Real-Time Speech-to-Text Translation From One Language to Many Using One-Shot Approach

Introduction:

On a variety of auto caption generation tasks, advanced deep learning neural network algorithms for speech to text translation have achieved state-of-the-art accuracy. However, the methods which convert a voice into a text in different languages, need too many samples and still do not get an upto the mark accuracy. To get better accuracy with fewer samples, we introduce a real-time speech-to-text translation from one language to another where we plan to test this challenge using an unique learning method known as One-Shot Learning.

Literature Survey:

The advancement in science and technology has resulted in the creation of an automatic speech-to-text translation device that converts speech into a visual text format. This paper[1] has developed a Speech translation model that transforms speech data into text in the language of the user's choice. This system was created by incorporating multilingual functionality into the existing Google Speech Recognition model, which is based on natural language processing principles.

Idea/Plan:

The aim of this research is to develop a speech-to-text translation model which learns through a one-shot approach and it enables a person to understand the contents of a real-time video in his native language. The structure of this learning model is that it compares three training word samples using a siamese network, with two of the word samples being identical and one being different. We will train our model using this method. The research samples are then fed into our model. One sample per word will be used in the testing, and the model will determine if the word is identical to a translated text or not. One of the implementations of this model can be, in online classes where there are foreign students who are having trouble understanding what the teacher is saying.

Challenges:

Since we are mixing two separate tasks and teaching the model a new methodology, training and testing this model can be challenging. These tasks must be performed correctly in order for the model to function.When it comes to implementing such training on a broad scale corpus, it can be very difficult. In certain cases, words that sound similar but have different meanings or words with different accents can appear in a speech, and the model would be prone to mixing them up.

Conclusion:
In this paper, we propose real-time speech-to-text translation to multiple languages using an approach “one-shot learning”, which will take fewer numbers of samples to get tested with better accuracy. We anticipate that this model would outperform other state-of-the-art models. This model can be used for any form of speech-to-text conversion. With the aid of this model, one can translate regional speech to text, and it can even be used in telecommunication and multimedia.

Bibliography:
[1] Bano, S. , Jithendra, P. , Niharika, G.L., Sikhi, Y. "Speech to Text Translation enabling Multilingualism",source:2020 IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020
[2] Shadiev, R. , Reynolds, B.L. , Huang, Y. , Shadiev, N. , Wang, W., Laxmisha, R. , Wannapipat, W. “Applying Speech-to-Text Recognition and Computer-Aided Translation for Supporting Multi-lingual Communications in Cross-Cultural Learning Project”, source: 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)

Idea Proposed by:
Group - 13
Priata Nowshin - 20141035
Shahriar Rumi Dipto - 20141036
Galib Abdun Noor - 20141037