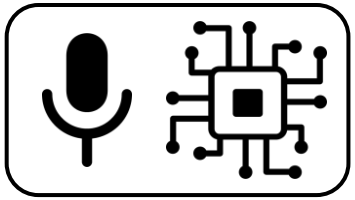
Computational Analysis of Sound and Music

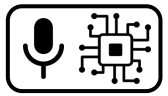


Music Information Retrieval – Rhythmic Analysis

Dr.-Ing. Jakob Abeßer

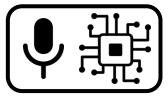
Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de



Outline

- **Introduction**
- Traditional Methods
- DL-based Methods



Tempo Detection

Introduction

- Tempo [beats / minute]
  - Frequency with which humans tap along the beat

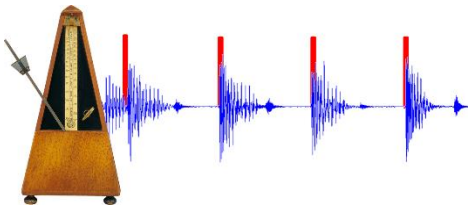


Fig-M1-1

Easy

Aud-M1-1

Harder

Aud-M1-2

- Beat tracking
  - Estimating precise beat positions

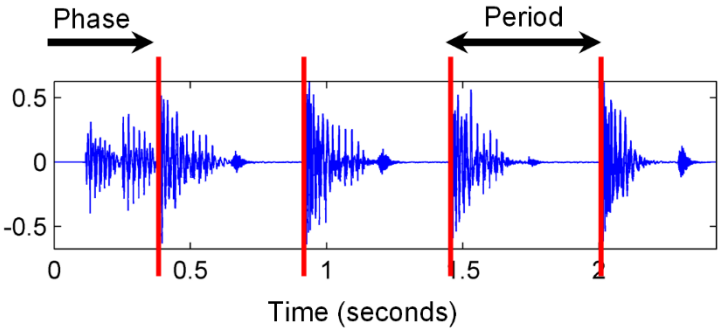
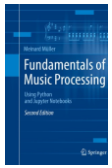


Fig-M1-2

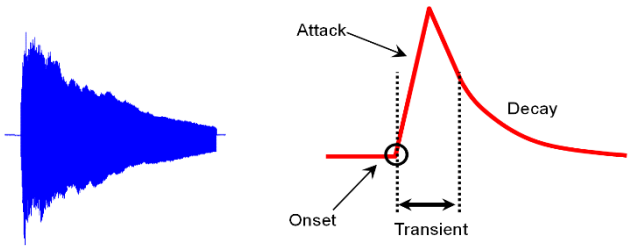


FMP Notebooks

Tempo Detection

Introduction

- Note onsets → note beginning times
  - Clearly defined for plucked string and percussion instruments
  - Ambiguous for wind & brass instruments
- Onset detection
  - Onset detection function
  - Peak picking



Audio samples → Note envelope

Fig-M1-3

Audio samples

↓

Onset detection function & peaks

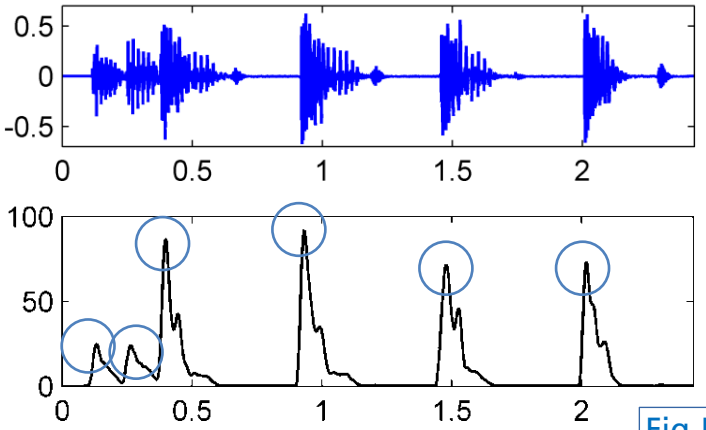
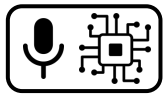


