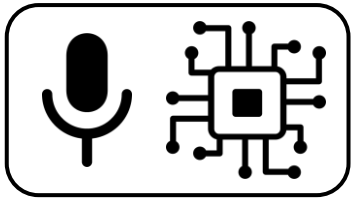
Computational Analysis of Sound and Music

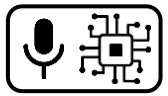


Music Information Retrieval – Music Transcription 2/2

Dr.-Ing. Jakob Abeßer

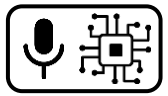
Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de



Outline

- **Drum Transcription**
- Polyphonic Music Transcription



Drum Transcription

Motivation

- Rhythmic Foundation
  - Drum sets and percussion instruments serve as the rhythmic backbone of music
    - Meter
    - Tempo
    - Structure
  - Provide a steady pulse and groove that guides other musicians and engages listeners
- Percussion instruments offer varied timbres and tonal qualities music

Own

Aud-M1-1

Drum Transcription

Vocabulary

- Drum class vocabulary
  - Bass drum (kick drum)
  - Snare drum
  - Hi-hat

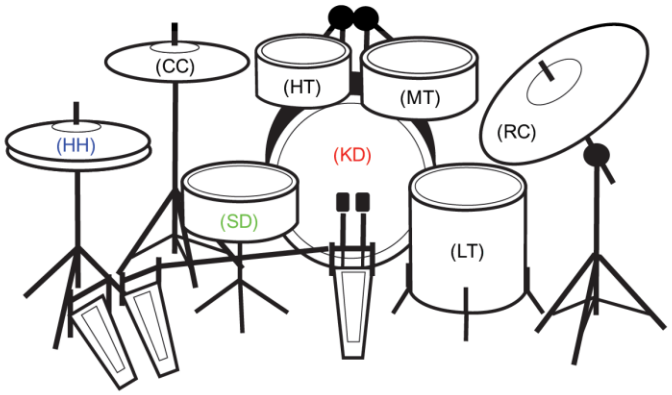


Fig-M4-1

- Sound characteristics

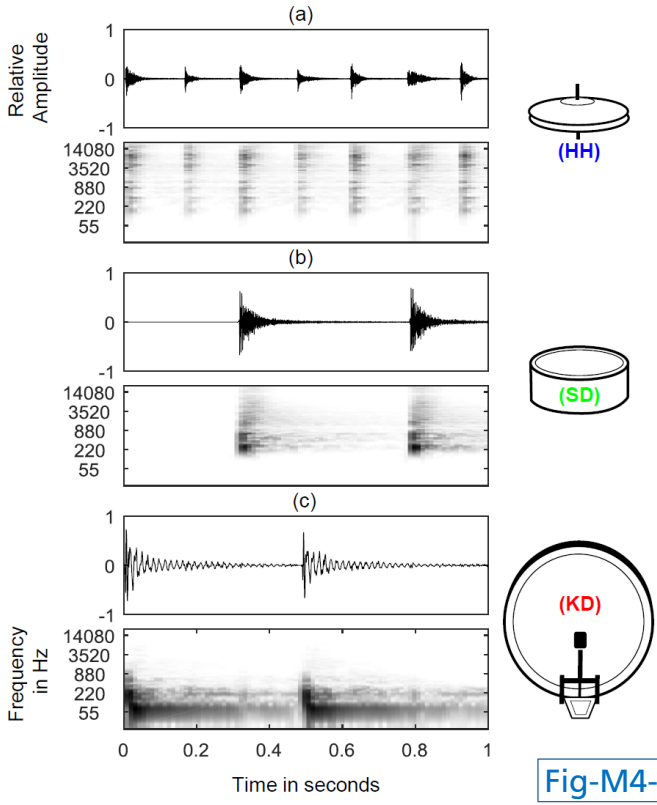
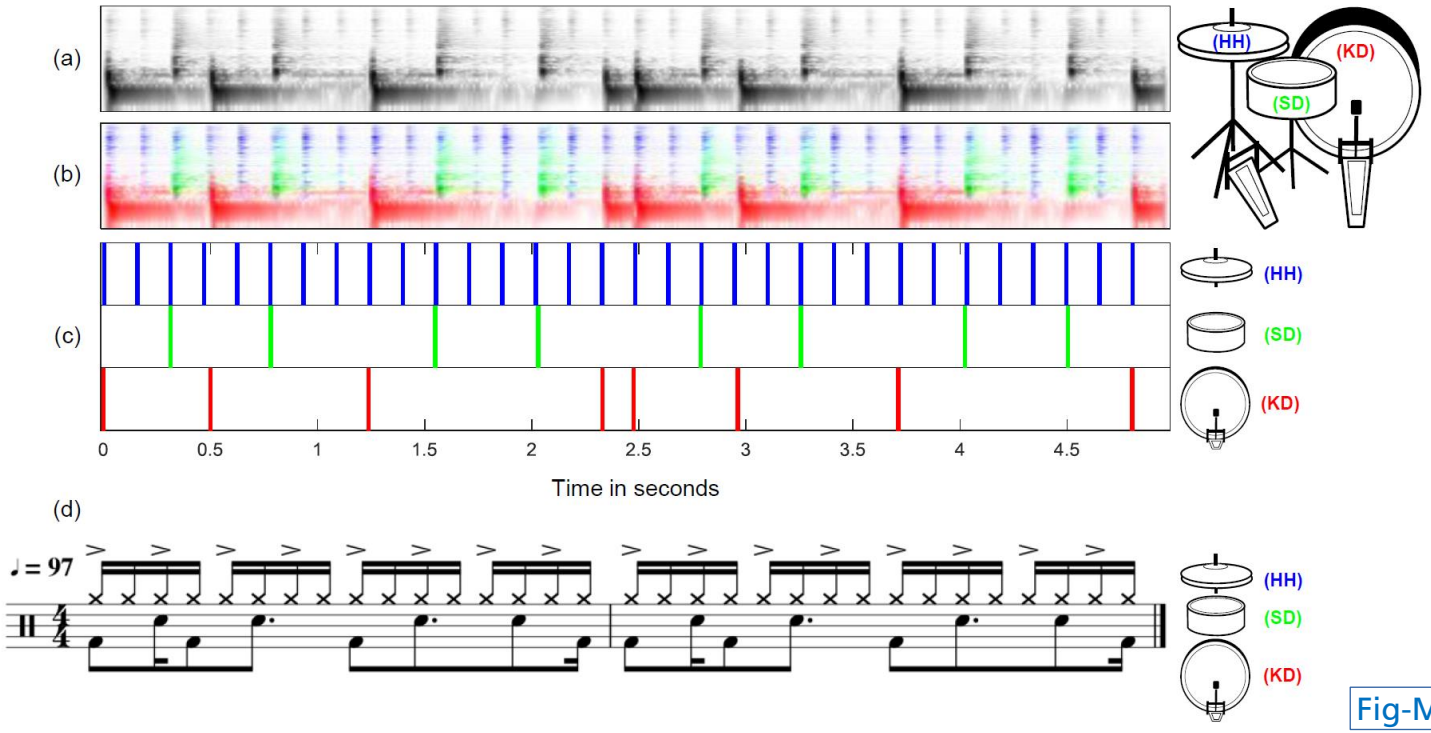


Fig-M4-2

Drum Transcription

Timbre

- Temporal coherence (same metrum)
- Spectral overlap (HH-SN, SN-KD)



