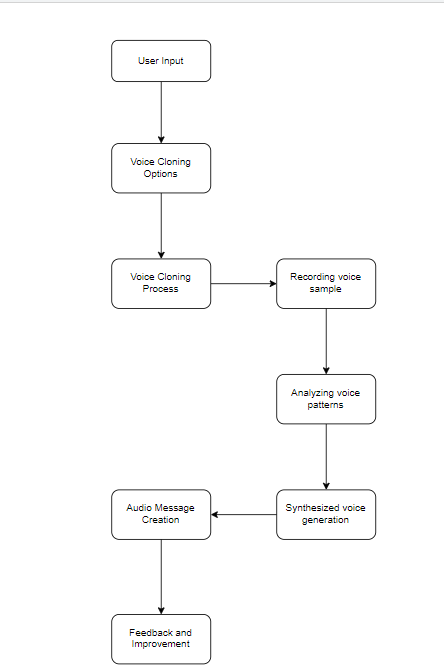
**TITLE**: VoiceChat - Bringing chats to life using Deep Learning

**Team Number**: CSM06

**Team Members**: K. Laxman Vikas, A. Pavan Kalyan, Yogeshwar Rayudu

**Github Link**: https://github.com/laxmanvikas463/VoiceChat-Bringing\_chats\_to\_life\_using\_DeepLearning

**Project Overflow:**



**Aim:-**

This project aims to build a mobile application that seamlessly integrates both app development and deep learning techniques to help bring chats to life. This app can read out text messages in respective sender’s voice using voice cloning technology.

**Objectives**:-

This project unfolds into 3 stages.

Stage 1:

Building a deep learning model that can provide:

- Accurate audio output that closely resembles sender’s voice.

- High inference speed

- A model that can convert Text-to-speech(TTS).

- supports custom voice conversion

- Audio that is emotionally rich and provides context and meaning to the message sent.

Stage 2:

To build a mobile application that is:

- Easy to use

- consists all the features of a traditional chat application

- provides a fresh-and-new feel with aesthetic design

- is able to integrate the above mentioned deep learning model within it.

Stage 3:

To integrate both deep learning model and mobile application that supports:

- Front-end - Back-end intercommunication

- Database of users and their voice models

- Privacy and security features

**Work Progress:-**

1. **Work Completed**

We have explored multiple deep learning models which have their pros and cons:

Traditional TTS model(TorToiSe):

Youtube refernce:

[https://www.youtube.com/watch?v=7SXuFVCIqS4&t=29s](https://www.youtube.com/watch?v=7SXuFVCIqS4&t=29s" \t "https://mail.google.com/mail/u/0/" \l "search/proj/_blank)

Colab link:

[https://colab.research.google.com/drive/1N8lXviiwfmrS9vYzuFdQnLHlddXMMMvU#scrollTo=4N9LSJ24IqOp](https://colab.research.google.com/drive/1N8lXviiwfmrS9vYzuFdQnLHlddXMMMvU" \l "scrollTo=4N9LSJ24IqOp" \t "https://mail.google.com/mail/u/0/" \l "search/proj/_blank)

Pros:

- Has TTS feature

- Supports custom voice conversion

Cons:

- Very slow

- Less accurate

RVC v2 crepe:

Explanation document:

<https://docs.google.com/document/d/13_l1bd1Osgz7qlAZn-zhklCbHpVRk6bYOuAuB78qmsE/edit>

Colab link:

<https://colab.research.google.com/drive/1Gj6UTf2gicndUW_tVheVhTXIIYpFTYc7?usp=sharing#scrollTo=7vh6vphDwO0b>

Pros: Very accurate

Cons:

- No TTS

- Cannot upload custom dataset

- Only uses default/limited voices

TTS on Windows(1 min guide):

<https://www.youtube.com/shorts/b0H4CFlzGz8>

<https://www.youtube.com/watch?v=axkDQp6IZbY&t=1s>

Applio RVC fork which uses Edge-TTS or Bark-TTS:

<https://www.youtube.com/watch?v=jh83ViDUvJs>

Pros:

- Fast

- Accurate

- Provides emotions such as laughing and crying

Cons:

No custom voice

Faster TTS generation - Deepspeed installation:

<https://www.youtube.com/watch?v=RVzpjYOV-Tk>

Run TTS on your local computer:

<https://www.youtube.com/watch?v=mq9lBd8XMY4>

Tacotron2 model:

<https://www.youtube.com/watch?v=e71H--vxRvo>

<https://colab.research.google.com/github/justinjohn0306/FakeYou-Tacotron2-Notebook/blob/dev/FakeYou_Tacotron_2_Training.ipynb>

1. **Need to-be Completed**

- Need support in building the mobile application

- Need support in integrating the deep learning model with the mobile application.