

# A Comparative Study of EEG Representations and Channel Settings for EEG-to-Text Decoding during Reading Tasks

Eduardo V. Pérez-Hernández, Alejandro A. Torres-García, Luis Villaseñor-Pineda

Dept. of Computational Science  
Biosignals Processing and Medical Computing Laboratory  
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)  
Puebla, México

E-mail: {eduardo.perezhera, alejandro.torres, villasen}@inaoep.mx

**ABSTRACT:** Decoding natural language from electroencephalography (EEG) signals remains a major challenge in brain-computer interface research for assistive communication. This work presents a comparative study of EEG representations and channel configurations for EEG-to-text decoding during reading tasks. We evaluate Hilbert–Huang Transform (HHT) features, power spectral density (PSD), and temporal statistical descriptors across three spatial configurations: full-head EEG, a reduced 10–20 system, and a Broca–Wernicke subset. Experiments are conducted on the ZuCo v1 and v2 datasets using fully autoregressive decoding without teacher forcing. Results show that decoding performance varies across EEG representations and channel configurations. Temporal statistical features achieve the best overlap-based performance under the full-head configuration (BLEU-1 = 11.45, ROUGE-1 F1 = 11.52), while reduced channel subsets remain competitive despite lower spatial dimensionality. These findings highlight the importance of representation design, spatial selection, and task consistency for EEG-to-text decoding under short and variable EEG segments.

## INTRODUCTION

Human communication is a fundamental component of learning, social interaction, and personal autonomy. However, this ability can be severely impaired in individuals with neurological disorders affecting movement and speech, such as Amyotrophic Lateral Sclerosis (ALS). Brain-Computer Interfaces (BCIs) have emerged as a promising technology for restoring communication by directly decoding neural activity. Early BCI communication systems primarily relied on paradigms such as P300-based spellers, which allow users to select characters through visually evoked responses [1]. Although these systems can achieve reliable performance, they typically require structured interfaces and sequential character selection, resulting in slow communication rates. Another widely explored approach is imagined speech recognition, where users mentally rehearse words to induce discriminative neural patterns. However, such systems are

typically limited to closed vocabularies or predefined linguistic categories [2], and their performance degrades as vocabulary size increases or when generalizing to unseen words [3]. Recent advances in large language models have enabled new paradigms for open-vocabulary neural decoding, where brain activity is mapped directly to natural language sequences. In this context, electroencephalography (EEG) represents an attractive modality due to its non-invasive nature, relatively low cost, and high temporal resolution. The EEG2Text framework proposed by [4] introduced the first approach integrating word-level EEG features with a pre-trained language model to reconstruct sentences. This framework relies on the Zurich Cognitive Language Processing Corpus (ZuCo) datasets [5, 6], which combine EEG and eye-tracking recordings collected during natural reading. Although ZuCo provides a unique resource for studying neural language processing, it was not originally designed for text decoding tasks. The dataset aligns EEG signals with individual words using gaze duration information, producing short and highly variable word-level EEG segments. Earlier language decoding studies were primarily conducted using neuroimaging modalities such as fMRI [7]. Several recent studies have extended the EEG2Text framework by improving the transformation of EEG features into language-model embeddings and introducing semantic alignment strategies [8–13]. A common characteristic of many of these approaches is the use of the Hilbert–Huang Transform (HHT) as the primary EEG representation, motivated by its suitability for analyzing non-stationary signals. However, these methods typically do not consider how adaptive spectral decomposition behaves when applied to short-duration EEG segments. In word-level EEG extracted from natural reading, gaze-based segmentation produces highly variable temporal windows, often containing a limited number of samples. Under such conditions, decomposition-based representations may become unstable due to insufficient temporal support. Despite this limitation, the impact of EEG feature representation and spatial channel configuration on EEG-to-text decoding under short and variable-length conditions has not been systematically analyzed.

The main contributions of this work are:

- A controlled comparison of EEG representations for word-level EEG-to-text decoding, analyzing Hilbert–Huang Transform (HHT) features used in prior work, Power Spectral Density (PSD), and temporal statistical descriptors under short and variable EEG segments.
- A systematic evaluation of spatial electrode configurations, comparing full-head EEG (105 channels), a reduced 10–20 montage (20 channels), and a task-relevant subset covering Broca and Wernicke areas (10 channels).
- An analysis of how reading task composition influences decoding performance under realistic short-segment conditions.

Rather than proposing a new architecture, this study focuses on understanding representation–space interactions to inform the design of robust EEG-to-text systems.

