**Improving Intelligibility of Dysarthric Speech**

**ET60029**

**Term Project Report**

Under the guidance/supervision of

**Professor Manjira Sinha**

**Team** **Transformers**

**Group Members**

| Gunda Venkata Shanmukha Sainath | 19EC10026 |
| --- | --- |
| B Sai Vamshi | 19CS10016 |
| Jayanth Yindukuri | 19CS10068 |
| DG Pranathi | 19CS10025 |
| Aggi Sahithi | 19ME10004 |
| Dhanush Kumar Reddy | 19EE10024 |

[**Video Presentation**](https://youtu.be/qzmIHfjmNf0)

[**Demo Video**](https://youtu.be/xIa_t_Z2634)

**Contents**

1. Introduction………………………………………….03
2. Problem Statement…………………………………04
3. Approach…………………………………………….04
4. GAN Modelling………………………………………05
5. Deployment………………………………………….07
6. Tech Specs…………………………………………..09
7. References…………………………………………..10

**INTRODUCTION**

**Dysarthria:**

Dysarthria refers to a group of speech disorders resulting from disturbances in muscular control of speech mechanisms due to damage in the central or peripheral

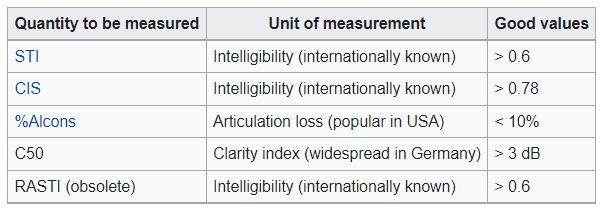
nervous system.

* Dysarthria occurs when the muscles you use for speech are weak or you have difficulty controlling them.
* Symptoms include slow voice, slurred voice, rapid speech, monotone speech, uneven speech volume, and strained voice
* Dysarthria does not affect intelligence or understanding, but a person with the condition may also have problems in these areas. Speech problems can also affect social interaction, employment, and education.
* Dysarthria is most common in neurological diseases such as
  + **Amyotrophic lateral sclerosis (ALS):** Up to 30% of people with ALS (Lou Gehrig’s disease) have dysarthria.
  + **Multiple sclerosis (MS):** Around 25% to 50% of people with MS get dysarthria at some point.
  + **Parkinson’s disease:** Dysarthria affects 70% to 100% of people with Parkinson’s disease.
  + **Stroke:** About 8% to 60% of people with stroke have dysarthria.
  + **Traumatic brain injury:** Some 10% to 65% of people with traumatic brain injury have dysarthria.
* Because of the communication problems dysarthria causes, complications can include:
* **Social difficulty.** Communication problems may affect your relationships with family and friends and make social situations challenging.
* **Depression.** In some people, dysarthria may lead to social isolation and depression.

**Intelligibility:**

Intelligibility is a measure of how comprehensible speech is in given conditions.

* Intelligibility is affected by the level (loud but not too loud) and quality of the speech signal, the type, and the level of background noise
* **Intelligibility standards:**



* Intelligibility measures give an indication of the functional communicative performance of dysarthric speakers which can be easily communicated to the speaker.
* Intelligibility score ranges from 70-90% for a normal non-dysarthric voice whereas the score ranges from 10-20% for a dysarthric voice.

**PROBLEM STATEMENT**

* To improve intelligibility and comprehensibility of dysarthric speech.
* Using a Machine Learning approach to transfer characteristics of normal speech to dysarthric speech.

**APPROACH**

* Training a GAN-based DL model (with dysarthric and control voices) which when given a dysarthric voice, converts to a non-dysarthric voice during inference
* Deploying the model on web app for live usage (with upload and download options)