Fig-M1-4

Outline

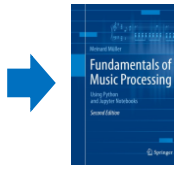
- Introduction
- **Traditional Methods**
- DL-based Methods



Tempo Detection

Traditional Method

- Predominant local pulse (PLP)
  - Correlation with local (windowed) periodic patterns
- Tempogram [Grosche & Müller, 2011]
  - Local likelihood of different tempo candidates
  - Allows to follow tempo changes (e.g., classical music)



FMP Notebooks

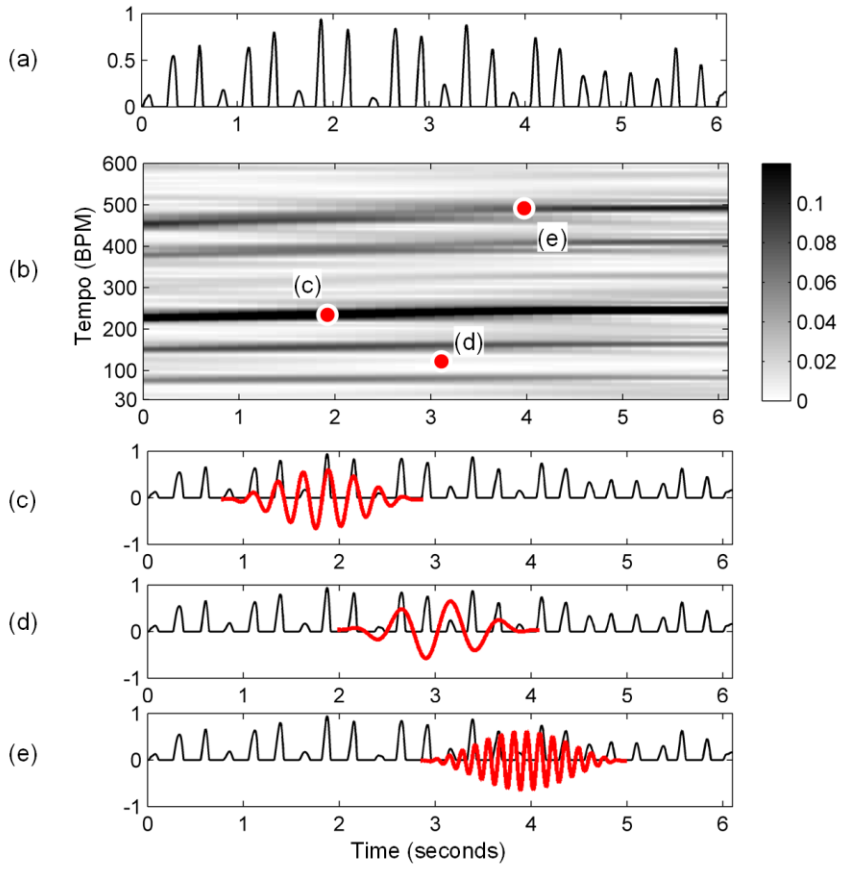
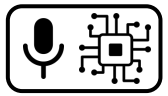


Fig-M1-5

Outline

- Introduction
- Traditional Methods
- **DL-based Methods**



Tempo Detection

DL-based Method

- Approach [Böck et al., 2015]

