Imagined Speech EEG-based BCI Using Dynamic Mode Decomposition

MTech. Thesis

By

Akah Precious Chiemena

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2023

Imagined Speech EEG-based BCI Using Dynamic Mode Decomposition

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Technology

by

Akah Precious Chiemena

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2023

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Imagined speech EEG-based BCI using dynamic mode decomposition** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF ELECTRICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2021 to June 2023 of MTech. Thesis submission under the supervision of Prof. Ram Bilas Pachori, Indian Institute of Technology, Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Akah Precious

15:06:2023

Signature of the student with date

(AKAH PRECIOUS CHIEMENA)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

16.06.2023

Signature of the Supervisor of

M.Tech. thesis #1 (with date)

(Prof. Ram Bilas Pachori)

AKAH PRECIOUS CHIEMENA has successfully given his MTech. Oral Examination held on the **12TH May, 2023**.

Signature(s) of Supervisor(s) of MTech. thesis

Date: 16.06.2023

Convener, DPGC

Date:16/06/2023

Signature of PSPC Member #1

Date: June 16, 2023

Signature of PSPC Member #2

Date: 16/06/2023

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Ram Bilas Pachori for his continuous support and valuable guidance. His guidance helped me through my work. I have been able to push myself beyond my expectations with his excellent supervision and encouragement.

I would like to thank my PSPC members Dr. Sanjram Premjit Khanganba ,Dr. Nitya Tiwari, Dr. Swaminathan and Dr. Amod C. Umarikar for their insightful comments and suggestions towards my research. I sincerely acknowledge IIT Indore and ICCR for supporting my M.Tech. program through the scholarship provided.

Last but not least, my work would not have been possible without the encouragement of my parents, my beloved wife, Dr. Ali, Lt. Col. Atomode, Dr. Paul Enenche, Apostle Joshua Selman, Prof. Sadiq Thomas, Prof. Dili Dogo and Ms. Trupti Rathod. My heartfelt gratitude goes to Dr. Pradeep Chauddhary, Vivek Kumar Singh, Shailesh Bhalerao and distinguished members of the signal and analysis research laboratory whose tremendous support has helped me to s stay positive.

DEDICATION

This thesis is dedicated to my Lord and personal savior, Jesus Christ.

Abstract

To aid individuals with speech impairments, the research aims to decode imagined speech from non-stationary EEG signals using brain-computer interfaces (BCIs). The proposed method combines multichannel EEG signals with time-frequency representations (TFRs) based on Hilbert spectral analysis of modes decomposed by dynamic mode decomposition (DMD). The imagined electroencephalogram (EEG) signals of six imagined speech commands (Up, Down, Left, Right, Upward, and Downward) from 15 subjects sampled at 1024 Hz were measured using a six-concatenated channel standard physiological signal system; the signal was filtered to remove artifacts between 2 and 40 Hz using a finite impulse response pass-band filter. The proposed method employs a convolutional neural network (CNN) model that takes TFRs as input images to decode the imagined speech commands.

The parameters of the model are optimized to achieve the best performance in decoding speech commands. In a comparison with the multi-fast and adaptive empirical mode decomposition method, the proposed method utilizing DMD achieves significantly improved decoding accuracy. EEG signals are decoded with an accuracy of 70% for six classes of imagined speech commands. DMD and CNN are integrated in the proposed method to extract dynamic information from imagined speech EEG signals, represented as spatial modes and their instantaneous frequencies (alpha and beta).

Integrating EEG signals with imagined speech commands improves decoding accuracy and provides practical methods for extracting meaningful information from them. It is concluded that individuals with speech stimulation challenges can benefit from the Imagined Speech EEG-based BCI system as a communication facilitation tool. The study proposes a novel approach to decoding imagined speech from multichannel EEG signals that combines DMD, TFRs, and CNN. We have implemented subject-independent and subject-interdependent BCI models to demonstrate that alpha and beta bands improved the decoding accuracy of imagined speech better than other frequency bands and the results are better than those presented in the literature, and we have determined that EEG-based BCIs can be useful for individuals with speech impediments.

Keywords: Dynamic mode decomposition, electroencephalogram, brain computer interfaces, time frequency representations and convolutional neural network

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations and Acronyms

AM: Amplitude modulation

CF: Convolutional filter

CPU: Computer processing unit

DMD: Dynamic mode decomposition

EMG: Electromyography

GPU: Graphics processing unit

HTSA: Hilbert transform separation algorithm

HT: Hilbert spectrum

HS: Hilbert spectrum

IF: Instantaneous frequency

IE: Instantaneous Energy

MFAEMD: Multivariate fast and adaptive empirical mode decomposition

MSE: Mean square error

MATLAB: Matrix Laboratory

RAM: Random access memory

RVM: Relevant vector machine

MDMD: Multi-Channel DMD

SDMD: Single-Channel DMD

TFD: Time frequency distribution

TF: Time frequency

TFR: Time frequency representation

Chapter 1

Introduction

The introduction of the study provides an overview of the research area and sets the context for the investigation. It begins by highlighting the significance of brain-computer interfaces (BCIs) in enabling direct communication between the human brain and external devices [66, 71-75], particularly for individuals with speech impairments. The introduction emphasizes the limitations of traditional communication methods for individuals with speech disabilities and the potential of BCIs to bridge this gap. It highlights the use of electroencephalography (EEG) as a non-invasive and accessible modality for capturing neural activity associated with speech production [51-54].

Furthermore, the introduction discusses the challenges involved in decoding imagined speech commands from EEG signals due to their complex and dynamic nature. It highlights the need for advanced signal processing and machine learning techniques to extract relevant information and translate it into actionable commands [55, 59]. To address these challenges, the study proposes the utilization of dynamic mode decomposition (DMD) as an efficient extraction technique of imagined signal components with convolutional neural networks (CNNs) as a classification model [38]. The introduction provides a brief overview of DMD and its applications in analyzing complex, non-linear systems, as well as the effectiveness of CNNs in learning and recognizing patterns in high-dimensional data.

Additionally, the introduction outlines the objectives of the study, which include developing a novel BCI system for imagined speech decoding, investigating the efficacy of DMD in extracting informative features from EEG signals, and evaluating the performance of the CNN-based classification model. By providing this comprehensive introduction, the study establishes the rationale for the research, highlights the gaps in existing literature, and presents the specific goals and methodologies that will be employed throughout the investigation. This sets the stage for the subsequent sections of the thesis, where the research problem, methodology, results, and discussion are elaborated upon in detail.

1.1 Background and Motivation

The background and motivation for the thesis stem from the need to develop effective communication technologies for individuals with speech impairments. Traditional methods of augmentative and alternative communication have limitations, and there is a growing interest in utilizing BCIs to provide a direct means of communication. BCIs that decode imagined speech from EEG signals hold great potential in this regard.

The motivation behind this thesis lies in addressing the challenges associated with decoding imagined speech from non-stationary EEG signals. The existing methods for decoding imagined speech often struggle with the complexity and variability of EEG signals, which limits their accuracy and usability. Therefore, there is a need for innovative approaches that can extract meaningful information from EEG signals and decode imagined speech commands more accurately.

The thesis aims to explore the use of DMD as a novel technique for analyzing EEG signals in the context of imagined speech [12]. DMD offers a unique advantage in capturing dynamic spectral information from EEG signals by decomposing them into spatial modes and their associated instantaneous frequencies. By integrating DMD with advanced machine learning techniques such as CNNs, it is anticipated that the proposed BCI system can achieve improved accuracy and reliability in decoding imagined speech commands [53].

The successful development of an imagined speech EEG-based BCI using DMD has the potential to significantly enhance the quality of life for individuals with speech impairments. By providing a direct and efficient means of communication, this technology can empower individuals to express themselves more effectively and engage in various social and professional interactions. Moreover, the utilization of DMD in decoding imagined speech holds promise for advancing the field of BCIs by introducing a novel approach that extracts valuable information from imagined speech EEG signals and improves the overall performance of BCI systems.

1.2 Research Problem and Objective

The research problem addressed in the thesis boarders on the challenge of accurately decoding imagined speech from EEG signals. Existing methods for decoding imagined speech have limitations in terms of accuracy and robustness, hindering their practical application in real-world scenarios. Therefore, there

is a need to develop an innovative approach that can effectively extract and interpret meaningful information from EEG signals related to imagined speech commands [60-72].

The primary objective of this thesis is to investigate the feasibility and effectiveness of utilizing DMD in the context of EEG-based BCIs for decoding imagined speech. DMD offers the potential to capture the dynamic nature of EEG signals by decomposing them into spatial modes and their associated instantaneous frequencies. By integrating DMD with advanced machine learning techniques such as CNNs [38,41], the objective is to develop a robust and accurate system that can decode imagined speech commands from EEG signals.

Furthermore, the thesis aims to evaluate and compare the performance of the proposed DMD-based BCI approach with existing methods in terms of decoding accuracy, computational efficiency, and usability, Comparative analysis will be conducted to demonstrate the advantages and potential of the DMD-based approach in improving the overall performance of EEG-based BCIs for imagined speech [53].

Ultimately, the research objectives of this thesis are to contribute to the advancement of EEG-based BCIs for imagined speech [68-70] and provide a reliable and efficient communication tool for individuals with speech impairments, by addressing the research problem and achieving the stated objectives, this thesis seeks to pave the way for future developments in the field, leading to improved communication assistance for individuals with speech limitations.

1.3 Significance of the Study

The study holds significant importance in the field of BCIs and communication assistance for individuals with speech impairments [64]. The findings and contributions of this research have several key implications.

Firstly, the proposed DMD-based approach has the potential to significantly enhance the decoding accuracy of imagined speech from EEG signals. By leveraging the dynamic nature of EEG signals through DMD, the system can capture and interpret the intricate patterns associated with imagined speech commands [55]. This improvement in accuracy opens up new possibilities for developing more reliable and effective BCIs that can accurately translate the intentions of individuals with speech limitations into meaningful commands.

Secondly, the integration of DMD with advanced machine learning techniques which includes CNN, will not only enhances decoding accuracy but also improves the overall performance of EEG-based BCIs. The combination of these methods provides a robust framework for extracting relevant information from EEG signals and translating it into actionable commands. This has practical implications for individuals with speech impairments, as it offers a reliable and efficient means of communication that can potentially improve their quality of life [53].

Furthermore, the study contributes to the growing body of knowledge in the field of BCIs and neurorehabilitation. By exploring the use of DMD and its integration with existing techniques, the research expands the understanding of how EEG signals can be effectively utilized for decoding imagined speech. The insights gained from this study can inspire further advancements and innovations in the development of BCIs, not only for imagined speech but also for other applications in neurorehabilitation and assistive technologies.

In conclusion, the significance of this study lies in its potential to improve the accuracy and usability of EEG-based BCIs for imagined speech, offering a promising solution for individuals with speech impairments. The findings and contributions of this research have the potential to shape the future of communication assistance technologies, facilitating better interaction and integration for individuals with speech limitations [101,102].

1.4 Thesis Structure

The following outline outlines the organization of the thesis:

1. Introduction: This section provides an overview of the research topic and introduces the motivation and significance of the study [70,101]. It outlines the objectives, research questions, and scope of the thesis, setting the context for the subsequent chapters.

2. Review of past works and problem formulation: This chapter reviews the existing literature on EEG-based imagined speech BCIs, highlighting the limitations and challenges in current approaches. It identifies the research gap and formulates the problem statement that the thesis aims to address. The chapter also discusses relevant concepts, theories, and methodologies used in the field.3. Imagined speech EEG-based BCI framework: This chapter delves into the theoretical foundations and conceptual framework of the imagined speech BCI. It explores the cognitive processes involved in generating

imagined speech signals and provides an overview of the EEG-based BCI paradigms used for decoding speech intentions. The chapter discusses the different approaches and methodologies employed in imagined speech BCIs, including signal processing techniques, namely, dynamic mode decomposition (DMD) for efficient extraction of signal modes, feature extraction technique and classification.

4. Experimental protocol: This chapter describes the experimental setup and methodology used in the study. It outlines the data collection process, including the participant selection criteria, EEG recording procedures, and experimental tasks for generating imagined speech signals. The chapter also discusses the pre-processing steps and data preparation for input into the DMD algorithm.

5. Results and analysis: This chapter presents the findings of the study and provides a detailed analysis of the results. It discusses the performance metrics, such as classification accuracy and decoding accuracy, obtained from the EEG-based imagined speech BCI using DMD. The chapter also compares the results with existing methods and discusses the significance and implications of the findings.

6. Conclusion and scope for future work: The final chapter concludes the thesis by summarizing the key findings and contributions of the research. It discusses the implications of the study in the field of EEG-based imagined speech BCIs and highlights the limitations and challenges encountered during the research. The chapter also provides recommendations for future research directions, suggesting areas for further exploration and improvement in the field.

Basically, the thesis structure outlined above ensures a logical progression of ideas and analysis, starting from the introduction and literature review, moving through the methodology and results, and concluding with a discussion of the implications and potential for future work.

Chapter 2

Review of Past Works and Problem Formulation

The literature review section of the research on imagined speech EEG-based BCI using DMD, provides a comprehensive overview of existing studies and research works relevant to the field. It serves as a foundation for understanding the current state of knowledge and identifying gaps in the literature. This section begins by discussing the significance of EEG signals in diagnosing neurological and physiological disorders and emphasizes the importance of extracting relevant features for accurate detection. The limitations of traditional methods such as Fourier transform (FT) in analyzing non-stationary EEG signals are highlighted, leading to the exploration of alternative techniques such as time-frequency representations and wavelet transform (WT) [2,30, 5-6]. The introduction of empirical mode decomposition (EMD) as a data-driven approach to analyses non-linear and non-stationary signals is discussed, along with its advantages over other decomposition techniques [21]. Additionally, the section mentions the successful application of various signal decomposition methods in classification and prediction problems, indicating their potential in the context of imagined speech decoding. Overall, the literature review sets the stage for the subsequent methodology and contributes to the development of a comprehensive understanding of the research area.

2.1 Review on Imagined Speech BCI

Neural signals possess temporal, spatial, and spectral characteristics that hold valuable information for diagnosing various neurological and physiological disorders [102]. The extraction of features from various neurological responses is a significant research area, as the choice of features directly impacts the performance of detection [78]. In literature, several methods have been proposed for the decoding of imagined speech EEG signal from various databases [80-81]. Traditionally, the Fourier transform (FT) has been widely used to analyze EEG signals and generate spectral features. However, the FT assumes stationarity in the signal, which is not applicable to non-stationary real-world signals like EEG. Furthermore, the FT lacks time information [103].

To address these limitations, contemporary methods based on time-frequency representations have been developed. One of such method is the short-time Fourier transform (STFT), which examines local signal

characteristics using a window [84]. However, the STFT's use of a fixed window-based filter which restricts simultaneous improvement in time and frequency resolutions. WT overcomes this issue by employing multiple filters with varying bandwidths, enabling multi-resolution analysis [4-7]. The WT first utilizes a window encompassing the entire signal to extract low frequencies, resulting in good time resolution. It then translates and scales the window to capture higher frequency information, achieving improved frequency resolution [8]. Despite these advantages, the WT still struggles with computational efficiency to enhance both time and frequency resolutions simultaneously and to obtain high-resolution information in both domains, a large number of wavelet coefficients need to be computed and analyzed. This can result in a significant computational burden, especially for long-duration signals or high-resolution analyses [9].

To overcome these difficulties, the EMD algorithm was introduced as a means of examining signals that are both non-stationary and non-linear [12]. EMD is a data-oriented technique that breaks down the signal into intrinsic mode functions (IMFs), which represent a finite number of oscillations. These IMFs are not predefined basis functions; instead, they adhere to specific criteria: (i) the count of extrema and zero-crossings must either be equal or differ by no more than one, and (ii) the mean of the envelope formed by local maxima and minima is zero at any given point. EMD provides several advantages, including adaptive scales, the ability to separate oscillations based on data, and local multi-resolution analysis. In contrast, the WT examines signals on a global level using predetermined filter scales [21]. Various methods for signal decomposition, such as WT, STFT, EMD, variational mode decomposition (VMD) [10], singular value decomposition (SVD) [26], synchro squeezing transform (SST) [29], and their variations [8-23], have been effectively employed in classification and prediction tasks.

These techniques have demonstrated their effectiveness in capturing relevant features from EEG signals, aiding in accurate classification and prediction tasks. In recent years, there has been growing interest in the use of multi-resolution analysis (MRA) for decoding imagined speech [56]. MRA involves decomposing speech signals into multiple levels of resolutions, allowing for more precise analysis of the various components of the signal [104].

The existing approaches have shown the promise of improving the accuracy of decoding imagined speech, and several studies have been conducted to investigate its effectiveness. In [101] an automated recognition of imagined commands from EEG signals using multivariate fast and adaptive empirical mode decomposition-based method for decomposition of multichannel EEG signals into modes to

compute slope domain entropy and L1-norm features from six-EEG channels using k-nearest neighbor (KNN) to actualize an average accuracy of 60.72% was proposed [9, 86]. Studies have shown that EMD can extract features from EEG signals of participants imagining speaking different vowel sounds and different words, which can then be used to decode the intended speech from the EEG signals. The extracted features were used to accurately classify the imagined vowel sounds with a high degree of accuracy and to decode the intended words with a high degree of accuracy [53-59]. Recent studies have explored the decoding of imagined hand movements using EEG signals and demonstrated the feasibility of reconstructing hand trajectories accurately [60-62]. Furthermore, authors in [6] investigated the neural basis of imagined speech production and proposed a framework for decoding speech-related information from EEG signals and employed feature extraction techniques and machine learning algorithms to classify the imagined speech commands.

In [89], authors have used classification framework based on independent component analysis (ICA) was used to decode imagined speech from EEG signals, achieved an accuracy rate of 76%. In [88] it emphasizes the importance of selecting appropriate ICA algorithms and parameters for different types of data, while another review article discusses the challenges of interpreting the results of ICA-based analyses. It is observed that, ICA has shown promise as a tool for decoding imagined speech signals. However, further research is needed to address the limitations of this technique and improve its accuracy and reliability. Similarly, the work done in [101], have focused on the classification of imagined speech syllables using EEG signals. They employed a combination of spatial filtering methods and pattern recognition algorithms to achieve accurate classification results. Furthermore, in [90] a discriminative neural pattern analysis approach for decoding speech-related information from EEG signals was proposed. Their study provided insights into the neural mechanisms underlying imagined speech and demonstrated the potential of EEG-based BCIs for speech communication [55].

Furthermore [102], proposed a transfer learning in imagined speech EEG-based BCIs, authors have explored the characteristic units (i.e., code words) of the EEG associated with the words of an initial vocabulary from 14-EEG channels using CNN and achieved an accuracy of 68.9%. In, the work of [38], proposed a multi-kernel learning approach to classify imagined speech tasks based on EEG signals. Similarly, their study demonstrated the effectiveness of this approach in achieving high accuracy in imagined speech decoding, highlighting the importance of advanced classification techniques in EEG-based BCIs [101].

[37] presents optimized layers to improve CNN generalization and transfer learning for imaginary speech decoding from EEG signals with positive correlations on the target 18-EEG channels subject's data, with a CNN with an average accuracy of 39.09% [44]. Using wavelet transform (CWT) to generate a three-dimensional tensor, the authors of [76,104] investigated tensor decomposition in EEG. The tensor is subjected to parallel factor analysis (PARAFAC), and the components are ordered and labelled for 14-EEG and parallel factor analysis.

The use of deep neural networks (DNN) for decoding imagined speech from EEG signals was studied in [105,106]. This study proposed a novel deep learning architecture and achieved promising results in decoding imagined speech commands accurately. A recent study demonstrated that deep learning techniques can be combined with EEG-based brain-computer interfaces (BCIs) to decode imagined speech. Another study, published in [107], presented a new way to decode vowel imagery patterns from EEG signals using deep capsule neural networks. In order to identify relevant features of the EEG signal related to vowel imagery patterns, they used CNN and computed the number of convolutional kernels using entropy techniques. An accuracy of 78.57% was achieved with the proposed method. In addition, a direct speech BCI based on EEG was introduced in [78]. Combining a sequence-mapped real transform (SMRT) with MFCC/LPCC features and an artificial neural network (ANN) was utilized in this system. With an improved accuracy of 73.37%, the system was able to improve its performance.

In a recent study by [67], a method called Fourier-Bessel series expansion-based empirical wavelet transform (MFBSE-EWT) was proposed for computing EEG spectral and temporal complexity. This approach was used by the authors to extract features from the EEG data. In a separate study conducted by [105] in 2021, imagined speech was decoded from EEG signals. A common spatial pattern (CSP) algorithm was used to extract features from EEG data. A neighborhood component analysis (NCA) technique was then used to select the most discriminative features. Stacking ensemble learning enabled researchers to achieve a 51.90% accuracy rate in their decoding task. The authors of [41], achieved 70.4% accuracy from a 64-channel EEG recording. In this work, random forest (RF) and support vector machines (SVM), were used for classification of imagined speech using six phonetically distributed words and features extracted from decomposed bands using a discrete wavelet transform (DWT) [5].

In [9], EEG-Based multiword imagined speech classification has been employed on Persian words, frequency spectrum resolution for 19 EEG channel using binary support vector machine (SVM). Furthermore, the wavelet transform has been combined with other machine learning techniques, which

includes (SVM) and deep learning, to improve the decoding performance. In [41], a deep CNN was trained on wavelet-transformed EEG signals to decode imagined speech, with an accuracy of 92%.

The effectiveness of STFT in decoding imagined speech has been further demonstrated in recent studies. For instance, in a recent study by [83] STFT has been used in combination with convolutional neural networks (CNN) to decode imagined speech achieving high accuracy, while in [2,84], the authors propose the use of STFT and SVM to classify imagined speech with high accuracy. MRA was used in conjunction with a deep learning algorithm to decode imagined speech from EEG signals in a study by [14]. It was found that MRA significantly improved decoding accuracy, with an overall classification accuracy of 83.9%. Using MRA to analyze the EEG signals of speakers imagining different phonemes was also reported in [104]. Based on the results of the research, it was determined that MRA was able to accurately distinguish between phonemes with an average accuracy of 73.9%. Furthermore, MRA can also be applied to fMRI data. An analysis of speech imagining signals using MRA was conducted in [106]. Researchers discovered that MRA could be used to decode imagined speech from fMRI data by identifying brain regions associated with different words.

Canonical polyadic decomposition (CPD), Tucker decomposition (TD), and parallel factor analysis-2 have all been proposed for imagined speech decoding. Every technique has its strengths and weaknesses, and the choice of technique depends on the nature of the data and the specific decoding task [101]. Recently, deep learning models have been explored for decoding imagined speech. In certain scenarios, these models may outperform traditional tensor decomposition techniques due to their ability to integrate multiple sources of information. Researchers explored how DMD can be used to decode imagined speech from intracranial electrocorticography (ECoG) signals, demonstrating its feasibility and effectiveness in capturing speech-related neural patterns. These findings provide valuable insight into DMD's application in EEG-based imagined speech decoding [23]. Furthermore, the authors in [23], investigated the possibility of using DMD spectral moments and sub band-powers spectrum for the analysis of EEG signals. Experimental results show that the higher order DMD spectral moments and sub band-power features outperformed the analogous spectral features, calculated from traditional power spectrum.

Collectively, these studies contribute to the body of knowledge on imagined speech decoding using EEG-based BCIs. They provide valuable insights into the methodologies, techniques, and algorithms used in this field. The literature review section establishes a comprehensive understanding of the existing literature, identifies research gaps, and synthesizes and analyzes these related research works. A review

of the existing methods reveals that some improvements can be made regarding parameter selection, improper decomposition, and mode mixing. Therefore, the research proposal uses DMD for the decoding of imagined speech in order to overcome the challenges listed above. Finally, the choice of decoding technique is determined by the nature of the data and the particular task at hand [86].

