Homework Assignment # 5

Due Date: Friday, November 22, 2019 by 11:59PM
Total Points: 100

This assignment introduces you to Google's API for converting speech to text (e.g. automatic speech recognition). More specifically, you will test Google's ASR system using data that you generate from the TIMIT speech corpus. **For the purposes of this assignment, please work individually**. However, feel free to ask basic questions to the instructor and classmates, but complete the assignment on your own. For specific instructions about Google Cloud Credits, be sure to contact the AI (Yuchen Liu/liu477@iu.edu)

You must submit a typed report (or Juypter Notebook file with adequate comments) that introduces the problem and discusses all results. All documents should be submitted through Canvas.

All assignments must be submitted on time to receive credit. No late work will be accepted, unless you have a prior arrangement with the instructor.

Question 1. [100 POINTS]

1. **Download TIMIT dataset**
   - The dataset is located at: Canvas → Box Course Folders. Go to COURSE FILES → Speech Data → TIMIT_full.zip
   - Download and unzip the file, then store the contents locally.
   - **Read the documentation to become familiar with the dataset**
2. **Google Cloud Account and Credits**: In Canvas, there is a Jupyter notebook file. Carefully follow the steps in the file to setup a Google Cloud account, project and to enable the Google speech-to-text API.
3. **Google Cloud Speech-to-Text**: In the Juypter notebook file, sample code is provided that shows how to use the Google Cloud Speech-to-text API. Read through this code carefully to understand what is happening.
   - Modify the code to predict text transcriptions for the TIMIT testing data. Compute/print the mean and standard deviation of the word-error rate across all testing signals. Discuss how these results compare with the results from the traditional ASR approaches (e.g. Kaldi assignment HW #4).
   - ASR performance can suffer when competing speech is present. Perform the following to determine how well Google's ASR system performs in adverse environments:
     - Choose a single speech signal from the TIMIT training set to serve as 'noise' (e.g. competing speech). You may need to replicate this signal to ensure that it is long enough.
     - Combine this signal with the TIMIT test data signals at -5, 0, 10, and 25 dB SNRs.
     - Use Google's ASR system to generate text transcriptions of the 'noisy' signals. Compute/print the mean and standard deviations of the WER separately at each SNR. Meaning, you should have 4 different mean/standard deviation results.
     - Discuss how performance varies as SNR decreases, including a comparison with the 'clean' signals.