Deep Speech Processing (DSP) Assignment: 5

Short-Term Frequency Domain Processing of Speech

Feb 2025

Instructions

- When uploading to Google Classroom, compress your files into a ZIP archive. Name the ZIP file as `SRN_Name.zip`
- All students are required to submit their assignments before coming for evaluation
- From now on, there is no need to upload a separate report. Instead, include your observations directly within the IPython notebook. For each experiment, create a text cell to write your observations. In addition, embed audio files directly into the notebook.
- During the the evaluation, you will present the concepts using the IPython notebook exclusively.
- Any deviation from the guidelines cannot be considered during the evaluation.
- For all the questions in the Basics section, ensure that both the time-domain representation and the magnitude spectrum plots are included in your Ipython notebook.
- If any doubts, please mail to `kishorks@iitdh.ac.in`

Limitation of DTFT [Q1]

Concepts to read: Stationary v/s Non-stationary

- Record audio with fs=16 kHz, 16bit.
- plot time and frequency magnitude spectrum [linear and log].
- What information is lost when applying the Discrete-Time Fourier Transform (DTFT) to an entire signal

Spectrogram [Q2]

Concepts to read: Framing, Frame size/Hop size, STFT

- Extract 30 ms of voiced segment and plot its waveform and frequency magnitude spectrum (linear and log).
- Extract 30 ms of unvoiced segment and plot its waveform and frequency magnitude spectrum (linear and log).
- Extract 30 ms of silence and plot its waveform and frequency spectrum (linear and log).
- Plot the STFT spectrogram of the full audio.

Spectrogram with windowing [Q3]

Concepts to read: Rectangular, Hamming, Hanning window
Repeat the above [Q2] experiment using Hamming and Hanning windows, then record the observations.

Effect of window size [Q4]

Consider $x[n]$ as a **300 ms voiced segment**, and define $y[n] = x[n] \cdot w[n]$ for the following cases. For each case, plot the time-domain signal and the frequency magnitude spectrum (both linear and logarithmic):

- $w[n] = \begin{cases} 1 & \text{for } 0 \leq n < 3\,\text{ms}, \\ 0 & \text{for } 3\,\text{ms} \leq n < 300\,\text{ms}. \end{cases}$
- $w[n] = \begin{cases} 1 & \text{for } 0 \leq n < 30\,\text{ms}, \\ 0 & \text{for } 30\,\text{ms} \leq n < 300\,\text{ms}. \end{cases}$
- $w[n] = 1 \quad \text{for } 0 \leq n < 300\,\text{ms}.$

Utils

- lab1: `https://colab.research.google.com/drive/1nX20djsBuHpdy29TNpbDCXc_6TCzyMlo?usp=sharing`
- lab3: `https://colab.research.google.com/drive/1yDGsctDdYIyCzTv2hJPCsRkzFJB9aOO-?usp=sharing`
- For recording audio, use wavsurfer or Audacity