DRAFT TECHNICAL REPORT

Neural voice activity detection with high-gamma ECoG signal correlation structure using a chronically implanted brain-computer interface in an individual with ALS

Kamil A. Grajski[1]

Abstract

Chronically implanted brain-computer interfaces (BCIs) for speech decoding hold promise for individuals with severe motor impairment. A key translational challenge to wider adoption is understanding and mitigating long-term neural signal variability. This study evaluated whether correlation-based features of electrocorticographic (ECoG) high-gamma signals (HG-C) provide greater long-term stability and robustness than high-gamma log-power (HGLP) features for neural voice activity detection (NVAD).

We analyzed an open-source dataset from Angrick et al., [1] of an individual with amyotrophic lateral sclerosis performing a syllable repetition task. Long short-term memory (LSTM) models were trained on HGLP or HG-C features and evaluated across sessions separated from training by up to six months. Feature importance was estimated with permutation-based marginal effects analysis and robustness was tested by simulated electrode disconnections.

HG-C achieved comparable to superior NVAD performance, with reduced temporal degradation (-8% vs. -23% F1 score decline). HG-C models leveraged distributed correlation structure and remained stable under both random and systematic electrode disconnections. These findings support HG-C as a candidate neural signal representation for speech BCIs which require long-term stability and robustness while maintaining ease of use for users and caregivers.

[1] NuroSci, LLC. Corresponding author: kgrajski@nurosci.org

Keywords

Neural Voice Activity Detection, Electrocorticography, High-Gamma, Correlation, Brain-Computer Interface, Speech Decoding, Signal Stability, Amyotrophic Lateral Sclerosis, Neural Prosthesis, Feature Robustness

Introduction

Chronically implanted brain-computer interfaces (BCIs) for speech decoding and speech synthesis offer transformative potential for individuals with severe motor impairment, including those with amyotrophic lateral sclerosis (ALS) [1-15]. Yet despite encouraging results, widespread clinical deployment faces numerous engineering challenges [16-18]. One of the most critical is maintaining system performance over extended periods of use. Existing BCI systems may resort to the application of dedicated recalibration or normalization sessions or trial-by-trial or session-block normalizations to compensate for changes in neural signal properties over time. These procedures, while technically effective, may present a burden in time and energy for both patients and caregivers, and they increase the operational complexity of the BCI system.

From a Translational Medicine perspective, minimizing the need for calibration is essential to building BCIs that are sustainable outside of research settings. From a Human Factors perspective, reducing the time and effort required from the user and achieving ease of use by patients and their caregivers is crucial for adoption. And from a Bioengineering standpoint, active areas of investigation are focused on how to characterize and best to handle long-term signal variability caused by mechanical shifts (e.g., micro-movements of the electrode grid), biological processes (e.g., inflammation, encapsulation, impedance drift), or hardware disconnection (e.g., electrode, channel and mechanical disconnection) [16-18].

To support more robust systems, we may list key requirements of neural signal features to be used in neural voice activity detection (NVAD) and neural speech decoding, whether for speech-to-text, text-to-speech, or speech-to-speech. Neural signal feature requirements include:

- Capture stable correlates of behavior over time,
- Be robust to spatial reorganization of the underlying neural signal,

- Be compatible with widely used machine learning models,
- Support interpretability and physiological insight,
- Require minimal session-level normalization and retraining,
- Resist disconnection of individual or subsets of signal sources (electrodes).

The point of departure for this study is the ECoG high-gamma (HG) power neural signal which undergirds recent milestone studies [1, 11-15]. Specifically, we hypothesize that ECoG correlation-based features (HG-C) are robust with respect to signal amplitude drift (regardless of root cause in biological or electrical or mechanical factors) and focal and distributed electrode disconnection. We evaluate this hypothesis by comparing the performance of neural voice activity detection (NVAD) models trained on HG-C (correlation-based) features to those trained on HG (log) power envelope features (HGLP), using an open-source dataset [1]. The goal of this technical report is to propose methods, to present results, and to discuss the potential viability of a correlation-based approach available for BCI signal processing and feature development.

The neural data and anonymized speech audio data used in this study were obtained from a publicly available source (http://www.osf.io/49rt7) originated by Angrick, et al. (2024) [1]. The data was recorded from an individual living with ALS who at the time was a participant in a clinical trial of an implantable BCI (ClinicalTrials.gov, NCT03567213). The implantable BCI consisted of a pair of subdural 8x8 electrode grids (2.0 mm diameter; 4.0 mm center-to-center spacing). The device(s) were surgically implanted in the left hemisphere and placed on the pial surface of the sensorimotor representation areas of speech and upper extremities. Angrick, et al., [1] applied this study data towards the estimation of models to generate online (real-time) speech synthesis from the 1000Hz sampled electrocorticograms (ECoG). Their end-to-end system processed input comprised of digitally filtered ECoG high-gamma power (HGP) envelope activity. The system utilized three neural network stages: a) a unidirectional or causal neural network model that identified speech segments; b) a bidirectional model that translated the ECoG in voice-labelled segments to LPC coefficients; and c) LPCNet which converted LPC coefficients to acoustic speech.

In the present study we focused on neural voice activity detection. NVAD provides a means to test the hypothesis that correlation-based features of ECoG meet requirements for effective

neural speech signal representation. The study design was in four stages. The first stage was to reproduce published findings obtained from the original open-source data. This entailed training an LSTM model comparable with that applied by Angrick, et al., [1] using ECoG HGLP as input features and measuring performance (F1 score) as a function of temporal distance from last training set. These results validated that the data was properly handled and set the stage to test the study hypothesis.

The second stage was to generate and to explore ECoG correlation spatial and temporal structure. This stage answered basic questions. What does ECoG HG-C data look like temporally and spatially? Could one "see" the neural speech signal in ECoG HG-C? Could one quantify the spatial extent of ECoG HG-C such as through a spatial decay constant?

The third stage entailed training an LSTM model of the same architecture as in the first stage. In this third stage, the ECoG HG-C input feature data was first passed through a two-layer Linear Network for dimensionality reduction to match ECoG HGLP input dimensionality (N=64). These results include a comparison of ECoG HGLP and ECoG HG-C NVAD performance stability and spatial distribution of marginal effects.

The final stage explored the effect on ECoG HGLP and ECoG HG-C NVAD model performance in response to randomized and systematic electrode disconnection.

Methods

Data source

The study and data were designed as a collection of thirteen (13) experimental sessions over a period of approximately seven months [1]. We followed Angrick, et al., [1] and organized the sessions into four groups. The training group consisted of the first eight sessions recorded over a one-month period. The testing session consisted of one session recorded approximately one week following the last training session. A single validation session followed the testing session by one day. Finally, three sessions were recorded approximately six months following the validation session.

Each recording session consisted of blocks of task-specific trials. The first block of trials consisted of a syllable repetition task (SRT). In an SRT trial, the participant was presented with

an acoustic stimulus consisting of one of twelve consonant-syllable pairs. In pseudo-random sequence, each of the twelve stimuli was presented five times, yielding a total of 60 trials. The second through n^{th} block of trials consisted of a keyword repetition task (KRT). In a KRT trial, the participant was prompted by a visual cue on a computer screen to read aloud one of six keywords or to remain silent. In pseudo-random sequence, each of the six stimuli were presented ten times, yielding a total of 60 trials. Sessions contained at least one and up to eight KRT blocks.

In their study, Angrick, et al., [1] utilized the SRT data for the purposes of gathering statistics to apply as baseline normalization of ECoG HGP features in the KRT data. In this study, we focused initially and exclusively on the SRT trials (N=780) as a source of NVAD training, testing, validation, and long-term stability testing. See Table 1. Metadata provided with the publicly available dataset included trial type, stimulus identifier, stimulus onset and offset times, and time-synchronized microphone audio recordings. With this information we defined a standardized NVAD trial shown in Figure 1.

Signal processing

Signal processing of ECoG data was implemented on a session-independent, task-independent, block-independent, trial-independent, and electrode channel-independent basis. There were three stages of signal processing. In the first stage, raw ECoG sample data time series were imputed for “bad channels” by simple nearest-neighbor spatial averaging. Angrick, et al., [1] identified 0-3 bad channels for certain sessions which this study accepted. Additional detailed quality checking did not identify any additional bad channels.