Online Demo

Fig-M4-3

Drum Transcription

Traditional Methods

- Building blocks

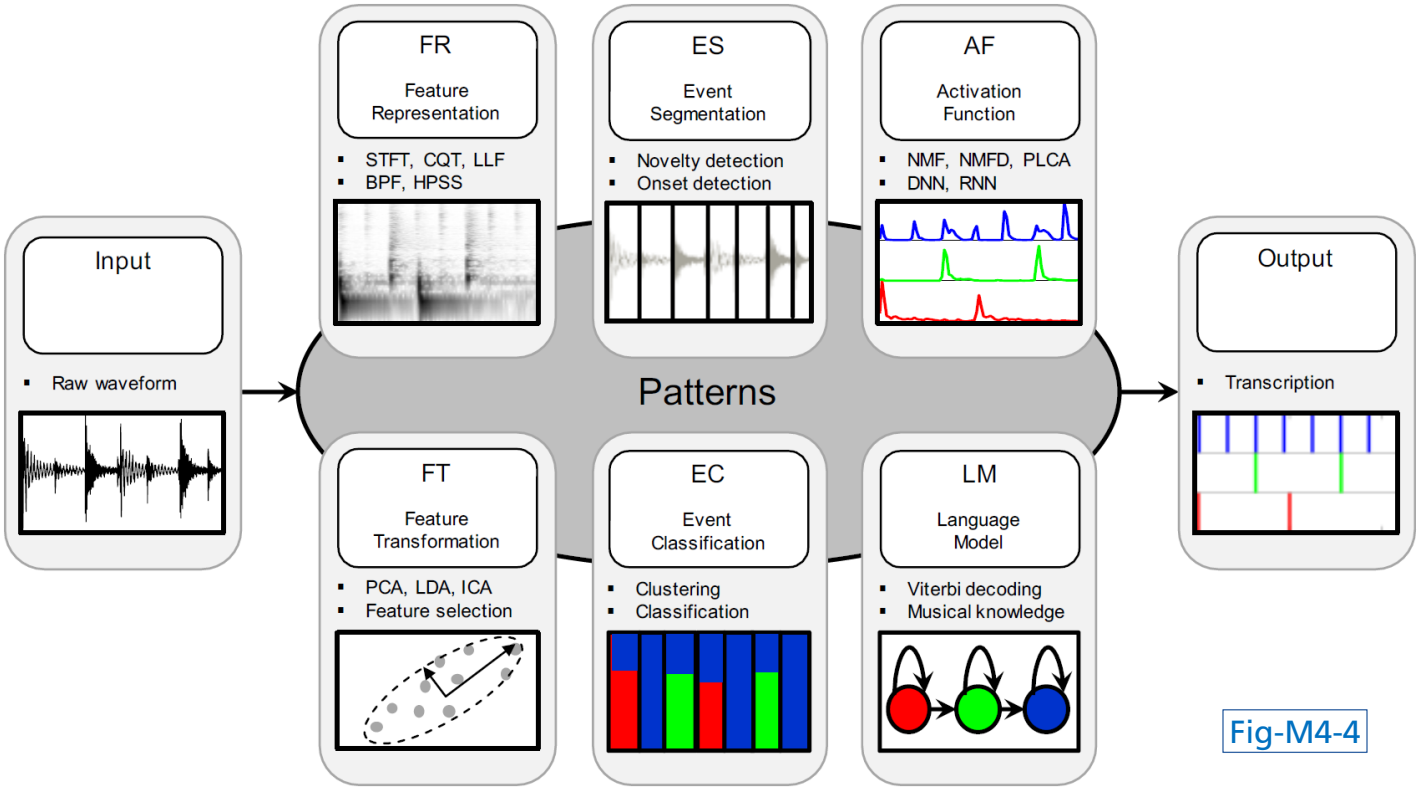


Fig-M4-4

Drum Transcription

Traditional Methods

- Non-negative matrix factorization (NMF)

$$X \approx B \cdot G$$

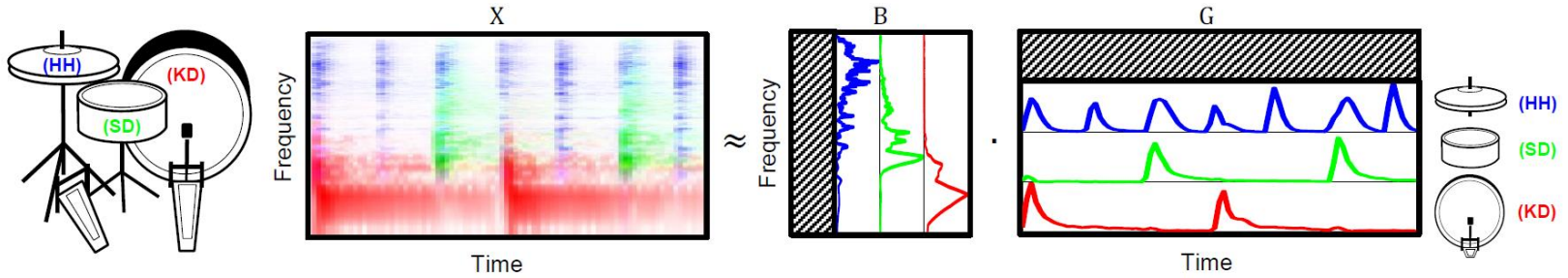


Fig-M4-5

- Non-negative matrix factor deconvolution (NMFD)
  - Convolutive approximation of X

