2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

Contributions of Stereotactic EEG Electrodes in Grey and White Matter to Speech Activity Detection

P. Z. Soroush[1], C. Herff[2], S. Ries[3], J. J. Shih[4], T. Schultz[5], and D. J. Krusienski[1]

***Abstract*— Recent studies have shown it is possible to decode and synthesize speech directly using brain activity recorded from implanted electrodes. While this activity has been extensively examined using electrocorticographic (ECoG) recordings from cortical surface grey matter, stereotactic electroencephalography (sEEG) provides comparatively broader coverage and access to deeper brain structures including both grey and white matter. The present study examines the relative and joint contributions of grey and white matter electrodes for speech activity detection in a brain-computer interface.**

I. INTRODUCTION

Invasive recordings of brain activity using intracranial arrays, such as electrocorticography (ECoG) [1], have shown promise in the design of brain-computer interfaces (BCIs) for a variety of applications including speech decoding and synthesis [2], [3], [4], [5], [6], [7]. Because ECoG only samples grey matter from the cortical surface, there has been little investigation into the potential contributions to BCI decoding of recordings from white matter, which makes up approximately 50% of human brain volume. Additionally, information from white matter recordings has been reported to be distinctive from that of grey matter [8]. The growing popularity of stereotactic electroencephalography (sEEG) [9] for clinical applications provides the opportunity to examine neural activity from broader brain regions and deeper structures, including both grey and white matter.

A few recent studies have investigated the role of sEEG recordings from white matter in the design of BCIs. It has been shown that including sEEG channels from both grey and white matter can help distinguish between various upper limb movements and rest, or between different movement types [10]. Other studies have shown that sEEG channels in both grey and white matter contribute to models for speech activity detection and speech production for BCIs [7], [11], [12]. While these studies highlight the potential contributions of both grey and white matter to speech production, a comprehensive characterization of the grey and white matter channels has yet to be conducted.

The present study investigates channels from both grey and white matter, individually and jointly, in the design of a speech activity detection model to discriminate *speech* from *non-speech* for a BCI.

*This work was supported by NSF 2011595/1608140 and BMBF 01GQ1602.

[1]P. Z. Soroush and D. J. Krusienski are with Virginia Commonwealth University, Richmond, VA, USA zanganehp@vcu.edu
[2]C. Herff is with the University of Maastricht, Maastricht, Netherlands
[3]S. Ries is with San Diego State University, San Diego, CA, USA
[4]J. J. Shih is with UCSD Health, San Diego, CA, USA
[5]T. Schultz is with the University of Bremen, Bremen, Germany

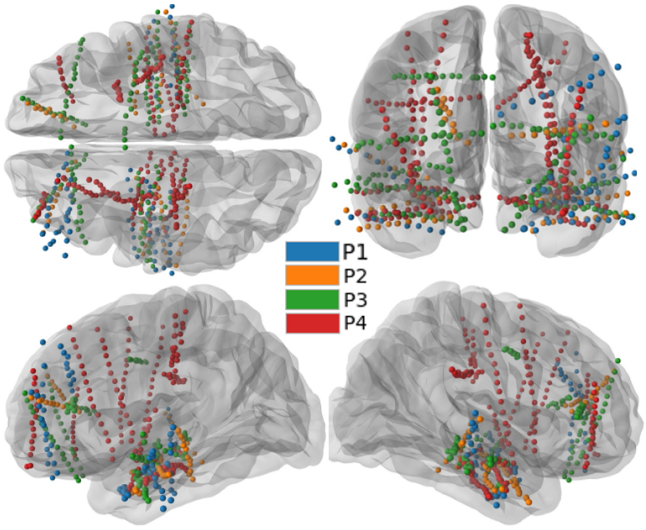

Fig. 1: The sEEG depth electrode locations for the 4 participants from different perspectives using an averaged brain model.

II. METHODOLOGY

A. Participants and Data Collection

sEEG data were collected from 4 native English-speaking participants being monitored as part of treatment for intractable epilepsy at UCSD Health. The study design was approved by the Institutional Review Boards of Virginia Commonwealth University and UCSD Health. The locations of sEEG electrodes were determined solely based on the participants' clinical needs. A subset of the implanted electrodes for each participant was determined to be in or adjacent to brain regions associated with speech and language processing. Fig. 1 shows the sEEG electrode locations for the 4 participants, with electrode (channel) counts provided in Table I. Anatomical location of the channels, including brain region and binary location (white or grey matter), were identified using the FreeSurfer software package [13].

