Isolated vowel recognition using Gaussian Mixture model report

Mathav raj - 194102311

   A Gaussian mixture of 16 densities was trained for each digit utterance by expectation maximization algorithm using sklearn library in python. 39 dimensional MFCC features has been extracted from each signal for training. In the testing phase the posterior probability of each utterance to each of the 10 digit models is computed as the per sample average log likelihood. Then the argmin of the maximum aposterior gives the digit model to which the sample belongs to. A confusion matrix was plotted for the test data.

Experiment:

- Training data - Digit recordings of ee624 class
- Test data - Digit recordings of 194102311

Parameters and results:

- Sampling frequency = 16 kHz
- Window length for mfcc computation = 20ms
- Frame shift = 10 ms
- Accuracy = 0.54

Confusion matrix:

	0	1	2	3	4	5	6	7	8	9
0	23	0	0	0	0	0	0	1	1	0
1	5	19	0	0	0	0	0	0	0	1
2	0	0	25	0	0	0	0	0	0	0
3	1	0	0	9	0	3	0	0	12	0
4	4	13	4	0	4	0	0	0	0	0
5	0	1	0	0	0	20	0	0	0	4
6	0	0	0	2	0	3	7	0	13	0
7	0	0	0	0	0	18	1	6	0	0
8	4	0	0	2	0	11	1	0	7	0
9	0	5	0	0	0	11	0	3	0	0