A Word Recognition Paradigm Through EEG Analysis: Imagined Speech Classification

Francesco Iacomi[a], Andrea Farabbi[b], Maximiliano Mollura[c], Edoardo Maria Polo[d], Riccardo Barbieri[e] and Luca Mainardi[f]

Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy

Keywords: Electroencephalogram (EEG), Machine Learning, Brain Computer Interface (BCI), Imagined Speech, Biomedical Signal Processing.

Abstract: This study presents an innovative approach for decoding imagined speech using EEG signals. The proposed analysis aims at revealing phonetic and semantic properties of imagined words through brain activity. The experimental protocol involves presenting words to subjects while recording EEG signals via a 64-channels cap. Each word is associated with three specific properties: length, presence of doubles, and category of meaning. The protocol includes fixation, cue presentation, thinking, and rest phases. EEG signals undergo meticulous preprocessing stage to mitigate noise and artifacts. Features are extracted from the processed signal, including statistical, spectral, and fractal domain measures. The dimensionality of features is reduced through statistical means. Several classifiers, (e.g., MLP, KNN, LDA, QDA), are trained and evaluated to predict mentioned properties of imagined words. An ensemble model (LDA) comprising the best 3 models mentioned above is then employed to enhance classification accuracy. Results illustrate the effectiveness of decoding imagined word properties with average accuracies of 35.2% for "Category", 57.2% for "Doubles", and 55.8% for "Length". By aggregating all predictions we are able to decode each single word with a mean accuracy of 11.8% (random accuracy = 8.33%) and an average word distance of 1.54. Post-classification studies on the most relevant variables and on the most discriminating channels further deepen our understanding of the proposed Imagined Speech cognitive process, showing different brain activations for different linguistic aspects.

1 INTRODUCTION

The central objective of this research is to decode Imagined Speech (IS), a cognitive process where individuals mentally simulate speech production without any actual vocalization or physical muscle movement (Martin, 2018). The significance of IS lies in its potential applications, particularly in the realm of Brain-Computer Interfaces (BCIs), where it can facilitate natural and efficient communication among individuals (Gu et al., 2021).

The electroencephalogram (EEG) is the chosen neurophysiological signal for this study due to its non-invasiveness, exceptional temporal resolution, and the minimal instrumentation required for data acquisition (Yi et al., 2013). EEG's portability and ease of use make it a compelling choice for BCIs, aligning with the study's goal of creating BCIs that are accessible to a wide audience (Laureys, 2005).

The proposed study also delves into the current state of research in IS decoding. This process can be distilled into four key steps: signal acquisition, pre-processing, feature extraction, and classification (Lopez-Bernal et al., 2022). Researchers have employed various IS prompts, including visual and auditory cues, and have employed several techniques for feature extraction, such as statistical analysis (Panachakel and Ramakrishnan, 2021), frequency domain analysis (Fitriah et al., 2022), and spatial analysis (Lopez-Bernal et al., 2022). Classification, the final step, has seen the application of both machine learning and deep learning models, each with its strengths and areas of application. In Table 1 are summarized the state of the art classifications as a term of comparison.

[a] https://orcid.org/0009-0005-2839-8414
[b] https://orcid.org/0000-0001-5582-4654
[c] https://orcid.org/0000-0002-2248-8145
[d] https://orcid.org/0000-0003-0432-1314
[e] https://orcid.org/0000-0001-9381-3833
[f] https://orcid.org/0000-0002-6276-6314

Iacomi, F., Farabbi, A., Mollura, M., Polo, E., Barbieri, R. and Mainardi, L.
A Word Recognition Paradigm Through EEG Analysis: Imagined Speech Classification.
DOI: 10.5220/0012372500003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 622-628
ISBN: 978-989-758-688-0; ISSN: 2184-4305

Table 1: State of the art classification performance.

