*Experiment Pre-Registration Proposal*: Speaking with the Brain: Decoding of Imagined Speech Spanish Vowels from EEG data using Deep Learning Models.

*Short Introduction*

Language is a fundamental tool for communication among individuals (Krishna et al., 2021). However, certain diseases pose significant challenges to communication, rendering it difficult or even impossible. Conditions such as amyotrophic lateral sclerosis (ALS), advanced stages of multiple sclerosis, or cerebrovascular infarctions affecting brainstem regions can disrupt the neural pathways responsible for language production (Coretto, Gareis & Rufiner, 2017). Specifically, individuals with ALS often lose the ability to move while their cognitive functions remain intact. Those suffering from this locked-in state (Bauer et al., 1979) report that the loss of the ability to communicate with others is their greatest concern (Wandelt et al., 2024).

Alternative speech systems have been developed to facilitate communication for individuals with limited mobility, relying on residual movements associated with linguistic expression, such as head or eye movements, to enable word or letter selection via a cursor (Koch-Fager et al., 2019). However, these systems are unnatural and have a very low information transmission capacity, resulting in a linguistic production rate of approximately 5-6 words per minute (Herff et al., 2015).

One promising avenue for addressing these challenges lies in the synthesis of language through Brain-to-Computer Interfaces (BCIs). BCIs are devices that can decode non-acoustic biosignals associated with language production, enabling individuals affected by various conditions to communicate through text, voice synthesizers, or cursor control for selection purposes (Wolpaw et al., 2002). Traditional BCIs have focused on motor imagery, event-related potentials (ERPs), and steady-state visually evoked potentials for language synthesis (McFarland et al., 2000; Farwell & Donchin, 1988; Sutter, 1992).

Nowadays, due to improvements in EEG recordings and the development of powerful Machine Learning (ML) systems, the possibility of directly decoding language from brain waves is emerging. This advancement is enabling the development of systems that provide individuals unable to produce language with a natural form of communication through imagined speech (Willet et al., 2023). Imagined speech involves the silent verbalization of phonemes, words, or sentences without the activation of facial muscles, offering a non-invasive and precise insight into the cognitive processes underlying language production. Imagined speech involves the silent verbalization of phonemes, words, or sentences without the activation of facial muscles, providing a non-invasive and precise insight into the cognitive processes underlying language production. Some studies have investigated the neural correlates of imagined speech, drawing parallels with the mechanisms of overt speech (LaRocco et al., 2023).

The imagined verbalization of phonemes, words, or even entire sentences solely through thought has been extensively studied using various techniques such as fNIRS (functional near-infrared spectroscopy), ECoG (electrocorticography), and EEG (electroencephalogram), with EEG being the focus of this study. Advances in EEG data collection and subsequent preprocessing have enabled us, through Machine Learning and Deep Learning (DL) models, to decode what a person is trying to say in their imagination (Koizumi et al., 2018; Panachakel & Ramakrishnan, 2021).

Recently, significant progress has been made in overcoming obstacles in the phoneme decoding of imagined speech, such as the high signal-to-noise ratio that previously hindered the decoding of less distinguishable phonemes. These advancements have been achieved through improvements in EEG data preprocessing, enhanced feature extraction methods, and advancements in classification models (Nitta et al., 2023; LaRocco et al., 2023).

*Objectives*

This study has three primary objectives and one secondary objective. The preliminary objective is to develop a robust DL model capable of decoding the linguistic content within EEG recordings. Initially, we aim to achieve above-chance decoding of the three main conditions: rest, overt speech, and covert speech.

Building on the results of the first objective, our second objective is to extend the model's decoding capability to classify Spanish vowels. This initiative represents the second attempt to decode vowels in the Spanish language. The first attempt, conducted by Coretto et al. (2017), yielded poor results, with accuracy rates near chance level: 22.72% for the Random Forest model and 21.94% for the Support Vector Machine model. This objective aims to improve upon these results and enhance reliability by including the overt speech condition, as the previous study only evaluated covert speech. Additionally, this objective encompasses the future goal of successfully decoding all Spanish phonemes. This initiative represents an unprecedented endeavor within the realm of Spanish linguistics. To the best of our knowledge, only one study in a different language (English) has achieved high accuracy rates in decoding various phonemes using EEG data (LaRocco et al., 2023). However, this pursuit will be continued in subsequent research, as the implementation of high-complexity models, such as those utilized in this study, requires a substantial volume of data that is not currently available as of the submission date of this work.

Third, another primary objective is the establishment of an openly accessible database in Spanish, fostering collaborative research efforts and facilitating methodological and model comparisons within the scientific community. This aims to cultivate a scholarly dialogue aimed at enhancing the accuracy and efficacy of EEG-based speech decoding methodologies.

