2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) | 979-8-3503-7149-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EMBC53108.2024.10782208

Speech Sensitivity Analysis of Spatially Distributed Brain Areas Using Stereotactic EEG

Darius Ivucic*, Michael Dexheimer†, Pedram Z. Soroush†, Stephanie Ries‡, Jerry Shih,§
Dean J. Krusienski†, Tanja Schultz*
*Cognitive Systems Lab, University of Bremen, Bremen, Germany
darivuci@uni-bremen.de
†Virginia Commonwealth University, Richmond, VA, USA
‡San Diego State University, San Diego, CA, USA
§UCSD Health, San Diego, CA, USA

***Abstract*—Electrocorticographic (ECoG) activity recorded from the speech cortex has been extensively characterized and used in the development of speech neuroprosthetics. A recent shift in clinical brain monitoring from ECoG to stereotactic electroencephalography (sEEG) provides the opportunity to examine the role of deeper brain structures in speech processing. This study investigates spectro-temporal brain patterns generated during a speech task using sEEG data from five epilepsy patients. The analysis shows significant correlations in left and right temporal and motor regions, consistent with prior research in ECoG. Furthermore, correlation effects in rostral frontal areas are observed. A time lag analysis demonstrates distinct and functionally plausible activation patterns. The results further support the viability of sEEG for studying speech processes and provide insights into the involvement of spatially distributed, deeper brain areas.**

***Index Terms*—stereotactic EEG, speech production, neural signals**

I. Introduction

Understanding the spatio-temporal processing of speech in the brain is critical for building robust speech neuroprostheses [1]–[3], systems that have the potential to restore natural communication for patients that have lost the ability to speak due to disease or injury. Current work on speech neuroprostheses utilizes electrocorticography (ECoG) [4] and microarrays [5] to achieve remarkable results in decoding and synthesizing textual and acoustic representations of speech via Deep Learning.

A recent shift in clinical brain monitoring from ECoG to stereotactic electroencephalography (sEEG) provides the opportunity to examine the role of deeper brain structures in speech processing [6]. Although sEEG has been used for Speech-Activity detection [7], acoustic reconstructions of speech [8], [9] and word decoding [10], a foundational spatio-temporal characterization of sEEG activity during speech has not been conducted as has been done with ECoG [11].

The present analysis investigates spectro-temporal brain patterns generated during a speech task using sEEG data from five epilepsy patients. The results reveal comparable spectro-temporal patterns in temporal and motor regions as shown in ECoG [11], and additionally show that rostral frontal electrode channels are activated. This analysis further establishes the viability of sEEG for studying speech processes and highlights the involvement of spatially distributed, deeper brain areas that are not accessible by ECoG or cortical microelectrode arrays.

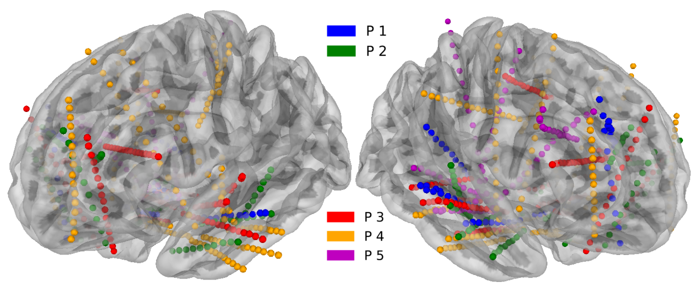

Fig. 1. Fronto-lateral views (left and right hemisphere, respectively) of superimposed electrode positions for all participants, mapped onto an average brain model.

II. Methodology

A. Participants and Electrode Positions

Data were collected from 5 patients hospitalized at UCSD Health for treatment of intractable epilepsy. All patients participated voluntarily and gave informed consent prior to the experiment. The study was approved by the Institutional Review Boards (IRBs) of Virginia Commonwealth University and UCSD Health. The sEEG electrode locations are solely based on clinical need and are unrelated to the research experiments. However, a subset of electrodes are determined to be in regions associated with language and speech processing. FreeSurfer [12] and MNE [13] were used to co-register CT and MRI and transform the electrode positions for each participant into an averaged brain (Figure 1). For this study, participants who did not complete all trials or for which no imaging data was available were excluded from the analysis. Participant information and electrode counts are summarized in Table I.