Author	N of prompts	Subjects	Trials per word	Accuracy
(Nguyen et al., 2018)	3 short words, 2 long words	6	100	50.1%(3 classes),66.2%(2 classes)
(Qureshi et al., 2018)	5 words	8	100	40.1%
(Dipti Pawar, 2020)	4 words	6	50	45.07%
(Chau, 2017)	2 words	12	180	66.35%
(Dipti Pawar, 2023)	5 words	8	180	64%

Yet, amidst these advancements, several challenges and limitations persist. These include a limited vocabulary, reduced accuracy, a predominant focus on offline approaches, and limited dataset sizes (Panachakel and Ramakrishnan, 2021). Overcoming these challenges need the creation of larger datasets, innovative prompts, refinements in feature extraction and machine learning techniques, a deeper exploration of connectivity during IS, and a shift toward inter-subject approaches (Panachakel and Ramakrishnan, 2021).

This study seeks to tackle these challenges head-on. It plans to do so by conducting a subject-specific classification of 12 carefully selected words, categorizing them based on word length, the presence of doubles, and semantic class. The goal is twofold: a) to determine which EEG features can effectively discriminate imagine words and b) expand the vocabulary space for IS decoding. Additionally, we aim at smoothing feature extraction processes, identify informative EEG channels, and enhance the overall efficiency and performance of IS classification.

2 MATERIALS AND METHODS

A total of 24 participants were recruited for EEG data acquisition, comprising 14 males and 10 females aged between 20 and 26 years. The protocol has been approved by Politecnico di Milano Research Ethical Committee (Opinion 80 n. 29/2021). This task primarily involved the mental imaging of twelve distinct words, carefully selected to enable unique encoding based on semantic and grammatical *Properties*. The *Properties* under consideration included word category (classified into three categories: Motion, Space Exploration and Unpleasantness), word length (categorized as short or long), and the presence of double letters in the word. The 12 words and their properties are reported in Table 2.

2.1 Experimental Protocol

Electroencephalogram (EEG) data were collected from 24 participants using the 64-channel EEG system BE Plus LTM by EBNeuro (Florence, Italy), with electrodes positioned across the scalp following the 10-20 International System.

The acquisition protocol is structured into 5 sessions, with each word presented once in a randomized sequence. Each word is associated with 4 distinct phases, (see Fig.1): an initial fixation period, the presentation of the word as a visual cue, a subsequent thinking period, and last, a rest period. The entire protocol lasts about 18 minutes, after which participants are required to fill in an online form to gather feedback on their perceptions and experiences.

2.2 Preprocessing and Feature Extraction

Following data acquisition, rigorous preprocessing steps were applied to the EEG recordings. These steps were crucial for mitigating artifacts effects and safeguarding the data's integrity.

In the preprocessing phase, entirely performed using EEGLAB (Delorme and Makeig, 2004), the EEG signals were subjected to frequency filtering, restricting their range to 0.05Hz to 120Hz. Notch filters were thoughtfully applied at 50Hz and 100Hz to effectively eliminate interference lines and their harmonics. Visual inspection was then meticulously employed to identify and exclude any corrupted channels from the dataset.

To further enhance data quality, the common average reference (CAR) was computed, which aids in minimizing common noise across all electrodes. The signals were then downsampled from their original 512Hz to 256Hz to enhance computational efficiency for the subsequent procedures.

Finally we applied Independent Component Analysis (ICA). This technique played a pivotal role in removing artifacts, particularly those arising from blinking. After ICA, any removed channels were reconstructed using spherical interpolation to ensure the spatial coherence of the EEG data.

Such thorough preprocessing stage ensured that EEG data artifacts -interference, and corruption- be minimized for subsequent analysis with the utmost data quality and reliability.

After preprocessing, the EEG signals underwent an epoching process, in which a 4-second window

Table 2: Italian selected words divided by Properties (columns), in brackets their English translation.