For the experiment, the participants were presented with a sentence displayed on a computer monitor and simultaneously narrated via computer speakers. When prompted, the participant was instructed to speak the sentence audibly while the acoustic speech and sEEG signals were simultaneously recorded. The participant was subsequently prompted to silently articulate (i.e., mouth) and imagine speaking without articulating, respectively, for the same sentences - although data from these components of the task were not used for the

present analysis. This structure was repeated for 50 unique Harvard sentences, which are phonetically-balanced based on conversational English [14].

The sEEG electrodes were referenced to a pair of subdermal needle electrodes in the scalp and digitized at 1,024 Hz. The audio signal recorded via an external microphone was digitized at 44,100 Hz. The data from the audible speech portions of the task were used to extract speech and non-speech segments from the audio recordings.

B. Labeling the Audio Files (Speech vs. Non-speech)

The recorded speech was manually transcribed using the Wavesurfer software package [15] for a separate analysis, but was found useful to provide precise labeling of the speech and non-speech segments. This was accomplished by shifting a 10 ms non-overlapping frame across the audio recording, with the resulting timings from the transcription word boundaries being used as the frame label. Each frame was labeled as *speech* if at least half of the frame length overlapped with a transcribed word and as *non-speech* otherwise. Because there were relatively few and short non-speech segments detected during the continuous vocalization of the sentence utterances, each 4-second interval encompassing the entire utterance was simply labeled as *speech* between the onset and offset of the entire utterance as determined by the transcription, and as *non-speech* before and after the onset and offset, respectively.

C. Data Preprocessing and Feature Extraction

All sEEG data were visually inspected and noisy or anomalous channels and segments were excluded from the analysis. The resulting sEEG channels were re-referenced using the Laplacian method [16], [17]. The sEEG channels were then normalized by removing the mean and scaled to unit variance. To ensure that perceptual information is not directly contributing to the decoding models, channels in the vicinity of the auditory cortex and belts exhibiting large correlations between gamma-band activity and the microphone recordings were excluded from the analysis. The number of channels excluded for each particpant is provided in Table I.

The sEEG channels were then resampled to 512 Hz and features were extracted in the alpha frequency band (8-12 Hz), which has been previously shown to be promising for distinguishing movement from rest [10] and *speech* from *non-speech* [11]. The sEEG channels over a 510-ms window around each audio frame (corresponding to 500 ms before the frame to the end of the frame) were zero-phase filtered from 8-12 Hz using a sixth-order Butterworth filter. This window length was chosen to capture sufficient cycles to reliably estimate the alpha band activity (i.e., 4 cycles $\times$ 0.125 s/cycle). Next, the narrow-band amplitude envelope of each sEEG channel was computed using the Hilbert transform. Finally, for each channel, the alpha-band features were computed every 10 ms as the natural logarithm of the resulting envelope energy over 210 ms, representing 10 ms overlapping the audio frame and 200 ms prior to the frame to emulate a causal design. The features from each included channel were concatenated to form the feature vector (# channels $\times$ 21 features) for the decoding models.

D. Model Training and Evaluation

1) Logistic Regression Model: For evaluating the features, a logistic regression model with L1 regularization was designed. A Proximal Adagrad Optimizer with SoftMax function was selected for training the model. This model has been shown to be effective for evaluating individual features and providing a convenient interpretation of the feature contributions [11].

To prevent training bias, the training data were normalized to zero mean and unit variance, and the same normalization parameters were applied to the test data. To establish the chance-level classification, a randomization test was performed where the labels of all trials were randomly shuffled and the 10-fold cross-validation process was repeated for 1000 separate randomization of the labels.

2) Single-channel Analysis: For each participant, using a 10-fold non-shuffled cross-validation analysis, performance of aforementioned decoding model was evaluated for each individual channel. This single-channel analysis was performed to compare the relative contributions of each channel.

Three groups of channels were selected: channels identified as located in white matter (WM), grey matter (GM), and the combination of grey and white matter channels (CM). To select the most relevant channels for each group, a threshold of the mean plus one standard deviation of the average balanced accuracy in the 10-fold cross-validation analysis was computed. Channels with an average balanced accuracy above the threshold were selected for inclusion in the multi-channel analysis.

3) Multi-channel Analysis: For each participant and group of selected channels (WM, GM, and CM), using a 10-fold non-shuffled cross-validation analysis, concatenated features of the channels were evaluated using the aforementioned decoding model. The final weights of the 10 decoding models of 10-fold cross-validation analysis were extracted to compute the averaged spatiotemporal patterns to compare the contribution of each feature in grey or white matters to the final model [18]. These patterns provide a convenient interpretation of individual feature's spatial and temporal contributions to speech activity detection.