In the second stage, pre-processing, the ECoG time series was centered to zero mean, digitally band-pass filtered to 70-170Hz, and then notch-filtered at 120Hz. A variation on strict zero-mean normalization was explored in which the top N% of values in the series were masked out of the mean estimation. A range of values of N (1-10%) were explored in a hyperparameter tuning stage (see below). The bandpass and notch filters were forward-backward digital filters using cascaded second order sections (SOS) which ensured that the output signals were temporally aligned with the input signals (including the audio signal). The output signal of the pre-processing stage was labelled as the ECoG high gamma (HG) time series.

The third stage implemented feature extraction from the HG signal. First, the HG envelope magnitude was estimated using the Hilbert transform. Second, the windowed instantaneous high-gamma log power signal was estimated in a moving window (48 msec, duration; 12 msec, stride) by computing the mean of the squared magnitude and then taking log10. The resulting signal was referred to in this study as the HGLP feature data. Each trial consisted of N=224 windows (frames).

Windowed correlation-based features were extracted from the same HG time series used for HGLP estimation. For each window, the Pearson cross-correlation matrix was derived from the covariance matrix. The Pearson correlation coefficients were Fisher Z-transformed and stored as a strict lower triangular matrix. This signal was referred to in this study as the HG-C feature data.

When working with both 8x8 electrodes, the dimensionality of the HGLP feature vector for each window was 128 and the dimensionality of the HG-C feature vector was 8,128. When working with just one 8x8 electrode, such as the ventral grid, the HGLP feature dimension was 64 and the HG-C dimension was 2,016.

Spatial correlation decay constants were estimated for HG-C (but not HGLP) by analyzing the relationship between inter-electrode distance and correlation estimate within each electrode grid (ventral and dorsal). For each session-stimulus combination, correlation matrices were computed using sliding time windows with Fisher Z-transformation for statistical validity during averaging across multiple trial repetitions. Physical distances between electrode pairs were calculated using the known 8×8 grid geometry with 4mm inter-electrode spacing, and correlation values were grouped by exact distance values (0, 4, 5.66, 8, 11.31... mm). An exponential decay model ($\rho(d) = \rho_0 \cdot e^{(-d/\lambda)}$) was fitted to the distance-correlation data using least-squares optimization to extract the spatial decay constant λ, which quantified how rapidly neural correlations diminish with increasing physical separation.

NVAD Labelling

This study leveraged the closed-loop nature of the SRT behavioral task to generate labelled training data for the NVAD task. Labelled training data consisted of assigning to each windowed feature vector (frame) the label 0 if “no speech” and 1 if “speech”. Automated labelling was by application of an adaptive threshold detector to the energy envelope of the audio signal. ECoG

data was not referenced in any way during NVAD labelling procedure: no data leakage. Following labelling and using the format of Figure 1, each trial was individually visually inspected and judged to be "correct" to within 1-2 window frames (12-24 msec) for speech onset and speech offset. To maximize reproducibility, there was no manual adjustment of frame labels. The labels were fixed once and only once at the start of the study and prior to any modeling or performance results from any feature representation. Summary statistics of speech activity for the 780-trials in this study dataset are listed in Table 1. No trials were excluded from the dataset.

Neural Network Model

The neural network model selected was a stacked two-layer LSTM (with dropout between the layers) followed by a linear fully connected output layer with two units (one signaling "speech"; the other signaling "no speech"). Ease of comparison with Angrick, et al., [1] results drove this decision. The nature of the NVAD task, including its time scale(s), dimensionality, and degree of difficulty, did not immediately indicate a need for more sophisticated architectures.

For HGLP models, the input layer was fixed at either 128 (dual grid) or 64 (single grid). For HG-C models, a multi-layer Linear Network (internal stages with ReLu output, but not final Linear Network stage) was incorporated ahead of the LSTM model. In the case of a single grid (ventral) model, the Linear Network mapped 2,016-dimensional input features to 64 output features then used as input to the LSTM model. In the case of a dual grid model, the Linear Network mapped 8,128-dimensional input features to 128 as input to the LSTM model. An identified alternative architecture was identified, but not explored, namely, to generate through an independent unsupervised procedure an HG-C VQ-VAE dimensionality reduction representation. The avenue remains open for further exploration.

Using Ray Tunes and Optuna an extensive hyperparameter search (Ray trial counts explored included: N=32, 128, 256) was executed to estimate optimal signal pre-processing, network architectural, and training parameters. For both HGLP and HG-C, the LSTM architectural hyperparameters included number of hidden units. For HG-C, the Linear Network architectural hyperparameters included number of linear layers and number of hidden units per linear layer. For HGLP, training parameters included alternative normalization methods, such as masking n^{th} percentile high-power values to better estimate a time series noise floor. For both HGLP and HG-C, training hyperparameters included batch size, number of training epochs, and learning

rate (with fixed *ReduceLROnPlateau* parameters). The HGLP NVAD model and HG-C NVAD model selected for analysis were the final model trained subject to the number of epochs indicated by the hyperparameter search. The final ECoG HGLP LSTM model consisted of two-layers with 256 hidden units each and dropout 50% between layers, and was trained with batch size equal to four, for 30 epochs, and learning rate 8.0×10^{-5}. The final ECoG HG-C LSTM model had the same architecture as for ECoG HGLP, the input Linear Network had 256 hidden units with 50% dropout which achieved the mapping from 2,016 to 256 to 64 dimensions, and was trained with batch size equal to eight, for 60 epochs, and learning rate 3.0×10^{-5}.

Training and Performance Metrics

The neural network NVAD models were trained using only the training sessions. The training utilized cross-entry loss estimation and the *RMSprop* optimizer implemented in PyTorch. LSTM internal state was reset to zero for each input trial. The standardized trial permitted batch operations for rapid hyperparameter search, training, and inference.

The key performance metric applied in this study was the F1 score for the per frame accuracy of classification of speech and no-speech. The study data had a class imbalance of 75% non-speech and 25% speech frames. In addition to F1 score, accuracy, precision, recall, AUC-ROC, and AUC-PR were recorded.

To assess the performance of the models following training, the test set was used to identify the optimal probability threshold with which to label a frame as “speech”. Thereafter, the probability threshold was frozen and applied in the evaluation of the validation and three online sessions. This method supported evaluation of model training stability on progressively time-removed evaluation sessions. For hyperparameter tuning, rather than training set loss, the figure-of-merit (FOM) was test-set F1 score (following optimal probability threshold search).

Marginal Effects

To further assess the performance of the NVAD models, we computed individual feature level marginal effects. The marginal effect of an individual feature was defined as the change in test set F1 score when the feature was permuted while all others were held constant: baseline test set F1 score minus permuted test set F1 score. Permutation preserves marginal distribution while randomizing its relationship to signal. A positive value indicated that permuting the feature

degraded performance, implying it was “useful” to the model. The distribution of test set F1 scores under model training with correct labels was compared to that of the same model trained with permuted target labels. In an initial pass, we pooled both 8x8 electrode arrays. We found marginal effects computed using the F1 score approach to be largely, but not completely comparable to the findings reported by Angrick, et al., [1] who used a partial derivative-based saliency method. Where Angrick, et al., [1] registered saliency for four electrodes in the dorsal grid (Figure 3, [1]), our method measured negative marginal effect. Consequently, the present study reports NVAD training and performance evaluation based on ventral grid inputs only.

Electrode Disconnection Analysis

To simulate performance impact of post-training neural signal loss, we simulated the disconnection of electrodes. In the case of HGLP, electrode disconnection simulation entailed setting the value of that electrode to zero for the trials in a session. In the case of HG-C, electrode disconnection meant setting the value of each of the correlation pairs involving that electrode to zero. This study explored two patterns of electrode disconnection: random; and systematic. In one approach, random sets of electrodes were disconnected while keeping the remainder of the signals intact. The mean and standard deviation of session F1 scores were computed over a set of repeated trials whose count was adaptively set in proportion to the number of possible electrode selection sequences. In the second approach, individual or small local neighborhoods of electrodes were sequentially disconnected while keeping the remaining signals intact. The NVAD F1 score for each post-training session was plotted against the number of electrodes disconnected. These analyses were conducted for HGLP and HG-C and model performance was compared to models trained with correctly labeled target data or permuted target data.

