Non-Negative Matrix Factorization for Drum Transcription

Jon Downing

University of Rochester

ABSTRACT

This poster presents a system for transcription and source separation of polyphonic drum recordings. Such a system may find applications in music education, music production, or entertainment. The system's methods for detection and decomposition are based on the well-known Non-Negative Matrix Factorization (NMF) approach. The basic multiplicative update rules are modified to capture the spectral variation over time of the percussive sounds per frame by using semi-adaptive update rules for the spectral templates. Additionally, two dictionary atoms are stored for each drum sound contained in the mixture, corresponding to the initial transient and steady-state decay of the drum sound. State-of-the-art onset detection methods are examined and applied to the initial decomposition. The proposed modification is shown to improve the f-score of the transcription given an identical onset detection function. We compare the transcription statistics over a dataset generated from acoustic and electronic drum samples.

BACKGROUND

This work in this paper is centered around the transcription and separation of single drum instruments from single monaural, polyphonic recordings of drum kit performances. In an educational context, this can enable the user to obtain feedback on a drum performance. Additionally, a live drum performance could be transcribed for later analysis or performance. In each of these applications, we require separate subsystems for source separation and transcription; here, we examine Non-Negative Matrix Factorization (NMF) and onset detection for each task respectively.

Non-Negative Matrix Factorization

A common approach to audio source separation is Non-Negative Matrix Factorization, as proposed in [1]. In this approach, the magnitude spectrogram of the signal is decomposed into a lower-rank approximation, consisting of a matrix B of r spectral vectors, or templates, and a matrix H of r time-varying gain vectors, or activations, for each source. A common approach is to choose a rank for the decomposition and initialize B and H randomly. Then the update rules below, from [1], are applied iteratively until convergence.

.

$$X \approx BH$$

$$B \leftarrow B \cdot \frac{\frac{X}{BH} B^T}{1B^T}$$

$$H \leftarrow H \cdot \frac{B^T \frac{X}{BH}}{B^T 1}$$

Fig. 1 – Equations of Non-Negative Matrix Factorization

Semi-Adaptive NMF and Head and Tail Modeling

One variation on NMF, applied to drum transcription in [2], is semi-adaptive NMF. Instead of a random initialization, the vectors of B are initialized to the time-averaged spectrograms of individual sound sources, forming a table of templates, B_P. Additionally, instead of allowing B to freely update according to Fig. 1, it is weighted more heavily toward the original B_p in early iterations, and is only allowed to freely adapt in later iterations. Intuitively, this means the algorithm prioritizes over matching the audio to the templates than the other way around.

In another study of modified NMF applied to drum sounds, [3] proposes a method of modeling the "heads" and "tails" of each source separately for drum sounds. This makes intuitive sense, since percussive onsets are almost always broadband and enharmonic, while drum decays often contain resonant harmonics. The system can be trained on isolated "sound check" recordings for each drum.

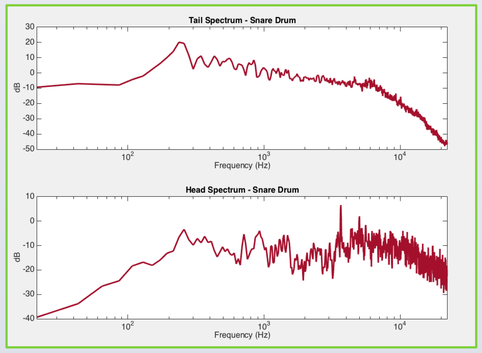

Fig. 2 – Spectral Tail and Head Templates for Snare Drum

Onset Detection

Usually, onset detection is performed on the reconstructed spectrograms of the components for transcription. Below is a simple method of extracting onsets from the magnitude spectrogram, employing a 1st-order difference function and simple thresholding.

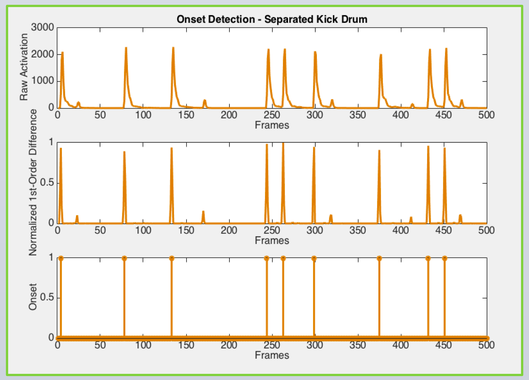

Fig. 3– Onset Detection Performed on Separated Kick Drum

PROPOSED METHOD

Our method makes use of semi-adaptive NMF with spectral templates learned from the isolated drums samples. However, we expand the rank to six templates, consisting of "heads" and "tails," as proposed in [3]. We obtain the "head" templates by using the onset detection on the training data to determine spectrogram frames corresponding to onsets, and generate our onset templates from only these frames. When semi-adaptive NMF is applied to the test data, it is hoped that crosstalk will be reduced since we now have more salient spectral information about the onsets of each individual drum, which is typically where most crosstalk occurs. The final transcription is done via onset detection on the reconstructed spectrogram of the B and H matrices from the NMF decomposition.