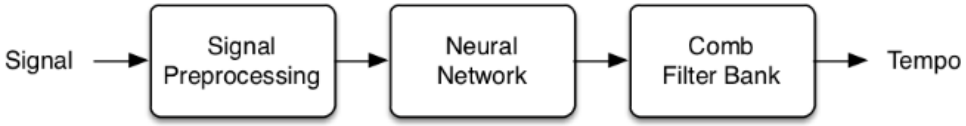
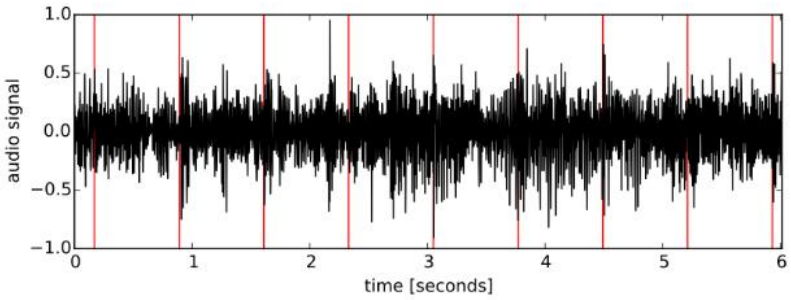
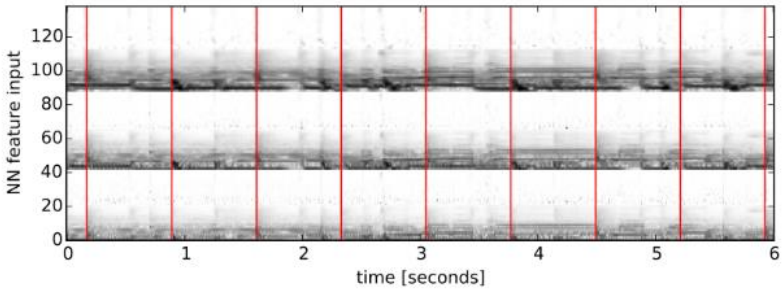


Fig-M1-6



(a) Input audio signal

- Signal representation
  - Stacking of 3 STFT magnitude spectrograms (N=1024, 2048, 4096)
  - Log-amplitude & log-frequency



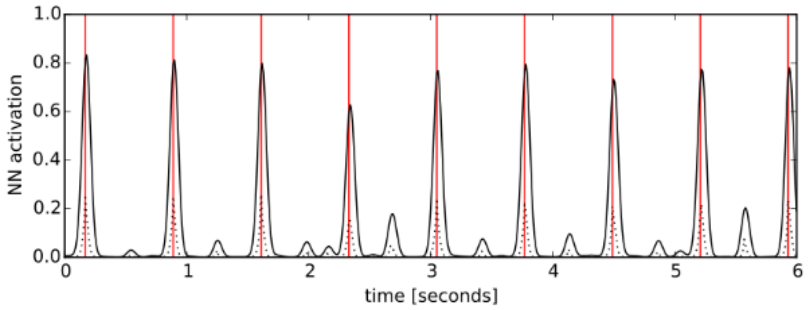
(b) Input to the neural network

Fig-M1-7

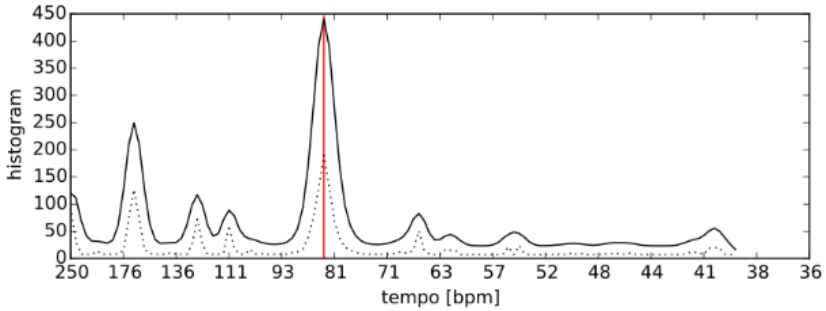
Tempo Detection

DL-based Method

- Neural Network
  - Recurrent (bi-directional LSTM) layer
  - Outputs beat activation function
- Comb filter bank
  - Multiple comb filters → detect periodicities
- Estimate tempo from histogram maximum



(c) Neural network output (beat activation function)