Results

NVAD with ECoG Power Features

This study defined a standardized trial and focused on NVAD of the voiced output recorded during the clinical trial Syllable Repetition Task (SRT). See Figure 1. Table 1 lists the study data trial counts and speech latency and duration in the N=780 trial SRT data set. At the outset

of this study, both dorsal and ventral 8x8 electrode grids were included, meaning feature input dimension N=128. The complete workflow described in Methods was applied to the 128-dimensional feature data, including marginal analysis (*not shown*). The results largely confirmed, but not completely, those reported by Angrick, et al., [1] who used a partial-derivative saliency analysis method. Consequently, all HGLP and HG-C NVAD performance results reported in this study were based on data from the ventral 8x8 electrode grid.

Training with ECoG HGLP progressed with the ventral electrode grid only as described in Methods. NVAD performance results across the set of post-training sessions are listed in Table 2 (second column labelled "F1 (HGLP)"). As noted, due to the class imbalance between speech (25%) and non-speech (75%) frames, this study preferred the F1 score. To foster direct comparison, it is noted that Angrick, et al., [1] reported a frame-wise accuracy of 93.4% on held-out validation data from the Keyword Repetition Task (KRT). In this study, we obtained 94.2% frame-wise accuracy on the held-out test set and 92.4% on the held-out validation set.

A marginal effects procedure was applied to the ventral electrode grid HGLP NVAD model. The distribution of marginal effect values is shown in Figures 2. Figure 3 shows the spatial distribution of marginal effect values. The results reproduced key aspects of those shown by Angrick, et al., in their paper in Figure 3 [1]. That there was conformance with published findings down to the individual electrode level, such as electrode #21, gave confidence that the present study properly handled the Angrick, et al., [1] data.

ECoG Power Correlation

The standard trial provided a convenient framework with which to perform initial data visualization of ECoG HG-C raw data. The temporal structure of HG-C is shown in Figure 4. Figure 4 shows overlays (spaghetti plots) of the pair-wise HG-C time series separately for the dorsal 8x8 grid and ventral 8x8 grid (N=2,016 overlaid pair-wise correlation series per grid). The results were visual evidence that as with ECoG HGLP it was possible visually to observe neural correlates of speech activity. There was visually observable greater activity (naively, a count of the number of plots) in the ventral compared to dorsal. The ECoG HG-C raw data spatiotemporal structure is visualized in Figure 5. Figures 4 and 5 provide an initial visual "signature" of the potential for correlation structure as BCI neural speech signal.

A quantitative analysis was performed to assess the stability and reproducibility of the visually observed HG-C signature. The quantitative measure selected was the correlation spatial decay constant measured in millimeters (mm). The analysis was performed independently for the dorsal and ventral electrode grids and further by grouped by pre-stimulus, stimulus, and post-stimulus period within the standardized trial framework. Figure 6 shows typical results for one stimulus ("BAH") across the five individual SRT trial repetitions and averaged response) recorded during a single training session. We noted the following. First, the ventral spatial decay constant (38-48mm) was 4-5x longer than the dorsal spatial decay constant (8-10mm), irrespective of pre-stimulus, stimulus, and post-stimulus time periods. Second, the ventral post-stimulus period showed an increase in correlation strength and longer spatial decay compared to stimulus and pre-stimulus periods. Third, the variability in the pre-stimulus period may be due in part to the shorter period (144 msec) compared to stimulus and post-stimulus time periods (>1.0 sec). The results were evidence of stability and reproducibility of neural signal correlations that could support input features for NVAD model training and evaluation.

NVAD with ECoG Power Correlation Feature

Training with ECoG HG-C progressed with the ventral electrode grid only as described in Methods. NVAD performance results across the set of post-training sessions are listed in Table 2. We noted the following. First, using HG-C feature data alone generated performance slightly to significantly improved compared with HGLP. Second, HG-C NVAD model performance decayed less with increasing temporal distance. For example, the F1 score change from test session to the last online session is -23% for HGLP and -8% for HG-C.

A marginal effects procedure was applied to the ventral electrode grid HG-C NVAD model. Results are shown in Figures 7 and 8. Figure 7 shows that the distribution of marginal effects is concentrated around a positive mean value and possesses a short negative tail and long positive tail. Models retrained with permuted speech target labels and analyzed with the above marginal effects procedure yielded the value of zero marginal effect. These results show that the models were not merely latching on to spurious features in the input data. Figure 8 shows the spatial distribution of HG-C marginal effects. The display is in the form of a symmetric correlation matrix, rotated for ease of comparison with Angrick, et al., [1] electrode labeling convention which placed electrode #1 on the rostral-most ventral-most position in the ventral grid. The HG-

C results were compatible with findings with HGLP. For example, correlation pairs that included electrode #21 appeared as high marginal effect features. See Figure 8 (left). However, HG-C feature marginal effects with comparably elevated marginal effects were observed at multiple spatial locations. These are visualized by plotting individual feature marginal effect relative to the global average marginal effect. See Figure 8 (right). This was consistent with the model leveraging distributed correlation structure.

Simulated Electrode Disconnection Analysis

The marginal effect procedure considered the case in which the marginal distribution of a feature was constant, but its relationship to target signal was permuted. As a test of feature robustness, we simulated disconnection of individual and subsets of electrodes. Figure 9 shows results when a single electrode at a time is disconnected by setting its input value to 0. The results underscored the central importance of electrode #21 and immediate neighbors to the performance of the HGLP NVAD model. By visual inspection, HG-C is less dependent on input features from individual electrodes. Figure 10 shows results when random subsets of electrodes ranging in size from 1-16 are disconnected. For HGLP NVAD, performance results degraded monotonically but with increasing standard deviation as disconnected subset size increased. For HG-C NVAD, there was a “knee” in the performance curve between 5-10 electrodes. The standard deviation grew significantly beyond but less so ahead of the knee.

Discussion

Summary

This study demonstrated that ECoG HG-C features support more stable and robust neural voice activity detection compared to high-gamma log-power (HGLP) features - *when applied to neural signal data from a single participant performing the SRT task.* Across multiple sessions spanning six months, HG-C models showed reduced temporal performance decay, greater resilience to electrode disconnection, and a distributed spatial representation of neural activity. These findings highlight HG-C as a viable signal representation for long-term BCI applications. We conclude this report with a discussion of study limitations, the significance of the results, and next steps.

Study Weaknesses

There are several potential weaknesses that could limit the further applicability of the findings. First, the task itself, consonant-vowel repetition, could be argued as too restrictive and artificial. That is certainly true. The reasoning to go forward was simply that if the hypothesized results could not be observed under such favorable circumstances, the hypothesis rather than the data would be questioned.

Second, it could be argued that the dataset is too small and the risk of overtraining is high. This study used 780 trials each with 224 frames for a grand total of 174,720 frames (of either 64-dimensional or 2,016-dimensional input features), of which 107,520 were reserved for training. The Methods applied took steps to avoid overtraining risk by intense hyperparameter tuning, aggressive dropout rates in the LSTM model, and in the case of HG-C input, the Linear Model stage. Overtraining was likely not responsible for the findings as demonstrated by the marginal effects analysis, the systematic electrode simulated disconnect experiments, the random subset electrode disconnection experiments, and comparison with feature data permutation (not shown) and target label permutation (see Figures 2, 3, 7, and 8).

A third potential weakness in this study is that the comparison between HGLP and HG-C was not fair, because the HGLP model had more trainable parameters. While the HGLP and HG-C NVAD LSTM model architectures were identical, the HG-C required a pre-input Linear Network to reduce input dimensionality from 2,016 to 64 (to match the 64 HG-C input dimensionality).

The hyperparameter search strategy likely made this less of an issue, but it remains a viable concern.

Last, the study was closed loop. The participant was able to generate speech response. With recorded time-aligned speech it was possible to label the data as speech and non-speech for use in supervised NVAD model training. This is a valid concern and is directed more towards the training method itself than the viability of HG-C as a candidate neural signal feature for BCI. Indeed, an unsupervised NVAD training procedure has been explored [19] and reported [20].

