

Independent Study

Sreenatha Bhatlapenumarthy

Roll no : 201001106

Training samples	Testing samples	Length of training sample (in sec)	Length of test sample (in sec)	Mixtures	Percentage of accuracy
6	2	90	10	16	12.96
6	2	90	10	32	12.9630
5	2	90	10	16	14.8148
4	2	90	10	16	16.6667
4	2	90	10	32	18.5185

Table 1. Results of language identification using GMM

Details

- Data source : [Prasar Bharathi](#)
- Speech parameter : Mel-frequency cepstrum(MLCC)
- Language model : Gaussian Mixture Model(GMM)
- Number of languages used : 27 (all Indian)
- Constraints enforced on the data
 - No 2 samples in training and test data should be of the same speaker.
 - Speakers equally distributed across gender(1 male, 1 female in test data).
 - No noise present in the samples.