Capstone Project Weekly Progress Report

Semester	Fall 2022
Course Code	AML 2404
Section	Section 2
Project Title	Generating Synthetic Voice
Group Name	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 2**
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.

- Search for the various dataset available online and try to learn what type of features it includes.
- Try to use these features for creating our own dataset.

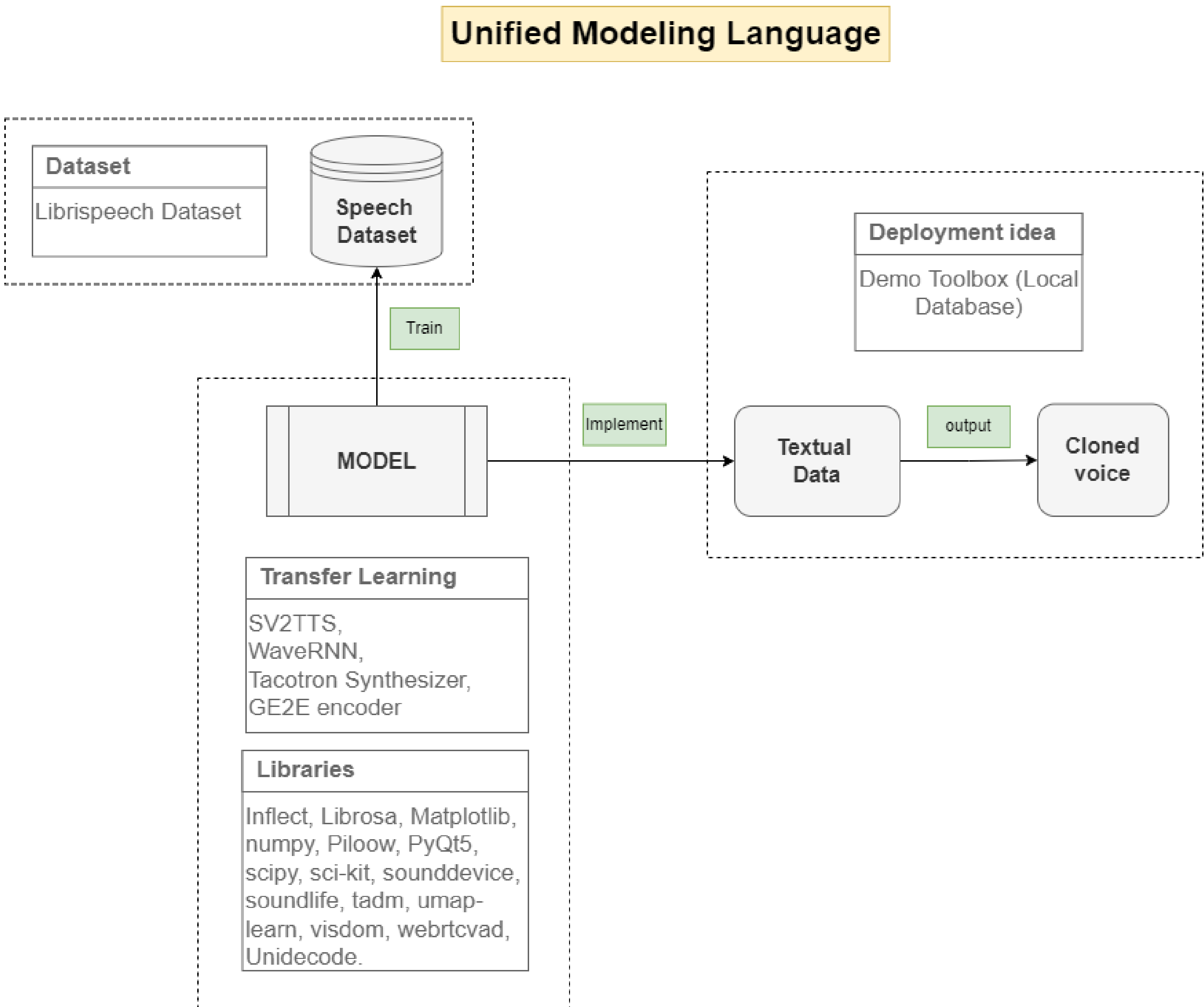
2. **Progress Made in Reporting Week** (Provide detailed information on the progress that you made in the reporting week. Limit your write-up to no more than two page)

This week's Task:

A brief description of our project:

Cloning or Imitating human behaviour has always been challenging in the field of AI. Our project is to explore to what extent we can achieve the near-perfect voice cloning of the 'human voice'. Our project "Generating Synthetic Voice" describes a neural network-based system for text-to-speech(TTS) synthesis that is able to generate speech audio in the voice of different speakers, including those unseen during training. Dataset will include noisy speech without transcripts from thousands of speakers. Our model will consist of three independent neural networks: a speaker encoder, a synthesizer and a vocoder. The goal of this work is to build a real-time TTS system which can generate natural speech for a variety of different speakers in a data-efficient manner.

UML diagram – It consists of a detailed flowchart of pipeline of our project.



Tech stack:

Version of python we are going to use?

Python 3.7 is recommended; version 3.5 < should work as well

Which model are we going to implement?

Transfer learning models from github (SV2TTS, WaveRNN vocoder, Tacotron synthesizer, GE2E encoder):

- SV2TTS – Trasnfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis

- WaveRNN Vocoder - For Efficient Neural Audio Synthesis
- Tacotron synthesizer – Towards End-To-End Speech Synthesis
- GE2E Encoder – Generalized End-To-End Loss for Speaker Verification

- Ffmpeg: to read audio files
- PyTorch: to speed up the process between research prototyping and deployment; simplifies creation of ANN

Information flow

From where are we going to use the data?

LibriSpeech dataset

LibriSpeech (LibriSpeech is a corpus of approximately 1000 hours of 16kHz read English speech, prepared by Vassil Panayotov with the assistance of Daniel Povey. The data is derived from read audiobooks from the LibriVox project, and has been carefully segmented and aligned.)

Components / Libraries you will use

- Inflect
- librosa
- matplotlib
- numpy
- Pillow
- PyQt5
- scikit-learn
- scipy
- sounddevice
- SoundFile
- tqdm
- umap-learn
- Unidecode
- urllib3
- visdom
- webrtcvad

Deployment ideas

A Demo toolbox (local db)

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:

- We faced difficulty on finalizing the dataset which we are going to use to train our model. As custom dataset training may require powerful resources and a lot of time, we decided to firstly use the already available dataset of voice recordings available on LibriSpeech. Further, we will later on add our customized voice recordings to be included in the dataset to fine tune our model.
- Secondly, we faced roadblocks on finalizing our deployment idea. We are planning to firstly deploy our model on a local database, then we will later on implement its functionality over cloud.
- Thirdly, the python libraries which we are going to use in our project are still not finalized completely. We have written all the libraries names with a rough idea and we will figure out along the way which additional libraries we are going to use.

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

- Studying and Exploring the LibriSpeech datasets.
- Beginning of the design 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.