

Towards an Integrated Model of Onset Detection and Multiclass Recognition

D.M. Carmona-Peña¹, C.A. Reyes-García¹, A.A Torres-García¹

¹Biosignals Processing and Medical Computing Laboratory, Instituto Nacional de Astrofísica,
Óptica y Electrónica, Puebla, Mexico

E-mail: david.carmona@inaoe.mx

ABSTRACT: Brain-Computer Interfaces (BCIs) based on imagined speech offer a promising alternative for intuitive communication; however, the transition to real-time asynchronous systems is hindered by the "onset detection" problem. This study evaluates the efficacy of Instant Wavelet Energies (IWE), Empirical Mode Decomposition (EMD), and Generalized Hurst Exponent (GHE) for identifying the transition between Idle State Segments (ISS) and Imagined Word Segments (IWS). We analyzed three datasets (DS1, DS2, and DS3) using Support Vector Machines (SVM), Random Forest (RF), k-NN, and Logistic Regression (LogReg). Our results demonstrate that binary intent detection consistently outperforms multiclass word recognition, with the latter remaining a significant challenge. The MCC remains < 0.35 for most configurations, indicating that while these features are effective for intention "gating"; however, they exhibit a lack of the requisite specificity to achieve decisive discrimination.

INTRODUCTION

A Brain-Computer Interface (BCI) is a system that establishes a direct communication channel between bioelectrical brain activity and external devices, enabling the control of assistive devices or prostheses without the intervention of conventional neuromuscular pathways [1]. These interfaces are fundamentally based on the acquisition and processing of biological signals, predominantly electroencephalograms (EEG) to decode cognitive processes. Their application is crucial in the development of assistive technologies aimed at restoring autonomy in individuals with severe motor disabilities.

In the state of the art, BCI systems are primarily categorized into synchronous and asynchronous modes. Most current developments operate under a synchronous paradigm, where the timing and speed of communication are predetermined by the experimental protocol [2, 3]. In these systems, users generate commands exclusively within temporal windows defined by external stimuli, such as P300 evoked potentials or Steady-State Visual Evoked Potentials (SSVEP) [2].

However, this dependence on external stimuli entails significant limitations. It not only requires precise visual stimulation hardware (e.g., LED matrices) that compli-

cates the technical setup [4], but also induces considerable visual fatigue after prolonged periods of use [5].

In contrast, asynchronous (or self-paced) systems offer a more natural interaction, allowing the user to issue commands voluntarily at any time. Since they do not rely on external stimuli, the system must continuously analyze the EEG stream to detect cognitive intent. The critical challenge in these systems is the "onset detection" problem, which consists of accurately discriminating between an active cognitive task and an inactive or resting state [6]. In the literature, this problem has been defined as intentional versus non-intentional control, primarily within the context of motor imagery [7, 8]. To address this asynchrony, various taxonomies of approaches have been developed, ranging from power threshold analysis and state-based models to the use of gating classifiers or evidence accumulation detectors [9–13].

As an alternative, the use of imagined speech has been proposed. While we acknowledge the pioneering advances of Chang et al. [14–16] in using electrocorticography (ECoG) to achieve superior Information Transfer Rates (ITR), our work focuses on accessibility and non-invasiveness. While Chang's methods require high-density surgical implants in patients with specific pathologies, our proposal seeks to optimize the detection of inactivity states through EEG. This enables a low-cost BCI technology with lower clinical risk, which remains an open challenge given the low signal-to-noise ratio inherent to non-invasive methods.

This neuroparadigm in EEG signals ranges from the identification of basic phonetic units such as vowels [17] and phonemes [18] to the decoding of syllables [19] and full words with semantic meaning [20, 21], as well as the identification of emotions transmitted through speech in EEG signals [22]. Despite its potential, current literature lacks a robust approach for detecting imagined speech segments independently of the specific words of interest (e.g., "up", "down").

