

Early Investigations Into

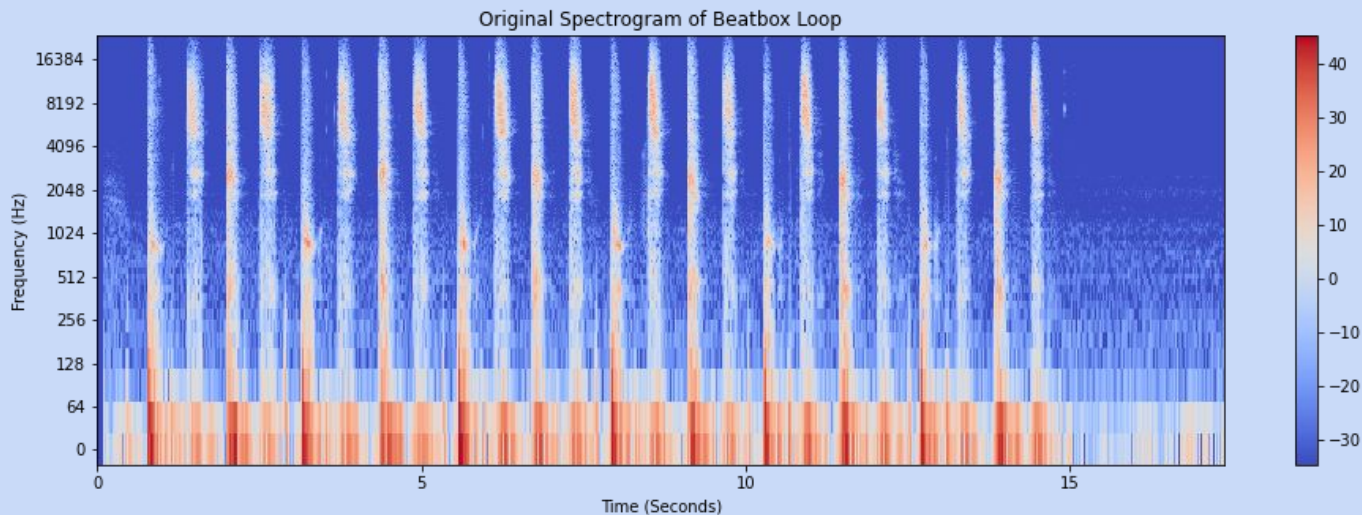
Real-Time Beatbox Resynthesis

Using Non-Negative Matrix Factorization

Joe Bumpus & Noah Miller
ECE 477 - Computer Audition
University of Rochester

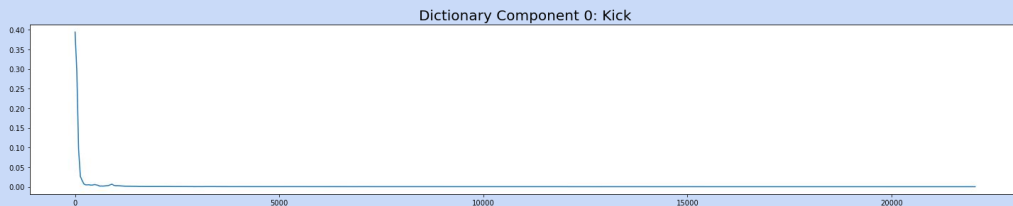
Project Statement & Purpose

- Use NMF to parse an input stream of human beatboxing and resynthesize an audio track using preloaded drum samples
- Facilitate quick form musical creation in a natural, intuitive, and inspiring way
- Propose methods for future real-time implementation