Significance of Findings

This study tested and confirmed the hypothesis that ECoG HG correlation structure has desirable properties required for long-term BCI applications. The significance of these findings is at least four-fold. First, the present results were achieved without recourse to session-based calibration or normalization. The signal pre-processing strategy simply was to Z-transform (for HGLP) or Fisher Z-transform (for HG-C) on a per-trial per-electrode basis.

Second, differential ECoG HG-C spatial and temporal structure was demonstrated between the dorsal and ventral 8x8 electrode grids, where principally, neural speech activity was observed in the ventral and not in the dorsal grid. The spatial correlation decay constant was 4-5x longer in ventral compared to dorsal grid.

Third, differential HG-C spatial organization was itself further differentiated within the same ventral grid across temporal phases of the standard trial. Specifically, in the post-stimulus period, the correlation vs distance curve shifted up and its spatial decay constant increased compared to the stimulus and pre-stimulus periods.

Last, ECoG HG correlation structure was shown to display robustness with respect to simulated electrode degradation. This observation is of particular importance as BCI devices trend towards higher electrode densities and electrode counts. Some thought experiments illustrate the significance. Suppose the underlying neurophysiology is such that the signal of interest is indeed a localized (or localizable) phenomenon. This appears to be what the Angrick, et al., [1] models leveraged (shown through saliency analysis) and that this study largely confirmed through marginal effects analysis. That is, within an 8x8 grid, a single electrode (#21) or small subset of 2-5 electrodes “excessively” contributed to HGLP NVAD model performance. In such a case,

BCI with higher densities certainly is robust to individual electrode disconnections - *under normal conditions*. However, were a biological process to occur in or around this key region and significantly impact impedance or other electrode properties, then the benefits of high density, high count BCI are impacted.

Suppose instead the underlying neurophysiology is distributed. Here, we do not mean distributed at brain-scale or even brain regional-scale. The specific meaning of distributed here remains on the order of single to tens of millimeters such as accessible within the spatial dimensions sampled in the Angrick, et al., [1] study. This is the configuration where higher densities and higher electrode counts have clear and robust benefit. Here, BCI attains the robustness required for long-term stable performance through oversampling of the signal.

Next Steps

This study tested and confirmed the hypothesis that ECoG HG correlation structure has desirable properties required for long-term BCI applications. Next steps should focus on addressing the identified weaknesses. First, will the proposed method generalize to other tasks and datasets? The Angrick, et al., [1] data offers an immediate opportunity by way of the companion keyword reading task (KRT) recorded in each session. Initial experiments with training on the KRT dataset alone (N=1,920 trials) were encouraging. However, it was observed that for the keyword "Enter" and the "long-tailed" keywords, "Left" and "Right", both HGLP and HG-C often marked the neural "silent" period between syllables (or phonemes) as not-speech or struggled to label the speech tail, whereas the automated labelling method labelled speech. This held down F1 scores (below those of SRT and in the range 0.80-0.85 typically). Operationally, filtering of the NVAD output labels could "fix" such issues. In this study, we eschew such post-processing steps and focus on the performance attained using the intrinsic signal feature properties.

Second, to address closed-loop concerns and to equalize trainable model parameter counts, it may be attractive to consider the application of unsupervised methods. These include VQ-VAE, which would eliminate the need for the input Linear Network ahead of the LSTM [19]. Angrick, et al., have also reported an approach to real-time detection of spoken speech from unlabeled ECoG signals [20].

Last, by design, this study focused on ECoG HG as the source signal. Further steps could consider partial correlation analysis of HG, additional ECoG bands, and more complex methods to model distributed spatiotemporal signals [16].

Historical Note

The present findings recall ECoG and EEG spatial and time series processing methods pioneered by Walter J. Freeman and his students at UC Berkeley during the so-called pre-Utah Age. Freeman, et al., recorded and analyzed ECoG using high-density high-throughput 8 x 8 electrode arrays (3.5mm x 3.5mm, 250 um diameter tungsten or stainless-steel platinum-tipped electrodes, 500 um center-center) chronically implanted epidurally on the lateral aspect of rabbit olfactory bulb in an animal behavioral model [21-24].

Code Availability

All source code supporting this study will be made publicly available on https:://github.com/kgrajski/ecog_stability upon acceptance of the manuscript. Reasonable requests for earlier release to foster collaboration(s) will be entertained.

Data Availability

The raw time series samples of neural data and anonymized speech audio were previously made publicly available at http:/www.osf.io/49rt7/ by the original study team [1]. The present study produced: feature data comprised of ECoG HGLP and correlation matrix-based features derived from HGLP (HG-C) as individual trials stored as .csv files; a metadata database for task, session, block, and trial metadata; and a database of experimental results. These may be made available upon reasonable request to foster collaboration.

References

1. Angrick, M., *et al.* Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS. *Sci. Rep.* **14**, 9617. (2024). https://doi.org/10.1038/s41598-024-60277-2.

2. Crone, N. E., et al. Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology*. **57**, 2045-2053. (2001).
3. Leuthardt, E., et al. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. *Front. Hum. Neurosci.* https://doi.org/10.3389/fnhum.2012.00099. (2012).
4. Bouchard, K. E., et al. Functional organization of human sensorimotor cortex for speech articulation. *Nature*. 495. 327-332. (2013).
5. Chartier, J., et al. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron*. **98**, 1042-1054.e4. (2018).
6. Ramsey, N. F., et al. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. **180**, 301-311. (2018).
7. Angrick, M, et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. *J. Neural Eng.* 16(3), 03619. (2019).
8. Mugler, E. M., et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. J. Neural Eng. 11(3):035015. (2014).
9. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* 385, 217-227 (2021).
10. Branco, M. P., et al. Nine decades of electrocorticography: A comparison between epidural and subdural recordings. *Eur. J. Neurosci.* **57**, 1260-1288. (2023).
11. Duraivel, S., et al. High-resolution neural recordings improve the accuracy of speech decoding. *Nature Comms.* 14:6938, 1-16. (2023).
12. Metzger, S. L., et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*. **620**. 1037-1046. (2023).
13. Willett, F. R., et al. A high-performance speech neuroprosthesis. *Nature*. **620**, 1031-1036. (2023).
14. Chen, X., et al. A neural speech decoding framework leveraging deep learning and speech synthesis. *Nature Mach. Intell*. **6**, 467-480. (2024).
15. Wairagkar, et al. An instantaneous voice-synthesis neuroprosthesis. *Nature*. **633**, 151-158. (2025).
16. Chao, Z.C., Nagasaka, Y., Fujii, N. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkeys. *Front. Neuroeng.* **3**:3. (2010).

17. Sillay K. A., et al. Long-term measurement of impedance in chronically implanted depth and subdural electrodes during responsive neurostimulation in humans. *Brain Stimul.* Sep;6(5):718-26. doi: 10.1016/j.brs.2013.02.001. (2013).
18. Wyse-Sookoo K., et al. Stability of ECoG high gamma signals during speech and implications for a speech BCI system in an individual with ALS: A year-long longitudinal study. *J Neural Eng.* 2024 Jul 12;21(4):10.1088/1741-2552/ad5c02.
19. Grajski, K. A., Speech BCI with VQVAE. https://github.com/kgrajski/speechBCI. (2025).
20. Angrick, et al. Real-time detection of spoken speech from unlabeled ECoG signals: A pilot study with an ALS participant. medRxiv preprint doi: https://doi.org/10.1101/2024.09.18.24313755. (2024).
21. Bressler, S.L. and Freeman, W.J. Frequency analysis of olfactory system EEG in cat, rabbit, and rat. *Electroencephalography and Clin. Neurophys. 50*(1), 19-24. (1980).
22. Grajski, K. A., et al. Classification of EEG spatial patterns with a tree-structured methodology: CART. *IEEE Trans Biomed Eng.* 33(12):1076-86. doi: 10.1109/TBME.1986.325684. PMID: 3817838. (1986).
23. Freeman, W. J. and Grajski, K. A. Relation of olfactory EEG to behavior: Factor analysis. *Behav. Neurosci.* 1987 Dec;101(6):766-77. doi: 10.1037//0735-7044.101.6.766. PMID: 3426793. (1987).
24. Freeman, W. J., & Baird, B. Relation of olfactory EEG to behavior: Spatial analysis. *Behavioral Neuroscience, 101*(3), 393-408. https://doi.org/10.1037/0735-7044.101.3.393. (1987).