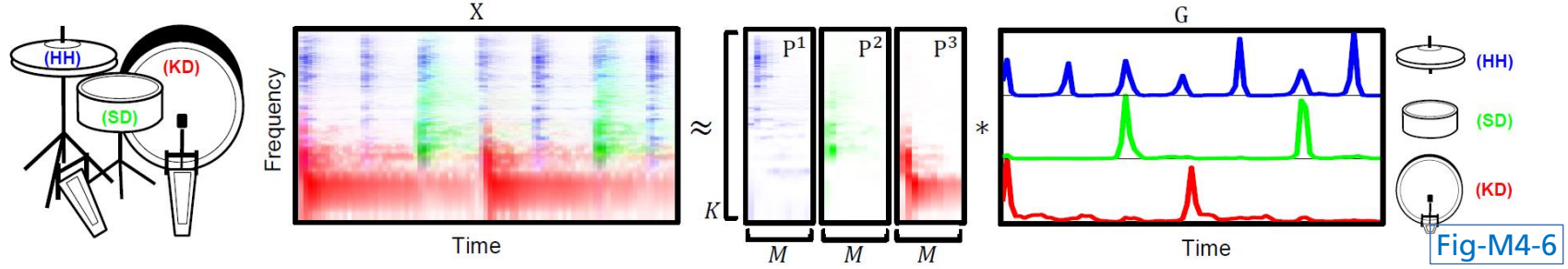


Fig-M4-6

Drum Transcription

DL-based Approach

- Onset-and-Frames (OaF-Drums) Model [Callender, 2020]
  - Joint prediction of note onset & velocity values
  - 12s long LogMel spectrograms (10 ms resolution, 250 Mel frequency bins)
  - Regularization (for better generalization)
    - Dropout (at multiple levels)
    - Mixup (2 random pairs)
    - Shuffled mixup (randomly concatenate 1 s excerpts to 12 s

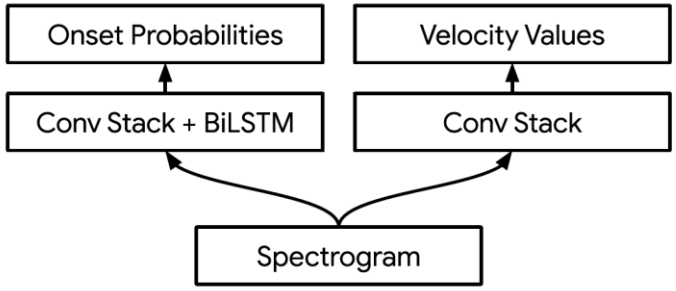


Fig-M4-7

Drum Transcription

DL-based Approach

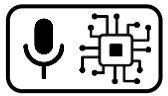
- Network architecture
  - CRNN model
  - Pooling only across frequency (keep time resolution!)
  - Sigmoid output activation function

Layer	Size	Filters	Stride
Log Mel Spectrogram	250 bins		
Conv	16	3x3	1x1
BatchNorm			
Conv	16	3x3	1x1
BatchNorm			
MaxPool		1x2	1x2
Dropout		Keep 25%	
Conv	32	3x3	1x1
BatchNorm			
MaxPool		1x2	1x2
Dropout		Keep 25%	
Dense	256		
Dropout		Keep 50%	
Bidirectional LSTM	64		
LSTM Dropout		Keep 50%	
Dense	88		
Sigmoid Cross Entropy			

Fig-M4-8

Outline

- Drum Transcription
- **Polyphonic Music Transcription**



Polyphonic Music Transcription

Motivation

- Related tasks
  - Multipitch Estimation
    - Frame-wise view → Identify all pitches
  - Streaming (into voices)
  - Polyphonic transcription
    - Further segmentation into note events

