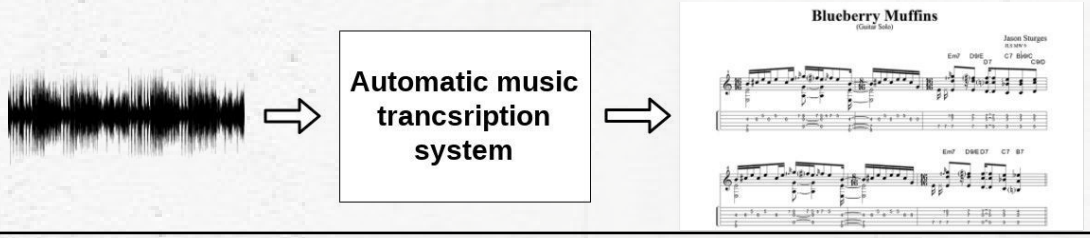
Coversion of Electric Guitar to MIDI

Author: Michal Klčo, Supervisor: Jan Černocký

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Problem

Automatic music transcription (AMT) is process where a music record is converted to a human-readable music notation. This is still challenging task and some automatic music transcription systems are limited to single music instrument to achieve more accurate transcription. The goal of this project was to create several AMT systems for conversion of electric guitar recordings to a MIDI representation based on different machine learning techniques, evaluate and compare them.



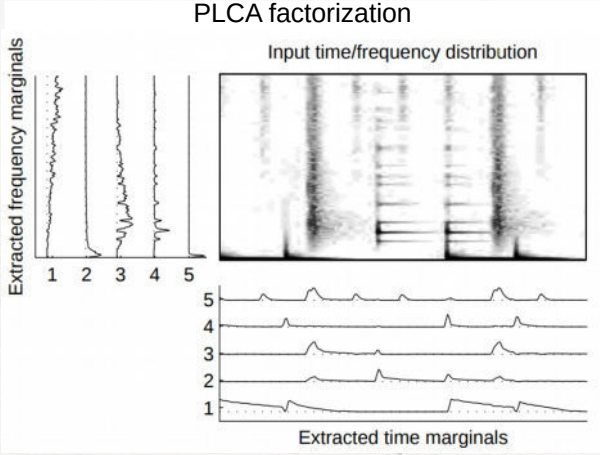
Designed Systems

4 different conversion systems were designed:

- PLCA2MIDI – based on spectrogram factorization technique called Probabilistic Latent Component Analysis (PLCA)
- DNN2MIDI – based on feedforward neural network
- LSTM2MIDI – based on recurrent neural network with Long Short-term Memory (LSTM) units
- HYBRID2MIDI – combination of PLCA and reccurent neural network with
- LSTM units

3 different sepctral analysis were used to create input spectrograms:

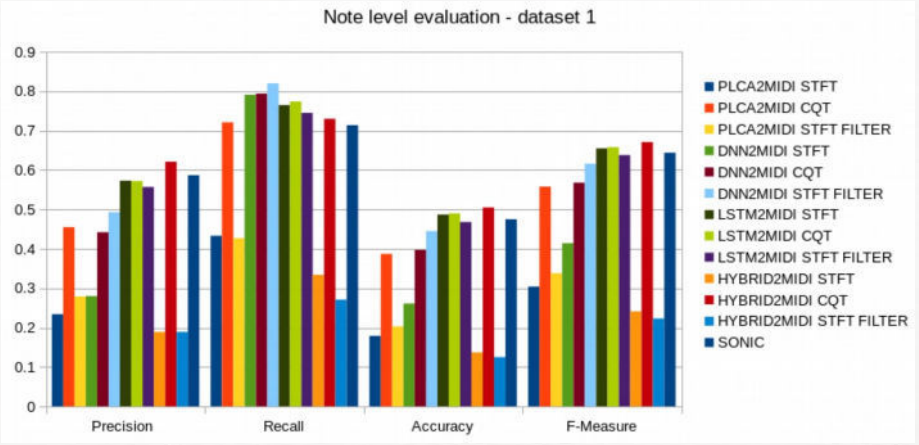
- Short Time Fourier Transform (STFT)
- Constant-Q Transform (CQT)
- STFT spectrogram filtered with
- semitone filterbank



Evaluation

To evaluate the systems, the IDMT-SMT-Guitar database containing electric guitar recordings provided by Fraunhofer IDMT was used. Frame level and note level metrics were calculated from the output. While frame level metrics measure the accuracy of transcription by observing correctly transcribed notes in 100 ms time frames, note level metrics measure the accuracy by observing transcribed notes as a whole. The following values were calculated for each type of metric:

- Precision
- Recall
- Accuracy
- F-Measure



Conclusion

Generally, the results showed that the systems that employ neural networks perform better than systems based only on PLCA. While the system with the DNN had the best results in frame level pitch estimation, the systems that incorporate the LSTM network achieved higher accuracy in note level evaluation due to their capability of temporal modeling. The combination of the neural network and PLCA achieved the highest accuracy in some cases but only with the CQT input spectrogram.