

Assignment 7

Problem Statement

To build the speech model using transformers and writing a report on pre-trained transformers.

Data:

https://huggingface.co/docs/transformers/model_doc/wav2vec2_phoneme

<https://www.kaggle.com/datasets/mfekadu/darpa-timit-acousticphonetic-continuous-speech>

Task 1

Write a report on each of following speech transformers,

- Hubert <https://arxiv.org/pdf/2106.07447.pdf>
- SEW <https://arxiv.org/pdf/2109.06870.pdf>
- Speech2Text <https://arxiv.org/pdf/2010.05171.pdf>
- Wav2Vec <https://arxiv.org/pdf/2006.11477.pdf>

Explain the advancements, features, and comparative theoretical significance of the transformer.

Task 2:

1. Use DR1 and DR2 dialect for training the model and DR3 dialect for evaluating the model.
2. Use wav2vec2 model to identify the phoneme from audios.
3. Calculate the distance between predicted and actual phonemes
4. Apply the same process in the testing set, provide your inferences.
5. Publish the solution notebook in that Kaggle dataset.