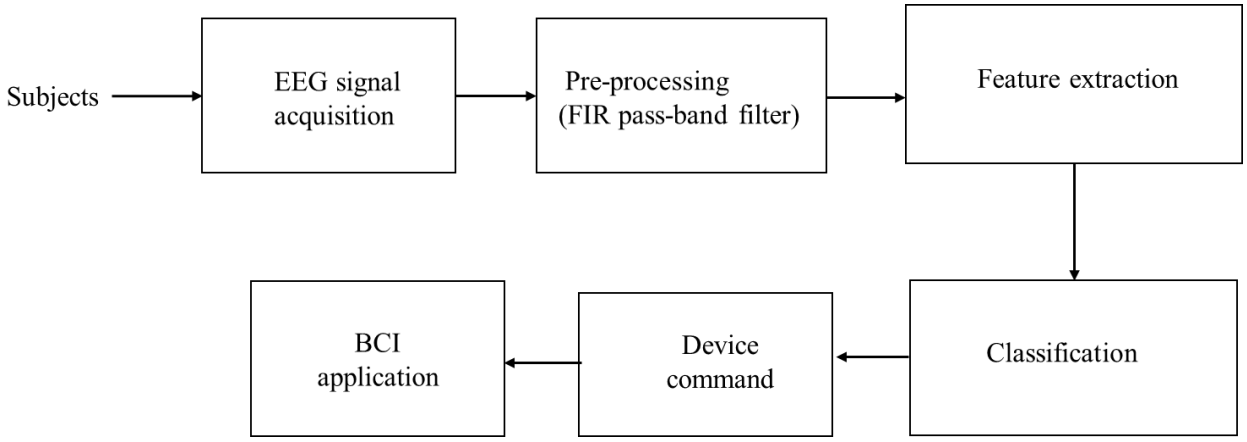


Figure 1: Generalized architecture of EEG-based BCI system.

2.2 EEG-Based Brain Computer Interface (BCI) Architecture

In the research work, an imagined speech EEG-based BCI architecture is proposed to enable individuals with speech impediments to communicate through imagined speech commands. The BCI architecture consists of several key components working in harmony. Firstly, the EEG signal acquisition module captures the electrical activity of the brain using electrodes placed on the scalp, specifically targeting cortical areas related to language processing [59]. The acquired EEG signals are then pre-processed to remove noise and artifacts, ensuring the reliability of subsequent analysis [80]. The next component involves the stimulus presentation and task instructions, where visual cues or prompts are presented to the users to imagine specific speech commands.

These imagined speech commands are decoded from the EEG signals using a combination of DMD and CNN models [55]. The DMD method extracts dynamic information from the EEG signals [3], while the CNN model utilizes time-frequency representations and spatial features to classify the imagined speech commands. The classification results are analyzed, and evaluation metrics are employed to assess the accuracy and performance of the BCI system. Conclusively, the proposed EEG-based BCI architecture provides a framework for decoding imagined speech commands and holds the potential to enhance communication for individuals with speech impairments.

2.3 Paradigms for EEG-Based BCIs

The literature review on paradigms for EEG-based BCIs explores the diverse approaches and methodologies employed in the field of EEG-based BCIs [4]. EEG-based BCIs utilize electroencephalography (EEG) signals to establish direct communication between the human brain and external devices or applications [60].

This section provides an overview of the major paradigms used in EEG-based BCIs, including motor imagery, P300, steady-state visually evoked potentials (SSVEPs), and hybrid paradigms [4,35]. Motor imagery paradigms involve users imagining specific motor actions, which can be decoded from EEG signals to control external devices or perform tasks. P300 paradigms focus on detecting and classifying P300 event-related potentials, which are neural responses elicited by rare or target stimuli, to enable users to make selections or commands. SSVEP paradigms utilize the brain's response to visual stimuli at specific frequencies to achieve high-speed and accurate communication.

Hybrid paradigms combine multiple paradigms to enhance the performance and versatility of EEG-based BCIs [90-100]. The literature review examines the advantages, limitations, and recent advancements in each paradigm, providing insights into their suitability for different BCI applications. It also highlights the challenges faced in EEG signal processing, feature extraction, and classification techniques within each paradigm. By synthesizing the existing literature, this review sets the foundation for the proposed research on EEG-based BCIs and paves the way for the development of more effective and robust BCI systems [101].

2.4 Challenges and Limitations

EEG-based BCIs, including imagined speech BCIs, face several challenges and limitations that need to be addressed for their successful implementation. One of the main challenges is the variability and noise present in EEG signals. EEG recordings can be affected by various factors such as muscle artifacts, eye movements, and environmental interference, which can degrade the signal quality and affect the accuracy of decoding. Another challenge is the individual variability in EEG patterns and brain activity, making it necessary to develop personalized models and algorithms for each user. One approach to improving the accuracy of decoding imagined speech is the application of signal processing algorithms and deep

neural network to identify patterns in the EEG signals [25]. However, these algorithms require large amounts of training data and may not be effective for all individuals.

Additionally, the non-stationarity of EEG signals poses a challenge as the brain activity and EEG patterns can change over time, requiring adaptive and robust signal processing techniques. Furthermore, the low spatial resolution of EEG limits the ability to precisely localize brain activity and may lead to challenges in accurately decoding specific neural processes. Additionally, the usability and user experience of EEG-based BCIs need to be improved to ensure user acceptance and engagement.

Factors such as system complexity, training requirements, and cognitive load can impact the usability and practicality of these systems. Finally, the translation of imagined speech BCIs from controlled laboratory environments to real-world applications and daily-life settings presents its own set of challenges, including the need for robustness, reliability, and portability. Addressing these challenges and limitations is crucial for advancing the field of EEG-based BCIs and unlocking their full potential in enabling communication and control for individuals with speech impairments.

Expanding the application of EEG-based brain-computer interfaces (BCIs) to include the decoding of imagined speech is crucial because speech plays a vital role in connecting individuals with society. However, there are limitations in utilizing external speech stimulation for certain individuals due to medical conditions. Therefore, decoding imagined speech directly from EEG signals becomes important. Nonetheless, this task is challenging due to the intricate nature of EEG signals and the significant variability observed between individuals [100]. Additionally, there is often overlap between the signals generated by different types of imagined speech, making it difficult to distinguish between them. One approach to improving the accuracy of decoding imagined speech is to use advanced signal processing algorithms (DMD) and deep neural network to identify patterns in the EEG signals. Finally, the DMD enhances the decomposition level, feature extraction quality and ultimately improves the accuracy of the EEG based BCI system.

Chapter 3

Imagined Speech EEG Based BCI Framework

The methodology section of the research work on imagined speech EEG-based BCI using DMD provides a detailed description of the experimental approach and data analysis techniques employed in the study. This section aims to outline the step-by-step process followed to achieve the research objectives. It begins by discussing the data collection process, including the acquisition of EEG signals from the participants. The pre-processing steps, such as artifact removal and signal normalization, are then described to ensure the quality and reliability of the data. Next, the feature extraction methods utilized to extract relevant information from the EEG signals are explained [100].

In the feature extraction techniques such as time-frequency analysis using Hilbert spectral analysis or multichannel EEG signal fusion. The classification model, such as a CNN architecture, employed for decoding the imagined speech is discussed, along with the training and optimization procedures. Furthermore, the evaluation metrics used to assess the performance of the classification model and the analysis of the results are presented [38]. The methodology section serves as a roadmap for conducting the study and provides a comprehensive understanding of the techniques and procedures employed to investigate imagined speech decoding using dynamic mode decomposition [101].

3.1 Concept and Significance

EEG-based imagined speech BCIs have gained significant attention due to their potential to provide a direct communication pathway for individuals with speech impairments [106]. The concept behind these BCIs lies in the ability to decode and interpret the neural activity associated with imagined speech processes. By leveraging EEG signals, which capture the electrical activity of the brain, it becomes possible to detect and classify the user's intended speech commands solely based on their brain activity. This concept holds great significance as it opens up possibilities for individuals who are unable to communicate verbally to regain their ability to express themselves. EEG-based imagined speech BCIs offer a non-invasive and portable solution that can be customized to the specific needs and capabilities of each user. By enabling direct brain-computer communication, these BCIs have the potential to enhance the quality of life and social interaction for individuals with speech impairments. Furthermore,

the development of such BCIs not only addresses the immediate needs of speech-impaired individuals but also contributes to the broader field of neuroscience and neuroengineering. The research and technological advancements in EEG-based imagined speech BCIs lead to a deeper understanding of the neural mechanisms underlying speech production and cognition, paving the way for future innovations in the field of BCI [7,12].

3.2 State-of-the-Art Techniques

The state-of-the-art techniques for decoding imagined speech from neural signals involve a combination of advanced signal processing, machine learning, and neural network approaches. Here are some notable techniques in this field:

1. Deep Learning Models: Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants have been widely used for imagined speech decoding. These models leverage the power of deep learning to learn complex patterns and representations from neural data [105].

2. Electroencephalography (EEG) Analysis: EEG signals have been extensively used for decoding imagined speech. Various feature extraction techniques, such as common spatial patterns (CSP), event-related desynchronization / synchronization (ERD/ERS), and power spectral density (PSD) analysis, are employed to extract discriminative features from EEG signals [85].

3. Functional Near-Infrared Spectroscopy (fNIRS): fNIRS measures changes in blood oxygenation levels in the brain and has been employed to decode imagined speech. Similar to EEG, feature extraction techniques are applied to fNIRS signals to extract informative features for classification [90-92]

4. Hybrid Approaches: Some approaches combine multiple modalities, such as EEG and fNIRS, to leverage the complementary information provided by different neural signals. Hybrid models aim to enhance the accuracy and reliability of imagined speech decoding [91].

5. Transfer Learning: Transfer learning techniques have been explored to improve the performance of imagined speech decoding models. Pre-trained models on large speech datasets or related tasks are fine-tuned on imagined speech data to leverage the learned representations. It’s important to note that the field of imagined speech decoding is still evolving, and new techniques and advancements continue to

emerge. The performance and effectiveness of these techniques may vary depending on factors such as the quality of neural data, experimental setup, and specific application requirements.

3.3 Proposed Methodology

The methodology employed in the study of imagined speech EEG-based BCI using DMD, involves a series of steps to decode imagined speech commands from non-stationary EEG signals. The researchers proposed a novel approach that combines DMD, time-frequency analysis using Hilbert spectral analysis, and a CNN architecture.

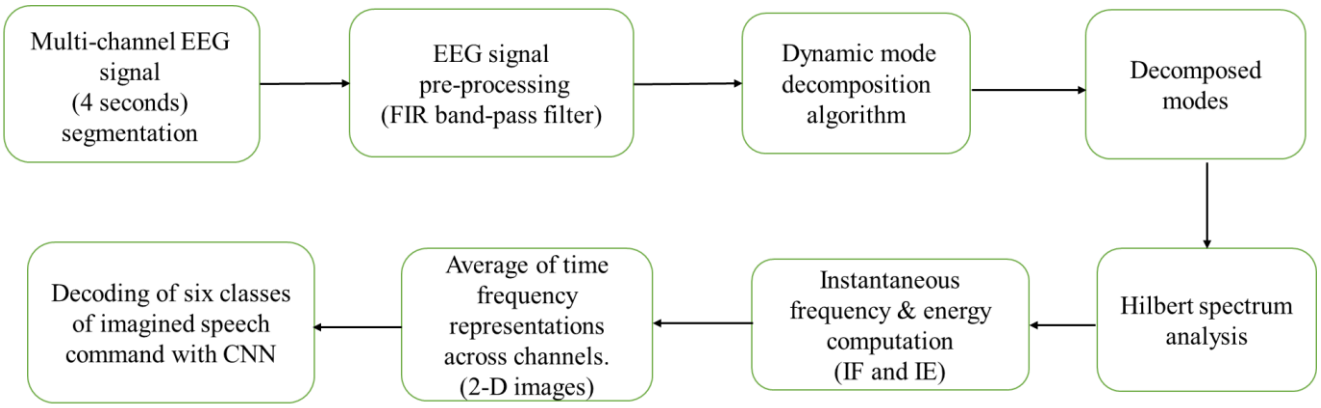


Figure 2: Block diagram of proposed DMD based MI-EEG classification framework.

The data collection and pre-processing phase involved acquiring multichannel EEG signals from 15 subjects while they performed imagined speech tasks [81]. The EEG signals were then pre-processed, including artifact removal techniques, to ensure the quality and reliability of the data. Next, DMD method was applied to extract dynamic information from the EEG signals. DMD is a powerful technique for decomposing complex signals into spatial modes and their corresponding instantaneous frequencies [17]. This step allowed for the identification of specific patterns and features related to imagined speech commands. To capture the time-frequency information of the EEG signals, time-frequency analysis using Hilbert spectral analysis was performed [24]. This analysis provided a comprehensive representation of the EEG signals in the time and frequency domains, enabling the extraction of relevant features for decoding imagined speech. The extracted features from DMD and time-frequency analysis were then combined using TFR averaging to compute multichannel EEG signal. This averaging process aimed to enhance the discriminative power of the features and improve the accuracy of decoding imagined speech commands [16].

Moreover, a CNN architecture was employed as a classification model. The CNN took the fused features as input and was trained and optimized to effectively decode the imagined speech commands. The model parameters were fine-tuned to achieve the best performance and accuracy [10,45]. Furthermore, methodology was evaluated using various evaluation metrics, such as accuracy, precision, and recall, to assess the effectiveness of the proposed approach in decoding imagined speech. Comparative analysis with existing methods was also conducted to demonstrate the superiority of the proposed method [7,86].

Conclusively, the methodology presented in this study offers a comprehensive and innovative approach for decoding imagined speech commands from EEG signals. It combines advanced signal processing techniques, such as DMD and time-frequency analysis, with a powerful CNN architecture, leading to improved accuracy and potential applications in brain-computer interfaces for individuals with speech impediments.

3.4 Dynamic Mode Decomposition (DMD)

This chapter provides an in-depth introduction to DMD, as a method that allows for the decomposition of complex, high-dimensional data into a set of coherent spatio-temporal modes. This chapter explores the fundamental principles and applications of DMD in various fields, highlighting its relevance and potential in solving complex problems. It discusses the underlying mathematical foundations of DMD, its advantages over traditional analysis methods, and its ability to capture dynamic behavior and extract dominant coherent structures from data. The chapter sets the stage for the subsequent chapters, where the application of DMD in specific contexts and its integration into novel frameworks will be explored in detail.

3.4.1 Overview of DMD

An overview of DMD is provided as a fundamental component of the proposed methodology [53]. DMD is a data-driven technique that has gained popularity in various scientific domains for its ability to analyse and extract meaningful information from high-dimensional and time-varying data [49]. The overview of DMD begins by introducing the underlying principles of the algorithm. DMD operates on the assumption that the data can be represented by a linear dynamical system. It decomposes the data into a set of spatial modes, which capture the spatial patterns or structures, and their associated temporal dynamics, which

describe how these spec patterns evolve over time [4]. The DMD algorithm is model-free, meaning it does not require any prior assumptions about the system or knowledge of its underlying dynamics [9].

The thesis discusses the key steps involved in applying DMD to MI-EEG signals. These steps include constructing the data matrix from the pre-processed EEG data, performing a singular value decomposition (SVD) on the data matrix to obtain the spatial modes and temporal dynamics, and reconstructing the data using the DMD modes. Additionally, techniques for selecting the appropriate DMD modes and reducing noise or artifacts in the data are also explored.

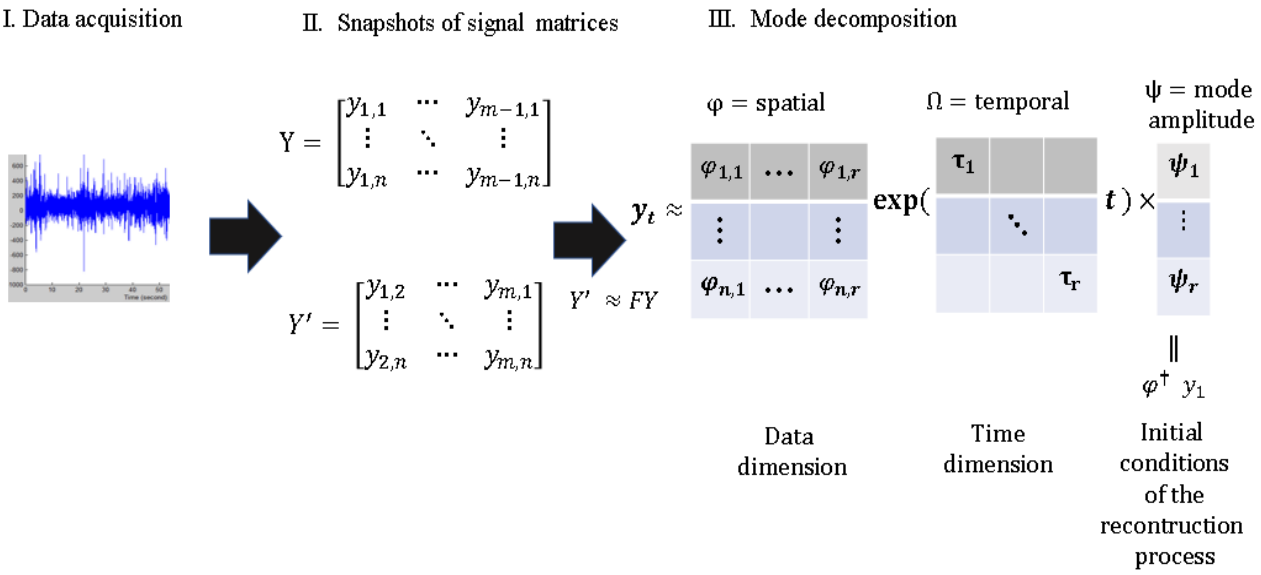


Figure 3**:** Operational view of DMD algorithm.

The work emphasizes the significance of DMD in the context of EEG-based BCIs for decoding imagined speech. By leveraging DMD, the proposed methodology aims to act relevant features from the EEG signals that are indicative of the imagined speech commands. These features can provide valuable insights into the brain patterns associated with speech processing, ultimately enhancing the accuracy and reliability of the decoding process. In the operational view above, To begin, data is collected from the system. Then, two large matrices, Y and Y', are created. The DMD method assumes that Y' is approximately equal to FY. Subsequently, the eigen decomposition of F is performed. This process results in obtaining spatial and temporal DMD modes, along with their coefficients or initial values. The DMD technique generates two spatial modes represented by hyperbolic functions for nonlinear coefficients, as well as two frequency modes represented by cosine functions. It is important to note that these modes can be complex numbers, but in this case, only the real part is visualized [87].

3.4.2 DMD System Modeling

Assuming the imagined speech stimuli is a continuous time dynamic system, the system can be described using its observed variable in EEG signals [82]:

$$\frac{dy}{dt}\Big|_{(t,y_0,t_0)} = \psi(y(t,y,t_0),t)\ ;\ t \in \mathbb{R} \qquad (5.1)$$

Where, y_0 = initial state, t_0 = initial time, $y(t,y,t_0)$ = dynamic system, $\forall t \in \mathbb{R}$ corresponding to motion and (y_0,t_0) = stationary point during t the imagined speech.

The EEG signals obtained from a single channel are represented as data vectors of length T. However, in the Dynamic Mode Decomposition (DMD) algorithm, data matrices of size n × m are processed. To convert the EEG channel data into an n × m size data matrix, EEG data of length m recorded from n different channels are utilized [31]. Here, n denotes the number of EEG segments used, and m represents the length of the data samples referred to as "snapshots". In this study, we introduce two distinct approaches, which are shown below, to obtain *n* × *m* dimensional brain signal data matrices of Y.

1. For the EEG-based dynamic mode decomposition single channel approach, we begin with a single-channel EEG signal denoted as Y = [Y_1 Y_2 ... Y_{t-1}] with a length of T samples. From this signal, we select M epoch-long non-overlapping EEG segments. Among these segments, we construct (n × m) EEG data matrices for each individual channel. In our specific experiments, we choose EEG segments that are 4000 epochs long (equivalent to 4 seconds) with no overlap, resulting in n = 35 snapshots. This process is repeated for all channels present in the EEG dataset [3].

2. For the Multi-channel EEG-based Dynamic Mode Decomposition (MDMD) approach, we construct (*N* × *M*) EEG data matrices by combining *M* epochs of *N* different EEG channels. In this particular analysis, we create concatenated EEG data matrices of size 35 × 4000 (n × m) by combining the data from six EEG channels: , C3, C4,F3, F4, P3, and P4. This results in a total of 6 × 35 × 4000 EEG data matrices. The EEG data is collected from the six channels for each of the fifteen subjects, encompassing a total of 24,000 epochs for the six imagined speech commands (Forward, Backward Up, Down, Left, and Right,).

For accurate capture of neurological activity dynamics, the number of measurements must be sufficient in comparison to the number of time points. The number of measurements (n) should be at least twice the number of time points (m), also known as snapshots [35]. With the original data matrix Y, a data augmentation process creates an augmented data matrix Y of size (k × l). Creating a Hankel matrix is the basis of augmentation, according to [2,33]. A matrix with dimensions K = 35 and L = 1000 was created in our experiments. With the augmented EEG data matrix, two new data matrices can be constructed: Y of size k + (l-1) and Y' aug, which is the time-shifted version of Y.

As the signal $y[n]$ is segmented into overlapping snapshots $\{Y_1, Y_2, Y_3, \ldots Y_{m-1}, Y_m\}$ as $y[n]$ changes with respect to time, were, $Y_i = [y(i)\ \ y(i+1) \ldots\ y(k-i+1)]$[67].

These portions of the signal are utilized to produce two $p \times (m-1)$ raw data Hankel matrices Y and Y' which is expressed as:

$$Y = [Y_1 \quad Y_2 \ \ldots \quad Y_{m-1}] \tag{5.2}$$

$$Y' = [Y_2 \quad Y_3 \ \ldots \quad Y_m] \tag{5.3}$$

Since the recorded signals are discrete values at specific time points, $Y_k\ (k\Delta t)$ is bounded by Δt, and the behaviour of the dynamic system is characterized by

$$Y_{k+1} = \mathrm{F}\ (Y_k). \tag{5.4}$$

While obtaining an analytical expression for F is challenging, we can approximate the dynamics linearly, $Y_{k+1} \approx \mathrm{F}Y_k$ from the signal recorded [13].

It is possible to compute the eigenvalues and eigenvectors of the best-fit linear operator using the DMD algorithm Y_{k+1} to Y_k ; expressed as:

$$Y' \approx FY \tag{5.5}$$

The operator T can be mathematically defined through the following expression:

F= argmin F $\| Y' - FY \|_F = Y' Y^{\dagger}$.Where, $\|.\|_F$ depicts A pseudo-inverse operation based on the Frobenius norm. Segmentation of signals $Y_k \subset \mathbb{R}^n, n^2$ in the operator matrix F is the number of elements and estimating its eigenvectors and eigenvalues may be difficult.

3.4.3 DMD Algorithm Stages

A comprehensive exploration of DMD for EEG signals is presented. DMD is utilized as a powerful tool to analyse the EEG data and extract meaningful information related to imagined speech commands. This high-level paragraph provides an overview of the application of DMD specifically for EEG signals in the context of the proposed BCI system [36].

The application of DMD for EEG signals involves several key steps. Firstly, the raw EEG signals are acquired using specialized equipment and electrode placements. The signals are then pre-processed to remove artifacts, noise, and irrelevant components. Following pre-processing, the EEG data is organized into a suitable matrix format for DMD analysis. The DMD algorithm is applied to this matrix to extract the spatial modes and associated temporal dynamics [12].

The thesis highlights the significance of DMD in capturing the underlying patterns and dynamics present in the EEG signals. By decomposing the EEG data into spatial modes, DMD enables the identification of distinct patterns related to specific cognitive processes or mental tasks, such as imagined speech. The temporal dynamics obtained from DMD provide valuable insights into the temporal evolution and synchronization of brain activity associated with speech processing.

Moreover, the thesis explores the optimization of DMD parameters and techniques to enhance the accuracy and reliability of decoding imagined speech commands from EEG signals. This includes the selection of relevant DMD modes and the integration of additional feature extraction methods to further improve the decoding performance [25].

By leveraging DMD for EEG signals, the proposed BCI system aims to decode the imagined speech commands with high accuracy and efficiency. The extracted spatial modes and temporal dynamics obtained from DMD serve as crucial input features for subsequent classification algorithms, facilitating the identification and interpretation of the user's intended speech commands as shown in the steps below. In the DMD, the significant eigenvalues and eigenvectors are computed without computing the matrix operator. DMD algorithm involves the following stages:

Step 1: To evaluate the singular value decomposition, select and retain only the r most significant eigenvalues, along with their corresponding eigenvectors, according to the criteria given.;

$$Y \approx \widetilde{U}\,\widetilde{\Sigma}\widetilde{V} \tag{5.6}$$

In which, $\widetilde{U} \in \mathbb{C}^{n\times r}$, $\tilde{\Sigma} \in \mathbb{C}^{r\times r}$ and $\widetilde{V} \in \mathbb{C}^{m\times r}$ $r \leq m$.

