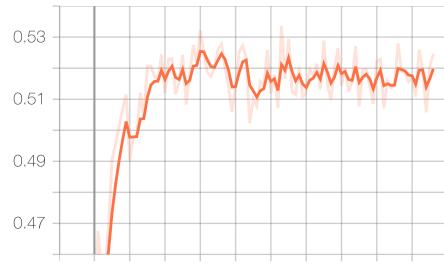
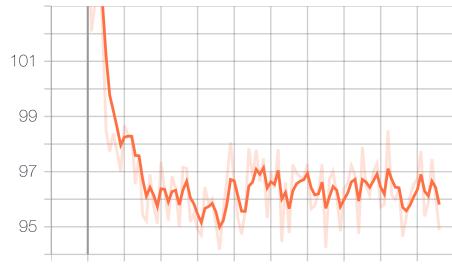
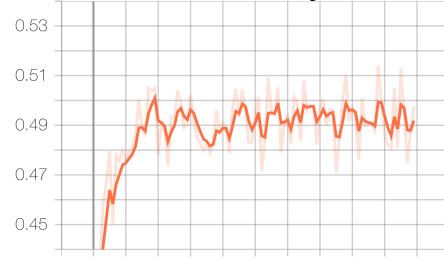


HOME WORK # 5

TRAINING # 2 (MAX ACCURACY: 53%) Sequence length = 10, hidden size 10

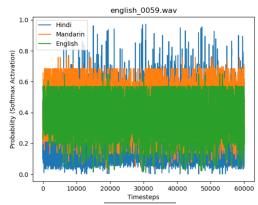
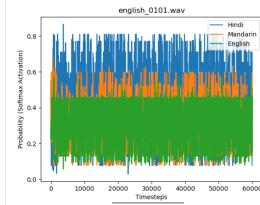
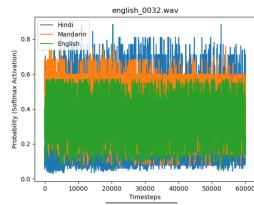


TRAINING # 1 (MAX ACCURACY 51%) Sequence length = 4, hidden size 5

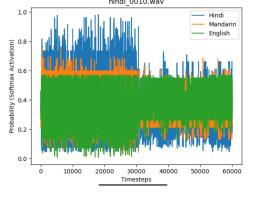
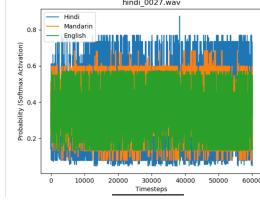
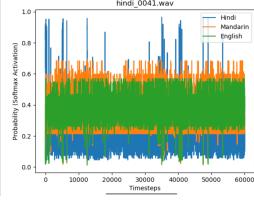


PREDICTIONS IN STREAMING

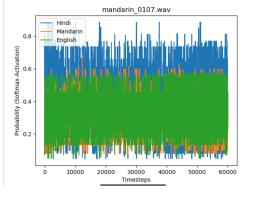
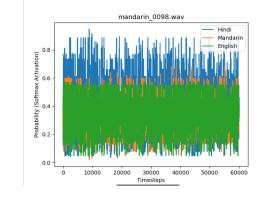
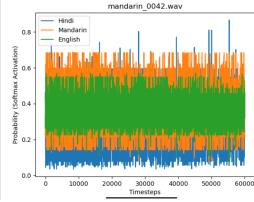
ENGLISH



HINDI



MANDARIN



Summary :

The model uses GRUs with input features of size 64, hidden size 10, output size 3, sequence length was varied from 20-100 and settled on 20. I had started with the RNN example notebook from the lectures to make a barebone model. The data loader has samples from all three languages in one item shuffled across for better training samples. We oversample the lower numbered languages to match English. I had used a preprocessor script to convert faulty audio files (different sampling rates, Voice Activity Detection and silence removal, stereo to mono conversion). After which I trained the model after reserving some of the audio files for testing/validation. Some files below the required 10 minute limits were removed and some had dialogues, tv shows which had to be removed to attempt to improve accuracy. I had tried different approaches to overcome silence (li Rosa trim and voice activity detector using Sox) and trimmed silence.