch candidates

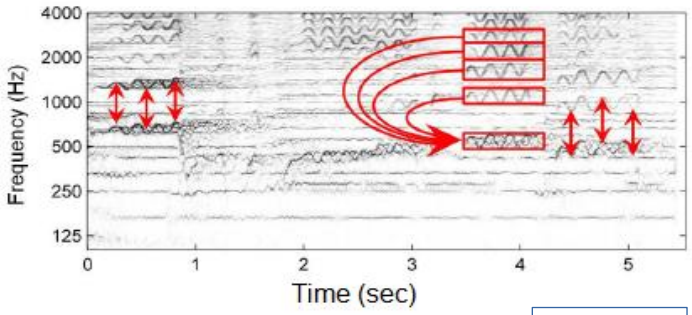
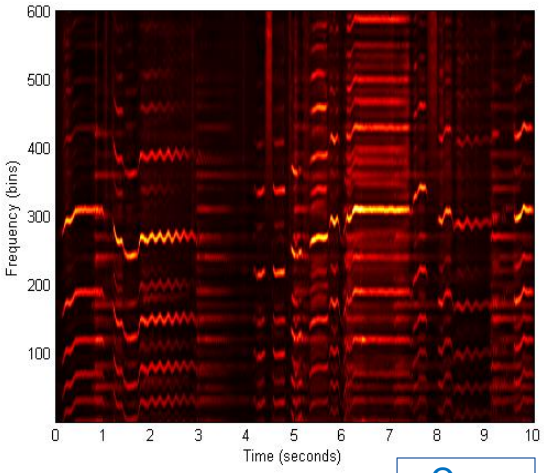


Fig-M3-5

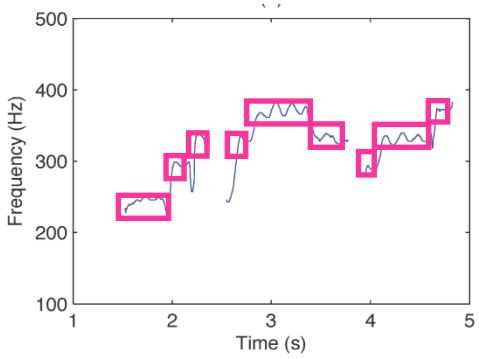


Own

Pitch Detection

Traditional Method

- Steps
  - Pitch contour creation & melody selection
    - Auditory streaming cues → group peaks to continuous paths (pitch contours)
    - Select melody contours using features (e.g. average pitch / salience, vibrato)
    - Note formation (one pitch value)



Pitch contour(s)