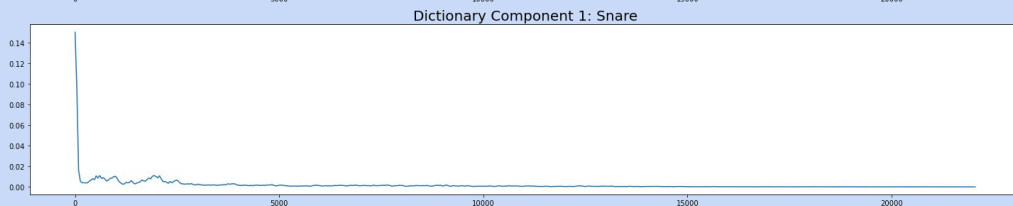


User Input & Dictionary Matrix

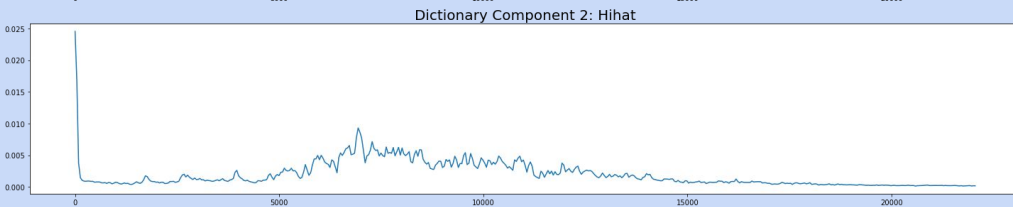
Columns of W (Dictionary Matrix)



Kick



Snare



Hihat

- Record using Pyaudio
- Convert from bytes object to Numpy array for processing
- Record each sound (3) individually to ensure NMF calculates individual dictionary elements in desired order
- Allows for users to listen back and re-record as quality check

Frequency (Hz)

Amplitude

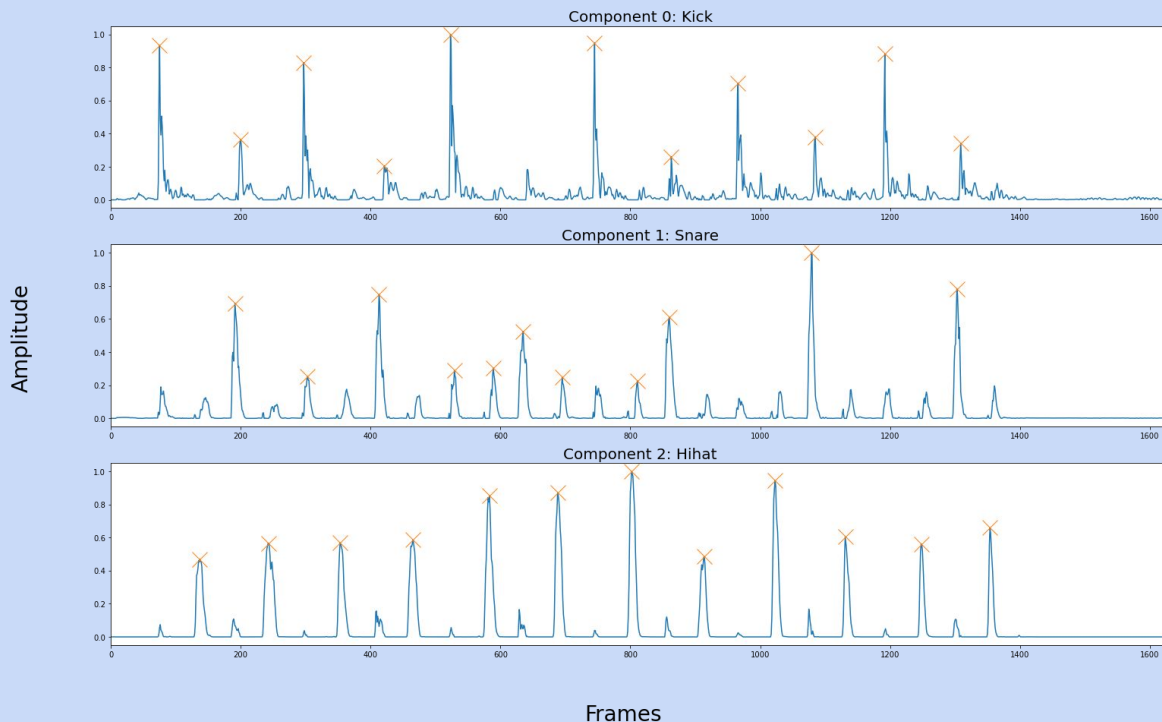
User Input & Initial NMF Activation Matrix