The contribution of this study lies in the proposal of a two-phase BCI structure. In this scheme, the first stage acts as a crucial stochastic filter that conditions the system's flow: the transition to secondary classification only occurs after effective isolation of the intention segment. This hierarchical model aims to optimize the allocation of computational resources and highly specialized algo-

rhythms, prioritizing the analysis of time windows with substantial informational weight, such as imagined word segments (IWS), and separating them from idle state segments (ISS).

This work builds upon the feature extraction framework proposed by Hernández-del-Toro et al. [20], applying a 1–50 Hz bandpass filter and evaluating its effectiveness on three different datasets across three distinct tasks: (1) discrimination between imagined word segments (IWS) and inactive state segments (ISS), (2) multiclass classification of imagined words in Spanish, and (3) a combined classification of words versus the inactive state, which more closely resembles a real-world application. Model performance is assessed using the Matthews Correlation Coefficient (MCC). Unlike the F1-score, whose values are not directly comparable across problems with different numbers of classes, MCC maintains a fixed scale of $[-1, +1]$ where zero corresponds to random prediction regardless of the class distribution. This property makes it particularly suitable for jointly evaluating the binary task [23] and the multiclass tasks [24], both of which exhibit class imbalance due to the inclusion of the inactive state.

MATERIALS

Datasets:

The imagined speech datasets employed in this study correspond to those previously used by Hernández-del-Toro et al. [20], with the exception of their third database, which has been replaced by the dataset from Nieto et al. [25], totaling three databases (DS1, DS2, and DS3). Datasets DS1 and DS2 comprise the imagined words "up," "down," "left," "right," and "select" ("arriba", "abajo", "izquierda", "derecha", "seleccionar"). Meanwhile, DS3 includes the commands "up," "down," "left," and "right" ("arriba", "abajo", "izquierda", "derecha").

The third database from [20], discarded due to technical inconsistencies identified during the data adaptation phase, specifically regarding the synchronization and alignment of event markers with the internal targets of the recorded signals.

DS1: Data were recorded using the EMOTIV EPOC+ device (14 channels, 128 Hz sampling rate) with the participation of 27 subjects. For each participant, 20 repetitions per word were obtained, resulting in a total of 100 samples per subject [26].

DS2: Recorded using the EMOTIV EPOC+ device (14 channels, 128 Hz sampling rate) with 27 subjects distinct from those in DS1. For each subject, 32 repetitions per word were obtained, totaling 160 samples per subject. In this data acquisition protocol, words were presented in a randomized order [26].

DS3: This dataset comprises multimodal recordings of electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) signals acquired with the BioSemi ActiveTwo high-resolution

biopotential measurement system. The configuration consisted of 128 active EEG channels and 8 external active EOG/EMG channels, with 24-bit resolution and an initial sampling rate of 1024 Hz. Subsequently, the data were decimated by a factor of four, resulting in a final sampling rate of 256 Hz. This set integrates data from 10 subjects across three different sessions. For each session, 50 repetitions per word were obtained, recording signals for imagined speech, spoken speech, and visualization, for a total of 200 samples per session per subject [25].

METHOD

Data Cleaning and Epoching:

Before the preprocessing stage, a data cleaning and segmentation phase was performed to isolate the epochs of interest from the continuous recordings. This process was implemented using MATLAB for all three datasets, ensuring a standardized approach to identifying and extracting trials based on event markers. Each essay was structured to cover the period leading up to the imagined speech, the imagined speech period, and the subsequent rest phase:

DS1: A total of 100 trials were extracted per subject with a mean duration of $4.3 \text{ s} \pm 1 \text{ s}$.

DS2: A total of 160 trials were extracted per subject with a mean duration of $5.6 \text{ s} \pm 1 \text{ s}$.

DS3: From the original 5,640 trials and over 9 hours of multimodal recordings, a specific subset was isolated. For this study, only the imagined speech condition was selected, excluding spoken speech and visualization trials. This resulted in 200 trials per subject across three sessions (50 repetitions per word), each with a standardized duration of 4.5 s.

The choice of preprocessing techniques was dictated by the electrode density of each system. For DS1 and DS2 (14 channels), CAR was applied as the primary spatial filtering strategy due to its robustness and suitability for low-density montages, where the application of ICA is limited by insufficient dimensionality for reliable source separation.

