

Name: Anubhav Jaiswal

Roll No.: 2016014

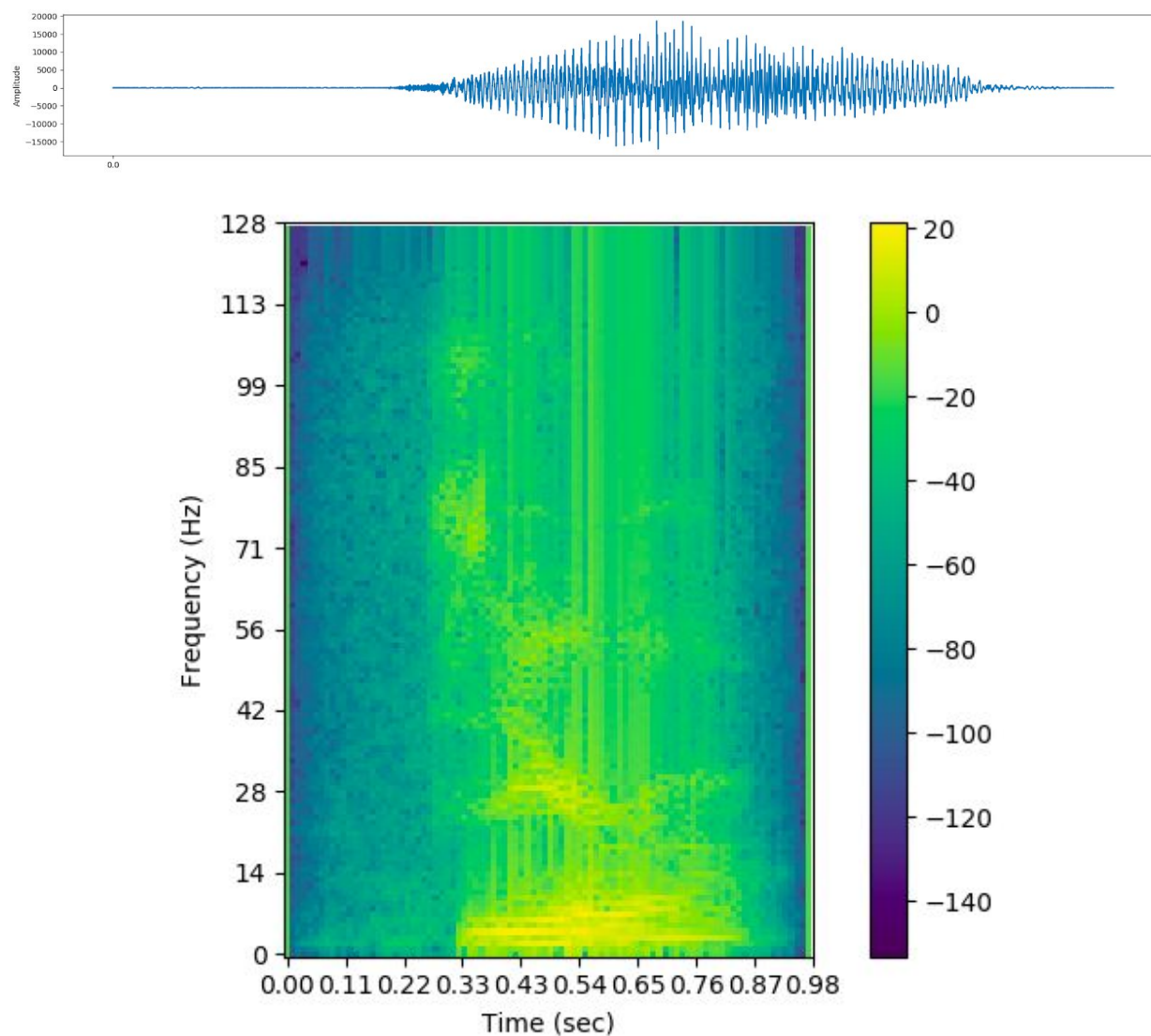
References:

