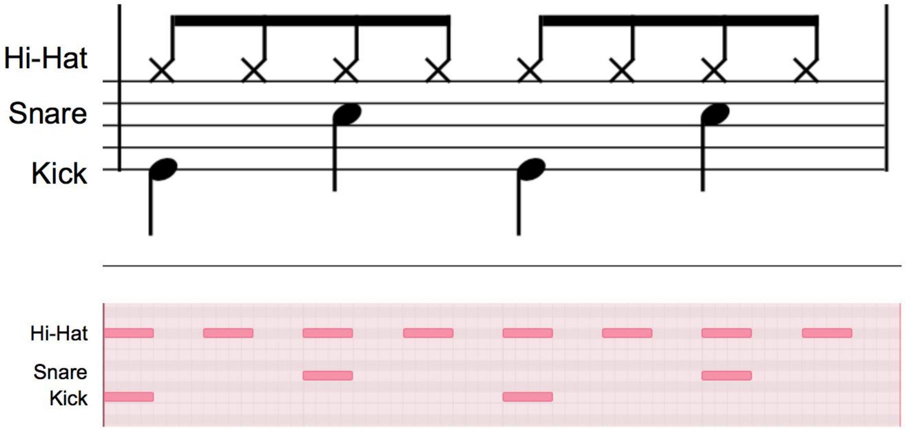
Musical Transcription of Drum Patterns Using Main Audio Features in KNN

Charles Jayson L. Dadios
BS Computer Science

Introduction

Drums is a prominent instrument in our musical culture.

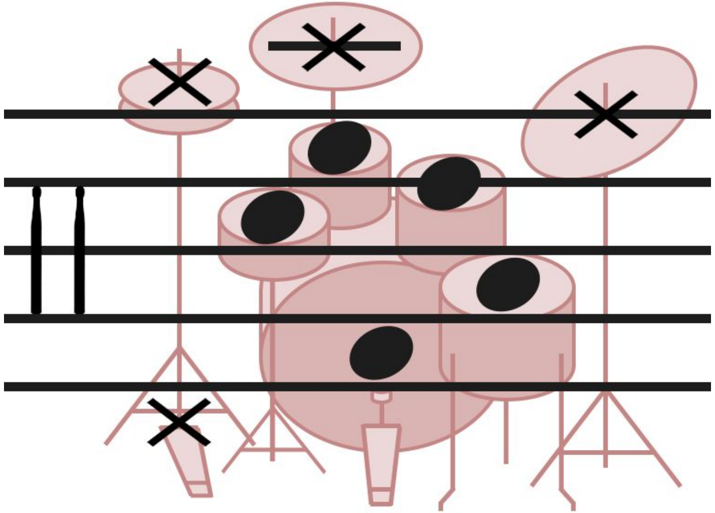
Automatic Drum Transcription (ADT) in simple terms is the process of converting a drum performance into a record usually as a drum notation printed as a music sheet.



KNN is a classification algorithm which uses data points to find the k-number of closest classes as the basis of prediction.

Statement of the Problem

Unlike for chordal instruments which use lyrics with chords, most popular songs have no available drum transcriptions for learners.



Although reading charts can be easy to pickup, writing charts is difficult and time-consuming.

This challenge encourages drummers to learn by ear and memorize, but people tend to forget the correct parts and play inconsistently.

Objectives

General: To produce a PDF and MIDI transcription of an drum recording

Specific:

1. To produce physiologically realistic drum sound combinations using downloaded drum instrument audio samples
2. To extract main audio features from the data set
3. To classify the onsets in a given drum recording
4. To produce a PDF and MIDI transcription out of the classified onsets
5. To evaluate the accuracy of the produced transcription