(f) Weighted histogram with summed maxima

Fig-M1-8

Tempo Detection

DL-based Method

- Approach [Schreiber & Müller, 2018]
  - Sample rate ~ 11 kHz, 40-band mel spectrogram
- Main contributions
  - End-to-end tempo without intermediate novelty function
  - 4 multi-filter modules → compress along frequency & find periodicities
  - Dense layers → tempo classification
    - 256 classes: 30 – 285 bpm

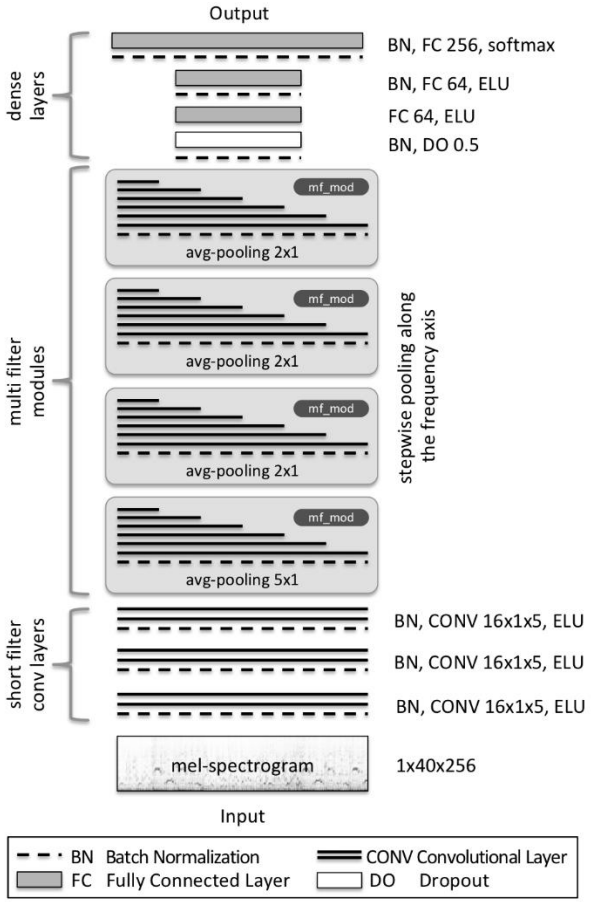


Fig-M1-9

Tempo Detection

DL-based Method

- Temporal Convolutional Networks (TCN) [Davies & Böck, 2019]
  - Use dilated convolution for larger temporal context

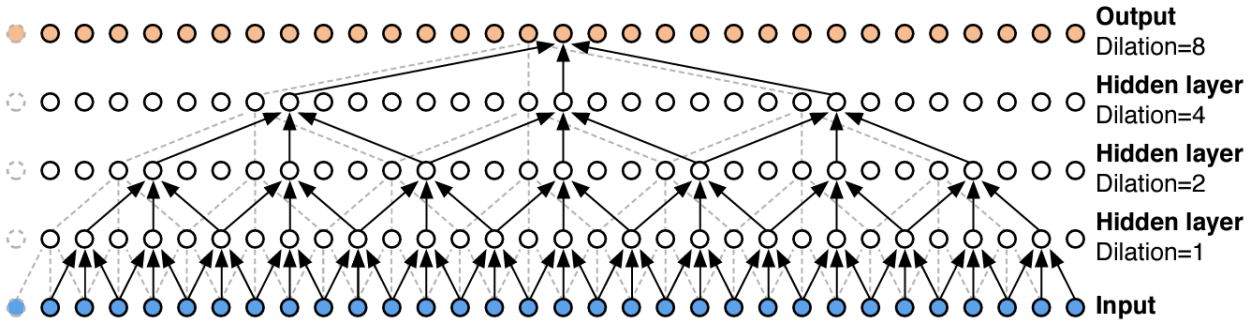
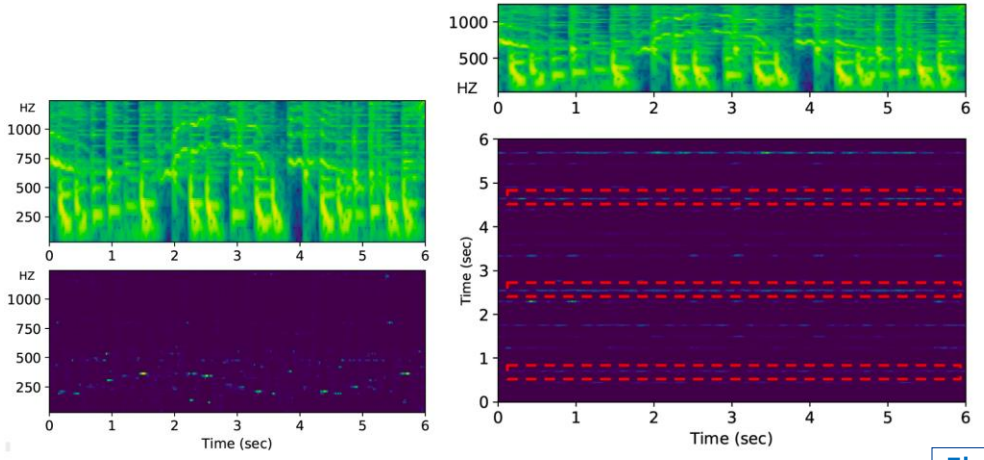


