DSP-1 Assignment 1

Speech Processing Problems

This assignment is per student

Your answers should be in a pdf format + speech data for problem 6 (before and after the VOCODER). Your code should be in a separate file.

1. Analyze the windows given below:
   a. The rectangular window is defined as:
   $$W_{Rec} = 1 \qquad 0 \leq n \geq N\text{-}1$$
   $$= 0 \qquad \text{otherwise}$$
   b. The Hanning window is defined as:
   $$W_{Han} = 0.5 - 0.5\cos(2\pi n/N)$$
   c. The Hamming Window is defined as:
   $$W_{Ham} = 0.54 - 0.46\cos(2\pi n/N)$$
   Sketch W_{Rec}, W_{Han}, W_{Ham} in the frequency domain on the same graph (use log magnitude scale), and give your comments.

2. The following autocorrelation data are given:
   R(0) = 1, R(1) = 0.7, and R(2) = 0.4.
   Calculate the linear predictive coefficients (a1 and a2) and the mean squared error by solving the matrix equation.

3. A speech waveform S has the values:
   s0, s1, s2, s3, s4, s5, s6, s7, s8 = [1,4,0,-4,-1,2,4,-1,2,5]
   The frame size is 6 with no pre-emphasized,
   a. Find auto-correlation parameters r0, r1, and r2 for the first frame. Then find the parameters of LPC of order 2 (i.e. calculate a1 and a2).
   b. If the number of overlapping samples for two frames is 2, find the LPC coefficients of the second frame.
   c. Repeat the above if the pre-emphasis constant is 0.96

4. An all-pole system is modeled by an 8th-order model. $$H(z) = \frac{G}{1 + \sum_{k=1}^{8} a_k z^{-k}}$$

The system has the following poles (four complex-conjugate pairs) at a sampling frequency Fs = 16000 Hz:

$p_{1,2} = 0.965500 \pm 0.192050j$ $\quad p_{3,4} = 0.812108 \pm 0.542633j$

$p_{5,6} = 0.534176 \pm 0.799451j$ $\quad p_{7,8} = 0.183930 \pm 0.924681j$

   a. Use MATLAB to plot the magnitude of the system spectrum in dB assuming a sampling frequency of 16000 Hz.
   b. Estimate the formants and their BWs.

5. Suppose a vocal tract transfer function has the following 8th-order denominator polynomial A(z), which corresponds to the same poles as in Problem 4:

$A(z) = 1 - 4.99142683z^{-1} + 12.37177836z^{-2} - 19.81615903z^{-3} + 22.40030463z^{-4} - 18.31127304z^{-5} + 10.60283765z^{-6} - 3.99936958z^{-7} + 0.75965617z^{-8}$

   a. Use MATLAB to plot the magnitude of the LPC spectrum in dB
   b. Estimate the formants and their BWs.
   c. verify that the poles obtained from this A(z) match those given in Problem 4.

6. Building a Speech Vocoder (Speech Analysis and Synthesis)

Objective:

To design and implement a basic **speech vocoder** that analyzes and reconstructs a recorded Arabic utterance using standard signal processing techniques.

You are free to select your parameters (like filter order, window size, ..etc.). You also can test some selections, and you can change these options to get better results. In this case you are advised to mention in your report these actions.

a. Speech Recording

Record your own voice saying the following Arabic sentence:

"دع الأيام تفعل ما تشاء ... وطب نفسًا إذا حكم القضاء"

Make sure the recording is clear, noise-free, and sampled at an appropriate rate (e.g., 16 kHz, mono).

b. Speech Analysis

Perform a detailed analysis of the recorded speech, including the following steps:

Pre-processing:

Divide the signal into overlapping frames (*windowing*) using a Hamming window.
Apply *pre-emphasis filtering* to enhance high frequencies.

Feature Extraction:

Perform **Linear Predictive Coding (LPC)** analysis for each frame to obtain LPC coefficients.
Estimate **pitch (fundamental frequency)** for voiced frames.
Compute **gain (energy)** for each frame.

Voiced/Unvoiced Classification:

Determine whether each frame is voiced or unvoiced.
You may use your own algorithm or an existing advanced tool for comparison.

c. Speech Synthesis (Reconstruction)

Using the extracted parameters (LPC coefficients, pitch, voiced/unvoiced decisions, gain), reconstruct the speech waveform through a **production model** of the vocoder.

d. Evaluation

Evaluate the perceptual quality of your synthesized speech using the **Mean Opinion Score (MOS)** method: https://en.wikipedia.org/wiki/Mean_opinion_score

Based on the MOS value, assign a qualitative mark to your vocoder's performance.

e. Improvements

check with the available advanced techniques on the web and substitute one or more of them with two conditions:

1. To get real improvements
2. To understand the basic concept behind this (these) improvement(s). you must explain this concept(s) in PPT slides.

You must regenerate the advanced version of the voice.

7. Speech Recognition System for Spoken Digits

Task:

Design and implement a complete **speech recognition system** for recognizing spoken digits (0–9).

Instructions:

1. **Feature Extraction:**
    - Compute the **Mel-Frequency Cepstral Coefficients (MFCCs)** for all the provided speech recordings of the 10 digits.
2. **Model Construction:**
    - Use **Dynamic Time Warping (DTW)** with the MFCC features to build your recognition system.
    - The system should be able to recognize a new spoken input corresponding to one of the ten digits.
3. **Evaluation:**
    - Test your system using **separate test data** (not used during training).
    - Construct and present a **confusion matrix** to evaluate recognition performance. (For reference, see: Confusion matrix – Wikipedia)