	Motion	Space Exploration	Unpleasantness
1	Resistenza (Endurance)	Astronave (Spaceship)	Delusione (Delusion)
2	Scatto (Sprint)	Razzo (Rocket)	Guerra (War)
3	Allenamento (Training)	Navicella (Shuttle)	Depressione (Depression)
4	Sforzo (Effort)	Cosmo (Cosmos)	Odio (Hate)

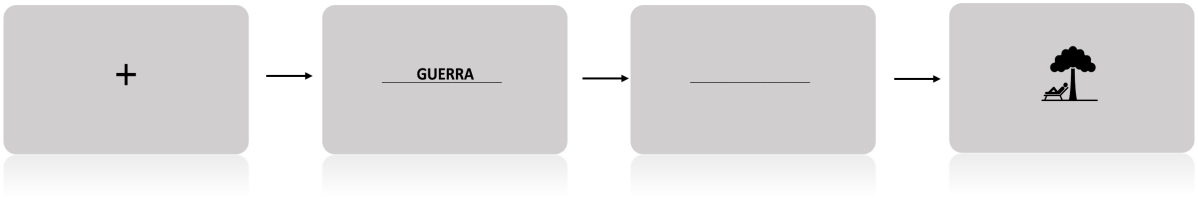

Figure 1: Experimental Protocol Pipeline.

was selected.

A comprehensive set of methods for feature extraction was meticulously employed to uncover intricate patterns within the data. The extracted features included three groups of variables: statistical (Mean, Autoregressive Coefficients, Variance...), spectral (band powers, spectral skewness...) and fractal domain features (Hurst Exponent, Fractal Dimension...) amounting to a total of 34 variables. These features were consistently derived from each channel of the signal in the time domain, as well as from the information obtained through the Discrete Wavelet Transform (DWT) and the Empirical Mode Decomposition (EMD), from which we extracted five Intrinsic Mode Functions (IMFs). The ”Db4” wavelet served as the mother wavelet for the DWT, and seven details were computed, from which the features were extracted.

Furthermore, to ensure comparability, all features were normalized with respect to a baseline, 4-second window preceding the start of the protocol.

This approach ensures that a consistent set of features is extracted from different perspectives, thus capturing a diverse range of information from the EEG data.

Once the dataset is prepared and processed, it assumes a shape of 60x26,963 [12 words*5 repetitions] x [61 channels*(1 time-domain signal+7 DWT details+5 IMFs)*34 features + target].

2.3 Classification

In this section, we will delve into the feature selection and classification procedures conducted on the training set, which comprises 80% of the original dataset. The remaining 20% is exclusively reserved for the blind test and subsequent evaluation.

2.3.1 Features Selection

The extracted features underwent a feature selection process, aiming to identify the most promising ones and their utility in classifying each *Property*. Additionally, topographic map of an EEG (the so called TopoPlot) are employed to visually analyze the most relevant channels.

It is important to note that three independent classifications, one for each *Property* as the target, are conducted using the same.

Firstly, the dataset undergoes a data quality check to identify missing values. Successively, the Kruskal-Wallis test (divergence from normality was assessed through Kolmogorov-Smirnov test) is applied to each variable to search for features whose medians show the greatest difference when stratifying by the target variable. Only features with $p<0.05$ are retained.

Next, each feature is standardized using a standard scaler, and a correlation analysis is performed to eliminate variables that exhibit strong correlations (>0.8) with the others preference is given to variables based on the H-statistics of the Kruskal-Wallis test, used to measure the diversity of the variable's distribution: variables with lower H-statistics are discarded first.

The final step involves the application of Minimum Redundancy Maximum Relevance (MRMR) feature selection algorithm. This step aims to identify variables that are less redundant and more relevant for the classification task. Importantly, all the presented procedures are initially performed on the training dataset, consisting in 80% of the original dataset.

The selection of the number of variables is carried out through an assessment involving five models, including Linear Discriminant Analysis (LDA), Decision Trees, Random Forest, Support Vector Machines (SVM), and Logistic Regression. By evaluating the mean accuracy from cross-validation across

