Capstone Project Weekly Progress Report

Semester	Fall 2022
Course Code	AML 2404
Section	Section 2
Project Title	Generating Synthetic Voice
Group Name	Group A: Goal Diggers
Student names/Student IDs	1. Abhimanyu Sharma (C0859053) 2. April Alcantara (C0858528) 3. Harsharan Singh Raina (C0857931) 4. Krati Rastogi (C0858753)
Reporting Week	Week 3
Faculty Supervisor	William Pourmajidi

1. **Tasks Outlined in Previous Weekly Progress Report** (Provide detailed information on the tasks to be completed in this week)

These are the tasks outlined to be completed this week.

- Studying and Exploring the LibriSpeech datasets.
- Beginning of the designing phase.
- Setting up the virtual environment for our project using Conda/Cmd.
- Getting insights from the dataset and getting our dataset ready for model training.

All the task outlined in the previous weekly report for this week has been accomplished within the decided time frame. Every team member contributed and did their part as assigned to them.

- Started with Studying and exploring the various branches of LibriSpeech datasets available online. We thought of using the pre-existing datasets would be more feasible given the time we have right now and making our own dataset can be achieved once we have a complete understanding of e dataset we are working on, so we put a pause on building our own dataset.
- Set up the environment and installed various libraries. Faced few problems which are discussed in the further report.

- Downloaded various datasets on Kaggle notebook for test-check. Got to know about the various required features for our project and after going through a few of the datasets we finalized Confmer-ctc-Librispeech dataset.

2. **Progress Made in Reporting Week** (Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)

This week's Task:

Abhimanyu and April suggested Libri-Speech-Data which was 2.1GB, Krati suggested Conformer-ctc-librispeech which was around 39GB and Harsharan recommended English multi-speaker corpus for voice cloning which was 11GB of data. All the datasets were from Kaggle and after looking at the datasets suggested by the team members, we decided to go with the Conformer-ctc-Librispeech dataset.

- It was the only dataset which fitted our requirements.
- Limited and necessary features with different voice data of single speaker
- It can help us improving the performance of our model

Explored the dataset. The following information is listed below: (**Harsharan and Krati**)

- The size and shape is 45537 rows and 11 features.
- There are no null values
- Head function was used to display the data
- Describe function displayed the information about 'speaker_id', 'chapter_id' and 'file_id' which turned out to be not so helpful individually, but we can use these features in combination with other features to identify a single speaker as a primary key. The the 'minute' column it was helpful as it describes how long an individual speaker's voice data we have.
- Different features of the dataset are:
    - Audiopath
    - Filename
    - Subset
    - Speaker_id
    - Chapter_id
    - File_id
    - Id
    - Sex
    - Minute
    - Speaker name
    - Sentence

April tried setting up the virtual environment for the project where she upgraded the pip version to install the pytorch framework (Conformer-ctc-librispeech, 2022).