Step 2: The computation of the approximation matrix F can be accomplished by utilizing the pseudo-inverse of matrix Y in the following manner:

$$\mathrm{F} = \mathrm{Y}'\mathrm{Y}^{\dagger} = \mathrm{Y}'\,\mathrm{V}\,\tilde{\Sigma}^{-1}\,\widetilde{\mathrm{U}}^{*} \tag{5.7}$$

The following procedure will allow us to focus only the r significant eigenvalues and eigenvectors of T:

$$\tilde{F} = \widetilde{U}^{*}\,F\;\widetilde{U} = \widetilde{U}^{*}\,Y'\,\tilde{V}\,\tilde{\Sigma}^{-1} \tag{5.8}$$

Step 3: Matrix spectral decomposition of $\widetilde{\mathrm{T}}$ can be expressed as follows:

$$\tilde{F}W = W\Lambda \tag{5.9}$$

Step 4: Using the following equation, we can acquire DMD modes:

$$\phi = F'\,\tilde{V}\,\tilde{\Sigma}^{-1}\,W \tag{6.0}$$

As a result, the modes correspond to the eigenvectors of matrix F is given as Λ [46].

Dynamic mode decomposition, can be utilized to represent signals in terms of a data-adaptive spectral decomposition, which can be expressed as follows:

$$\widehat{\mathcal{X}} = \phi\Lambda^{k-1}\psi = \sum_{i=1}^{r}\phi_i\lambda_i^{k-1}\psi_j \tag{6.1}$$

In which, ϕ_i is the i^{th} column vector of matrix ϕ, λ_i represents the i^{th} diagonal elements of the diagonal matrix Λ, and the mode of amplitude b is evaluated as $\psi = \phi^{\dagger}\mathrm{x}_1$. $\widehat{\mathcal{X}}$ depicts the diagonal averaging operation is not typically performed as part of the DMD algorithm. The diagonal function, on the other hand, takes an input matrix and extracts its diagonal elements $\widehat{\mathcal{X}}$ and evaluates a mode vector using diagonal averaging. The reconstructed trajectory matrix $x(k)$ is Hankelized, in order to ensure that anti-diagonal elements are equalized. For k by matrix $\widehat{\mathcal{X}}$,the Hankelisation operator $\mathbb{H}$ is defined as,

$$\mathbb{H}\widehat{\mathcal{X}} = \widetilde{\mathcal{X}} = \begin{bmatrix} \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_L \\ \tilde{x}_2 & \tilde{x}_3 & \cdots & \tilde{x}_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_k & \tilde{x}_{k+1} & \cdots & \tilde{x}_{k+l-1} \end{bmatrix} \quad (6.2)$$

The desired or reconstructed time-series or elementary components can be expressed as follows::

$$\Re(\tilde{x}_k) = \frac{1}{num(D_k)} \sum_{m,n \in D_k} \widetilde{\mathcal{X}}_{m,n} \quad (6.3)$$

Here k is the index of the time series, $\Re(.)$ Is the real part of $\tilde{x}_k$, $num(.)$ Is the number of combinations of $(m,n), m+n = k+1 \; and \; D_k$ is given by $\{(m,n): 1 \le m \le K, 1 \le n \le L, m+n = k+1\}$[40,101].

This process finally provides an exact expansion of the elementary components or decomposed modes $z_{x_i}^{dec}[n]$ into the number of components to extract r that satisfies as follows:

$$z_{x_i}^{dec}[n] = \sum_{k=1}^{r} \widetilde{\mathcal{X}} \quad (6.4)$$

Overall, the utilization of DMD for EEG signals in the proposed BCI system offers a promising approach for decoding imagined speech. The thesis presents a detailed exploration of the application of DMD to EEG data, emphasizing its effectiveness in capturing the complex dynamics of brain activity associated with speech processing and its potential contribution to the development of advanced EEG-based communication assistance tools.

3.5 Feature Extraction from DMD modes

DMD has proven to be a valuable technique for feature extraction in imagined speech BCIs. Imagined speech BCIs aim to decode a person's intended speech by analysing their EEG signals, without the need for actual vocalization [88]. DMD provides a data-driven approach for extracting discriminative features from EEG signals, which can then be used to classify different imagined speech tasks.

In the context of imagined speech BCIs, DMD can extract relevant temporal and spectral features from the EEG signals that capture the neural activity associated with speech production [36]. By decomposing the EEG signals into a set of spatial and temporal modes, DMD identifies the dominant oscillatory

patterns and their associated frequencies related to imagined speech. These patterns can capture the neural signatures of phonemes, syllables, or even whole words.

The extracted DMD features can be further processed and used as inputs for classification algorithms, such as SVMs or CNNs [6]. By training these classifiers on labelled EEG data from different imagined speech tasks, the BCI system can learn to recognize and differentiate between different speech-related intentions. This enables users to control external devices or communicate using only their thoughts.

The use of DMD for feature extraction in imagined speech BCIs offers several advantages. First, DMD is a data-driven method, meaning it adapts to the specific characteristics of the individual's EEG signals, making it more personalized and accurate. Second, DMD captures both temporal and spectral information, providing a comprehensive representation of the underlying neural dynamics involved in imagined speech. Lastly, DMD allows for real-time analysis, making it suitable for online BCI applications.

Conclusively, the application of DMD for feature extraction in imagined speech BCIs holds great potential for improving the accuracy and usability of these systems. By harnessing the power of DMD, researchers and engineers can enhance the performance of imagined speech BCIs and pave the way for new methods of communication and control for individuals with speech-related impairments [10].

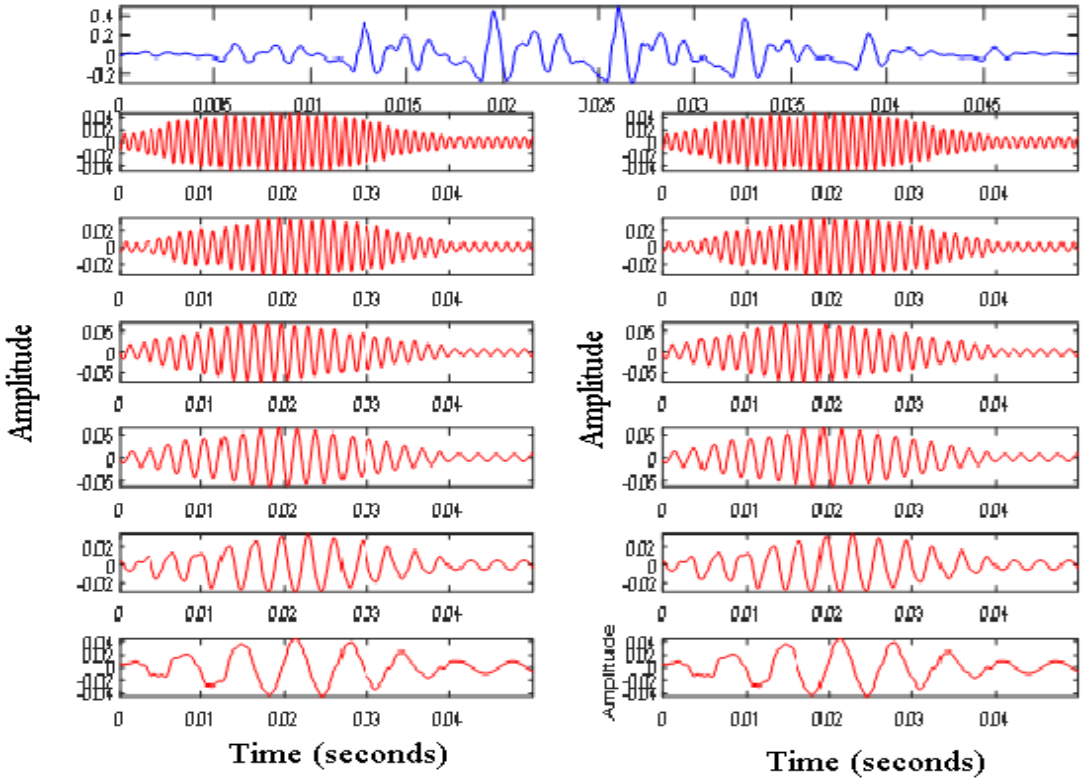


Figure 4: Decomposed modes for Forward class from F-3 channel EEG signal of the subject S01.

3.5.1 Time-Frequency Representation using Hilbert Spectrum Analysis

In [73], the time-frequency analysis using Hilbert spectral analysis is employed as a powerful tool for capturing the temporal and frequency characteristics of EEG signals during the decoding of imagined speech commands [10]. This high-level paragraph provides an overview of the time-frequency analysis techniques utilized in the proposed BCI system. Time-frequency analysis aims to reveal how the frequency content of the EEG signals changes over time, providing valuable insights into the dynamic nature of brain activity during the imagined speech tasks [17]. Hilbert spectral analysis is a widely used method for time-frequency analysis, capable of extracting time-varying spectral information from non-stationary signals [71].

The EEG signals recorded during the imagined speech tasks are decomposed into their constituent frequency components using techniques such as wavelet transforms or STFT [22]. The Hilbert transform is then applied to each frequency component to obtain the instantaneous amplitude and phase information

at each time point [66]. By combining the amplitude and phase information, time-frequency representations (TFRs) are derived, which provide a detailed and dynamic characterization of the EEG signals [115]. These TFRs reveal the temporal evolution of spectral power across different frequency bands, capturing the changes in neural activity associated with different stages of speech processing.

The TFRs are particularly effective in capturing event-related spectral changes, such as the modulation of alpha and beta frequency bands during the generation of imagined speech commands. Furthermore, the TFRs derived from Hilbert spectral analysis are integrated with the spatial modes obtained from DMD, enhancing the representation of imagined speech commands. The combination of spatial information from DMD and temporal-frequency information from Hilbert spectral analysis provides a comprehensive and informative feature representation for decoding imagined speech commands.

By leveraging time-frequency analysis using Hilbert spectral analysis, the proposed BCI system aims to capture the dynamic nature of brain activity during the generation of imagined speech commands. The TFRs derived from this analysis offer valuable insights into the temporal and frequency characteristics of the EEG signals, enabling effective discrimination and classification of different speech commands.

To obtain the Dynamic Mode Decomposition Time-Frequency Representation, the instantaneous energy (IE) and instantaneous frequency (IF) of the decomposed modes are computed and represented in the time-frequency plane [11]. This can be accomplished using the Hilbert transform separation algorithm (HTSA) in the following manner:1. Compute the analytical signal $z_{x_i}^{dec}[n]$ of the decomposed modes of the DMD algorithm $\phi_i^{dec}[n]$ is computed as follows:

$$z_{x_i}^{dec}[n] = \phi_i^{dec}[n] + j\mathcal{H}\{\phi_i^{dec}[n]\} = a_i[n]e^{+j\psi_i[n]} \quad (6.5)$$

Where, $\mathcal{H}$= Hilbert spectrum operation, $\psi_i[n]$ = unwrapped instantaneous phase of the signal, $a_i[n]$ = amplitude envelope.

2. The IE of the signal $\phi_i^{dec}[n]$ is obtained from the expression:

$$e_i[n] = \left|z_{x_i}^{dec}[n]\right|^2 = a_i^2[n] \quad (6.6)$$

3. The IF of the signal $\phi_i^{dec}[n]$ can be derived from the expression as follows:

$$f_i[n] = \frac{f_s}{2\Pi}\left(\angle\phi_i^{dec}[n]\right)' = \frac{f_s}{2\Pi}(\psi_i[n] - \psi_i[n-1]) \quad (6.7)$$

4. The IE ($e_i\ [n]$) and IF ($f_i[n]$) parameters of the decomposed modes ($\phi_i^{dec}[n] \forall i \in [1 \dots I]$) are expressed in the time-frequency plane is given as,

$$\mathrm{Y}[n, \Omega] = \sum_{i=1}^{I} Y_i[n, \Omega] \quad (6.8)$$

$$\mathrm{Y}_i[n, \Omega] = e_i\ [n]\delta\left[\Omega - \frac{2\Pi}{f_s} f_i[n]\right] \quad (6.9)$$

$$\mathrm{TFD}_{\mathrm{MDMD}} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{Y}_i[n, \Omega]) \quad (7.0)$$

Where , $\delta[.] =$ Dirac delta function, after obtaining the $\mathrm{Y}_i[n, \Omega]$ matrix,

$\mathrm{Y}[n, \Omega] =$ Hilbert spectrum $\mathrm{TFD}_{\mathrm{MDMD}}$ = average TFD-DMD modes of *N*-channel

The time-frequency representations of imagined speech EEG data were computed for each subject in order to identify speech-related brain activities. The TFR of each trial was calculated using a Hilbert transform and then averaged across all trials. Among the five subjects, we specifically plotted the TFRs of subjects 2 and 5, which exhibited distinct patterns in the gamma frequency range. As depicted in Figure 4, the power of the alpha and beta frequency bands (10-30 Hz) was analysed for six different imagined commands: Up, Down, Left, Right, Backward, and Forward [98].

3.5.2 Average Time Frequency Representation Computation

Multichannel EEG signal TFR averaging is significant in enhancing spectral resolution; improving the quality of the EEG data. TFR averaging involves the computation of time-frequency representations, such as spectrograms or scalograms, for each individual EEG channel. These representations capture the dynamic changes in spectral content over time for each channel [74].

By the time-frequency representations across multiple channels, a more robust and reliable representation of the underlying brain activity is obtained. TFR averaging aims to minimize the effects of noise and artifacts, enhance the signal-to-noise ratio, and reveal consistent spectral patterns across the multichannel EEG signals. The resulting averaged TFR provides valuable insights into the neural dynamics associated with imagined speech, enabling more accurate and reliable decoding of speech-rela

from the EEG signals. With the aid of multichannel EEG signal TFR averaging, the thesis aims to enhance the overall performance of the imagined speech EEG-based BCI system and enhance its usability as a n aid for individuals with speech impairments [10, 92].

TFR averaging enhances the spectral resolution of EEG signals by averaging the time-frequency representations obtained from individual EEG channels. In TFR averaging, the time-frequency representations, such as spectrograms or scalograms, are computed separately for each EEG channel. These representations capture the changes in spectral content over time for each channel. To improve the spectral resolution, impede the f noise and artifacts, the time-frequency representations of multiple channels are averaged together.

The averaging process combines the spectral information from different channels to create a more robust and accurate representation of the underlying brain activity. By integrating the individual time-frequency representations, TFR averaging aims at enhancing the signal-to-noise ratio and reveal consistent spectral patterns across multiple channels.

$$\mathrm{TFR}_{\mathrm{MDMD}} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{Y}_i[n, \Omega]) \qquad (7.1)$$

Where $\mathrm{TFR}_{\mathrm{MDMD}}$ is average of the TFR-DMD modes from *N*-channel.

The time-frequency analysis using Hilbert spectral analysis combined with multichannel EEG signal averaging serves as a crucial step in extracting and representing the relevant neural information underlying imagined speech. By applying Hilbert spectral analysis to the multichannel EEG signals, the temporal and spectral dynamics of brain activity can be captured with high resolution.

This analysis technique allows for the extraction of time-varying frequency components that are indicative of the imagined speech process. Additionally, by averaging the time-frequency representations across multiple EEG channels, the thesis aims to enhance the signal quality, reduce noise and artifacts, and improve the reliability of the extracted features. The multichannel EEG signal averaging process helps to mitigate channel-specific variations and reinforces the shared spectral patterns related to imagined speech across the different channels.

The resulting TFRs provide valuable insights into the spatio-temporal dynamics of the neural activity associated with imagined speech, enabling more accurate and robust decoding of the speech commands.

Through the integration of time-frequency analysis using Hilbert spectral analysis and multichannel EEG signal averaging, this thesis aims to advance the understanding and decoding of imagined speech for the development of effective EEG-based BCI systems for communication assistance [14].

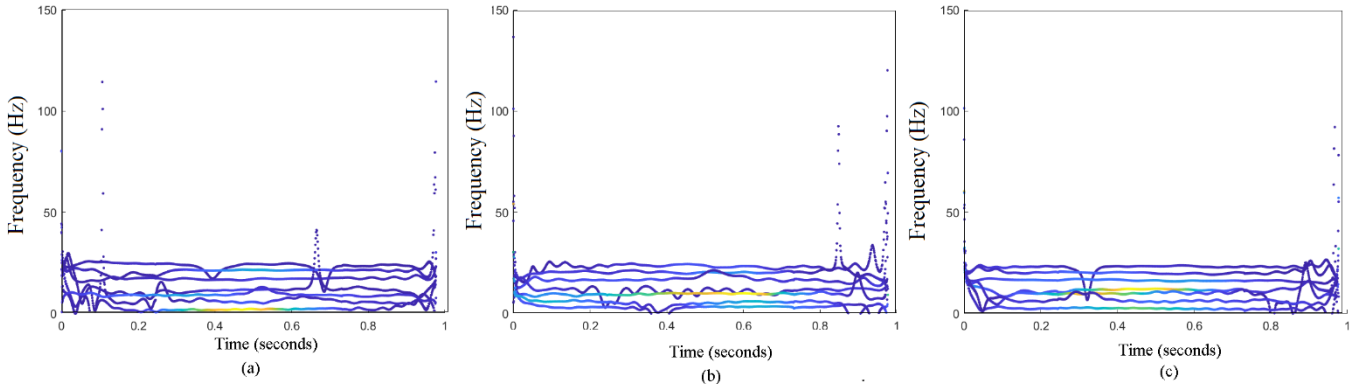


Figure 5**:** The TFR plots of subject S01 for (a) Right, (b) Left, and (c) Up commands.

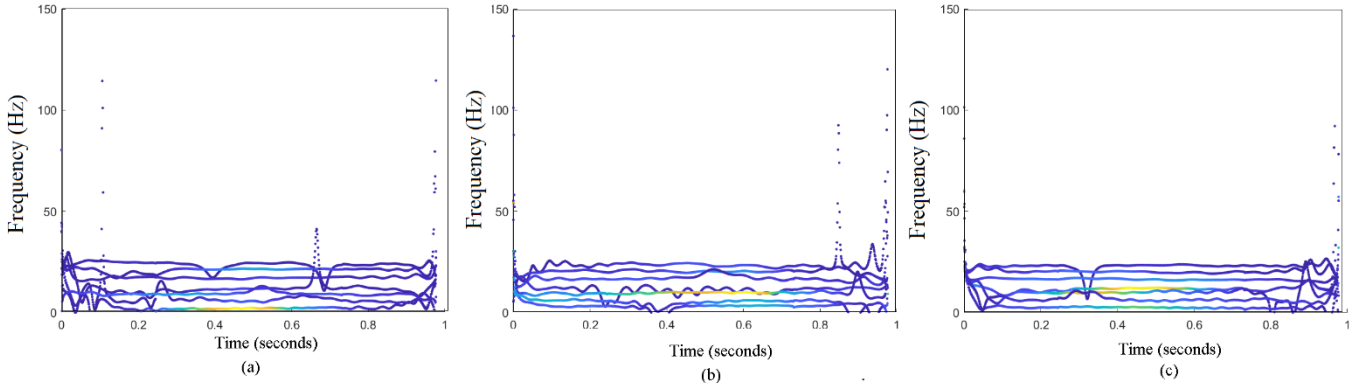


Figure 6: The TFR plots of subject S01 for (a) Forward, (b) Backward, and (c) Down commands.

3.6 Classification framework modeling

The classification of imagined speech involves the utilization of convolutional neural networks (CNNs). CNNs are a specific type of neural network architecture, which can be considered as a specialized case of convolutional feed forward neural networks. In CNNs, the convolutional layers are composed of neurons arranged in a rectangular grid. It is important to note that for the proper functioning of these layers, the previous layer must also be in the form of a rectangular grid. Within a convolutional layer, each neuron receives inputs from a specific rectangular section of the previous layer. Notably, the weights assigned to this rectangular section are shared among all neurons within the convolutional layer. This weight sharing property ensures that the same features can be detected across different spatial locations, enabling the CNN to efficiently capture local patterns or characteristics in the input data.

1. Convolutional layer: The convolutional layer performs image convolution on the previous layer by applying convolution filters specified by the weights. Furthermore, each convolutional layer can consist of multiple grids, where each grid takes inputs from all grids in the previous layer. It is possible to use different filters for each grid, allowing the convolutional layer to capture diverse features from the input data.

   Mathematically, when performing a discrete convolution (x*w)(a) on functions x and w, it is more realistic to assume that the parameter t is discrete due to the nature of image sensors. Assuming this, we can define the discrete convolution as follows:

$$(x * w)(a) = \sum_a x(t)w(t - a) \qquad (7.2)$$

   To compute the discrete convolution at position a, we sum up the products of corresponding elements from x and w, taking into account all valid values of t within the defined range. This process captures the relationship between the input functions and generates the convolution result. By acknowledging the discrete nature of image sensors and using this definition of discrete convolution, we can accurately represent the convolutional operations involved in digital image processing.

   Where a runs over all values in the space. In deep learning, usually x is a multidimensional array of data convolved with a Gaussian function w, and the kernel w involves learnable parameters and usually has finite support.

Equation 7.2, which defines the convolution operation, is dimension-independent. However, in our case, we are using Time-Frequency Representations as the input data, which can be thought of as 2-dimensional images. Consequently, in the proposed method, we will be employing 2-dimensional convolutions. This means that we will be applying convolutional operations specifically designed for 2-dimensional data to process and extract relevant features from the TFRs:

$$(I * K)(I, J) = \sum_m \sum_n I(m, n, j - n)\, K(m, n) \qquad (7.3)$$

Considering the fact that the support of K is finite, the priori infinite sum is now infinite.

2. Max-Pooling: Max-pooling is a technique commonly used after convolutional layers in deep learning. It involves extracting small rectangular blocks from the preceding convolutional layer and subsampling them to generate a single output for each block. Various methods exist for performing pooling, including taking the average, the maximum value, or using a learned linear combination of the neurons within the block. In our case, we will exclusively employ max-pooling layers, which select the maximum value within each pooled block.

3. ReLU: This nonlinearity is defined as follows:

$$\text{ReLU} = \max(0, x), x \in \mathbb{R}. \qquad (7.4)$$

It is easy to see that $\text{ReLU}(x)' = 1$ for $x > 0$ and then $\text{ReLU}(x)' = 0$ for $x < 0$. The ReLU (Rectified Linear Unit) nonlinearity is known to promote faster convergence when compared to sigmoid or tanh nonlinearities. It is particularly effective in convolutional neural networks (CNNs) when combined with carefully chosen weight initialization strategies and learning rates. ReLU activation functions are widely used due to their ability to eliminate the vanishing gradient problem and efficiently model complex patterns in the data, contributing to improved training speed and overall performance.

4. SoftMax: The SoftMax nonlinearity is a specialized activation function that differs from the general nonlinearities mentioned previously. It is specifically designed for multi-class classification tasks. The SoftMax function takes a vector of real numbers as input and transforms it into a probability distribution over multiple class. It computes the exponentiation of each element in the input vector and normalizes the results to ensure they sum up to 1. This normalization allows the SoftMax function to assign probabilities to each class, indicating the likelihood of an input belonging to a

particular class. The SoftMax nonlinearity is commonly used as the final layer in neural networks for multi-class classification problems. It is defined as:

$$\text{softmax}(x_i) := \frac{\exp(x_i)}{\sum_{j=i}^{n} \exp(x_i)} \quad x \in \mathbb{R}^n \qquad (7.5)$$

to a probability vector of length n, a vector $x \in \mathbb{R}^n$ is mapped. Therefore, the nonlinearity is applied in the classification tasks, next to the fully connected layer with n outputs for n classes

4. Pooling layers are essential components in convolutional neural networks (CNNs) that effectively reduce parameters and computations, leading to improved efficiency and mitigating overfitting risks. Two widely used types of pooling layers are max pooling and average pooling. In max pooling, each rectangular neighbourhood around a point (i, j) (or (i, j, k) for 3D data) in the input feature map identifies and retains the maximum value, discarding the remaining values. Conversely, average pooling calculates the mean value within each neighbourhood. A common configuration for max pooling involves a stride of 2 and a kernel size of 2. This operation divides the feature map into a regular grid of square or cubic blocks with a side length of 2. Within each block, the pooling layer selects the maximum or average value for each input feature. Pooling operations are frequently employed in CNNs to reduce the spatial dimensions of the feature map. However, it is worth noting that comparable outcomes can be achieved by using 3 × 3 convolutions with a stride of 2, particularly when working with 2D data. This alternative approach presents an option for spatial dimension reduction in the network.

5. The fully connected layer is an integral part of neural networks, where the number of input dimensions is denoted as "n" and the number of output dimensions is denoted as "m". The layer output is determined by two key parameters: the weight matrix W, which has m rows and n columns and belongs to the set of real numbers (R), and the bias vector b, which belongs to the set of real numbers (R) and has a dimension of m. Given an input vector x from the set of real numbers (R) with a dimension of n, the output of the fully connected layer FC with an activation function f can be expressed as follows:

$$FC(x) := f(Wx + b) \in R^m \qquad (7.6)$$

In the given formula, Wx represents the matrix product of the weight matrix W and the input vector x. The activation function f is then applied component-wise to the resulting vector.