Author Contributions

KAG conceived of and designed the experiments to assess ECoG HG-C reported in this manuscript, implemented all code underlying the processing methods and results, analyzed the data, and wrote the manuscript.

Acknowledgements

Acknowledgements first and foremost to the ALS patient whose enrollment in the clinical trial (NCT03567213) resulted in the creation of this study dataset. Acknowledgements and

appreciation to the Nathan E. Crone Lab and Miguel Angrick at the Department of Neurology at the Johns Hopkins University School of Medicine who elected to make even the raw time series study dataset available as open source.

Competing Interest

The author (KAG) declares that there are no competing interests.

Revision History

Version	Date	Changes
0.0	22 August 2025	Initial draft. Subject to change.

Tables

NVAD Standardized Trial Speech Statistics Syllable Repetition Task			
Session	**Trials With Speech (N, %)**	**Speech Latency (s)**	**Speech Duration (s)**
All Trials (N=780)	777, 99%		
Mean, Std. Dev.		0.347, 0.162	0.721, 0.123
Range		0.032 - 1.616	0.132 - 1.164

Table 1. NVAD standardized trial speech statistics. In the Syllable Repetition Task, the participant is instructed to repeat the audio prompt immediately upon termination of the audio prompt. The study dataset consists of 60 recorded trials from each of 13 sessions (N=780 trials). An automated voice activity labelling method (followed by visual inspection) is used to create "ground truth" speech activity. Ninety-nine percent of trials contain speech. There are three trials with no speech and five trials with speech cutoff at end of trial. These trials remained in the study dataset. At the trial frame level (224 per trial; 174,720 total) approximately 25% of frames are labelled as speech and 75% as non-speech.

NVAD Model F1 Score Comparison High-gamma Log Power (HGLP) vs High-gamma Correlation (HG-C) Input Features				
Session	F1 (HGLP)	F1 (HG-C)	Δ F1	AUC-PR (HG-C)
Test	0.87	0.90	+0.03	0.96
Validation	0.86	0.91	+0.05	0.97
Online 1	0.78	0.85	+0.07	0.94
Online 2	0.76	0.81	+0.05	0.90
Online 3	0.67	0.82	+0.15	0.89

Table 2. NVAD model F1 score performance comparison. High-gamma Log Power (HGLP) results shown are in line with previously published reports for this dataset [2]. Correlation-based feature (HG-C) F1 score performance is improved over HGLP for all sessions. Under the standardized trial data, speech comprises 25% of windowed data frames, a slight imbalance, and for that reason the area under curve (AUC) precision-recall (PR) statistic is quoted for the HG-C models. Test session F1 score is used as the figure-of-merit during hyperparameter grid search. For each model the Test session output probability threshold for speech vs non-speech frame labelling is optimized and then fixed for follow-on sessions. The results shown are purely based on NVAD model output. There is no post-processing or filtering or padding of predicted labels. There is no session-level normalization or calibration or group normalization. The only normalization is per-electrode per-trial Z-score normalization for HGLP and Fisher Z-transform for HG-C.

Figures

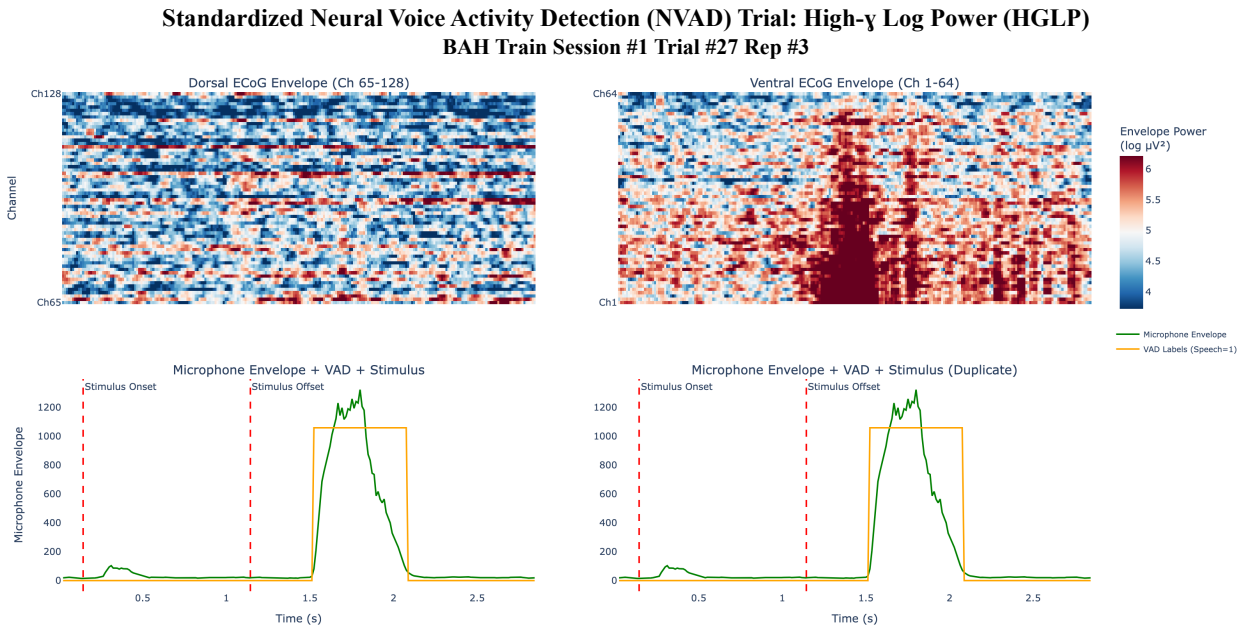

Figure 1. Standardized syllable repetition task neural voice activity detection trial example (*BAH*). A standardized trial spans the time 144 msec pre-stimulus, continuing through the 1.0 second stimulus delivery period, and ending 1728 msec post-stimulus. In each of thirteen (13) sessions, the patient completes a syllable repetition task comprised of randomized trials (N=60) 2.5-3.5 seconds duration resulting in five (5) repetitions of twelve (12) consonant-vowel audio stimuli played through a loudspeaker. ***Top row***: *y-axis* dorsal 8x8 electrode grid (left) and ventral 8x8 electrode grid (right) windowed high-ɣ log power (HGLP) envelope; *x-axis* is time-aligned with bottom row panels. ***Bottom row***: (*left and right repeated for ease of comparison with spatial data shown on top row):* stimulus onset and offset (*dotted red lines*); microphone sample data windowed average (48 msec; 12 msec stride) (*solid green line*); and automatically labelled (see Methods) speech ON/OFF (0/1) target label (*orange line*). A key observation is that one can "see" neural correlates of voice activity.

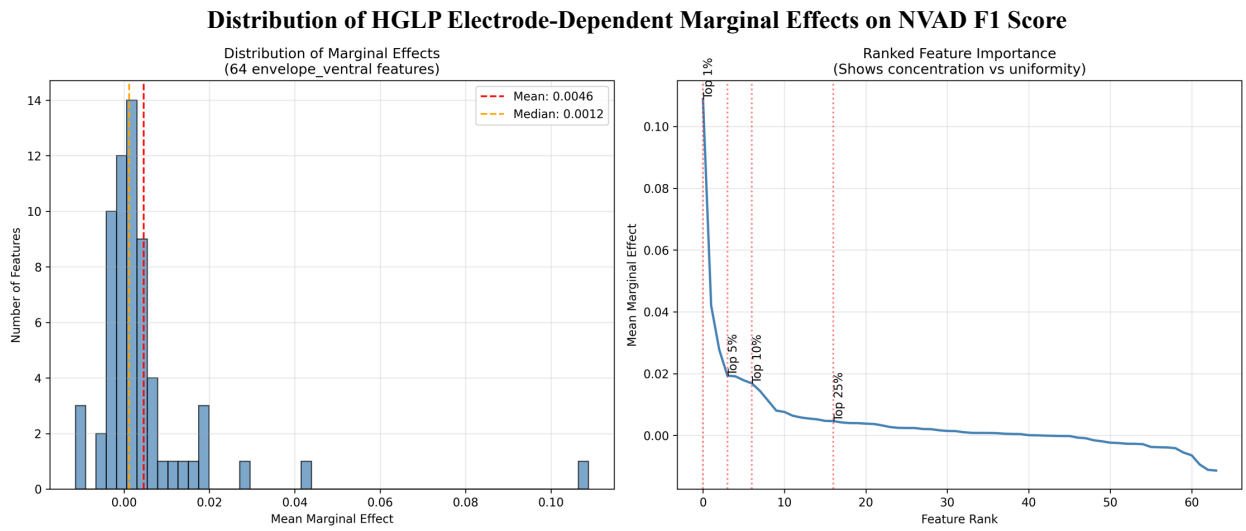

Figure 2. Distribution of empirical marginal effect values of individual electrode HGLP features (N=64) on test session F1 score performance of the NVAD model. ***Left***: histogram of values; mean and median score (*dotted lines*). ***Right***: ranked values; percentiles (*dotted red lines*). A key observation is that the distribution of empirical marginal effect scores is non-uniform with a top 5th-10th percentile of "excessively" important electrodes (relatively high positive value). As a control, the NVAD model training procedure is repeated with permuted target (non)speech labels which yield the constant zero for individual electrode marginal effect (as expected).