↓

Note events

Own

Pitch Detection

Traditional Method

- Melodia plugin available for Sonic Visualiser

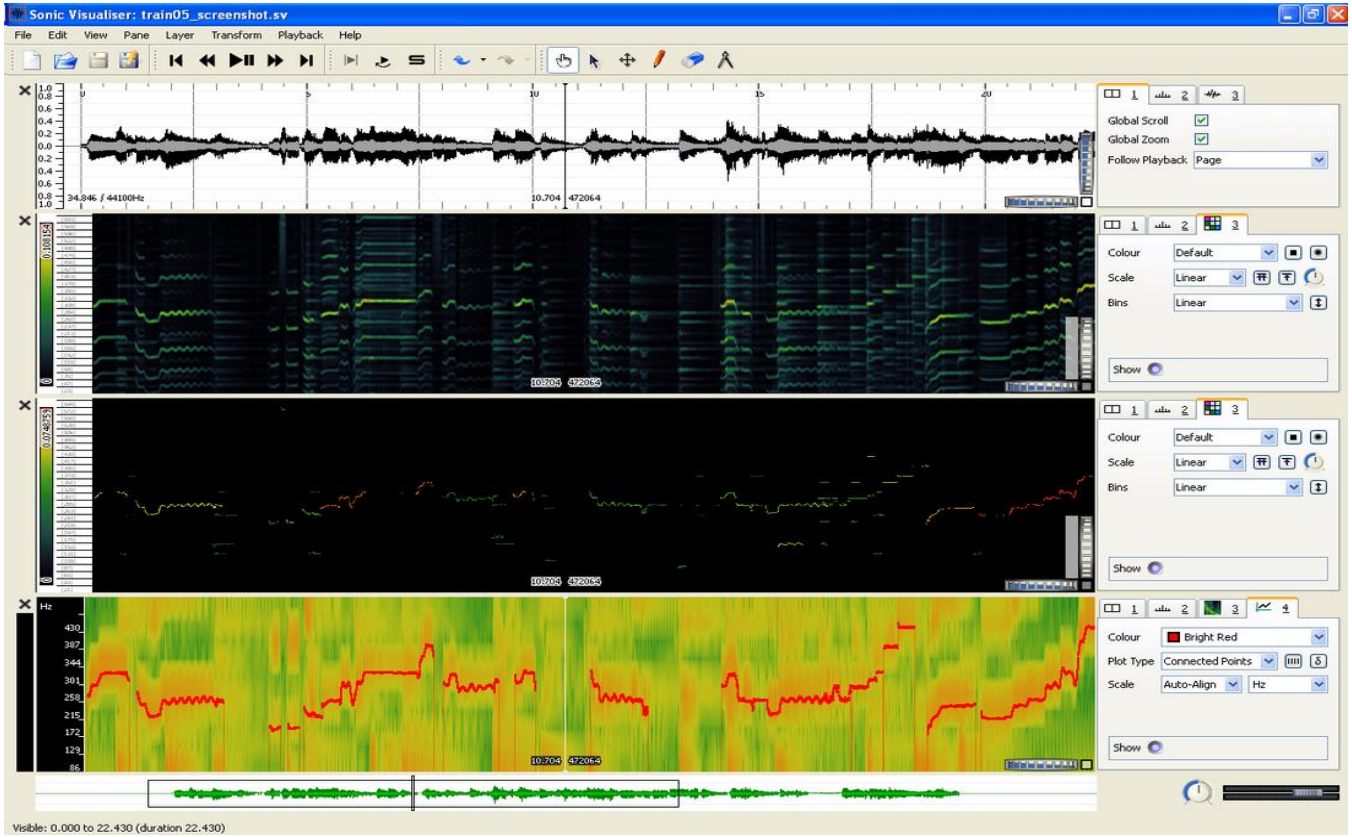


Fig-M3-6

Pitch Detection

DL-based Method

- CREPE (Convolutional Representation for Pitch Estimation) [Kim et al., 2018]
  - Monophonic pitch tracker
  - End-to-end modeling
    - Audio samples → pitch likelihoods
    - 20 cent resolution (5 pitch bins per semitones)

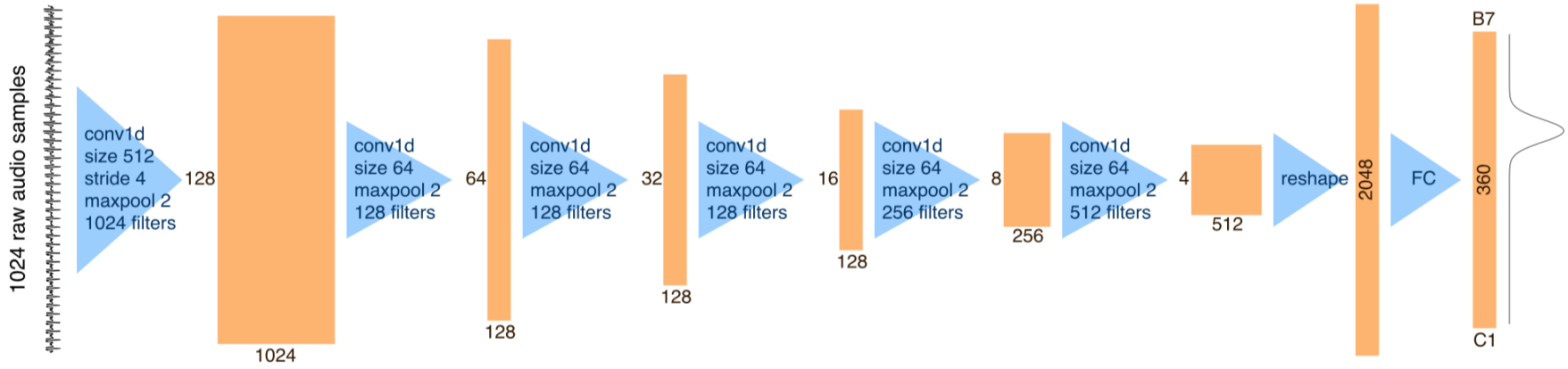


Fig-M3-7

Pitch Detection

DL-based Method

- Auto-encoder structure (U-Net) [Hsieh et al., 2019]
  - Time-frequency representations (2D) → pitch saliency map (2D)
  - (Bottleneck) embedding encodes pitch voicing (melody activity)