## MATERIALS AND METHODS

*Task Definition:* As illustrated in Fig. 1, EEG and eye-tracking signals are simultaneously recorded during reading tasks. Eye-tracking information is used to temporally align EEG recordings with individual words, enabling the extraction of word-level EEG segments. These raw EEG segments are subsequently preprocessed and transformed into structured EEG representations, which serve as inputs to a sequence-to-sequence language model for sentence generation.

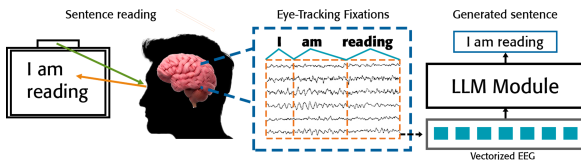


Figure 1: Overview of the EEG-to-text decoding task

The EEG-to-text decoding task aims to generate a sentence  $S$  from a sequence of word-level EEG feature vectors  $E = [e_1, e_2, \dots, e_K]$ , where each  $e_k$  corresponds to the EEG activity associated with the  $k$ -th word. The decoding process follows a sequence-to-sequence formulation that models the conditional probability of the target sentence given the EEG sequence:

$$p(S|E) = \prod_{k=1}^K p(s_k|E, s_{<k}) \quad (1)$$

where  $K$  denotes the length of the target text sequence and  $s_{<k}$  represents previously generated tokens.

*Database:* This study uses the Zurich Cognitive Language Processing Corpus (ZuCo 1.0) [5] and ZuCo 2.0 [6], which provide synchronized EEG and eye-tracking recordings collected during multiple reading tasks: Sentiment Reading (SR), Normal Reading (NR),

and Task-Specific Reading (TSR). In the SR and NR tasks, participants were instructed to read sentences naturally for comprehension, which was verified through randomly presented questions. In the TSR task, participants were required to determine whether a sentence corresponded to a predefined semantic category.

Table 1 summarizes the task-level statistics of the benchmark datasets. Here,  $N$  denotes the number of subjects, and %Fix represents the percentage of words containing at least 50 EEG samples. Across all tasks, approximately 62% of words are represented by fewer than 50 EEG samples. The average word-level EEG segment length is 141 samples, with a standard deviation of 88 samples, indicating substantial variability in segment duration.

Table 1: Task-level summary statistics. Values are reported as mean  $\pm$  standard deviation across subjects.

| Task   | N  | Sentences | % Fix           | Av. Samples |
|--------|----|-----------|-----------------|-------------|
| SR v1  | 12 | 400       | 67.3 $\pm$ 8.6  | 145.0       |
| NR v1  | 12 | 300       | 66.3 $\pm$ 11.8 | 138.7       |
| TSR v1 | 12 | 407       | 50.3 $\pm$ 13.1 | 131.7       |
| NR v2  | 18 | 349       | 66.6 $\pm$ 10.2 | 149.3       |

EEG signals were recorded using a 128-channel Geodesic Hydrocel system at a sampling frequency of 500 Hz. Twenty-three channels were removed due to artifact contamination, primarily from neck, chin, and electrooculogram electrodes. The artifact-corrected EEG signals provided by Hollenstein et al. [5] were used in all experiments.

Word-level EEG segments extracted from gaze duration exhibit highly variable duration, with many segments containing fewer than 50 samples (100 ms at 500 Hz). Such short windows can limit the stability of spectral estimates, particularly for low-frequency components.

*EEG Channel Configurations:* To analyze the effect of spatial channel selection, three EEG channel configurations were evaluated, as illustrated in Fig. 2: (i) full-head EEG with 105 channels, (ii) a reduced 10–20 system comprising 20 channels, and (iii) a task-relevant subset of 10 channels covering Broca and Wernicke areas.

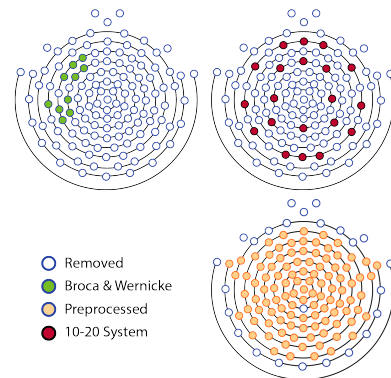


Figure 2: EEG channel configurations evaluated in this study: (i) full-head (105 channels), (ii) reduced 10–20 system (20 channels), and (iii) Broca–Wernicke subset (10 channels).

These configurations allow controlled analysis of spatial information reduction.

