Beatbox to Drums using Support Vector Machines

Arvind Ramanathan (email - arm@ur.rochester.edu)

Introduction

Beatboxing is a vocal imitation of musical instruments mainly drums and percussive instruments. The goal of this project is to convert the input beatboxing audio waveform with the synthetic drum sounds using SVMs. As SVM classifiers are much faster to train and less computationally expensive.

Overview

The overall idea is to segment the input beatboxing audio waveform into individual slices based on the detected onsets from the audio.

Once the audio is segmented into slices, the pre-trained SVM classifier is employed to detect the group name for each segmented slice of beatboxing.

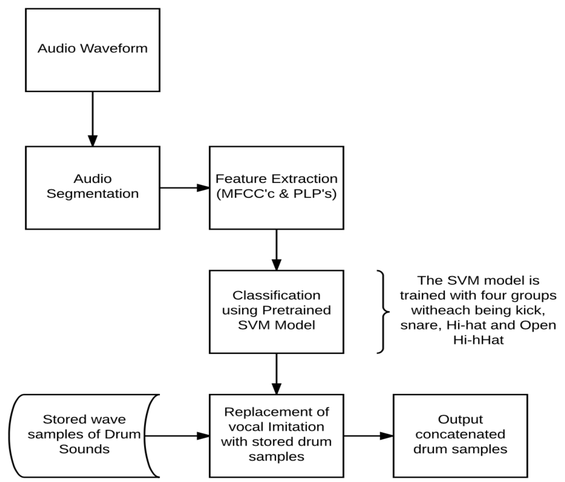

Figure 1: Simple block diagram of the system

Onset Detection

Spectral Flux method is used to detect the onsets. Spectral Flux measures the change in magnitude in each frequency bin, and if this is restricted to the positive changes and summed across all frequency bins.

It gives the onset function SF. H(x) is the half-wave rectifier function.

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|)$$

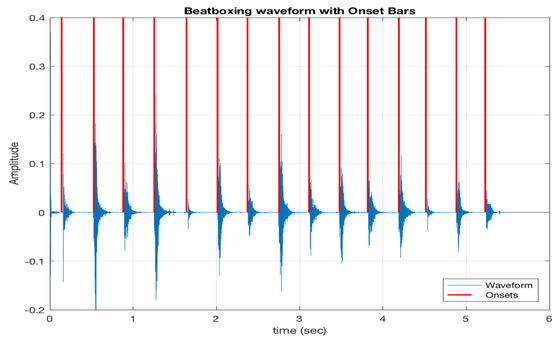

Figure 2: Audio waveform with detected onsets

Feature Extraction

As MFCCs are good in capturing the acoustic features of audio and speech efficiently. The training and segmented audio frame is processed to obtain the Mel-frequency coefficients. MFCCs is a representation of the short-term power spectrum of a sound.

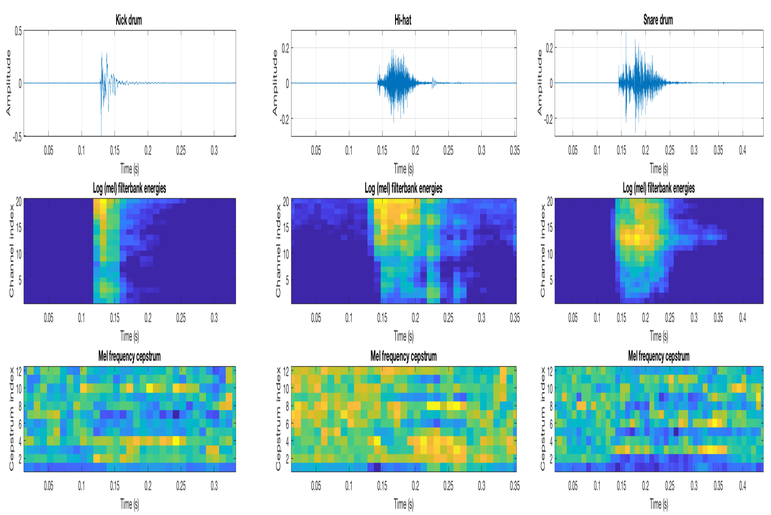

Figure 3. MFCCs of the drum vocal imitation

SVM

SVM's are supervised learning models, associated learning algorithms that analyzes data used for classification and regression analysis.

This system uses libsvm, a MATLAB executable C and C++ library for multiclass classification using SVM's.

Data size used to train the model consist of 400 observations and equally split among the groups.

Results

The trained SVM model achieved an accuracy of about 95.7% in the test set. Which is almost in par with the results obtained through neural networks.

Since the dataset is it minimal, the system is not robust to different vocal differences and onset detection falters in noisy recordings.

Future Work

Work has to be done on generating a bigger dataset, As there are no good dataset readily available now.

One other future extensions to the proposed system could be adding different classes to classify (i.e., having more drum sounds as vocal imitation).

The system accuracy however, is severely affected in the presence of noise. Which has to be addressed in future in order to make this system to be more robust to all environments.

UNIVERSITY of ROCHESTER