Methodology

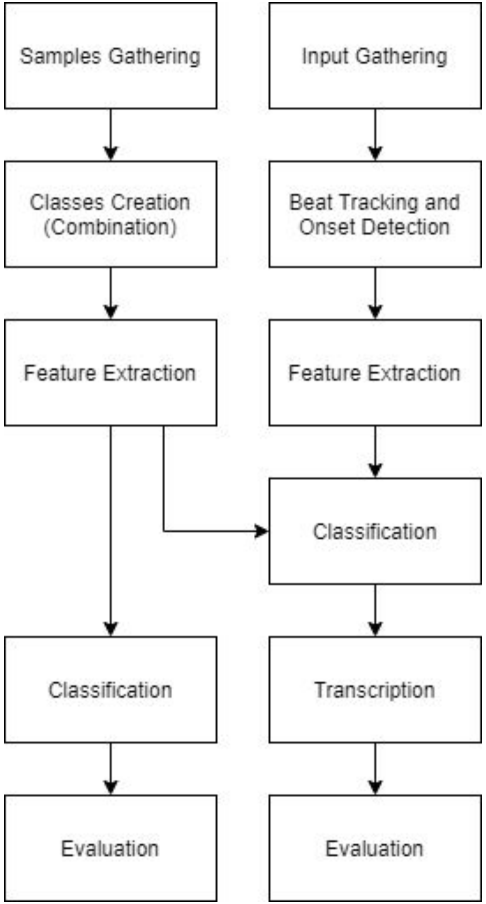
A. Data Gathering
   1. Create classes - combination
   2. Detect classes - Librosa

B. Feature Extraction - Librosa

C. Classification - Scikit-learn

D. Transcription - MIDIUtil, Lilypond

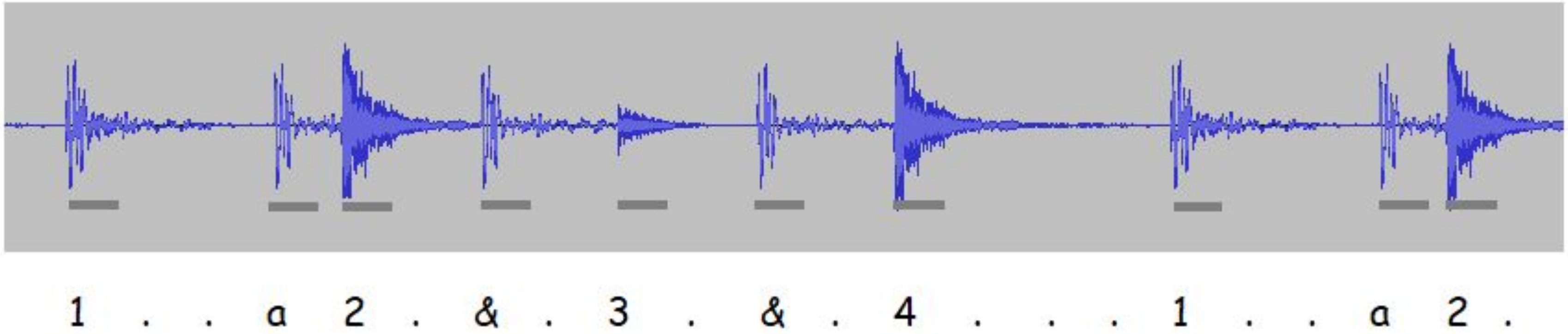
E. Evaluation - Scikit-learn, confusion matrix



Methodology

1. Combination
2. Beat and onset detection

Single	Double	Triple
Bass	Bass Snare	
Snare	Bass Hihat	Bass Snare Hihat
Hihat	Snare Hihat	



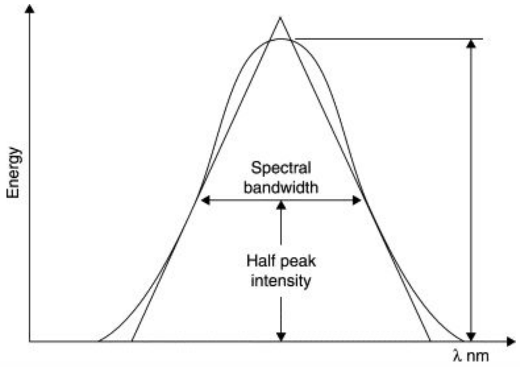
Methodology

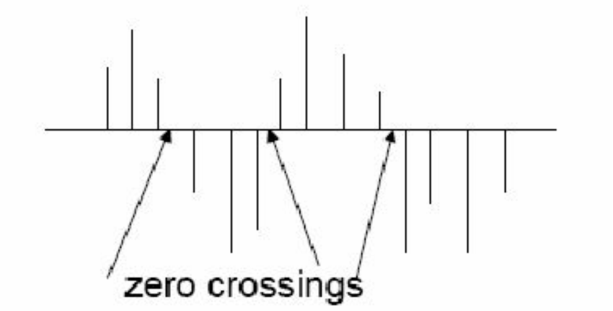
Audio Features:

