EE624: Speech Technology Course Projects

Submission date: On or before May 02, 2023; Strict deadline: May 13, 2023

Project 1: Perform DTW based isolated vowel recognition

- Record **twenty five** utterances of each of the isolated English digits (***0-9***), vowels (***a,e,i,o,u***), and the given sentence (***I'm registered with speech technology course and recording this data for course project***) in your voice. Record speech data over PC using ***Audacity*** software in PCM format with a sampling frequency of 16 kHz and a precision of 16-bits. Store the recorded data as: **<Roll.No>_<digit>/<vowel>/sentence_<recording_num>.wav.** Human voice characteristics usually differ slightly over the time and it is referred to as the **session variation**. So you are advised to record different entities at least in two sessions.

- For each of the vowels, choose one of the utterance as the reference template and the remaining ones as the test templates. Compute 39-dimensional MFCC features comprising base, delta, and delta-delta components for all the recordings using MATLAB. Using DTW command of MATLAB, write a script that computes DTW scores for any two similar or dissimilar test utterances and also plots the corresponding DTW curve.

Project 2: Perform GMM based isolated digit recognition

- Given the training and testing file lists, compute 39-dimensional MFCC features and learn a GMM having 16/32 densities for each of the isolated digits using training data.
- Identify the digits in test data using the maximum likelihood rule over the trained GMMs and report the confusion matrix for the recognition task.

Project 3: Develop context dependent subword GMM-HMM and DNN-HMM based ASR systems for given speech data using Kaldi toolkit

1. GMM-HMM based ASR systems
    - Compute MFCC features for the given continuous speech database
    - Train mono-phone model for the vocabulary
    - Create tri-phone HMM models and train tied system with 8 Gaussian per state
    - Train a bigram LM using the training transcript
    - Perform Viterbi based decoding of the test set
    - Evaluate performance

2. Develop a hybrid DNN-HMM system for continuous speech recognition.
    - Train a feed-forward neural network to map MFCC features to the triphone posterior probabilities produced by the GMM-HMM system developed in part 1
    - Perform Viterbi based decoding of the test set using bigram LM
    - Evaluate performance

Project 4: Perform DTW based speaker verification with and without GMM posteriogram mapping.

- Implement Section 3 in paper (*Sarfaraz Jelil, et al., "Speaker Verification Using Gaussian Posteriorgrams on Fixed Phrase Short Utterances", Interspeech, 2015*).

****** **Deliverables for each course project will be pointed out during the course** ******