**EEG Representations:** Word-level EEG segments are defined using gaze duration information obtained from eye-tracking data. For each word, consecutive fixations are grouped, and the longest fixation sequence is selected. The EEG samples corresponding to this fixation group are concatenated to form a continuous EEG segment associated with the word. This procedure ensures temporal alignment between neural activity and linguistic units.

**Temporal Statistical Features:** Eight time-domain features are computed per channel: mean, standard deviation, skewness, kurtosis, Hjorth mobility, Hjorth complexity, sample entropy, and root mean square (RMS) [14]. This results in 840 features for the full-head configuration (105 channels), 160 for the reduced 10–20 configuration (20 channels), and 80 for the Broca–Wernicke configuration (10 channels).

Due to the short duration of word-level EEG segments, low-frequency bands that require longer temporal windows may produce unstable estimates. Therefore, spectral features are computed only within frequency bands that remain robust under short-window conditions.

**Power Spectral Density (PSD) Features:** Frequency-domain features are extracted using Welch’s method applied independently to each word-level EEG segment and channel. A 50-sample window (100 ms at a sampling rate of 500 Hz) with 50% overlap and a 256-point FFT is used. The mean spectral power is computed within five frequency bands: alpha2 (10.5–13 Hz), beta1 (13.5–18 Hz), beta2 (18.5–30 Hz), gamma1 (30.5–40 Hz), and gamma2 (40–49.5 Hz).

PSD features are computed for the same three electrode configurations as the temporal features. Each channel contributes one feature per frequency band, resulting in 525 features for the full-head configuration, 100 features for the reduced 10–20 configuration, and 50 features for the Broca–Wernicke configuration.

For comparability with prior EEG-to-text frameworks, the baseline Hilbert–Huang Transform (HHT) representation follows the original configuration used in previous work, which includes additional lower-frequency bands (theta1: 4–6 Hz, theta2: 6.5–8 Hz, and alpha1: 8.5–10 Hz).

**Proposed Architecture:** Following prior EEG-to-text decoding approaches [4, 12], a sequence-to-sequence architecture is employed to map word-level EEG representations to natural language text. Let  $E = [e_1, e_2, \dots, e_K]$  denote the sequence of word-level EEG feature vectors extracted from a sentence. These features are first processed by a dedicated Transformer-based EEG encoder composed of six layers, each using five attention heads. The output is passed through a linear projection layer with ReLU activation to a 1024-dimensional space, matching the embedding dimension of the downstream language model. The projected embeddings are provided as input to a pre-trained BART-large encoder–decoder model [15], replacing the standard textual embeddings

at the BART encoder input. The overall architecture is illustrated in Fig. 3.

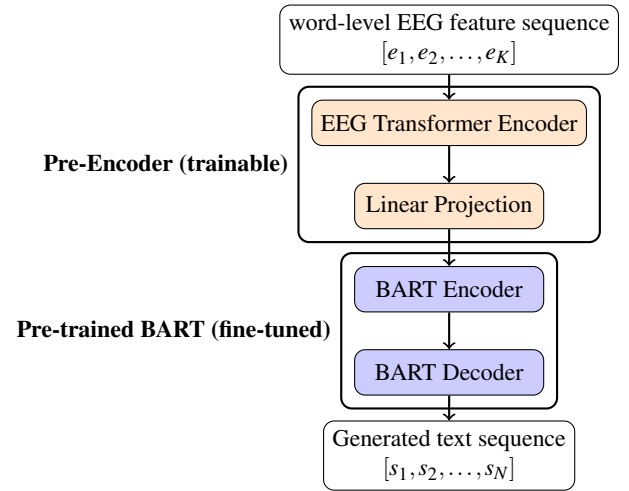


Figure 3: Overview of the EEG-to-text decoding architecture, illustrating the word-level EEG feature sequence, projection to the language model embedding space (Pre-Encoder), and pre-trained BART.

All model components are trained jointly for 30 epochs using the Adam optimiser with a learning rate of  $2 \times 10^{-5}$  and gradient clipping with a maximum  $\ell_2$  norm of 1.0. The ZuCo dataset is partitioned using an 80%–10%–10% split (train–development–test) with a fixed random seed of 42. The split is performed at the sentence level across the entire dataset. This setup corresponds to an intra-subject evaluation protocol, where samples from the same subjects may appear in both training and test sets, but no sentence appears in more than one split. The objective of this work is controlled representation-level comparison under a fixed evaluation setting. The complete set of hyperparameters is summarised in Table 2.

