E9 261 – Speech Information Processing

Homework # 2
Due date: February 21, 2022

Please upload (in the course webpage under section 'Your Voice/files to upload:') your recordings and codes as a zipped folder (or in multiple zip files with filenames have part1 part2 etc. each not exceeding 10Mb). In the zipped folder the program names should be self explanatory. Include a README file with every program (mandatory). Filename of each program should contain the question number it is associated with (follow instructions and examples as was given in the first HW).
Course Webpage: https://ee.iisc.ac.in/∼prasantg/e9261_speech_jan2022.html

1. **Best Band**
   Use PRAAT to record and save (as .wav file) five different sentences from five of your friends (both male and female; one sentence from one friend). Use sampling rate of 16kHz for recording. Now design a band-stop filter with stop-band from FHz to $(F + 1000)$Hz. For each recording do the following:

   Vary F from 0Hz to 7000Hz in a step of 500Hz and for each of the choice of F, band-stop filter the original recording and save the filtered signal as .wav file. Listen the filtered .wav file and compare with the original .wav file to give your score of distortion in the filtered signal in a scale of 0-10 (10 - filtered and original identical, 0 - they are different).

   Tabulate the score for all five recordings and different F and report which 1kHz band results in the best score for each of five recordings? If the best band is different for different recordings, argue why that could be so.

2. **Speaking rate**
   Use PRAAT to record and save (as .wav file) a paragraph containing at least 50 words. Make five different recordings of the same paragraph. Use CMU dictionary to convert the paragraph into a sequence of phonemes. Compute average and standard deviation of the following two quantities: A) word rate (i.e., number of words per second), B) phoneme rate (i.e., number of phonemes per second).

3. **CMU dictionary**
   Download the dictionary from
   http://svn.code.sf.net/p/cmusphinx/code/trunk/cmudict/cmudict-0.7b
   Then write programs to report the followings:

(a) Generate a plot of $C(k)$ vs k where k=1, 2, 3, $\cdots$, 30 and $C(k)$ is the number of words having phonetic transcription with length less than equal to k.

(b) The histogram of the phonemes in the entire dictionary

(c) A phonemic sequence for a given sentence with words from the dictionary. (i.e., your program should take a sentence as input)

(d) Number of words in the dictionary for which number of alphabets in the word is identical to the number of phonemes in the corresponding phoneme transcription.

4. **g2p**
Create a grapheme-to-phoneme converter using CMU dictionary and the open source tool available at:
http://www-i6.informatik.rwth-aachen.de/web/Software/g2p.html

Use this converter and get phoneme sequence for any ten words which are not there in the CMU dictionary.

5. **Know your formants and change them**
Record (using PRAAT) sustained phonation (by yourself) of all vowels (at 8kHz) listed in the classnote. Analyze the recording in PRAAT to obtain first three formants (F1, F2, F3) for each vowel uttered by you. Are they identical to those given in the lecture notes in the class? If not, can you guess why?
For each vowel (referred to as source vowel) create a piecewise linear frequency map with every other vowel (with formants F1', F2', F3'), referred to as target vowel, such that F1, F2, F3 are mapped to F1', F2', F3'. Create a spectrum for the target vowels by frequency scaling the spectrum of the source vowel using this map and then use inverse Fourier transform to obtain time domain signal for the target vowel. Listen to these converted signals and summarize which all converted signals are perceived identical to the target vowels. Discuss your observations.

6. **Long term average spectrum (LTAS)**
Go the section titled 'Transcript for recording' in the course webpage. You will find sentences in six groups in
https://ee.iisc.ac.in/∼prasantg/downloads/final/HW2_recording.txt
each group having 50 sentences. Read setences in each group (following instructions given in the webpage for recording and upload) and record (in PRAAT). Take every 0.02 sec signal segment and compute power spectral density (PSD) for each sentence recording. Average all PSDs from all segments. This is called Long term average spectrum (LTAS). Plot the LTAS of your voice separately for each group. Similarly, record four different sounds (preferably non-stationary) each for at least 5 minutes (e.g., traffic noise, restaurant noise etc.). Compute and plot LTAS for each of these four recordings. Which LTAS among these four is the closest to the six LTASs of your voice?