NLP/ASR

Unit 5: Advanced Topics in ASR/NLP

5.1.3

Attention Mechanisms and Transformers

Transformer models in ASR

Transformers in Speech Recognition

- Transformers excel in processing sequential data, including audio for speech recognition

- Their ability to handle sequences in parallel and capture contextual relationships between data points at different positions makes them well-suited for the dynamic requirements of speech processing

How Transformers Process Speech

- Speech as input is first converted into a spectrogram or a sequence of feature vectors, which represent the audio signal's power at various frequencies over time

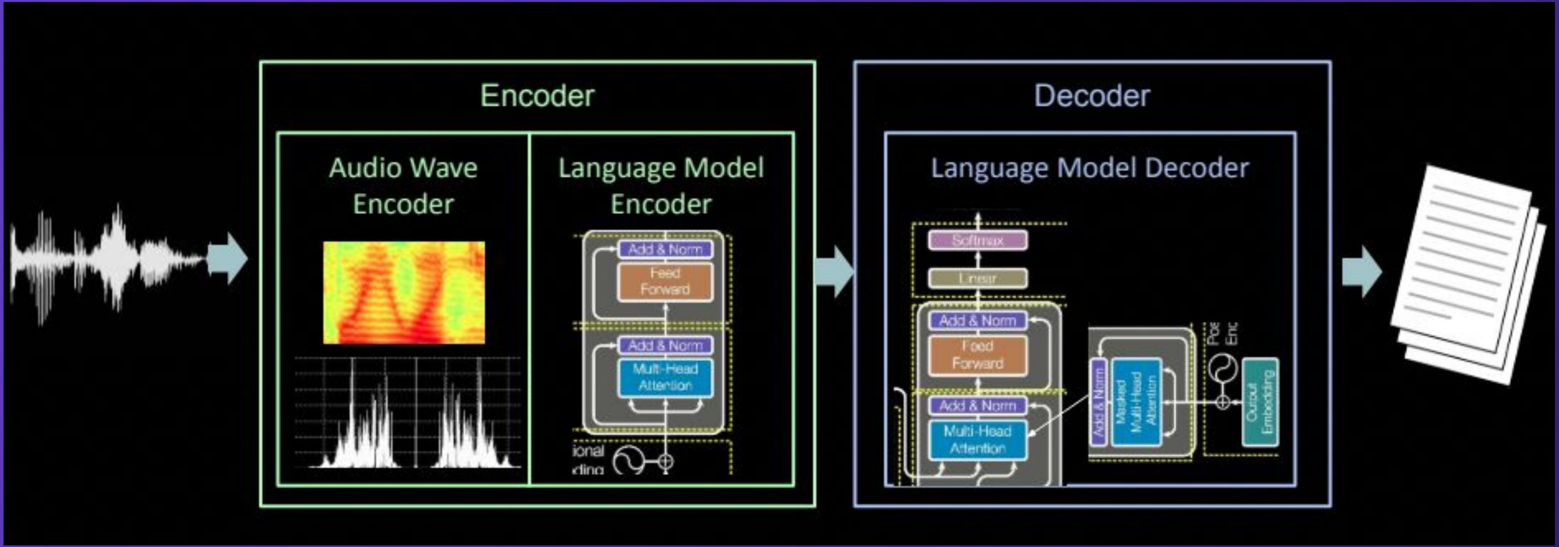
- Transformers apply multi-headed self-attention to this input, allowing the model to focus on different parts of the speech input at once, recognizing patterns like phonemes, syllables, and words concurrently

- The decoder combines the acoustic and language model outputs to generate the final text transcript

(Refer to 5.1.2 for a more detailed explanation of transformer components)

TIL-AI
TODAY I LEARNED-AI

How Transformers Process Speech



(Refer to 5.3.1 for a more in-depth review of different kinds of Transformer architectures e.g. Encoder-Decoder, Encoder-only)

Advantages in Speech Recognition

- Transformers handle long-range dependencies in speech better than traditional methods, such as Hidden Markov Models (HMMs)

- They offer improved accuracy and faster processing times because they can be trained in parallel