E9 261 - Speech Information Processing

Midterm# 1

Total Points: 100

Due date: 11:59PM, April 18, 2021

Please upload (in the course webpage) your recordings and codes (or in multiple zip files with filenames having part1 part2 etc. each not exceeding 10Mb). In the zipped folder, the program names should be self explanatory. Mention the filenames you upload in your answer book.

Record as you speak each of the following six vowels (V1, V2, V3, V4, V5, V6)
V1 as in Hid
V2 as in Head
V3 as in Had
V4 as in Hudd
V5 as in Hod
V6 as in Hood
in a sustained manner for at least 1 second at 16kHz. Record each vowel separately for ten times. Trim 1second of vowel from each recording. First eight recordings of all vowels form the training set and remaining two from all vowels form the testset. You need to carry out a frame-level six-class vowel classification experiment using kNN classifier. The accuracy of the classification is computed as the percentage of frames correctly classified in the test set. For this purpose, consider a frame length of 20msec and a frame shift of 10msec. As a representative feature vector in each frame, you need to experiment with three features: 1) 2-dimensional comprising first and second formants, 2) 13-dimensional mel frequency cepstral coefficients (MFCCs) , 3) 12-dimensional MFCCs excluding the 0-th coefficient.

A. *(15 points)* Report the classification accuracy using all three features on the test set for K=1, 5, 10, 50, 100, 500, 1000 in the kNN classifier. Which feature vector resulted in the highest classification accuracy? What in this feature vector makes it achieve the highest classification accuracy? Consider only the best performing feature vector for the rest of the questions.
B. *(10 points)* Report the confusion matrix for the value of K which resulted in the best classification accuracy. Which vowels are mis-classified the most? Provide your explanation in this regard. For the rest of the questions, use the value of K which resulted in the best classification accuracy.
C. *(15 points)* Take all recordings in the test set and compute their spectrogram. Reconstruct the signal back from the spectrogram using the Griffin-Lim algorithm, as discussed in the class. Report the classification accuracy using these reconstructed signals in the test set. Does the classification accuracy change? Argue why.

D. *(10 points)* Can you design a linear time invariant (LTI) system (excluding trivial systems, e.g., a system that outputs zero irrespective of the input) such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy drops near to the chance level, i.e., 16.67%? If yes, provide a complete description and program for such an LTI system and report the classification accuracy. If no, argue why.

E. *(10 points)* Can you design an LTI system such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy does not change? If yes, provide a complete description and program for such an LTI system and report the classification accuracy. If no, argue why.

F. *(10 points)* Can you design a non-linear system (excluding trivial systems, e.g., a system that outputs a fixed value irrespective of the input) such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy drops near to the chance level, i.e., 16.67%? If yes, provide a complete description and program for such a non-linear system and report the classification accuracy. If no, argue why.

G. *(10 points)* Can you design a non-linear system such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy does not change ? If yes, provide a complete description and program for such a non-linear system and report the classification accuracy. If no, argue why.

H. *(10 points)* Can you design a time-varying system (excluding trivial systems, e.g., a system that outputs a fixed value irrespective of the input) such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy drops near to the chance level, i.e., 16.67%? If yes, provide a complete description and program for such a non-linear system and report the classification accuracy. If no, argue why.

I. *(10 points)* Can you design a time-varying system such that when all train and test recordings are passed through this system and the classification experiment is repeated, the classification accuracy does not change ? If yes, provide a complete description and program for such a non-linear system and report the classification accuracy. If no, argue why.

Few references/links:

MFCC implementations

http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/mdoc/v_mfiles/v_melcepst.html
https://librosa.org/doc/main/generated/librosa.feature.mfcc.html

kNN classifier

https://www.mathworks.com/help/stats/classificationknn.html;jsessionid=c3d882cfb40baf5102de384a5ea3
https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html