In contrast, for DS3 (128 EEG channels plus auxiliary channels capturing EOG/EMG), ICA was employed as a preprocessing step for artifact removal, leveraging the higher dimensionality and the availability of reference signals to achieve a more accurate decomposition of neural and non-neural sources.

Following the segmentation stage, each trial was delimited by specific start and end markers identifying the Imagined Words Segment (IWS) and its corresponding word label. To prepare these segments for feature extraction, the processing and validation pipeline (Figure 1) was implemented. This workflow encompasses signal filtering, experimental scenario definition, sliding window segmentation, and the final machine learning evaluation.

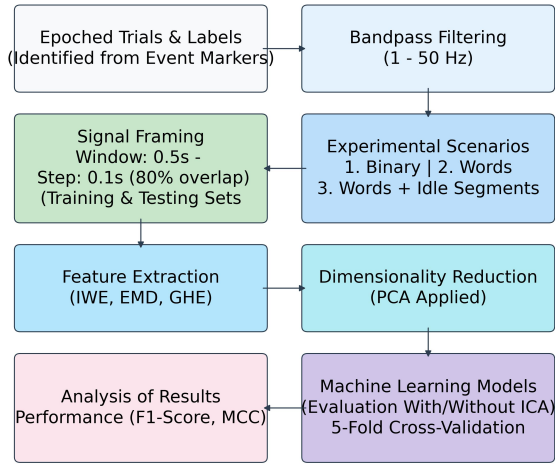


Figure 1: Block diagram of the methodology for imagined speech detection and classification. The process spans from initial epoching to the final machine learning evaluation, comparing performance with and without PCA across different experimental tasks.

This bandpass filter was applied because previous frequency analysis revealed significant saturation below 1 Hz. Additionally, this filter effectively eliminates power line interference (50/60 Hz) and high-gamma frequencies, which are outside the scope of this study.

Unlike the original work, we evaluate the performance of the three scenarios under the extracted feature sets. The sliding window segmentation (Signal Framing) and model training were conducted independently for each scenario using a hybrid workflow: feature extraction was executed in MATLAB (exporting data in .h5 format), while model training and validation were implemented in Python.

Experimental Scenarios Detail:

IWS vs. ISS (Binary): For the training phase, a specific division is made prior to Signal Framing: Class 0 (ISS) aggregates segments recorded before and after word imagination, while Class 1 (IWS) includes all imagined speech segments. For the testing phase, signal framing is applied to the entire trial; the label for each window is assigned based on a majority voting criterion, determined by the predominant label within the 0.5s window duration.

Words (Multi-class): This scenario exclusively isolates the imagined speech segments. Labels are assigned according to the specific word (e.g., "up", "down"), resulting in a number of classes that depends on the database (DS1/DS2/DS3). Signal Framing is applied only to these epochs for both training and testing sets.

Words + idle (Words + Inactive): This scenario follows a logic similar to the binary one but with higher granularity. Imagined speech segments retain their individual words labels (word classes), while all rest segments are consolidated into a final category designated as Idle State.

Feature extraction:

1.-Wavelet Transform (DWT):

A four-level DWT was implemented using the bior2.2 mother wavelet. This technique decomposes the signal into 1 set of approximation coefficients (low frequencies) and 4 set of detail coefficients (high frequencies) 1 per level. To quantify the power distribution in the time-frequency domain, the Instant Wavelet Energie (IWE) was calculated for each of the obtained coefficients [20]. This procedure generates a vector of 5 features per channel.

2. Empirical Mode Decomposition (EMD):

EMD is an adaptive technique that decomposes complex signals into basis functions called Intrinsic Mode Functions (IMF) and a final residue. Unlike other methods, EMD does not rely on a predefined basis function, making it ideal for non-linear and non-stationary EEG signals. Following the decomposition, the two IMFs exhibiting the greatest Euclidean distance relative to the original signal were selected, ensuring the capture of the most physically relevant components[20]. For each of these two IMFs, six non-linear metrics were calculated: Teager Energy (TE) and Instantaneous Energy (IE). Higuchi and Katz fractal dimensions. Generalized Hurst Exponent (GHE) for orders $q = 1$ and $q = 2$. This stage contributes a total of 12 features per channel (6 metrics $\times$ 2 IMFs).