1. <https://arxiv.org/pdf/1003.4083.pdf>
2. <https://kaldi-asr.org/doc/feat.html>
3. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>
4. <https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520>

Spectrogram

Graph

File: training/zero/0b56bcfe_nohash_0.wav



Best Result

	precision	recall	f1-score	support
0	0.68	0.71	0.69	260
1	0.53	0.57	0.55	230
2	0.43	0.49	0.46	236
3	0.55	0.60	0.57	248
4	0.72	0.67	0.69	280
5	0.57	0.62	0.59	242
6	0.76	0.77	0.77	262
7	0.69	0.56	0.62	263
8	0.67	0.63	0.65	243

9	0.60	0.53	0.56	230
accuracy			0.62	2494
macro avg	0.62	0.62	0.62	2494
weighted avg	0.62	0.62	0.62	2494

These results were achieved with -

1. Window Size: 256
2. Overlap: 84
3. Without Noise

Other Experiments

Without “Hanning” weights

showing results window:160 overlap:80

	precision	recall	f1-score	support
0	0.66	0.63	0.65	260
1	0.49	0.50	0.50	230
2	0.40	0.50	0.44	236
3	0.55	0.56	0.56	248
4	0.57	0.55	0.56	280
5	0.57	0.58	0.58	242
6	0.77	0.79	0.78	262
7	0.72	0.63	0.67	263
8	0.64	0.59	0.61	243
9	0.56	0.56	0.56	230
micro avg	0.59	0.59	0.59	2494
macro avg	0.59	0.59	0.59	2494
weighted avg	0.60	0.59	0.59	2494

showing results window:256 overlap:84

	precision	recall	f1-score	support
0	0.62	0.66	0.64	260
1	0.48	0.50	0.49	230
2	0.41	0.47	0.44	236
3	0.55	0.57	0.56	248
4	0.61	0.56	0.58	280

5	0.61	0.60	0.61	242
6	0.82	0.76	0.79	262
7	0.71	0.60	0.65	263
8	0.64	0.60	0.62	243
9	0.51	0.55	0.53	230

micro avg	0.59	0.59	0.59	2494
macro avg	0.59	0.59	0.59	2494
weighted avg	0.60	0.59	0.59	2494

Without Log reduction

showing results window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.41	0.63	0.50	260
1	0.34	0.48	0.40	230
2	0.39	0.50	0.44	236
3	0.48	0.47	0.47	248
4	0.77	0.59	0.67	280
5	0.45	0.52	0.48	242
6	0.79	0.57	0.66	262
7	0.60	0.41	0.49	263
8	0.74	0.45	0.56	243
9	0.56	0.52	0.54	230

accuracy			0.52	2494
macro avg	0.55	0.51	0.52	2494
weighted avg	0.56	0.52	0.52	2494

showing results window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.41	0.61	0.49	260
1	0.38	0.46	0.42	230
2	0.38	0.47	0.42	236
3	0.52	0.47	0.49	248
4	0.71	0.54	0.62	280
5	0.51	0.55	0.52	242
6	0.76	0.60	0.67	262
7	0.61	0.48	0.54	263
8	0.64	0.47	0.54	243

9	0.50	0.53	0.52	230
accuracy			0.52	2494
macro avg	0.54	0.52	0.52	2494
weighted avg	0.55	0.52	0.53	2494

