EN.520.680 Speech and Auditory Processing by Humans and Machines
Homework 5

Due: April 21^{st}

Attached are speech utterances of ten digits (zero to nine) from two male and two female speakers. Each speaker has 2 trials of the same utterance recorded in different sessions (for example, the folder `male1_trial1` has ten digits spoken by male speaker 1 recorded in trial 1).

1. Compute the mel-spectrogram features of each utterance from `male1_trial1`. Use a Dynamic Time Warping (DTW) algorithm to classify mel-spectrograms from all the remaining trials. In order to classify an utterance, compute DTW distance between the mel-spectrogram of that test utterance and all training utterances in `male1_trial1` and choose the class that has the smallest distance. Report the recognition accuracy for each speaker and each trial and analyse your results. (**Points:** 25)

2. Repeat part 1 with Perceptual Linear Prediction (PLP) features. Compare and analyse your recognition results with those obtained with mel-spectrogram (**Points:** 25)

You can use ANY python/MATLAB packages for this assignment. For example

- Mel-spectrogram can be computed with
  python - `https://librosa.org/doc/main/generated/librosa.feature.melspectrogram.html`
  MATLAB - `https://www.mathworks.com/help/audio/ref/melspectrogram.html`.
- PLP can be computed with the `rastaplp` methods in
  python - see attached
  MATLAB - `https://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/`