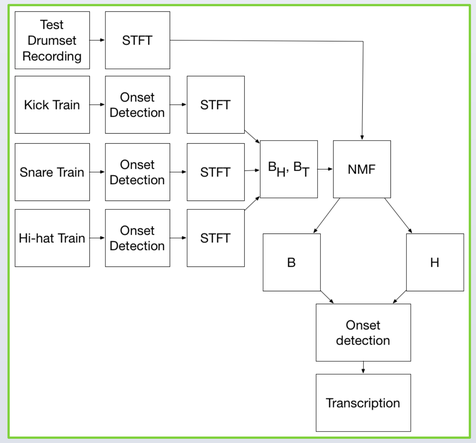

Fig. 4 – Block Diagram, Proposed Method

EXPERIMENTS

The system was tested on a database of synthesized and sampled drum recordings generated specifically for this study using Ableton Live. Below is a sample spectrogram of one of the test signals. The system was trained on a series of isolated drum hits. Multisampled instruments were used to ensure a realistic variation in timbre for single drums. Preset grooves and "hand played" MIDI sequences were used.

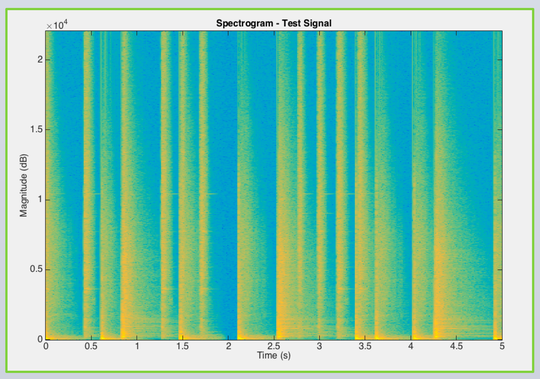

Fig. 5 – Test Signal Spectrogram

RESULTS

Precision and recall statistics are presented below for the system with no supervision, with semi-adaptive templates from the training data, and with separate head and tail templates from the training data.. It was found that the best subset of onset activations to use for transcription were kick tail, snare tail, and hi hat head. The spectrogram shows the snare drum's onset prominently in the resynthesized snare drum spectrogram. Precision is generally below recall due to mistaken hi hats.

Table 1 – Experimental Results

Condition	Precision	Recall	F-Score
Blind NMF	0.66	0.69	0.67
Semi-Adaptive Templates	0.72	0.74	0.73
Fixed Templates Heads and Tails	0.74	0.80	0.77
Semi-Adaptive Heads and Tails	0.74	0.81	0.77

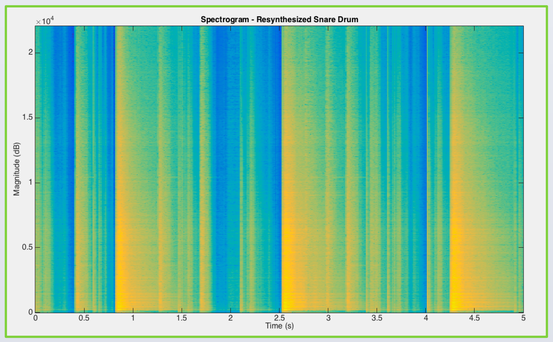

Fig. 6 – Separated Snare Drum Spectrogram

CONCLUSIONS

- With simple onset detection, separating heads and tails can improve precision
- Recall is marginally improved with semi-adaptive NMF
- These results represent simplified onset detection methods
- Hi-hat the biggest problem, causes false positive due to crosstalk
- Should be tested on more robust acoustic dataset

REFERENCES

[1] D. Lee and H. Seung. "Algorithms for Non-negative Matrix Factorization," *Advances in neural information processing systems*,Vol.~13, 2001.

[2] C. Dittmar and D. Gartner. "Real-Time Transcription and Separation of Drum Recordings Based on NMF Decomposition," *Proceedings of the 17th International Conference on Digital Audio Effects*, 2014.

[3] E. Battenberg, V. Huang, and D. Wessel. "Live Drum Separation Using Probabilistic Spectral Clustering Based on the Itakura-Saito Divergence," *AES 45th International Conference*, 2012.

ACKNOWLEDGEMENTS

Dr. Zhiyao Duan, Dr. Jakob Abeßer, and the University of Rochester Computer Audition class of Fall 2015.