EE679 Computing Assignment 4 Due: Nov 5, 2017

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** 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. (Optional: in the subsequent parts, compare performance with/without spectral dynamics related features)

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 (i) training vectors directly, and (ii) a VQ codebook for each digit with different numbers of clusters (e.g. 4, 8). Observe the common confusions, and comment on your results.

4. 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.