Fig-M1-10

Tempo Detection

DL-based Method

- Transformer Networks [Hung et al., 2022]
  - Joint beat & downbeat tracking
  - Combine TCN & Spectral-Temporal Transformer in Transformer (SpecTNT)
  - Separate encoder allow for separate spectral attention (timbre) & temporal attention (beat / downbeat positions)



(a) Spectral attention (b) Temporal attention

Fig-M1-11

Programming session

Fig-A2-13

References

Images

Fig-M1-1: [Müller, 2021], p. 309, chapter 6 (cover image)

Fig-M1-2: [Müller, 2021], p, 310, Fig. 6.1(b)

Fig-M1-3: [Müller, 2021], p. 311, Fig. 6.2

Fig-M1-4: [Müller, 2021], p. 313, Fig. 6.3(a)&(b)

Fig-M1-5: [Grosche & Müller, 2009], p. 2, Fig. 1(e-g) & p. 3, Fig. 2 (a)

Fig-M1-6: [Böck et al., 2015], p. 2, Fig. 1

Fig-M1-7: [Böck et al., 2015], p. 3, Fig. 2 (a) & (b)

Fig-M1-8: [Böck et al., 2015], p. 3, Fig. 2 (c) & (f)

Fig-M1-9: [Schreiber & Müller, 2018], p. 3, Fig. 2

Fig-M1-10: [Davies & Böck, 2019], p. 4, Fig. 2

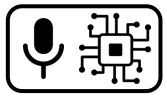
Fig-M1-11: [Hung et al., 2022], p. 403, Fig. 3

References

Audio

Aud-M1-1: orangefreeso..., “Disco Funky Beat”, Website https://freesound.org/people/orangefreesounds/sounds/257993/ , CC BY 3.0 licence, 2014.

Aud-A2-2: launchnlyrics, "120_D_OrchestraStab.wav", Website https://freesound.org/people/launchnlyrics/sounds/631756/, CC BY 4.0 licence, 2022.



References

References

Böck, S., Krebs, F., & Widmer, G. (2015). Accurate Tempo Estimation Based on Recurrent Neural Networks and Resonating Comb Filters. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 625-631. Málaga, Spain.

Davies, M., & Böck, S. (2019). Temporal convolutional networks for musical audio beat tracking. Proceedings of the 27th European Signal Processing Conference (EUSIPCO), 1-5, A Coruna, Spain.

Hung, Y. -N. , Wang, J. -C. , Song, X. , Lu, W. -T., & Won, M. (2022) Modeling Beats and Downbeats with a Time-Frequency Transformer. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 401-405. Singapore, Singapore.

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

Schreiber, H., & Müller, M. (2018). A Single-Step Approach to Musical Tempo Estimation using a Convolutional Neural Network. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 98–105. Paris, France.