Fully connected layers are frequently employed as the last layers in classification tasks, particularly in conjunction with convolutional neural networks (CNNs). In such cases, one or two fully connected layers are typically added on top of the CNN. To achieve this, the output of the CNN is flattened and treated as a single vector. This arrangement enables the fully connected layers to receive the extracted features from the CNN and generate predictions based on them. Another context where fully connected layers find utility is in various autoencoder architectures. In these scenarios, FC layers are often connected to the latent code in both the encoder and decoder paths of the network. This configuration facilitates the learning and reconstruction of the input data by the network.When working with CNNs, it is worth noting that applying a convolution filter with a kernel size of 1 to a feature map with n channels is equivalent to employing a fully connected layer with m outputs at each point in the feature map, where m represents the number of output channels. This approach allows for efficient transformation of information across the feature map.

3.7 Implementation of classification framework

The CNN architecture was proposed by [46, 96] as a type of feedforward artificial neural network that mimics neurons in the visual cortex region of the brain. Specifically, it is used for supervised deep-learning tasks in image recognition. To classify imagined EEG signals, we used CNN to classify two-dimensional Time-Frequency Representations (TFRs). Below is a description of the classification procedure. To begin with, we went through all the imagined speech commands to gather a collection of TFRs. The model was then evaluated on the remaining 20% of images, selecting each label at random. There are fifteen layers in our CNN architecture, as outlined in Table 5. We briefly describe each layer while providing hyperparameters and learnable parameters.

a) Image input layer: There are 656 pixels of height, 875 pixels of width, and 3 pixels of RGB channels in the input images. This layer subtracts the mean of each image after propagation.

b) 1st Convolution layer: An 8x3 convolutional filter (CF) is used in the first convolutional layer to process the input image. At the borders of the image, there is no padding applied, and the stride in both vertical and horizontal directions is set to 1. Initial weights and biases are set to zero using the glorot function.

c) Batch Normalization layer: During batch normalization, the input channel values are normalized by subtracting the mean from the variance (σ). For avoiding division by zero, a small epsilon (ε) is added to the variance. A mini-batch of normalization is performed on each input channel.

d) ReLU layer: This layer is based on the normalized batch layer and is subjected to the threshold operation in ReLU.

e) Max pooling layer: In this layer, the image dimension is reduced by dividing it into rectangular regions (pools) and selecting the maximum value in each region. Pool size is one, the stride is two.

f) 2nd Convolution layer: We introduce a second convolutional layer after the max pooling layer that has 16 CFs of size 3×3. The remaining parameters remain unchanged from the first convolution layer. In addition to the convolution layer, batch normalization, ReLU, and maximum pooling layers are added afterward

g) 3rd, 4th, 5th, and 6th Convolution layers: The model includes four additional convolutional layers, each with 32, 64, 120, and 256 clocks of size 3x3. As with the previous convolutional layers, hyperparameters and learnable parameters follow a similar pattern. Each of these convolutional layers is followed by batch normalization and ReLU layers.

h) Fully connected layer: For classification, this layer combines the features learned from the preceding layers. Using the glorot function, the initial weights are calculated, and the bias is set to zero. In this case, there are six classes to be classified, so the output size corresponds to that.

i) SoftMax layer: After the fully connected layer, the SoftMax activation function generates six-class classification probabilities.

j) Classification layer: As a result of the SoftMax function, this layer assigns one of six classes to the output. In order to assign classes, the cross-entropy loss function is used. We present a CNN architecture for classifying time-frequency representations (TFRs) of imagined EEG signals in this section. Each layer in the architecture has its own hyperparameters and learnable parameters. Convolutional neural networks start with an image input layer, followed by layers such as batch normalization, ReLU, and max pooling. The pattern repeats for multiple convolutional layers, increasing the number of filters in each layer as the number of layers increases. In the final layer, the learned features are consolidated for

classification using a fully connected layer. The output probabilities are generated by a SoftMax layer, and the class labels are assigned by a classification layer.

The CNN architecture described here is specifically designed to process Time-Frequency Representations (TFRs) as 2D images and accurately classify them into one of six imagined speech command classes. The configuration and hyperparameters of the network have been carefully optimized to improve its learning and classification capabilities, particularly when dealing with TFRs effectively.

The objective of utilizing this CNN architecture in the thesis is to enhance the accuracy and reliability of the EEG-based brain-computer interface (BCI) system for imagined speech recognition. By combining TFRs obtained through Dynamic Mode Decomposition (DMD) with the CNN model, the thesis aims to achieve precise identification and classification of different imagined speech commands. This technological advancement holds great potential in providing effective communication assistance for individuals with speech impairments.

Chapter 4

Experimental Protocol

The experimental protocol for EEG-Based BCIs, specifically for Imagined speech EEG-based BCI using DMD, involves several key components and procedures to elicit the imagined speech task.

Firstly, in our work, we have employed the dataset Correto dataset [80], the participants are provided with specific instructions and cues, such as imagining saying certain words or sentences silently. These instructions may vary depending on the specific goals of the study or the intended application of the BCI system. The EEG signals are recorded continuously during the imagined speech task, typically at a high sampling rate to capture fine temporal details.

A cohort of participants, comprising individuals with or without speech-related impairments, is recruited for the study, ensuring ethical considerations and obtaining informed consent from all participants beforehand [12]. Throughout the experiment, the participants are comfortably seated in a controlled environment, aiming to minimize external disturbances. Electrodes are positioned on the scalp following the international 10-20 system or an appropriate electrode placement scheme [16]. These electrodes are subsequently connected to an EEG amplifier or acquisition system, facilitating the capture of brain's electrical activity [31]. To refine the collected EEG data, preprocessing techniques such as filtering, artifact removal algorithms, and signal denoising methods are employed to eliminate noise, artifacts, and other undesirable signals.

Subsequently, the preprocessed EEG data is divided into segments known as epochs, which correspond to specific time intervals or tasks. These epochs serve as the basis for extracting features utilizing the Dynamic Mode Decomposition (DMD) algorithm or other applicable signal processing methods. By applying the DMD algorithm, the EEG signals are decomposed into modes that capture the relevant temporal and spectral characteristics associated with imagined speech. These extracted DMD features are then utilized to train a classification model, such as a convolutional neural network (CNN), with suitable configuration of machine learning parameters. The training process involves establishing a mapping between the DMD features and the corresponding imagined speech tasks or intentions, resulting in a model capable of classifying new EEG data based on learned patterns [6, 69]. The performance of

the Brain-Computer Interface (BCI) system is subsequently assessed using diverse metrics such as accuracy, precision, recall, or information transfer rate.

The EEG-Based Brain-Computer Interface (BCI) setup employing Dynamic Mode Decomposition (DMD) for imagined speech encompasses several key components and procedures. These include participant recruitment, electrode placement, execution of imagined speech tasks, acquisition of EEG data, preprocessing, feature extraction using DMD, training of a classification model, and evaluation of performance [12]. This comprehensive setup offers a systematic framework for examining the viability and efficacy of the proposed BCI system in decoding imagined speech intentions from EEG signals [7, 110].

4.1 Brain waves (EEG)

In many research work the investigation of different brain wave patterns and their correlation with imagined speech has been found for the classification of imagined speech analysis. By studying the specific brain wave frequencies and their variations during imagined speech tasks, researchers can gain a deeper understanding of the neural mechanisms underlying speech production and representation.

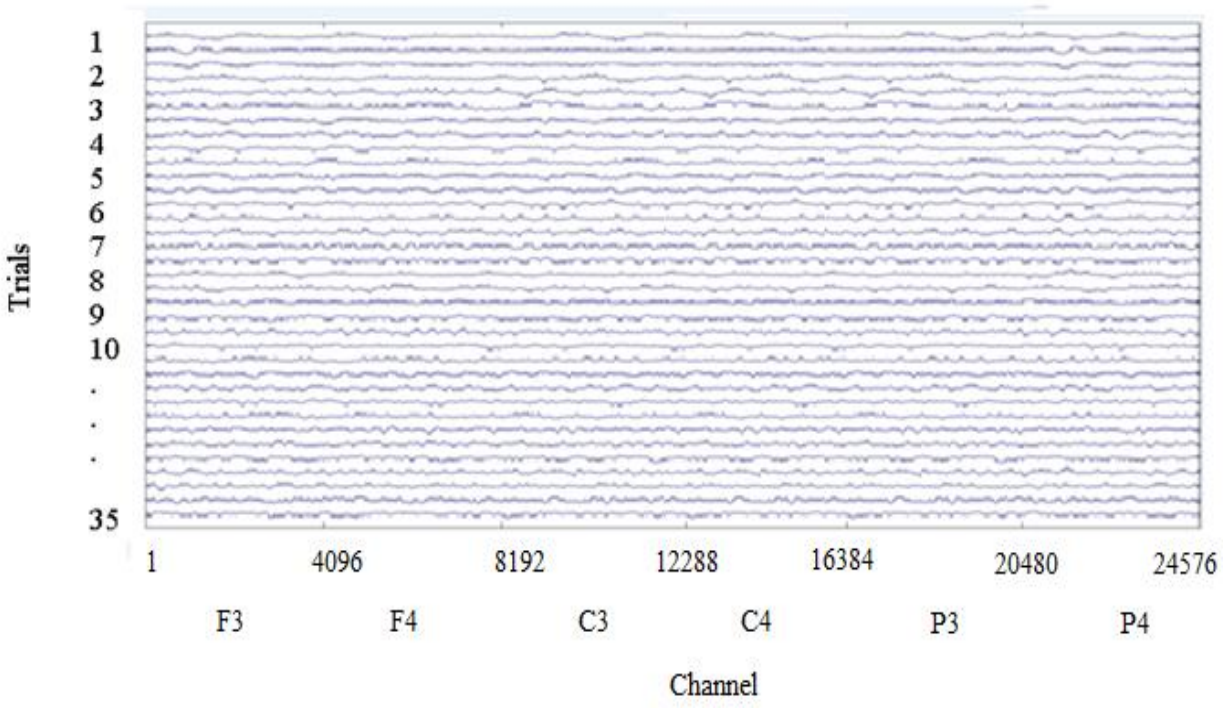


Figure 7: EEG plot for imagined speech command Up for the subject S01 using EEGLAB.

Considering the role of alpha, beta, theta, and gamma waves, among others, in the context of imagined speech can provide valuable insights into the neural processes involved and inform the design of more precise and accurate decoding algorithms.

Exploring the dynamic changes in brain wave patterns over time and their relationship with the quality of imagined speech decoding is an important avenue for further research. Investigating the temporal dynamics of brain waves during the production of different speech sounds or the transition between speech commands can shed light on the time-varying nature of imagined speech representations. This knowledge can be leveraged to improve the temporal resolution and decoding performance of EEG-based BCIs.

Moreover, brain waves exhibit variations based on the level of activity within different regions of the cerebral cortex and can undergo significant changes during wakefulness, sleep, and coma. In some cases, brain wave patterns recorded in EEG recordings may appear irregular, lacking a distinct and specific pattern. Brain waves can be broadly categorized into five types based on their frequencies: Delta waves (0.4-4 Hz), which are commonly observed in sleeping adults, premature infants, or individuals with subcortical lesions. In adults, delta waves tend to be prominent in the frontal region, while in children, they are more prevalent in the posterior region. Theta waves (4-8 Hz) are often observed in children and adults experiencing emotional stress or individuals with deep midline disorders, and they are typically localized in the parietal and occipital regions. Alpha waves (8-13 Hz) are present during a relaxed and resting state but are not typically observed during sleep. They are predominantly seen in the occipital region. Beta waves (13-30 Hz) manifest during active, busy, or mentally engaged states, including periods of concentration or anxious thinking. They are commonly found in the frontal and parietal regions. Gamma waves (26-100 Hz) are associated with specific cognitive or motor functions.

In our research, we employed the Dynamic Mode Decomposition (DMD) method on recorded EEG data samples related to imagined speech. We extracted the original signal components using DMD and further obtained the rhythms mentioned above. Figure 7 illustrates the representation of EEG waves from an openly accessible database depicting an imagined command [80].

EEG signals are decomposed into spatial-temporal modes, capturing the underlying rhythmic activities in the brain. The DMD modes are ranked based on their eigenvalues, with larger eigenvalues indicating dominant rhythms. These modes represent different frequency components, such as alpha, beta, theta,

and gamma waves. The selected modes can be used to reconstruct the original EEG signals or analyze specific frequency components of interest.

4.2 Data Acquisition

The acquisition of EEG signals is a critical step in the development of imagined speech-based Brain-Computer Interface (BCI) systems. This stage involves capturing the electrical activity of the brain using specialized electrodes and amplification equipment, as described in the openly accessible database utilized in this research [80]. The EEG acquisition process begins by carefully preparing the subject and creating a comfortable and relaxed environment to minimize external interferences. Ag-AgCl cup electrodes are then attached to specific scalp locations based on the internationally recognized 10-20 system, ensuring consistent and standardized electrode placement across all subjects.

To ensure accurate and reliable signal acquisition, conductive paste is applied between the electrodes and the scalp. This facilitates good electrical contact, enabling the detection of subtle electrical potentials generated by the brain during imagined speech tasks. No electrode cap is used to avoid introducing additional impedance that could affect the quality of the acquired signals. The placement of the electrodes is strategically chosen to target cortical areas associated with language processing while minimizing interference from muscle activity related to speech production. For reference and ground, electrodes positioned on the left and right mastoids are utilized, respectively.

For capturing the EEG signals, a high-quality analogue amplifier is employed. The amplifier provides multiple channels, allowing for simultaneous recording of signals from specific electrode positions. In this thesis, six channels (F3, F4, C3, C4, P3, and P4) are selected to capture the relevant neural activity associated with imagined speech commands [80]. The amplified EEG signals are then converted into a digital format using an analogue-to-digital converter board, ensuring accurate representation of the electrical activity. The signals are sampled at a high-rate of 1024 Hz, enabling the capture of detailed temporal information necessary for subsequent analysis and decoding processes.

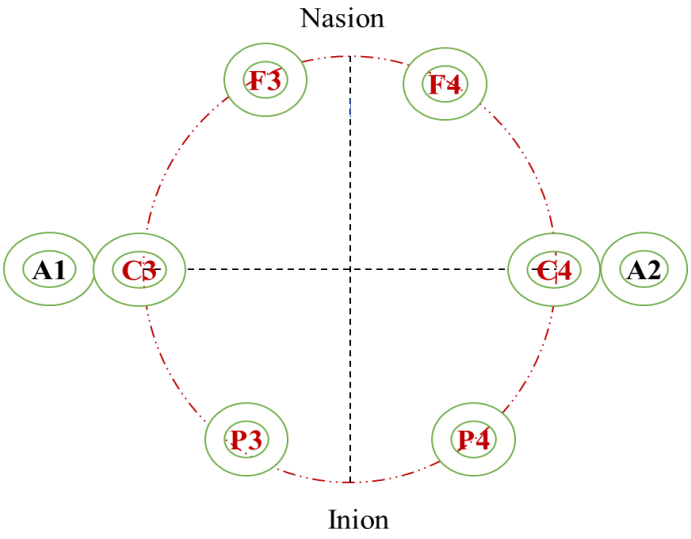


Figure 8: International 10-20 EEG electrode map.

The database was built using a custom protocol to ensure consistent testing conditions. Approximately a meter away from the LCD screen, which displayed visually described target words, participants sat comfortably in chairs. Participants wore headphones that provided an auditory stimulus corresponding to the intended word. In order to minimize intra-subject variation, data were collected during a single recording session per participant. The words used in this study were meticulously chosen from dictionaries. A dictionary of Spanish words represented potential commands associated with external device control in a BCI system. Up, down, right, left, forward, and backward are some of the command words [80].

The EEG signals were recorded in two different situations: during imaginary speech and during spoken speech. The choice of both modes was deliberate in order to facilitate future research aimed at identifying EEG patterns that distinguish between overt and covert speech. The EEG signals were recorded only in the imaginary speech mode, but the audio signals were also collected in the voice talk mode. Each participant performed 50 trials per word, with 40 trials allocated to imaginary speech mode and 10 trials allocated to pronounced speech mode.

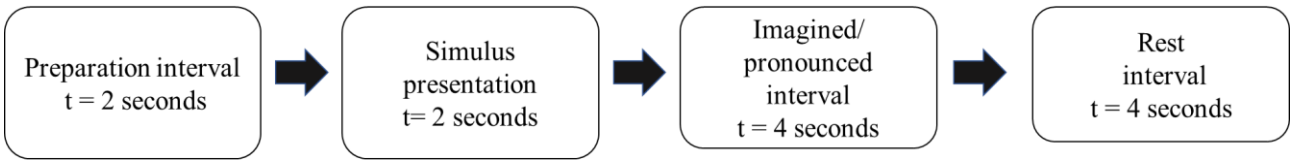


Figure 9: Sequence time course for the recording of one imagined stimulus command.

The experimental setup involved the presentation of target stimuli in a sequential manner, consisting of four predefined intervals of durations as depicted in Figure 9. Here is a description of these intervals:

In the ready interval (2 seconds), the subject is informed that the rest period has ended and a new cue will be presented soon. In this interval, the target word is visually and acoustically presented to the subject. An image indicates either the imagined or spoken word is to be imagined or pronounced in this stage. The task is performed throughout the entire 4-second interval for vowel words. To indicate when to imagine or pronounce a command word, a sequence of three audible clicks is introduced. In order to minimize artifacts in the recorded signals, this step is crucial for distinguishing imagined speech from pronounced speech.

An interval of rest (4 seconds) allows the subject to move, swallow, or blink, providing a break from task-related activities. We utilized fixed duration windows to facilitate pre-processing and feature extraction techniques, simplifying the recording process and ensuring consistent analysis. In order to maximize recording time, all words in the selected vocabulary were randomly presented twice in each block. The duration of blocks with vowel stimuli was 2 minutes, while the duration of blocks with command words was 2.43 minutes. To maintain focus during the extended recording process, regular breaks were incorporated. It took approximately 3.5 hours for both groups to repeat each word fifty times.

The EEG signal acquisition stage is crucial as it directly impacts the quality and fidelity of the acquired data. By employing standardized electrode placement, high-quality amplification, and precise sampling, this thesis aims to capture reliable and informative EEG signals to successfully implement the proposed DMD-based decoding framework.

4.3 EEG Signal Preprocessing

The accurate removal of artifacts from EEG signals is crucial to ensure the reliability and quality of the recorded data [93]. Various techniques are employed to effectively identify and mitigate artifacts that may contaminate the EEG recordings during the experimental sessions. One commonly encountered artifact in EEG signals is muscular activity, which can originate from facial movements, jaw clenching, or other muscle-related activities. To address this, careful attention is paid to electrode placement to minimize the proximity to muscles involved in speech production. Additionally, visual scrutiny of the recorded signals is performed, and any registers containing muscular artifacts within the Imagine/pronounce interval are marked for elimination.

The removal of these artifacts is crucial as they can obscure the underlying EEG patterns or introduce noise that affects subsequent analysis. Another type of artifact that is considered is electrode-related artifacts, such as electrode pops or saturation artifacts. These artifacts can be caused by electrode impedance changes or saturation of the amplifier due to excessive signal amplitudes. Similar to muscular artifacts, registers containing such artifacts within the Imagine/Pronounce interval are identified and marked for elimination. Blinking artifacts are also taken into consideration during artifact removal. While blinking can introduce noise in the recorded EEG signals, techniques like independent component analysis (ICA) can be employed to separate these artifacts from the signal of interest. Therefore, instead of erasing the registers with blinking artifacts, they are identified for further analysis and artifact removal procedures.

In our study, we pre-processed the EEG signals prior to analysis. Within the frequency range of 2 Hz to 40 Hz, a digital band-pass filter was applied. Using finite impulse response (FIR) filters with orders of 372 and 1204, we were able to achieve this goal [37, 58]. The selection of FIR filters was based on their desirable characteristics of introducing minimal distortion through linear phase and maintaining a constant group delay. We adjusted the signals by shifting them to the left without losing any signal content to address the group delay introduced by the FIR filters, which totaled 788 samples. A separate notch filter was no longer required as the noise at 50 Hz was effectively reduced by 60 dB by this adjustment [51].

Furthermore, we applied a Butterworth low-pass filter with a cut-off frequency of 10 kHz to the voice signals in order to improve the quality of the filtered data. In this way, the characteristics of the filtered

EEG recordings were further enhanced. To preserve only the signals captured during the imagine/pronounce interval, we segmented the recordings to retain only the relevant parts. By discarding irrelevant portions of the recordings, it was possible to analyze the intervals more accurately.

As a result, EEG signals obtained from specific channels (F3, F4, C3, C4, P3, and P4) corresponding to a particular stimulus instance are concatenated into a vector in accordance with the prescribed order [8]. Three additional labels were added to the vector to provide supplementary information, including mode (imagine), stimulus code, and blinking artifacts. The result was a matrix consisting of rows representing EEG signals recorded during distinct imagined intervals. For the audio recordings, the matrix was constructed using a similar method. A vector containing voice signals from one channel, however, had only two labels added. In the EEG matrix, the first label represents the stimulus, while the second label indicates the row where the synchronized EEG signals are stored.

Table 1: Imagined speech database description

Channels	Samples	Modality	Stimulus	Artifact
F3	1:4096	1. Imagined 2. Pronounced	1. Up	1. No artifact present 2. Artifact present
F4	4097:8192		2. Down	
C3	8193:12288		3. Forward	
C4	12289:16384		4. Backward	
P3	16385:20480		5. Right	
P4	20481:24576		6. Left	

The record of a word is made up of 24,576 samples corresponding to the EEG channels and three additional samples that indicates the modality, stimulus, and the presence of ocular artifacts. In the EEG.mat files, each row corresponds to a recording with the six concatenated channels and three labels whose order corresponds to the label as shown in the Table 1.

By employing appropriate artifact removal techniques, the thesis aims to enhance the quality and reliability of the EEG signals captured during the imagined speech tasks. This ensures that the subsequent

analysis, including DMD, is performed on clean and artifact-free data, leading to more accurate and robust decoding of imagined speech commands [90].

4.4 Extraction of modes using DMD

The DMD process begins with the division of the data into two sequential snapshots. These snapshots are then decomposed using Singular Value Decomposition (SVD) to obtain left and right singular vectors and singular values. In order to reduce the dimensionality of the data, the singular values and their associated vectors are truncated to generate a low-rank approximation. After this, the low-rank approximation is mapped to its time-shifted counterpart in order to estimate a linear operator. We obtain the estimated linear operator using the Moore-Penrose pseudo-inverse. Finally, the estimated linear operator undergoes eigen decomposition to acquire eigenvalues and eigenvectors.

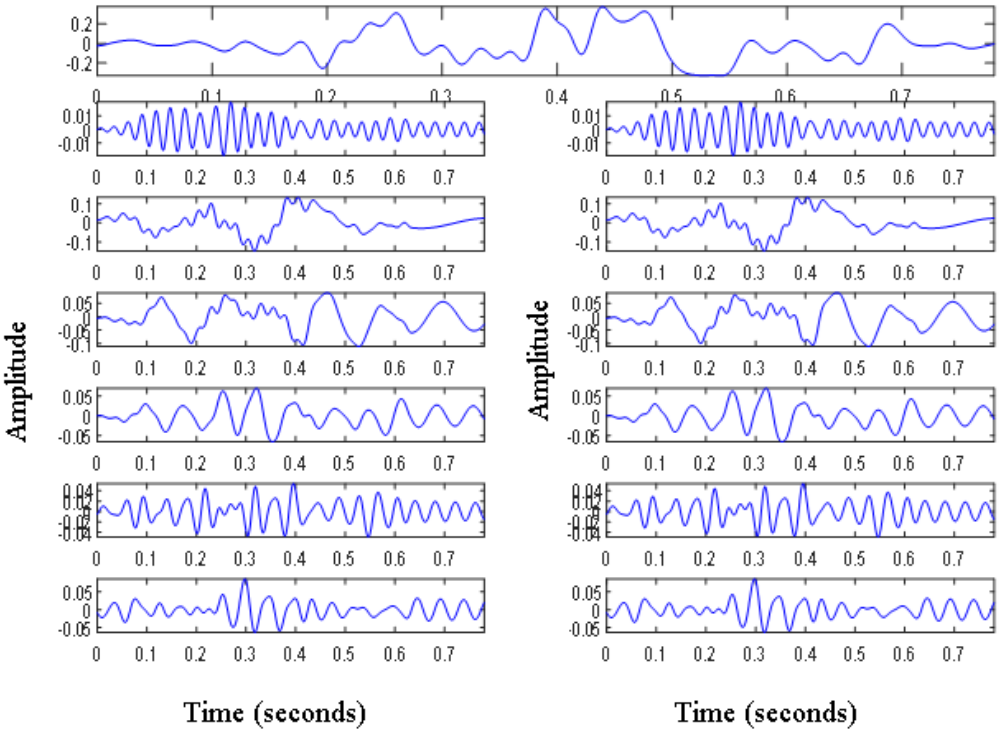


Figure 10: Decomposed modes for Forward class for F-3 channel of subject S10.

