

Course Project Proposal: Brain-to-Text '25 Kaggle Challenge

Anonymous COMP433/6331 submission

Paper ID 0000

1. Problem Statement and Application

We are investigating the problem of linguistic neural decoding for communication restoration in individuals who have lost speech ability, for example due to stroke or ALS. This project is based on the **Brain-to-Text '25 Kaggle Challenge**, which provides EEG recordings of participants listening to spoken sentences and requires models to decode these into text.

The problem is important because the ability to restore speech would drastically improve quality of life for patients suffering from severe communication impairments. Challenges include reproducing an end-to-end pipeline on EEG data, handling complex input formats, ensuring reproducibility of baselines, and understanding how model components contribute to performance. Our expectation is to first reproduce working benchmarks and then explore improvements. Possible improvement strategies include data augmentation, architectural changes such as switching from RNNs to Transformer variants, experimenting with loss functions, and investigating different tokenization approaches. Throughout, we will document reproducibility issues, bottlenecks, and both successful and unsuccessful attempts.

2. Reading Material

We selected four main references that ground our understanding of the problem:

EEG→Speech with auxiliary phoneme prediction: [2]. This paper trains an EEG encoder–decoder to compress EEG into latent features and reconstruct signals. A phoneme predictor then outputs phoneme probabilities from the latent space, forming a self-supervised constraint on EEG representations. While their model also uses a speech module not directly applicable to our dataset, it is highly relevant to our setup.

Deep learning architectures for speech recognition: [4]. This review surveys RNNs, CTC-based models, and seq2seq approaches. It contextualizes why the Kaggle baseline model uses causal RNNs and CTC, and compares architectures for performance and trade-offs.

Transformer for EEG decoding: [3]. This study shows how Transformer-based architectures can outperform RNNs for EEG sequence modeling, suggesting a promising improvement direction.

Host team's neuroprosthesis baseline: [1]. This paper, from the competition organizers, outlines their ensemble causal RNN with 5-gram language modeling. It provides critical context for the baseline, evaluation, and neuroprosthesis motivation.

3. Possible Methodology

Our methodology involves structured steps to reproduce and extend results:

Technical process:

- **Run existing benchmarks:** load provided Kaggle baselines without training and verify inference. These include Stanford-NPTL causal RNN ensemble (with TTA+5-gram) and UCD-NPL causal RNN (+5-gram). If missing, implement a submission script for Kaggle output.
- **Simple baselines:** implement fast-to-train models such as random, mean, median, linear regression, logistic regression, kNN, and SVM.
- **Improvement strategies:**
 - Hyperparameter search
 - Data preprocessing and augmentation
 - Loss function alternatives
 - Alternative architectures (GRU, Transformer)
- **Repository workflow:** clone Kaggle baselines into our repo for version control, track modifications, and swap components such as optimizers or tokenizers.
- **Kaggle operations:** submit a dummy CSV (done), submit loaded baseline predictions, then trained models for leaderboard evaluation.

4. Metric Evaluation

The primary evaluation metric is *word error rate* (WER), computed as edit distance at the word level (substitutions, insertions, deletions). The official deliverable is a CSV file containing predictions for 1,450 test sentences.

During training, we will track loss and visualize learn-

078 ing curves. For model comparison, we will use WER
079 across baselines and improvements, report leaderboard per-
080 formance, and qualitatively analyze example outputs. Plots
081 of loss trajectories and WER trends over time will be in-
082 cluded in the final report.

083 **Supplement: Gantt Chart (1 page)**

Phase / Task	Description	Responsible	W1	W2	W3	W4	W5	W6	W7	W8	Milestone
Project Setup	Choose competition, scope, roles	All	✓								Selected competition
Proposal Draft	Write and format proposal	All (Lead: Ion)	✓	✓							Submission-ready draft
Literature Review	Collect and summarize key papers	Kirill		✓	✓	✓					Curated reading list
Environment Setup	Kaggle/Colab/Codespaces; clone baselines	David		✓	✓						Reproducible env
Benchmark Exploration	Run/analyze baseline models	All			✓	✓					Initial submission
Baseline Validation	Train/validate baselines	Elion & Ion				✓	✓				Leaderboard score
Model Improvement	Arch changes, tuning, losses	All					✓	✓	✓		Improved model
Evaluation	WER and qualitative outputs	Kirill & David						✓	✓		Eval report
Final Report	Paper & slides	Elion & Ion							✓	✓	Final submission

084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101

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, 2024. [1](#)
- [2] Jihwan Lee, Tiantian Feng, Aditya Kommineni, Sudarsana Reddy Kadiri, and Shrikanth Narayanan. Enhancing Listened Speech Decoding from EEG via Parallel Phoneme Sequence Prediction, 2025. [arXiv:2501.04844 \[eess\]](#). [1](#)
- [3] Young-Eun Lee and Seo-Hyun Lee. EEG-Transformer: Self-attention from Transformer Architecture for Decoding EEG of Imagined Speech, 2021. [arXiv:2112.09239 \[cs\]](#). [1](#)
- [4] Ilias Papastratis. Speech Recognition: a review of the different deep learning approaches, 2021. [1](#)