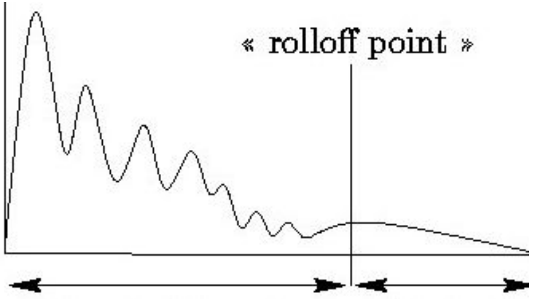
1. Spectral bandwidth -
2. Spectral contrast* - “ Spectral contrast is defined as the level difference between peaks and valleys in the spectrum”
3. Zero crossing rate -

4 MFCC* - “human perception graph”

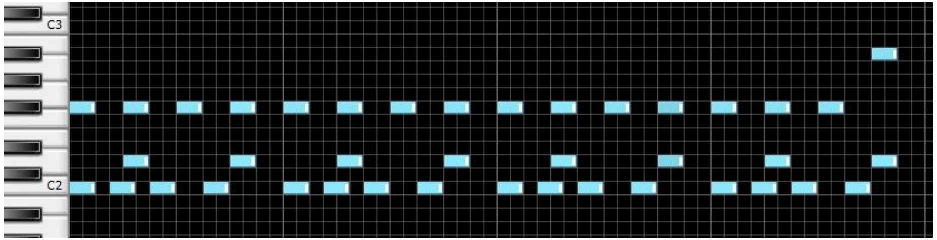
5. Spectral centroid - spectral center of mass
6. Spectral flatness - tone-likeness
7. Spectral rolloff -

