Name : Muhid Qaiser

Roll No. : 22i-0472

Section : AI-B

Assignment : 5

Introduction:

In this assignment, I designed a Text-to-Speech (tts) program to output the text of the diary in the form of an audio. I experimented with multiple techniques and models to find the best combination of techniques and model for the best audio output.

Methodology:

Preprocessing:

I took 3 diaries from the diaries dataset. I split each diaries by line and stored them in a list so that I can feed my model one line at a time and later combine the outputs. Each line was preprocessed using the multiple preprocessing functions I created such lowercase, remove first index, remove non-alphanumeric symbols etc. This outputted a list of lists of strings where each string is a line from one diary. And there are a total of 3 diaries in my testing dataset.

Model Selection:

Various models where tested such as variants of whisper, Mars5 for real-time voice cloning, speecht5_finetuned_urdu_TTS, mms-tts-urd-script_arabic, Guy-Urdu-TTS … etc but the one that stood out the most with the best audio output for roman urdu text input was seamless-m4t-v2-large as it was clear with correct annunciation. It was easy to integrate using the huggingface transformer's library and required minimum set-up.

Customization:

Customization was made on Mars5 model using it's real-time voice cloning feature but it did not produce as good results as did seamless-m4t-v2-large. So for customization, I employed 2 strategies.

1. Noise Reduction
2. Slight Pauses between each narration of line in combined audio.

Data Feeding:

Each line was fed into the model to generate an audio, audio was stored in a list with addition of slight delay/pause to make it seem organic and natural. These audios were than combined in sequential order to create the combined audio of the lines of that diary. This process was repeated for 3 diaries.

Results:

One of the best TTS models of Roman Urdu was created using these methodologies and the outputs of the model are according to the requirements of the assignments.