B. Experimental Design and Data Acquisition

All participants performed an experiment designed to investigate brain activity during different modes of speech production (see [14] and [15] for more details). A sentence from the Harvard sentence Corpus [16] was displayed on a

Participant	Sex	Age	#Total Electrodes	#Excluded Electrodes	#Speech Sensitive
P1	M	60	70	3	16
P2	M	32	121	5	24
P3	F	42	175	14	11
P4	M	21	234	11	18
P5	M	22	116	2	21

TABLE I
DEMOGRAPHIC INFORMATION AND NUMBER OF sEEG ELECTRODES FOR EACH PARTICIPANT

computer screen for 4 seconds while it was simultaneously narrated through loudspeakers. Patients were instructed by a visual cue to speak the sentence aloud from memory, followed by a cue to silently mouth the sentence, and finally, a cue to imagine speaking the sentence. The duration of each response period was 4 seconds. This sequence was repeated for 50 unique sentences. Throughout the experiment, time-synchronized sEEG and audio data were recorded.

sEEG signals were digitized at a sample rate of 1,024 Hz, with the audio signal captured using an external microphone and digitized at 44,100 Hz. In the present study, only data from the vocalized trials were analyzed since these offered a clear ground truth for the labeling process.

C. Preprocessing and Feature Extraction

The signals for each sEEG shaft were visually inspected and excluded in case of excessive noise or anomalous behavior, listed as '#Exluded Electrodes' in Table I. The signals from the remaining sEEG electrode were re-referenced shaft-wise using the Laplacian method [17]. Broadband-Gamma (BG) power features, a frequency feature known to reflect neural spiking [18] and to contain highly localized information about speech processes [19], [20], were extracted from each electrode channel. To extract the BG features, for each channel a band-pass filter from 70 to 170 Hz is applied followed by a notch-filter at 120 Hz, removing the second harmonic of the 60 Hz power line noise, and finally down-sampled to 400 Hz. The resulting BG-signal was windowed in 50 ms frames using 50 ms time shifts. Features were computed as the log-squared sum in each window and z-normalized per channel and trial.

D. Speech Labeling

The simultaneously recorded audio data was manually labeled for epochs of silence and speech. Following the *Verbmobil II* guideline for spontaneous speech labeling [21] pauses shorter than 2-syllable words ($\sim$ 350 ms), as well as *filled pauses* (i.e *Hmm* or *Uhm*), were included in the speech category when labeling the audio data.

To extract speech features aligned to the BG-features, 50-ms windows with 50-ms shifts of the annotated audio were processed, labeling a window as speech if the majority of the duration was annotated as a speech epoch and otherwise labeled as silence.

E. Speech Sensitivity

To assess the sensitivity of targeted areas to a given behavior, former work utilized p-value based activation measures [11], [22]–[24]. Here we highlight the effect size by calculating the Spearman correlation between BG-Features and Speech Labels for each channel. A positive correlation corresponds to a BG increase during speech, while a negative correlation indicates a decrease. For averaging across participants and comparison to activation measures in prior work, the squared Spearman-ρ is reported.

To assess the significance of the monotonic relation between BG features and speech labels, a randomization test was performed. The speech labels were randomly scrambled and correlated with the BG features 10,000 times for each channel. The resulting values of unsigned chance correlations were fit to a beta distribution for each channel, with the goodness of fit confirmed using a Kolomogorov-Smirnov test (p<0.05). Using the stochastic BG behavior of task-unrelated regions as an exclusion criterion [25], a channel was deemed to have a significant sensitivity if its correlation value from the respective beta-distribution yielded a p-value < 0.05, Bonferroni-corrected for the number of total channels across participants (N=761).

F. Time Lag Analysis

To examine the evolution of spatio-temporal activity during speaking, the BG-feature windows were analyzed in 50 ms steps from -200 ms to +200 ms with respect to the fixed speech data. This procedure resulted in 9 time-lagged correlations for each channel. The maximum correlation over all time steps is then taken as the activation time of the particular brain region [11]. Channels not significantly correlated over any time lag were excluded from further analysis (see Table I).

G. Regions of Interest

Volumetric labels for the electrode positions were obtained with FreeSurfer [12] using the Desikan-Killiany atlas [26] and including contacts positioned in white matter [7]. The labels were used to group the electrode channels into four anatomical regions of interest (ROI) based on relevance to speech and coverage: Frontal (F), Premotor and Motor (M), Right Temporal (RT) and Left Temporal (LT). Electrodes labeled as outside of the specified ROIs were excluded from the analysis.

III. RESULTS AND DISCUSSION

A. Spatial distribution of Speech Correlations

Figure 2 illustrates the spatial activation representing the largest significant R^2 value across all time lags for each respective channel. A total of 90 channels were observed to have significant correlations with speech (18$\pm$4.4 per subject, see '#Speech Sensitive' in Table I). Strong correlations are observed in left and right temporal regions, areas associated with hearing and speech processing [27]. It should be noted that for participants with bilateral coverage of the temporal lobes, slightly larger correlation values were observed over the left temporal region compared to the right temporal region.

