

E9 261 – Speech Information Processing

Homework # 1

Due date: February 11, 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.

Course Webpage: https://ee.iisc.ac.in/~prasantg/e9261_speech_jan2022.html

1. Select any five English words and another five sentences in English. Get each of these words and sentences recorded when one of your friends (or family members) speaks them three times in each of slow, normal and fast rates. Do this from a total of 10 of your friends. Normal rate is a rate your friend is comfortable speaking with. Slow and fast rates are expected to be at 1/2 and 2 times the normal rate. Compute the average and standard deviation of the duration of each word and sentence in each rate from all of your ten friends’ recordings and summarize your observations. Upload the recordings in the course webpage. Each recording file must be named properly. For example if you (AJAY) get a recording from your friend (RAHUL) who spoken the first English word in the slow rate, then three files will be
HW1_Q1_AJAY_RAHUL_WORD1_SLOW_1.wav
HW1_Q1_AJAY_RAHUL_WORD1_SLOW_2.wav
HW1_Q1_AJAY_RAHUL_WORD1_SLOW_3.wav
Similarly, for the recordings of third sentence in fast rate, the filenames will be
HW1_Q1_AJAY_RAHUL_SENT3_FAST_1.wav
HW1_Q1_AJAY_RAHUL_SENT3_FAST_2.wav
HW1_Q1_AJAY_RAHUL_SENT3_FAST_3.wav
2. Record ten different English words spoken by you. Also, record a noise of your choice. Add different realizations from the noise recording at 30dB SNR to each of the ten words’ recordings. Make these noisy files listened by five of your friends (or family members) and find out how many of your friends could correctly recognize each word (this gives you the word recognition accuracy). Compute the word recognition accuracy for each word as you reduce the SNR from 30dB to -10dB with a stepsize of 10dB. Plot the word recognition accuracy vs SNR and summarize your observations. You may want to choose a different set of five friends for each SNR;

otherwise, your friend will already know the word when they listen the recording at high SNR. Upload your recordings (original and NOT noisy) in the course webpage. (you can use online option, e.g., Google forms, for carrying out listening experiments). Follow an appropriate filename protocol: HW1_Q2_AJAY_WORD1.wav (for example)

3. Take a set of ten English sentences. Record as you speak these sentences at a distance of 6 inches from the microphone. Repeat all sentences separately when you are at a distance of 2.5 feet and 5 feet from the microphone. Present these 30 recordings in random order and ask one of your friends (or family members) to listen and classify each recording into one of the three distances' categories. Do this with a total of five friends and report the classification accuracy. Summarize your observations. Can you do this classification task automatically by a computer? You can use any classifier you know for this purpose. Upload your recordings in the course webpage. Follow an appropriate filename protocol: HW1_Q3_AJAY_SENT6_2p5FEET.wav (for example)
4. Decide a topic and request two of your friends (or family members) to have a conversation in English for 5min on that and record the same. Make sure both of your friends are at equal distance from the microphone (note down the distance and report it). Save this as conv_english.wav. Have another recording of a conversation in native language of the friends on the same topic and save as conv_nl.wav. Now prepare a paragraph on the same topic having 400-500 words in both English and the same native language. Record as you (or your friend) read the paragraph out and save them as read_english.wav and read_nl.wav, respectively. Carry out sentence level labeling using Audacity separately for each of these four wavfiles and save them as conv_english.lab, conv_nl.lab, read_english.lab and read_nl.lab, respectively. Upload your recordings and label files (Add HW1 and Q4 and your name in both filenames as in previous questions) in the course webpage.
 - (a) Compute the number of words per second and report for each of the four recordings. Summarize your observations.
 - (b) Compute absolute value of the difference in pitch between two consecutive frames in all voiced segments. Report the mean and standard deviation of it for all of your four recordings. Summarize your observations.
 - (c) Select any other parameter(s) of your choice and report the same for each of the four recordings. Summarize your observations.