Spatial Distribution of HGLP Electrode-Dependent Marginal Effects on NVAD F1 Score (Ventral 8x8 Electrode Grid)

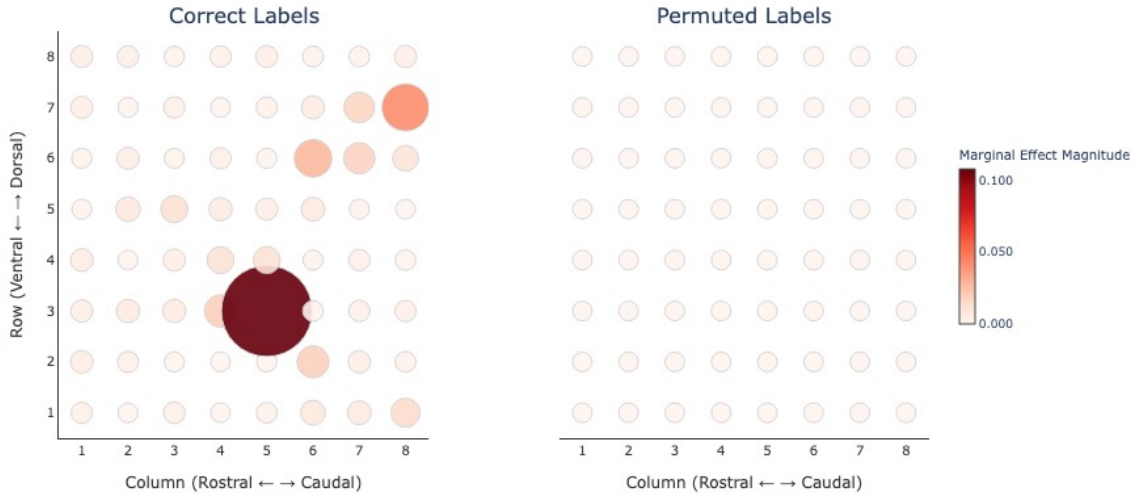

Figure 3. Spatial distribution and weighting of empirical marginal effect values of individual electrode HGLP features (N=64) on test session F1 score performance of the NVAD model. ***Left***: heatmap of values for ventral 8x8 grid when NVAD model is trained with correct target labels; diameter and color proportional to marginal effect magnitude. ***Right***: heatmap of values for ventral 8x8 grid when NVAD model is trained with permuted target labels. A key observation is that these results - using independent analysis and complementary methods and data - almost exactly reproduce Angrick, et al., saliency analysis results, particularly regarding electrode 21. (See Figure 3 in ref. [2]). These results confirm that we properly handle the data and have set the stage for comparative analysis of correlation-based features as a basis for syllable repetition task NVAD.

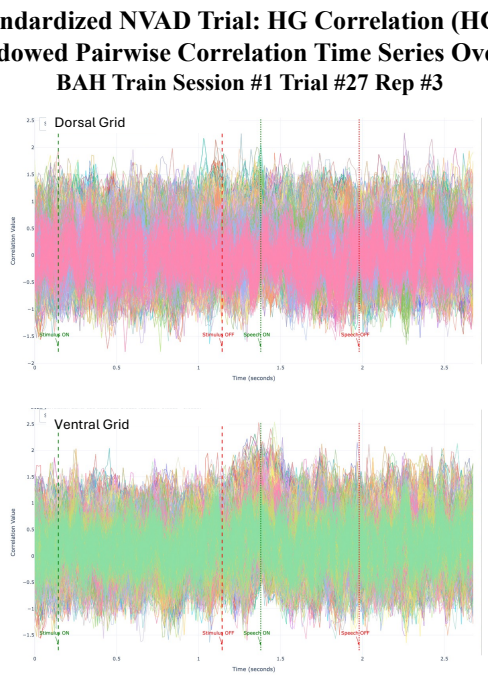

Figure 4. Standardized neural voice activity detection trial (*same as Figure 1 trial*) shows overlays of windowed pairwise correlation time series. ***Top***: overlay of pairwise correlation time series *within* dorsal 8x8 grid (N=2016 pairs). ***Bottom***: overlay of pairwise correlation time series *within* ventral 8x8 grid (N=2016 pairs). The x-axis shows time from trial start in seconds. The y-axis shows Fisher Z-transformed Pearson correlation coefficient of windowed high-ɣ log power envelope time series. A key observation is the increase in correlation values immediately post-stimulus and preceding speech onset largely, if not exclusively local to the ventral grid.

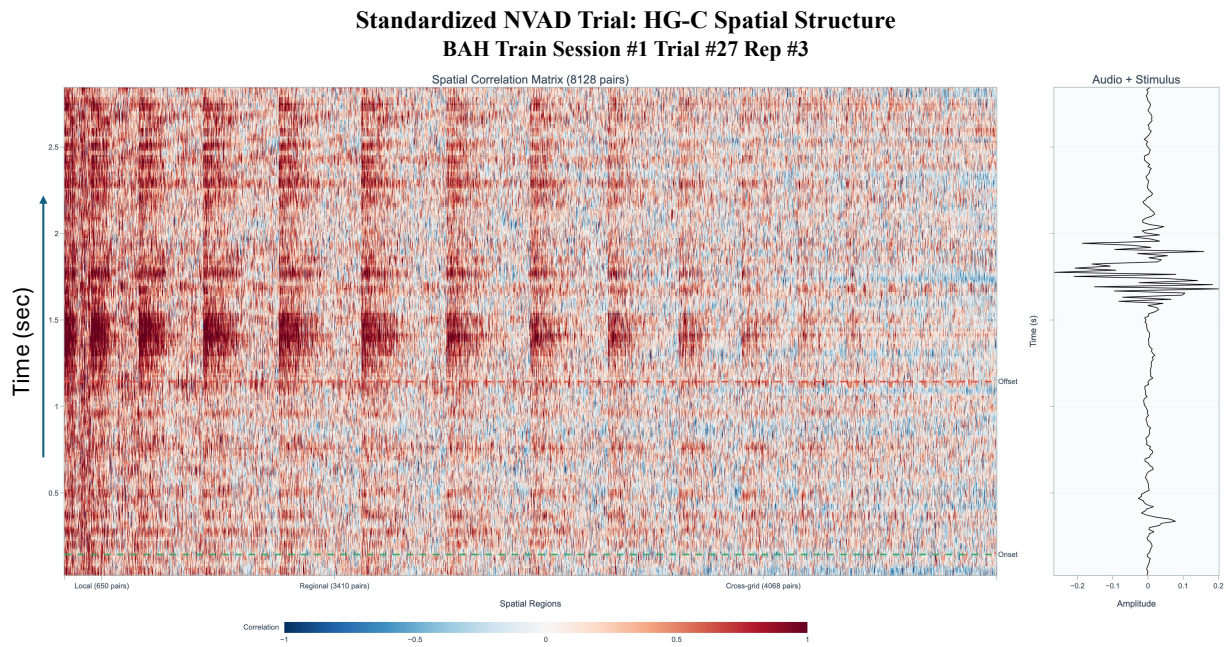

Figure 5. Standardized neural voice activity detection trial (*same as Figure 1 trial*) shows spatial correlation structure (HG-C). ***Left***: y-axis shows time from trial start (*bottom left*) to 1768 msec post stimulus (*top left*); x-axis shows Pearson correlation coefficient of high-ɣ log power envelope windowed time series (48 msec; 12 msec stride). ***Right***: time-aligned (*y-axis*) windowed audio average signal (*x-axis*) with spatial correlation structure shown at left. A key observation is the spatial structure evident at both local (near-neighbors), regional (distant within-grid), and to a far lesser extent (visually) between grid distance scales immediately preceding, during, and following the patient's consonant-vowel voicing.