Table 2: Model configuration and experimental settings.

| Parameter           | Value                            |
|---------------------|----------------------------------|
| EEG Input Dimension | Representation-dependent         |
| EEG Encoder         | 6-layer Transformer (5 heads)    |
| Projection Layer    | Linear + ReLU (1024 dimensions)  |
| Language Model      | BART-large (1024-dim embeddings) |
| Training Epochs     | 30                               |
| Batch Size          | 32                               |
| Beam Search Size    | 5                                |
| Repetition Penalty  | 5.0                              |
| Max Sequence Length | 56 tokens                        |

**Inference and Evaluation:** Model evaluation is conducted using autoregressive generation without teacher forcing during inference. No ground-truth tokens are provided to the decoder at test time. Model performance is evaluated using complementary automatic metrics commonly adopted in EEG-to-text studies. BLEU scores with  $n$ -grams ranging from 1 to 4 are reported to measure lex-

Table 3: Feature representation comparison under the SR+NR+NRv2 setting using the full-head configuration (sentence-level mean  $\pm$  standard deviation).

| Representation | BLEU-1                            | BLEU-2                           | BLEU-3                           | BLEU-4                           | ROUGE-1 F1                        | ROUGE-1 P                         | ROUGE-1 R                         | WER                                 |
|----------------|-----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|
| Baseline HHT   | 9.9 $\pm$ 5.9                     | 2.2 $\pm$ 1.9                    | 1.2 $\pm$ 0.8                    | 0.9 $\pm$ 0.5                    | 10.0 $\pm$ 7.0                    | 11.3 $\pm$ 8.1                    | 9.8 $\pm$ 7.4                     | 119.9 $\pm$ 48.7                    |
| PSD            | 10.0 $\pm$ 6.3                    | 2.3 $\pm$ 2.3                    | 1.3 $\pm$ 1.3                    | 0.9 $\pm$ 0.9                    | 9.9 $\pm$ 7.2                     | 12.0 $\pm$ 9.3                    | 9.4 $\pm$ 7.5                     | <b>112.7 <math>\pm</math> 41.6*</b> |
| Statistical    | <b>11.4 <math>\pm</math> 6.6*</b> | <b>2.8 <math>\pm</math> 3.1*</b> | <b>1.5 <math>\pm</math> 1.6*</b> | <b>1.1 <math>\pm</math> 1.1*</b> | <b>11.5 <math>\pm</math> 7.8*</b> | <b>12.8 <math>\pm</math> 9.0*</b> | <b>11.3 <math>\pm</math> 7.8*</b> | 113.3 $\pm$ 35.8*                   |

\*  $p < 0.05$  with respect to HHT features. Bold values indicate the best result for each metric.

ical overlap between generated and reference sentences. ROUGE scores are used to assess recall- and precision-oriented overlap at the sequence level. Word Error Rate (WER) is additionally employed to quantify token-level transcription errors. For readability, BLEU and ROUGE scores are reported as percentages. WER values may exceed 100% because autoregressive generation can produce insertion, deletion, and substitution errors whose cumulative count surpasses the number of reference tokens.

## RESULTS

The experimental analysis compares EEG representations, channel configurations, and task composition under identical decoding conditions. The Hilbert–Huang Transform (HHT) representation is used as the reference baseline because it corresponds to the feature extraction strategy adopted in prior EEG-to-text frameworks [4, 12]. Results are reported as mean  $\pm$  standard deviation across sentences, and statistical significance was assessed using paired bootstrap resampling with 10,000 iterations at the sentence level. Asterisks indicate statistically significant differences ( $p < 0.05$ ). The reported standard deviations capture sentence-level variability caused by differences in segment duration, sentence complexity, and error propagation in autoregressive generation.

*Impact of EEG Representation Type:* To isolate representational effects, we compare three feature types under a controlled setting: (i) HHT features as a baseline representation, (ii) PSD features, and (iii) temporal statistical features proposed in this study.

All representations are evaluated under identical decoding conditions. Table 3 reports results using the full-head spatial configuration and the SR+NR+NRv2 task combination, allowing the effect of EEG representation to be analyzed while keeping spatial configuration and dataset composition fixed.

Temporal statistical features achieve the highest BLEU and ROUGE scores, indicating more stable lexical reconstruction under short EEG segments. In contrast, PSD features obtain the lowest WER, suggesting slightly lower token-level reconstruction error. The HHT baseline consistently produces lower overlap-based performance under the evaluated conditions.