Beatboxing Testing File

- Simple local maxima detection with thresholding
- Doesn't take into account distance or the other components

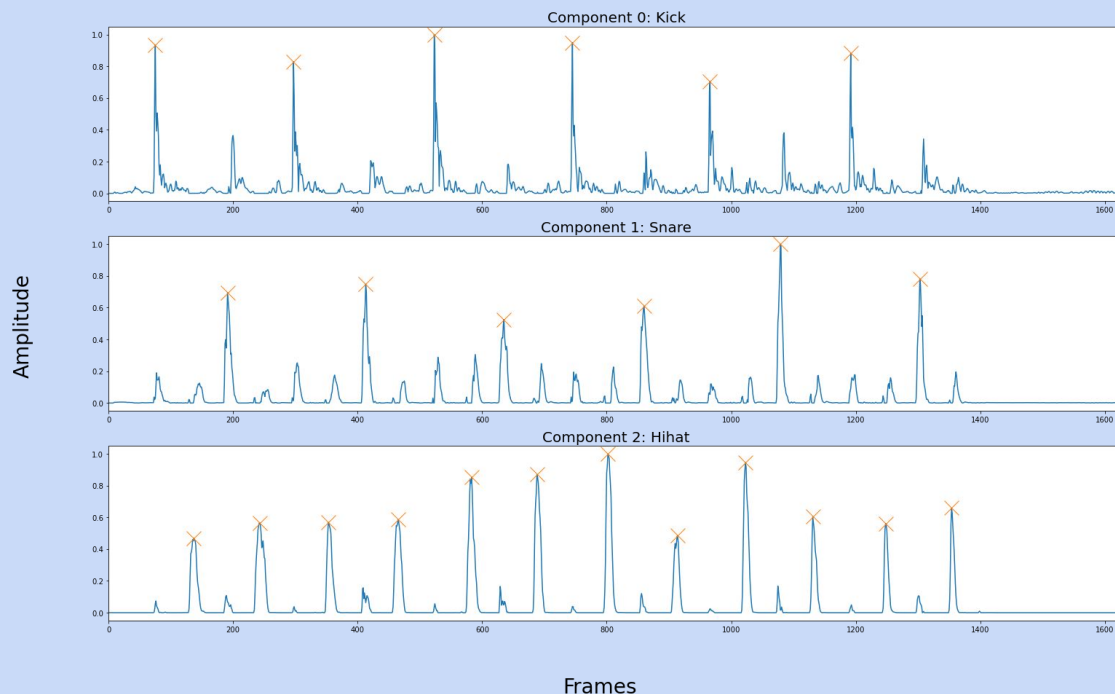
Rows of H (Activation Matrix)



Activation With Amplitude Comparison

- For every detected maxima point, the next maxima (from any component) is examined
- If it is within a certain distance threshold, the amplitudes of both are compared, and the largest one is determined to be the correct detection
- The false detection is then removed and the algorithm continues through the rest of the maxima