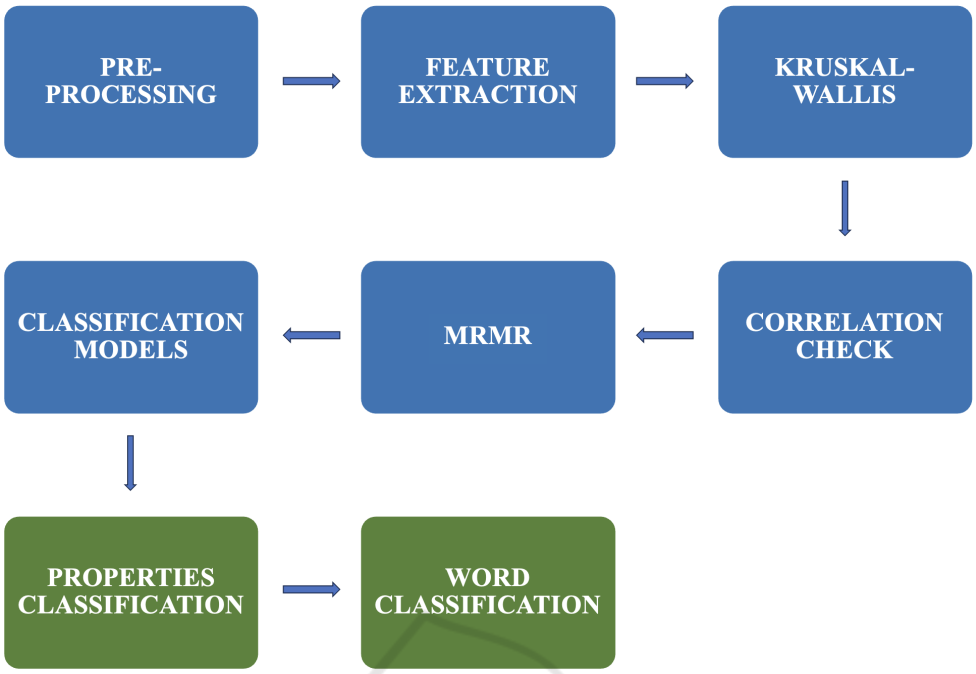

Figure 2: Classification Pipeline.

these models, the number K of features to retain is chosen. The range considered for K typically spans from 1 to 4, ensuring a favorable sample-to-feature ratio of about 10.

2.3.2 Classification Models

Following the selection of K features, Multilayer Perceptrons (MLP), k-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forest, Naive Bayes, and Logistic Regression Models are trained with hyperparameter tuning in order to predict each word Property. The best-performing model for the classification of each Property, as determined by the highest validation accuracy, is then selected to carry out the final classification.

Furthermore, an ensemble model is employed for classification. This ensemble model is a Linear Discriminant Analysis (LDA) that takes as input the predicted probabilities generated by the three best individual models. The final prediction is made by considering either the prediction from the individual model or the ensemble model, depending on their respective validation accuracies. This approach ensures that the final classification is based on the most reliable and accurate model, as determined by rigorous evaluation.

Once the three predictions of the individual *Properties* for each word have been generated, they can be collectively analyzed to decode the original word. This decoding process is facilitated by the unique encoding of each word based on its distinctive properties.

In addition, another metric known as Word Distance (D) was defined. It provides a unique perspective, quantifying the distance between each word and its predicted counterparts in the Property space.

$$D = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{n}(\delta(a_{ij},b_{ij}))\right) \quad (1)$$

Where:

- D represents the distance metric.
- N is the total number of words in the test set.
- n is the number of properties being evaluated (3 properties in our case: Semantic Category, Length, Doubles).
- a_{ij} represents the value of the i-th word for the j-th property in the reference dataset.
- b_{ij} represents the value of the i-th word for the j-th property in the predicted dataset.
- $\delta(a_{ij},b_{ij})$ is a function that returns 0 if a_{ij} is equal to b_{ij} and 1 if they are different.

The formula calculates the average differences between the properties of words in the reference dataset and the predicted words, with D representing the average distance across all words in the test set.

3 RESULTS

The statistics obtained from the online form showed no significant correlations with the accuracies of individual *Properties*. The average accuracy across participants for the three *Properties* is displayed in Table 3.

The average accuracy, obtained by aggregating predictions across participants, was 11.8% [8.1%; 15.3%] for a 12-class problem, against an 8.33% random chance. The confusion matrix is shown in Fig.3

Furthermore, in this study, the average word distance between classified words across subjects was 1.54, in contrast to the random distance of 1.67.