*Impact of Spatial Configuration:* To assess the impact of spatial electrode selection, we analyze the SR+NR+NRv2 configuration across all spatial settings while keeping task composition fixed. Table 4 summarizes the results obtained using PSD and statistical feature

representations.

For PSD features, reduced spatial configurations achieve comparable or slightly higher overlap-based metrics than the full-head setup, with the 10–20 configuration obtaining the highest BLEU-1 and ROUGE-1 F1 scores and the Broca–Wernicke subset yielding the lowest WER. A similar trend appears for temporal statistical features: the 10–20 configuration produces the highest BLEU-1 score, the full-head configuration achieves the best ROUGE-1 F1, and the Broca–Wernicke subset again obtains the lowest WER. Overall, performance differences between spatial configurations remain small, suggesting that reducing the number of electrodes does not systematically degrade decoding performance.

*Impact of Task Composition:* We further analyzed the impact of task composition across feature representations using the full-head spatial configuration. This experimental design allows evaluation of the effect of progressively increasing dataset size and task heterogeneity while maintaining identical spatial conditions. Table 5 summarizes the obtained results.

Under the SR setting, temporal statistical features achieve the highest BLEU-1 and ROUGE-1 scores, while PSD features produce the lowest WER within this task configuration. When NR is incorporated, temporal statistical features achieve the strongest overall overlap-based performance, whereas both PSD and HHT representations produce slightly lower scores.

The inclusion of TSR changes the relative behavior across representations. Under this setting, PSD features achieve the highest ROUGE scores and the lowest WER, while temporal statistical features exhibit lower overlap-based performance compared with the SR+NR configuration.

Finally, the SR+NR+NRv2 configuration yields the strongest overall performance for statistical representations, with improvements across most BLEU and ROUGE metrics relative to HHT and PSD features.

Overall, the results indicate that decoding performance depends not only on dataset size, but also on the cognitive consistency of the reading tasks. Combining tasks with similar reading behavior generally produces more stable performance, whereas incorporating TSR introduces additional variability that may affect the temporal consistency of word-level EEG segments.

Table 4: Impact of spatial configuration under the SR+NR+NRv2 setting using PSD and temporal statistical features (sentence-level mean  $\pm$  standard deviation).

| PSD features                  |                                   |                                  |                                  |                                  |                                   |                                   |                                   |                                     |
|-------------------------------|-----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|
| Spatial Configuration         | BLEU-1                            | BLEU-2                           | BLEU-3                           | BLEU-4                           | ROUGE-1 F1                        | ROUGE-1 P                         | ROUGE-1 R                         | WER                                 |
| Full-head                     | 10.0 $\pm$ 6.3                    | 2.3 $\pm$ 2.2                    | 1.3 $\pm$ 1.3                    | 0.9 $\pm$ 0.9                    | 9.9 $\pm$ 7.2                     | 12.0 $\pm$ 9.3                    | 9.4 $\pm$ 7.5                     | 112.7 $\pm$ 41.6                    |
| 10–20                         | <b>11.1 <math>\pm</math> 6.9*</b> | <b>2.9 <math>\pm</math> 3.4*</b> | <b>1.5 <math>\pm</math> 2.2*</b> | <b>1.2 <math>\pm</math> 1.7*</b> | <b>11.9 <math>\pm</math> 8.3*</b> | <b>14.0 <math>\pm</math> 10.4</b> | <b>11.9 <math>\pm</math> 8.3*</b> | 111.7 $\pm$ 42.2                    |
| Broca–Wernicke                | 11.0 $\pm$ 6.1*                   | 2.5 $\pm$ 2.4*                   | 1.3 $\pm$ 0.9*                   | 1.0 $\pm$ 0.5                    | 10.8 $\pm$ 7.1*                   | 12.3 $\pm$ 8.4                    | 10.5 $\pm$ 7.8*                   | <b>111.5 <math>\pm</math> 36.5</b>  |
| Temporal statistical features |                                   |                                  |                                  |                                  |                                   |                                   |                                   |                                     |
| Spatial Configuration         | BLEU-1                            | BLEU-2                           | BLEU-3                           | BLEU-4                           | ROUGE-1 F1                        | ROUGE-1 P                         | ROUGE-1 R                         | WER                                 |
| Full-head                     | 11.4 $\pm$ 6.6                    | <b>2.8 <math>\pm</math> 3.1</b>  | 1.5 $\pm$ 1.6                    | 1.1 $\pm$ 1.1                    | <b>11.5 <math>\pm</math> 7.8</b>  | <b>12.8 <math>\pm</math> 9.0</b>  | <b>11.3 <math>\pm</math> 7.8</b>  | 113.3 $\pm$ 35.8                    |
| 10–20                         | <b>11.9 <math>\pm</math> 6.5*</b> | 2.7 $\pm$ 2.3                    | 1.4 $\pm$ 1.1                    | 1.1 $\pm$ 0.8                    | 11.2 $\pm$ 7.5                    | 12.5 $\pm$ 8.7                    | 11.0 $\pm$ 8.0                    | 111.2 $\pm$ 33.9*                   |
| Broca–Wernicke                | 11.2 $\pm$ 6.7                    | 2.8 $\pm$ 2.8                    | 1.5 $\pm$ 1.4                    | 1.1 $\pm$ 1.0                    | 10.0 $\pm$ 8.0                    | 12.8 $\pm$ 9.6                    | 10.36 $\pm$ 7.9                   | <b>107.6 <math>\pm</math> 31.3*</b> |