**GAN MODELLING**

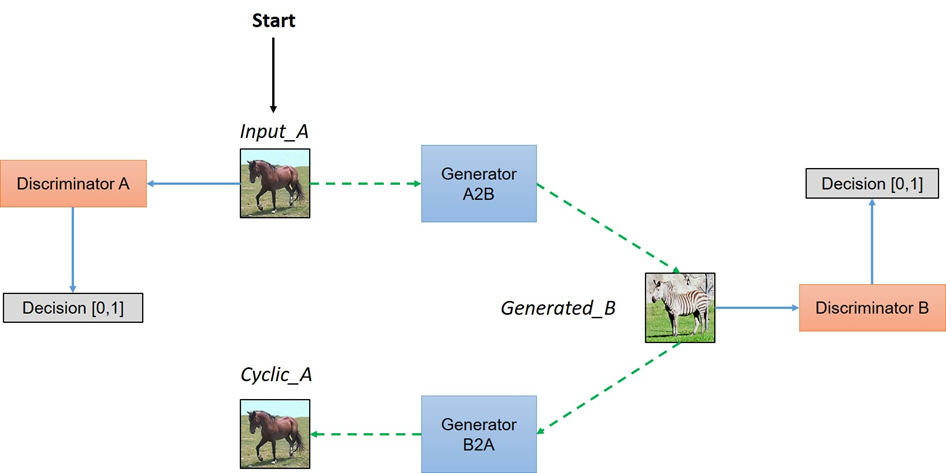
**Dataset:**

* There are many publicly available datasets of dysarthric speech.
  + [TORGO Database](http://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html)
  + [UASpeech Database](http://www.isle.illinois.edu/sst/data/UASpeech/)
* Dataset mainly contains the following
  + Audio samples of a person with a Dysarthric voice (Dysarthric speaker)
  + Audio samples of a person with a Normal voice (Normal speaker)
  + All digits, letters, common words, and Uncommon words are recorded
* We will be using TORGO Dataset for our project. The data structure is as follows
  + dysarthria\_female :Female Dysarthric speech files (500 wav files)
  + non\_dysarthria\_female : Female Control speech files (500 wav files)
  + dysarthria\_male : Male Dysarthric speech files (500 wav files)
  + non\_dysarthria\_male : Male Control speech files (500 wav files)

