Speech Signal Processing (EC5.408)

Assignment 1

January 9th, 2025 - January 23rd, 2025

Guidelines

- Do not copy or plagiarise. If you're caught for plagiarism, the penalty will range from zero in the assignment to an F grade in the course.
- Always cite your sources (be it images, papers, or existing libraries).
- Mention clearly if any assumptions are being considered.
- Only MATLAB or Python can be used for the coding part.
- For this assignment, you might use Audacity or Wavesurfer software.

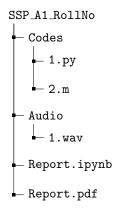
Submission Format

Make a directory using the naming format SSP_A1_RollNo. The submission might include:

- Codes (.py/.m) to answer the coding problems.
- Reports (.pdf) to answer the theory questions.
- Notebooks (.ipynb) to answer both coding and theory questions together.

Place the files in their respective folders and zip the main directory using the naming format SSP_A1_RollNo.zip and upload this zip file to Moodle.

Example Directory Structure



Questions

I. Define the following:

- 1. Epoch
- 2. Pitch
- 3. Formant
- 4. Sampling
- 5. Quantization

II. Explain the following:

- 1. Differentiate between Fourier series and Fourier transform. What is their significance respectively?
- 2. What are the assumptions we take while analyzing a speech signal? How do we handle the non-stationarity of the speech signal?
- 3. "Female pitch is more when compared to male pitch." True or False. Justify the statement with a proper explanation.

III. Record a sentence as: "My name is your name"

- 1. Plot the waveform.
- 2. Identify and mark the voiced, unvoiced, silence, and plosive regions (plot it over the waveform).
- 3. Provide an acoustic-phonetic description of the regions (MOA and POA).
- 4. Plot the Mel spectrogram and identify the formants.

Use Audacity or Wavesurfer for this question. You are expected to submit:

- A .wav file of the recording.
- Annotated transcription on top of the waveform.

For example, for the word kitAb (/k/, /i/, /t/, /A/, /b/):

• It is an unvoiced unaspirated velar stop followed by a front vowel followed by an unvoiced unaspirated dental stop followed by a middle vowel followed by a voiced unaspirated bilabial stop.

IV. Write a code for the following:

- 1. Use the above audio file and the corresponding phoneme segment boundaries
- 2. Compute short-time energy and short-time zero-crossing rate (ZCR) of the recorded audio with a 20ms window size and a 10ms window shift.
- 3. Compute the average of short-time energy and short-time ZCR values within each manually marked phoneme segment. Compare the averaged values across vowel phonemes and comment on the observed values.

