Decoding of Speech Features from Single Neuron Recordings from the Human Brain

Idan Kanat[1], Eadan Schechter[1], Prof. Neta Rabin[2], Dr. Ariel Tankus[3,4,5]

[1] The Joint Data Science Program, Tel Aviv University, Israel.

[2] School of Industrial & Intelligent Systems Engineering, Tel Aviv University, Israel.

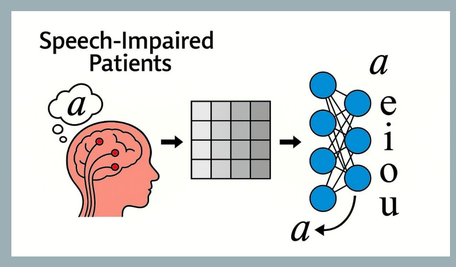
[3] Functional Neurosurgery Unit, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

[4] Department of Neurology and Neurosurgery, School of Medicine, Tel Aviv University, Tel Aviv, Israel

[5] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

Raymond & Beverly Sackler Faculty of Exact Sciences
Tel Aviv University

School of Industrial & Intelligent Systems Engineering
The Iby and Aladar Fleischman Faculty of Engineering
Tel Aviv University



Goal

Our project aims to develop and compare Deep Learning (DL) models for offline decoding of vowel articulations directly from the electrical activity of single neurons in the brain of epilepsy patients. This project paves the way for brain-computer interfaces (BCIs) that will restore speech communication in completely paralyzed patients.

Background & Experiment Setup

People worldwide suffer from neurological conditions that severely limit their ability to speak, leaving them unable to communicate their basic needs. This study extended previous research (Tankus et. al. 2024), which classified two vowels for a single patient. Our project expands this to more vowels, patients and models.
Neural recordings were collected from seven patients with epilepsy using depth electrodes implanted during neurosurgery. Following implantation, as seen in Figure 1 each patient performed trials in which they articulated one of five vowels (/a/, /e/, /i/, /o/, /u/) in response to a beep, while single-neuron activity was recorded from 1 second before to 1 second after the cue.

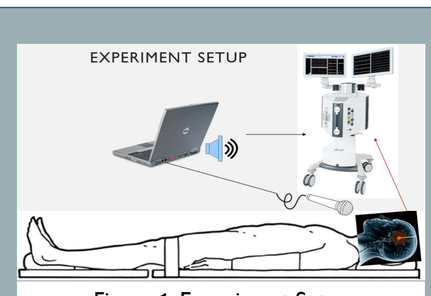
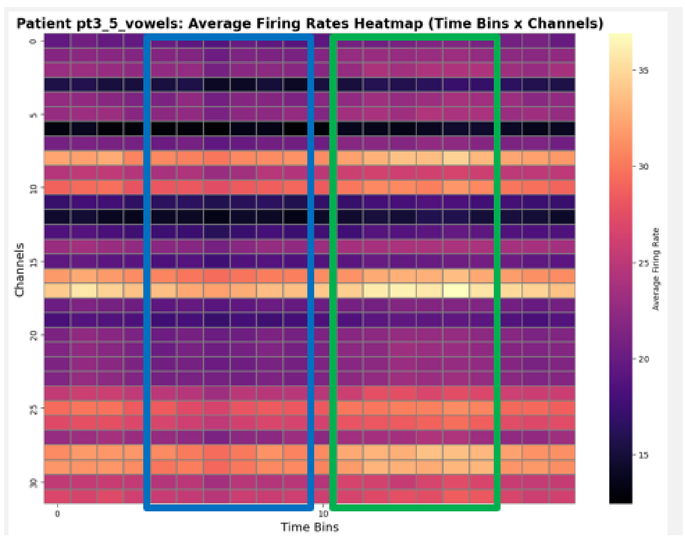


