RECURRENT NEURAL NETWORKS FOR DRUM TRANSCRIPTION

Richard Vogl, Matthias Dorfer and Peter Knees

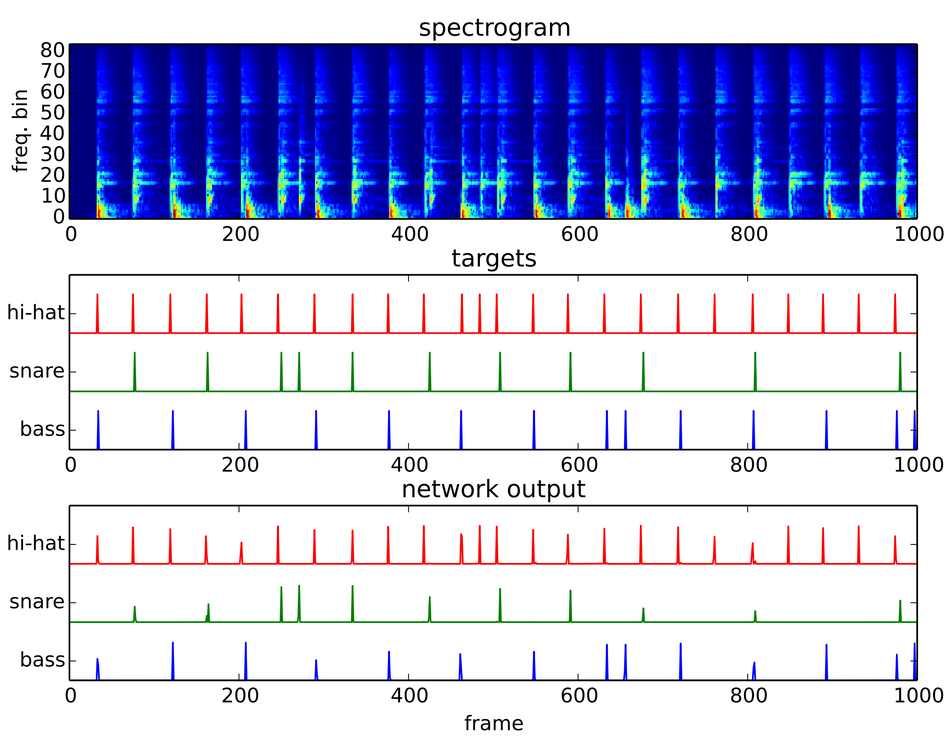
Department of Computational Perception

JOHANNES KEPLER UNIVERSITY LINZ

Department of Computational Perception

1 introduction

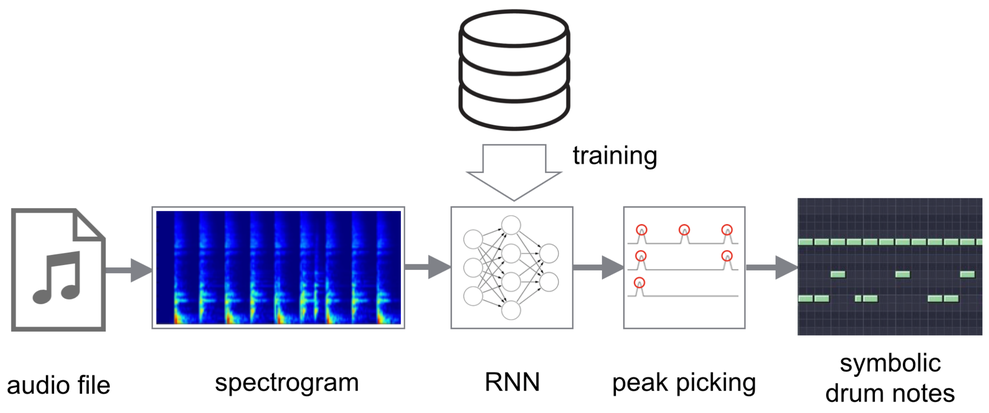
- method for transcription of solo drum audio recordings based on neural networks
- comparison of four RNN architectures for drum transcription
- evaluation on two well-known datasets



3 evaluated architectures

- simple RNN
- backward RNN (bwRNN)
- bidirectional RNN (bdRNN)
- RNN with time shift (tsRNN)