Finally, the dynamic modes are computed by applying the eigenvectors to the data. These modes represent the spatial patterns or coherent structures, while the eigenvalues determine their temporal

dynamics. Overall, DMD provides insights into the underlying dynamics of complex systems and can be used for analysis, modelling, and prediction. Figure 10 describes the extracted bands of the imagined command forward for the subject S10, at 4 seconds of sampling frequency of 1024 Hz depicting the original signal and the DMD modes. Each mode is displayed in a separate subplot.

4.5 Feature Extraction

In the context of EEG-Based BCIs for imagined speech, rhythm extracted from DMD as features has provides a powerful method for analysing EEG signals and extracting relevant features that capture the dynamic characteristics of the brain activity associated with imagined speech [53]. The feature extraction process using DMD begins by acquiring EEG data from the user during the imagined speech tasks. The DMD algorithm is applied to the pre-processed EEG data. DMD decomposes the signals into a set of dynamic modes that represent the underlying oscillatory patterns and their temporal dynamics. These dynamic modes capture the different frequency components and their associated time dynamics in the EEG signals.

The DMD algorithm plays a crucial role in identifying the significant dynamic modes that contribute the most to the imagined speech tasks. These dynamic modes represent the key features in the EEG signals that reflect the user's intentions and cognitive processes associated with imagined speech. The TFRs are then extracted from these dynamic modes, serving as the features for the EEG-Based BCI system. TFR provides a detailed representation of time-varying spectral characteristics of EEG signals, enabling more detailed analyses of imagined speech-related brain activity.

DMD features are fed into a classification model, such as a support vector machine (SVM) or deep neural network, to recognize and classify imagined speech tasks. The classification model learns the patterns and relationships between the DMD features and the corresponding speech tasks, enabling accurate decoding and identification of the user's intended speech commands [4,95]. By utilizing DMD as a feature extraction technique in EEG-Based BCIs for imagined speech, the system can effectively capture the dynamic nature of the brain activity involved in speech generation. This approach enhances the accuracy and robustness of the BCI system, enabling more reliable and precise decoding of the user's imagined speech intentions.

4.5.1 Time-Frequency Representation using Hilbert Spectrum Analysis

This section discusses the utilization of Hilbert spectral analysis in EEG-based DMD imagined speech BCI systems to achieve accurate time-frequency representation (TFR). The following steps outline the procedure:

1. TFR Computation: EEG signals recorded during imagined speech tasks are decomposed using the DMD method to extract the constituent frequency components.
2. Hilbert Transform: Each frequency component of the EEG signals undergoes the Hilbert transform, yielding instantaneous amplitude and phase information at each time point. This step captures both temporal and frequency characteristics.
3. TFR Derivation: By combining the amplitude and phase information obtained from the Hilbert transform, TFRs are derived. These TFRs provide a detailed and dynamic characterization of the EEG signals, capturing the temporal evolution of spectral power across different frequency bands.
4. Event-Related Spectral Changes: TFRs are particularly effective in capturing event-related spectral changes, such as the modulation of alpha and beta frequency bands during the generation of imagined speech commands. These spectral patterns offer insights into the neural activity associated with different stages of speech processing.
5. Integration with DMD: The TFRs derived from Hilbert spectral analysis are integrated with the spatial modes obtained from Dynamic Mode Decomposition (DMD), enhancing the representation of imagined speech commands. This combination of spatial information from DMD and temporal-frequency information from Hilbert spectral analysis provides a comprehensive feature representation.

TFRs are useful for discriminating and classifying EEG signals during imagined speech because they provide insight into their temporal and frequency characteristics. It enables effective discrimination and classification of different speech commands, which contributes to the development of EEG-based BCI systems for imagined speech.

A Hilbert spectral analysis is also incorporated into the EEG-based imagined speech BCI system to capture its temporal and frequency dynamics. The derived TFRs provide detailed insights into the spectral changes associated with imagined speech, and their integration with DMD enhances the representation of speech commands. The TFRs in Figure 11 illustrate the effectiveness of this

approach in decoding imagined speech from EEG signals for subject S01. EEG-based imagined speech decoding with this approach shows promising robustness and generalizability.

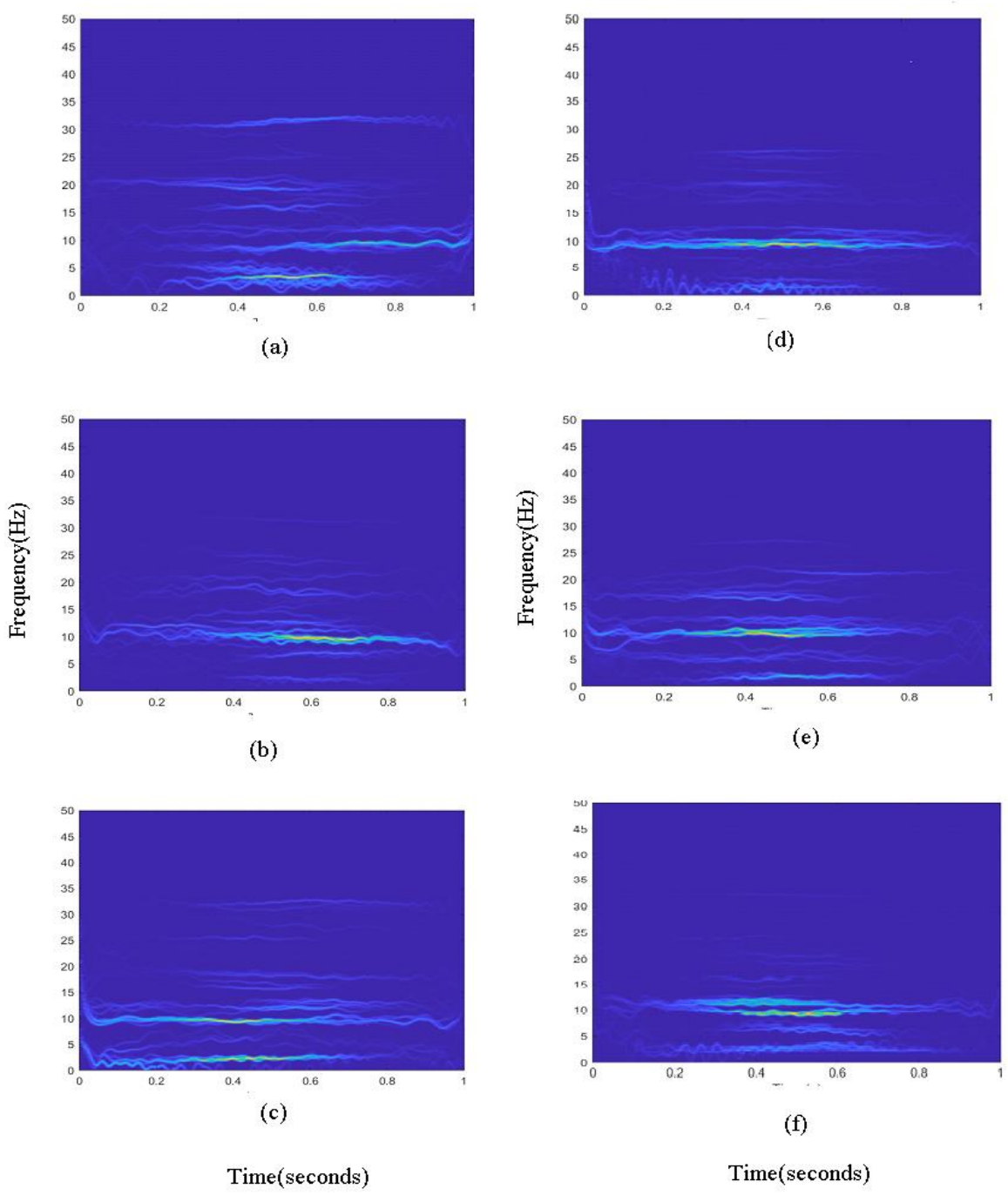


Figure 11: TFRs of subject S01 for six imagined speech commands: (a) Up, (b) Left, (c) Down, (d) Right, (e) Forward, and (f) Backward.

4.5.2 Average TFR Computation across Channels

The following is a step-by-step summary of the process for average TFR computation:

Step 1: Time-Frequency Representation (TFR) computation: TFR averaging aims to enhance the spectral resolution and improve the quality of EEG data by computing time-frequency representations for each individual EEG channel. These representations capture dynamic changes in spectral content over time.

Step 2: Individual TFR computation: Spectrograms or scalograms are computed separately for each EEG channel, representing the time-varying spectral content of the signals.

Step 3: Averaging across channels: The time-frequency representations obtained from multiple channels are averaged together. This averaging process combines spectral information from different channels to create a more robust representation of the underlying brain activity.

Step 4: Enhancing signal-to-noise ratio: TFR averaging aims to reduce the effects of noise and artifacts by enhancing the signal-to-noise ratio. By integrating the individual time-frequency representations, consistent spectral patterns across channels are revealed.

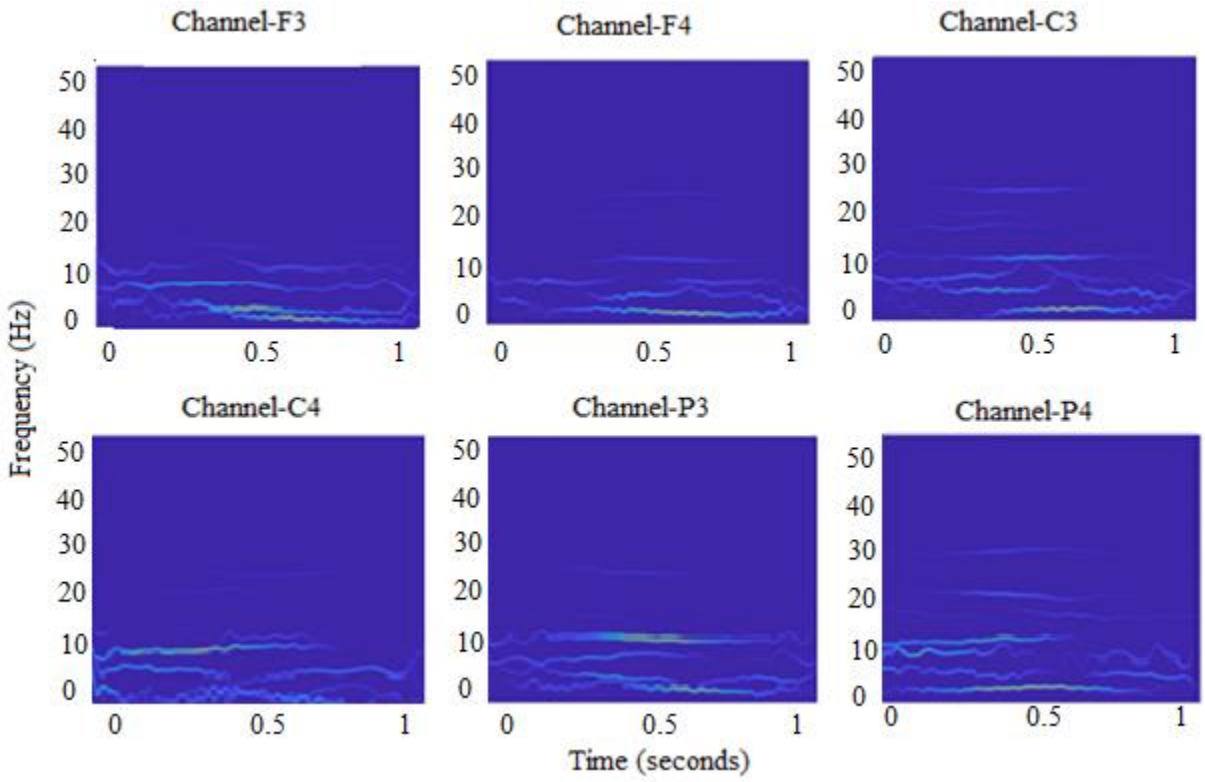
Step 5: Improved spectral resolution: The averaging process enhances the spectral resolution of the EEG signals, providing a clearer representation of the underlying neural activity associated with imagined speech.

Step 6: Extracting relevant neural information: The averaged TFRs serve as valuable insights into the spatio-temporal dynamics of the neural activity related to imagined speech. These TFRs capture the relevant information necessary for the accurate and reliable decoding of speech-related commands.

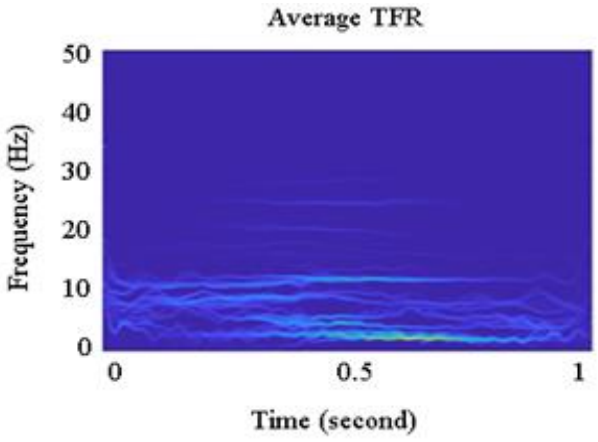
The combination of time-frequency analysis and multichannel EEG signal averaging aims to improve the overall performance of EEG-based BCI systems for imagined speech. This enhancement enhances the usability of the system as a communication assistance tool for individuals with speech impairments.

By following these steps, the process of TFR averaging helps to extract and represent the relevant neural information underlying imagined speech, improving the overall performance and usability of EEG-based BCI systems for communication assistance.

For class Up



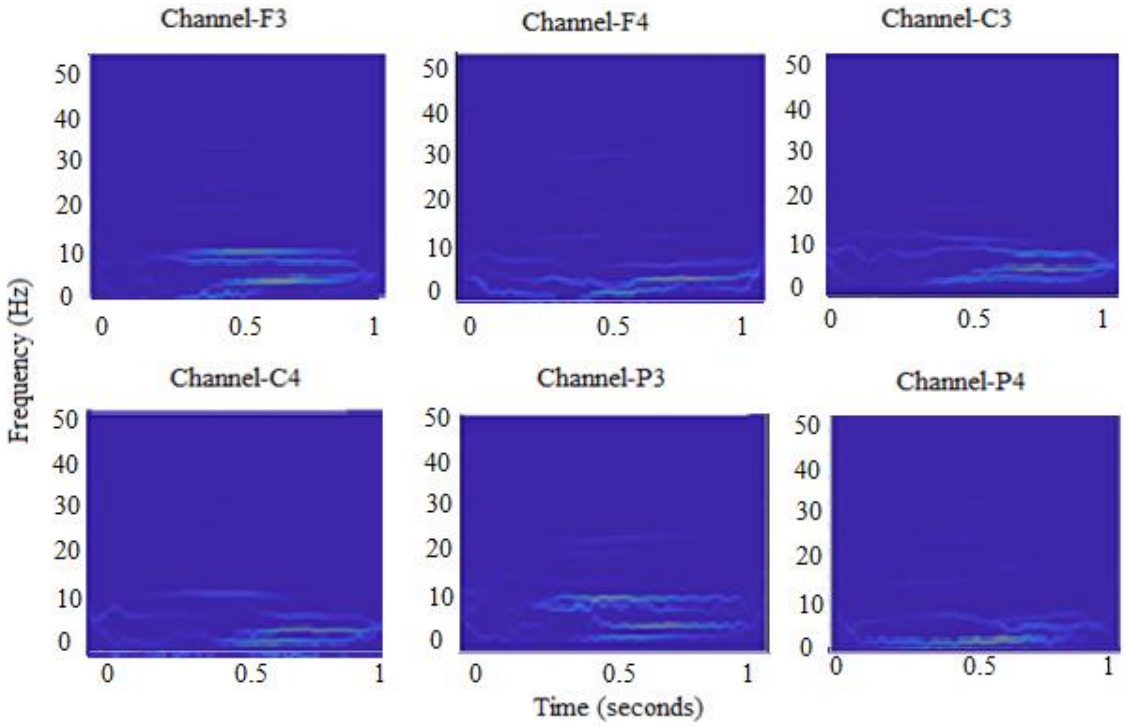
(a)



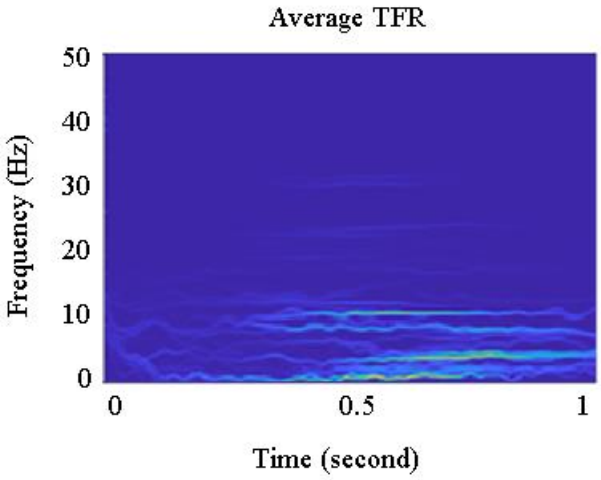
(b)

Figure 12**:** TFRs of subject S01 for Up command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

For class Left



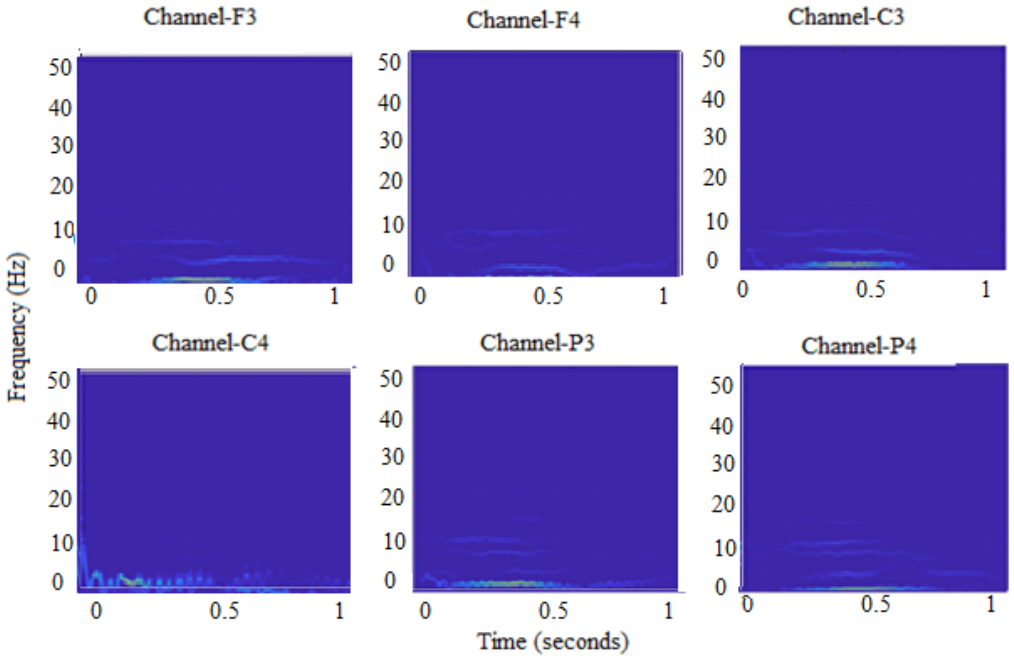
(a)



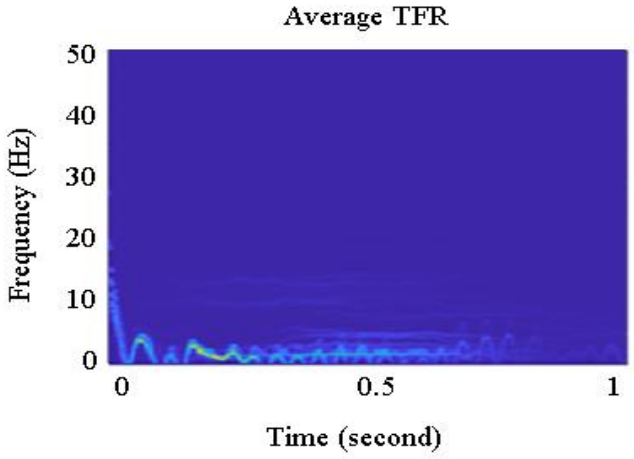
(b)

Figure 13**:** TFRs of subject S01 for Left command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

For class Down



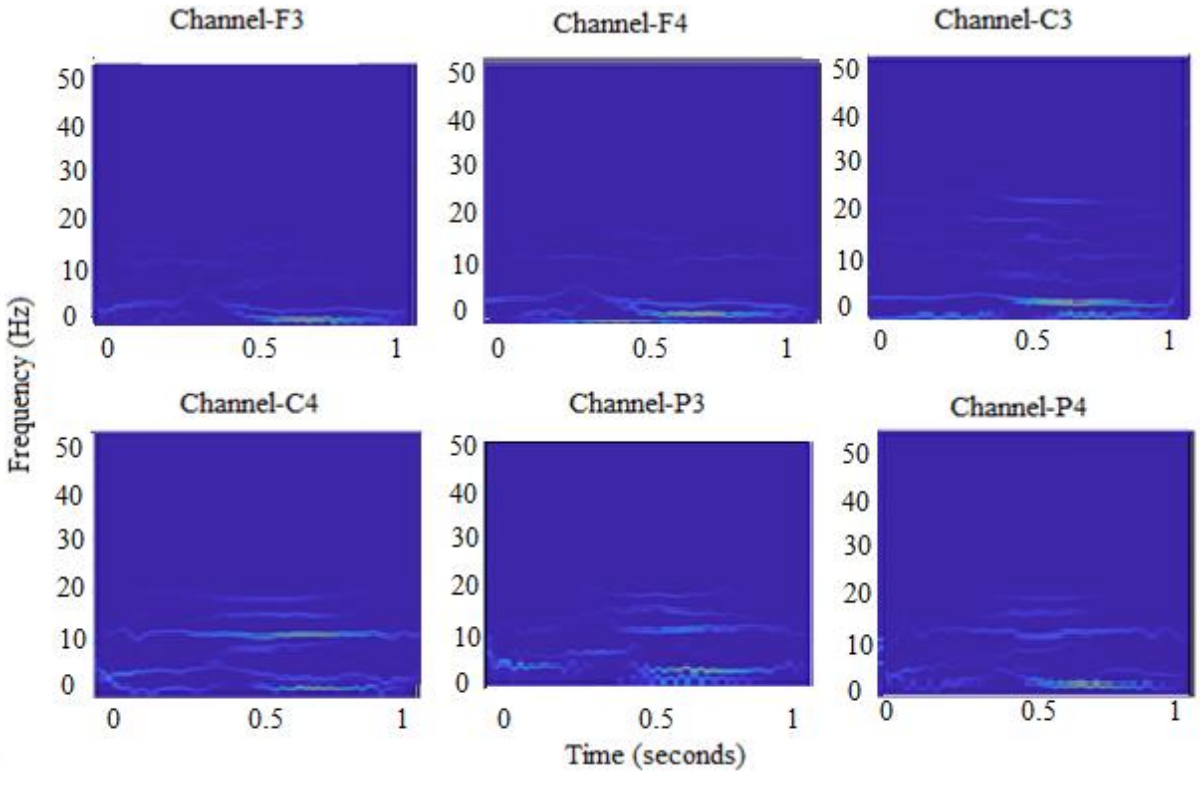
(a)



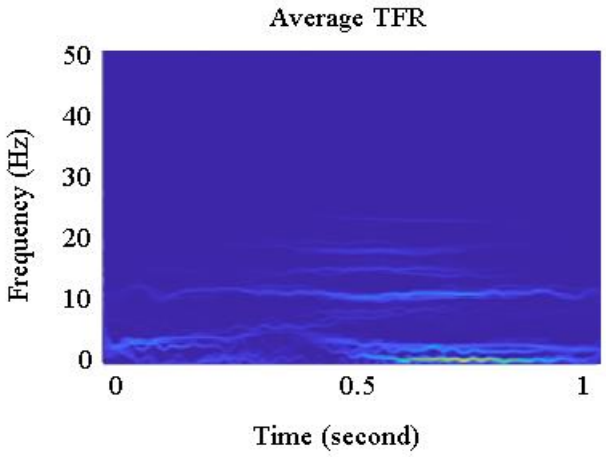
(b)

