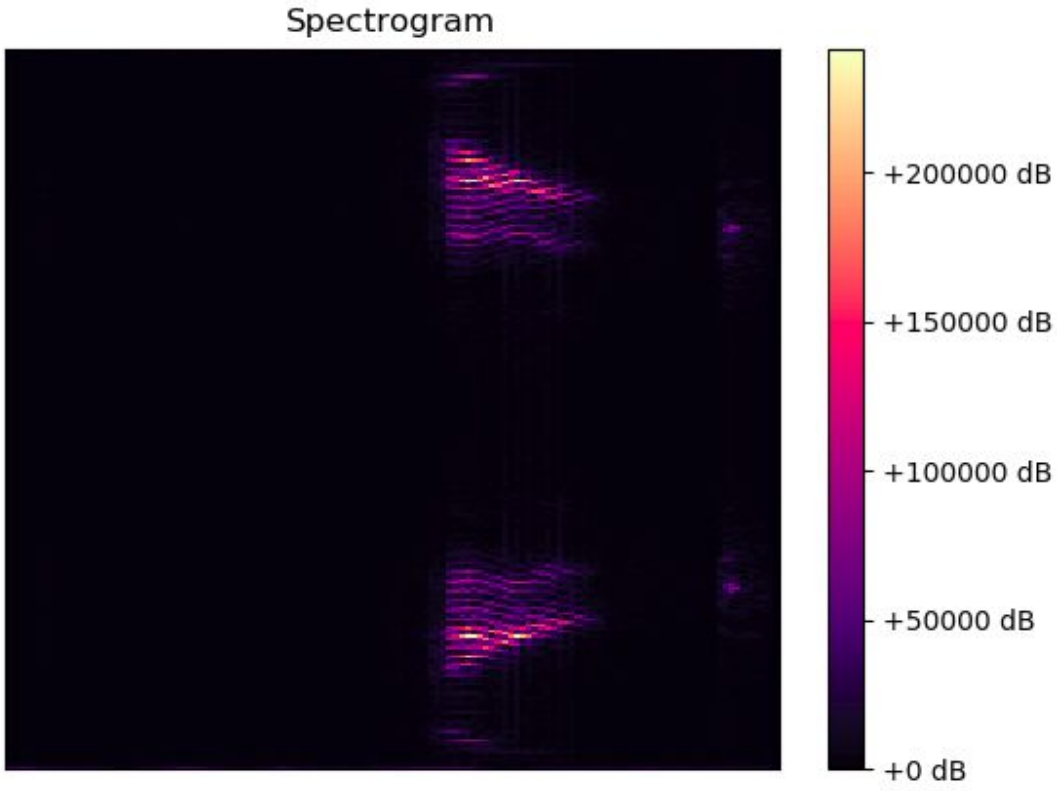
Assignment 2

By Rohan Chhokra
2016080

1. For an audio file, a window of 256 length was taken, with subsequent windows having an overlap of 50%.
2. On each such frame a custom defined FFT was implemented.
3. The following spectrogram was achieved for an audio file saying ‘eight.



For my MFCC implementation, I followed the following approach ->

1. To save time and computation I pickled the previous spectrograms and since the first steps for MFCC are spectrograms.
2. I got a power spectrum of the said spectrograms.
3. This spectrogram was matrix multiplied with a custom filterbank of the following parameters -

    Number of Filter points - 20
    Window Size - 256
    Overlap percentage - 50

In both the implementations the noise files have been appended to the signal with a weight of 0.001. Noise files were selected at random.

While training an SVM due to computation limits, only the default parameters and kernels could be tried.

The results are as follows ->

Spectrogram -

Class wise Precision,Recall Report

	precision	recall	f1-score	support
0	0.81	0.50	0.62	260
1	0.48	0.59	0.53	229
2	0.33	0.51	0.40	235
3	0.55	0.39	0.45	248
4	0.78	0.63	0.69	280
5	0.67	0.53	0.59	242
6	0.41	0.85	0.55	262

7	0.76	0.50	0.61	263
8	0.59	0.55	0.57	242
9	0.69	0.46	0.55	230
accuracy			0.55	2491
macro avg	0.61	0.55	0.56	2491
weighted avg	0.61	0.55	0.56	2491

MFCC -

Class wise Precision,Recall Report

	precision	recall	f1-score	support
0	0.58	0.42	0.49	260
1	0.47	0.35	0.40	229
2	0.16	0.76	0.27	235
3	0.37	0.23	0.28	248
4	0.59	0.43	0.50	280
5	0.65	0.44	0.52	242
6	0.59	0.39	0.47	262
7	0.69	0.25	0.37	263
8	0.66	0.28	0.39	242
9	0.62	0.41	0.49	230
accuracy			0.39	2491
macro avg	0.54	0.40	0.42	2491
weighted avg	0.54	0.39	0.42	2491

Resources Referred -

Spectrogram(FFT and DFT)

https://towardsdatascience.com/fast-fourier-transform-937926e591cb
https://www.youtube.com/watch?v=htCj9exbGo0

MFCC -
http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/