

Andrew Smith

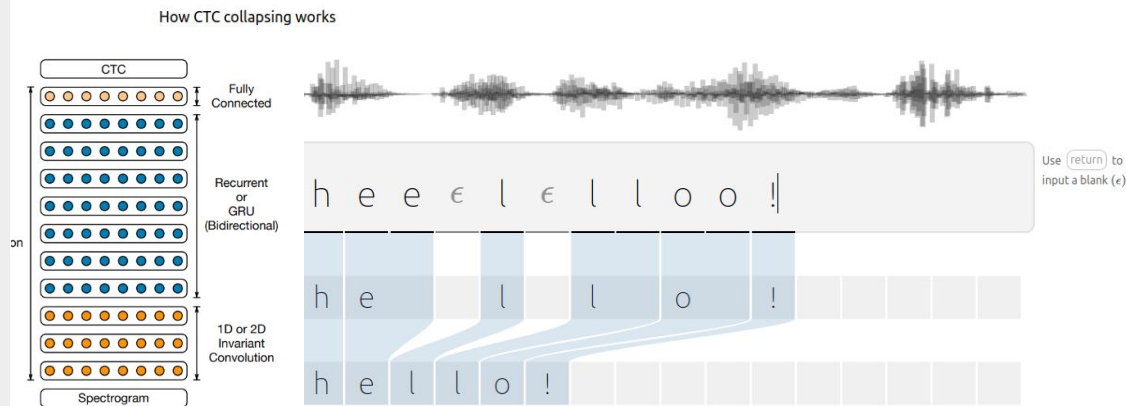
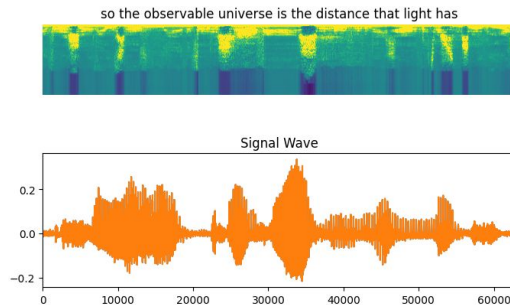
Project Name: Application of Artificial Neural Networks (ANN) to Automatic Speech Recognition (ASR) on a Novel Dataset created using YouTube

Problem: Training a model, trained on a novel data set created using YouTube, that converts speech to text.

Approach: Download (audio,transcript) pairs from various YouTube videos (about 10 hours worth or 12000 input-target pairs). Partition data into training, validation, testing sets that are i.i.d. Train a model with a similar architecture to Deep Speech 2 using the CTC loss function. Test model on testing data, evaluating with Word Error Rate (WER) and human inspection of target-prediction pairs.

Result: On 10% of the YouTube dataset (1200 input-target pairs), our model achieves a WER of 66.9. Additionally, we have some example target prediction pairs shown here.

Comment: DS1 and DS2 were trained on over 7000 hours of audio. Our model was trained on only 10 hours of audio. Thus, it is likely we are severely limiting information available to the model during training. Training our model on our dataset took about an hour per epoch or 50 hours for 50 epochs. Increasing the size of the dataset is likely to increase the time of training as well. We provided a comparison of our model to DS1 and DS2; however, our comparison was not based on the same dataset. In the future, we could, for example, test a pre-trained Deep Speech 2 on our novel dataset from YouTube for a better comparison. The method for creating dataset seems to be efficient and scalable.



- (1) Target: in eighteen eleven williams who murdered the marrs in ratcliffe highway having committed suicide in jail to escape hanging
- (2) Prediction: in aite a leven wiloms wo murdered the marrs and relit hywa hing commisted suasid in jail to sca aning
- (3) Target: on his release an uncle a slopseller in chatham gave him a situation as barker or salesman
- (4) Prediction: on his relys and oncal ay sloselr in shaterm ga himmasituation as barker or salsmon
- (5) Target: even after sentence and until within a few hours of execution he was buoyed up with the hope of reprieve
- (6) Prediction: the vin after senten sand until within of forsobex-cution he as boio with thu hal profrprve
- (7) Target: oswald's marine training
- (8) Prediction: oswalds moain tranning
- (9) Target: these materials are used by the animal in the manufacture of new protoplasm to take the place of that which has been used up
- (10) Prediction: thes materials ar us by the animol in the many-factur of n protoplasm to tak the place of that which has beedusdu