Voice Cloning

The voice cloning has been implemented mainly using pre-existing models as most publicly available data had quality issues, the pretrained models we experimented with were already trained on publicly available data and training takes too long (the training process is complex and involves training 3 neural networks the encoder, vocoder, synthesizer). For this multi-GPU hardware is required as we get out of memory errors most of the time while attempting to train. Consider experimenting with different models for encoder, vocoder, sythesizer. Also, ensure to high quality audio recordings without pauses or stuttering or background noise, so that the quality of voice generated is high. This field is still an active research area, hence we couldn't improve much over the state of the art results. Instead we just found the combination that works the best which can be found in the voice_cloning subdirectory of the cloud folder.

Most of the relevant documentation and other research insights can be found in the cloud folder especially in comments of the ipynb notebooks and py files.