MIR Task: Beatbox Dataset Evaluation Using Automatic Drum Transcription

Pedro Ricardo

Github Link: https://github.com/PedroMCRicardo/MIR

1. Introduction and Task Proposal

According to MIREX2018 [1], drum transcription is defined as the task of detecting the positions in time and labeling the drum class of drum instrument onsets in polyphonic music. The article A Review of Automatic Drum Transcription [2] offers an in-debt explanation about automatic drum transcription as well as an extremely up to date state of the art.

The present task proposes to test and evaluate a beatbox dataset using an automatic drum transcription algorithm - ADTLib. To do so, a second dataset ENST-Drums (used initially to develop ADTLib) will be used as a term of comparison. Further conclusions about the beatbox dataset and its elements will be made.

2. Data

2.1 Beatboxset1

The beatboxset1: beatboxing audio data set available in the Audio Content Analysis [6] contains beatboxing recordings from various contributors, who recorded the clips themselves in various conditions. There is a spreadsheet file "beatboxset1.csv" accompanying the dataset that provides metadata for the recordings. Further annotations of the recordings are also included: these mark the positions of onsets as well as categorizing the events into a handful of standard classes.

```
* k  = kick
* hc = hihat, closed
* ho = hihat, open
* sb = snare, "bish" or "pss" -like
* sk = snare, "k" -like  (may sound like a "clap" or "rimshot" snare)
* s  = snare but not sure which of the above types (or isn't either of them)
* br = a breath sound (not intended to sound like percussion)
* m  = humming (or similar, a note with no drum-like or speech-like nature)
* v  = speech or singing
* x  = miscellaneous other sound (identifiable, but not fitting one of
                                  the other categories)
* ?  = unsure of classification
```

Labels used in Annotation of beatboxset1

2.2 ENTS-Drums

The ENST-Drums database is a large and varied research database for automatic drum transcription and processing.

Label	Description	Label	Description
bd	Bass drum	lmt	Low-mid tom
sweep	Brush sweep	mt	Mid tom
sticks	Sticks hit together	mtr	Mid tom, hit on the rim
sd	Snare drum	lt	Low tom
rs	Rim shot	ltr	Low tom, hit on the rim
cs	Cross stick	lft	Lowest tom
chh	Hi-hat (closed)	rc	Ride cymbal
ohh	Hi-hat (open)	ch	Chinese ride cymbal
cb	Cowbell	cr	Crash cymbal
c	Other cymbals	spl	Splash cymbal

Labels used in Annotation of ENTS-Drums

3. ADTLib

Automatic Drum Transcription Library (ADTLib) is a library that contains open source ADT algorithms to aid other researchers in areas of music information retrieval (MIR). The algorithms returns both a .txt file of kick drum, snare drum, and hi-hat onsets location. ADTLib was develop as an open source implementation of the different systems and algorithms initially presented in the research paper Automatic drum transcription for polyphonic recordings using soft attention mechanisms and convolutional neural networks (2017). The paper by Carl Southall, Ryan Stables and Jason Hockman focus in implementing different techniques using soft attention mechanisms (SA), convolutional neural networks (CNN) and bidirectional recurrent neural networks(BRNN) on the automatic drum transcription problem.

However, the available implementation of ADTLib does not train neural networks. The music files are processed through a pre-trained neural network to give the automatic drum transcriptions.

After some research on the topic, there wasn't found any type of algorithm implemented directly in Essentia that deals with this particular problem. Exploration of Techniques for Automatic Labeling of Audio Drum Tracks' Instruments [9] by Perfecto Herrera and Fabien Gouyon, produced in the MTG in 2001, is used as a reference for the implementation of the algorithm StrongDecay [10] in Essentia. Although this algorithm is related to the recognition of percussive instruments, it does not tackle this problem directly.

4. Evaluation Method

The evaluation method focus in understanding how well the implemented algorithm performs by comparing the hand-made annotations, available in the data sets, to the annotations produced by ADTLib. A F-measure value will be computed for each different sound: kick, snare and hi-hat. The detected onsets are accepted as true positives if they fall within 50ms of the hand-made annotations provided by each dataset. The individual instrument F-measures are calculated as the mean F-measure across test tracks and the mean instrument F-measure is calculated as the mean F-measure across the individual instruments. mir_eval.onset.evaluate: determine which estimated onsets are “correct”, where correctness is defined as being within a small window of a reference onset. More information about mir_eval can be red in mir_eval: A Transparent Implementation of Common MIR Metrics

5. Results

Although the implementation of the task is not fully concluded, some tests were made in order to understand the feasibility of the task.
Figure 1 and Figure 2 show the computation of the 3 individual f-measures for each instrument (Kick, Snare and Hi- hat) across all different tracks in both datasets.

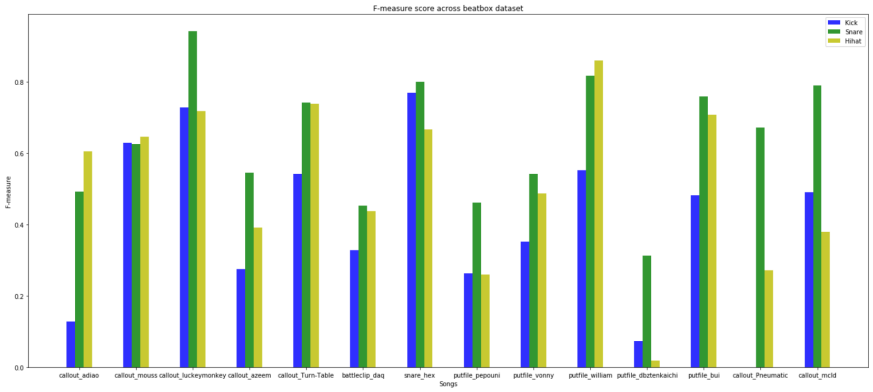

Fig2). F-measures for beatboxset1 dataset

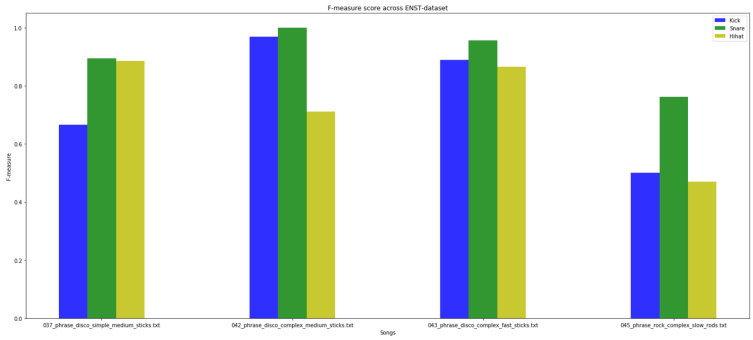

Fig2) F-measures for ENTS dataset

The way data is organized doesn't let us make any conclusions about the beatbox dataset. Further analysis about each instrument and each contributor to the beatbox dataset should be made in order to reach interesting conclusions.

6. Conclusions

Other test must be made before final conclusions can be made.