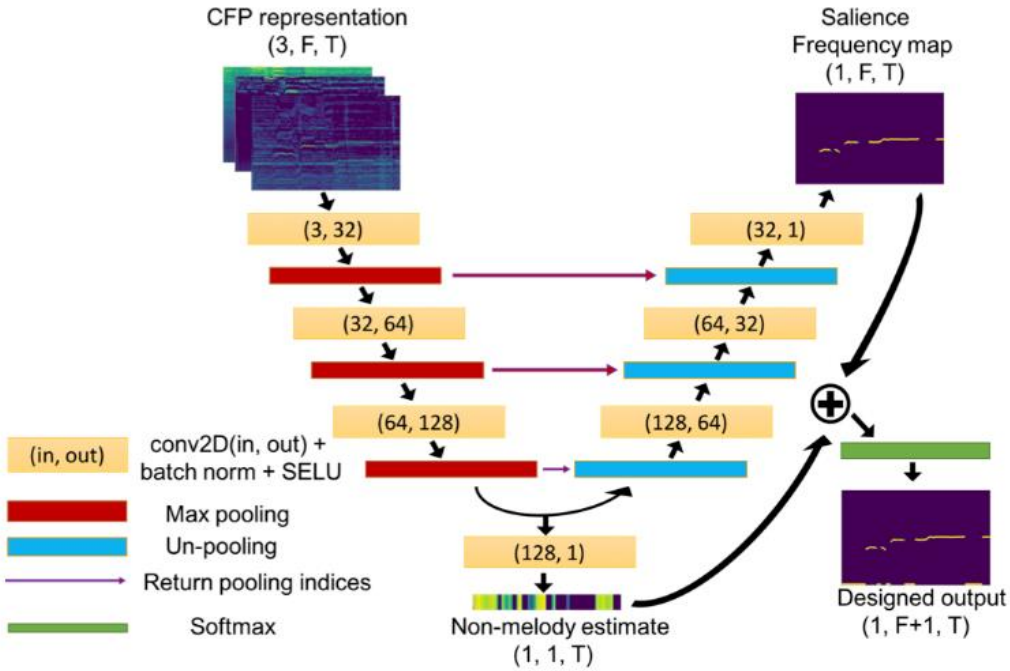


Fig-M3-8

Programming session

Fig-A2-13

References

Images

Fig-M3-1: [Müller, 2015], p. 449, Fig. 8.15(b)

Fig-M3-2: https://unsplash.com/de/fotos/person-die-braune-e-gitarre-spielt-bAHwQEJqAb8

Fig-M3-3: https://cdn2.whatoplay.com/screenshots/2631slide-4.jpg

Fig-M3-4: https://unsplash.com/de/fotos/tierfotografie-flamingoschwarm-ub1sSvJ_Tbs

Fig-M3-5: [Müller, 2015], p. 449, Fig. 8.15(a)

Fig-M3-6: https://www.upf.edu/documents/8071534/8190069/melodia_vamp_screenshot.png

Fig-M3-7: [Kim et al., 2018], p. 2, Fig. 1

Fig-M3-8: [Hsieh et al., 2019], p. 2, Fig. 2

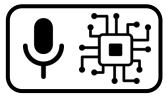
References

Audio

AUD-1: Aislinn – Capclear (2013), https://freemusicarchive.org/music/Aislinn/Aislinn/10_-_Aislinn_-_Capclear

AUD-2: Aislinn – Fourteen Days (2013), https://freemusicarchive.org/music/Aislinn/Aislinn/11_-_Aislinn_-_Fourteen_days

AUD-3: Anonymous Choir – Amicus Meus (2009), https://freemusicarchive.org/music/Anonymous_Choir/Toms_Luis_de_Victorias_Amicus_Meus/Amicus_Meus



References

References

Müller, M. (2021). Fundamentals of Music Processing - Using Python and Jupyter Notebooks (2nd ed.). Springer.

Kim, J. W., Salamon, J., Li, P., & Bello, J. P. (2018). Crepe: A Convolutional Representation for Pitch Estimation. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 161–165. New Orleans, USA.

Salamon, J., & Gomez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. IEEE Transactions on Audio, Speech and Language Processing, 20(6), 1759–1770.

Hsieh, T. H., Su, L., & Yang, Y. H. (2019). A Streamlined Encoder/Decoder Architecture for Melody Extraction. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 156–160. Brighton, UK.