With Hanning Without Log

showing results window:160 overlap:80

precision recall f1-score support

0	0.64	0.60	0.62	260
1	0.48	0.50	0.49	230
2	0.43	0.49	0.46	236
3	0.56	0.56	0.56	248
4	0.55	0.55	0.55	280
5	0.55	0.58	0.57	242
6	0.78	0.80	0.79	262
7	0.71	0.60	0.65	263
8	0.63	0.61	0.62	243
9	0.54	0.53	0.54	230

accuracy			0.59	2494
macro avg	0.59	0.58	0.58	2494
weighted avg	0.59	0.59	0.59	2494

showing results window:256 overlap:84

precision recall f1-score support

0	0.66	0.63	0.64	260
1	0.47	0.48	0.48	230
2	0.41	0.48	0.44	236
3	0.53	0.58	0.55	248
4	0.59	0.56	0.58	280
5	0.60	0.61	0.60	242
6	0.80	0.77	0.79	262
7	0.69	0.61	0.65	263
8	0.63	0.59	0.61	243
9	0.52	0.54	0.53	230

accuracy			0.59	2494
----------	--	--	------	------

macro avg	0.59	0.59	0.59	2494
weighted avg	0.59	0.59	0.59	2494

Noise Augmented

Sound:noise augmentation ratio 1:1

showing results specs: window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.43	0.45	0.44	260
1	0.35	0.41	0.38	230
2	0.29	0.40	0.34	236
3	0.37	0.42	0.39	248
4	0.49	0.40	0.44	280
5	0.47	0.44	0.45	242
6	0.65	0.56	0.60	262
7	0.49	0.40	0.44	263
8	0.46	0.38	0.42	243
9	0.35	0.39	0.37	230

accuracy			0.43	2494
macro avg	0.43	0.42	0.43	2494
weighted avg	0.44	0.43	0.43	2494

showing results specs:window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.42	0.45	0.44	260
1	0.31	0.39	0.35	230
2	0.26	0.35	0.30	236
3	0.37	0.40	0.39	248
4	0.50	0.42	0.46	280
5	0.47	0.45	0.46	242
6	0.63	0.54	0.58	262
7	0.44	0.38	0.41	263
8	0.53	0.43	0.48	243
9	0.37	0.36	0.36	230

accuracy			0.42	2494
----------	--	--	------	------

macro avg	0.43	0.42	0.42	2494
weighted avg	0.43	0.42	0.42	2494

Sound:noise augmentation ratio 7:3

showing results specs: window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.60	0.62	0.61	260
1	0.48	0.51	0.50	230
2	0.39	0.45	0.42	236
3	0.49	0.57	0.53	248
4	0.67	0.64	0.65	280
5	0.54	0.57	0.55	242
6	0.72	0.69	0.71	262
7	0.70	0.51	0.59	263
8	0.62	0.64	0.63	243
9	0.57	0.50	0.54	230

accuracy		0.57		2494
macro avg	0.58	0.57	0.57	2494
weighted avg	0.58	0.57	0.58	2494

showing results specs: window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