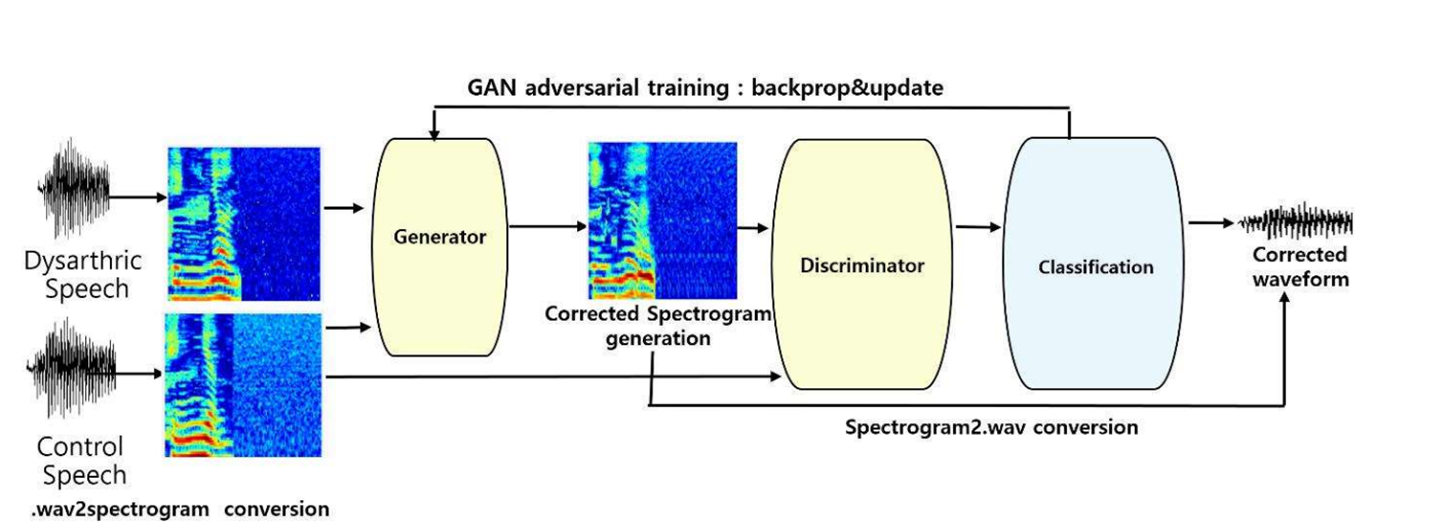
**Data Preprocessing:**

* Audio signal is converted into wideband spectrogram using Short-Time Fourier Transform (STFT) with windows of 512 frames and 33% overlap.
* Converted to wideband spectrograms ranging from 100-200Hz, which allows for high-quality output in the time resolution.
* Spectrogram is then converted to dB amplitude scale and saved as a grayscale image.
* Finally the obtained gray-scale image is converted to RGB image and resized to a 256x256x3 image