Furthermore, a secondary objective is to contribute to the scholarly discourse surrounding the utilization of DL frameworks in the task of decoding imagined speech. This stems from an observed disparity within the neuroscientific community, where the adoption of more rudimentary machine learning (ML) models prevails, thereby constraining clinical applicability (Shah et al., 2019).

*Methods*

*Sampling Plan:* Due to the extended duration of the experiment, which will allow us to collect a substantial amount of data, and based on the existing literature on imagined speech decoding in EEG, a minimum of 10 participants will be recruited. This number is higher than that typically used in other experiments (e.g., Coretto et al., 2017; Koizumi et al., 2018; Abdulghani et al., 2023; Nitta et al., 2023; LaRocco et al., 2023), but given the complexity of the DL models employed, we believe a larger sample size is necessary.

*Participants:* The participants will primarily be students from the University of Granada (UGR). They will be administered the Edinburgh Handedness Inventory (Oldfield, 1971) to determine manual dominance, as the selection of EEG channels will prioritize the language area in the left hemisphere. Participants must have 20/20 vision or corrected vision. Informed consent will be obtained from all participants. They will be compensated with €40 for their participation in the experiment.

*Materials:* The stimuli presented in the first block will include all the phonemes of the Spanish language (95 in total) plus five silences. The phonemes will be presented both auditorily and visually. In the second block, 30 pseudowords and 60 words from six different semantic categories will be presented: 1- Kitchen utensils; 2- Body parts; 3- Food; 4- Musical instruments; 5- Clothes; 6- Animals. For each semantic category, half of the words will be high frequency (>10/1,000,000) and the other half low frequency (<5/1,000,000). This will be calculated according to the CORPES XXI open database from the Royal Spanish Academy (Corpus del Español del Siglo XXI [CORPES XXI], October 2013). The presentation of words will be conducted in three different ways: auditory, pictorial, and visual.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamenteBoth the first and second blocks will be repeated twice.

**Un conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza media**

**Figure 1.** a) Order of presentation of the first block. b) Order of presentation of the second block.

*Presentation Procedure:* The presentation procedure for both Block One and Block Two is detailed in Figure 1. The duration of the experiment, including helmet placement and gel preparation beforehand, will be 3 hours per participant.

*EEG Recordings:* EEG data will be obtained using a 64-channel system mounted on a cap (actiCAP snap, Brain Products) and a computer running Brain Vision Recorder (version 1.20.0601), Lab Stream View, and ActiChamp. Electrode placement will follow the 10-20 system. Efforts will be made to maintain impedances below 5 kΩ, with particular emphasis on areas of interest (Left Temporal Hemisphere and Motor Areas). The digitized signal will have a sampling rate of 1000 Hz, with an earlobe electrode used as the reference electrode.

*EEG Preprocessing:* The MNE Python library (Gramfort et al., 2013) will be utilized for data analysis and preprocessing using the Visual Studio Code editor. The data will be segmented into epochs of 1000 ms, starting 200 ms before Speech Onset.

Escala de tiempo

Descripción generada automáticamente con confianza mediaThe following steps will be followed for data preprocessing: 1- The sample will be resampled to 250 Hz, based on existing literature; 2- A notch filter at 50 Hz and its harmonics (100 Hz and 150 Hz) will be applied; 3- A bandpass filter between 1 and 125 Hz will be applied; 4- Independent Component Analysis will be conducted to discard noise sources that may affect the signal (e.g., eye blinks); 5- Electrodes of interest will be selected, including those near the Broca and Wernicke areas. Motor areas electrodes will also be included, following the recommendations of Li et al. (2021) for accurate decoding of imagined speech; 6- Time-Frequency analysis will be performed to determine frequency bands showing increased activity for subsequent decoding; 7-Within Time-Frequency analysis, the Fourier Transform with a Morlet wavelet mother function (TFM; Min et al., 2016) will be applied. Additionally, inter-trial coherence will be evaluated to assess temporal neural coherence across different frequency activities; 8- Finally, the selected channels will be evaluated using the power.plot function (picks=), followed by applying the logratio function to adjust results to the baseline.

**Figure 2.** Architecture of the CNN deep learning model and the process flow

*Deep Learning Models:* Two DL models have been pre-designed for this study. The first model is a one-dimensional convolutional neural network (CNN 1D; see Figure 2), optimized for extracting features from input data. This model excels in capturing intricate temporal dependencies within sequential data, as well as in feature extraction (Roy et al. 2019; Panachakel & Ramakrishnan, 2021).

The second model combines the CNN 1D architecture with a recurrent neural network model known as the Gated Recurrent Unit (GRU; Figure 3). This hybrid model enhances performance by integrating the CNN's ability to learn hierarchical features with the GRU's capability to handle sequential dependencies, making it adept at understanding complex temporal patterns in the data.