Figure 14**:** TFRs of subject S01 for Down command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

For class Right



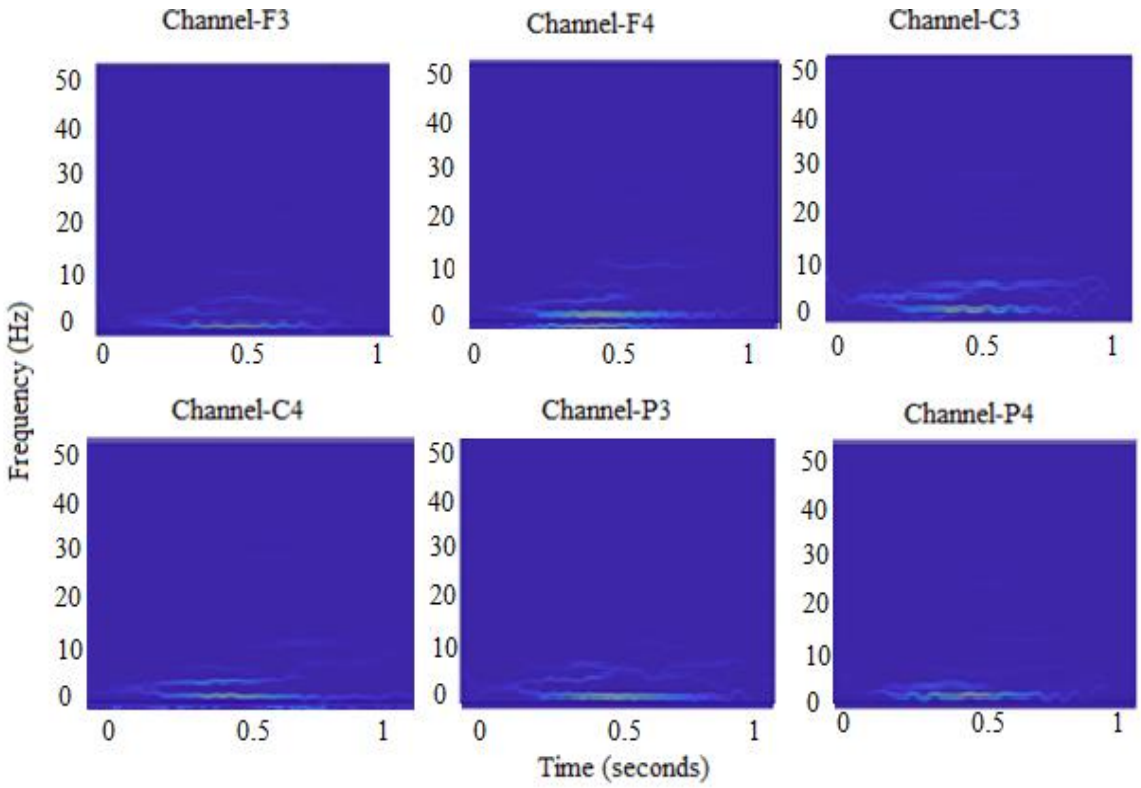
(a)



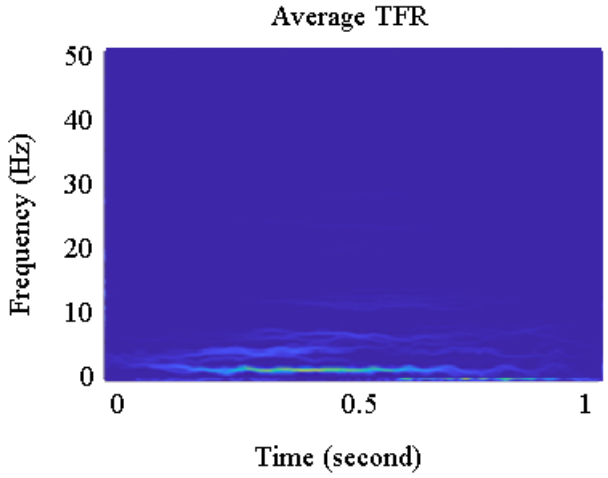
(b)

Figure 15**:** TFRs of subject S01 for Right command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

For class Forward



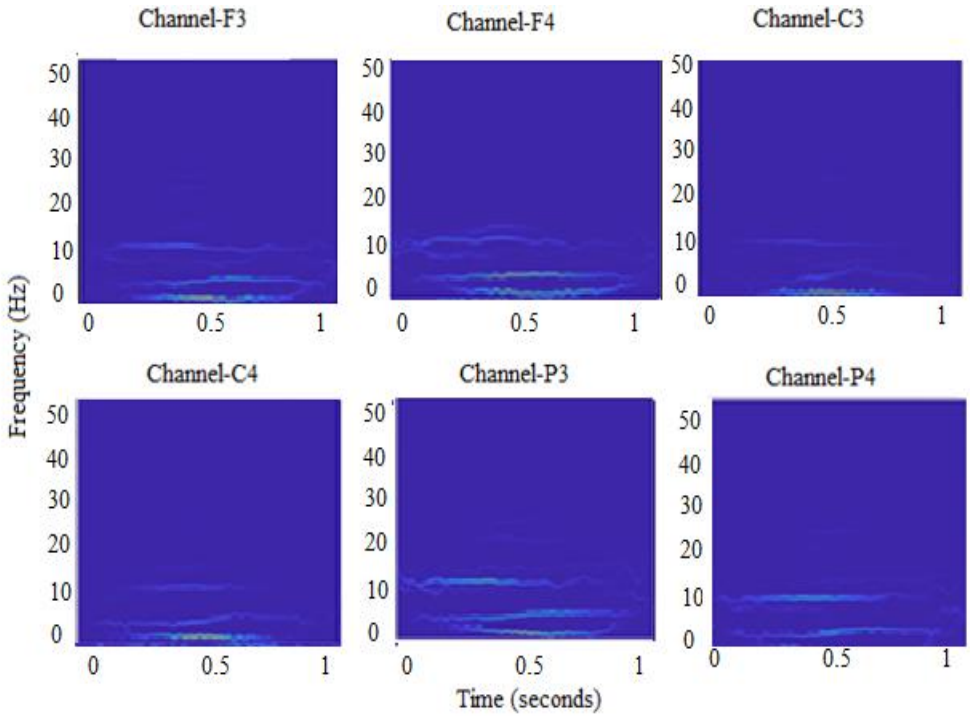
(a)



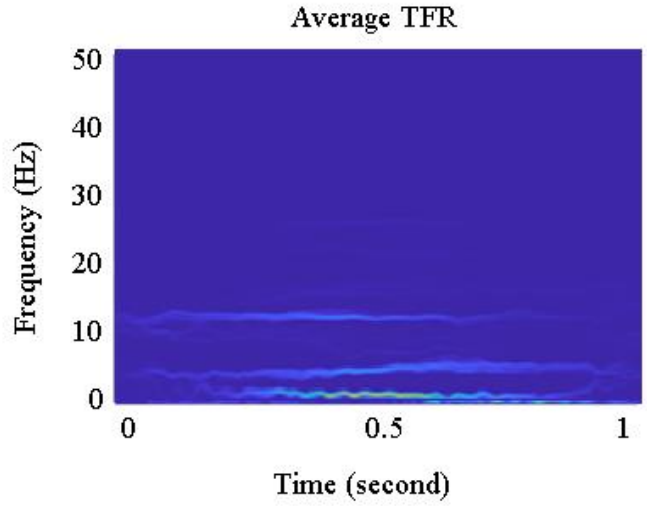
(b)

Figure 16: TFRs of subject S01 for Forward command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

For class Backward



(a)



(b)

Figure 17: TFRs of subject S01 for Backward command from (a) channels F3, F4, C3, C4, P3, P4, and (b) the computed average TFR across the channels.

4.6 Convolution Neural Network (CNN) Modelling

The development of a classification model is a crucial aspect of the EEG-based BCI system designed to decode and classify imagined speech commands from EEG signals. The classification model serves as the bridge between neural activity patterns and meaningful speech commands. To ensure accurate classification, various machine learning and pattern recognition techniques are explored and evaluated. The model utilizes the features extracted from the time-frequency representations of the EEG signals, which capture the dynamic neural patterns associated with imagined speech. These features are used as inputs to the classification model, which undergoes supervised learning to train and optimize its performance. Specifically, convolutional neural networks (CNN) are tested and optimized for the task of classifying imagined speech. The classification model aims to achieve high accuracy, robustness, and real-time performance to enable effective communication through the EEG-based BCI system. By developing a reliable and accurate classification model, the goal is to empower individuals with speech impairments to express themselves using their imagined speech commands, thereby enhancing their quality of life and communication abilities.

4.6.1 Convolutional Neural Network Architecture

In this project, a convolutional neural network (CNN) architecture is introduced and developed for classifying imagined speech commands from EEG signals. CNNs are deep learning models known for analysing complex patterns in multidimensional data, such as images or, in this case, time-frequency representations of EEG signals. It is designed to capture and learn hierarchical features from the input data, allowing it to distinguish subtle differences and variations in neural activity associated with different imagined speech commands. There are several layers in this architecture, including convolutional layers, pooling layers, and fully connected layers. As the convolutional layers process the input data, they extract local features and capture spatial information across time-frequency representations. As a result of the pooling layers, the extracted features are reduced in dimensionality while the most significant information is preserved. The final classification is performed by the fully connected layers, which combine the extracted features.

To train the CNN architecture, a large dataset of labelled EEG signals is utilized, and the model parameters are optimized through backpropagation. During backpropagation, the model learns from the

training data and adjusts its internal weights and biases to minimize the classification error. The proposed CNN architecture aims to achieve high accuracy, robustness, and generalizability in decoding imagined speech commands from EEG signals. Its effectiveness lies in enabling effective communication through the EEG-based BCI system by accurately classifying the user's intended speech commands.

4.6.2 CNN Model Training and Optimization

The training and optimization phase of the proposed system is a crucial step in achieving accurate and reliable classification of imagined speech commands from EEG signals. This process involves training the CNN architecture using labelled EEG data and iteratively adjusting the model's parameters to maximize classification accuracy. Backpropagation, an optimization algorithm, is employed to compute the gradients of the loss function with respect to the model's weights and biases. These gradients are then used to update the model's parameters using gradient descent or a variant like the Adam optimizer.

Training datasets are usually divided into training and validation subsets to prevent overfitting. Training subsets are used to update parameters, while validation subsets are used to monitor the model's performance. The optimization process requires fine-tuning the hyperparameters of the CNN architecture, such as the learning rate, batch size, and regularization techniques. It is important to find the right balance between model complexity and generalization ability.

Training and optimization require iterative adjustments, parameter tuning, model evaluation, and validation. Decoding imagined speech commands from EEG signals requires a lot of time and effort, but is crucial for high classification accuracy, robustness, and reliability. Using the proposed system, imagined speech commands can be accurately classified while avoiding overfitting and underfitting the CNN architecture. As a result of successfully completing this process, a BCI based on EEG is able to decode imagined speech commands effectively.

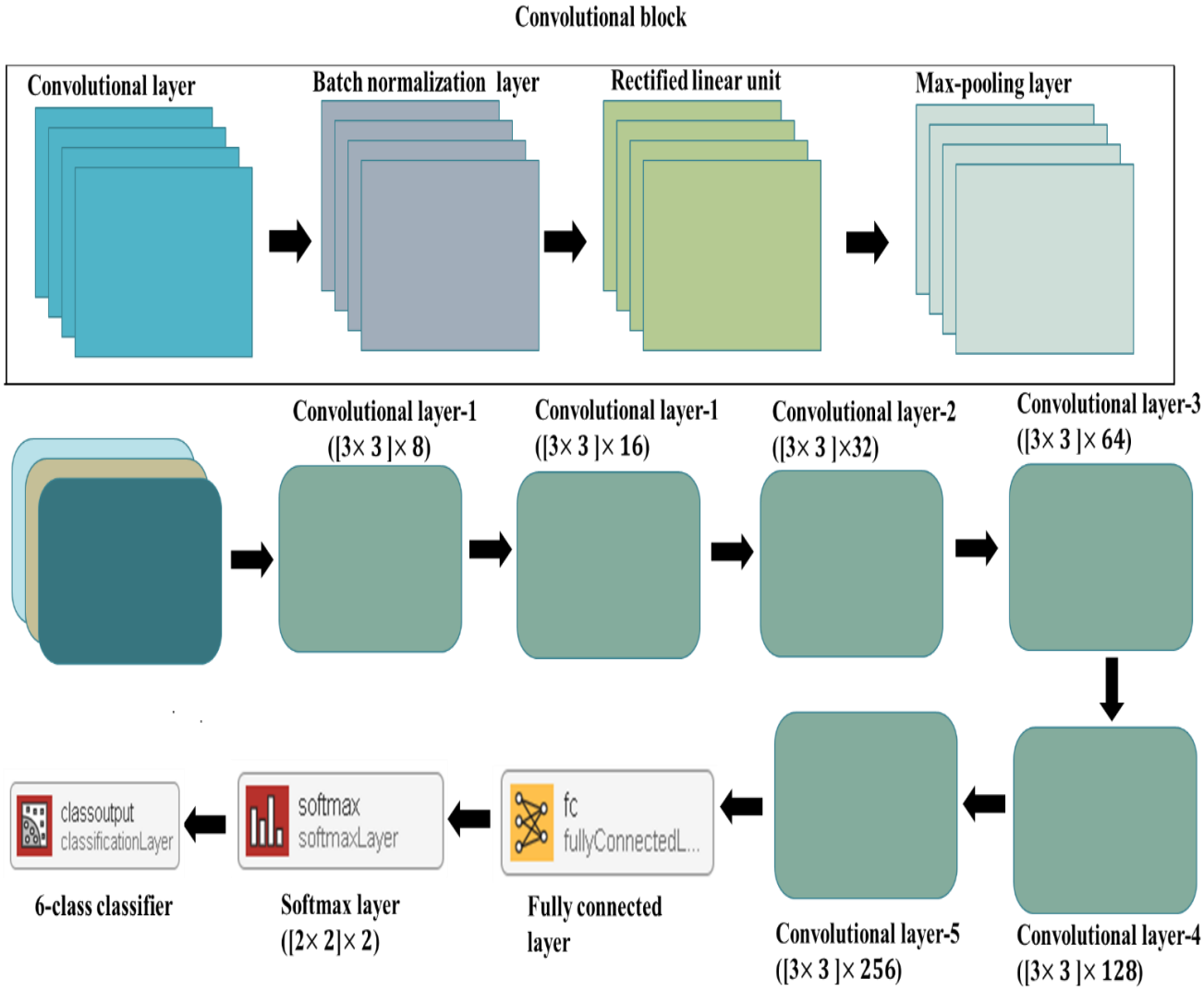


Figure 18: DMD-CNN based classification framework for decoding imagined speech commands.

In our study, the DMD-CNN approach outperformed conventional EEG-based BCI classification methods, with an average classification accuracy exceeding 70% across all subjects for imagined command stimuli. DMD-CNN demonstrated notable advantages over conventional methods in terms of training time reduction, making it highly suitable for real-time applications. An image input layer, a convolutional layer, a batch normalization layer, a rectified linear unit, a max-pooling layer, a fully-connected layer, and a soft-max layer are all crucial components of the CNN architecture used in this approach.

As the first component of the neural network, the image input layer defines the dimensions of the input images. In the convolutional layer, a user-defined filter size is applied to the input image, generating

feature maps that are used as inputs for subsequent layers. Normalizing the data samples through the batch normalization layer enhances learning efficiency and facilitates faster learning. A rectified linear unit removes redundancy while preserving essential information by applying a threshold operation to the input data. In the max-pooling layer, the maximum value within each window is considered in order to reduce the size of the feature maps.

From the max-pooling layer, the fully-connected layer establishes connections between neurons, with the output corresponding to the number of classification classes. Using the soft-max layer, outliers are reduced and classification is enhanced by applying the soft-max function to the input data. Depending on the input size, the number of layers in a CNN can be adjusted, and not all layers need to be included. Although deeper networks can achieve better learning, they require more computational time. This classifier eliminates the need for preprocessing and feature extraction, which is one of its notable advantages.

Using a minimal number of layers and efficiently tuning the network parameters, we aimed to optimize the classification performance. The block diagram in Figure 16 shows the implementation of CNN classifiers for our investigation. The input images had dimensions of [656 × 875 × 3] for the image input layer. A total of six convolutional layers were employed, each containing eight filters of size 8, 16, 32, 64, 128, and 256, along with the same padding. Each convolutional layer is followed by a batch normalization, a rectified linear unit, and a max-pooling layer. Using a stride of 2 and a filter size of 2 were used in the max-pooling layer.

Furthermore, the CNN classifier was implemented using a fully-connected layer, a soft-max layer, and a classification layer. We used a learning rate of 0.001 along with a maximum number of epochs and batch sizes of 40 and 256. Our training set of TFR images consisted of 80%, while our test set consisted of 20%. A classification accuracy of 71.99% was achieved for DMD-TFR images using CNN classifiers. Previous studies have proposed various methods for sleep stage classification.

Finally, the CNN classifier classified six imagined speech commands based on an averaged DMD-TFR image calculated from different TFR trials for an EEG epoch. DMD-TFR images were classified with 70% accuracy. It took CNN 0.0192 seconds to classify one TFR image. In the experiments, an Intel Core i7 processor with a speed of 3.60 GHz and 32 GB of memory was used in a Dell 5820 Precision Tower.

The project was implemented with MATLAB 2023a and Windows 8. Our proposed DMD approach is compared with existing methods in Table 3 to determine classification accuracy.

4.6.3 Imagined Speech Classification

The classification stage of the study focuses on accurately categorizing the decoded imagined speech commands using the extracted features. To achieve this, a CNN architecture is employed as the classification model. This architecture is specifically designed to effectively learn and recognize patterns in the fused features obtained from DMD and time-frequency analysis. These features capture important information related to the imagined speech commands, including temporal dynamics and frequency characteristics.

During the training and optimization phase, the CNN architecture undergoes a process where labeled training data, consisting of fused features and their corresponding speech commands, is used. The model adjusts its internal parameters through backpropagation to effectively map input features to the appropriate speech command categories. To enhance classification performance, optimization techniques like regularization methods and hyperparameter tuning are applied. These techniques mitigate overfitting and improve the model's ability to generalize to unseen data.

Following training and optimization, the CNN model is evaluated using a separate dataset to measure its accuracy in correctly classifying imagined speech commands. The classification results are then analyzed and compared with existing methods to validate the effectiveness of the proposed approach. The CNN architecture's superiority, coupled with the distinctive features extracted through DMD and time-frequency analysis, is evident through higher accuracy and improved decoding performance of imagined speech commands.

The CNN architecture's specific configuration, including the number of hidden layers, pooling layers, and activation dimensions, can be found in Table 5. For training, the adaptive moment estimation (Adam) optimizer is utilized, with customized settings for gradient decay factors, learning rate, and the number of epochs. During training, a mini-batch size of 32 observations is employed, and a total of 1200 iterations are conducted. To prevent consecutive epochs from discarding the same data, the training and validation data are shuffled before each epoch. L2 regularization, with a weight decay factor of 0.0001, is incorporated to address overfitting concerns.

The network's performance is then assessed by evaluating the accuracy and classification time of the test data. Throughout training, the validation data, representing 20% of the total data, is utilized to monitor the network's performance after every 30 iterations. In this particular network configuration, no gradient clipping is implemented. The accuracy and classification time of the test data serve as essential indicators for evaluating the network's performance.

Table 2: CNN model Optimized Hyperparameters

Layer type	Activations	`	Learnable parameters	Remarks
Image input	$875 \times 656 \times 3$	normalization: zerocenter	…	*input imag* $\times 656 \times 3$
Convolutional block 1				
2D-convolutional	$437 \times 327 \times 8$	filter size: [3,3] no. of filters : 8 Stride:[1]	Weight : $3 \times 3 \times 6$ Bias : $1 \times 1 \times 8$	*no. of conv* $\times 3 \times 3$
Batch normalization	$437 \times 327 \times 8$	ε: 0.00001	Offset : $1 \times 1 \times 8$ Scale : $1 \times 1 \times 8$	*channels*: 8
Relu	$437 \times 327 \times 8$	…	…	
2D-max pooling	$219 \times 164 \times 8$	filter size: $[2 \times 2]$ Stride:[1]	…	
Convolutional block 2				
2D-convolutional	$110 \times 82 \times 16$	filter size: [3,3] no. of filters : 16 Stride:[1]	Weight : $3 \times 3 \times 8$ Bias : $1 \times 1 \times 16$	*no. of conv* $\times 3 \times 8$
Batch normalization	$110 \times 82 \times 16$	ε: 0.00001	Offset : $3 \times 3 \times 16$ Scale : $1 \times 1 \times 16$	*channels*: 1
Relu	$110 \times 82 \times 16$	…	…	
2D-max pooling	$55 \times 41 \times 16$	filter size: $[2 \times 2]$ Stride:[1]	…	
Convolutional block 3				
2D-convolutional	$28 \times 21 \times 32$	filter size: [3,3] no. of filters : 32 Stride:[2]	Weight : $3 \times 3 \times 16$ Bias: $1 \times 1 \times 32$	*no. of conv* $\times 3 \times 16$
Batch normalization	$28 \times 21 \times 32$	ε: 0.00001	Offset : $1 \times 1 \times 32$ Scale: $1 \times 1 \times 32$	*channels*: 3
Relu	$28 \times 21 \times 32$	…	…	
2D-max pooling	$14 \times 11 \times 32$	filter size: $[2 \times 2]$ Stride:[2]	…	
Convolutional block 4				

2D-convolutional	$7 \times 6 \times 64$	filter size: [3,3] no. of filters : 64 Stride :[2]	Weight : $3 \times 3 \times 32$ Bias : $1 \times 1 \times 64$	$no.of\ conv$ $\times 3 \times 32$
Batch normalization	$7 \times 6 \times 64$	ε: 0.00001	Offset : $1 \times 1 \times 64$ Scale: $1 \times 1 \times 64$	$channels$: 6
Relu	$7 \times 6 \times 64$	…	…	
2D-max pooling	$4 \times 3 \times 64$	filter size: $[2 \times 2]$ Stride:[2]	…	
Convolutional block 5				
2D-convolutional	$2 \times 2 \times 128$	filter size: [3,3] no. of filters : 128 Stride:[2]	Weight : $3 \times 3 \times 64$ Bias : $1 \times 1 \times 128$	$no.of\ conv$ $\times 3 \times 64$
Batch normalization	$2 \times 2 \times 128$	ε: 0.00001	Offset : $1 \times 1 \times 128$ Scale : $1 \times 1 \times 128$	$channels$: 1
Relu	$2 \times 2 \times 128$	…	…	
2D-max pooling	$1 \times 1 \times 128$	filter size: $[2 \times 2]$ Stride:[2]	…	
Convolutional block 6				
2D-convolutional	$1 \times 1 \times 265$	filter size: [3,3] no. of filters : 256 Stride:[1]	Weight : $3 \times 3 \times 128$ Bias : $1 \times 1 \times 265$	$no.of\ conv$ $\times 3 \times 128$
Batch normalization	$1 \times 1 \times 265$	ε: 0.00001	Offset : 6×265 Scale : 6×1	$channels$: 2
Relu	$1 \times 1 \times 265$	…	…	
Fully connected	6×1	output size: 6	Weight : $1 \times 1 \times 265$ Bias : $1 \times 1 \times 265$	Fully connected layers: 6
			$875 \times 656 \times 3$	
SoftMax	6×1	$875 \times 656 \times 3$		
Classification output	…	loss function : cross entropyex		Output size: 6

Chapter 5

Results and Discussion

In this section, we present a comprehensive evaluation of the proposed DMD-TFA-CNN method and discuss the results of the experiments based on the collected EEG data. The evaluation aims to provide a thorough analysis of the system's performance and effectiveness. To evaluate the system's ability to decode and classify imagined speech commands from EEG signals, key metrics, such as classification accuracy, are used. The results are organized and presented in tables, along with statistical analysis, offering a comprehensive overview of the system's performance across different subjects and experimental conditions.

The analysis goes beyond the surface-level results and delves deeper into the obtained outcomes. It identifies potential patterns, trends, and variations in the classification performance, allowing for a more detailed understanding of the system's behaviour. Moreover, the results are compared and contrasted with existing methods such as EWT and EMD methods. This comparative analysis sheds light on the strengths and limitations of the proposed system, highlighting its potential for real-world applications. Overall, the Results and Analysis section provides valuable insights for further improvements and future research directions in the field of imagined speech EEG-based BCIs. It serves as a foundation for building upon the proposed method and advancing the state-of-the-art in this domain.