III. Results

Fig. 2 illustrates violin plots of the distributions of averaged balanced accuracy of the 10-fold cross validation analysis of the models trained in the single-channel analysis. The black dots on Fig. 2 represent the channels that yielded an averaged balanced accuracy above the threshold of the group and were selected for the multi-channel analysis.

Fig. 3 shows the grey and white matter channels selected by the single channel analysis. Channels from a wide range of cortical and sub-cortical brain regions, including inferior and middle frontal gyrus, superior temporal gyrus, and sulcus were selected for the multi-channel analysis. It can be seen that these channels lie in both right and left hemispheres,

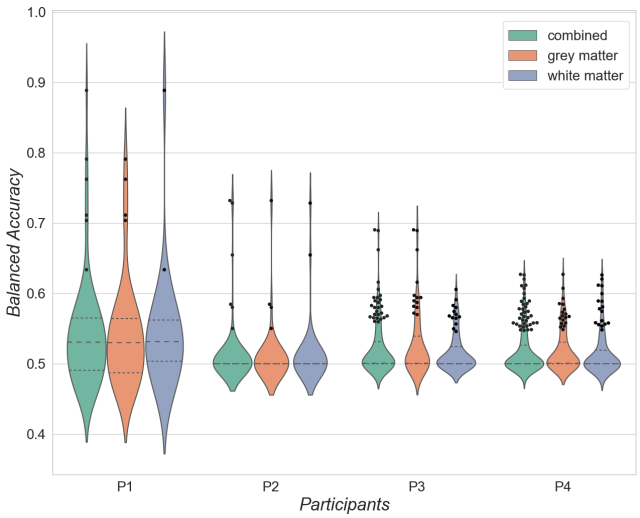

Fig. 2: Distributions of the cross-validated classification performance of decoding models for channels in grey matter, white matter, or combined. The black dots represent the selected channels for each group.

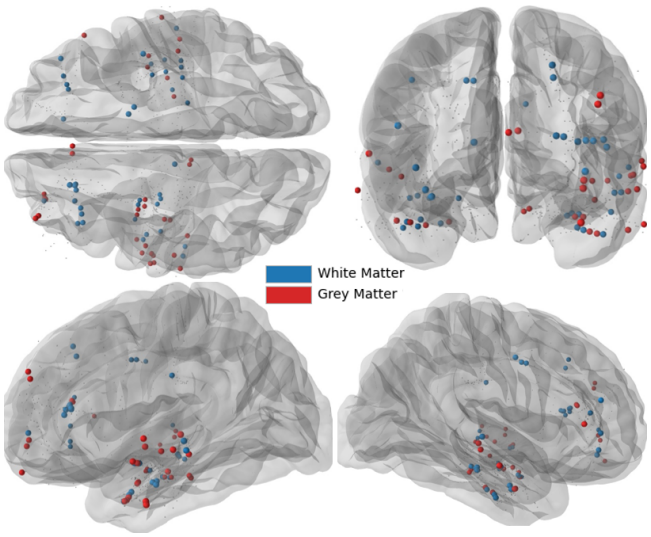

Fig. 3: The selected grey and white matter channels across all 4 participants.

which supports recent studies that have reported speech-related activity in both hemispheres [11], [12], [19], [20].

The number of selected channels utilized by each decoding model is provided in Table I. Fig. 4 shows the averaged balanced accuracy of the decoding models in the 10-fold cross-validation analysis for each group. All models performed significantly better than chance level ($p < 0.0001$, Bonferroni-Holm corrected Wilcoxon signed-rank test). For all four participants and the average across participants, the results of the CM group were significantly larger than the results of both WM and GM groups ($p < 0.01$). No significant difference was observed between the WM and GM model performance ($p = 0.0913$).

Fig. 5 shows the averaged spatiotemporal patterns derived

Participant	# Recorded	# Excluded	# Model Selected
P1	70	4	4 (GM), 2 (WM), 6 (CM)
P2	69	4	4 (GM), 2 (WM), 6 (CM)
P3	173	13	13 (GM), 13 (WM), 27 (CM)
P4	227	6	18 (GM), 17 (WM), 37 (CM)

TABLE I: Numbers of channels for each participant. The first column lists the total number of channels recorded during the experiment, the second column lists the number of channels excluded as located in auditory regions, and the third column lists the number of channels selected for each decoding model.

from the decoding model weights for the WM, GM, and CM groups of an example participant (P2) [18]. The feature contributions tend to be more prominent near speech onset (i.e., 0 ms). It is also observed that the relevant channels in the CM group are determined to be a subset of channels from both the WM and GM groups. Channels from GM group had a relatively lower impact in the CM models compared to channels from WM group. This is likely due to the nature of logistic regression with L1 regularization, which shrinks a feature's weight to zero based on relative importance. The WM features with the largest contribution to the CM model are consistent for both the WM and CM models. For the other participants, it was observed that channels from either WM or GM were more prominent in the CM model.