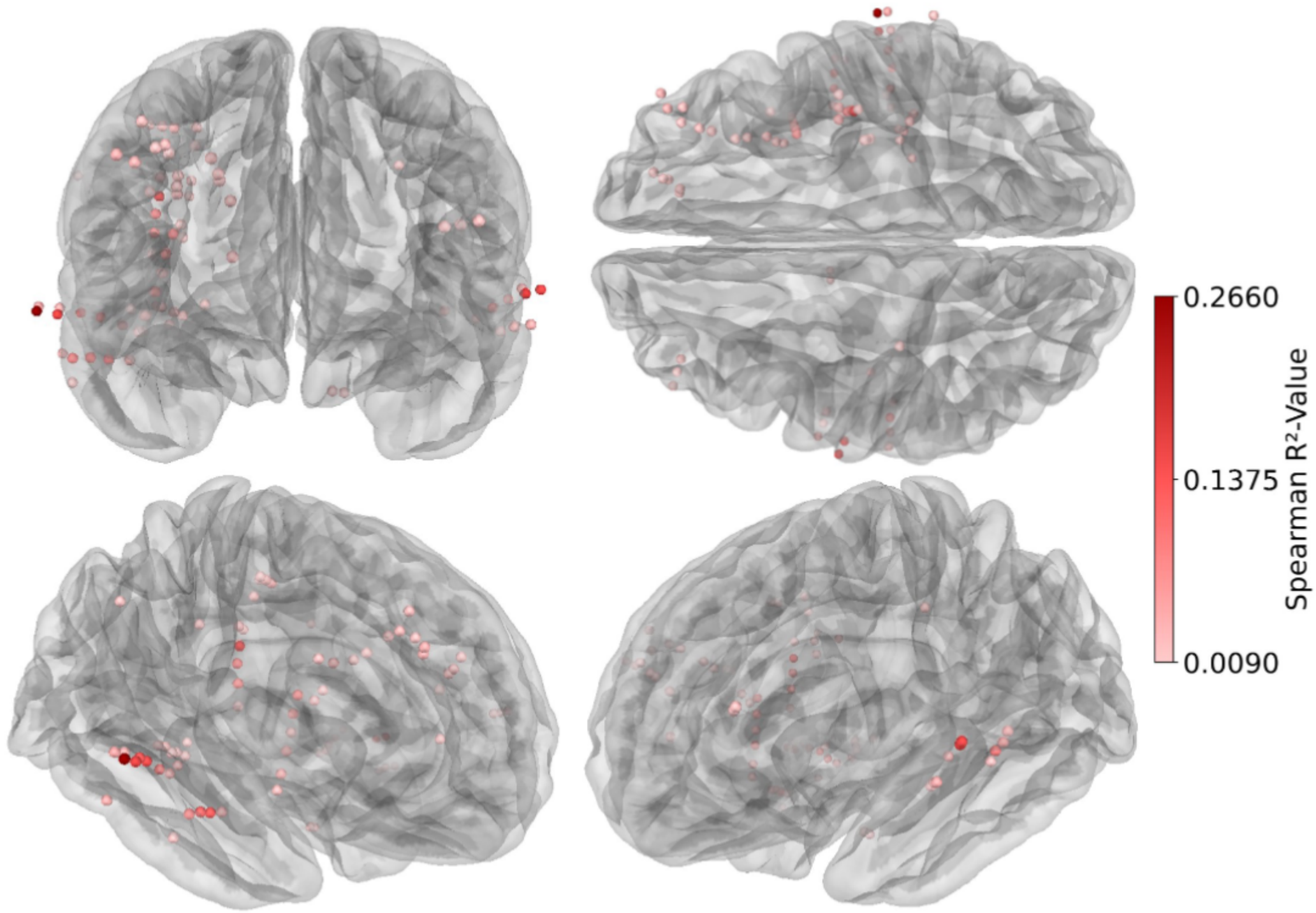

Fig. 2. Maximum R^2-values over all time lags for each channel, super-imposed for all participants. The lower views are the same as shown in Figure 1. The upper views show the rostral (left) and dorsal (right) views, respectively. Contacts with non-significant correlations to speech are omitted. Channels most sensitive to speech activity mainly lie in left and right temporal as well as motor areas. Frontal regions also show significant activation.

Despite limited coverage of speech-associated premotor and motor regions, channels near these areas show activations corresponding to the speech task. While the strongest activations are observed within the temporal lobe, weaker but significant effects are observed in the frontal, parietal, and other brain regions. These spatial activation patterns align with findings from related ECoG research [11] as well as prior sEEG studies [14].