3. Fractal Analysis and Generalized Hurst Exponent (GHE):

Given the non-stationary nature and dynamic complexity of EEG signals, the GHE was employed to quantify self-similarity and temporal scaling properties. The GHE extends the classic Hurst concept by evaluating different statistical moments of signal fluctuations, allowing for the identification of persistent, random, or anti-persistent behaviors.

Exponents were calculated for two specific orders:

$H(q = 1)$: Characterizes the dynamics associated with absolute fluctuations of the signal[20]. $H(q = 2)$: Linked to second-order dynamics and the variance of fluctuations[20].

The analysis of these two orders captures complementary information regarding the multifractal structure of the EEG signal, generating 2 additional features per channel.

Feature Vector Summary:

By integrating the aforementioned techniques (5 DWT + 12 EMD + 2 GHE), a total of 19 features per channel is obtained. Based on the electrode configurations of each database, the resulting feature vector per trial is:

DS1 and DS2: 266 features (14 channels $\times$ 19).

DS3: 2432 features (128 channels $\times$ 19).

Classification:

The classification phase was conducted using four distinct machine learning architectures. RF was implemented as an ensemble of 100 decision trees. The kNN

algorithm was configured with $k = 100$ neighbors. For the SVM, a non-linear decision boundary was established using the Radial Basis Function (RBF) kernel with a default regularization parameter ($C = 1.0$). Finally, LogReg was employed as a probabilistic classifier with an L2 penalty for weight regularization; notably, the maximum number of iterations was increased to 10,000 to guarantee model convergence across all experimental scenarios.

Data Leakage Prevention:

Because the original trials underwent differentiated segmentation (as described in Experimental Scenarios Detail), a rigorous data integrity control was implemented. Strict trial-level separation was guaranteed: any trial selected for the test set was entirely excluded from the training database. This procedure prevents the model from being evaluated on overlapping segments or signals derived from data already seen during the learning phase, thereby ensuring the generalization capability of the results when presented with unseen data.

Table 1: Experimental structure and training configuration per dataset and scenario

Scenario / Task	Feature Set	PCA	ML Models
Binary	All	Yes / No	4 / 4
	IWE, EMD, GHE	No	4 × 3
Words*	All	Yes / No	4 / 4
	IWE, EMD, GHE	No	4 × 3
Words-all**	All	Yes / No	4 / 4
	IWE, EMD, GHE	No	4 × 3

Note: A total of 60 models were trained per subject across the three datasets (DS1, DS2, and DS3) * 5 classes for DS1/DS2 and 4 for DS3 ** 6 classes for DS1/DS2 and 5 for DS3 .

To evaluate the performance of the classifiers across the different imagined speech scenarios, two metrics were used: the F1-score and the Matthews Correlation Coefficient (MCC).

RESULTS AND DISCUSSION

The application of Principal Component Analysis (PCA) in this study was intended to manage the high dimensionality of the features extracted from the EEG signals, aiming to mitigate noise and reduce spatial redundancy. However, a comparative analysis between the results obtained using the original feature set (Table 2) and those following the application of PCA (Table 3) reveals a distinct performance pattern. This behavior appears to be strictly dependent on the channel density and the specific nature of the experimental task.

Statistical validity was assessed using the Wilcoxon signed-rank test with Bonferroni correction. For DS1 and DS2, no significant differences were found ($p \geq 0.05$), justifying the use of the original feature set to preserve data integrity. In DS3, despite a marginal improvement in the binary scenario with PCA, the change was not statistically significant. This suggests that inter-subject and inter-session variability remain the dominant factors in performance, outweighing the benefits of linear dimen-

sionality reduction even in high-density (128-channel) configurations.

Table 2: Best performing classifier for each scenario using the combined feature set (All) without PCA.