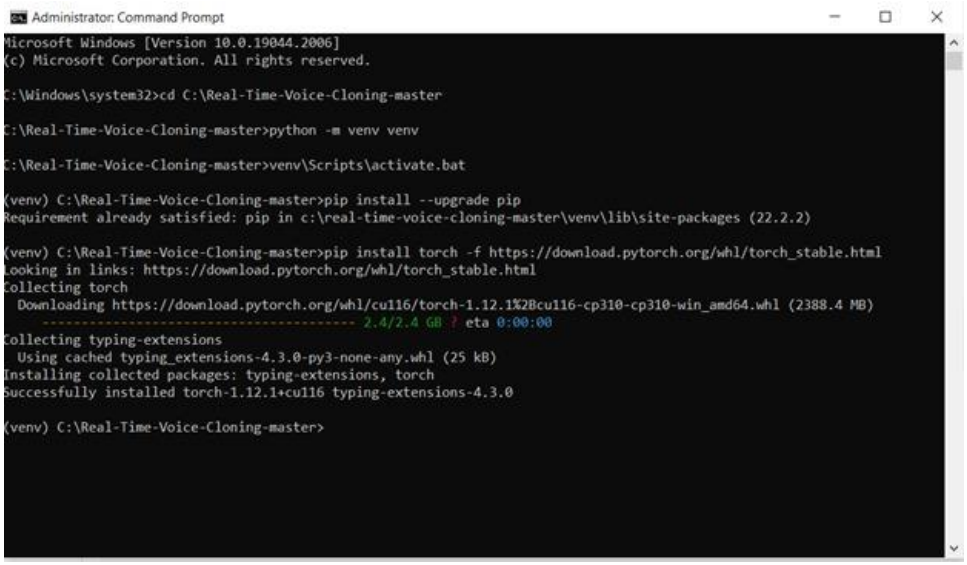

Figure 1: Installation of the necessary library

3. **Difficulties Encountered in Reporting Week** (Provide detailed information on the difficulties and issues that you encountered in the reporting week. Limit your write-up to no more than one page)

   The difficulties we faced this week are given below:

   - Imported the dataset, due to the large size of the data (~40GB) it took a fair amount of time. So we decided to use Google colab instead of jupyter notebook. (**Abhimanyu and Krati**)
   - Faced difficulties in setting up the virtual environment for the project. The requirements file was not accessible and there was admin rights issue as well. (**April**)