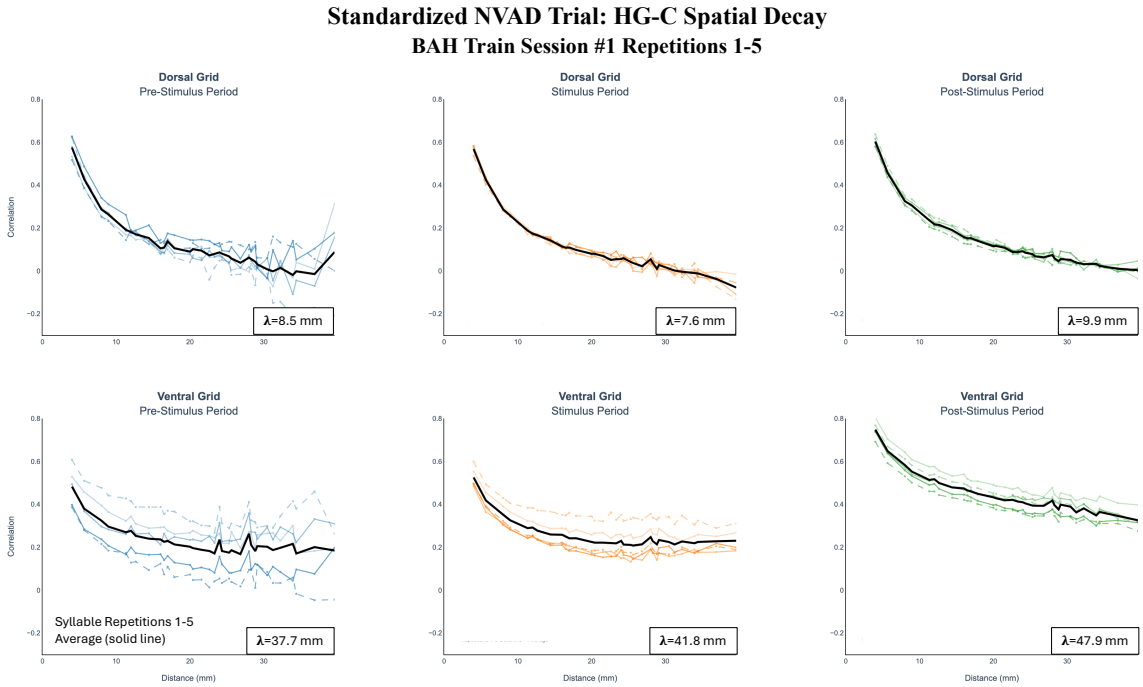

Figure 6. Standardized neural voice activity detection trial time periods showing empirically estimated HG-C spatial decay constants. Each plot shows results for each of the five utterances of the given stimulus (BAH) recorded during one session. The y-axis shows Pearson correlation coefficients value. The x-axis shows distance between electrode pairs. The spatial decay constant is estimated by fitting an exponential function. See Methods. ***Top-row***: (*left*) pre-stimulus period; (*middle*) stimulus period; and (*right*) post-stimulus period. ***Bottom-row***: same as top row but for ventral grid. The dorsal and ventral grid analyses are conducted independently as they are physically embodied as a pair of 8x8 electrodes offset in distance and rotated one with respect to the other. The results shown here are typical across stimulus trials and sessions.

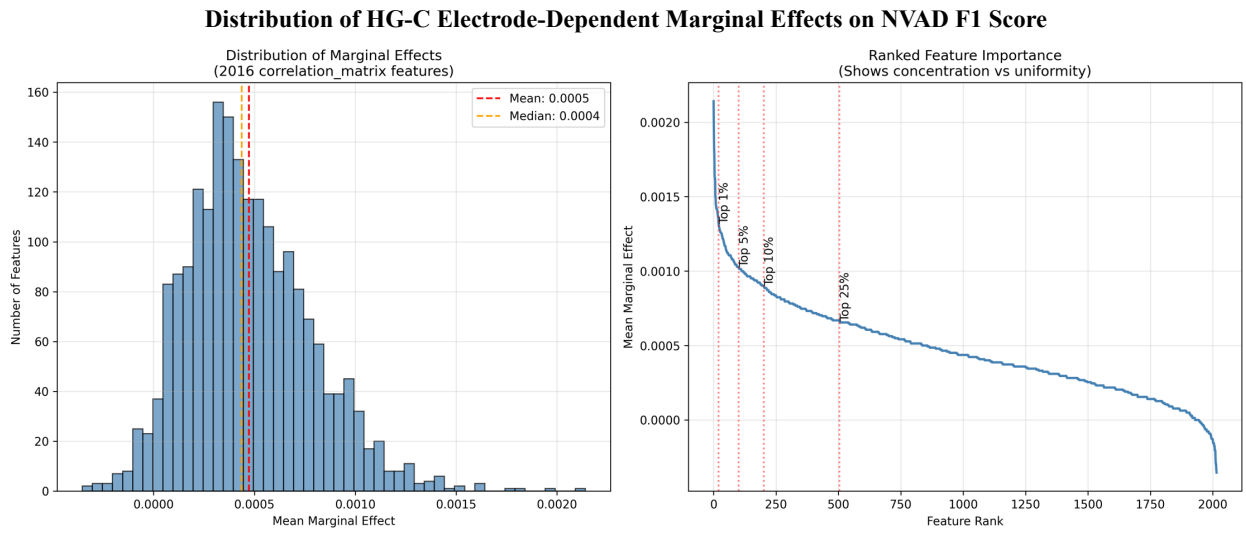

Figure 7. Distribution of empirical marginal effect values of individual ventral 8x8 electrode grid pairwise correlation HG-C features (N=2,016) on test session F1 score performance of the NVAD model. ***Left***: histogram of values; mean and median score (*dotted red lines*). ***Right***: ranked values; percentiles (*dotted red lines*). A key observation is that the distribution of empirical marginal effect scores is non-uniform with a top 5th-10th percentile of very "important" electrodes (relatively high positive value). As a control, the NVAD model training procedure is repeated with permuted target (non)speech labels which yields the constant zero for individual electrode marginal effect (as expected).