Dataset / Scenario	Best Classifier	MCC $\pm$ SD
DS1 (Binary)	LogReg	0.250 $\pm$ 0.136
DS1 (Words)	LogReg	0.423 $\pm$ 0.169
DS1 (Words + Inactive)	LogReg	0.049 $\pm$ 0.072
DS2 (Binary)	LogReg	0.153 $\pm$ 0.118
DS2 (Words)	SVM	0.032 $\pm$ 0.059
DS2 (Words + Inactive)	SVM	0.071 $\pm$ 0.053
DS3 (Binary)	SVM	0.299 $\pm$ 0.116
DS3 (Words)	LogReg	-0.009 $\pm$ 0.048
DS3 (Words + Inactive)	SVM	0.131 $\pm$ 0.050

Table 3: Best performing classifier for each scenario using the combined feature set (All) with PCA.

Dataset / Scenario	Best Classifier	MCC $\pm$ SD
DS1 (Binary)	LogReg	0.253 $\pm$ 0.141
DS1 (Words)	LogReg	0.402 $\pm$ 0.168
DS1 (Words + Inactive)	LogReg	0.050 $\pm$ 0.075
DS2 (Binary)	LogReg	0.143 $\pm$ 0.117
DS2 (Words)	SVM	0.030 $\pm$ 0.057
DS2 (Words + Inactive)	SVM	0.069 $\pm$ 0.053
DS3 (Binary)	LogReg	0.317 $\pm$ 0.114
DS3 (Words)	RF	-0.006 $\pm$ 0.040
DS3 (Words + Inactive)	SVM	0.128 $\pm$ 0.050

The superior performance achieved in DS1 (MCC = 0.423) should be interpreted considering that the data were acquired through a continuous stimulation protocol without randomization of stimuli. This sequential structure likely enhanced the stability of neural patterns, enabling the classifier to optimally leverage the integrity and relevance of the original extracted features.

For DS2, the LogReg model in the Binary task without PCA achieved an average F1-score of 0.704 ± 0.052 across 27 subjects, with a corresponding MCC of 0.153 ± 0.118 . This significant discrepancy between metrics suggests that the model is biased toward the majority class, which in this case represents the inactive states. This phenomenon is attributed to the acquisition methodology, where the duration of the imagined speech trials was self-regulated by the subjects through a keypress at the beginning and end of the mental task, resulting in a temporal class imbalance. Similarly, for the Words + Inactive scenario using SVM, the observed F1-score of 0.256 ± 0.055 and MCC of 0.071 ± 0.053 further indicate that the classifier predominantly identifies the majority class.

In contrast, DS3 featured stricter control over the timing of the imagined speech task; however, a class imbalance persists. In the Binary task with PCA, the LogReg model with PCA yielded the highest performance with an MCC of 0.317 ± 0.114 and an associated F1-score of 0.746 ± 0.043 . For the Words + Inactive dataset, the SVM model with PCA achieved an MCC of 0.128 ± 0.050 and an F1-score of 0.339 ± 0.036 . These results demonstrate that while improved temporal control enhances classification

quality, the underlying class imbalance continues to influence the relationship between these performance metrics.

Evaluation by Feature Type:

To identify the individual contribution of each extraction method (IWE, EMD, GHE), independent tests were conducted. The comparative performance across different datasets and tasks is summarized in Table 4.

Table 4: Top performing configuration (Classifier and Feature Extraction method) for each dataset and experimental scenario (without PCA).

Dataset / Scenario	Rank 1: Model + Feature (MCC $\pm$ SD)
DS1 (Binary)	LogReg + IWE (0.242 $\pm$ 0.157)
DS1 (Words)	LogReg + IWE (0.422 $\pm$ 0.156)
DS1 (Words + Inactive)	LogReg + EMD (0.053 $\pm$ 0.075)
DS2 (Binary)	LogReg + IWE (0.157 $\pm$ 0.114)
DS2 (Words)	SVM + GHE (0.027 $\pm$ 0.037)
DS2 (Words + Inactive)	RF + GHE (0.054 $\pm$ 0.042)
DS3 (Binary)	LogReg + IWE (0.325 $\pm$ 0.112)
DS3 (Words)	LogReg + IWE (0.003 $\pm$ 0.046)
DS3 (Words + Inactive)	SVM + EMD (0.133 $\pm$ 0.039)