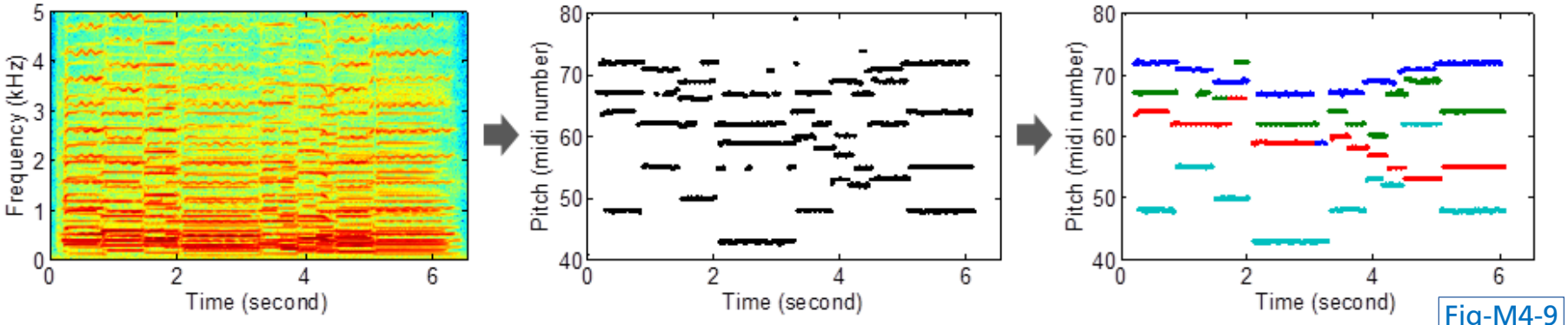


Fig-M4-9

Polyphonic Music Transcription

Traditional Method

- Decomposition with Non-Negative Matrix Factorization (NMF)
  - One basis function per pitch (F0 + harmonics)

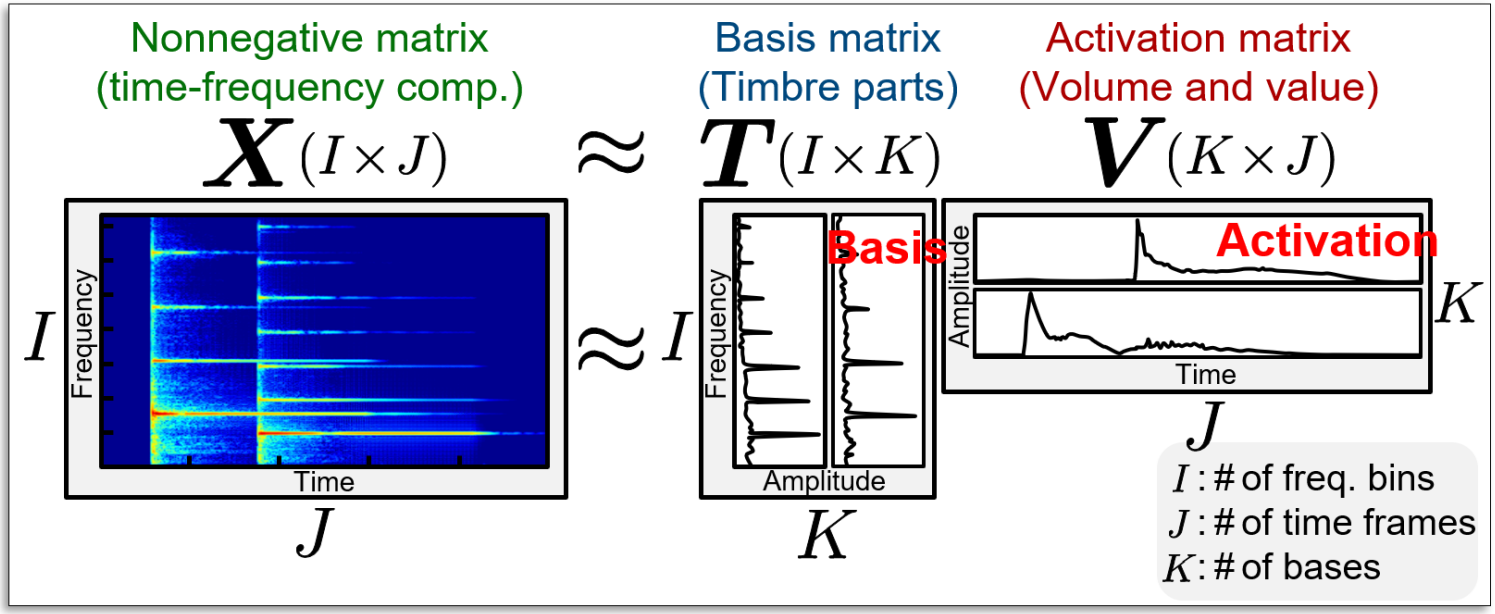


Fig-M4-10

Polyphonic Music Transcription

DL-based Method

- Onset and Frames (OaF) Piano Transcription [Hawthorne, 2018]
  - Separate modelling of note onset times and note pitch values
  - CRNN architecture
  - Onset informs pitch

