NEURAL VOICE CLONING

Timeframe

Approx time of review	Expected Status
Phase One (Sept end)	Learning Phase to understand the concept of sequential neural network, the basics of voice cloning and to look into the various methods to convert text to speech.
Phase Two (Nov mid)	Learning Phase to have a proper understanding of the various methods of voice cloning that can be implemented and to choose the method that seems to be the most feasible one and gives the best results.
Phase Three (Jan end)	We plan to complete the following two objectives in the following time frame: 1)Train a model that can learn a voice and replicate it from a minute or so of voice sample 2)Use the voice that is generated by the model to read the text that has been supplied by the user.
Phase Four (Mar end)	Design the app that is required and to finalise how the transfer of data between the model and app will take place.

Monitoring and Evaluation

Phase 1 will be marked by: i) KSS and hackathons that are conducted in order to deepen the understanding of the basics of neural nets and sequential neural nets.
ii) A few sessions will be held to understand the underlying concepts in the papers that we are using as reference for the project.

Phase 2 will be marked by: A method must zeroed in on after the study of the feasibility and limitations of the various methods hat we have considered. This can done after trying out the various approaches that we have seen through our literature survey and thus choose the one that delivers the most conclusive results.

Phase 3 will be marked by: We will have to train the model by this phase and it must be capable of replicating the voice that we want to clone from the voice sample that we have recorded for a minute or so.

Phase 4 will be marked by: A method to use the voice that we have cloned to convert the text that we have supplied to speech must be created by this phase. There are a few few github repositories on the text to speech converts and taking them as reference a method will be deployed to achieve the same.

By the final evaluation the model will be integrated into an app that will act as the interface for the model to take input from the user (which here will be the voice to be cloned and the text to be read) and give the desired output which is the text read out in the cloned voice.

Deadline Date	Topic	Reading/ Video
7th Oct	Papers to Read	• Neural Voice Cloning with a few Samples • Attention Is All You Need • Online samples
14th Oct	Data, Processing, Corpora	Blog Posts - Speech Processing • Speech Synthesis Techniques • Interspeech Report Critical Reading of Scientific Literature • Harvard • Rice • Waterloo • Regina
21st Oct	Audio Synthesis (1)	Tacotron • Tacotron: Towards End-to-End Speech Synthesis *

		• Learning Filter Banks using Deep Learning for Acoustic Signals • Highway Networks • Learning Phrase Representations using RNN Encoder–Decoder for SMT Deep Voice • Deep Voice: Real-time Neural Text-to-Speech
28th Oct	Audio Synthesis (2)	Tacotron 2 • Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions * Deep Voice (cont) • Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning * Char2Wav • Char2Wav: End-To-End Speech Synthesis *
4th Nov	Audio Vocoder (1)	Tacotron 2 • Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions *

		Deep Voice (cont) ● Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning * Char2Wav ● Char2Wav: End-To-End Speech Synthesis *
11th Nov	Audio Vocoder (2)	Additional ● Sound Samples ● A Wavenet for Speech Denoising ● Fast Wavenet Generation Algorithm
18th Nov	Speech Embeddings (1)	Speaker Embeddings ● Deep Voice 2: Multi-Speaker Neural Text-to-Speech * ● Deep Speaker: an End-to-End Neural Speaker Embedding System * Projects ● Text-independent voice vectors ● Deep Speaker: an End-to-End Neural Speaker Embedding System