EE679 Computing Assignment 4 Due: Nov 15, 2018

Spoken Digit Recognition

You are provided recorded audio suitable for developing a template-based digit recognition system. Four utterances of each of the 10 digits ("Zero" to "Nine") sampled at 8 kHz recorded by each of several speakers are provided in the accompanying folder. With N speakers, you can train and test your digit recognition system in leave-one-speaker out (or "N-fold cross-validation") mode. You can thus test your digit recognizer on 4xNx10 = 40N words.

1. Develop an **end-pointer** using speech/silence detection that enables the automatic segmentation of the individual digit utterances from the continuous audio record. Obtain the pre-emphasised signal corresponding to each utterance.

2. Develop a **feature extractor** that computes an MFCC feature vector for every 10 ms frame of an utterance.

3. Develop a digit recognizer based on the "bag of frames" approach with a codebook for each digit created out of training set speakers' data. Provide the achieved word error rate (WER) in terms of %words incorrectly detected in the N-fold CV testing using a VQ codebook for each digit obtained via K-means clustering. Provide the achieved WER for with different numbers of clusters (e.g. 4, 8, 16, 64). Observe the common confusions, and comment on your results.

4. Repeat 3. above at K=64 but with different features, e.g. MFCC computed after pre-emphasis, concatenating spectral dynamics related features. Provide WER for train-test matched and mismatched conditions (train on male and test on female data).

5. Develop a template-matching digit recognizer based on DTW alignment and distance computation. Provide the achieved WER in N-fold CV evaluation. Observe the common confusions, and comment on your results.

Submit a single report describing your methods, observations, results and critical discussion along with your code snippets.