Table 4 reports the best-performing configuration (classifier and feature extraction method) for each dataset and classification scenario in the absence of PCA. Logistic Regression combined with IWE features consistently ranks first in five of the nine evaluated combinations, suggesting that this pair generalizes well across binary tasks. The lowest non-trivial performance (0.003 $\pm$ 0.046) corresponds to DS3 in the Words scenario, highlighting the substantial inter-dataset variability that characterizes imagined-speech decoding. For the most ecologically realistic task (Words + Inactive), the best configurations achieve modest MCC values (0.053 for DS1, 0.054 for DS2, and 0.133 for DS3), with SVM combined with EMD features producing the highest result. These figures confirm that the inclusion of the inactive class introduces substantial difficulty due to class imbalance, and that no single classifier–feature combination dominates across all datasets and scenarios.

A noteworthy observation is the relatively high MCC obtained on DS1 in the Words scenario (0.422 $\pm$ 0.156 using LogReg + IWE), which exceeds the values reported for the other two datasets in the same scenario by a wide margin. As previously described, DS1 was acquired using a non-randomized (sequential) stimulation protocol, in contrast to the randomized protocols employed in DS2 and DS3. This sequential structure likely provided a more stable cognitive framework and reduced cognitive load for the participants, favoring the consistency of EEG patterns and allowing linear models such as Logistic Regression to identify discriminative features more effectively. This result should therefore be interpreted with caution: the higher MCC reflects the specific acquisition protocol rather than an intrinsic advantage of the classification pipeline, and it is not directly comparable to the values obtained on DS2 and DS3.

Across all datasets, Logistic Regression and SVM emerged as the most robust classifiers. This consistency,

combined with the strong performance of IWE and EMD features, suggests that discriminative information related to imagined speech can be effectively captured through linear relationships in the time–frequency domain. The competitive performance of EMD features in the Words + Inactive scenario further indicates their capacity to extract non-stationary components of the EEG signal that retain class-discriminative information even under the more challenging conditions imposed by class imbalance.

CONCLUSION

Results demonstrate that, regardless of whether the complete feature space or isolated subsets are employed, superior performance—quantified by the Matthews Correlation Coefficient (MCC)—is consistently achieved for the intention identification task (IWS vs. ISS) relative to multiclass scenarios. This trend is particularly salient when contrasting the binary onset task against the continuous classification paradigm (Words + Inactive), which serves as the most proximal model for real-time BCI applications.

The evaluation in the multiclass word scenario confirmed that the proposed feature sets are primarily optimized for detecting the cognitive intent of speech (IWS vs. ISS). This experiment corroborated that these descriptors are not designed to efficiently discriminate between individual words, proving insufficient for an adequate classification of imagined speech words. While the observed MCC values consistently remain below 0.35—which exceeds the chance threshold but recognizes non-optimal performance—these findings demonstrate that the features implemented from the baseline work of [20] provide a viable operational basis for the first gating stage. Consequently, this scenario underscores that the path to enabling a successful transition toward word classification does not lie in the simple tuning of classifiers, but in the strategic evolution of this first phase of intent detection.

To perfect this isolation of IWS from ISS and break current limits, future research must maintain the gating architecture but expand the analytical framework beyond isolated-channel paradigms. While evidence demonstrates that the current feature sets, specifically IWE and EMD, are effective as initial filters, it is essential to leverage the intrinsic high-dimensionality and spatiality of the EEG manifold without imposing restrictive linearizations. Directly integrating spatial covariance through adaptive methods and multiresolution analysis thereby allows transforming the BCI from a conventional pattern classifier into a dynamic system, capable of faithfully decoding the complex cortical activity of speech within non-invasive neural interfaces.

ACKNOWLEDGMENTS

The authors thankfully acknowledge the computer resources, technical expertise and support provided by the Laboratorio Nacional de Supercómputo del Sureste de

México, SECIHTI member of the network of national laboratories. I wish to express my gratitude for the postgraduate scholarships awarded by SECIHTI, Mexico. These scholarships have enabled me to conduct the research presented in this work.