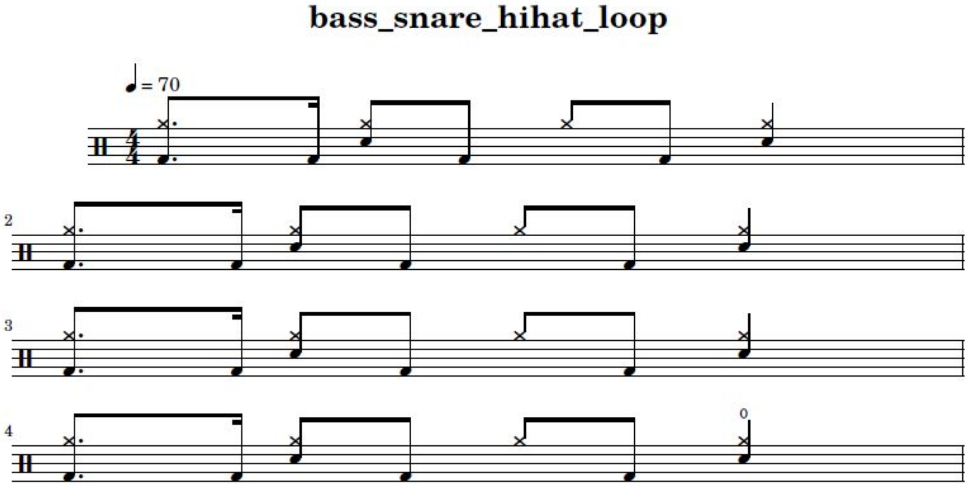


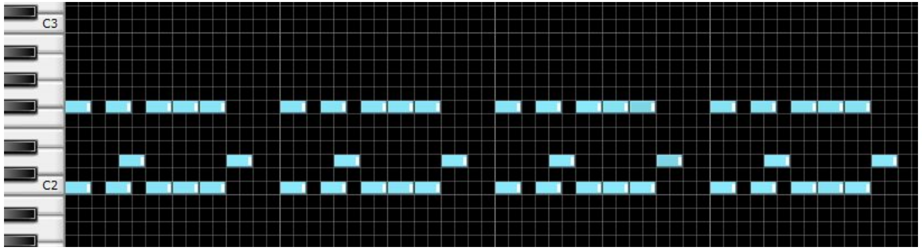


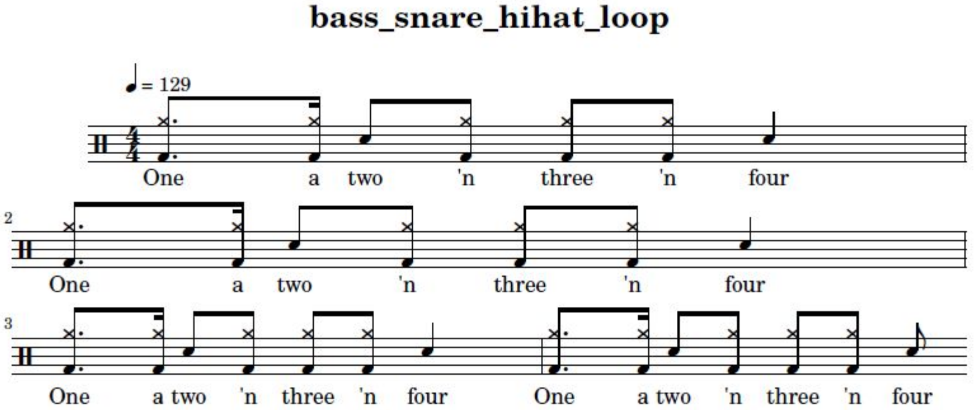


Results


C3
C2


bass_snare_hihat_loop
♩ = 70
2
3
4
0


C3
C2


bass_snare_hihat_loop
♩ = 129
One a two 'n three 'n four
2
One a two 'n three 'n four
3
One a two 'n three 'n four One a two 'n three 'n four

Results

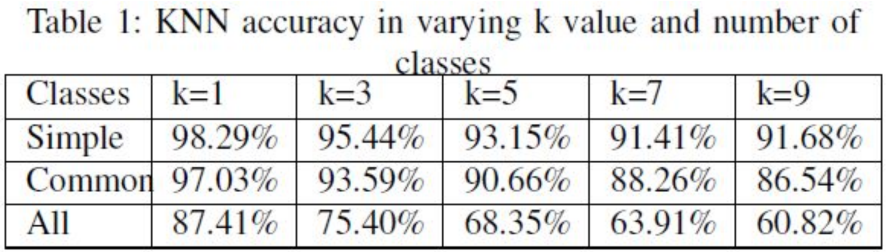


Table 1: KNN accuracy in varying k value and number of classes

Classes	k=1	k=3	k=5	k=7	k=9
Simple	98.29%	95.44%	93.15%	91.41%	91.68%
Common	97.03%	93.59%	90.66%	88.26%	86.54%
All	87.41%	75.40%	68.35%	63.91%	60.82%

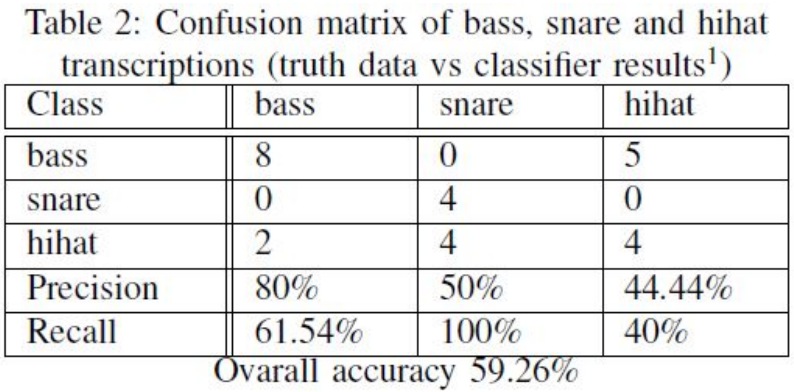


Table 2: Confusion matrix of bass, snare and hihat transcriptions (truth data vs classifier results[1])

Class	bass	snare	hihat
bass	8	0	5
snare	0	4	0
hihat	2	4	4
Precision	80%	50%	44.44%
Recall	61.54%	100%	40%

Ovarall accuracy 59.26%

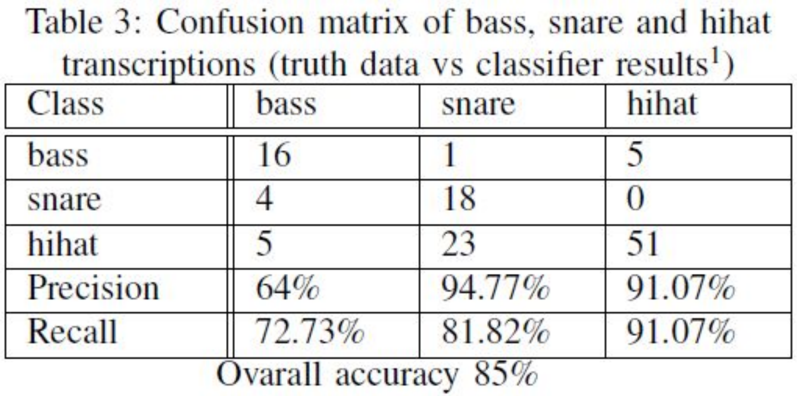


Table 3: Confusion matrix of bass, snare and hihat transcriptions (truth data vs classifier results[1])