Figure 1: Experiment Setup

Exploratory Data Analysis

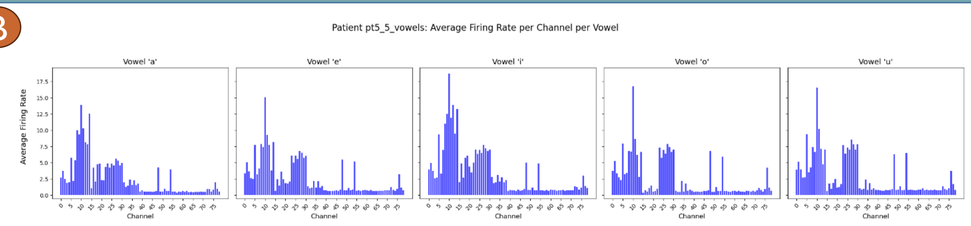
The neuronal recordings of each experiment varied amongst the patients in numerous aspects: number of channels (electrodes), trial counts, and vowel distributions, reflecting real-world clinical heterogeneity.

A



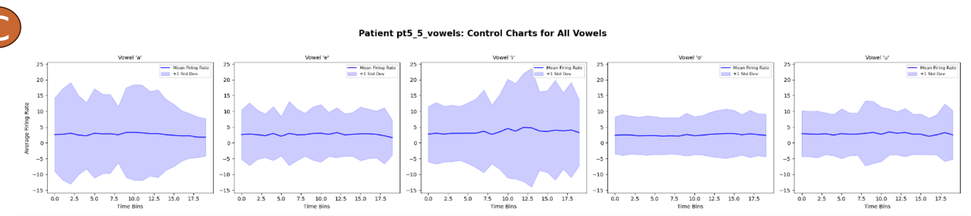
A. Average Firing Rates Heatmap for Patient 3 (Time Bins × Channels): This heatmap displays the average firing rate across all trials for Patient 3, with time bins on the x-axis and recording channels on the y-axis. Brighter colors indicate higher neuronal activity. The blue rectangle highlights the pre-cue window—time bins before the beep sound that signals the start of vowel articulation—representing baseline neuronal activity. The green rectangle marks the post-cue window, corresponding to time bins immediately following the beep. The most notable feature of this heatmap is the pronounced increase in firing rate across several channels within the green (post-cue) window. This surge in activity is time-locked to the auditory cue and likely reflects speech related processes.

B



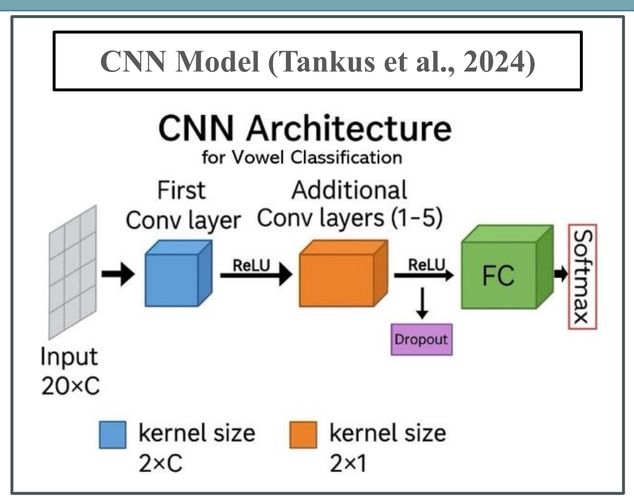
B. **Average Firing Rate per Channel per Vowel for Patient 5:** This visualization shows the average firing rate across channels separately for each vowel class. For patient 5, the overall firing rate distributions remained mostly consistent across vowels, with similar patterns of channel activation regardless of the vowel articulated. This indicates that, at the level of the average firing rate, most channels do not exhibit strong discrimination for individual vowels.

