Spectrogram

The audio files were zero-padded to make their length 16000 samples or 1 second in duration
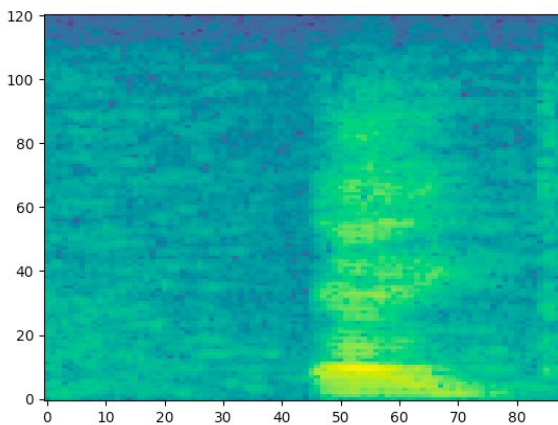
After due experimentation, the following values were chosen -
1. Window Size - 15ms
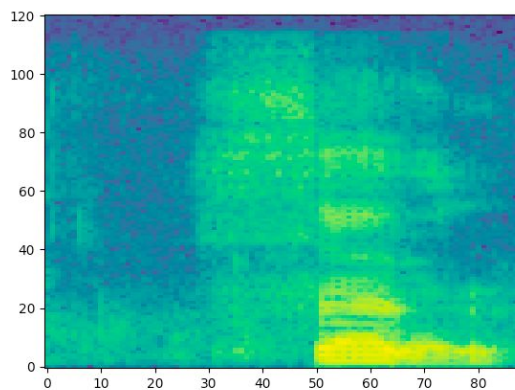2. Overlap - 25%

This resulted in feature vectors of size 121 x 88

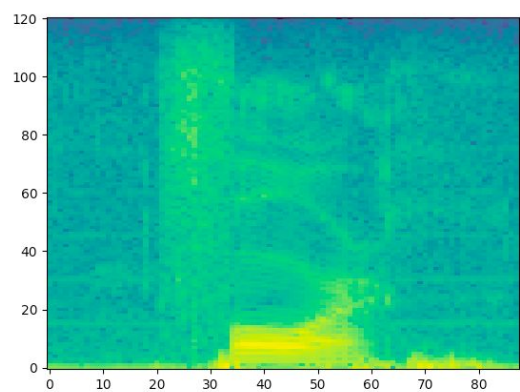Sample Spectrograms from the training data

Class - Zero

Class - Five

Class - Nine

Results

Training Results

	precision	recall	f1-score	support
0	0.78	0.59	0.67	1000
1	0.50	0.61	0.55	1000
2	0.46	0.59	0.52	1000
3	0.72	0.52	0.61	1000
4	0.71	0.62	0.66	1000
5	0.55	0.62	0.59	1000
6	0.65	0.83	0.73	1000
7	0.67	0.65	0.66	1000
8	0.64	0.65	0.65	1000
9	0.60	0.49	0.54	1000
micro avg	0.62	0.62	0.62	10000
macro avg	0.63	0.62	0.62	10000
weighted avg	0.63	0.62	0.62	10000

Validation Results

	precision	recall	f1-score	support
0	0.75	0.59	0.66	260
1	0.39	0.50	0.44	230
2	0.38	0.51	0.44	236
3	0.58	0.52	0.55	248
4	0.62	0.60	0.61	280
5	0.56	0.52	0.54	242
6	0.73	0.78	0.76	262
7	0.68	0.62	0.65	263
8	0.59	0.60	0.59	243
9	0.48	0.41	0.45	230
micro avg	0.57	0.57	0.57	2494
macro avg	0.58	0.57	0.57	2494
weighted avg	0.58	0.57	0.57	2494

Changes in Result on Addition of Noise

Adding Noise to 25 % of samples

Training Result

	precision	recall	f1-score	support
0	0.74	0.46	0.57	1000
1	0.50	0.48	0.49	1000
2	0.45	0.45	0.45	1000
3	0.73	0.40	0.51	1000
4	0.71	0.48	0.57	1000
5	0.46	0.55	0.50	1000
6	0.31	0.86	0.45	1000
7	0.62	0.51	0.56	1000
8	0.59	0.51	0.55	1000
9	0.63	0.37	0.47	1000
micro avg	0.51	0.51	0.51	10000
macro avg	0.58	0.51	0.51	10000
weighted avg	0.58	0.51	0.51	10000

