ASR Assignment

Problem Statement:
Performing phoneme classification on the TIMIT dataset by training and testing gmm models for each phoneme.

Procedure :
i) *Preprocessing*
a) Downloaded and extracted the timit dataset
b) Converted the wav files from NIST “SPHERE” file format to WAVE PCM format using “sphn_to_wav.py” file

ii) *Training*
a) Used “import_timit.py” file to generate mfcc (13 dim), mfcc + delta (26 dim), mfcc + delta + delta (39 dim) features. They are saved in “features/” as train1.hdf, train2.hdf and train3.hdf respectively.
b) train.ipynb reads the mfcc features, extracts the feature and label array, and fits a gmm for each phoneme.
c) The gmms are then saved onto the system using the pickle library

iii) *Testing*
a) Similar to the training process, mfcc features are generated, for the test directory of TIMIT
b) Feature and label array are extracted. This label array acts as the ground truth for the test dataset.
c) The trained gmms are loaded from disk.
d) MAP rule is applied, to classify each feature vector in the sequence, for a given test file (40-way classification), and the predicted labels are stored in an array
e) Frame-level accuracy and Phoneme error rate are calculated using ground truth array and predicted labels array

Results:

MFCC feature dim	Mixture number	Frame-level accuracy	Phoneme error rate
With Energy co-efficients			
13	2	11.61936	89.15897
	4	13.65873	88.92773
	8	14.60678	88.56927
	16	15.02723	88.22339
	32	14.85941	88.05866
	64	14.34398	87.76315
	128	13.84426	88.08959
	256	13.51503	88.15495
26	64	18.93172	87.58913
39	**64**	**19.13807**	**82.85112**
Without Energy co-efficients			
12 (13-1)	64	13.23828	88.89605
24 (26-2)	64	17.12372	84.91502
36 (39-3)	64	17.14276	84.90527

Observations:
From the results tabulated, we can conclude that the gmm model with 64 components and running on mfcc + delta + delta features (39 dim), is the best, with the maximum training accuracy and the least phoneme error rate among all the 13 different models.

Best model : 64 component gmm on 39 dim mfcc+delta+delta features
Testing accuracy: 19.138068458575034 %
Phoneme error rate: 82.85112155638635

Team Members:
Pranav S – IMT2014039
Tummala Armitha – IMT2014060
Vutla Keerthi Chandra – IMT2014064