C

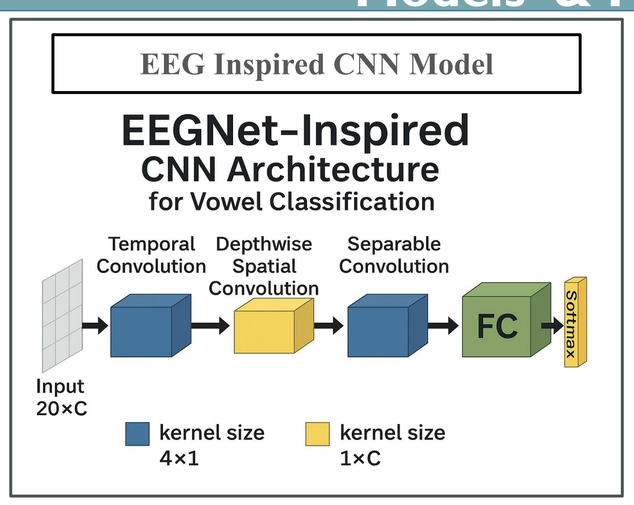


C. **Control Charts of Average Firing Rate Over Time per Vowel (Patient 5):** These control charts display the mean firing rate (blue line) and ±1 standard deviation (shaded) across time bins for each vowel. All vowels are centered around a low mean (~3 Hz), but the variability differs between them - most notably, vowel /i/ exhibits much greater fluctuation.

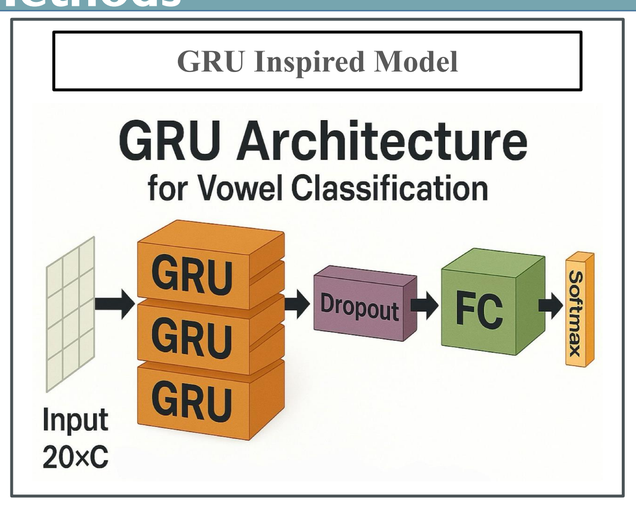
Models & Methods



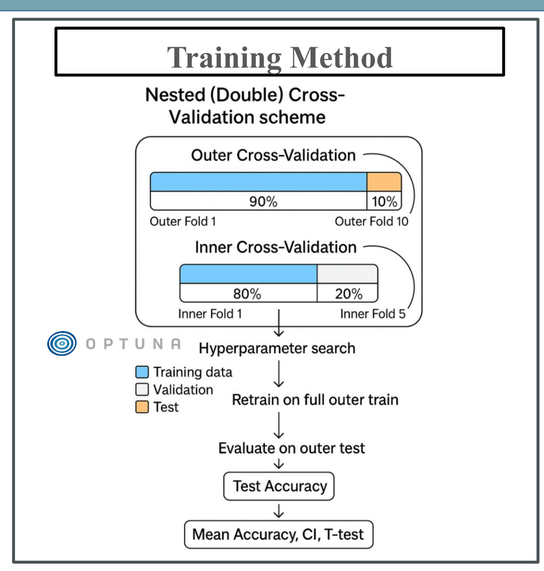
Adapted from "A Speech Neuroprosthesis in the Frontal Lobe and Hippocampus: Decoding High-Frequency Activity into Phonemes" (Tankus et al., 2024). This architecture uses a shallow convolutional network to classify vowels from neural activity. It begins with a spatial 2×C convolutional layers to capture joint time-channel features, followed by up to five 2×1 temporal convolutional layers that extract deeper temporal patterns.



Adapted from "EEGNet: A Compact Convolutional Neural Network for EEG-based Brain–Computer Interfaces" (Lawhern et al., 2018). This model applies domain-specific convolution operations: a temporal convolution (4×1), a depthwise spatial convolution (1×C), and a separable temporal convolution (4×1). It captures fine-grained temporal patterns while maintaining spatial filtering across channels.



