Speech Signal Processing
EC5.408

Assignment 2

Aug 19, 2023

Guidelines

- Do not copy or plagiarise. If you're caught for plagiarism, the penalty will range from **zero** in the assignment to **F** grade in the course.
- Cite your sources (be it images, papers or existing libraries) when necessary.
- Mention clearly if any assumptions are being considered.
- Only MATLAB or Python can be used for the coding part.
- Theory answers (in report) should be typed unless mentioned otherwise.

Submission Format

Make a directory using the naming format **SSP_A2_RollNo**. The submission might include codes **(.py/.m)** to answer the coding problems, reports **(.pdf)** to answer the theory questions or 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_A2_RollNo.zip** and upload this zip file to Moodle.

This is how the final directory structure might look like

```
SSP_A1_RollNo
├── codes
│   ├── code_1.py
│   ├── code_2.ipynb
│   └── code_3.m
├── reports
│   ├── report_1.pdf
│   └── report_3_4.pdf
└── wavs
    ├── audio_1.wav
    └── audio_3.wav
```

Questions

[Maximum marks: 30]

1. Explain briefly about the following **[4]**
   (a) Autocorrelation [1]
   (b) ZCR [1]
   (c) Mel spectrogram [1]
   (d) LP spectrum [1]
2. Explain **voiced** and **unvoiced** speech? Explain any three different methods used for identifying them. Write a function for each method that takes a signal and classifies it as voiced or unvoiced. **[1 + 2 + 2]**
3. Using the given audio file, write a code to analyse the following **[5]**
   (a) effect of **window length** in STFT [2]
   (b) effect of **window shape** in STFT [2]

   Comment on the plots for both the cases. [1]
   NOTE: You need to implement STFT using FFT from any library.
4. Using the given audio file, write a code to do the following **[16]**
   (a) Write a function that takes a **signal**, **window_length**, **hop_length** and calculates the frames accordingly [2]
   (b) Plot the time domain waveform and calculate the number of frames using the above function [1]
   (c) Plot the frames and label each frame as **voiced** or **unvoiced** using the functions you defined in **Q2**. [2]
   (d) Apply Fourier Transform [4]
   - on the entire signal
   - on the signal at the frame level.

   What are your observations in each case? Explain.
   (e) Take one **voiced** frame and one **unvoiced** frame and do the following [4]
   - Calculate the magnitude spectrum
   - Reconstruct the time domain signal back from the magnitude spectrum

   What are your observations in each case? Explain.
   (f) Calculate the **Pitch** contour. Do we need to do this on the frame level? If yes, are all frames considered? Explain. [3]

   NOTE: For each part, chose the parameters for calculating frames appropriately and mention the assumptions if any.