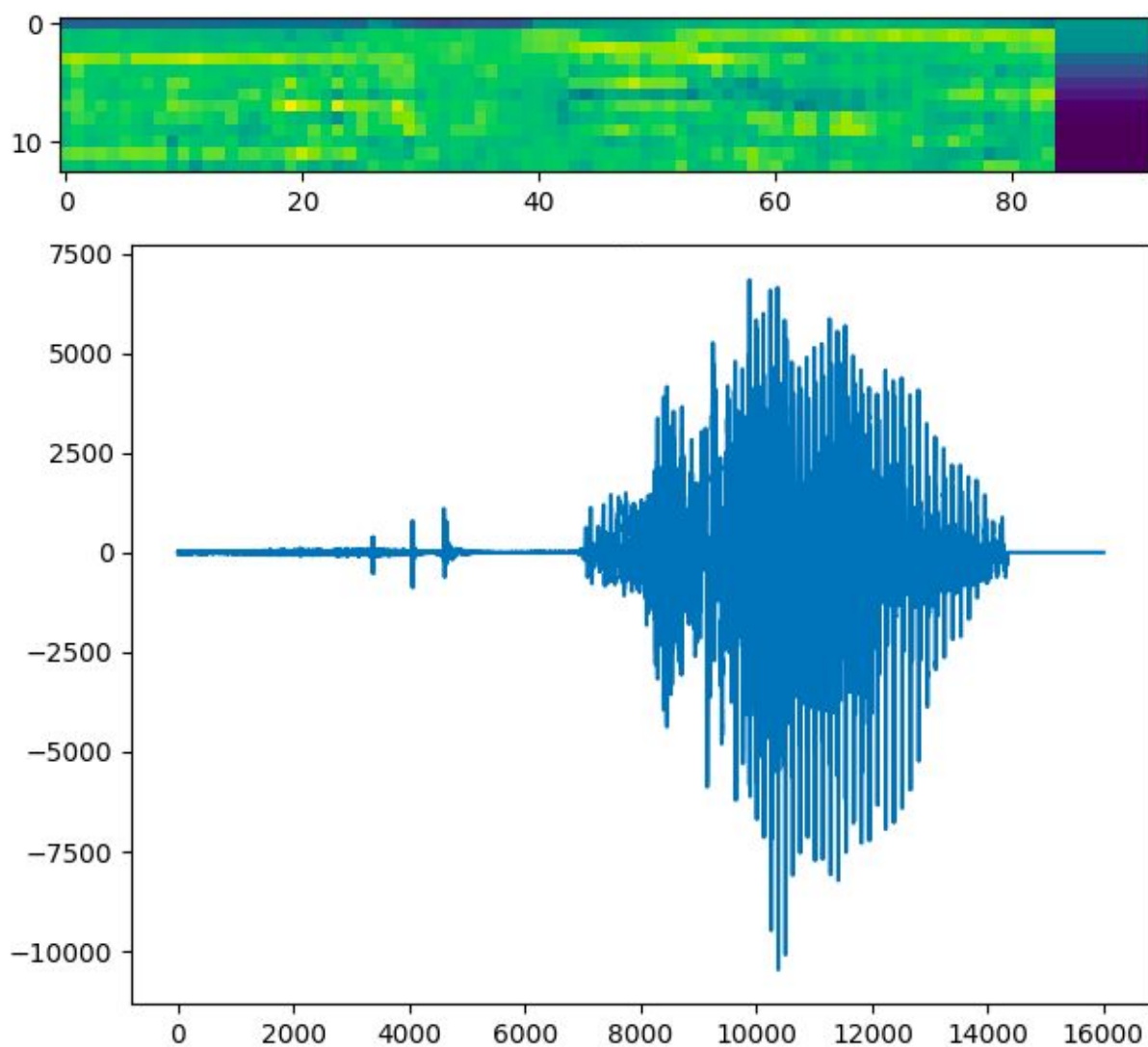
0	0.64	0.65	0.65	260
1	0.50	0.52	0.51	230
2	0.42	0.42	0.42	236
3	0.51	0.61	0.56	248
4	0.62	0.61	0.61	280
5	0.58	0.58	0.58	242
6	0.74	0.72	0.73	262
7	0.67	0.52	0.59	263
8	0.65	0.69	0.67	243
9	0.58	0.56	0.57	230

accuracy		0.59		2494
macro avg	0.59	0.59	0.59	2494
weighted avg	0.60	0.59	0.59	2494

MFCC

Graph

File: training/zero/0b09edd3_nohash_0.wav



Best Result

	precision	recall	f1-score	support
0	0.70	0.73	0.71	260
1	0.52	0.57	0.54	230

2	0.47	0.54	0.50	236
3	0.63	0.59	0.61	248
4	0.69	0.70	0.70	280
5	0.61	0.64	0.63	242
6	0.77	0.80	0.78	262
7	0.70	0.59	0.64	263
8	0.68	0.64	0.66	243
9	0.55	0.49	0.52	230

accuracy			0.63	2494
macro avg	0.63	0.63	0.63	2494
weighted avg	0.64	0.63	0.63	2494

This result was achieved with these hyperparameters

Mel filter bank bin count: 80

DCT cepstral features count: 13

Filter count: 22

Window Size for spectrogram features: 256

Overlap for spectrogram features: 84

Other Experiments

Basic

For below:

Mel filter bank bin count: 80

DCT cepstral features count: 13

Filter count: 22

showing results window:160 overlap:80

	precision	recall	f1-score	support
0	0.56	0.61	0.59	260
1	0.40	0.44	0.42	230
2	0.34	0.44	0.39	236
3	0.48	0.50	0.49	248
4	0.59	0.55	0.57	280
5	0.50	0.50	0.50	242
6	0.75	0.69	0.72	262
7	0.57	0.49	0.53	263
8	0.61	0.51	0.56	243

9	0.45	0.45	0.45	230
micro avg	0.52	0.52	0.52	2494
macro avg	0.53	0.52	0.52	2494
weighted avg	0.53	0.52	0.52	2494

showing results window:160 overlap:53

	precision	recall	f1-score	support
0	0.52	0.57	0.54	260
1	0.43	0.49	0.46	230
2	0.32	0.41	0.36	236
3	0.50	0.52	0.51	248
4	0.56	0.53	0.54	280
5	0.53	0.53	0.53	242
6	0.75	0.68	0.71	262
7	0.57	0.50	0.53	263
8	0.54	0.47	0.50	243
9	0.45	0.38	0.42	230
micro avg	0.51	0.51	0.51	2494
macro avg	0.52	0.51	0.51	2494
weighted avg	0.52	0.51	0.51	2494

showing results window:200 overlap:66

	precision	recall	f1-score	support
0	0.53	0.62	0.57	260
1	0.38	0.46	0.41	230
2	0.37	0.44	0.40	236
3	0.47	0.48	0.48	248
4	0.56	0.53	0.54	280
5	0.52	0.52	0.52	242
6	0.72	0.69	0.71	262
7	0.57	0.49	0.53	263
8	0.56	0.46	0.50	243
9	0.44	0.37	0.40	230
micro avg	0.51	0.51	0.51	2494

macro avg	0.51	0.51	0.51	2494
weighted avg	0.52	0.51	0.51	2494

showing results window:200 overlap:100

	precision	recall	f1-score	support
0	0.57	0.64	0.60	260
1	0.39	0.48	0.43	230
2	0.35	0.42	0.38	236
3	0.46	0.51	0.49	248
4	0.59	0.55	0.57	280
5	0.49	0.50	0.49	242
6	0.75	0.69	0.72	262
7	0.60	0.50	0.54	263
8	0.59	0.49	0.54	243
9	0.41	0.34	0.37	230

micro avg	0.52	0.52	0.52	2494
macro avg	0.52	0.51	0.51	2494
weighted avg	0.53	0.52	0.52	2494

showing results feature:mfcc specs:bins23_ceps13 window:160 overlap:80

	precision	recall	f1-score	support
0	0.56	0.60	0.58	260
1	0.41	0.47	0.43	230
2	0.36	0.47	0.41	236
3	0.46	0.51	0.48	248
4	0.57	0.55	0.56	280
5	0.49	0.45	0.47	242
6	0.75	0.70	0.72	262
7	0.58	0.49	0.53	263
8	0.58	0.48	0.53	243
9	0.44	0.41	0.43	230

accuracy			0.52	2494
macro avg	0.52	0.51	0.51	2494
weighted avg	0.52	0.52	0.52	2494

showing results feature:mfcc specs:bins23_ceps13 window:256 overlap:84

	precision	recall	f1-score	support
0	0.56	0.62	0.59	260
1	0.39	0.46	0.42	230
2	0.37	0.43	0.40	236
3	0.51	0.53	0.52	248
4	0.62	0.59	0.60	280
5	0.56	0.57	0.56	242
6	0.74	0.69	0.72	262
7	0.58	0.51	0.54	263
8	0.60	0.53	0.56	243
9	0.39	0.35	0.37	230
accuracy		0.53		2494
macro avg	0.53	0.53	0.53	2494
weighted avg	0.54	0.53	0.53	2494

