

# Outro

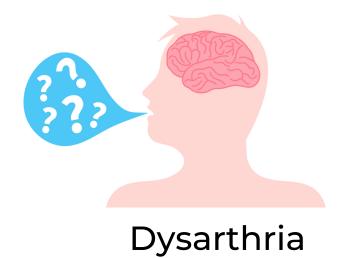




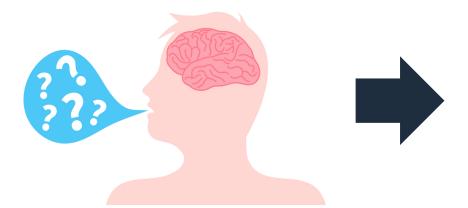


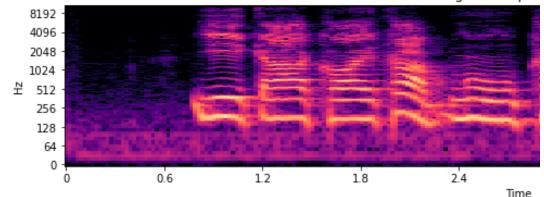


Normal speech

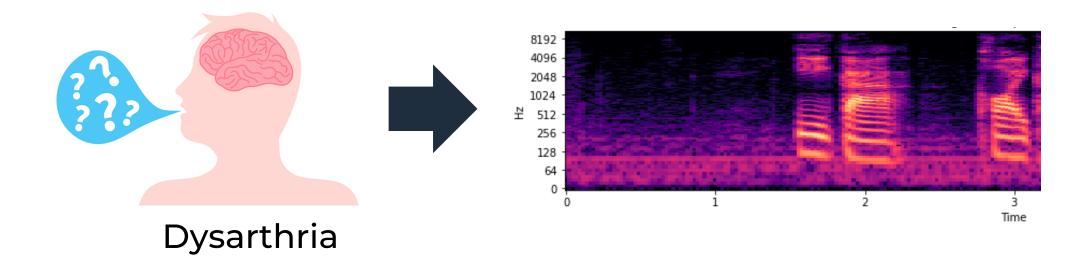






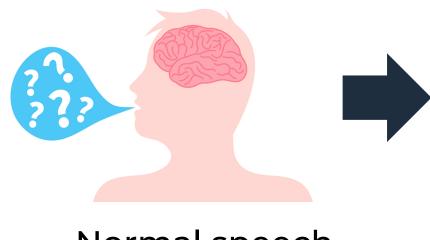


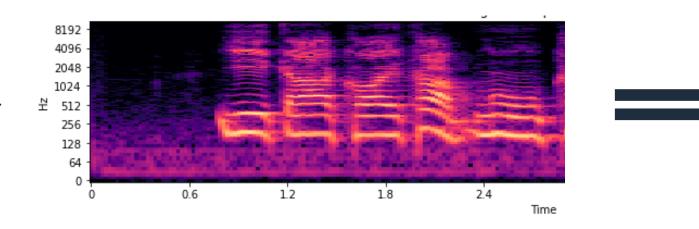
Normal speech

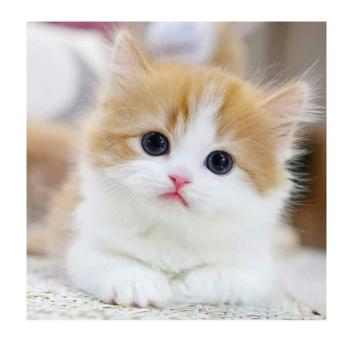






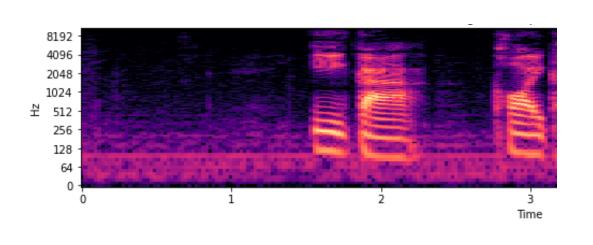






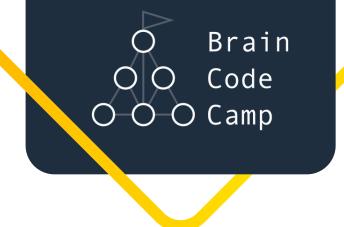
Normal speech





Session: Outro





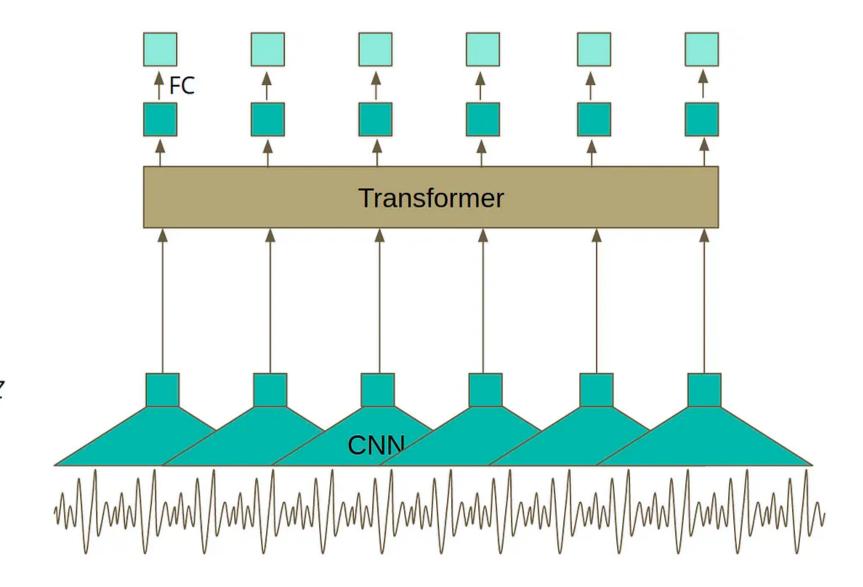
#### Wav2vec

Output Y

Context representations C

Latent speech representations *Z* 

Raw waveform X





