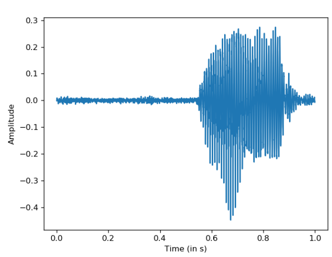
MCA Assignment 2

Anish Madan, 2016223

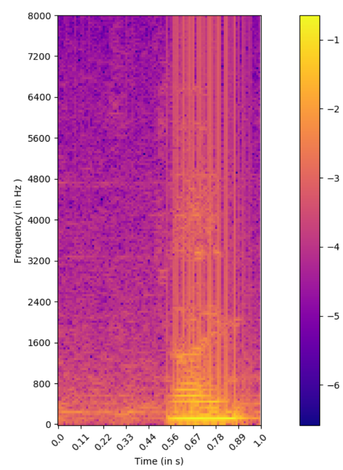
March 2020

1 Q1 - Spectrogram

Spectrogram is a way to visualize and analyse signals like audio signals and their behavior over time. The size of this feature depends on sampling rates of the signal and frame sizes chosen for evaluation. Here we see a visualization by plotting a spectrogram for an audio signal.



(a) Original Audio signal

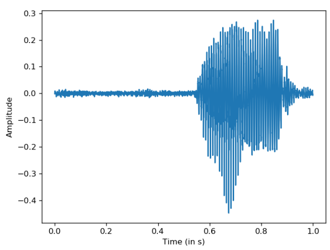


(b) Spectrogram Visualization of corresponding Audio Signal

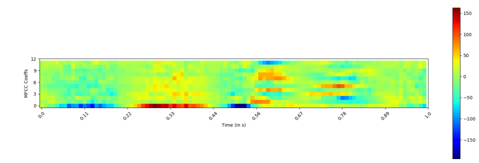
Figure 1: Visualizing the spectrogram of an audio signal

2 Mel-Frequency Cepstral Coefficients (MFCC)

MFCCs are an important technique to generate an audio signal's features, especially for tasks involving ASR(Automatic Speech Recognition). We compute filter banks using the power spectrogram and take the Discrete Cosine Transform (DCT) of these filter banks to de-correlate them, thereby giving us better feature representations. We then choose 12 MFCC coefficients as features for each window (or frame of audio) in consideration. Here we visualize MFCC features of an audio signal.



(a) Original Audio signal



(b) MFCC Visualization of corresponding Audio Signal

Figure 2: Visualizing the MFCC features of an audio signal

3 Audio Classification

We use SVMs to classify the audio in our dataset. We use both MFCC and Spectrogram features representations (with and without random noise). We tried a number of parameter settings for our SVMs and chose ones which gave the best performance. We use an RBF kernel for Spectrogram and Polynomial kernel for MFCC features, with default parameters provided in sklearn.

For cases where signal is augmented with random noise, we employ the following method to perform it. For each of the noise files, we randomly pick ten 1-second intervals. We then try to add multiple noise signals by simply adding the the weighted sum of 1-second noise signals chosen to the original signal. The weights of the noise signals are small numbers to maintain the concept of adding a noise, and not creating big changes to the original signal. We tried using one 1-second and three 1-second noise signals to be augmented with each audio. Though both cases resulted in a slight decrease in precision/recall values, we report it for the case of choosing one 1-second noise signal.

We report our precision/recall values obtained for the various models trained: olors).

Results obtained for Audio Classification				
Metric	Spectrogram	Spectrogram + Noise	MFCC	MFCC+Noise
Precision	0.6823	0.6657	0.8055	0.7807
Recall	0.6467	0.5942	0.8023	0.7762

Table 1: We note that precision/recall values are computed using the weighted average criteria. We observe that MFCC features work better than spectrogram . This is expected as MFCC features provide richer information as compared to spectrogram due to algorithm behind it. We observe that in our case using noise signal leads to a slight decrease in performance. This factor is quite dependent on the validation data and hence it is very hard to comment the reasons behind it. Also, there is a significant difference in precision and recall values for spectrogram features indicating more false negatives than false positives. Such a difference does not exist for MFCC features.