IV. Conclusion

The results of this pilot study show that an effective speech activity detection model can be created using alpha band activity from sEEG recordings. The results indicate that constructing the model using combination of grey and white matter channels is more effective than using exclusively grey or white matter channels. Moreover, on average, models constructed from white matter channels performed equivalently to models constructed from grey matter channels. These findings highlight the potential of features beyond the cortex, particularly from under-explored white matter, for improving the performance of intracranial speech BCIs. Further work is needed to explore the specific contributions of functional areas and networks with respect to speech production and perception, as well as whether these results can be generalized across a larger participant pool.

References

[1] G. Schalk and E. C. Leuthardt, "Brain-computer interfaces using electrocorticographic signals," *IEEE reviews in biomedical engineering*, vol. 4, pp. 140–154, 2011.

[2] C. Herff, D. Heger, A. De Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, "Brain-to-text: decoding spoken phrases from phone representations in the brain," *Frontiers in neuroscience*, vol. 9, p. 217, 2015.

[3] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," *Journal of neural engineering*, vol. 16, no. 3, p. 036019, 2019.

[4] C. Herff, L. Diener, M. Angrick, E. M. Mugler, M. C. Tate, M. Goldrick, D. J. Krusienski, M. W. Slutzky, and T. Schultz, "Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices," *Frontiers in neuroscience*, vol. 13, 2019.

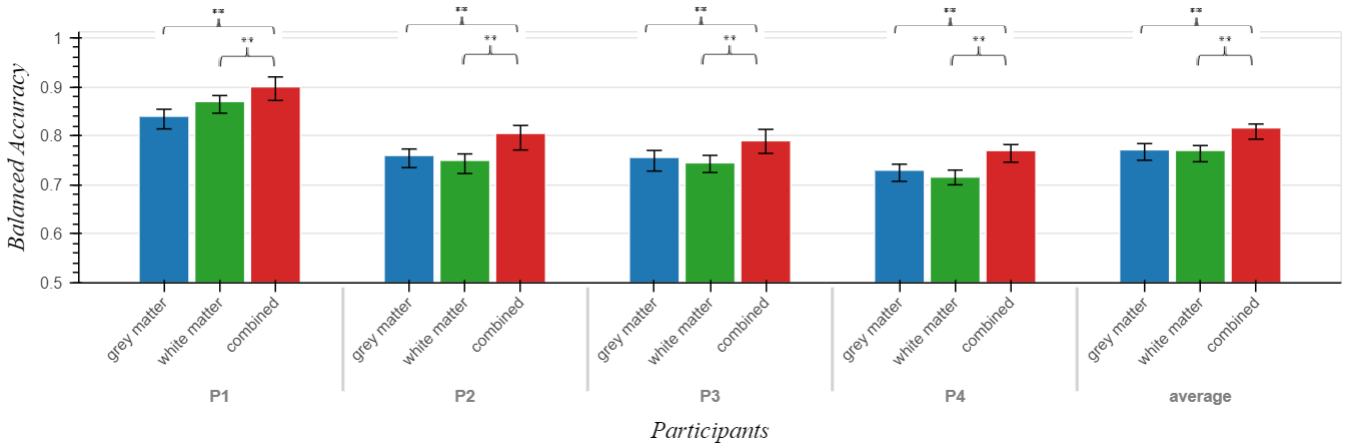

Fig. 4: Classification performance of decoding models for channels in grey matter, white matter, or combined. The bars show the averaged balanced accuracy of the 10-fold non-shuffled cross-validation analysis, and the error bars indicate the 95% confidence intervals. The dashed line shows the chance level results (50%). The right-most panel shows the average across all participants (**$p < 0.01$, Bonferroni-Holm corrected Wilcoxon signed-rank test).

