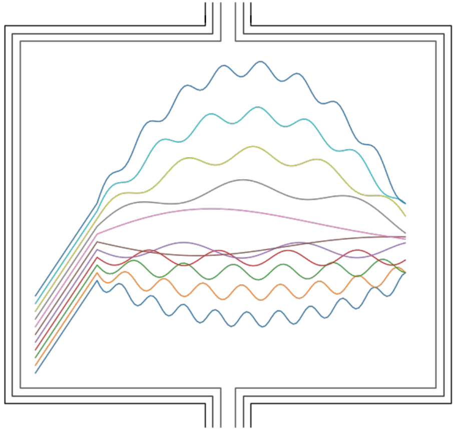
A Penny for Your Thoughts: Decoding Speech from Brain Signals

11-785 Final Course Project

Team*
Quentin Auster, Chuang Ma,
Kateryna Shapovalenko, Demaio Sun

*Ordered alphabetically by last name



Problem Description

Goal: Predict sequences of English words a person has just heard from brain activity.

Practical Impact: Speech-impaired accessibility and development of brain-computer interfaces.

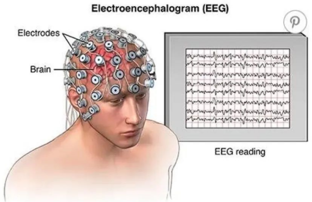
Previous Work: Défossez et al. (Meta) achieve promising results decoding speech from brain signals.

Task and Contribution: Propose new approach for brain-to-speech decoding.

Materials

Types of Brain Signal

EEG

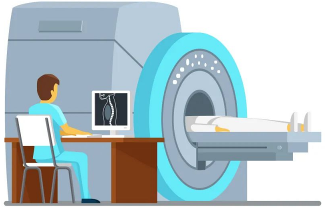


High Temporal
Low Spatial
Cheap/Easy

MEG

High Temporal
High Spatial
Expensive

fMRI



Low Temporal
High Spatial
Expensive

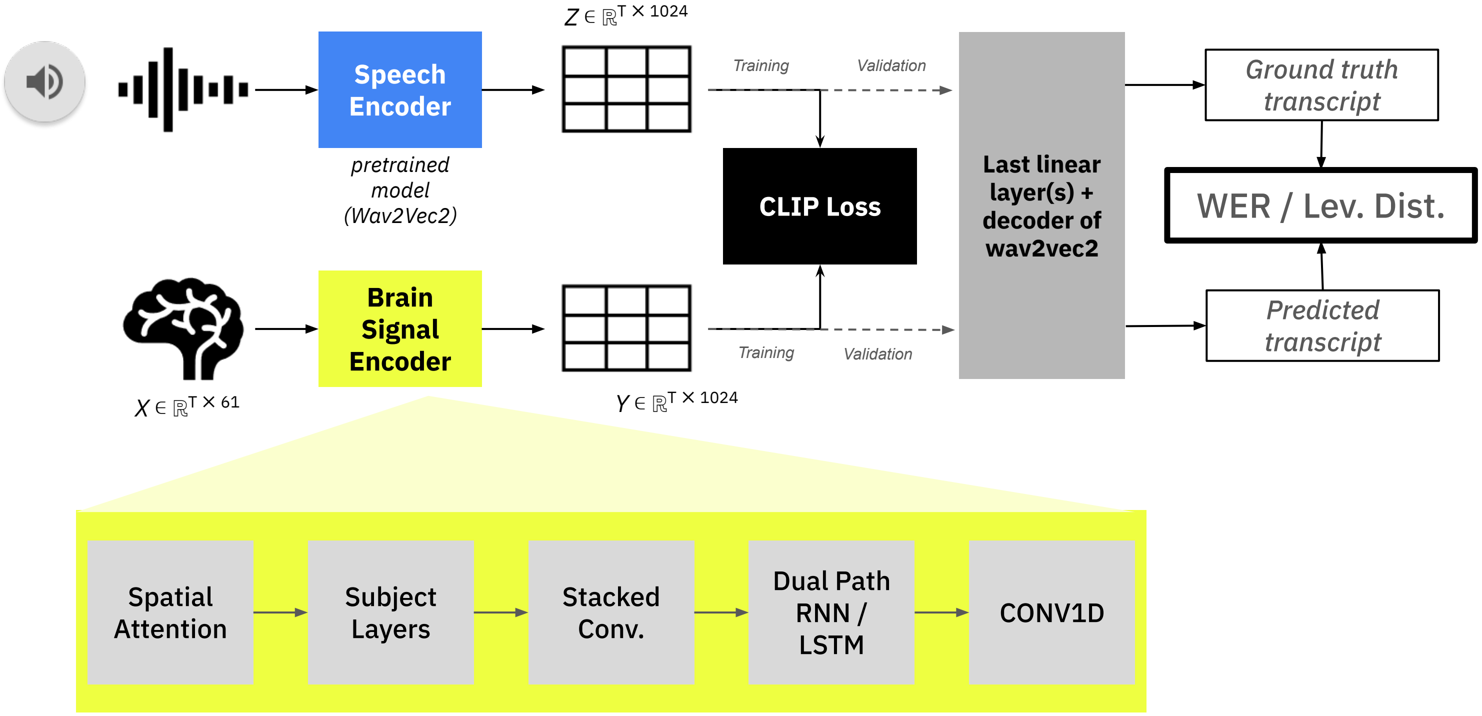
We rely on data from Brennan & Hale (2019):