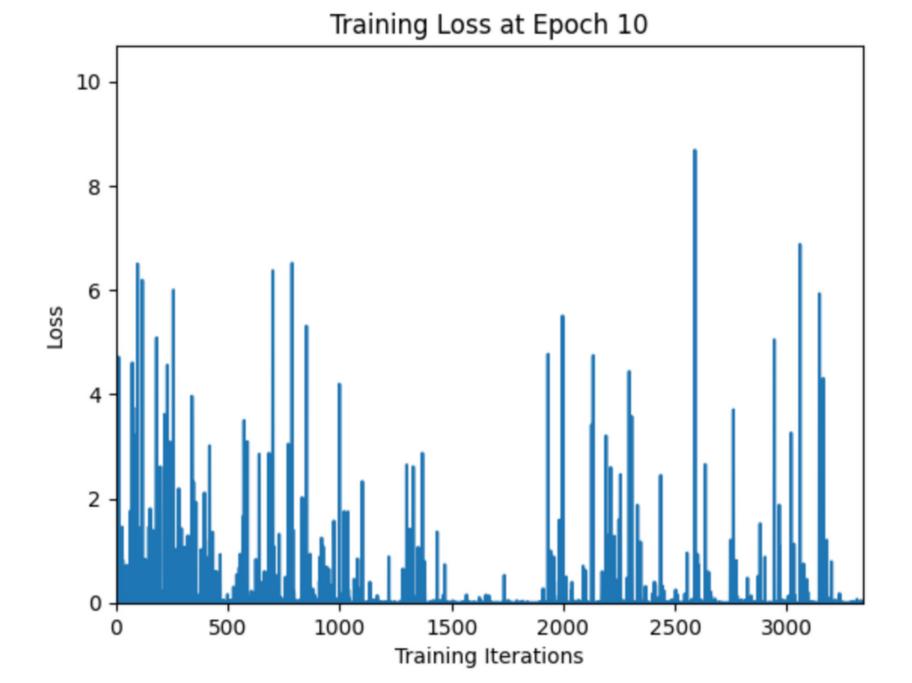
Speaker: Naphapa Panyanirun

Session: Outro



### Results

Speaker: Naphapa Panyanirun



Average Loss after Epoch 10: 0.10438593053260029

Session: Outro

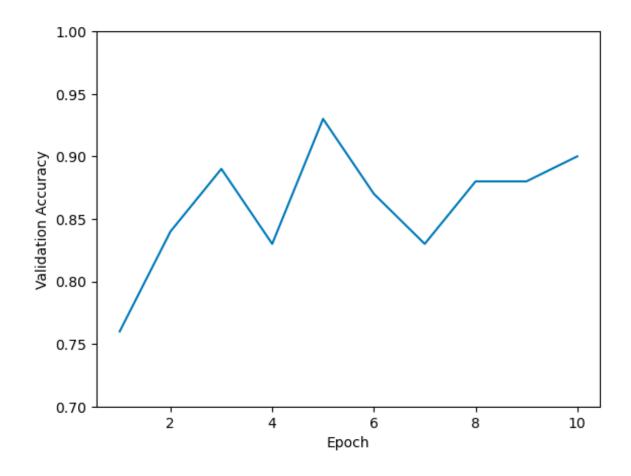
Epoch 10, Train Loss: 0.10438593053260029

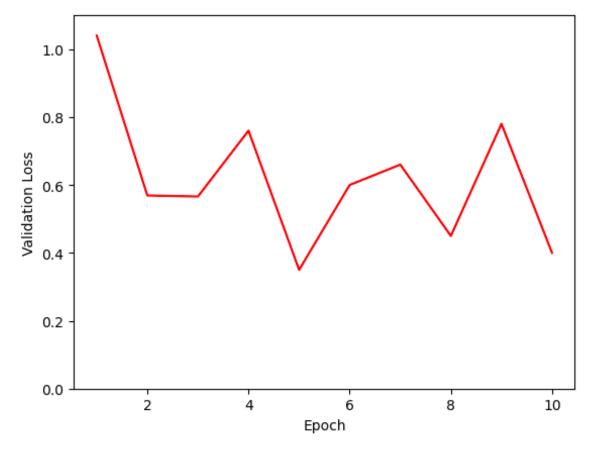
Epoch 10, Validation Loss: 0.3996981418384758, Validation Accuracy: 0.90



## Results

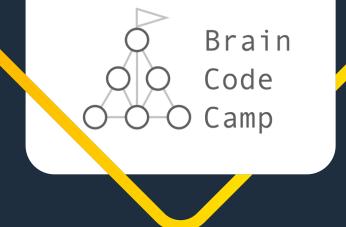
Speaker: Naphapa Panyanirun





Session: Outro Module: Deep learning





# Limitations



Require large amout of data to train effectively



High computational requirement



Costly and time-consuming