With All Cepstral features

showing results window:160 overlap:80

	precision	recall	f1-score	support
0	0.65	0.68	0.67	260
1	0.54	0.53	0.54	230
2	0.40	0.48	0.44	236
3	0.53	0.54	0.54	248
4	0.65	0.66	0.65	280
5	0.57	0.62	0.59	242
6	0.75	0.74	0.75	262
7	0.66	0.55	0.60	263
8	0.66	0.56	0.61	243
9	0.53	0.52	0.53	230
micro avg	0.59	0.59	0.59	2494
macro avg	0.59	0.59	0.59	2494
weighted avg	0.60	0.59	0.59	2494

showing results feature:mfcc specs:preemp_fullceps window:160 overlap:80

precision	recall	f1-score	support
-----------	--------	----------	---------

0	0.61	0.66	0.63	260
1	0.48	0.52	0.50	230
2	0.39	0.48	0.43	236
3	0.55	0.55	0.55	248
4	0.63	0.61	0.62	280
5	0.57	0.55	0.56	242
6	0.75	0.70	0.73	262
7	0.62	0.55	0.58	263
8	0.63	0.54	0.59	243
9	0.45	0.45	0.45	230

accuracy		0.56		2494
macro avg	0.57	0.56	0.56	2494
weighted avg	0.57	0.56	0.57	2494

showing results feature:mfcc specs:preemp_fullceps window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.59	0.67	0.63	260
1	0.39	0.43	0.41	230
2	0.39	0.45	0.42	236
3	0.52	0.54	0.53	248
4	0.65	0.62	0.63	280
5	0.59	0.57	0.58	242
6	0.76	0.72	0.74	262
7	0.61	0.57	0.59	263
8	0.62	0.53	0.57	243
9	0.47	0.42	0.44	230

accuracy		0.56		2494
macro avg	0.56	0.55	0.55	2494
weighted avg	0.56	0.56	0.56	2494

showing results feature:mfcc specs:preemp_fullceps window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.59	0.67	0.63	260
1	0.39	0.43	0.41	230
2	0.39	0.45	0.42	236

3	0.52	0.54	0.53	248
4	0.65	0.62	0.63	280
5	0.59	0.57	0.58	242
6	0.76	0.72	0.74	262
7	0.61	0.57	0.59	263
8	0.62	0.53	0.57	243
9	0.47	0.42	0.44	230

accuracy			0.56	2494
macro avg	0.56	0.55	0.55	2494
weighted avg	0.56	0.56	0.56	2494

showing results feature:mfcc specs:preemp_fullceps window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.61	0.66	0.63	260
1	0.48	0.52	0.50	230
2	0.39	0.48	0.43	236
3	0.55	0.55	0.55	248
4	0.63	0.61	0.62	280
5	0.57	0.55	0.56	242
6	0.75	0.70	0.73	262
7	0.62	0.55	0.58	263
8	0.63	0.54	0.59	243
9	0.45	0.45	0.45	230

accuracy			0.56	2494
macro avg	0.57	0.56	0.56	2494
weighted avg	0.57	0.56	0.57	2494

showing results specs:preemp_bins80_ceps_80 window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.65	0.68	0.67	260
1	0.54	0.53	0.54	230
2	0.40	0.48	0.44	236
3	0.53	0.54	0.54	248
4	0.65	0.66	0.65	280
5	0.57	0.62	0.59	242
6	0.75	0.74	0.75	262
7	0.66	0.55	0.60	263

8	0.66	0.56	0.61	243
9	0.53	0.52	0.53	230

accuracy			0.59	2494
macro avg	0.59	0.59	0.59	2494
weighted avg	0.60	0.59	0.59	2494

showing results specs:preemp_bins80_ceps_80 window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.62	0.67	0.65	260
1	0.46	0.51	0.49	230
2	0.38	0.47	0.42	236
3	0.56	0.56	0.56	248
4	0.67	0.65	0.66	280
5	0.56	0.57	0.57	242
6	0.77	0.73	0.75	262
7	0.68	0.59	0.63	263
8	0.65	0.58	0.61	243
9	0.52	0.47	0.50	230

accuracy			0.59	2494
macro avg	0.59	0.58	0.58	2494
weighted avg	0.59	0.59	0.59	2494