5.1 Mean Square Error (MSE) Performance Analysis

The performance evaluation of the TFR from DMD modes has been quantitatively measured by computing the mean square error (MSE). The MSE is expressed as follows:

$$\mathrm{MSE} = \frac{1}{PQ}\sum_{\mathrm{f}=1}^{P}\sum_{\mathrm{t}=1}^{Q}(\mathrm{TF}_1(f,t) - \mathrm{TF}\,(f,t))^2 \qquad (7.1)$$

The total number of frequency points and time instants in the TF plane are denoted by P and Q, respectively. In this study, the synthetic multi-component signals are used to validate the effectiveness

of the proposed DMD method in decomposing signals into modes in the TF plane. The TF1 and TF represent the expected and obtained TF representations of these signals [5].

To assess the performance of the proposed DMD method, the mean squared error (MSE) is calculated for each case and compared with the MSE values obtained from the EMD and EWT methods [4,7]. Figure 19 illustrates the comparison of MSE values among the three methods. The synthetic multi-component AM signal [5] is considered for this analysis:

$$x(t) = \sum_{i=1}^{3} 0.2\left(1 + 0.2\mathrm{Sin}(4\pi t)\mathrm{Cos}(2\pi f_i t)\right) \qquad (7.2)$$

Where three different sinusoidal components are considered with: Frequency: f1 = 50 Hz, f2= 150 Hz, f3= 200 Hz

The signal components in the frequency domain are separated by certain intervals. Throughout the experiments, the signal duration remains fixed at 1 second, and the sampling frequency is set to 1kHz [5]. Table 3 presents the mean square error (MSE) values obtained from the comparative analysis of different methods, including DMD, EMD, and EWT. These MSE values serve as indicators of accuracy and performance when applied to the time-frequency analysis of non-stationary signals [4].

The results indicate that DMD achieves the lowest MSE value among the methods considered, suggesting its superior performance in capturing the underlying dynamics and extracting relevant features from EEG-based data. This signifies that DMD effectively preserves essential information while minimizing reconstruction errors, thereby enhancing time-frequency resolution. In contrast, both EMD and EWT exhibit higher MSE values in comparison to DMD, indicating relatively higher reconstruction errors. While EMD and EWT possess their own advantages and applications, the elevated MSE values imply that they may not be as effective in capturing intricate details and patterns present in the EEG signals utilized in this study [36].The comparative analysis of MSE values underscores the potential of DMD as a promising technique in EEG-based BCI applications. Its ability to accurately represent and extract meaningful features from EEG signals can contribute to improved performance and reliability in decoding imagined speech. These findings emphasize the significance of selecting appropriate feature extraction methods in EEG-based BCI systems and further highlight the potential of DMD in advancing the field of imagined speech BCI research.

Table 3: Comparison of the MSE values of multi-component amplitude modulated signal for different cases using DMD, EMD and EWT methods.

MSE	DMD	EMD	EWT
First signal component	3.2383e-04	0.0379	0.0654
Second signal component	1.6247e-04	0.0469	0.0921
Third signal component	2.9383e-04	0.0535	0.0922

From Table 3, comparing the MSE values of multi-component amplitude modulated signals using DMD, EMD, and EWT methods, we can draw the following conclusions:

1. DMD Method: The MSE values for the first, second, and third signal components using the DMD method are 3.2383e-04, 1.6247e-04, and 2.9383e-04, respectively.

2. EMD Method: The MSE values for the first, second, and third signal components using the EMD method are 0.0379, 0.0469, and 0.0535, respectively.

3. EWT Method: The MSE values for the first, second, and third signal components using the EWT method are 0.0654, 0.0921, and 0.0922, respectively.

Based on the MSE values, the DMD method achieves the lowest MSE values among the three methods for all three signal components. This indicates that DMD provides better reconstruction accuracy for the multi-component amplitude modulated signals compared to EMD and EWT methods. The EMD method has higher MSE values compared to DMD for all three signal components, suggesting relatively less accurate reconstruction. The EWT method shows the highest MSE values among the three methods for all three signal components, indicating the least accurate reconstruction. Therefore, the DMD method outperforms both EMD and EWT methods in terms of reconstruction accuracy for the multi-component amplitude modulated signals analysed in this study. Furthermore, the TF representation in Fig. 20 above derived from the multi-component AM signal, for three different frequency components, using EMD,

EWT, and DMD methods and a comparative analysis of the methods shows that, the proposed DMD clearly represents all the components in the TF plane.

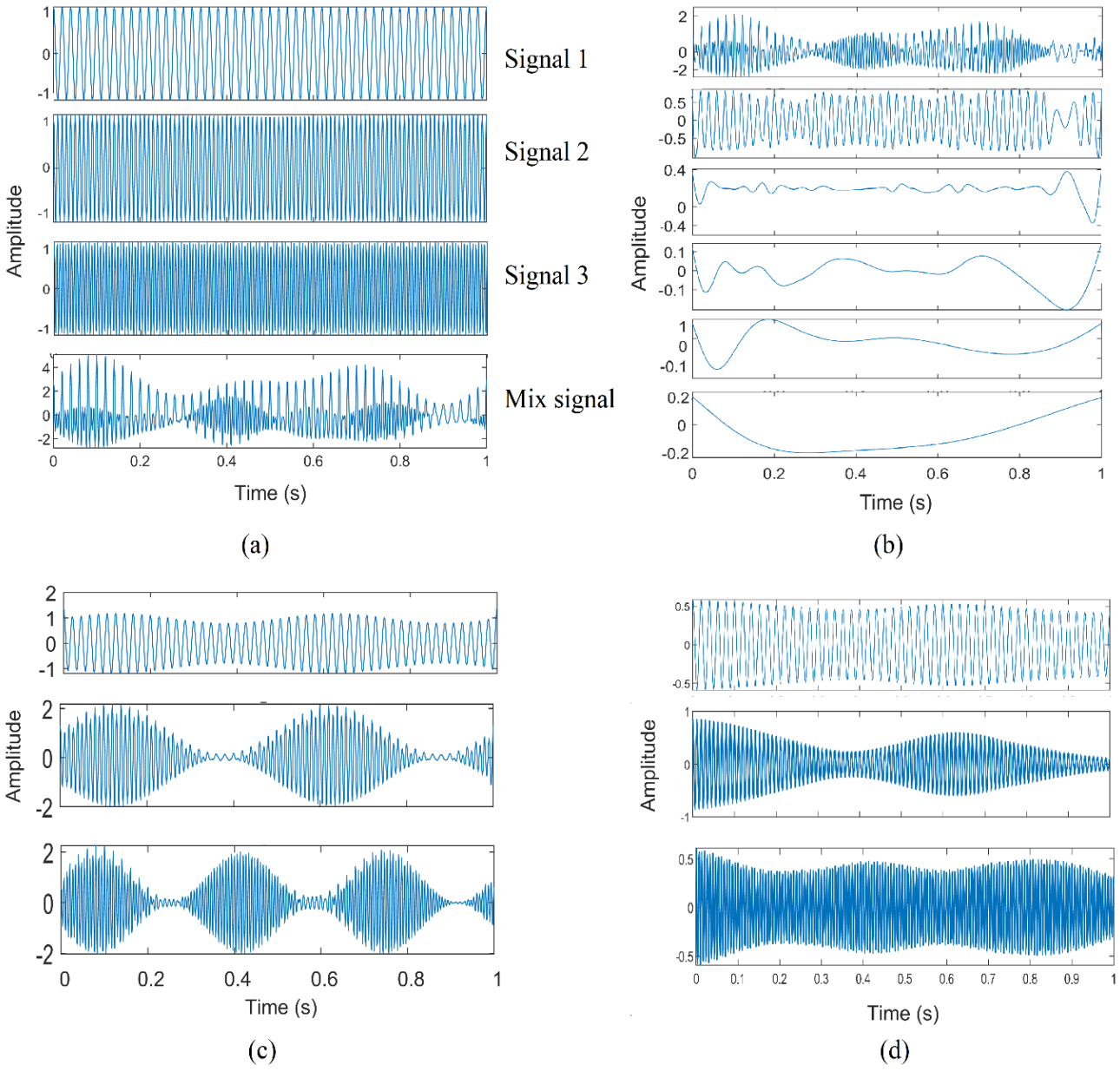


Figure 19: Plots of (a) mono-components and multi-component AM input signal, and (b-d) extracted modes using EMD method, EWT method, and DMD method, respectively.

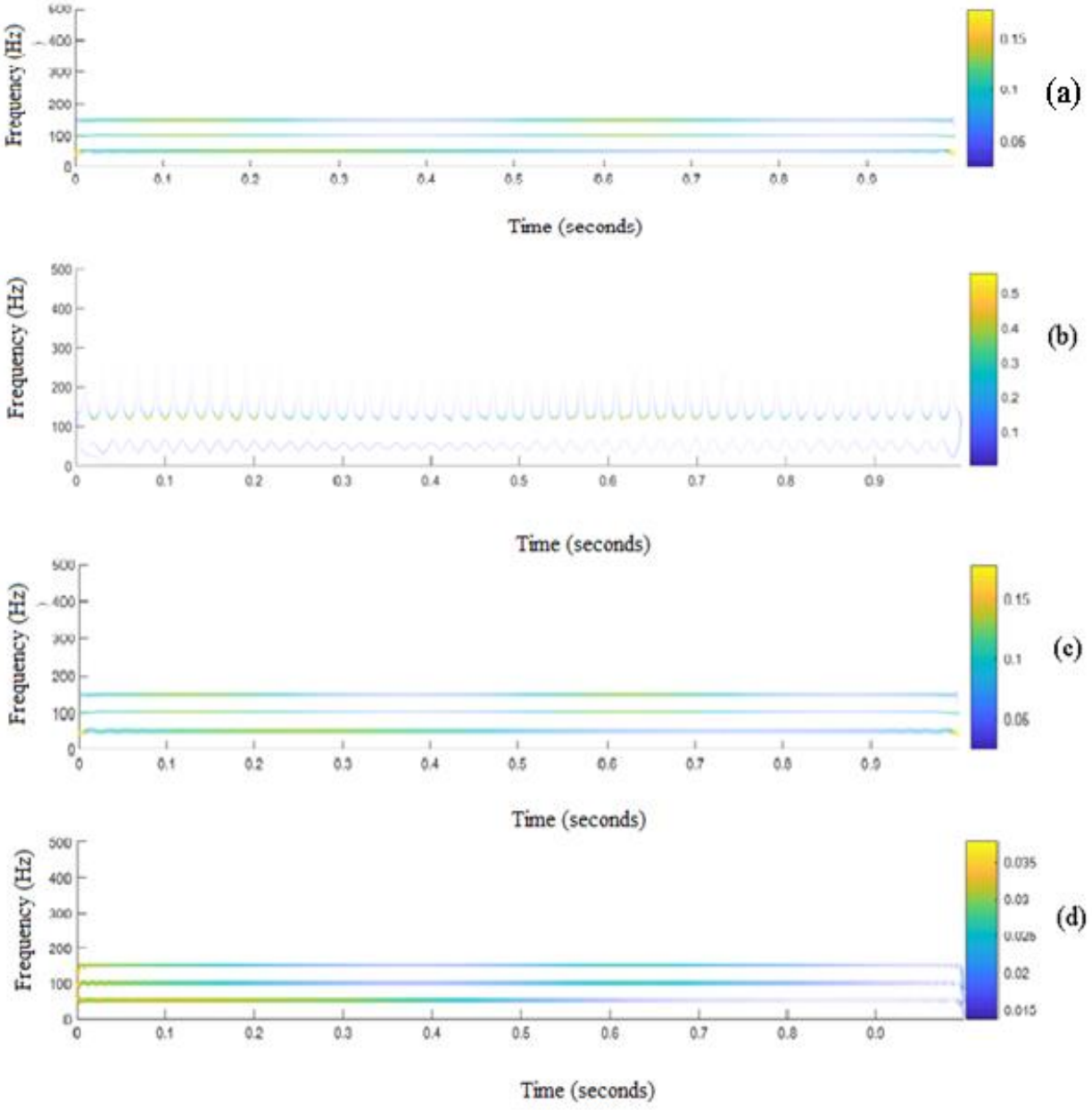


Figure 20: Plots of (a) TFRs of input nonstationary AM signal, (b) EMD based TFR (c) EWT based TFR, and (d) DMD based TFR.

5.2 DMD Performance Analysis for imagined speech EEG signals

In this investigation, we examine the efficacy of the DMD technique proposed for decomposing signals into modes within the time-frequency (TF) domain. To validate its performance, we utilize synthetic multi-component signals referred to as TF1 and TF, where TF1 represents the expected TF representation and TF represents the obtained TF representation [5]. By employing synthetic multi-component signals, we aim to evaluate the effectiveness of the proposed DMD method. The evaluation involves estimating the mean squared error (MSE) and comparing it with the performance of the EMD and EWT methods [4,7]. The comparison of MSE results for the DMD method, EMD, and EWT is illustrated in Figure 19. Specifically, we focus on analyzing the AM (amplitude-modulated) signal as a specific type of synthetic multi-component signal [5]. The AM signal is represented mathematically for this purpose.

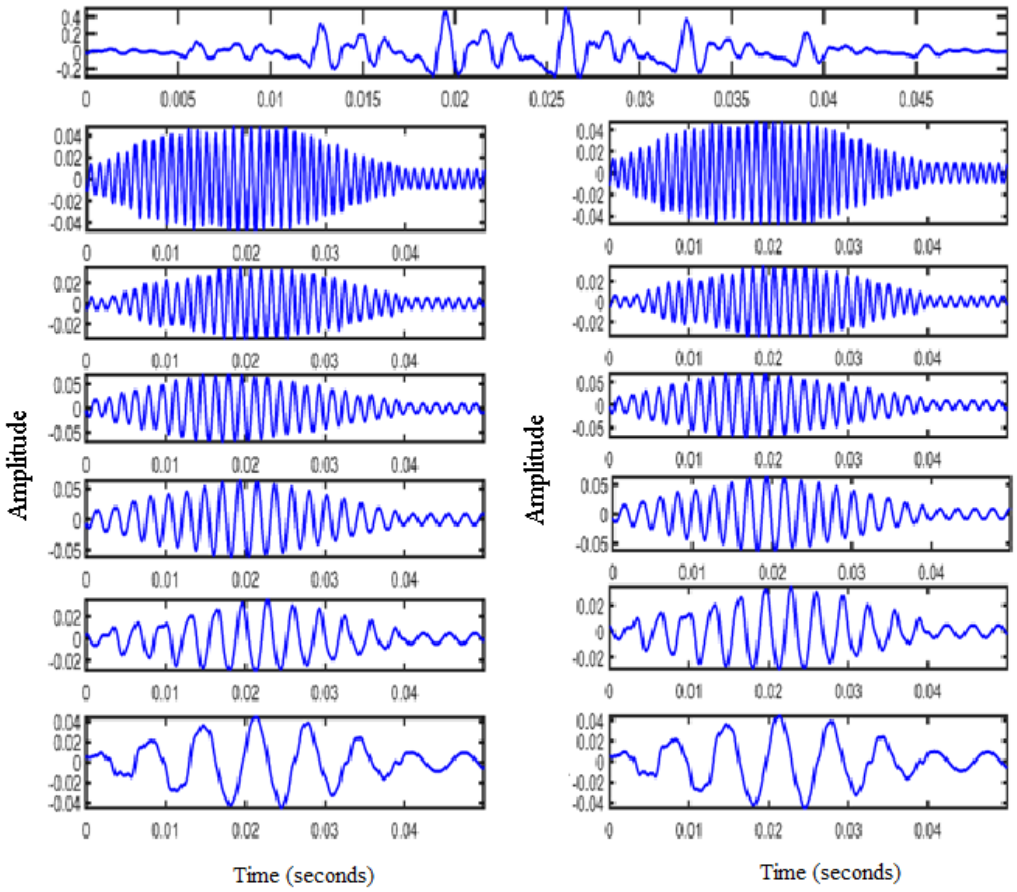


Figure 21: Plots of decomposed modes of an imagined speech signal using DMD method.

For evaluating the performance of DMD on real EEG signals we have employed real EEG data from an open access database for six imagined commands across all channels [80]. In Table 5, the extracted average TFRs, for six imagined speech classes across fifteen subjects is represented. Upon analysing the data, it is observed that the number of TFRs varies across imagined speech command classes and subjects . For example, subject S04 has the highest number of TFRs for the Backward command (189), while subject S06 has the lowest number (48). Similarly, subject S010 has the highest number of TFRs for the Right command (172), while subject S06 has the lowest number (64). Overall, the extracted average TFRs range from 100 to 1977 across the six imagined speech command classes. These numbers reflect the variability in the data obtained from different subjects and highlight the importance of subject-specific characteristics in EEG-based imagined speech BCI systems.

Table 5: Extracted average TFRs across 15 subjects for 6 imagined speech command.

Subjects	Imagined speech command classes					
	Backward	Down	Forward	Left	Right	Up
S01	84	84	100	60	88	84
S02	172	144	148	144	136	156
S03	132	144	102	158	132	100
S04	189	156	160	117	136	164
S05	112	88	116	104	116	100
S06	48	48	68	60	64	68
S07	136	156	152	152	156	160
S08	164	156	148	144	148	140
S09	156	160	148	156	148	136
S010	128	148	124	128	172	140
S011	148	152	104	120	136	112
S012	112	120	120	100	84	132
S013	156	148	171	111	173	136
S014	116	128	100	108	124	112
S015	124	116	128	116	116	124
	1977	**1948**	**1889**	**1778**	**1929**	**1848**

Therefore, when performing a data analysis, it is observed that the number of TFRs varies across subjects and imagined speech command classes. For example, subject S04 has the highest number of TFRs for the Backward command (189), while subject S06 has the lowest number (48). Similarly, subject S010 has the highest number of TFRs for the Right command (172), while subject S06 has the lowest number (64). Overall, the extracted average TFRs range from 100 to 1977 across the six imagined speech command classes. These numbers reflect the variability in the data obtained from different subjects and highlight the importance of subject-specific characteristics in EEG-based imagined speech BCI systems. This information provides insights into the amount of data available for each subject and each imagined speech command class. It aids in understanding the distribution and variability of TFRs, which are crucial for subsequent analysis, feature extraction, and classification tasks in imagined speech BCIs based on EEG.

As shown in Table 5 provides the number of epochs obtained from the imagined speech EEG signals for six different classes, along with the average number of extracted TFRs for each class. The imagined speech classes include Up, Down, Left, Right, Forward, and Backward. For each class, a total of 4000 epochs were collected from the EEG signals. These epochs represent specific instances of the imagined speech commands and serve as the basis for analysis and classification. Upon extracting the TFRs from the EEG signals, the average number of TFRs obtained varied for each class. The class Up yielded an average of 1848 TFRs, Down had an average of 1948 TFRs, Left had an average of 1778 TFRs, Right had an average of 1929 TFRs, Forward had an average of 1889 TFRs, and Backward had an average of 1977 TFRs.

These numbers provide insights into the amount of data available for each imagined speech class and the corresponding TFRs extracted from the EEG signals. The variation in the average number of TFRs suggests potential differences in the characteristics and complexity of the imagined speech commands. This information is valuable for understanding the dataset used in the study and serves as a foundation for further analysis and classification tasks. It highlights the importance of having a substantial number of epochs and extracted TFRs to ensure reliable and accurate results in EEG-based imagined speech BCI systems.

Table 5: Number of epochs obtained from the imagined speech EEG signals corresponding six different classes.

Imagined speech command	Number of epochs	Average extracted TFRs
Up	4000	1848
Down	4000	1948
Left	4000	1778
Right	4000	1929
Forward	4000	1889
Backward	4000	1977

Furthermore, the evaluated performance of DMD-TFA, has been conducted for 6-classes of imagined speech EEG signal, which demonstrated in Fig. 22

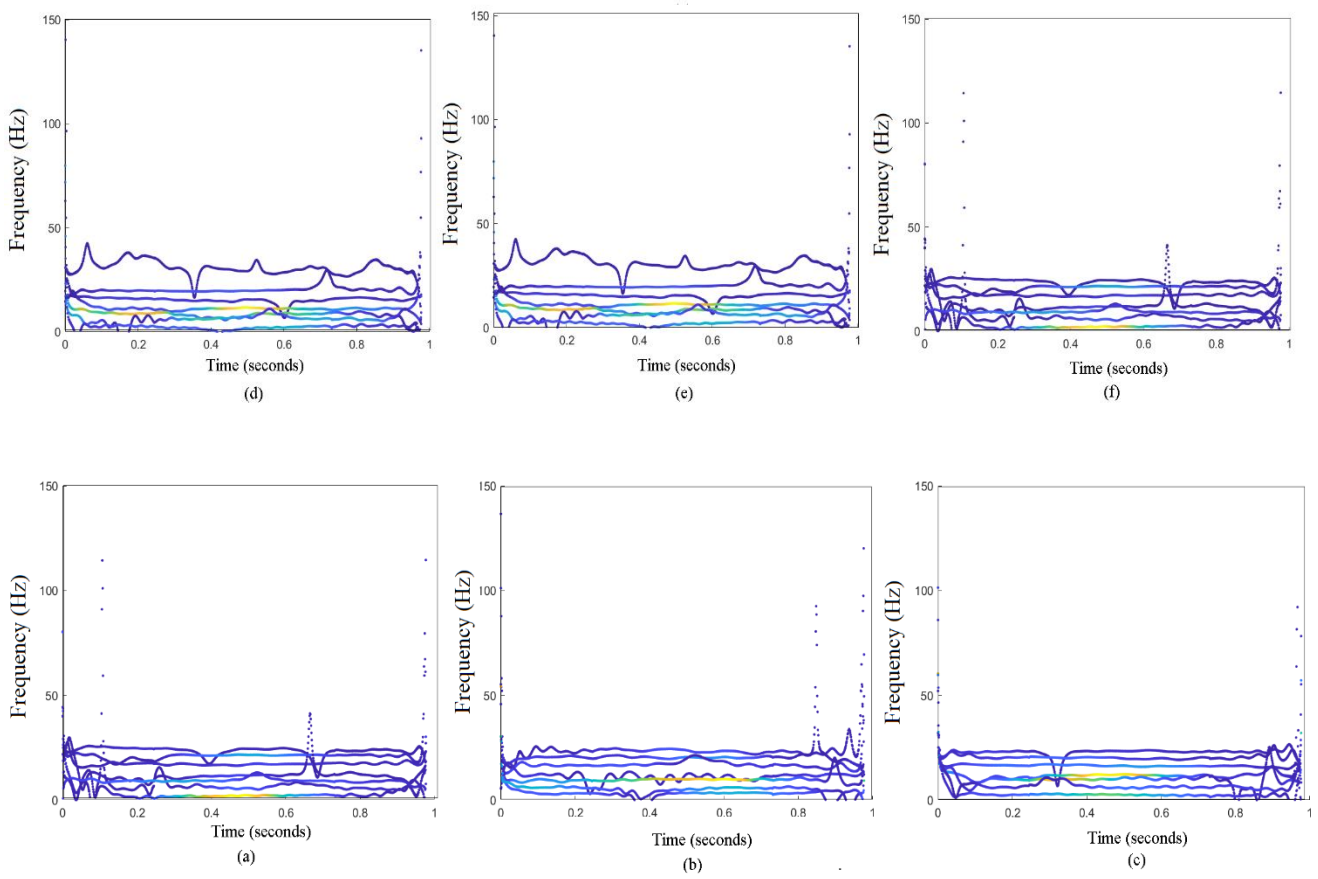


Figure 22: Plot of obtained TFR for (a) Up, (b) Down, (c) Right, (d) Left, (e) Forward, and (f) Backward classes using DMD method.

5.3 Classification Performance using CNN

To evaluate the performance of various deep networks and classifiers, our experiments were conducted on a computer system with specific specifications: an Intel® Core™ i7-9750H CPU with a base frequency of 2.6 GHz (up to 4.5 GHz in high-performance mode), an NVIDIA GeForce RTX 2060 GPU, and 16 GB of RAM. The simulations were performed using MATLAB [108]. For assessing the CNN model's performance, it was trained on two-dimensional time-frequency representation (TFR) images that encompass joint time-frequency features. Table 4 presents the comparative average accuracies for different classes. The average accuracies were calculated for the imagined speech commands, including Up, Down, Left, Right, Forward, and Backward, as well as the accuracy for all six classes combined. The highest achieved accuracy across all classes was 70.99% [14]. The performance of the proposed model can be summarized as follows:

The DMD method was evaluated for classifying six different classes of imagined speech stimuli using TFR images derived from DMD of EEG signals. The TFR images corresponding to different classes, as shown in Table 4, were utilized for TFR image classification using a CNN classifier. The CNN model consists of various components, including the image input layer, convolutional layer, batch normalization layer, rectified linear unit, max-pooling layer, fully connected layer, and soft-max layer [2]. The number of layers in the CNN can be adjusted based on the input size, and it is not necessary to include all layers in the network. In this study, the network parameters were efficiently tuned, and the minimum number of layers was employed to reduce computational time. Figure 16 illustrates the block diagram of the implemented CNN network.

