Title: Speech to Text Translation with Machine Learning

By: Adrian Sicaju Ruiz, Omar Suarez, Daniel Maiello

Introduction:

Speech to text translation is the process of converting spoken words into written text. This technology has a wide range of applications, including automatic speech recognition, voice-based search, and language translation. With the increasing use of virtual assistants and the proliferation of smart devices, the demand for speech to text translation has grown significantly in recent years. There has been a significant amount of research in the field of speech to text translation. Early approaches to this problem relied on hand-crafted rules and dictionaries to convert speech to text. However, these methods were limited in their accuracy and flexibility. With the advent of machine learning, researchers have been able to develop more sophisticated methods for speech to text translation. We decided to go along with this project and abandon our work for our Bipedal Walker. Issues that kept occurring and prevented us from utilizing this OpenAI gym was that most versions had been updated, deleted or didn't work with either our IDE's and Jupyter Notebook. Most instructions were of older versions that they had updated over the years and with the little amount of time we had, we thought it would be easier to implement something a little more easy to understand without having to see which versions needed to work with each other.

Problem Definition:

The goal of this project is to develop a machine learning model that can accurately translate speech to text and then take the text to translate. The model should be able to handle a wide range of accents, languages, and speaking styles. The performance of the model will be evaluated based on its ability to accurately transcribe speech to text.

Data:

The data for this project will consist of audio recordings of spoken words and their corresponding written transcriptions. The recordings will be sourced from a variety of sources, including public databases, crowd-sourced data, and real-world recordings. The data will be split into training, validation, and testing sets.

Methods:

The proposed approach for this project is to train a deep neural network to perform speech to text translation. The network will consist of multiple layers, including an input layer, hidden layers, and an output layer. The input to the network will be the audio recordings, and the output will be the corresponding written transcriptions. The network will be trained using supervised learning, with the goal of minimizing the difference between the predicted transcriptions and the ground truth transcriptions.

The pipeline for this project will consist of the following steps:

1. Preprocessing of the audio data: This will involve converting the .wav files into a form that is suitable for input to our model. This may include applying signal processing techniques such as filtering and noise reduction.
2. Feature extraction: We will extract relevant features from the preprocessed audio data that will be used as input to our model. This may include extracting characteristics such as pitch, volume, and spectral features.
3. Model training: We will train a machine learning model, such as a deep neural network, on the extracted features and corresponding transcriptions. This will involve optimizing the model's parameters to minimize the error between the predicted and actual transcriptions.
4. Model evaluation: We will evaluate the performance of our trained model on a held-out dataset of audio files and transcriptions. This will involve comparing the predicted transcriptions to the actual transcriptions and calculating metrics such as accuracy and precision.
5. Model deployment: Once our model has been trained and evaluated, we will deploy it for use in real-world applications. This will involve integrating the model into a system that can accept audio input and produce a transcription as output.

Works Cited

"Read 'Voice Communication between Humans and Machines' at Nap.edu." *Training and Search Methods for Speech Recognition | Voice Communication Between Humans and Machines |The National Academies Press*, https://nap.nationalacademies.org/read/2308/chapter/15#210.