Class	bass	snare	hihat
bass	16	1	5
snare	4	18	0
hihat	5	23	51
Precision	64%	94.77%	91.07%
Recall	72.73%	81.82%	91.07%

Ovarall accuracy 85%

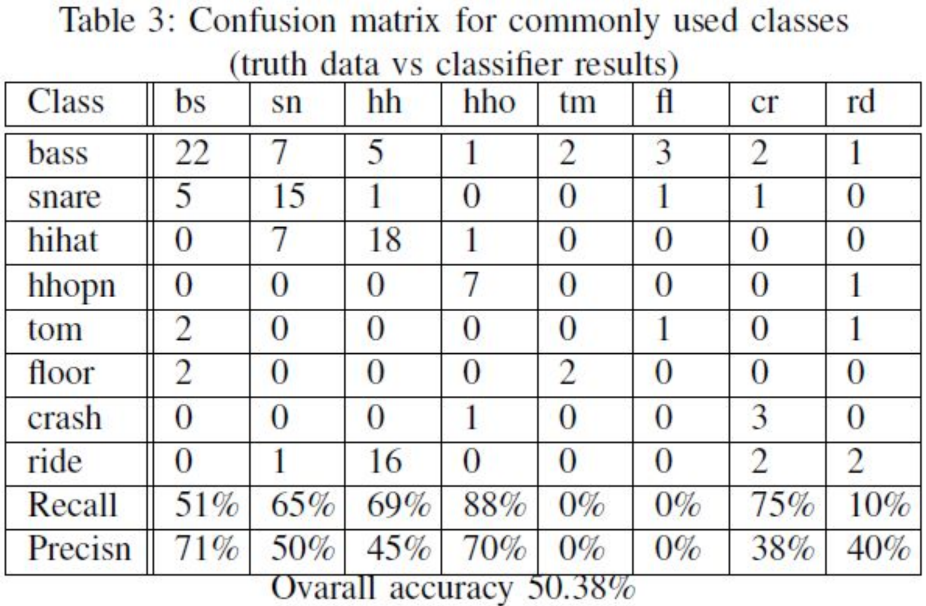


Table 3: Confusion matrix for commonly used classes (truth data vs classifier results)

Class	bs	sn	hh	hho	tm	fl	cr	rd
bass	22	7	5	1	2	3	2	1
snare	5	15	1	0	0	1	1	0
hihat	0	7	18	1	0	0	0	0
hhopn	0	0	0	7	0	0	0	1
tom	2	0	0	0	0	1	0	1
floor	2	0	0	0	2	0	0	0
crash	0	0	0	1	0	0	3	0
ride	0	1	16	0	0	0	2	2
Recall	51%	65%	69%	88%	0%	0%	75%	10%
Precisn	71%	50%	45%	70%	0%	0%	38%	40%

Ovarall accuracy 50.38%

Conclusion

The program successfully combined the drum samples to produce other physiologically valid classes.

Audio features were successfully extracted from every class samples.

Using main audio features in KNN is effective in identifying individual onset types from another, attaining the highest overall accuracy of 98% for k=1 and when considering only between 3 classes, even with the feature vector dimension of 77, in this case.

The algorithm was able to output same transcriptions but in both PDF and MIDI format.

The transcription performance was evaluated with an accuracy of 50% for k=1 when considering the commonly used classes, and the highest of 85% when considering just the three classes.

References

Spectral bandwidth image - https://ars.els-cdn.com/content/image/3-s2.0-B9780857092298500061-f06-15-9780857092298.gif

Zero crossing image - https://www.researchgate.net/profile/Buket_Barkana/publication/259828967/figure/fig1/AS:299377940811777@1448388671414/Definition-of-zero-crossings-rate.png

Spectral rolloff image - https://www.researchgate.net/profile/Julien_Pinquier/publication/278631244/figure/fig5/AS:669382405545993@1536604610274/Definition-du-Spectral-Rolloff-Point.ppm