B. Time Lag Correlations in ROIs

Figure 3 shows time averaged R^2-values for all channels, grouped by ROIs. Among the 90 channels assessed as speech sensitive, 75 were located within the selected ROIs: 12 electrodes in *Frontal*, 18 in *Motor*, 23 in *Right Temporal*, and 22 in *Left Temporal* areas.

The strongest correlation effects are found in the left temporal lobe, an area linked to complex auditory and speech processing [27]. Effects are maximal around a time shift of 0 to +100 ms, as is expected for auditory feedback while speaking [11]. While a similar temporal activation is observed in the right temporal lobe, the resulting correlations are slightly lower. As previously mentioned, effects in the temporal areas show participant-specific unilateral focus, which may be due to the asymmetric sampling of the temporal lobes.

As expected for speech motor planning and execution, the correlations in motor regions peak at a time shift of -100 to 0 ms. Despite the sparse coverage of these areas, typical of sEEG implants [6], the spatio-temporal activiations align with prior EGoG findings [11].

Channels in frontal areas (mostly right-rostral) comparatively show the lowest correlation values, spread from -150 to +100 ms. Since the speech task minimally involves working memory, the rostral-frontal cortex is expected to be involved to a limited degree [28].

IV. Conclusion

This study analyzed the temporal sensitivity of different brain areas during speech production and perception using sEEG. While previous studies used other modalities such as ECoG, with more focused coverage of cortical areas involved in speech processes, the present study offers unique insights into a variety of spatially distributed areas, including

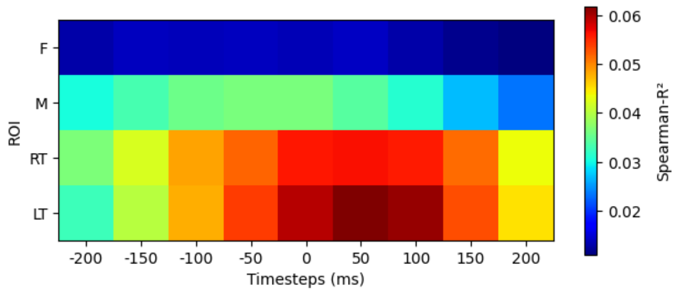

Fig. 3. Timestep averaged R^2-values for channels from chosen ROIs. (F=Frontal, M=Motor, RT=Right Temporal, LT=Left Temporal)

deeper structures. The temporally lagged correlation analysis revealed areas expected to be active in speech processing, such as the temporal lobes and motor areas, exhibit increased activity with expected time courses. These results reinforce the viability of sEEG for studying speech processes, which can be complementary to existing modalities such as ECoG and microelectrodes. Additional studies are needed to further characterize the unique contributions of sEEG beyond what can already be observed from the cortex, such as the potential for sEEG to decode speech prosody and affect, for instance.

References

[1] F. Bocquelet, T. Hueber, L. Girin, S. Chabardès, and B. Yvert, “Key considerations in designing a speech brain-computer interface,” *Journal of Physiology-Paris*, vol. 110, no. 4, pp. 392–401, 2016.

[2] C. Herff and T. Schultz, “Automatic speech recognition from neural signals: a focused review,” *Frontiers in neuroscience*, vol. 10, p. 429, 2016.

[3] S. Chakrabarti, H. M. Sandberg, J. S. Brumberg, and D. J. Krusienski, “Progress in speech decoding from the electrocorticogram,” *Biomedical Engineering Letters*, vol. 5, pp. 10–21, 2015.

[4] S. L. Metzger, K. T. Littlejohn, A. B. Silva, D. A. Moses, M. P. Seaton, R. Wang, M. E. Dougherty, J. R. Liu, P. Wu, M. A. Berger, *et al.*, “A high-performance neuroprosthesis for speech decoding and avatar control,” *Nature*, vol. 620, no. 7976, pp. 1037–1046, 2023.

[5] F. R. Willett, E. M. Kunz, C. Fan, D. T. Avansino, G. H. Wilson, E. Y. Choi, F. Kamdar, M. F. Glasser, L. R. Hochberg, S. Druckmann, *et al.*, “A high-performance speech neuroprosthesis,” *Nature*, vol. 620, no. 7976, pp. 1031–1036, 2023.

[6] 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.

[7] P. Soroush, C. Herff, S. Ries, J. Shih, T. Schultz, and D. Krusienski, “Contributions of stereotactic eeg electrodes in grey and white matter to speech activity detection,” in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4789–4792, IEEE, 2022.

[8] 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, p. 1055, 2021.