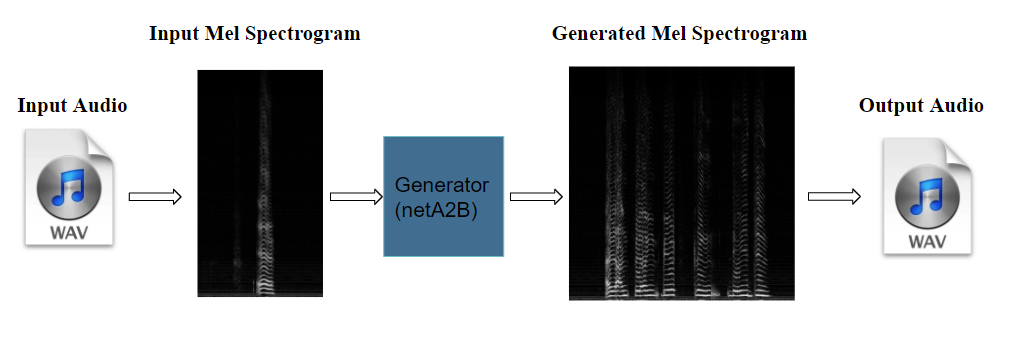
**Cycle GAN:**

****

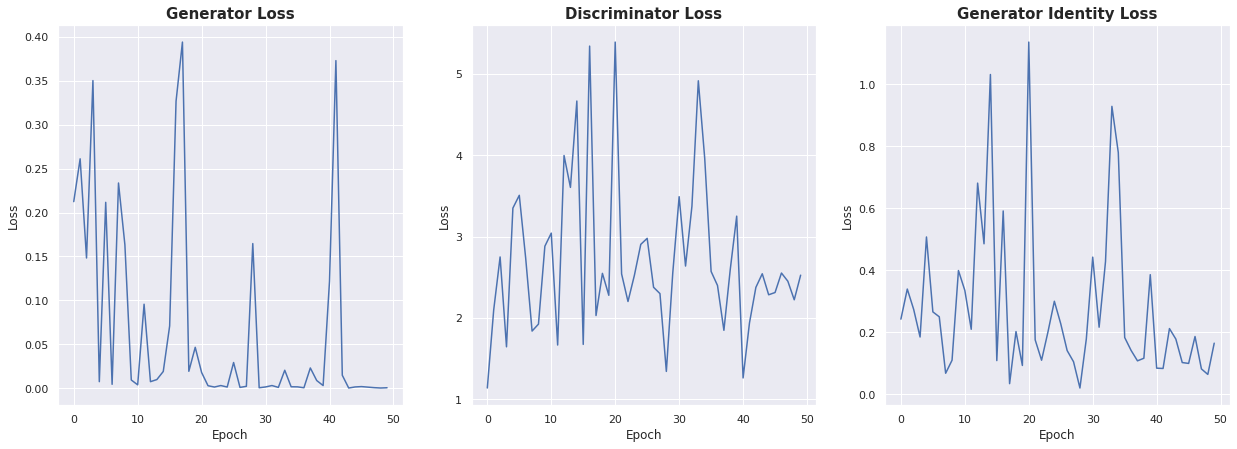
**Training:**

****

**Inference:**

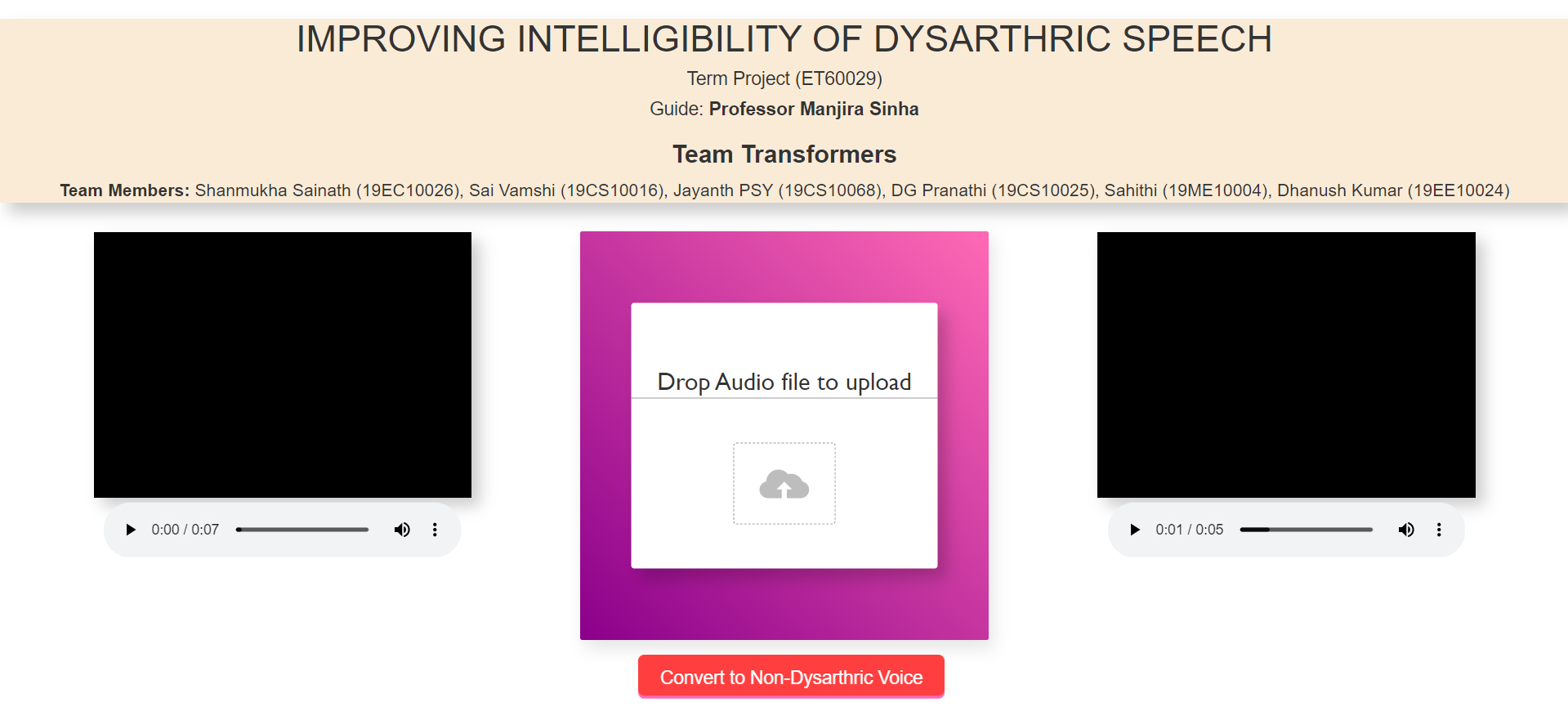
****

**Results:**

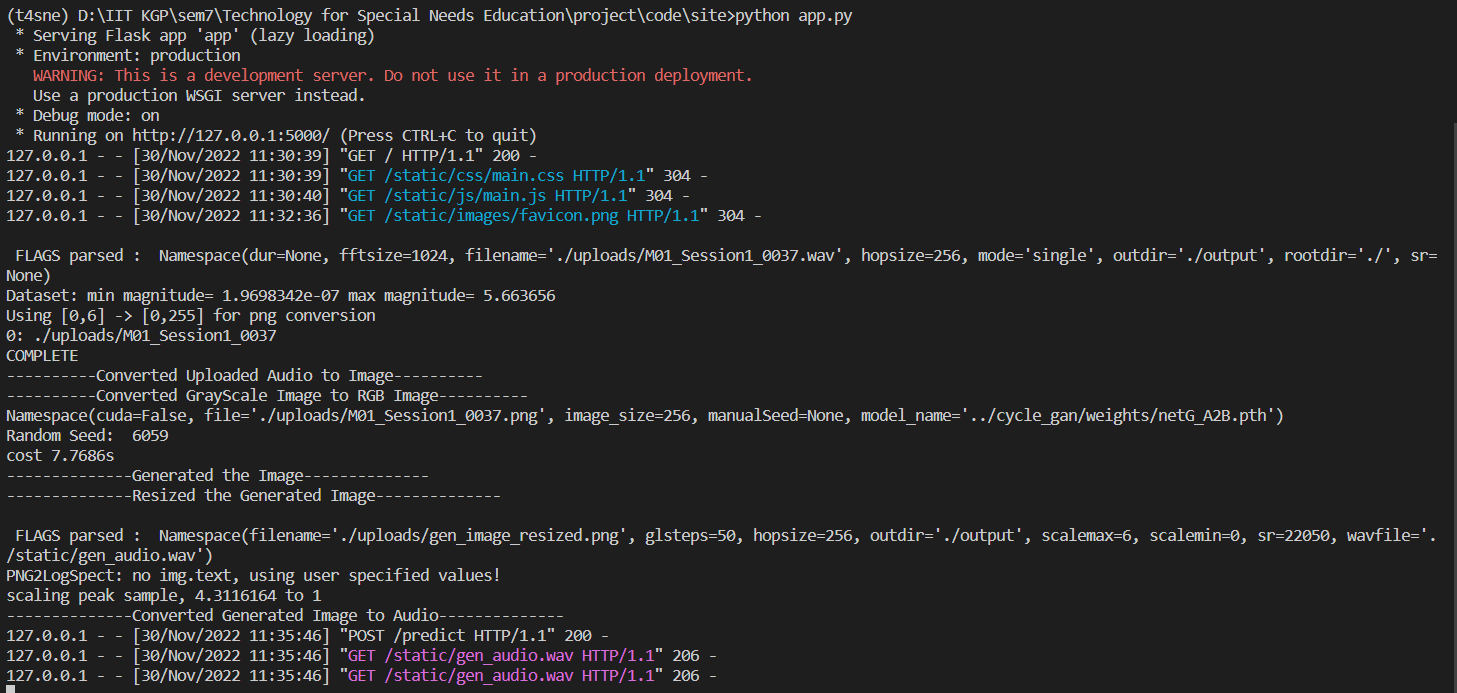


**DEPLOYMENT**

* A webpage with HTML, CSS, and Javascript is created.
* User can upload wav file using the upload button
* On the left side of the page, one can listen to uploaded audio can along with waveform display
* Upon clicking “Generate non-dysarthric voice”, uploaded audio is preprocessed and inference is done using Generator model of CycleGAN with pre-trained weights.
* Finally the generated spectrogram is converted to a wav file and displayed on the right side of the webpage.
* **Webpage:**



* **Terminal Output:**



**TECH SPECS**

**GAN Modelling:**

* Kaggle GPU (Nvidia P100)
* Libraries
  + Librosa
  + Numba
  + Numpy
  + Scipy
  + Pillow
  + OpenCV
  + PyTorch
  + Torchvision
  + Scikit-Learn

**Deployment:**

* HTML
* CSS
* Java Script
* Flask

**REFERENCES**

* Rudzicz, F., Namasivayam, A.K., Wolff, T. (2012) [**The TORGO database of acoustic and articulatory speech from speakers with dysarthria**](http://www.springer.com/alert/urltracking.do?id=Le4b3ccMb22f84Sade33dd). Language Resources and Evaluation, **46**(4), pages 523--541.
* Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.
* Yang, Seung Hee, and Minhwa Chung. "Improving dysarthric speech intelligibility using cycle-consistent adversarial training." arXiv preprint arXiv:2001.04260 (2020).