

# Abstract for RnD

Pranav Sankhe  
Indian Institute of Technology, Bombay

April 20, 2019

## Abstract

In the last semester(6th semester), we worked on NMF based approach aimed at solving the problem of transcription of drum recordings. Having implemented an NMF based system, we envisioned towards working on a (Deep Neural Networks) DNN based system to solve the same problem. The inspiration was fuelled by the impressive results obtained by team magenta on the transcription of Piano recordings. The issues which we saw while going ahead with this particular approach was the huge amount of data required to be able to get an accurate enough transcription system. The scarcity of the data was even more acute when it came to developing a DNN based transcription system for tabla sounds.

To get started we analyzed the state of art drum transcription system which was also based on DNN. The authors claim that the system has the means to utilize information on the rhythmical structure of a song. Their results showed that convolutional feature processing is beneficial for drum detection and employing multi-task learning proves beneficial and achieves superior accuracy. I found an implemented system on GitHub and I ran it on my laptop. The F-measure for audio recordings from the IDMT SMT dataset hovers around 0.94-0.95 which verifies their claim about the reported results and indeed these results are the state of art.

Given the fact that highly accurate transcriptions for drum have been implemented, we decided to use the ideas and rationale behind these systems and adapt them accordingly for Tabla recordings. Given the peculiarities of Tabla, we can use our understanding of the same to design a learning model.

The data scarcity of Tabla recordings was a huge problem and hence we decided to generate our own tabla compositions. Thus generated tabla sequences also helped us to evaluate the NMF based source separation system for drums. We generated lots of tabla sequences based on original compositions we had by changing the BPM and the bol sequence.

Having done that, we implemented the Magenta's transcription system for tabla recordings. The available code was suited for `Tfrecord` files and we don't have the same for our data. Hence we had to implement the entire model by ourselves. Having done that, another problem which arose was that the available sequences were long and it's difficult to train RNN models for long sequences. Hence we split the audio files into a length of 20 seconds each and stored the evaluated spectrograms, the corresponding onset and bol annotations.

As of now the entire model has been implemented and the data files have been split and the corresponding spectrograms and the onset and bol annotations have been stored. Now what remains to be done is to train the model and generate transcriptions.