\*  $p < 0.05$  with respect to the full-head configuration. Bold values indicate the best result for each metric within each feature representation.

Table 5: Impact of task composition across EEG feature representations using the full-head spatial configuration (sentence-level mean  $\pm$  standard deviation).

| Task       | Representation | BLEU-1                            | BLEU-2                           | BLEU-3                           | BLEU-4                           | ROUGE-1 F1                        | ROUGE-1 P                          | ROUGE-1 R                         | WER                                 |
|------------|----------------|-----------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|------------------------------------|-----------------------------------|-------------------------------------|
| SR         | HHT            | 12.9 $\pm$ 6.8                    | 3.0 $\pm$ 2.9                    | 1.6 $\pm$ 1.8                    | 1.2 $\pm$ 1.0                    | 12.1 $\pm$ 7.4                    | 13.9 $\pm$ 9.7                     | 12.1 $\pm$ 8.1                    | 128.0 $\pm$ 66.9                    |
|            | PSD            | 12.3 $\pm$ 6.2                    | 2.8 $\pm$ 2.4                    | 1.5 $\pm$ 0.9                    | 1.1 $\pm$ 0.5                    | 11.5 $\pm$ 6.9                    | 13.6 $\pm$ 9.2                     | 11.1 $\pm$ 7.2                    | <b>121.3 <math>\pm</math> 57.8</b>  |
|            | Statistical    | <b>13.5 <math>\pm</math> 6.8*</b> | <b>3.3 <math>\pm</math> 3.0</b>  | <b>1.7 <math>\pm</math> 1.6</b>  | <b>1.2 <math>\pm</math> 0.8</b>  | <b>12.8 <math>\pm</math> 7.3</b>  | <b>14.8 <math>\pm</math> 9.7</b>   | <b>12.7 <math>\pm</math> 7.9</b>  | 124.8 $\pm$ 67.3                    |
| SR+NR      | HHT            | 10.6 $\pm$ 6.4                    | 2.4 $\pm$ 2.1                    | 1.3 $\pm$ 0.9                    | 0.9 $\pm$ 0.6                    | 10.7 $\pm$ 7.4                    | 13.0 $\pm$ 9.0                     | 10.2 $\pm$ 7.9                    | 120.8 $\pm$ 47.6                    |
|            | PSD            | 11.2 $\pm$ 6.0                    | 2.2 $\pm$ 1.7                    | 1.2 $\pm$ 0.8                    | 0.9 $\pm$ 0.5                    | 10.5 $\pm$ 6.9                    | 11.0 $\pm$ 8.4                     | <b>11.2 <math>\pm</math> 7.3*</b> | 115.0 $\pm$ 41.9*                   |
|            | Statistical    | <b>11.5 <math>\pm</math> 6.8*</b> | <b>2.8 <math>\pm</math> 2.3</b>  | <b>1.5 <math>\pm</math> 1.1</b>  | <b>1.1 <math>\pm</math> 0.8</b>  | <b>11.6 <math>\pm</math> 7.7*</b> | <b>13.4 <math>\pm</math> 9.5</b>   | 11.1 $\pm$ 8.2                    | <b>111.5 <math>\pm</math> 37.6*</b> |
| SR+NR+TSR  | HHT            | <b>10.9 <math>\pm</math> 6.1</b>  | 2.6 $\pm$ 2.3                    | 1.4 $\pm$ 1.1                    | 1.0 $\pm$ 0.6                    | 10.9 $\pm$ 7.1                    | 13.2 $\pm$ 9.0                     | 10.3 $\pm$ 7.9                    | 109.0 $\pm$ 29.2                    |
|            | PSD            | 9.9 $\pm$ 8.6                     | <b>2.8 <math>\pm</math> 3.5</b>  | <b>1.4 <math>\pm</math> 1.6</b>  | <b>1.0 <math>\pm</math> 1.0</b>  | <b>12.4 <math>\pm</math> 8.3*</b> | <b>16.3 <math>\pm</math> 11.5*</b> | <b>11.0 <math>\pm</math> 8.0*</b> | <b>107.0 <math>\pm</math> 38.0*</b> |
|            | Statistical    | 10.4 $\pm$ 5.9                    | 2.4 $\pm$ 2.0                    | 1.3 $\pm$ 0.9                    | 1.0 $\pm$ 0.5                    | 10.6 $\pm$ 7.9                    | 12.1 $\pm$ 9.8                     | 10.5 $\pm$ 8.4                    | 112.7 $\pm$ 36.9                    |
| SR+NR+NRv2 | HHT            | 9.9 $\pm$ 5.9                     | 2.2 $\pm$ 1.9                    | 1.2 $\pm$ 0.8                    | 0.9 $\pm$ 0.5                    | 10.0 $\pm$ 7.0                    | 11.3 $\pm$ 8.1                     | 9.8 $\pm$ 7.4                     | 119.9 $\pm$ 48.7                    |
|            | PSD            | 10.0 $\pm$ 6.3                    | 2.3 $\pm$ 2.3                    | 1.3 $\pm$ 1.3                    | 0.9 $\pm$ 0.9                    | 9.9 $\pm$ 7.2                     | 12.0 $\pm$ 9.3                     | 9.4 $\pm$ 7.5                     | <b>112.7 <math>\pm</math> 41.6*</b> |
|            | Statistical    | <b>11.4 <math>\pm</math> 6.6*</b> | <b>2.8 <math>\pm</math> 3.1*</b> | <b>1.5 <math>\pm</math> 1.6*</b> | <b>1.1 <math>\pm</math> 1.1*</b> | <b>11.5 <math>\pm</math> 7.8*</b> | <b>12.8 <math>\pm</math> 9.1*</b>  | <b>11.3 <math>\pm</math> 7.8*</b> | 113.4 $\pm$ 35.8*                   |