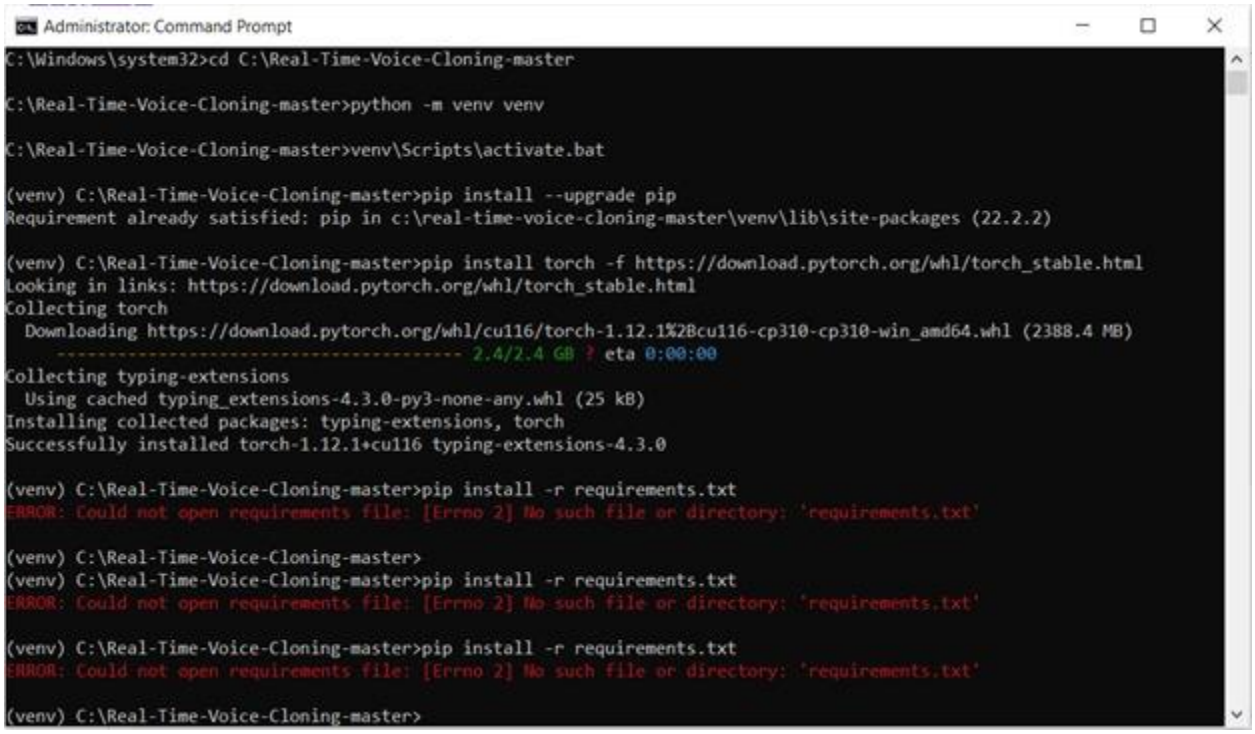

Figure 2: Error while setting up the virtual environment

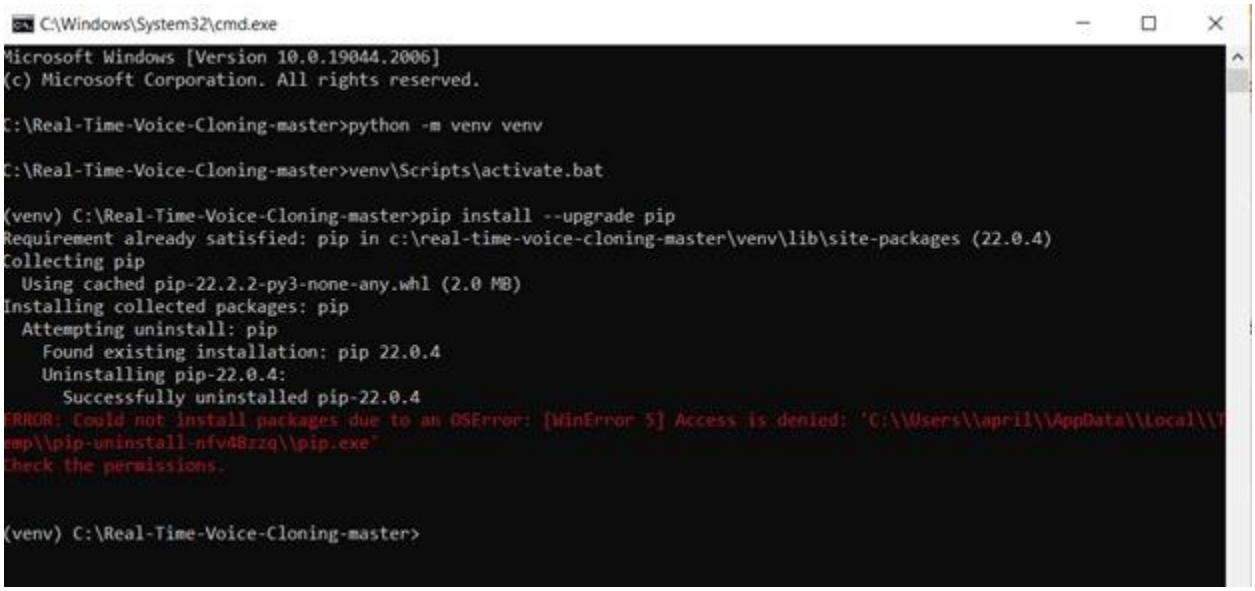

Figure 3: Admin right issues

4. **Tasks to Be Completed in Next Week** (Outline the tasks to be completed in the following week)

- Krati and April will be analysing the data further and in more broader perspectives.
- Harsh and Abhimanyu will be exploring and learning about the models (SV2TTS, WaveRNN vocoder, Tacotron synthesizer, GE2E encoder) with respect to our project from github.

References:

conformer-ctc-librispeech. (2022, March 9). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/code/tuannguyenvananh/conformer-ctc-librispeech/data

Starter: LibriSpeech 8da5c067-f. (2019, June 13). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/code/kerneler/starter-librispeech-8da5c067-f/data

English Multi-speaker Corpus for Voice Cloning. (2019, December 17). Kaggle. Retrieved October 2, 2022, from https://www.kaggle.com/datasets/mfekadu/english-multispeaker-corpus-for-voice-cloning