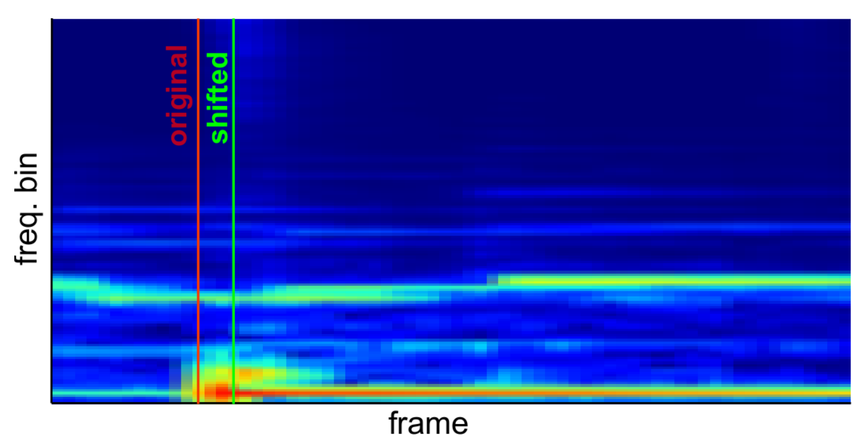
overview of method



2 method

- extraction of a log. 84-bin power spectrogram @ 100Hz
- single layer RNNs with 200 nodes
- output of three detection functions for bass drum, snare drum, and hi-hat
- peak picking yields notes for the three instruments

time shift of labels



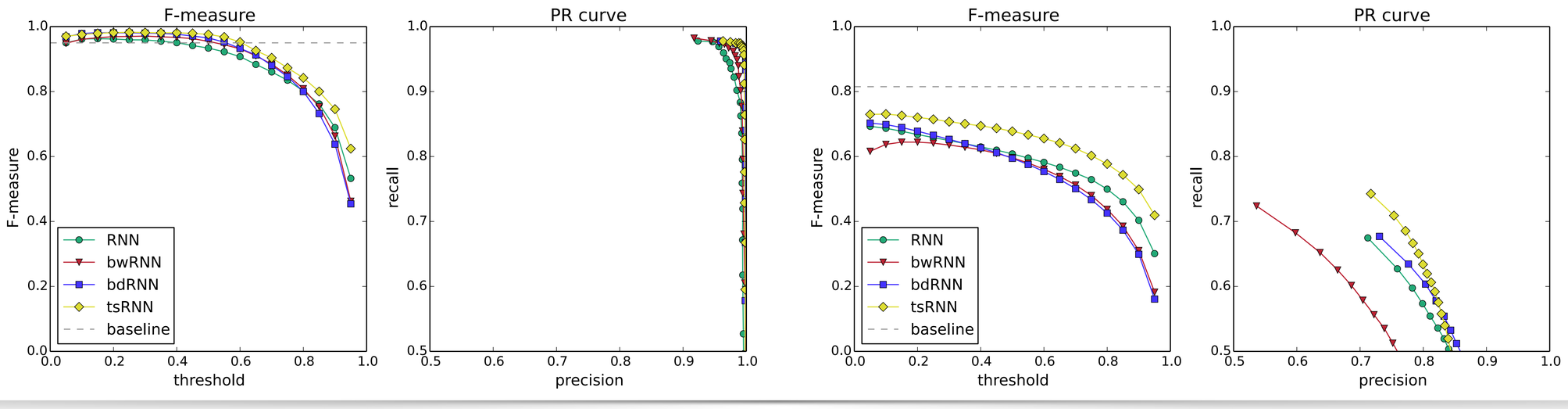
results

Results for IDMT-SMT-Drums

algorithm	best F-measure [%]	at threshold
RNN	96.3	0.15
bwRNN	97.1	0.30
bdRNN	98.1	0.15
tsRNN	**98.2**	0.25
NMF [5]	**95.0**	-

Results for ENST-Drums (trained on SMT)

algorithm	best F-measure [%]	at threshold
RNN	69.3	0.05
bwRNN	64.4	0.15
bdRNN	70.3	0.05
tsRNN	**73.1**	0.10
HMM [24]	**81.5**	-



richard.vogl@jku.at
cp.jku.at

giant_steps

www.giantsteps-project.eu