\*  $p < 0.05$  with respect to the HHT baseline within the corresponding task configuration. Bold values indicate the best result for each task configuration.

## DISCUSSION

The results show that EEG representation, channel configuration, and task composition all influence EEG-to-text decoding performance, although the absolute scores remain modest. This is consistent with the difficulty of the task: word-level EEG segments are short, heterogeneous, and only indirectly aligned with linguistic structure. Under these conditions, the relative trends across experimental settings become more informative than the absolute metric values themselves.

One of the main findings of this study is that temporal statistical features perform consistently well across experimental conditions, suggesting that compact time-domain descriptors may provide more stable representations than decomposition-based spectral methods for gaze-aligned EEG segments.

Another important observation concerns spatial configuration. Reducing the number of EEG channels does not systematically degrade decoding performance, and reduced channel subsets remain competitive with the full-head configuration across several metrics.

Task composition also plays an important role. The best performance is not always obtained by simply increasing the amount of training data. Instead, combinations of tasks with similar reading behavior tend to produce more stable decoding patterns, whereas Task-Specific Reading introduces additional cognitive variability. This variability likely affects fixation dynamics and temporal align-

ment between EEG activity and linguistic processing.

The relatively low BLEU and ROUGE scores observed across all experiments should be interpreted in the context of the evaluation protocol and the intrinsic difficulty of EEG-to-text decoding. Unlike constrained decoding settings, the present work evaluates fully autoregressive open-vocabulary generation without teacher forcing. Under this setting, decoding errors propagate across generated tokens, substantially increasing task difficulty. In addition, the ZuCo datasets provide short and highly variable EEG segments aligned at the word level rather than sentence-level neural representations, which further limits decoding stability.

Although direct quantitative comparison with previous EEG-to-text studies remains difficult due to differences in inference protocols, evaluation settings, and dataset composition, the HHT representation used in this work follows the same feature extraction strategy adopted in previous EEG-to-text frameworks [4, 12]. Under this controlled setting, the proposed temporal statistical representations achieved competitive or superior overlap-based performance relative to the HHT baseline across several experimental configurations.

Importantly, the observed limitations of HHT representations should not be interpreted as a general limitation of adaptive spectral decomposition methods. Instead, the results suggest that these methods may become less stable when applied to gaze-aligned word-level EEG recordings with limited temporal context.