With best Spectrogram Results

showing results specs:bins80_ceps13 window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.70	0.70	0.70	260
1	0.51	0.59	0.55	230
2	0.44	0.48	0.46	236
3	0.56	0.60	0.58	248
4	0.70	0.72	0.71	280
5	0.59	0.57	0.58	242
6	0.78	0.78	0.78	262
7	0.72	0.57	0.64	263
8	0.65	0.64	0.65	243
9	0.54	0.48	0.51	230

accuracy		0.62	2494
macro avg	0.62	0.61	0.61 2494
weighted avg	0.62	0.62	0.62 2494

showing results specs:bins80_ceps80 window:160 overlap:80

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.71	0.69	0.70	260
1	0.51	0.56	0.53	230
2	0.43	0.52	0.47	236
3	0.56	0.59	0.57	248
4	0.69	0.72	0.71	280
5	0.58	0.60	0.59	242
6	0.77	0.81	0.79	262
7	0.69	0.56	0.62	263
8	0.64	0.58	0.61	243
9	0.55	0.47	0.50	230

accuracy		0.61	2494
macro avg	0.61	0.61	0.61 2494
weighted avg	0.62	0.61	0.61 2494

showing results new_bins80_ceps80 window:256 overlap:84

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.73	0.69	0.71	260
1	0.54	0.56	0.55	230
2	0.45	0.52	0.48	236
3	0.58	0.61	0.59	248
4	0.73	0.72	0.73	280
5	0.59	0.60	0.60	242
6	0.75	0.80	0.78	262
7	0.70	0.60	0.64	263
8	0.66	0.63	0.64	243
9	0.56	0.52	0.54	230

accuracy		0.63	2494
macro avg	0.63	0.62	0.63 2494
weighted avg	0.63	0.63	0.63 2494

Noise Augmented

Sound:noise augmentation ratio 1:1

showing results specs:bins23_ceps13 window:160 overlap:80
support

precision recall f1-score

0	0.38	0.49	0.43	260
1	0.31	0.37	0.33	230
2	0.29	0.38	0.33	236
3	0.40	0.39	0.40	248
4	0.50	0.38	0.43	280
5	0.46	0.42	0.44	242
6	0.58	0.50	0.54	262
7	0.40	0.37	0.38	263
8	0.50	0.38	0.43	243
9	0.33	0.33	0.33	230

accuracy		0.40		2494
macro avg	0.41	0.40	0.40	2494
weighted avg	0.42	0.40	0.41	2494

showing results specs:bins23_ceps13 window:256 overlap:84
support

precision recall f1-score

0	0.40	0.52	0.45	260
1	0.33	0.41	0.37	230
2	0.28	0.36	0.32	236
3	0.36	0.39	0.37	248
4	0.52	0.41	0.46	280
5	0.46	0.48	0.47	242
6	0.61	0.50	0.55	262
7	0.44	0.35	0.39	263
8	0.52	0.40	0.45	243
9	0.34	0.29	0.31	230

accuracy		0.41		2494
macro avg	0.42	0.41	0.41	2494
weighted avg	0.43	0.41	0.42	2494

showing results specs:bins80_ceps13 window:160 overlap:80
support

precision recall f1-score

0	0.41	0.48	0.44	260
1	0.28	0.35	0.31	230
2	0.28	0.39	0.33	236
3	0.38	0.35	0.36	248
4	0.46	0.37	0.41	280
5	0.43	0.39	0.41	242
6	0.54	0.49	0.51	262
7	0.39	0.35	0.37	263
8	0.48	0.40	0.43	243
9	0.33	0.31	0.32	230