In both models, the hyperparameters to be applied, as well as the learning rate for each, are yet to be established.

Escala de tiempo

Descripción generada automáticamente

**Figure 3.** Architecture of the CNN + GRU deep learning model and simplified process flow

*Stadistical Analysis:* Classification accuracy of the three classes (overt, covert, and rest) will be calculated across 10 participants to assess significance using Friedman's test, a non-parametric method akin to a balanced two-way ANOVA (Lee et al., 2020). Confidence intervals will also be computed to validate the model's performance across these classes (Muller-Putz et al., 2008). Similar analyses will be conducted for the five vowels in the Spanish language.

*References*

Abdulghani, M. M., Walters, W. L., & Abed, K. H. (2023). Imagined Speech Classification Using EEG and Deep Learning. *Bioengineering, 10*(6), 649.

Bauer, G., Gerstenbrand, F., Rumpl, E. (1979). Varieties of the locked-in syndrome. *Journal of Neurology, 221*(2): 77–91.

Coretto, G. A. P., Gareis, I. E., & Rufiner, H. L. (2017, January). Open access database of EEG signals recorded during imagined speech*. In 12th International Symposium on Medical Information Processing and Analysis* (Vol. 10160, p. 1016002). SPIE.

Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE transactions on rehabilitation engineering*, *8*(2), 174-179.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, *70*(6), 510-523.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L. & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience, 7*, 70133.

Herff, C., Heger, D., De Pesters, A., Telaar, D., Brunner, P., Schalk, G., & Schultz, T. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in neuroscience*, *8*, 141498.

Koch-Fager, S., Fried-Oken, M., Jakobs, T., & Beukelman, D. R. (2019). New and emerging access technologies for adults with complex communication needs and severe motor impairments: State of the science. *Augmentative and Alternative Communication*, *35*(1), 13-25.

Koizumi, K., Ueda, K., & Nakao, M. (2018, July). Development of a cognitive brain-machine interface based on a visual imagery method. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1062-1065). IEEE.

Krishna, G., Tran, C., Carnahan, M., & Tewfik, A. H. (2021, May). Advancing speech synthesis using EEG. *In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)* (pp. 199-204). IEEE.

LaRocco, J., Tahmina, Q., Lecian, S., Moore, J., Helbig, C., & Gupta, S. (2023). Evaluation of an English language phoneme-based imagined speech brain computer interface with low-cost electroencephalography. *Frontiers in Neuroinformatics*, *17*.

Lee, S. H., Lee, M., & Lee, S. W. (2020). Neural decoding of imagined speech and visual imagery as intuitive paradigms for BCI communication. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(12), 2647-2659.

Li, F., Chao, W., Li, Y., Fu, B., Ji, Y., Wu, H., & Shi, G. (2021). Decoding imagined speech from EEG signals using hybrid-scale spatial-temporal dilated convolution network. *Journal of neural engineering*, *18*(4), 0460c4.

McFarland, D. J., Miner, L. A., Vaughan, T. M., & Wolpaw, J. R. (2000). Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, *12*, 177-186.

Min, B., Kim, J., Park, H. J., & Lee, B. (2016). Vowel imagery decoding toward silent speech BCI using extreme learning machine with electroencephalogram. *BioMed research international*, *2016*.

Nitta, T., Horikawa, J., Iribe, Y., Taguchi, R., Katsurada, K., Shinohara, S., & Kawai, G. (2023). Linguistic representation of vowels in speech imagery EEG*. Frontiers in Human Neuroscience, 17*, 1163578.

Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding covert speech from EEG-a comprehensive review. *Frontiers in Neuroscience*, *15*, 642251.

Roy, S., Kiral-Kornek, I., & Harrer, S. (2019). ChronoNet: A deep recurrent neural network for abnormal EEG identification. In *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17* (pp. 47-56). Springer International Publishing.

Shah, U., Alzubaidi, M., Mohsen, F., Abd-Alrazaq, A., Alam, T., & Househ, M. (2022). The role of artificial intelligence in decoding speech from EEG signals: a scoping review. *Sensors*, *22*(18), 6975.

Wandelt, S. K., Bjånes, D. A., Pejsa, K., Lee, B., Liu, C., & Andersen, R. A. (2024). Representation of internal speech by single neurons in human supramarginal gyrus. *Nature human behaviour*, 1-14.

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., Kamdar, F., Glasser, M., Hochberg, L., Druckmann, S., Shenoy, K. & Henderson, J. M. (2023). A high-performance speech neuroprosthesis. *Nature*, *620*(7976), 1031-1036.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical neurophysiology*, *113*(6), 767-791.