# 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:** [8]. 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:** [10]. 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:** [9]. This study shows how Transformer-based architectures can outperform RNNs for EEG sequence modeling, suggesting a promising improvement direction.

**Host team's neuroprosthesis baseline:** [2]. 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 learning curves. For model comparison, we will use WER across baselines and improvements, report leaderboard performance, and qualitatively analyze example outputs. Plots of loss trajectories and WER trends over time will be included in the final report.

COMP433/6331
#0000



# 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 |

## References

[1] Word error rate, 2025. Page Version ID: 1281062858.

[2] 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

[3] 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, Sergey D. Stavisky, and David M. Brandman. An Accurate and Rapidly Calibrating Speech Neuroprosthesis. *New England Journal of Medicine*, 391(7):609–618, 2024. Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa2314132.

[4] Xupeng Chen, Ran Wang, Amirhossein Khalilian-Gourtani, Leyao Yu, Patricia Dugan, Daniel Friedman, Werner Doyle, Orrin Devinsky, Yao Wang, and Adeen Flinker. A neural speech decoding framework leveraging deep learning and speech synthesis. *Nature Machine Intelligence*, 6(4):467–480, 2024. Publisher: Nature Publishing Group.

[5] Debadatta Dash, Paul Ferrari, and Jun Wang. Neural Decoding of Spontaneous Overt and Intended Speech. *Journal of speech, language, and hearing research: JSLHR*, 67(11): 4216–4225, 2024.

[6] Sheng Feng, Heyang Liu, Yu Wang, and Yanfeng Wang. Towards an End-to-End Framework for Invasive Brain Signal Decoding with Large Language Models. In *Interspeech 2024*, pages 1495–1499, 2024. arXiv:2406.11568 [cs].

[7] Erin M. Kunz, Benyamin Abramovich Krasa, Foram Kamdar, Donald T. Avansino, Nick Hahn, Seonghyun Yoon, Akansha Singh, Samuel R. Nason-Tomaszewski, Nicholas S. Card, Justin J. Jude, Brandon G. Jacques, Payton H. Bechefsky, Carrina Iacobacci, Leigh R. Hochberg, Daniel B. Rubin, Ziv M. Williams, David M. Brandman, Sergey D. Stavisky, Nicholas AuYong, Chethan Pandarinath, Shaul Druckmann, Jaimie M. Henderson, and Francis R. Willett. Inner speech in motor cortex and implications for speech neuroprostheses. *Cell*, 188(17):4658–4673.e17, 2025. Publisher: Elsevier.

[8] 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

[9] 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

[10] Ilias Papastratis. Speech Recognition: a review of the different deep learning approaches, 2021. 1

[11] Timothée Proix, Jaime Delgado Saa, Andy Christen, Stephanie Martin, Brian N. Pasley, Robert T. Knight, Xing Tian, David Poeppel, Werner K. Doyle, Orrin Devinsky, Luc H. Arnal, Pierre Mégevand, and Anne-Lise Giraud. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nature Communications*, 13:48, 2022.

[12] Sergey D. Stavisky. Restoring Speech Using Brain–Computer Interfaces. *Annual Review of Biomedical Engineering*, 27(Volume 27, 2025):29–54, 2025. Publisher: Annual Reviews.

[13] Pengfei Sun, Gopala K. Anumanchipalli, and Edward F. Chang. Brain2Char: A Deep Architecture for Decoding Text from Brain Recordings, 2019. arXiv:1909.01401 [cs].

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, 2023. arXiv:1706.03762 [cs].

[15] Francis R. Willett, Erin M. Kunz, Chaofei Fan, Donald T. Avansino, Guy H. Wilson, Eun Young Choi, Foram Kamdar, Matthew F. Glasser, Leigh R. Hochberg, Shaul Druckmann, Krishna V. Shenoy, and Jaimie M. Henderson. A high-performance speech neuroprosthesis. *Nature*, 620(7976): 1031–1036, 2023. Publisher: Nature Publishing Group.