

BIL735 Speech Recognition:
Assignment 3,
DTW for isolated word recognition
5 April 2017

In this assignment you will be required to record digits many many times and build DTW and HMM-based isolated word recognition systems. In total, you will need to record each of the digits 0,1,..9 ten times each. Each recording must be isolated, with no more than half a second of preceding and trailing silence. The details of the problem are below.

The first problem is on DTW-based recognition.

- Write a routine to perform DTW between a template feature vector sequence and a “data” feature vector sequence. The feature vectors must be 39-dimensional features (cepstra/delta cepstra/double delta) that are obtained with the code you wrote for assignment 1.
- Record one instance each of digits zero, one, two, three, four, five, six, seven, eight, nine as templates. You may use your code from assignment 1 to record the data. Compute feature vector sequences from them to act as templates.
- Record a further five instances each of the same digits (isolated words) as test data. Compute feature vectors from them and recognize them using the DTW code you just wrote for assignment 2. Report recognition accuracy.
- Redo the earlier step with time-synchronous DTW.
- Redo the above with pruning. Use relative pruning. Plot pruning threshold as a function of recognition accuracy.
- Record yet another set of four recordings of each of the digits. Including the first set of recordings you will now have five templates for each digit. Repeat the recognition experiment above using multiple templates for each word. Plot how the recognition accuracy changes as a function of the number of templates of the words (1 through 5). Do this using time-synchronous DTW, without pruning. Repeat it using a pruning threshold determined from the plot in the earlier experiment.

Due: 19 April 2017.