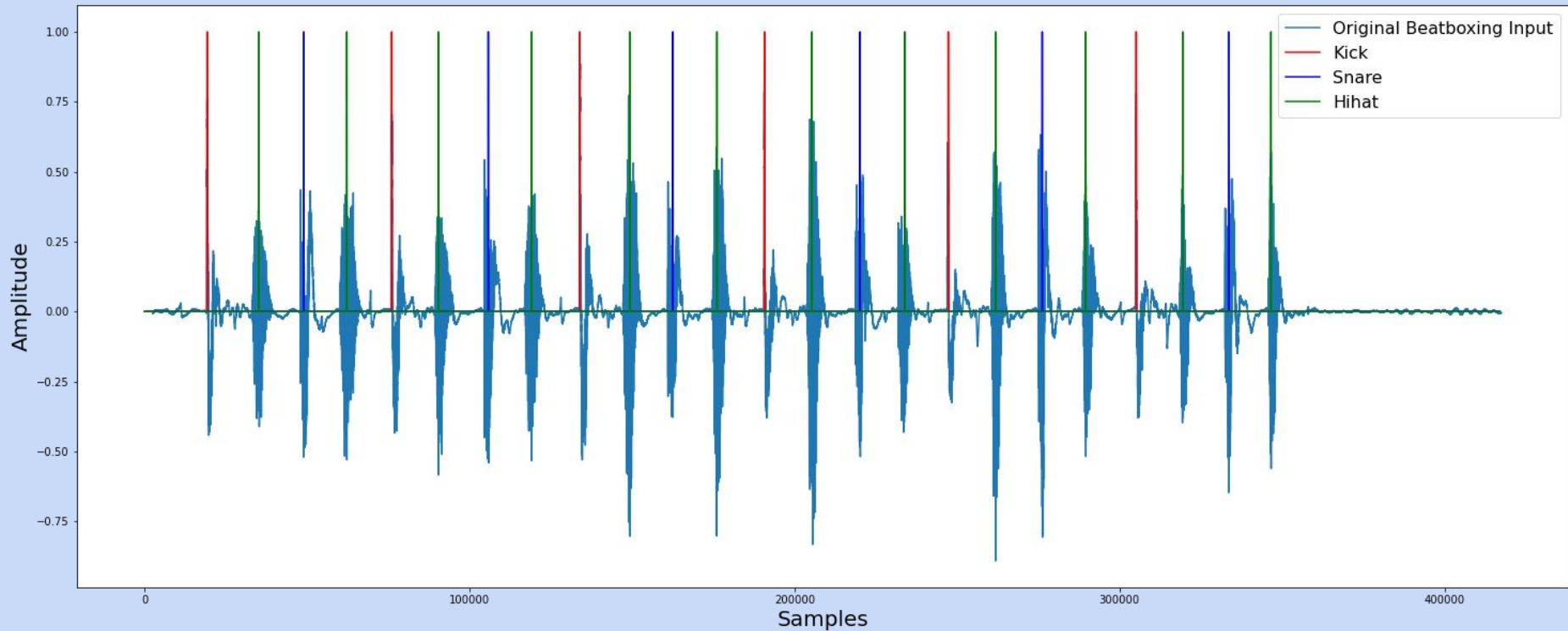
Rows of H (Activation Matrix)