Validation Result

	precision	recall	f1-score	support
0	0.75	0.59	0.66	260
1	0.40	0.52	0.45	230
2	0.35	0.49	0.41	236
3	0.63	0.50	0.56	248
4	0.62	0.59	0.60	280
5	0.56	0.55	0.55	242
6	0.74	0.76	0.75	262
7	0.66	0.60	0.63	263
8	0.54	0.60	0.57	243
9	0.49	0.40	0.44	230
micro avg	0.56	0.56	0.56	2494
macro avg	0.57	0.56	0.56	2494
weighted avg	0.58	0.56	0.57	2494

Adding Noise to 50 % samples

Training Result

	precision	recall	f1-score	support
0	0.72	0.33	0.45	1000
1	0.49	0.35	0.41	1000
2	0.44	0.29	0.35	1000
3	0.62	0.31	0.41	1000
4	0.67	0.35	0.46	1000
5	0.37	0.47	0.42	1000
6	0.22	0.87	0.35	1000
7	0.60	0.38	0.47	1000
8	0.54	0.41	0.47	1000
9	0.60	0.26	0.36	1000
micro avg	0.40	0.40	0.40	10000
macro avg	0.53	0.40	0.42	10000
weighted avg	0.53	0.40	0.42	10000

Validation Result

	precision	recall	f1-score	support
0	0.72	0.56	0.63	260
1	0.40	0.50	0.44	230
2	0.34	0.47	0.40	236
3	0.56	0.48	0.52	248
4	0.57	0.57	0.57	280
5	0.53	0.50	0.51	242
6	0.74	0.76	0.75	262
7	0.64	0.57	0.60	263
8	0.55	0.58	0.56	243
9	0.42	0.36	0.39	230
micro avg	0.54	0.54	0.54	2494
macro avg	0.55	0.53	0.54	2494
weighted avg	0.55	0.54	0.54	2494

MFCC

The audio files were zero-padded to make their length 16000 samples or 1 second in duration
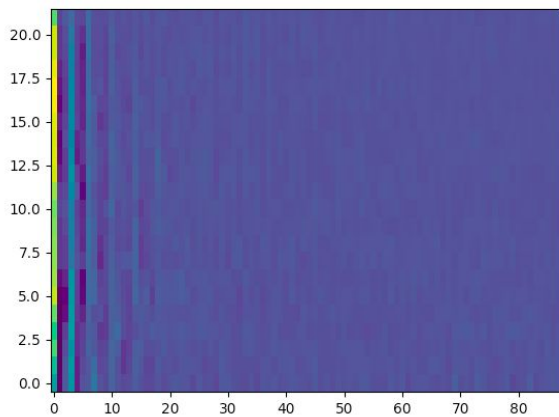
After due experimentation, the following values were chosen -
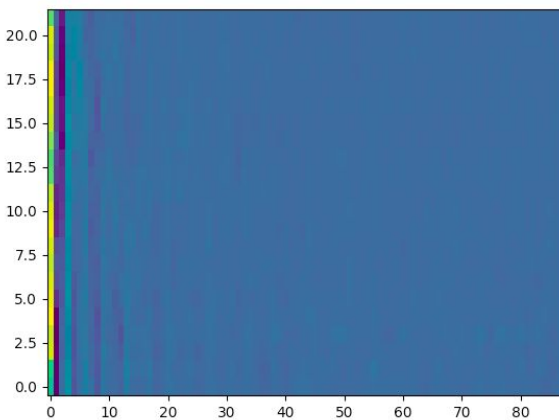1. Window Size - 15 ms
2. Overlap - 25%
3. Mel Filters - 22

This resulted in feature vectors of size 22 x 88
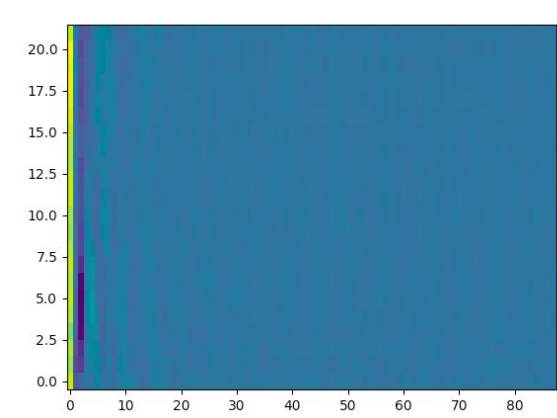
Sample MFCC from the training data