[9] 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.

[10] A. Petrosyan, A. Voskoboinikov, D. Sukhinin, A. Makarova, A. Skalnaya, N. Arkhipova, M. Sinkin, and A. Ossadtchi, “Speech decoding from a small set of spatially segregated minimally invasive intracranial eeg electrodes with a compact and interpretable neural network,” *Journal of Neural Engineering*, vol. 19, no. 6, p. 066016, 2022.

[11] J. S. Brumberg, D. J. Krusienski, S. Chakrabarti, A. Gunduz, P. Brunner, A. L. Ritaccio, and G. Schalk, “Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task,” *PloS one*, vol. 11, no. 11, p. e0166872, 2016.

[12] B. Fischl, “Freesurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[13] A. P. Rockhill, E. Larson, B. Stedelin, A. Mantovani, A. M. Raslan, A. Gramfort, and N. C. Swann, “Intracranial electrode location and analysis in mne-python,” *Journal of open source software*, vol. 7, no. 70, 2022.

[14] P. Z. Soroush, C. Herff, S. K. Ries, J. J. Shih, T. Schultz, and D. J. Krusienski, “The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings,” *NeuroImage*, vol. 269, p. 119913, 2023.

[15] S. Lesaja, M. Stuart, J. J. Shih, P. Z. Soroush, T. Schultz, M. Manic, and D. J. Krusienski, “Self-supervised learning of neural speech representations from unlabeled intracranial signals,” *IEEE Access*, vol. 10, pp. 133526–133538, 2022.

[16] E. Rothauser, “Ieee recommended practice for speech quality measurements,” *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, 1969.

[17] G. Li *et al.*, “Optimal referencing for stereo-electroencephalographic (seeg) recordings,” *NeuroImage*, vol. 183, pp. 327–335, 2018.

[18] N. E. Crone, D. Boatman, B. Gordon, and L. Hao, “Induced electrocorticographic gamma activity during auditory perception,” *Clinical neurophysiology*, vol. 112, no. 4, pp. 565–582, 2001.

[19] E. C. Leuthardt, X.-M. Pei, J. Breshears, C. Gaona, M. Sharma, Z. Freudenberg, D. Barbour, and G. Schalk, “Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task,” *Frontiers in human neuroscience*, vol. 6, p. 99, 2012.

[20] N. Crone, L. Hao, J. Hart, D. Boatman, R. Lesser, R. Irizarry, and B. Gordon, “Electrocorticographic gamma activity during word production in spoken and sign language,” *Neurology*, vol. 57, no. 11, pp. 2045–2053, 2001.

[21] S. Burger, *Lexicon for the Transliteration of Spontaneous Speech*. Interactive Systems Laboratories, Carnegie Mellon University, Pittsburgh, PA, USA. Subset for the transcription of the ISL Meeting Corpus.

[22] F. Lotte, J. Brumberg, P. Brunner, A. Gunduz, A. Ritaccio, C. Guan, and G. Schalk, “Electrocorticographic representations of segmental features in continuous speech. front hum neurosci. 2015; 09 (february): 1–13,” 2015.

[23] G. Schalk, J. Kubanek, K. J. Miller, N. Anderson, E. C. Leuthardt, J. G. Ojemann, D. Limbrick, D. Moran, L. A. Gerhardt, and J. R. Wolpaw, “Decoding two-dimensional movement trajectories using electrocorticographic signals in humans,” *Journal of neural engineering*, vol. 4, no. 3, p. 264, 2007.

[24] A. Gunduz, P. Brunner, A. Daitch, E. C. Leuthardt, A. L. Ritaccio, B. Pesaran, and G. Schalk, “Neural correlates of visual–spatial attention in electrocorticographic signals in humans,” *Frontiers in human neuroscience*, vol. 5, p. 89, 2011.

[25] A. Nourmohammadi, J. R. Swift, A. de Pesters, C. S. Guay, M. A. Adamo, J. C. Dalfino, A. L. Ritaccio, G. Schalk, and P. Brunner, “Passive functional mapping of receptive language cortex during general anesthesia using electrocorticography,” *Clinical Neurophysiology*, vol. 147, pp. 31–44, 2023.

[26] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et al.*, “An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest,” *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[27] G. Hickok, “Computational neuroanatomy of speech production,” *Nature reviews neuroscience*, vol. 13, no. 2, pp. 135–145, 2012.

[28] P. W. Burgess, S. K. Scott, and C. D. Frith, “The role of the rostral frontal cortex (area 10) in prospective memory: a lateral versus medial dissociation,” *Neuropsychologia*, vol. 41, no. 8, pp. 906–918, 2003.