EEG Data: 60 sensors from 33 subjects, totaling approximately 6.7 hours of recordings. Brennan & Hale excluded 16 participants due to noisy recordings and/or poor comprehension. We drop four additional subjects due to lack of annotations for segment start times.

Audio Data: Audio data recording the reading out loud of Chapter One of Alice in Wonderland is provided in 12 wav files (segments).

Updates: Instead of segmenting EEG at the word level, we segment by each recording.

Model Architecture



Results and Discussion (Ablations)

Key models / hyper-parameters	Pre-processing (signal clamping)	Subject- specific layers & embedding	Spatial attention	Convolutional sequence	Accuracy (100% - Word Error Rate)
Meta AI model (EEG)	Yes, clamp value of 20	Partially	No	Initial & final convs, GELU, skip connections	**25.75% - baseline**
Experiment #1 (EEG)	Yes, clamp value of 100	Yes, subject layers & embedding	No	Initial & final convs, GELU, skip connections	**33.27%**
Experiment #2 (EEG)	Yes, clamp value of 100	Yes, subject layers & embedding	Yes	Initial & final convs, GELU, skip connections	**34.98%**
Experiment #3(EEG)	Yes, clamp value of 20	Yes, subject layers & embedding	Yes	Initial & final convs, GELU, skip connections	**38.41% - best performance**
Key findings		**Subject-specific layer helps to account for individual differences between subjects**	**Spatial attention is key to minimize poor EEG spatial resolution**		

Results and Discussion (Modifications to the Modules)

Modules Modified	Explanation and Motivation	Performance Improvement
Subject Layer	The original subject layer does not have attention. **By adding the attention mechanism**, the this module now can learn the pattern in the brain signal of one specific subject that is important to the uniqueness of this very subject.	**19.1% improvement compared to the baseline model with the original subject layer.**
Spatial Attention	The original spatial attention module that processes the signals from the sensors treats every subject the same way. By enabling the heads mechanism, **each subject has their own attention score matrix**. It thus personalizes the spatial attention mechanism to individual subjects. It acknowledges the inter-individual variability in brain signal pattern.	**20.2% improvement compared to the baseline model with the original spatial attention.**
DualPathRNN	The original DualPathRNN serves as the last but one layer processing the brain signal before comparing. **Dual-direction LSTM** layers are now used and **attention mechanism is added** after each layer of LSTM, which strengthens its capacity in identifying pattern in a long sequence.	**11.4% improvement compared to the baseline model with the original DualPathRNN.**

Conclusions and Next Steps

- The ability to accurately decode speech from brain signals could have widespread implications for accessibility and brain-computer interfaces
- Current challenge: EEG signals, which are less expensive and invasive to record, are more noisy and have worse spatial precision than MEG → **much worse results for EEG than MEG**
- Performance on EEG can be **improved with model and hyperparameter modifications**.
- We are continuing to experiment with additional model architectures and preprocessing techniques to take advantage of the high temporal precision of EEG while minimizing the drawbacks of poorer spatial precision.

Thank you