REFERENCES

- [1] Moreno I et al. Los sistemas de interfaz cerebro-computadora basado en EEG: Características y aplicaciones. *I+D Tecnológico*. 2019;15(2):13–26.
- [2] Chailloux Peguero JD, Mendoza-Montoya O, Antelis JM. Single-option P300-BCI performance is affected by visual stimulation conditions. *Sensors*. 2020;20(24):7198.
- [3] Waytowich N et al. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of Neural Engineering*. 2018;15(6):066031.
- [4] Fazel-Rezai R et al. P300 brain computer interface: Current challenges and emerging trends. *Frontiers in neuroengineering*. 2012;5:14.
- [5] Saha S et al. Progress in brain computer interface: Challenges and opportunities. *Frontiers in Systems Neuroscience*. 2021;15:578875.
- [6] Mitra K et al. Characterizing the onset and offset of motor imagery during passive arm movements induced by an upper-body exoskeleton. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2023, 3789–3794.
- [7] Birch GE, Bozorgzadeh Z, Mason SG. Initial online evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2002;10(4):219–224.
- [8] Shahini N et al. Automatically identified EEG signals of movement intention based on CNN network (End-To-End). *Electronics*. 2022;11(20):3297.
- [9] Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*. 1999;110(11):1842–1857.
- [10] Blankertz B et al. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2003;11(2):127–131.
- [11] Obermaier B, Müller GR, Pfurtscheller G. "virtual keyboard" controlled by spontaneous EEG activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2003;11(4):422–426.
- [12] Perdakis S, Leeb R, Millán JdR. Context-aware adaptive spelling in motor imagery BCI. *Journal of Neural Engineering*. 2016;13(3):036018.
- [13] Faulkner HG et al. Sequential hypothesis testing for automatic detection of task-related changes in cerebral perfusion in a brain–computer interface. *Neuroscience Research*. 2015;100:29–38.
- [14] Silva AB et al. The speech neuroprosthesis. *Nature Reviews Neuroscience*. 2024;25(7):473–492.
- [15] Metzger SL et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*. 2023;620(7976):1037–1046.
- [16] Moses DA et al. Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature Communications*. 2019;10(1):3096.
- [17] DaSalla CS et al. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*. 2009;22(9):1334–1339.
- [18] Shevchenko AO, Vartanov AV. Comparison of the mechanisms of phonemic awareness and internal pronunciation of phonemes and syllables: EEG and fMRI study. *Russian Psychological Journal*. 2022;19(4):186–210.
- [19] Suyuncheva A et al. Reconstruction of words, syllables, and phonemes of internal speech by EEG activity. In: *Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics*. Springer International Publishing: Cham, 2021, 319–328.
- [20] Hernández-Del-Toro T, Reyes-García CA, Villaseñor-Pineda L. Toward asynchronous EEG-based BCI: Detecting imagined words segments in continuous EEG signals. *Biomedical Signal Processing and Control*. 2021;65:102351.
- [21] Cooney C, Folli R, Coyle D. A bimodal deep learning architecture for EEG-fNIRS decoding of overt and imagined speech. *IEEE Transactions on Biomedical Engineering*. 2022;69(6):1983–1994.
- [22] Chang J et al. Generative listener EEG for speech emotion recognition using generative adversarial networks with compressed sensing. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(4):2025–2036.
- [23] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(1):6.
- [24] Gorodkin J. Comparing two K-category assignments by a K-category correlation coefficient. *Computational Biology and Chemistry*. 2004;28(5–6):367–374.
- [25] Nieto N et al. Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. *Scientific Data*. 2022;9(1):52.
- [26] Torres-García AA. Análisis y clasificación de electroencefalogramas (EEG) registrados durante el habla imaginada. Español. Tesis de doctorado. Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE). Santa María Tonantzintla, Puebla, México. Apr. 2016. [Online]. Available: <https://repositorio.inaoe.mx/handle/123456789/100>.