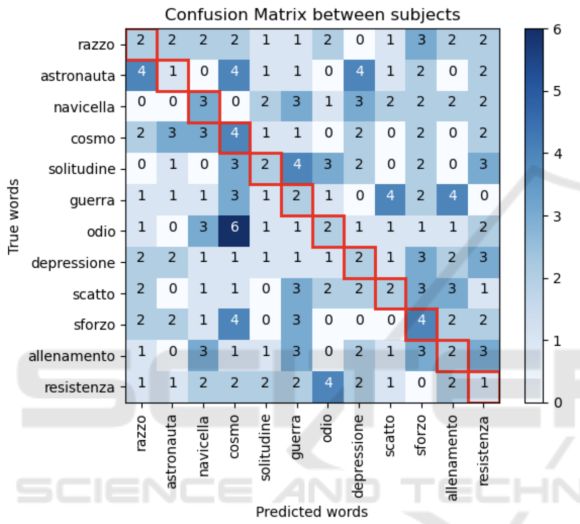

Figure 3: Confusion Matrix of the 12-class problem.

For the ”Category” property the most relevant variables among participants are: Lyapunov exponents, mean, Autoregressive coefficients, and Hjorth coefficients. However, for both binary problems (i.e., ”Lenght” and ”Double” properties), Autoregressive coefficients were by far the most selected ones.

It was observed that the best results on the test dataset were achieved when considering variables extracted from all sources or solely from those derived from DWT or IMF. Considering only time-domain variables did not yield favorable results.

The most active channels across participants were, for the ”Category” property, those related to the temporal lobes in both hemispheres. For ”Length,” it was the channel on the mid-line sagittal plane, and for ”Doubles,” it was channels related to the temporal and central lobes in both hemispheres. A visual comparison between the pertinent channels for ”Category” and ”Length” can be drawn by respectively examining Figure 4.

4 DISCUSSION

The results of this study provide valuable insights into the classification of word *Properties* from EEG data during language processing. The protocol involved five sessions, each with four distinct phases. The subsequent preprocessing ensured the quality and integrity of the data, making it suitable for detailed analysis.

One notable finding is that the statistics from the online form were not correlated with the accuracies of individual *Properties*. This suggests that participants' subjective perceptions and experiences during the experiment did not align with the objective accuracy of the property classifications. Such a lack of correlation highlights the complexity of the relationship between cognitive processes and self-reported experiences.

the average accuracies achieved by the developed models on data extracted from each participant in classifying *Properties* varied across the different *Properties*. For the ”Category” property, the accuracy was slightly above chance level (33%), indicating some ability to distinguish between word categories. In contrast, the binary *Properties* ”Doubles” and ”Length” displayed accuracies of 57.3% and 56.1%, respectively. These results highlight the inherent challenges associated with multi-class classification tasks compared to binary classification.

It is important to note that these lower accuracies, particularly in the ”Category” property, are attributed to the utilization of a dataset with limited recordings, thus impacting on the overall classification performance.

However, the fact that all average accuracies are above the random chance level supports the possibility of classifying the word *Properties*. Despite the relatively low accuracy rates, the models points toward the ability to discriminate between them.

Aggregating predictions from participants yielded a mean accuracy of 11.8%. This low value can be attributed to the increased complexity of the problem (12 classes so 8.33% random chance) and to the low number of repetitions of each word. Nevertheless, achieving an accuracy above random chance in a 12-class problem demonstrates the potential of EEG-based classification of semantic and grammatical words *Properties*.

It is crucial to highlight that this form of word encoding has paved the way for the introduction of the distance metric *D*. This represents one of the most innovative aspects of our study. In the past, each word was treated as entirely unique, devoid of shared *Properties*, resulting in equidistant relationships between all words. However, our approach has transitioned

Table 3: Classification Accuracy, in square brackets the 95% confidence intervals.