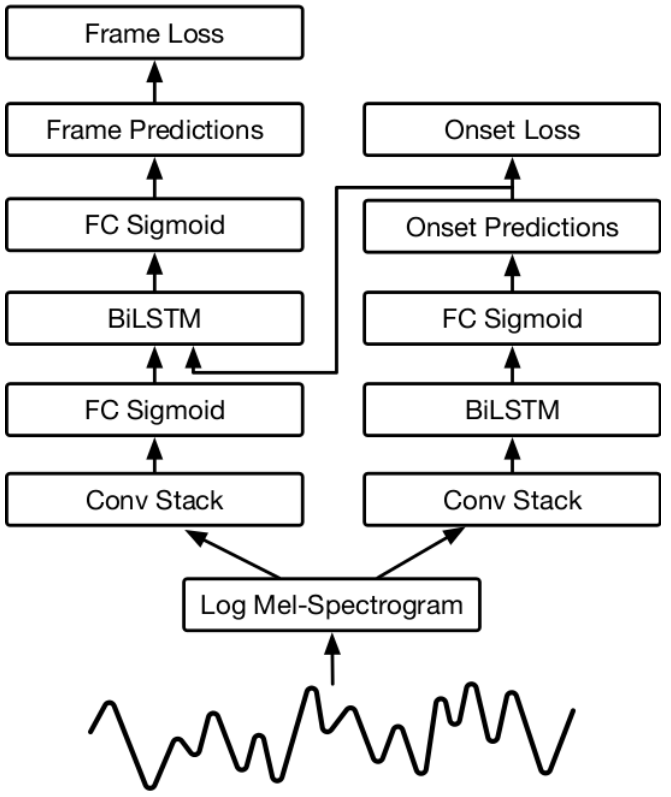


Fig-M4-11

Polyphonic Music Transcription

DL-based Method

- Example

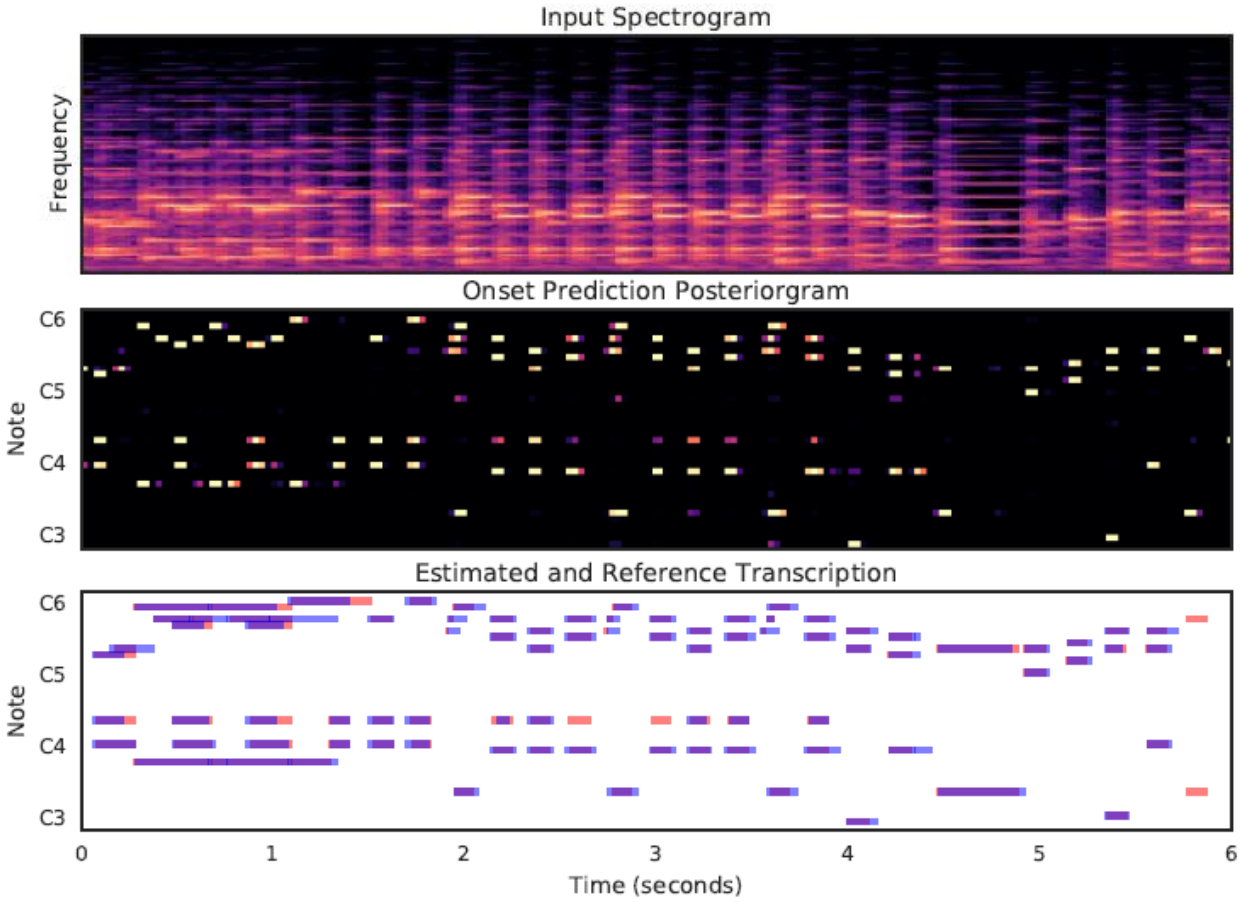


Fig-M4-12

Polyphonic Music Transcription

DL-based Method

- Online Demo

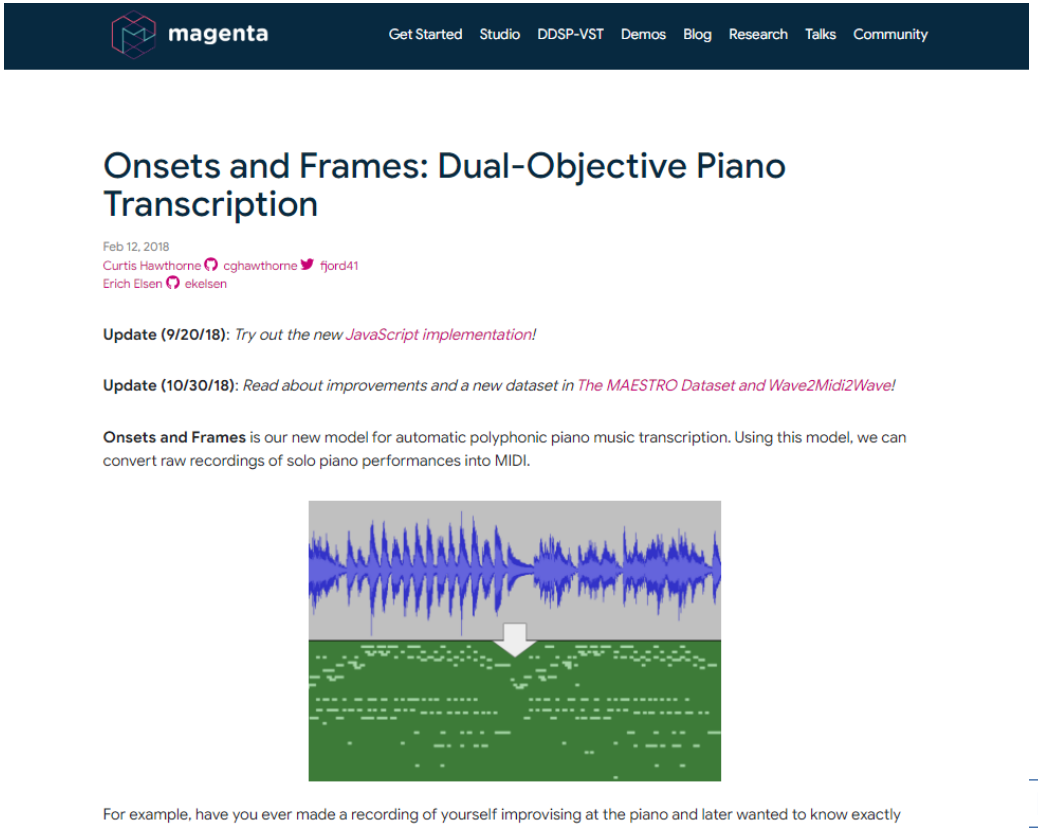


Fig-M4-13

Polyphonic Music Transcription

DL-based Method

- Music Transcription with Transformers [Hawthorne, 2021]
  - Single-instrument or multi-instrument Transcription
  - Model predicts MIDI event tokes (onset, velocity, pitch)

