SPEAKER RECOGNITION SYSTEM

Experiments:

In the experiments I have used the standard setting with variation in the parameter being experimented on.

Standard setting:
Window length = 30 ms
Window shift = 15 ms
Number of GMM mixtures = 16
Training/Testing split = 80/20
Voiced/Unvoiced threshold = 0.05
Iterations for GMM training = 100

(1) Varying number of GMM mixtures

GMM mixtures	Accuracy
8	0.836734693878
16	0.887755102041
32	0.867346938776
64	0.760204081633

Increasing the number of mixtures improves accuracy upto a certain point (16 in this case) and then decreases as the number of mixtures becomes too much in proportion to the number of speakers.

(2) Varying Training/Testing split

Split	Accuracy
80-20 split	0.887755102041
70-30 split	0.816326530612

Keeping other parameters same and decreasing training samples, decreases the accuracy.

(3) Using Delta features

In the standard setting adding the delta and delta-delta features decreases the accuracy, **0.867346938776** to **0.816326530612**.

(4) Varying Voiced/Unvoiced threshold

Threshold	Accuracy
0.03	0.887755102041
0.05	0.867346938776
0.07	0.882653061224

(5) Varying number of iterations for GMM training

Iterations	Accuracy
25	0.867346938776
50	0.867346938776
75	0.882653061224
100	0.887755102041

On increasing the number of iterations for the EM algorithm, the GMMs are better fitted and therefore the accuracy also increases.