	Property	Mean accuracy over subjects	Random Accuracy
1	Semantic Category (3-class)	35.2% [30.5%; 39.6%]	33.3%
2	Doubles (Binary)	57.2% [50.6%; 61.0%]	50.0%
3	Length (Binary)	55.8% [50.1%; 63.3%]	50.0%

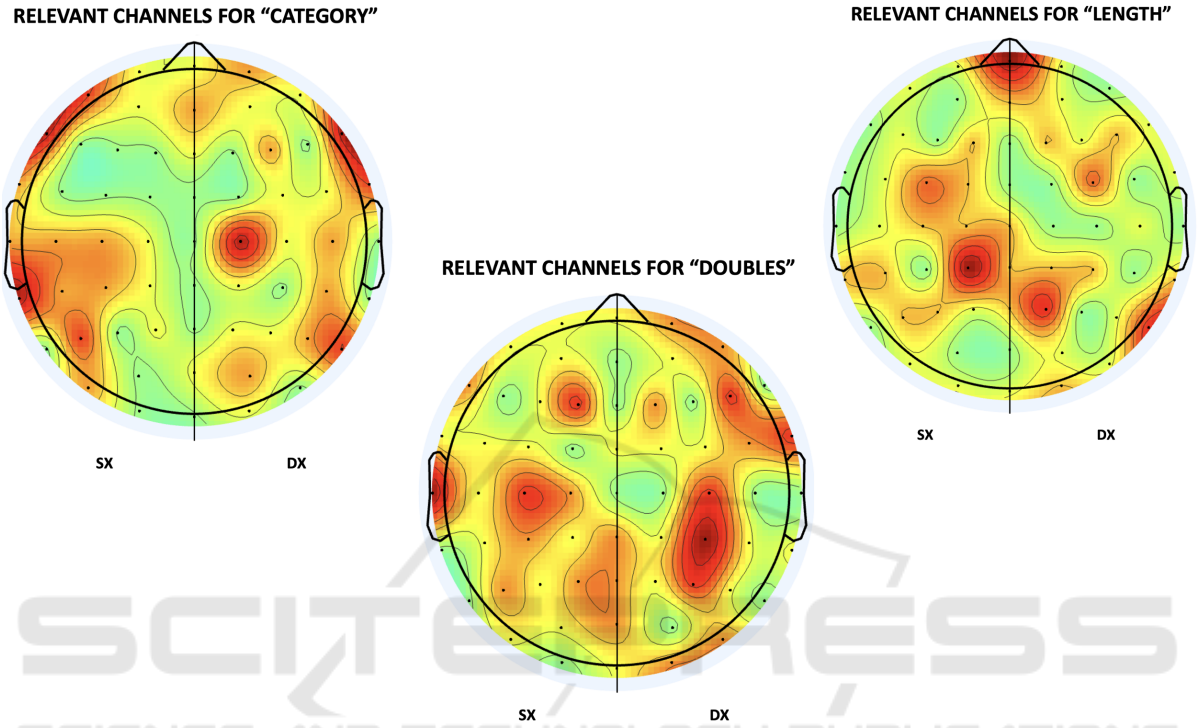

Figure 4: Comparative TopoPlots between the most relevant channels in the three *Properties*.

from one-hot encoding and a non-dense space to a more densely packed space, where words are encoded based on their semantic and grammatical *Properties*. Importantly, our results prove that these words can be effectively decoded.

Analyzing the relevance of variables across participants revealed interesting patterns. For the "Category" property, Lyapunov exponent, Mean, Autoregressive coefficients, and Hjorth coefficients were all found to be equally relevant. In contrast, Autoregressive coefficients dominated as the most relevant variables for both binary problems. This suggests that certain EEG features, such as Autoregressive coefficients, consistently play a critical role in classifying cognitive *Properties*, regardless of the specific property.

Furthermore, considering variables from different sources—time domain, DWT, and IMF—proved to be advantageous for achieving better classification results. This underscores the importance of incorporating a wide range of EEG features to capture the intricate dynamics of cognitive processes during language tasks.

Channel-wise analysis highlighted specific brain regions associated with each property. For the "Category" property, temporal lobes in both hemispheres were notably active. "Length" exhibited activity in the mid-line sagittal plane, and "Doubles" involved channels related to the temporal and central lobes in both hemispheres. These findings suggest the involvement of distinct brain regions in processing different linguistic aspects, aligning with existing neurocognitive theories (Price, 2012).

5 CONCLUSION AND FUTURE DEVELOPMENTS

In conclusion, our study on Imagined Speech Detection has marked a significant step forward in the challenging task of classifying imagined word *Properties* using EEG data.

