NLP703: Speech Processing (Spring 2024)
Week 6 Lab Assignment

Release Date: 14 February 2024
Due Date: 20 February 2024

- Submit your answers as an electronic copy on Moodle (pdf + jupyter notebook).
- No unapproved extension of deadline is allowed.
- Cite your sources if you are taking help (papers, websites, students etc.).
- Plagiarism is strictly prohibited. Negative mark will be assigned for plagiarism.
- **Remember to comment your code. And your answers should be detailed.**

Dataset for this lab is the BuckEye Corpus. Availble on HPC */l/users/BuckEyeCorpus/BuckEye*.

Question 1: Feature Extraction

(a) Extract 10 second audio from the dataset

(b) Plot the audio

(c) Generate and plot the mel-spectrogram

(d) In the spectrogram, try to manually identify any visible phonetic features (e.g. stops, vowels, fricatives)

Question 2: Data Preparation

(a) Split and save all audios in the dataset by silences

(b) Generate MFCC features for each audio segments and align them with the corresponding phones

(c) Create a Dataset of MFCC,phone input output pairs

(d) Split the data into test and train set

Question 3: Train and Evaluate Classifiers

(a) Train multiple (2 or more) classifiers. One classifier can be a simple classifier with an individual input frame as input. At least one classifier should take more context (e.g. a context window around the target frame, or a recurrent model)

(b) Evaluate each classifier. Evaluation should be reported in frame error rate, in addition to qualitative evaluation by looking at specific utterances and their predicted labels. Show ane example prediction in your report.

Question 4:

Write a paragraph explaining the advantages and drawbacks of this phone classification approach for ASR.