Course Project Proposal: EEG-to-Speech Kaggle Challenge

Team Members: Ion Turcan, Kirill, David, Elion

October 5, 2025

1 Problem Statement and Application (Ion Turcan 40154098)

In this project we are going to investigate the topic of linguistic neural decoding. The problem is interesting and important because many people that had a brain stroke or suffer from ALS could lose the capability of talking. To restore speech, the patient would need an interface that would capture the brain activity related to speech. Next, this brain activity data can be decoded into the message the person wanted to say through an RNN model. Our main goal is to try to improve existing models. To achieve that we would need to be able to run either an existing benchmark or a baseline model. That would be a challenge since we would need to make sure that the whole pipeline is working. After accomplishing this task, we will try to improve this model through different improvement strategies. These strategies might include data augmentation, model architecture change, using different loss function and using different tokenization strategies. The results are going to be reported and discussed. Trying to improve the model is really challenging because that involves deeper understanding of the whole pipeline and it is very likely that we will need to consult additional materials to have an idea on how these potential improvements could be implemented.

2 Reading Material (Kirill)

- Lee et al. [2]: The model in this paper is very similar to the Kaggle baseline model. It uses a speech module which is not applicable for our project but serves as a good overview of the current SOA for EEG/speech decoding. It contains:
  - EEG Module: trains the model to understand EEG structure.
  - EEG Encoder: compresses raw EEG signals into “latent representations”.
  - EEG Decoder: reconstructs EEG latent representation into raw EEG, acts as a self-supervised learning constraint.
  - Phoneme Predictor: outputs phoneme probabilities by trying to predict phoneme sequences (/p/, /a/, /t/) from compressed EEG signals.
- Papastratis [4]: Reviews/explains various DL architectures for speech recognition. Explains and justifies the use of RNNs and CTC-based models (which is what

the Kaggle baseline uses), and compares the performance of different models while providing suggestions.

- Lee and Lee [3]: Explores how Transformer-based architectures can outperform RNNs for EEG sequence decoding, which could be explored as an alternative to GRUs for modeling temporal relationships in precomputed EEG features.
- Card et al. [1]: Authored by the same team hosting the Kaggle competition. It largely forms the basis of the competition's baseline model architecture.

3 Possible Methodology

- Run existing benchmarks: load without training and infer
  - Stanford-NPTL causal RNN Ensemble + 5gram
  - Stanford-NPTL causal RNN TTA-Ensemble + 5gram
  - UCD-NPL causal RNN + 5gram
  - Write a submission script/module (if missing).
- Build simple baselines:
  - random, mean, median, linear regression, logistic regression, kNN, SVM
- Gather suggested improvements:
  - Data augmentation, preprocessing, supplementing
  - Model architecture changes (GRUs, Transformers, etc.)
  - Hyperparameter tuning methods
  - Loss functions
- Clone and track benchmarks in repository for version control, swap out parts (e.g., optimizer).
- Familiarize with Kaggle leaderboard submissions:
  - Submit dummy csv (done).
  - Submit benchmark results (loaded, not trained).
  - Train and submit benchmark models.

4 Metric Evaluation (David)

We expect a .csv file of our model predictions for the dataset of 1,450 test sentences for this challenge. Our output is therefore sentence predictions based on the input of neural recordings. We will use the performance metric *word error rate* (WER), defined as the edit distance between the decoded sentence and the actual sentence, computed over words. This counts the number of edits such as substitutions, insertions, and deletions necessary to make the predicted sentence match the true sentence. We will also use loss as an ongoing metric during training to determine the confidence of our model in its predictions and to evaluate how loss decreases as the training progresses.

5 Gantt Chart (David)

Phase / Task	Description	Responsible	W1	W2	W3	W4	W5	W6	W7	W8	D
Project Setup	Choose competition, define scope, assign roles	All	✓								S
Proposal Draft	Write and format proposal	All (Lead: Ion)	✓	✓							S
Literature Review	Collect and summarize key papers	Kirill		✓	✓	✓					C
Environment Setup	Configure Kaggle/Colab, clone baselines	David		✓	✓						R
Benchmark Exploration	Run/analyze provided benchmarks	All			✓	✓					In
Baseline Validation	Train and validate baseline models	Elion & Ion				✓	✓				L
Model Improvement	Explore architecture changes, tuning	All					✓	✓	✓		In
Evaluation	Analyze WER trends, outputs	Kirill & David						✓	✓		E
Final Report	Compile final paper/presentation	Elion & Ion							✓	✓	F

6 Bibliography (Elion)

References

[1] Nicholas S. Card, Maitreyee Wairagkar, Carrina Iacobacci, Xianda Hou, Tyler Singer-Clark, Francis R. Willett, Erin M. Kunz, Chaofei Fan, Maryam Vahdati Nia, Darrel R. Deo, Aparna Srinivasan, Eun Young Choi, Matthew F. Glasser, Leigh R. Hochberg, Jaimie M. Henderson, Kiarash Shahlaie, David M. Brandman, and Sergey D. Stavisky. An accurate and rapidly calibrating speech neuroprosthesis. *medRxiv*, page 2023.12.26.23300110, April 2024. doi: 10.1101/2023.12.26.23300110. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11030484/.

[2] Jihwan Lee, Tiantian Feng, Aditya Kommineni, Sudarsana Reddy Kadiri, and Shrikanth Narayanan. Enhancing Listened Speech Decoding from EEG via Parallel Phoneme Sequence Prediction, January 2025. URL http://arxiv.org/abs/2501.04844. arXiv:2501.04844 [eess].

[3] Young-Eun Lee and Seo-Hyun Lee. EEG-Transformer: Self-attention from Transformer Architecture for Decoding EEG of Imagined Speech, December 2021. URL http://arxiv.org/abs/2112.09239. arXiv:2112.09239 [cs].

[4] Ilias Papastratis. Speech Recognition: a review of the different deep learning approaches, July 2021. URL https://theaisummer.com/speech-recognition/.