The present study should be interpreted as a controlled comparative analysis of EEG representations and spatial configurations under a fixed autoregressive decoding framework rather than as a direct state-of-the-art comparison. Taken together, the results suggest that future EEG-to-text systems should not focus exclusively on model complexity, since representation design, spatial parsimony, and cognitively consistent datasets may be equally important for improving neural language decoding performance.

## CONCLUSIONS

This study investigated the influence of EEG representation, spatial channel configuration, and task composition on word-level EEG-to-text decoding using the ZuCo v1 and v2 datasets.

The results show that temporal statistical features provide competitive and stable decoding performance across several experimental settings, suggesting that compact time-domain descriptors may be more robust than decomposition-based spectral representations for short EEG segments.

The spatial analysis further demonstrates that reduced electrode subsets can achieve performance comparable to full-head EEG configurations. In particular, the 10–20 and Broca–Wernicke configurations remain competitive while substantially reducing spatial dimensionality.

The experiments also highlight the importance of task composition. Combining reading paradigms with similar cognitive characteristics generally improves decoding stability, whereas heterogeneous tasks such as Task-Specific Reading introduce additional variability that can negatively affect performance.

## ACKNOWLEDGMENT

The first author of this work acknowledges the Secretaría de Educación, Ciencia, Tecnología e Innovación (SECI-HTI) of Mexico for the postgraduate scholarship they received. This support made it possible to successfully complete this work.

## REFERENCES

- [1] Pan J, Chen X, Ban N, He J, Chen J, Huang H. Advances in p300 brain–computer interface spellers: Toward paradigm design and performance evaluation. *Frontiers in Human Neuroscience*. 2022;16.
- [2] Kheddar H, Hemis M, Himeur Y. Automatic speech recognition using advanced deep learning approaches: A survey. *Information Fusion*. 2024;109:102422.
- [3] Lopez-Bernal D, Balderas D, Ponce P, Molina A. A state-of-the-art review of EEG-based imagined speech decoding. *Frontiers in Human Neuroscience*. 2022;16:867281.
- [4] Liu H, Hajialigol D, Antony B, Han A, Wang X. EEG2text: Open vocabulary EEG-to-text translation with multi-view transformer. In: *Proc. 2024 IEEE Int. Conf. Big Data (BigData)*. IEEE: Washington, DC, USA, 2024, 1824–1833.
- [5] Hollenstein N, Rotsztein J, Troendle M, Pedroni A, Zhang C, Langer N. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific Data*. 2018;5(1):180291.
- [6] Hollenstein N, Troendle M, Zhang C, Langer N. ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation. In: *Proc. 12th Int. Conf. Language Resources and Evaluation (LREC)*. 2020.
- [7] Tang J, LeBel A, Jain S, Huth AG. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*. 2023;26:858–866.
- [8] Feng X, Feng X, Qin B, Liu T. Aligning semantic in brain and language: A curriculum contrastive method for EEG-to-text generation. *IEEE Trans. Neural Syst. Rehabil. Eng.*. 2023;31:3874–3883.
- [9] Zhou J, Duan Y, Chang YC, Wang YK, Lin CT. BELT: Bootstrapped EEG-to-language training by natural language supervision. *IEEE Trans. Neural Syst. Rehabil. Eng.*. 2024;32:3278–3288.
- [10] Amrani H, Micucci D, Napoletano P. Deep representation learning for open vocabulary EEG-to-text decoding. *IEEE J. Biomed. Health Inform.*. 2024.
- [11] Wang J, Song Z, Ma Z, Qiu X, Zhang M, Zhang Z. Enhancing EEG-to-text decoding through transferable representations from pre-trained contrastive EEG-text masked autoencoder. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL: Bangkok, Thailand, 2024, 7278–7292.
- [12] Jo H, Yang Y, Han J, Duan Y, Xiong H, Lee WH. Evaluating EEG-to-text models through noise-based performance analysis. *Scientific Reports*. 2025.
- [13] Zheng H, Wu Y, Qian T, Yue W, Wang X. Guiding LLMs to decode text via aligning semantics in EEG signals and language. *Expert Systems with Applications*. 2025;299:130300.
- [14] Jin Z, Li D, Huang S. A systematic review of eeg-based imagined speech decoding. *Applied Soft Computing*. 2025;183:113563.
- [15] Lewis M et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. ACL: Online, Jul. 2020, 7871–7880.