It has not only provided valuable insights but has

also introduced innovative concepts like word *Properties* and distance metrics into the realm of Imagined Speech detection that can open up new avenues for research and offers a novel direction for enhancing the accuracy and robustness of future classification algorithms.

Additionally, our exploration into the neurophysiological aspects of word perception has revealed intriguing insights. EEG features extracted from Discrete Wavelet Transform (DWT) and Intrinsic Mode Functions (IMF) have shown promise in classification, emphasizing the importance of considering variables from multiple sources.

Our channel-wise analysis has shed light on the brain regions associated with different *Properties*, emphasizing the multifaceted nature of word perception. These insights contribute to our understanding of the neural basis of language.

Expanding our data collection protocol to gather more extensive and diverse datasets is a crucial step. These richer datasets will empower machine learning models with a deeper understanding of imagined speech recognition.

Furthermore, an exciting avenue for future research lies in the fusion of Natural Language Processing (NLP) techniques with the novel concepts introduced in this study, such as word distance and the comprehensive characterization of words based on their semantic, grammatical, and phonetic properties.

In the realm of semantics, an intriguing exploration involves distinguishing synonyms to uncover if it's feasible to capture the precise meaning of words, thereby refining the concept of semantic category.

Lastly, our study's successful achievement in discrimination word length as a classification task opens a new dimension for research: a regression to predict the exact number of letters in a word. This represents a more intricate and challenging facet of Imagined Speech Detection, offering exciting possibilities for future investigations.

In conclusion, we envision a future marked by exciting developments in Imagined Speech Detection where the convergence of artificial intelligence, neuroscience, and linguistics offers immense promise. In this context, our study provides a solid foundation for further exploration towards a more comprehensive understanding of Imagined Speech Detection. The path ahead promises deeper insights, increased functionality, and broader applications at the intersection of human cognition, language, and technology.

REFERENCES

Chau, A. R. S. R. T. A. B. T. (2017). Eeg classification of covert speech using regularized neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Dipti Pawar, S. D. (2020). Multiclass covert speech classificationusing extreme learning machine. *Biomedical Engineering Letters*.

Dipti Pawar, S. D. (2023). Eeg-based covert speech decoding using random rotation extreme learning machine ensemble for intuitive bci communication.

Fitriah, N., Zakaria, H., and Rajab, T. L. E. (2022). Eeg-based silent speech interface and its challenges: A survey. *International Journal of Advanced Computer Science and Applications*, 13(11).

Gu, X., Cao, Z., Jolfaei, A., Xu, P., Wu, D., Jung, T. P., and Lin, C. T. (2021). EEG-Based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and Their Applications. *IEEE/ACM Trans Comput Biol Bioinform*, 18(5):1645–1666.

Laureys (2005). The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless?

Lopez-Bernal, D., Balderas, D., Ponce, P., and Molina, A. (2022). A State-of-the-Art Review of EEG-Based Imagined Speech Decoding. *Front Hum Neurosci*, 16:867281.

Martin (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. *Frontiers*.

Nguyen, C. H., Karavas, G. K., and Artemiadis, P. (2018). Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. *J Neural Eng*, 15(1):016002.

Panachakel, J. T. and Ramakrishnan, A. G. (2021). Decoding Covert Speech From EEG-A Comprehensive Review. *Front Neurosci*, 15:642251.

Price, C. J. (2012). A review and synthesis of the first 20 years of pet and fmri studies of heard speech, spoken language and reading. *Neuropsychologia*, 50(11):2625–2641.

Qureshi, M. N. I., Min, B., Park, H.-j., Cho, D., Choi, W., and Lee, B. (2018). Multiclass classification of word imagination speech with hybrid connectivity features. *IEEE Transactions on Biomedical Engineering*, 65(10):2168–2177.

Yi, G., Wang, J., Bian, H., Han, C., Deng, B., Wei, X., and Li, H. (2013). Multi-scale order recurrence quantification analysis of EEG signals evoked by manual acupuncture in healthy subjects. *Cogn Neurodyn*, 7(1):79–88.