POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

A Word Recognition Paradigm Through EEG Analysis: Imagined Speech Classification

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Author: Francesco Iacomi

Advisor: Prof. Barbieri Riccardo, Prof. Mainardi Luca

Co-advisor: Ing. Farabbi Andrea, Ing. Mollura Maximiliano, Ing. Polo Edoardo Maria

Academic year: 2022-2023

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 movements. 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.
The electroencephalogram, (EEG), is the chosen neurophysiological signal for this study. Its selection is rooted in its non-invasiveness, exceptional temporal resolution, and the minimal instrumentation required for data acquisition[1]. 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[2].

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 [3]. Researchers have employed various IS prompts, including visual and auditory cues, and have employed diverse techniques for feature extraction, such as statistical analysis [4], frequency domain analysis [5], and spatial analysis [3]. Classification, the final step, has seen the application of both machine learning and deep learning models, each with its strengths and areas of application.
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. 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[4].
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, see Table 1. The goal is twofold: a) to determine which characteristics of an imagined word can be effectively discriminated and b) expand the vocabulary space for IS decoding. Additionally, the study aims at streamlining feature extraction processes, identify informa-

tive EEG channels, and enhance the overall efficiency and precision of IS classification.

2. Material 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, during which they engaged in the Imagined Speech task. The protocol has been approved by the Politecnico ethical committee 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), word length (categorized as short or long), and the presence of double letters in the word.

	Motion	Space	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)

Table 1: Italian selected words, in the bracket their English translation

2.1. Experimental Protocol

Electroencephalogram (EEG) data were collected from 24 participant using a 64-channel EEG system, 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 finally, a rest period. The entire protocol lasts about 18 minutes, after which participants are required to complete an online form.

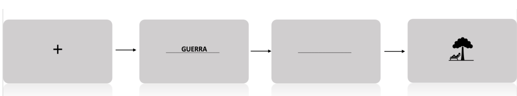


Figure 1: Protocol Visualization

2.2. Preprocessing and Feature Extraction

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

In the preprocessing phase, entirely performed using EEGLAB [6], the EEG signals underwent a series of essential steps. Initially, they 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.

The pinnacle of the preprocessing endeavor involved the application of Independent Component Analysis (ICA). This advanced technique played a pivotal role in removing artifacts, particularly those arising from blinking. Post-ICA, any removed channels were meticulously reconstructed using spherical interpolation to ensure the spatial coherence of the EEG data.

This thorough preprocessing protocol ensured that the EEG data were devoid of artifacts, interference, and corruption, rendering them ready for subsequent analysis with the utmost data quality and reliability.

Following the meticulous preprocessing steps, the EEG signals underwent an epoching process, marking the commencement of the feature extraction procedure. A diverse array of feature extraction methods was thoughtfully applied to uncover intricate patterns within the data. These methods encompassed Autoregressive coefficients, Lyapunov exponents, fractal dimension, Hjorth coefficients, Skewness, Kurtosis, and various others.

The same features, normalized by a baseline, are extracted from each channel of the signal in the time domain, as well as from the details obtained through the Discrete Wavelet Transform (DWT) and the Intrinsic Mode Functions

(IMF). 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.
The extracted features will undergo a feature selection process, as detailed in the subsequent section. This selection aims to identify the most promising feature sources and their utility in classifying each *Property*. Additionally, TopoPlots will be employed to visually analyze the most relevant channels derived from the selected features.

2.3. Classification

Once the dataset is prepared and processed, it assumes a shape of 60x27.084 (12 words*5 repetitions x number of variables). The classification pipeline can then commence. It is important to note that three independent classifications of the *Properties* are conducted using the same dataset differentiated only by the unique property-specific target.
Firstly, the dataset undergoes a data quality check to identify null columns and missing values. Following this, the feature selection phase begins. Initially, the Kruskal-Wallis test is applied to each variable to assess their statistical significance concerning the target variable. Only variables with a p-value from the test lower than 0.05 are retained.
Next, the dataset is standardized using a standard scaler, and a correlation analysis is executed to eliminate variables that exhibit strong correlations with others. The threshold for correlation removal is set at 0.8, with preference 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. This step aims to identify variables that are less redundant and more relevant for the classification task. Importantly, all these procedures are initially performed on the training dataset. Then, using only the information gleaned from the training set, the same procedures are applied to the test dataset to ensure consistency.
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 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.
Following the selection of K features, additional models are explored with more aggressive hyperparameter tuning. This includes models such as Multilayer Perceptrons (MLP) and k-Nearest Neighbors (KNN). The best-performing model, 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 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 that can be introduced known as Word Distance (D), provides a unique perspective. In fact it specifically quantifies the distance between each word and its predicted counterparts in the Property space.

3. Results

The statistics obtained from the online form showed no significant correlations with the accuracies of individual *Properties* (p-value less than 0.05). The average accuracy across participants for the three-class "Category" property was 34.4%, while for binary "Doubles" and "Length," it was 57.3% and 56.1%, respectively. The average accuracy obtained by aggregating predictions across participants was 11.8% for a 12-class problem, Fig. 2, against an 8.33%

random chance.
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.