accuracy		0.39		2494
macro avg	0.40	0.39	0.39	2494
weighted avg	0.40	0.39	0.39	2494

showing results specs:bins80_ceps13 window:256 overlap:84
support

precision recall f1-score

0	0.39	0.50	0.44	260
1	0.31	0.41	0.35	230
2	0.27	0.36	0.30	236
3	0.44	0.42	0.43	248
4	0.49	0.40	0.44	280
5	0.45	0.41	0.43	242
6	0.52	0.45	0.48	262
7	0.40	0.39	0.40	263
8	0.52	0.40	0.45	243
9	0.40	0.33	0.36	230

accuracy		0.41		2494
macro avg	0.42	0.41	0.41	2494
weighted avg	0.42	0.41	0.41	2494

showing results specs:bins80_ceps_80 window:160 overlap:80
f1-score support

precision recall

0	0.43	0.47	0.45	260
1	0.30	0.36	0.33	230
2	0.28	0.41	0.33	236
3	0.36	0.35	0.35	248
4	0.52	0.37	0.43	280
5	0.46	0.45	0.45	242
6	0.55	0.51	0.53	262
7	0.42	0.40	0.41	263
8	0.51	0.37	0.43	243
9	0.37	0.38	0.37	230

accuracy		0.41		2494
macro avg	0.42	0.41	0.41	2494
weighted avg	0.42	0.41	0.41	2494

showing results specs:bins80_ceps_80 window:256 overlap:84
f1-score support

precision recall

0	0.40	0.50	0.44	260
1	0.31	0.37	0.34	230
2	0.26	0.34	0.29	236
3	0.36	0.34	0.35	248
4	0.52	0.36	0.43	280
5	0.46	0.40	0.43	242
6	0.57	0.47	0.52	262
7	0.39	0.43	0.41	263
8	0.52	0.40	0.45	243
9	0.33	0.34	0.34	230

accuracy		0.40		2494
macro avg	0.41	0.40	0.40	2494
weighted avg	0.42	0.40	0.40	2494

Sound:noise augmentation ratio 7:3

showing results specs: bins80_ceps13 window:160 overlap:80
precision recall f1-score support

0	0.67	0.71	0.69	260
1	0.49	0.63	0.56	230
2	0.50	0.53	0.51	236
3	0.58	0.62	0.60	248
4	0.69	0.63	0.66	280
5	0.62	0.68	0.65	242
6	0.79	0.75	0.77	262
7	0.66	0.54	0.59	263
8	0.63	0.57	0.60	243
9	0.56	0.50	0.53	230

accuracy		0.62		2494
macro avg	0.62	0.62	0.62	2494
weighted avg	0.62	0.62	0.62	2494

showing results specs: bins80_ceps13 window:256 overlap:84
precision recall f1-score support

0	0.74	0.71	0.72	260
1	0.51	0.59	0.55	230
2	0.47	0.48	0.47	236
3	0.56	0.64	0.60	248
4	0.68	0.69	0.68	280
5	0.64	0.64	0.64	242
6	0.76	0.76	0.76	262
7	0.69	0.56	0.62	263
8	0.66	0.62	0.64	243
9	0.57	0.54	0.55	230

accuracy		0.63		2494
macro avg	0.63	0.62	0.62	2494
weighted avg	0.63	0.63	0.63	2494

Analysis

The MFCC features performed slightly better than Spectrogram features.

One of the main highlights was the validation set was not correct, i.e., the sound clips are not categorized correctly in the folders. All classes performed similarly, however, class "2" was not

able to achieve a similar score. Class "6" achieved the highest precision of all. The features improved a lot by the addition of Hanning weights. Only 13 cepstral features were taken, representing 13 energy bands. These performed better than taking all of them, due to the elimination of redundancy and noise bands. Emphasizing on the signal, we only wanted the higher frequencies, as the lower frequencies are background noises. Thus came in pre-emphasis coefficient. Even after augmenting noise onto the files, the model wasn't robust and the performance was as per before.