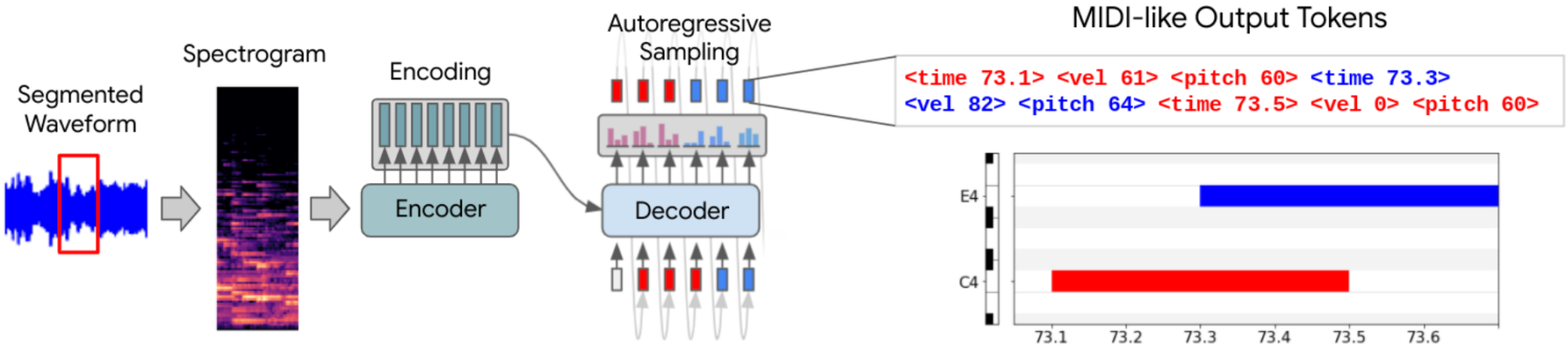


Fig-M4-14

- Online Demo

Programming session

Fig-A2-13

References

Images

Fig-M4-1: [Wu et al., 2018], p. 1, Fig. 1

Fig-M4-2: [Wu et al., 2018], p. 2, Fig. 2

Fig-M4-3: [Wu et al., 2018], p. 3, Fig. 3

Fig-M4-4: [Wu et al., 2018], p. 6, Fig. 4

Fig-M4-5: [Wu et al., 2018], p. 15, Fig. 5

Fig-M4-6: [Wu et al., 2018], p. 16, Fig. 6

Fig-M4-7: [Callender et al., 2020], p. 3, Fig. 1

Fig-M4-8: [Callender et al., 2020], p. 10, Tab. 11

Fig-M4-9: https://labsites.rochester.edu/air/projects/multipitch/MPET.png

Fig-M4-10: http://d-kitamura.net/demo/defNMF/nmf_en.png

Fig-M4-11: [Hawthorne et al., 2018], p. 3, Fig. 1

Fig-M4-12: [Hawthorne et al., 2018], p. 5, Fig. 2

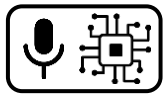
Fig-M4-13: Screenshot https://magenta.tensorflow.org/onsets-frames

Fig-M4-14: https://magenta.tensorflow.org/assets/transcription-with-transformers/architecture_diagram.png

References

Audio

Aud-M4-1: szegvari, "DrumJam Conga Solo Sample Ethno Music Drums 119bpm_2022-07-15_19.12.40.wav", Website https://freesound.org/people/szegvari/sounds/641823/, CC0 1.0 licence, 2022



References

References

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

Wu, C.-W., Dittmar, C., Southall, C., Vogl, R., Widmer, G., Hockman, J., Müller, M., & Lerch, A. (2018). A Review of Automatic Drum Transcription. IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 9, pp. 1457-1483.

Callender, L., Hawthorne, C., & Engel, J. (2020). Improving Perceptual Quality of Drum Transcription with the Expanded Groove MIDI Dataset. arXiv:2004.00188.

Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I., Raffel, C., Engel, J., Oore, S., & Eck, D. Onsets and Frames (2018). Dual-Objective Piano Transcription. arXiv:1710.11153

Hawthorne, C., Simon, I., Swavely, R., Manilow, E., & Engel, J. (2021). Sequence-to-Sequence Piano Transcription with Transformers. Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Online.