

Course Project Proposal: EEG-to-Speech 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. An interface records brain activity signals related to intended speech, and machine learning models attempt to decode these into text output.

This problem is both interesting and important: the ability to restore speech would drastically improve quality of life for patients suffering from severe communication impairments. The challenges are significant. They include reproducing an end-to-end pipeline on EEG data, handling complex input formats, ensuring reproducibility of benchmarks, and understanding deeply how each model component contributes to overall performance. Our main expectation is to first reproduce working baselines or benchmark models, then to explore and implement 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 the project, we will document reproducibility issues, highlight the bottlenecks, and report 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 problem 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 archi-

tectures 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 for us.

Host team’s neuroprosthesis baseline: [1]. This paper, from the Kaggle competition organizers, provides critical context for the baseline model, experimental setup, 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 **References**

- 085 [1] Nicholas S. Card, Maitreyee Wairagkar, Carrina Iacobacci,
086 Xianda Hou, Tyler Singer-Clark, Francis R. Willett, Erin M.
087 Kunz, Chaofei Fan, Maryam Vahdati Nia, Darrel R. Deo,
088 Aparna Srinivasan, Eun Young Choi, Matthew F. Glasser,
089 Leigh R. Hochberg, Jaimie M. Henderson, Kiarash Shahlaie,
090 David M. Brandman, and Sergey D. Stavisky. An accurate
091 and rapidly calibrating speech neuroprosthesis. *medRxiv*, page
092 2023.12.26.23300110, 2024. [1](#)
- 093 [2] Jihwan Lee, Tiantian Feng, Aditya Kommineni, Su-
094 darsana Reddy Kadiri, and Shrikanth Narayanan. Enhancing
095 Listened Speech Decoding from EEG via Parallel Phoneme
096 Sequence Prediction, 2025. [arXiv:2501.04844 \[eess\]](#). [1](#)
- 097 [3] Young-Eun Lee and Seo-Hyun Lee. EEG-Transformer: Self-
098 attention from Transformer Architecture for Decoding EEG
099 of Imagined Speech, 2021. [arXiv:2112.09239 \[cs\]](#). [1](#)
- 100 [4] Ilias Papastratis. Speech Recognition: a review of the different
101 deep learning approaches, 2021. [1](#)