Resulting Activations



Beatbox Audio and Drum Activations

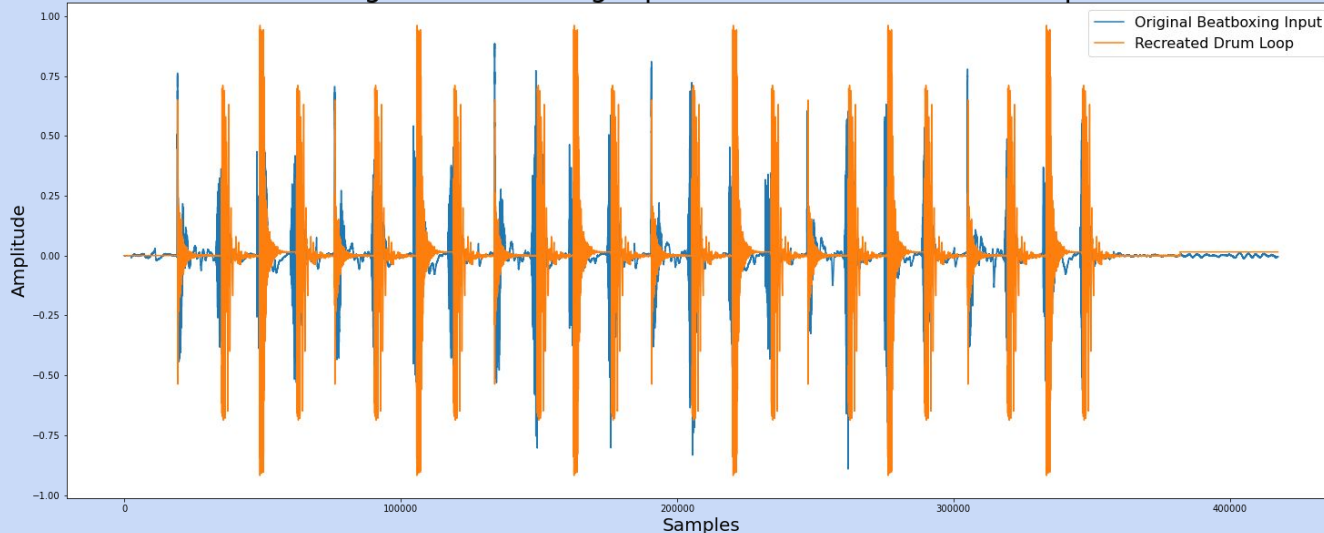


Output

The output audio signal is shown line up properly with the output.

The triggered drum samples line up with the corresponding components from the input beatboxing.

Original Beatboxing Input and Recreated Drum Loop

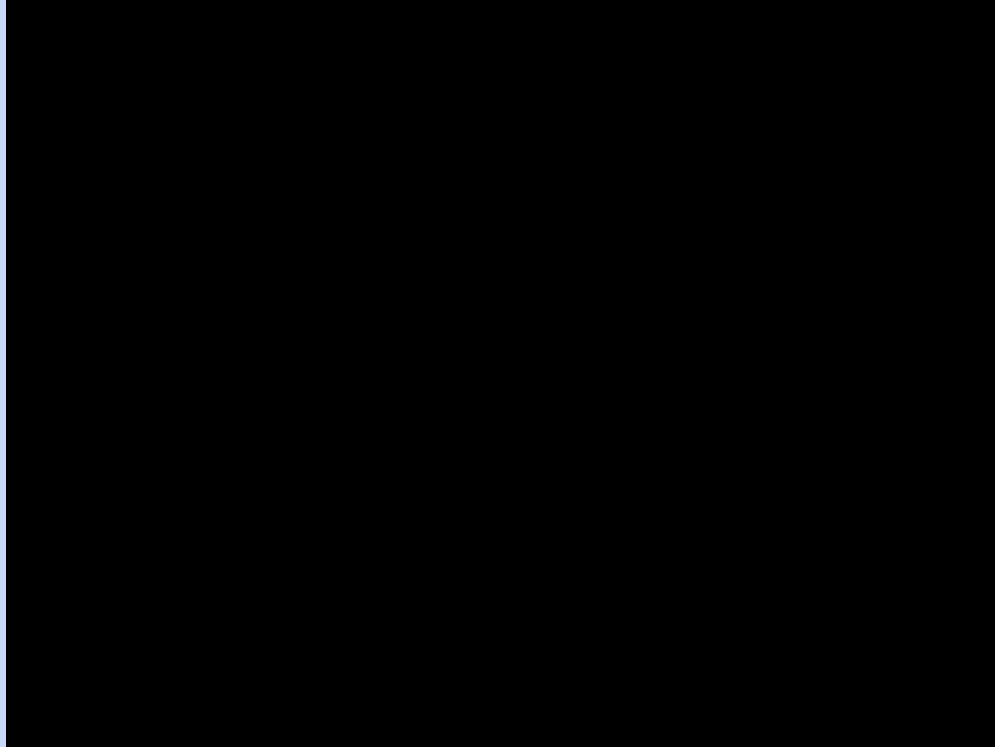


Mixed Signal



Drum Output

Script Demo



Please excuse the bad beatboxing.

Limitations & Future in Real-Time



Current Limitations:

- Analytically determined local maxima distance within activation matrix element and across elements in comparison can result in not detected activations if performing too fast or slurred

Future in Real-Time:

- How to determine local maxima and take representative STFT without look ahead
- Track activations for analytically determined number of frames
- Trigger based on strongest activation
- Limiting latency will be a significant issue

Realistically, don't use NMF. It's iterative and relies on STFTs. Use an actual source identification algorithm, not source separation.

Questions?

