Total Marks: 5

Speech Understanding
Programming Assignment 1

Due date: 28 January 2024

Assignment Guidelines

1. Any kind of plagiarism is not accepted. We will strictly follow institute policies for plagiarism.
2. Recommended programming languages: Python + PyTorch.
3. You may use any external libraries or GitHub codes. However, the evaluation will test your knowledge of the algorithm and the choice of hyperparameters. Do cite the libraries/codes.
4. **A single zip file containing the report, codes and readme if required. The zip file should be named as Rollno_PA1.zip.**

Question 1. Introduction to Audio Applications

Goal: The aim is to make you familiar with different audio related Tasks.

Tasks: —- Record two sentences in both English and your native language in your own voice.

—- Utilize the pre-trained Massively Multilingual Speech (MMS) Language Identification (LID) models by following the instructions in the provided Colab notebook. Evaluate the model's accuracy in identifying languages by comparing predicted and ground-truth languages. Analyze the model's performance. [1 Mark]

—- Use the pre-trained Massively Multilingual Speech (MMS) Text-to-Speech (TTS) models, as shown in the provided Colab notebook, to generate speech from both English and native language sentences. [2 Marks]

—- Employ the pre-trained Massively Multilingual Speech (MMS) Automatic Speech Recognition (ASR) models by following the instructions in the provided Colab notebook. Report the transcription performance using Character Error Rate (CER) and Word Error Rate (WER) as metrics.

1. Calculate CER and WER for the transcription generated by ASR from recorded audios, comparing it with the ground truth transcription (text) for both English and your native language.
2. Calculate CER and WER for the transcription generated by ASR from the previously generated audios, comparing it with the ground truth transcription for both English and your native language.

Evaluate the model's performance in the Automatic Speech Recognition (ASR) task by comparing its effectiveness in English and native language texts, as well as between recorded and generated audios texts. [2 Marks]

Total: 5 Marks