



Epoch
IIIT Delhi

Project Report

Project **Kritrim Vani**

By Team Non-IIITians

AKSHAY PRATAP SINGH

ARPIT GUPTA

PRASOON PANDEY

HARSH KUMAR

Problem Statement

The goal is to train a low-footprint GAN vocoder for generating multilingual audios from discrete unit/token representations from a generative audio model

Our Approach

01.

Speech Identification

Decode and extract features from audio and convert its sampling rate to match the model

02.

Text Tokenization

Converting the text To numpy arrays and audio spectrograms to Feed it to the OPENAi-Whisper Model to Process it

03.

Language Translation

Utilizing the Already Tried and Tested Google Translate python package to detect the audio language and further translate it to multiple regional Languages

04.

Reversing Back to Audio

Reversing the Translate Audio to Spectograms and further transform them to waveforms for easy access



Tech Stack Used

openai-whisper

googletrans

gTTS

Transformers

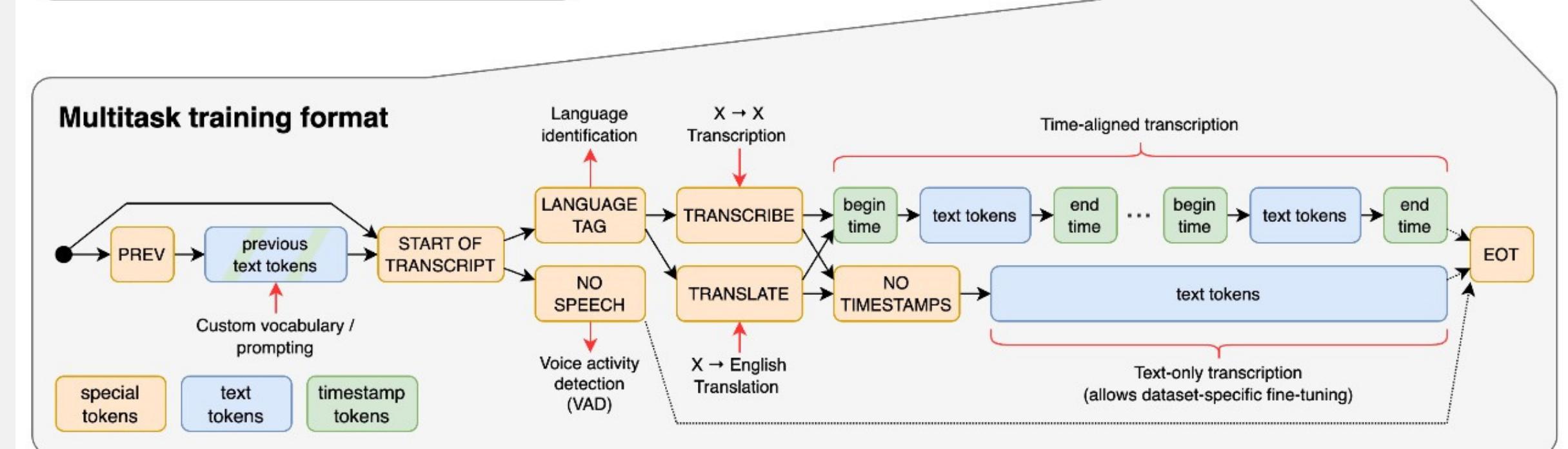
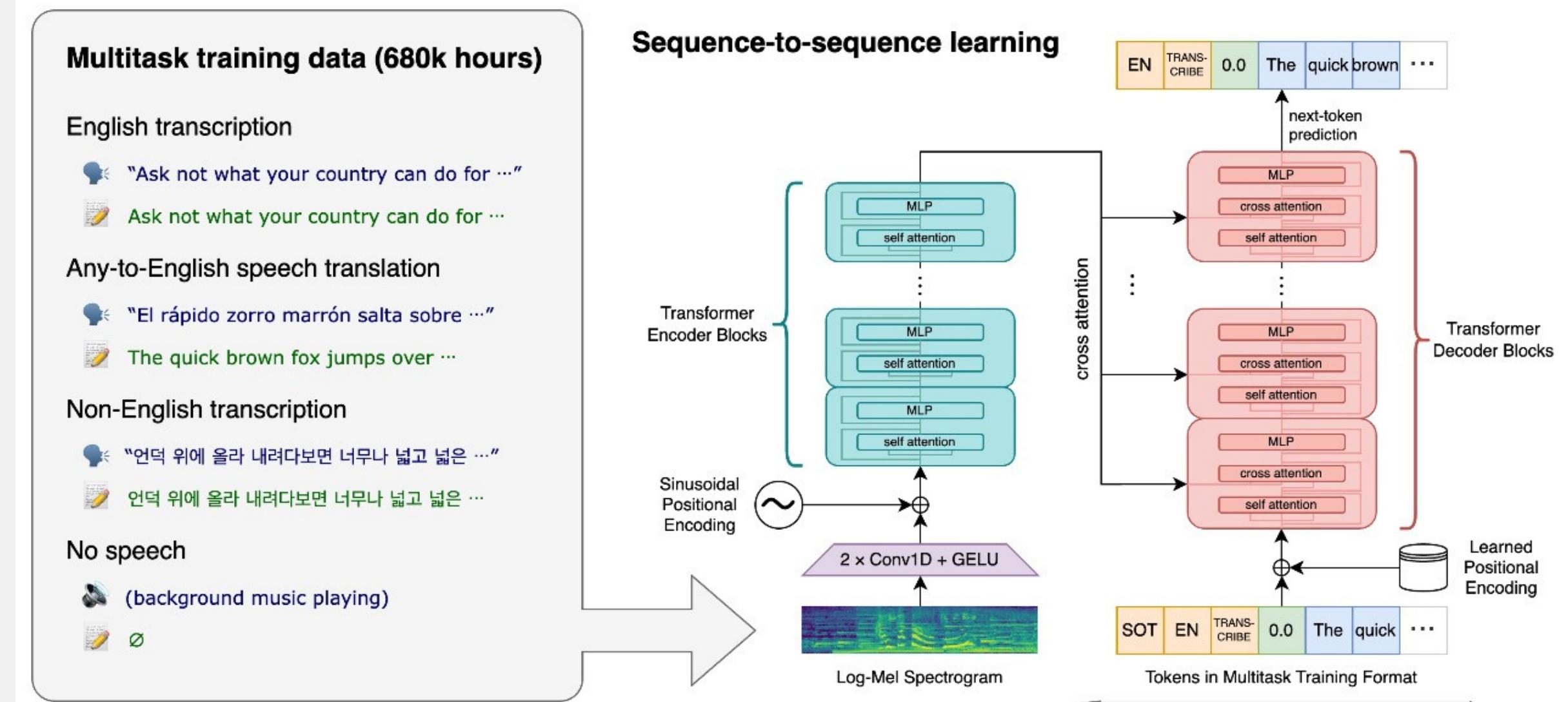
Tensorflow