Pair-wise Correlation Marginal Effects on NVAD F1 Score

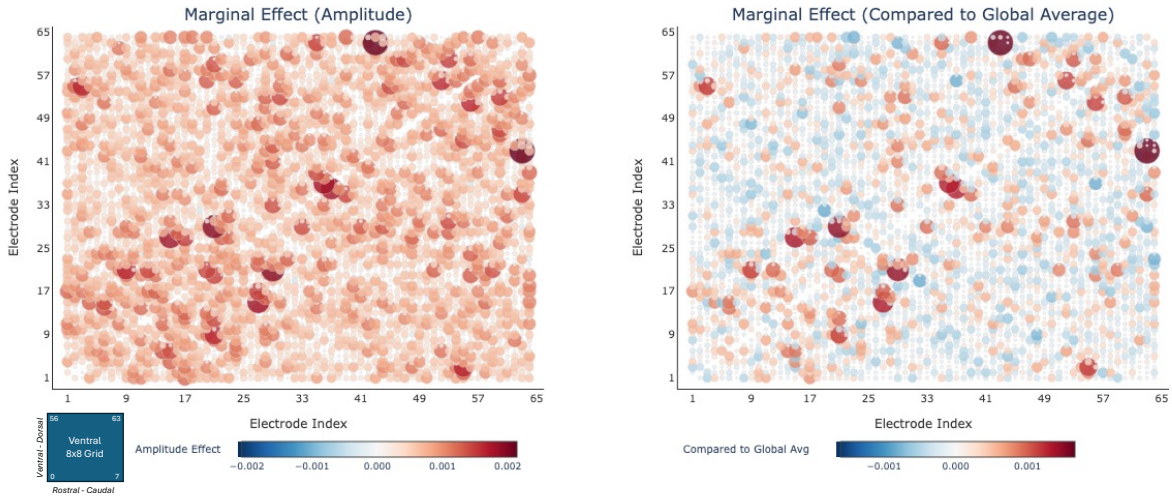

Figure 8. Spatial distribution of empirical marginal effect of individual HG-C features on test session F1 score performance of the HG-C NVAD model. Each subplot is in the form of a symmetric correlation matrix. Each point in the subplot represents a pairwise correlation computed in the ventral 8x8 electrode grid in a single typical trial. The correlation matrix is rotated to maintain the convention in the Angrick, et al., [1] study data that electrode #1 is in the rostral- and ventral-most position in the ventral electrode grid. ***Left***: Pair-wise correlation marginal effects when NVAD model is trained with correct target labels. Diameter and color proportional to marginal effect *magnitude*. ***Right***: The same pair-wise correlation marginal effects plotted as values relative to the global average correlation marginal effect. Diameter and color proportional to relative marginal effect *amplitude*. A key visual observation is the multiplicity of high marginal effect locations.

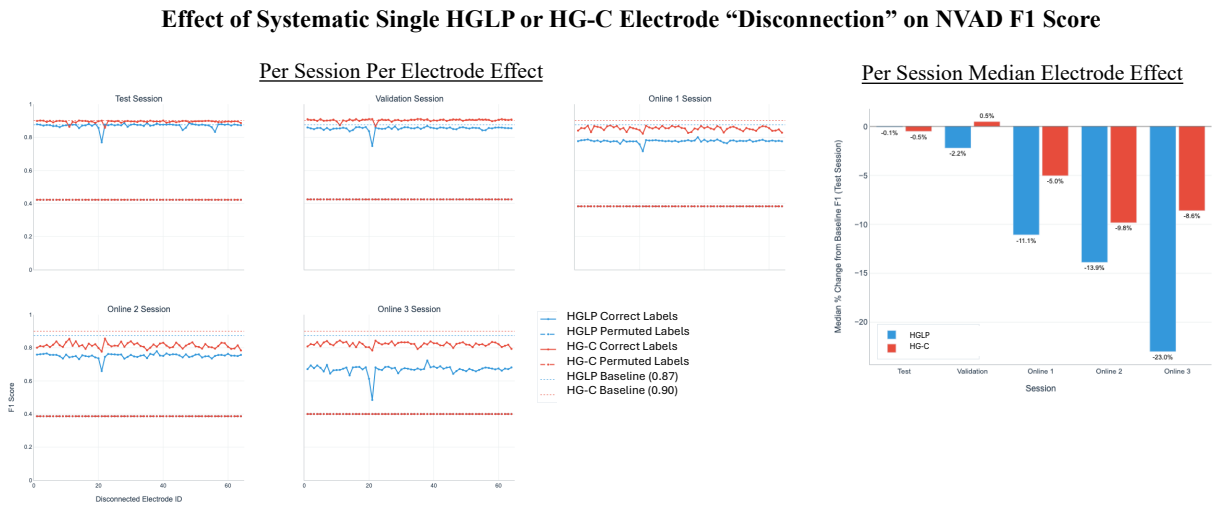

Figure 9. Effect of simulated systematic disconnection of individual grid electrodes on session NVAD F1 score. Disconnection is simulated by setting all HGLP values or all pairwise HG-C values involving that electrode to zero and then evaluating NVAD performance. **Per Session Per Electrode Effect Panel *(left).*** F1 score is measured per session under several conditions. Sessions are progressively later in time from the training sessions: Test (*top left*); Validation (*top middle*); Online 1 (*top right);* Online 2 (*bottom left*); and Online 3 (*bottom middle*). The plotted conditions are: HGLP NVAD with correct and with permuted labels; HG-C NVAD with correct and with permuted labels; and the F1 score obtained on Test set by the HGLP and HG-C NVAD models. The F1 score is plotted on the *y-axis* and the electrode ID number disconnected is shown on the *x-axis*. **Per Session Median Electrode Effect Panel (*right*).** Summary plot showing the median effect of systematic electrode disconnection across sessions and by feature type, HGLP, or HG-C. The y-axis plots the median percentage change in F1 score compared to the baseline F1 score, where baseline F1 score is that achieved on the Test session. The main observation is that under the conditions of the syllable repetition task in this dataset, with no session-based normalization and with no dedicated calibration trials, HG-C trained NVAD models performed nearly ~2.5X better (degraded less) than HGLP NVAD models in the session farthest in time from training data.

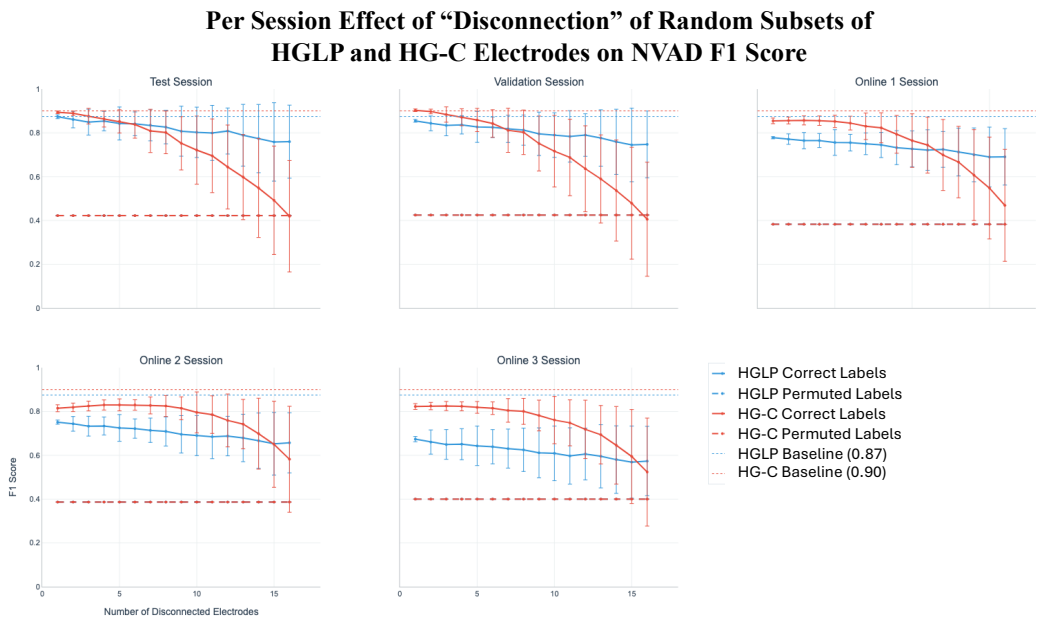

Figure 10. Effect of simulated disconnection of random subsets of size N of grid electrodes on session NVAD F1 score. Electrode disconnection is simulated by setting all HGLP values or all pairwise HG-C values involving that electrode to zero and then evaluating NVAD performance. F1 score is measured per session under several conditions. Sessions are progressively later in time from the training sessions: Test (*top left*); Validation (*top middle*); Online 1 (*top right*); Online 2 (*bottom left*); and Online 3 (*bottom middle*). Each plot overlays results for several conditions. These are: HGLP NVAD with correct and with permuted labels; HG-C NVAD with correct and with permuted labels; and the F1 score obtained by the NVAD model on Test Set. (***y-axis***). The F1 score is plotted on the *y-axis*. The error bars are computed as the standard deviation over a set of S samples for a given value of N, where the S is adaptively set in proportion to the number of possible combinations, "64 choose N", but restricted to the range 10 to 200 (where applicable). (***x-axis***). The number of electrodes disconnected per sample is shown on the *x-axis*. The main observation is that under the conditions of the syllable repetition task in this dataset, with no session-based normalization and with no dedicated calibration trials, the HG-C trained NVAD model performance was more stable (lower standard deviation in performance up to about 12% of electrodes disconnected) than HGLP NVAD models in the temporally distant "online" trials.