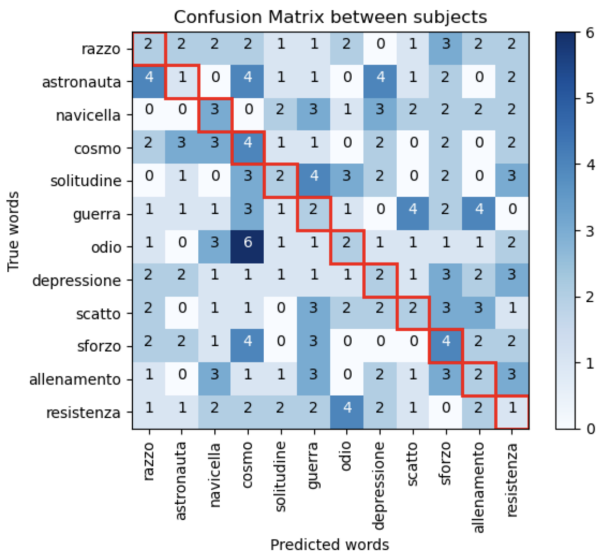


Figure 2: Confusion Matrix for the 12-class problem

Regarding the most relevant variables among participants, for the "Category" property, Lyapunov exponent, mean, Autoregressive coefficients, and Hjorth coefficients were all equally relevant. However, for both binary problems, Autoregressive coefficients were by far the most common.
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 variables extracted from the time domain 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 3.

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 underscores the complexity of the relationship between cognitive processes and self-reported experiences.
The average accuracies achieved by participants 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, especially in the "Category" property, are influenced by the use of a patient-specific dataset with limited data. This limitation could have impacted the overall classification performance.
However, the fact that all the mean accuracies are above the random chance level confirms the possibility of classifying the word *Properties*. Despite the relatively low accuracy rates, the models demonstrate the capability 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 in previous studies, devoid of shared *Properties*, resulting in equidistant relationships between all words. However, our approach has transitioned 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, it has been demonstrated that these 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.

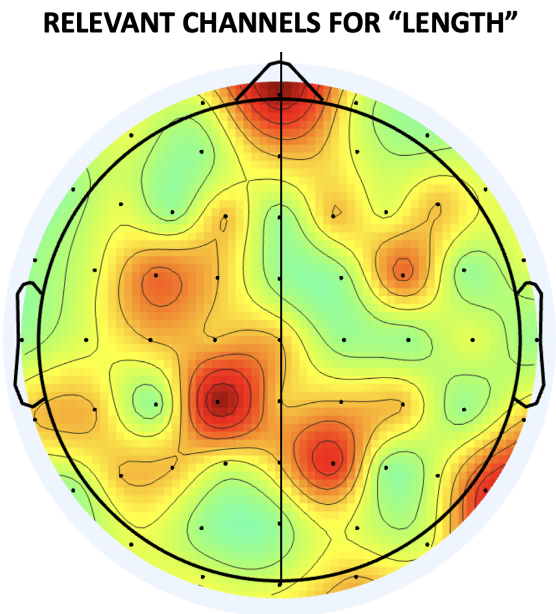
Channel-wise analysis highlighted specific brain regions associated with each property. For the "Category" property, temporal lobes in both hemispheres were notably active. "Length" exhibited activity in the mid-line sagittal plane, and "Doubles" involved channels related to the temporal and central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

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



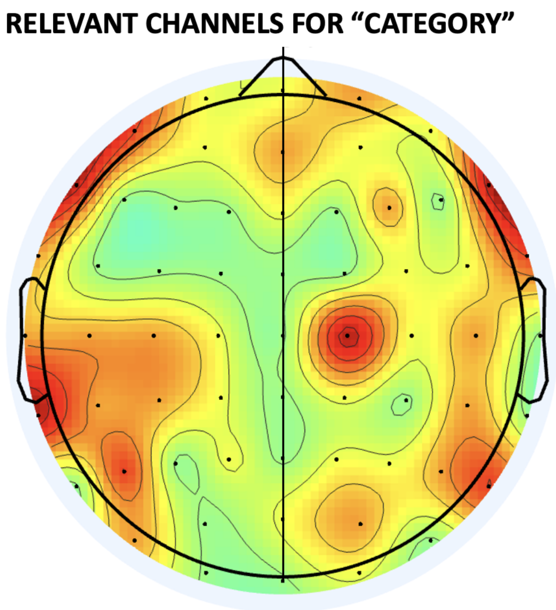


Figure 3: Comparative TopoPlots between most relevant "Category" and "Length" channels

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.
Additionally, our results bring evidence that there is a significant potential in investigating the discrimination between articles, nouns, and verbs, offering a deeper understanding of how grammatical properties can be decoded from EEG signals.
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.
As we conclude this study, we envision a future marked by exciting developments in Imagined Speech Detection. The convergence of artificial intelligence, neuroscience, and linguistics offers immense promise, and our study provides a solid foundation for further exploration.
In this era of rapid advancements, our study serves as a foundation for 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

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

[2] Laureys et. al. The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless? 2005.

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

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

[5] Nilam Fitriah, Hasballah Zakaria, and Tati Latifah Erawati Rajab. Eeg-based silent speech interface and its challenges: A survey. *International Journal of Advanced Computer Science and Applications*, 13(11), 2022.

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