Adapted from "A high-performance speech neuroprosthesis" (Willett et al., 2023). This architecture uses stacked Gated Recurrent Units (GRUs) to model sequential dependencies in short 20×C neural recordings. The final GRU hidden state is regularized via dropout, during training only, and fed into a fully connected layer. The model is suited for capturing temporal dynamics in single-trial neural data.



A 10-fold outer cross-validation loop partitions the dataset into training and test splits. For each outer fold, a 5-fold inner cross-validation is performed on the training data to optimize hyperparameters via Optuna, a hyperparameter optimization framework. The best model is then retrained on the entire outer training set and evaluated on the held-out outer test set.

Results

PATIENT 1 - RESULTS

Phoneme Pair	Model	Mean Accuracy (%)	Standard Deviation (%)	95% CI (%)	p-value
/a/ vs /e/	**CNN**	**0.8125**	0.1473	[0.7071, 0.9179]	0.0000
	EEGNet-Inspired CNN	0.5750	0.1208	[0.4886, 0.6614]	0.0406
	GRU	0.7625	0.1713	[0.6400, 0.8850]	0.0005
/a/ vs /i/	**CNN**	**0.7750**	0.0986	[0.7045, 0.8455]	0.0000
	EEGNet-Inspired CNN	0.5000	0.1021	[0.4270, 0.5730]	0.5000
	GRU	0.7375	0.1497	[0.6304, 0.8446]	0.0004
/a/ vs /o/	CNN	0.7750	0.1646	[0.6573, 0.8927]	0.0003
	EEGNet-Inspired CNN	0.5875	0.1324	[0.4928, 0.6822]	0.0331
	GRU	**0.8625**	0.1243	[0.7736, 0.9514]	0.0000
/a/ vs /u/	**CNN**	**0.8500**	0.1291	[0.7576, 0.9424]	0.0000
	EEGNet-Inspired CNN	0.6750	0.1467	[0.5700, 0.7800]	0.0022
	GRU	0.8125	0.1350	[0.7159, 0.9091]	0.0000

The results table compares the performance of all 3 models on binary vowel classification tasks involving /a/ versus each of the four other vowels. Overall, the CNN consistently achieved the highest or near-highest mean accuracy across most phoneme pairs, while the GRU outperformed the CNN for the specific pair of /a/ vs /o/. The EEGNet-Inspired CNN showed the lowest accuracy throughout.

PATIENT 2 - RESULTS

Phoneme Pair	Model	Mean Accuracy (%)	Standard Deviation (%)	95% CI (%)	p-value
/e/ vs /i/	CNN	0.5900	0.1524	[0.4810, 0.6990]	0.0473
	EEGNet-Inspired CNN	**0.6200**	0.1476	[0.5144, 0.7256]	0.0151
	GRU	0.5800	0.1229	[0.4921, 0.6679]	0.0349
/i/ vs /u/	CNN	0.5500	0.1509	[0.4420, 0.6580]	0.1611
	EEGNet-Inspired CNN	**0.6100**	0.1524	[0.5010, 0.7190]	0.0242
	GRU	0.5600	0.1075	[0.4831, 0.6369]	0.0557

For the /e/ vs /i/ and /i/ vs /u/ phoneme pairs, the EEGNet-Inspired CNN achieved the highest mean accuracy among the three models, outperforming both the standard CNN and GRU. However, all models showed only slightly above chance accuracy (ranging from 0.55 to 0.62), indicating these pairs were more challenging to classify. Statistically significant p-values were observed, except for the CNN and GRU in the /i/ vs /u/ task.

Conclusions

- ❑ None of the models that we tested appears to be a universal "best" model.
- ❑ Due to the variability in electrode placements and brain anatomy between patients further analysis is to be considered.
- ❑ Our work demonstrates the technical feasibility of decoding speech elements from single-neuron recordings using deep learning, specifically CNN and GRU-based neural network architectures. These findings support the future development of communication-restoring BCIs.
- ❑ Our study extended the existing literature of decoding single neuron activity in the human frontal lobe to infer speech features using deep learning models never explored before on data from these brain areas.