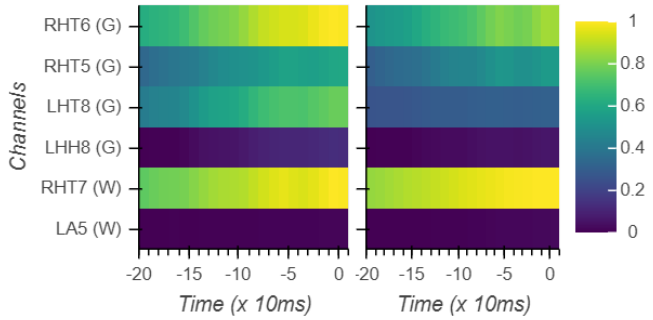

Fig. 5: Absolute value of average spatiotemporal patterns derived from the decoding model weights for an example participant (P2). The channel labels are grouped by electrode shaft, with left (L) and right (R) hemisphere designations and numbering 1 through 10 ranging from depth to superficial. The left panel shows the GM channels in the upper rows and the WM channels in the lower rows. The right panel shows the CM channels, which are a subset of the channels from the WM and GM models.

[5] P. Sun, G. K. Anumanchipalli, and E. F. Chang, "Brain2char: a deep architecture for decoding text from brain recordings," *Journal of neural engineering*, vol. 17, no. 6, p. 066015, 2020.

[6] D. A. Moses, M. K. Leonard, J. G. Makin, and E. F. Chang, "Real-time decoding of question-and-answer speech dialogue using human cortical activity," *Nature communications*, vol. 10, no. 1, pp. 1–14, 2019.

[7] M. Angrick, M. C. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, J. Saal, A. J. Colon, L. Wagner, D. J. Krusienski *et al.*, "Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity," *Communications biology*, vol. 4, no. 1, pp. 1–10, 2021.

[8] A. Y. Revell, A. B. Silva, D. Mahesh, L. Armstrong, T. C. Arnold, J. M. Bernabei, J. M. Stein, S. R. Das, R. T. Shinohara, D. S. Bassett *et al.*, "White matter signals reflect information transmission between brain regions during seizures," *BioRxiV*, 2021.

[9] C. Herff, D. J. Krusienski, and P. Kubben, "The potential of stereotactic-EEG for brain-computer interfaces: current progress and future directions," *Frontiers in neuroscience*, vol. 14, p. 123, 2020.

[10] G. Li, S. Jiang, S. E. Paraskevopoulou, G. Chai, Z. Wei, S. Liu, M. Wang, Y. Xu, Z. Fan, Z. Wu *et al.*, "Detection of human white matter activation and evaluation of its function in movement decoding using stereo-electroencephalography (sEEG)," *Journal of Neural Engineering*, vol. 18, no. 4, p. 0460c6, 2021.

[11] P. Soroush, M. Angrick, J. Shih, T. Schultz, and D. Krusienski, "Speech activity detection from stereotactic EEG," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2021, pp. 3402–3407.

[12] J. Kohler, M. C. Ottenhoff, S. Goulis, M. Angrick, A. J. Colon, L. Wagner, S. Tousseyn, P. L. Kubben, and C. Herff, "Synthesizing speech from intracranial depth electrodes using an encoder-decoder framework," *arXiv preprint arXiv:2111.01457*, 2021.

[13] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[14] E. Rothauser, "IEEE recommended practice for speech quality measurements," *IEEE Trans. on Audio and Electroacoustics*, vol. 17, pp. 225–246, 1969.

[15] K. Sjölander and J. Beskow, "Wavesurfer-an open source speech tool," in *Sixth International Conference on Spoken Language Processing*, 2000.

[16] G. Li, S. Jiang, S. E. Paraskevopoulou, M. Wang, Y. Xu, Z. Wu, L. Chen, D. Zhang, and G. Schalk, "Optimal referencing for stereo-electroencephalographic (sEEG) recordings," *NeuroImage*, vol. 183, pp. 327–335, 2018.

[17] M. R. Mercier, S. Bickel, P. Megevand, D. M. Groppe, C. E. Schroeder, A. D. Mehta, and F. A. Lado, "Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: a glimpse on white matter signal," *Neuroimage*, vol. 147, pp. 219–232, 2017.

[18] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96–110, 2014.

[19] J. A. Tourville, K. J. Reilly, and F. H. Guenther, "Neural mechanisms underlying auditory feedback control of speech," *Neuroimage*, vol. 39, no. 3, pp. 1429–1443, 2008.

[20] A. M. Alexandrou, T. Saarinen, S. Mäkelä, J. Kujala, and R. Salmelin, "The right hemisphere is highlighted in connected natural speech production and perception," *NeuroImage*, vol. 152, pp. 628–638, 2017.