SPEECH TO TEXT

So, this project takes a speech (voice) and then converts it to text. The biggest hurdle here is converting the speech and text to numbers so the model can remember the context of the words at different times. Speech recognition is a very hard field because we are dealing with 2 things which are the speech itself and the conversion to text.

So, let’s break it down. First what the model does is, convert the speech into a spectrogram (don’t get trouble, spectrogram is just a graph or image. We will extract the pixels from that graph or image, and we will get the numbers there. That is what the model will use to remember the context of the words at different times - spectrogram is a type of graph, it is uniquely powerful because it visualizes three dimensions of data—**time, frequency, and amplitude**—allowing for detailed analysis and interpretation of signals). In this project, we will be using the transformer architecture to convert the text to numbers (if you remember NLP, then you can remember that there are other techniques as well like TFIDF etc etc) but we will be using the transformer architecture so the model can remember the context of the word (for example, head and headmaster are 2 different words with different meaning). Transformers or TFIDF are not the only model you can use for speech to text. Some others are: Rule based machine learning translation, statistical machine translation, hidden Markov model, deep learning etc. Okay so now this is how we are going to approach this problem.

1. We are going to load the LJ speech data along with-it meta data
2. Create custom embedding for both the speech and the text
3. We are going to create a custom transformer (encoder and decoder) using TensorFlow
4. Download the data from S3 bucket
5. Train the model and evaluate the score

So, first thing first, let build the transformer in jupyter notebook to have some sense of understanding of how it works. So, I’ll advise you to start there (so see C:\Users\midof\OneDrive\Desktop\INeuron\Data\_Science\_Project\Industry\_Ready\_Proj\STT-main\ SpeechToTextTransformersColab.ipynb).