Class - Zero

Class - Five

Class - Nine

Results

Training Results

	precision	recall	f1-score	support
0	0.77	0.61	0.68	1000
1	0.47	0.62	0.54	1000
2	0.44	0.56	0.49	1000
3	0.67	0.52	0.58	1000
4	0.70	0.63	0.66	1000
5	0.60	0.63	0.61	1000
6	0.71	0.81	0.75	1000
7	0.73	0.60	0.66	1000
8	0.65	0.64	0.65	1000
9	0.57	0.55	0.56	1000
micro avg	0.62	0.62	0.62	10000
macro avg	0.63	0.62	0.62	10000
weighted avg	0.63	0.62	0.62	10000

Validation Results

Adding Noise to 25 % samples

Training Result

	precision	recall	f1-score	support
0	0.75	0.48	0.59	1000
1	0.48	0.48	0.48	1000
2	0.30	0.50	0.37	1000
3	0.67	0.40	0.50	1000
4	0.70	0.50	0.58	1000
5	0.47	0.56	0.51	1000
6	0.42	0.78	0.55	1000
7	0.68	0.47	0.55	1000
8	0.62	0.53	0.57	1000
9	0.54	0.43	0.48	1000
micro avg	0.51	0.51	0.51	10000
macro avg	0.56	0.51	0.52	10000
weighted avg	0.56	0.51	0.52	10000

Validation Result

	precision	recall	f1-score	support
0	0.76	0.66	0.71	260
1	0.43	0.49	0.46	230
2	0.31	0.42	0.36	236
3	0.60	0.52	0.56	248
4	0.68	0.60	0.64	280
5	0.57	0.57	0.57	242
6	0.73	0.79	0.76	262
7	0.68	0.59	0.63	263
8	0.62	0.59	0.60	243
9	0.47	0.48	0.47	230
micro avg	0.58	0.58	0.58	2494
macro avg	0.58	0.57	0.58	2494
weighted avg	0.59	0.58	0.58	2494

Adding Noise to 50% Samples

Training Result

	precision	recall	f1-score	support
0	0.71	0.36	0.48	1000
1	0.45	0.36	0.40	1000
2	0.25	0.36	0.30	1000
3	0.52	0.35	0.42	1000
4	0.67	0.36	0.47	1000
5	0.40	0.46	0.43	1000
6	0.27	0.77	0.40	1000
7	0.57	0.38	0.46	1000
8	0.49	0.43	0.46	1000
9	0.56	0.31	0.40	1000
micro avg	0.42	0.42	0.42	10000
macro avg	0.49	0.42	0.42	10000
weighted avg	0.49	0.42	0.42	10000

Validation Result

	precision	recall	f1-score	support
0	0.76	0.65	0.70	260
1	0.39	0.51	0.45	230
2	0.29	0.34	0.31	236
3	0.54	0.50	0.52	248
4	0.64	0.56	0.60	280
5	0.55	0.53	0.54	242
6	0.70	0.79	0.74	262
7	0.66	0.56	0.60	263
8	0.61	0.57	0.59	243
9	0.44	0.47	0.45	230
micro avg	0.55	0.55	0.55	2494
macro avg	0.56	0.55	0.55	2494
weighted avg	0.57	0.55	0.56	2494

Analysis

According to the results, the F1 score of MFCC 0.59 is slightly higher than that of Spectrogram 0.57 though the training results are the same. This might be due to the fact of considering Mel filter banks

For both the spectrogram and MFCC, adding 25% Noise did not decrease the validation results much but decreased the training results considerably. But adding 50% reduced both the results. Adding 25 % noise, the model is still able to learn the features and on removing the noise only the training results get better indicating a certain amount of overfitting

Decreasing the overlap and window length results in a decrease in inaccuracy. Increasing the overlap increase the number of features.

MFCC features are significantly smaller than the Spectrogram features but the results are almost similar. Due to fewer features, it takes considerably less time to train MFCC than Spectrogram