Keras

PyTorch

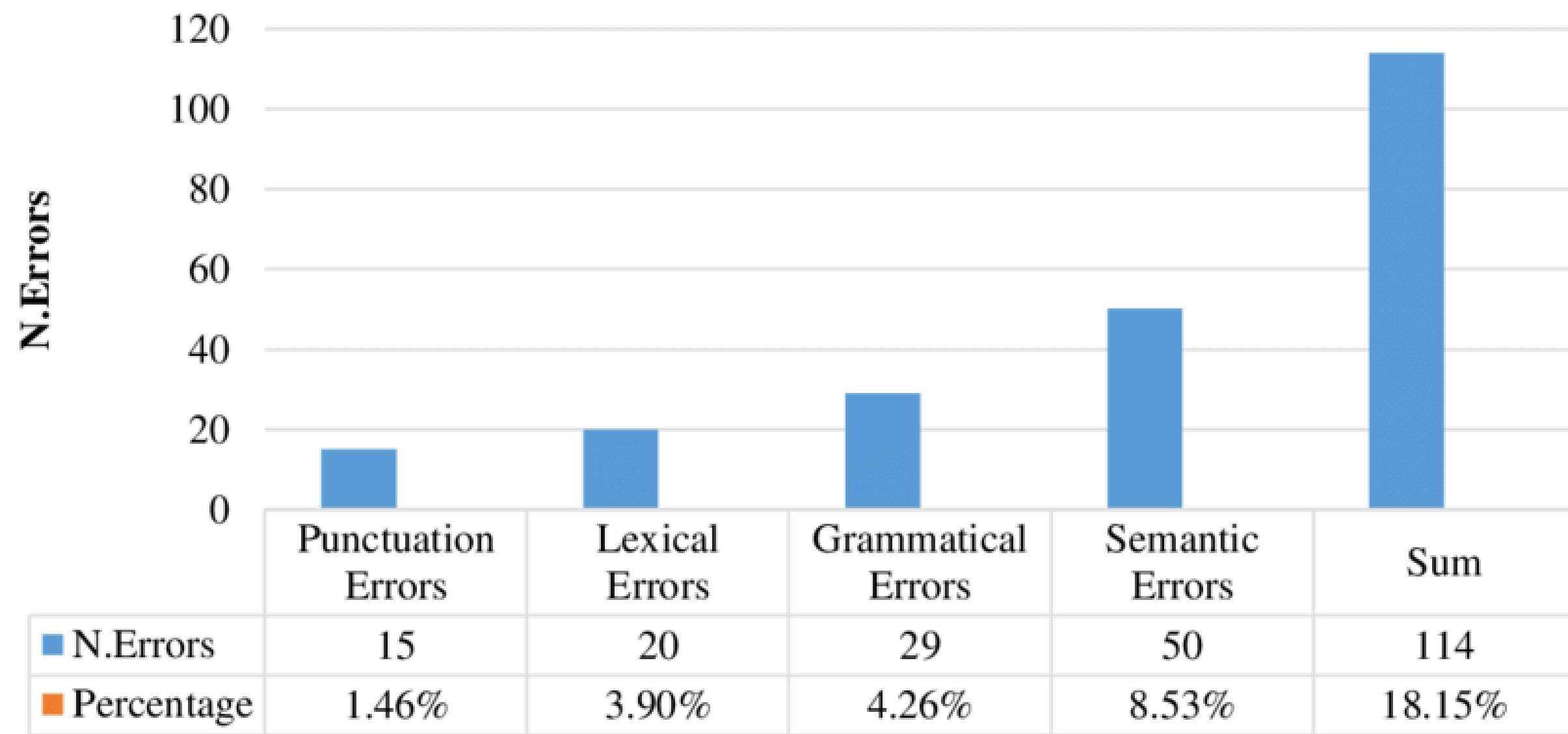
librosa

Model Architecture



90% + Accuracy/Precision

Google Translate



Challenges Faced

01 - AUDIO PROCESSING

The challenge was to Load the audio Files and process them to feed them to the pretrained Model

02 - AUDIO TRANSCRIPTION

Extracting Features from Audio File Data using Voice Encoders and Decode them according to usecases

03 - AUDIO TRANSLATION

Converting Natural Language From audio to Machine Language and further converting it to different human languages

04 - AUDIO REVERSE ENGINEERING

Finally Converting the final output Audio to multiple audio File Format

Resources Utilized

<https://arxiv.org/abs/2212.04356>

https://www.researchgate.net/publication/347009752_A_STUDY_OF_THE_TRANSLATION_OF_GOOGLE_TRANSLATE

THANK YOU

Team Non-IITians