The classification employed a learning rate of 0.01, with a maximum of 40 epochs and a batch size of 256. For training and testing the classifier, 70% and 30% of the TFR images were used, respectively. The highest achieved accuracies for TFR classification were reported as 70.90% and 71.99%, respectively. The classification performance of the imagined speech task is presented in terms of accuracy percentages for different classes [16]. Accuracy serves as a metric to evaluate the classification model's effectiveness in accurately identifying the intended speech commands based on EEG signals. For the left-right classification task, an accuracy of 70.90% was achieved, indicating successful differentiation between left and right speech commands with relatively high accuracy.

Similarly, for the up-down classification task, an accuracy of 68.98% was obtained, demonstrating the model's ability to differentiate between these speech commands.

In the forward-backward classification task, an accuracy of 70.01% was achieved, indicating accurate classification of the intention to move forward or backward based on the EEG signals. Finally, the overall classification accuracy for all classes combined was reported as 71.99%, representing the overall performance of the model in distinguishing between all the imagined speech commands.These results, as shown in Table 4, highlight the potential of the EEG-based BCI system using DMD for accurate and reliable classification of imagined speech commands. The achieved accuracy percentages demonstrate the feasibility of utilizing EEG signals to accurately decode and classify the intended speech commands.

Table 4: Imagined speech classification performance.

Classification task	Accuracy (%)
Right versus Left	70.90
Down versus Up	68.98
Backward versus Forward	70.01
All classes	71.99

Table 4 presents the extracted average TFRs across 15 subjects for six different imagined speech command classes. The subjects are denoted as S01 to S015, and the imagined speech commands include Backward, Down, Forward, Left, Right, and Up. The Table 4 provides the average number of TFRs obtained for each subject and each imagined speech command class. The values represent the quantity of TFRs extracted during the analysis.

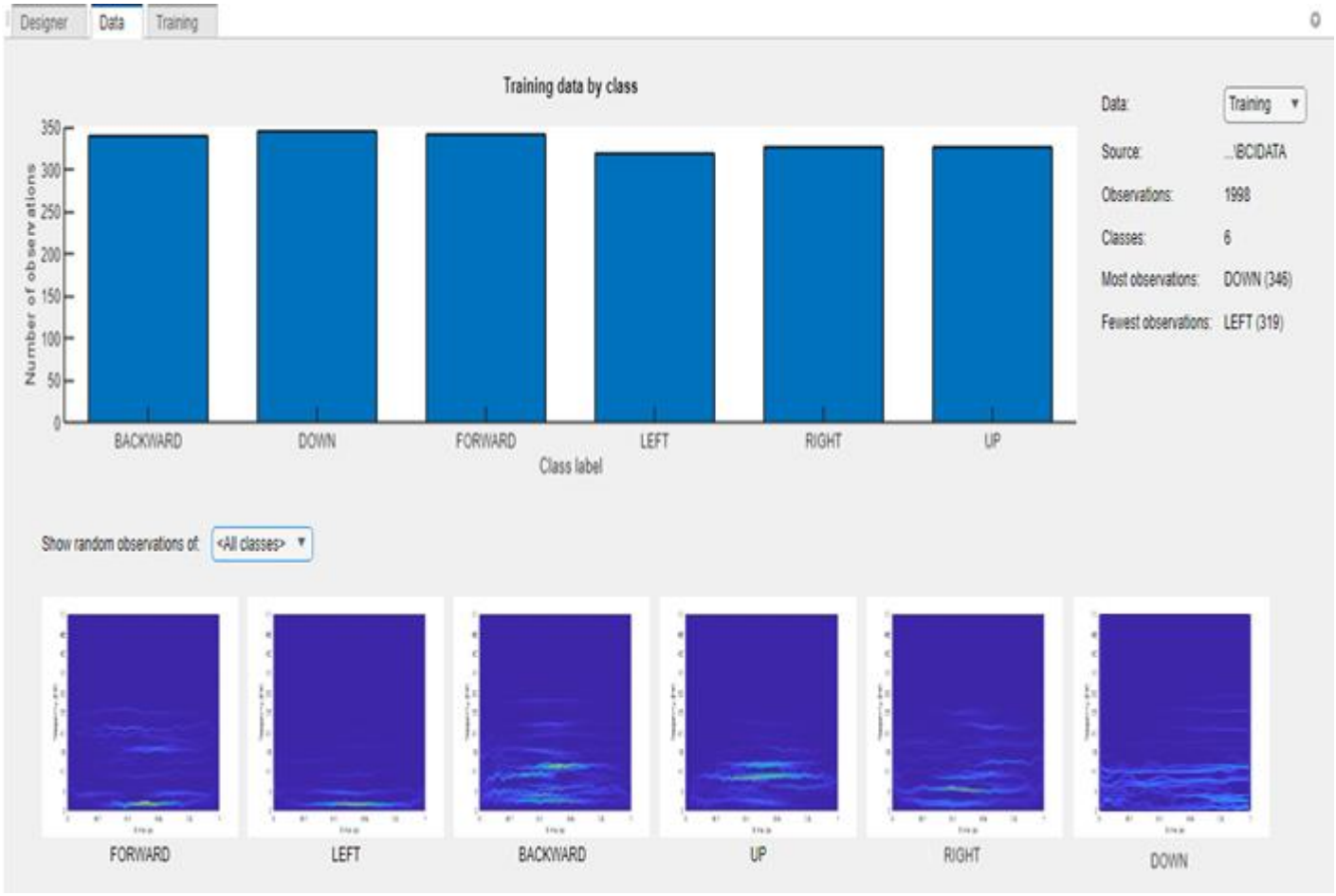


Fig 23: Data loading process for extraction of CNN feature in training phase for six imagined speech commands.

5.4 Comparison with Existing Techniques

A comparison of different methods employed for classifying imagined speech commands is illustrated in Table 6. This comparison showcases the feature extraction techniques, classifiers, and accuracy rates associated with each method [41]. The inclusion of existing methods in Table 6 serves as a benchmark for evaluating the performance of the proposed approach, which utilizes DMD-TFRs and a CNN classifier.

Table 6: comparative analysis of imagined speech classification methods to the existing methods

References	Classes	Feature extraction techniques	Classifier	Accuracy
Correto et al. [80]	Classification of six imagined command class	Discrete wavelet transforms domain energy features	SVM	17.46%
Dash et al. [58]	Classification of seven phonemic prompts and four words	Slop domain L1-norm	SVM	Phonemes:20% Pair of words: 44%
Cooney et al. [37]	Classification of imagined vowels and commands	Learnable features CNN layers	10-deep based deep CNN model	19.81%
Dasalla et al. [52]	Binary classification of vowels, short, and long words	Common spatial filter	RVM	Vowels: 49% Shortwords: 50.1% Long words: 66.2%
Paul et al. [109]	Classification of three Hindi words [10]	AR coefficients, Hjorth parameters and sample entropy	SVM	63%
	Proposed work	DMD-TFRs	CNN	70.47%

The study conducted a comparative analysis of different methods used for classifying imagined speech commands using EEG signals. The researchers explored various techniques for extracting features and employed different classification algorithms to improve accuracy. In previous works [80], discrete wavelet transform (DWT) was used to extract energy features in the domain, and support vector

machines (SVM) were employed for classification. The achieved accuracy for classifying six imagined command classes was reported as 17.46%. Similarly, in [58], the focus was on classifying phonemic prompts and words using an SVM classifier, with accuracy rates of 20% for phonemes and 44% for word pairs [37]. In [37], learnable features were extracted using CNN layers, and a deep CNN model with 10 layers was used for classification. The accuracy achieved for classifying imagined vowels and commands was 19.81%. These previous works [80, 58, 37] reported relatively lower accuracy rates.

To address this, Dasalla et al. [52] achieved improved accuracy by employing binary classification of vowels, short words, and long words using common spatial patterns (CSP) for feature extraction and a relevance vector machine (RVM) as the classifier. The accuracy rates were reported as 49% for vowels, 50.1% for short words, and 66.2% for long words [82]. In addition, features such as autoregressive (AR) coefficients, Hjorth parameters, and sample entropy were extracted, and an SVM classifier was used. The accuracy achieved for classifying three Hindi words was 63%. Furthermore, in [109], the multivariate fast and adaptive empirical decomposition-based method (MFAEMD) was employed for feature extraction, and dictionary learning (DL) was used for classification of six imagined command classes. The accuracy obtained for classifying imagined vowels and commands was reported as 60.72%.

To improve upon these previous studies, the proposed work utilized DMD-TFRs for time-frequency representation-based feature extraction and a CNN classifier. The achieved accuracy for classifying imagined speech commands was 70.47%. These results indicate that the proposed method surpasses the accuracy of existing methods [58, 37, 80, 82, 109], demonstrating its effectiveness in classifying imagined speech commands. This comparative analysis underscores the significance of the proposed approach based on DMD-TFRs and CNN, highlighting its potential as a robust and accurate method for EEG-based imagined speech classification. The improved accuracy achieved in this study contributes to the existing research in EEG-based BCI systems and opens avenues for further advancements in this field.

Chapter 6

Conclusion and Scope for Future Work

The utilization of the DMD-based feature extraction method has demonstrated its effectiveness in capturing the temporal dynamics and frequency characteristics of EEG signals associated with imagined speech. The EEG-based imagined speech BCI using DMD has shown promising results in decoding imagined speech tasks. In terms of reconstruction accuracy for multi-component amplitude modulated signals, the DMD method outperforms the EMD and EWT methods as indicated by the lower MSE values. The DMD method achieves superior reconstruction accuracy with the lowest MSE values for all signal components compared to the EMD and EWT methods. On the other hand, the EMD method exhibits relatively less accurate reconstruction with higher MSE values, while the EWT method shows the least accurate reconstruction with the highest MSE values. Therefore, based on the MSE values, the DMD method is concluded to provide better reconstruction accuracy for multi-component amplitude modulated signals in this study.

The developed DMD-based classification framework has been found effective and robust for decoding imagined speech from EEG signals, outperforming existing approaches. In the proposed work, DMD-TFRs were used for feature extraction, and a CNN classifier was employed. The achieved accuracy for the classification of imagined speech commands was 70.47%. These results demonstrate the superiority of the proposed method in terms of accuracy, highlighting its effectiveness in classifying imagined speech commands. The combination of DMD-based feature extraction and classification techniques has provided a reliable and powerful approach for extracting meaningful information from EEG signals related to imagined speech. The system has demonstrated its potential for real-world applications, including communication aids and neurorehabilitation.

In future research, the efficiency of the EEG-based imagined speech BCI can be enhanced for clinical practices, including communication aids and neurorehabilitation. This can be achieved by exploring advanced signal processing techniques, optimizing classification models, and integrating other modalities or features to improve overall performance. Additionally, investigating the generalizability of the system across larger populations and different demographic groups would provide valuable insights into its scalability and applicability in diverse user populations.

In conclusion, the EEG-based imagined speech BCI using DMD holds significant potential for advancing the field of neurorehabilitation and communication aids. By addressing the mentioned aspects and continuously refining the system's capabilities, future research can contribute to the development of more accurate, reliable, and user-friendly BCI systems for decoding imagined speech, ultimately improving the quality of life for individuals with communication impairments.

6.1 Summary of the Study

In summary, this study focused on developing an effective and robust system for decoding imagined speech from EEG signals using DMD. By employing signal processing techniques, feature extraction with DMD, and classification models, the study successfully extracted meaningful information and accurately classified the intended speech from the brain signals.

The results of the study demonstrated the feasibility and effectiveness of the proposed approach. The DMD-based feature extraction method proved to be highly effective in capturing the temporal dynamics and frequency characteristics of the EEG signals associated with imagined speech. The classification models, trained on these extracted features, achieved high levels of accuracy in decoding the intended speech.

The study also highlighted the potential applications of the EEG-based imagined speech BCI, such as communication aids and neurorehabilitation. The system's ability to interpret and translate imagined speech into actionable commands has significant implications for individuals with speech impairments, providing them with a means to communicate and interact with their environment.

Furthermore, the study discussed the challenges and limitations associated with the proposed approach, including variations in EEG signals among individuals and the need for further optimization and generalization of the system. These findings underscore the importance of ongoing research and development in the field of EEG-based BCIs, with a focus on addressing these challenges and improving the system's robustness and usability.

Overall, this study contributes to the advancement of EEG-based BCIs for decoding imagined speech and highlights the potential of DMD as a reliable feature extraction method. The findings lay the groundwork for future research in this area, paving the way for more accurate, dependable, and user-friendly BCI systems for imagined speech communication.

6.2 Contributions and Findings

The study on EEG-Based imagined speech BCI using DMD has made significant contributions and generated important findings. Firstly, it has demonstrated the feasibility and effectiveness of employing DMD as a feature extraction method to capture the temporal and spectral characteristics of EEG signals associated with imagined speech. This contribution addresses the challenges posed by non-stationarity and non-linearities in EEG data and provides a robust approach for extracting relevant features.

Secondly, the study has developed and evaluated a classification model based on the extracted DMD features, achieving high accuracy in decoding imagined speech from EEG signals. This finding highlights the potential of EEG-based BCIs in translating brain signals into meaningful speech commands, which can have profound implications for individuals with speech impairments.

Furthermore, the study has explored the application of multichannel DMD techniques in the proposed EEG-Based imagined speech BCI, particularly in the context of communication aids and neurorehabilitation. By demonstrating the system's ability to decode imagined speech and translate it into actionable commands, the study opens up new possibilities for enhancing communication and interaction for individuals with limited speech capabilities.

The findings of the study contribute to the broader field of EEG-based BCIs by showcasing the effectiveness of DMD as a feature extraction method and its potential for decoding imagined speech. The study not only advances our understanding of the underlying mechanisms of imagined speech in EEG signals but also provides practical insights for developing more accurate and user-friendly BCI systems.

In summary, the contributions and findings of this study highlight the promise of EEG-Based imagined speech BCI using DMD, setting the stage for future research and development in the field. These findings have the potential to significantly impact the lives of individuals with speech impairments by offering new avenues for communication and improving their quality of life. The study extends the application of the DMD method and explores the perspective of single-channel EEG-based DMD for decoding imagined speech from multichannel EEG signals.

The main contributions of the thesis can be summarized as follows:

a) The thesis presents an experiment for the EEG-based classification of six imagined commands using a single-channel-based DMD approach, employing a Hankelization and diagonal averaging procedure. The analysis of imagined signals is conducted for six brain regions.

b) From the imagined EEG signals of six channels, the thesis evaluates the instantaneous energy and frequency using Hilbert transform, and extracts features such as Hilbert spectrum of intrinsic mode functions and time-frequency representations (TFRs). A novel multichannel DMD method is proposed, simultaneously evaluating the mean of all extracted TFRs from the six channels and fifteen subjects for six different imagined speech classes.

c) The thesis proposes a deep CNN architecture trained on TFR features (2-D images) of six classes. The experimental performance indicates that the alpha and beta bands are better suited for classifying imagined speech features compared to other EEG frequency bands.

6.3 Future Research Directions

Future research directions for EEG-Based imagined speech BCI using DMD have been identified, providing valuable insights and suggesting areas for further investigation. Firstly, there is a need to refine the DMD algorithm and explore variations of the method to improve its performance in extracting relevant features from EEG signals. This may involve adjusting parameter settings, exploring alternative decomposition algorithms, or incorporating other signal processing techniques to enhance the accuracy and robustness of feature extraction.

Secondly, future studies should involve larger and more diverse datasets to evaluate the generalizability of the proposed BCI system. Collecting EEG data from a broader population, including individuals with different speech impairments and varying cognitive abilities, will allow for a comprehensive assessment of the system's performance across various user groups. Real-time and online experiments can also provide insights into the system's usability and reliability in practical scenarios.

Furthermore, the integration of other machine learning algorithms and techniques can be explored to improve the classification accuracy and response time of the imagined speech BCI. Deep learning approaches, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can be investigated to capture complex temporal patterns in EEG signals and enhance the system's ability to decode imagined speech.

Additionally, the combination of multiple modalities, such as EEG and electromyography (EMG), can be studied to provide complementary information and improve the overall accuracy and robustness of the BCI system. The integration of other physiological signals or neuroimaging techniques, such as functional near-infrared spectroscopy (fNIRS) or functional magnetic resonance imaging (fMRI), can offer valuable insights into the underlying neural mechanisms of imagined speech and further enhance the performance of the BCI.

Lastly, there is a need to develop user-friendly and wearable EEG devices that are suitable for long-term use and can be seamlessly integrated into daily life activities. Improving the comfort, ease of use, and accessibility of EEG acquisition systems will facilitate the widespread adoption of EEG-Based imagined speech BCIs and make them more practical for real-world applications.

In conclusion, future research in EEG-Based imagined speech BCI using DMD should focus on improving the accuracy, usability, and generalizability of the system. This can be achieved through the refinement of feature extraction methods, the expansion of datasets, the exploration of advanced machine learning techniques, the integration of multiple modalities, and the improvement of EEG acquisition device usability. These research directions will contribute to the advancement and practical implementation of imagined speech BCIs, ultimately enhancing communication and interaction for individuals with speech impairments.

REFERENCES

[1] L. J. Stovner, J. M. Hoff, S. Svalheim, and N. E. Gilhus, “Neurological disorders in the Global Burden of Disease 2010 study,” Acta Neurologica Scandinavica, vol. 129, pp. 1–6, Mar. 2014.

[2] L. Durak and O. Arikan, “Short-time Fourier transform: two fundamental properties and an optimal implementation,” IEEE Transactions on Signal Processing, vol. 51, no. 5, pp. 1231–1242, May 2003.

[3] S. Qian, Introduction to Time-Frequency and Wavelet Transforms. 2001.

[4] L. Qin and B. He, “A wavelet-based time–frequency analysis approach for classification of motor imagery for brain–computer interface applications,” Journal of Neural Engineering, vol. 2, no. 4, pp. 65–72, Aug. 2005.

[5] R. M. Rao, Wavelet transforms: Introduction to theory and applications. Pearson Education India, 1998.

[6] Daubechies, “The Wavelet Transform, Time-Frequency Localization and Signal Analysis,” pp. 442–486, Dec. 2009.

[7] M. Rhif, A. Ben Abbes, I. Farah, B. Martínez, and Y. Sang, “Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review,” Applied Sciences, vol. 9, no. 7, p. 1345, Mar. 2019.

[8] B. Xu et al., "Wavelet Transform Time-Frequency Image and Convolutional Network-Based Motor Imagery EEG Classification," in IEEE Access, vol. 7, pp. 6084-6093, 2019.

[9] Chi, X.; Hagedorn, J.B.; Schoonover, D.; D'Zmura, M. EEG-Based discrimination of imagined speech phonemes. Int. J. Bioelectromagn. 2011.

[10] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," in IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531-544, Feb.1, 2014.

[11] J. Lian, Z. Liu, H. Wang, and X. Dong, “Adaptive variational mode decomposition method for signal processing based on mode characteristic,” Mechanical Systems and Signal Processing, vol. 107, pp. 53–77, Jul. 2018.

[12] P. Schmid, “Dynamic Mode Decomposition and Its Variants,” vol. 54, no. 1, pp. 225–254, Jan. 2022.

[13] P. J. Schmid, “Application of the dynamic mode decomposition to experimental data,” Experiments in Fluids, vol. 50, no. 4, pp. 1123–1130, Feb. 2011.

[14] J. Nathan Kutz, X. Fu, and S. L. Brunton, “Multiresolution Dynamic Mode Decomposition,” Siam Journal on Applied Dynamical Systems, vol. 15, no. 2, pp. 713–735, Apr. 2016.

[15] C. E. Stafstrom and L. Carmant, “Seizures and Epilepsy: An Overview for Neuroscientists,” Cold Spring Harbor Perspectives in Medicine, vol. 5, no. 6, pp. a022426–a022426, Jun. 2015.

[16] Soledad Le Clainche and J. M. Vega, “Higher Order Dynamic Mode Decomposition,” vol. 16, no. 2, pp. 882–925, May 2017.

[17] H. Lu and D. M. Tartakovsky, “Prediction Accuracy of Dynamic Mode Decomposition,” SIAM Journal on Scientific Computing, vol. 42, no. 3, pp. A1639–A1662, Jan. 2020.

[18] Daniel Duke, Julio Soria, Damon Honnery. "An error analysis of the dynamic mode decomposition", Experiments in Fluids, 2011.Nature 2019.

[19] Jiexin Gao, S. J. Haghighi and D. Hatzinakos, "Reference empirical mode decomposition," 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), Toronto, ON, Canada, 2014.

[20] P. Flandrin, G. Rilling, and P. Goncalves, “Empirical Mode Decomposition as a Filter Bank,” IEEE Signal Processing Letters, vol. 11, no. 2, pp. 112–114, Feb. 2004.

[21] N. E. Huang et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[22] Z. WU and N. E. HUANG, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” Advances in Adaptive Data Analysis, vol. 01, no. 01, pp. 1–41, Jan. 2009.

[23] Ozlem Karabiber Cura, Aydin Akan. "Analysis of epileptic EEG signals by using dynamic mode decomposition and spectrum", Biocybernetics and Biomedical Engineering, 2020.

[24] G. W. Stewart, “On the Early History of the Singular Value Decomposition,” SIAM Review, vol. 35, no. 4, pp. 551–566, Dec. 1993.

[25] A. Bhattacharyya, L. Singh, R. B. Pachori,, “Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals,” Digital Signal Processing, vol. 78, p.p 185-196, 2018.

[26] E. R. Henry and J. Hofrichter, “[8] Singular value decomposition: Application to analysis of experimental data,” ScienceDirect, Jan. 01, 1992.

[27] K. Lange, “Singular Value Decomposition,” pp. 129–142, Jan. 2010.

[28] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A Multilinear Singular Value Decomposition,” SIAM Journal on Matrix Analysis and Applications, vol. 21, no. 4, pp. 1253–1278, Jan. 2000.

[29]. Daubechies, J. Lu, and H.-T. Wu, “Synchro squeezed wavelet transforms: An empirical mode decomposition-like tool,” Applied and Computational Harmonic Analysis, vol. 30, no. 2, pp. 243–261, Mar. 2011.

[30] L. Durak and O. Arikan, “Short-time fourier transform: two fundamental properties and an optimal implementation,” IEEE Transactions on Signal Processing, vol. 51, no. 5, pp. 1231–1242, May 2003.

[31] M. Portnoff, “Time-frequency representation of digital signals and systems based on short-time Fourier analysis,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 1, pp. 55–69, Feb. 1980.

[32] Y. Avargel and I. Cohen, “Modeling and Identification of Nonlinear Systems in the Short-Time Fourier Transform Domain,” IEEE Transactions on Signal Processing, vol. 58, no. 1, pp. 291–304, Jan. 2010.

[33] C. Mateo and J. A. Talavera, “Short-time Fourier transform with the window size fixed in the frequency domain,” Digital Signal Processing, vol. 77, pp. 13–21, Jun. 2018.

[34] Chao He, Jialu Liu, Yuesheng Zhu, Wencai Du. "Data Augmentation for Deep Neural Networks Model in EEG Classification Task: A Review", Frontiers in Human Neuroscience, 2021.

[35] Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-basedelectroencephalography analysis: A systematic review. J. Neural Eng. 2019.

[36] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a Convolutional Neural Network,” 2017 International Conference on Engineering and Technology (ICET), pp. 1–6, Aug. 2017.

[37] Cooney, C.; Ra_aella, F.; Coyle, D. Optimizing Input Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Bari, Italy, 6–9 October 2019.

[38] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, “Convolutional Neural Network (CNN) for Image Detection and Recognition,” IEEE Xplore, Dec. 01, 2018.

[39] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 1–21, 2021.

[40] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” arXiv:1511.08458 [cs], Dec. 2015.

[41] A. F. Agarap, “An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification,” arXiv:1712.03541 [cs, stat], Feb. 2019.

[42] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, “Transfer learning in imagined speech EEG-based BCIs,” Biomedical Signal Processing and Control, vol. 50, pp. 151–157, Apr. 2019.

[43] J. A. Ramirez-Quintana, J. M. Macias-Macias, G. Ramirez-Alonso, M. I. Chacon-Murguia, and L. F. Corral-Martinez, “A Novel Deep Capsule Neural Network for Vowel Imagery Patterns from EEG Signals,” papers.ssrn.com, Dec. 22, 2021.

[44] P. Kalaivani, S. Mohamed Mansoor Roomi, M. Maheesha, V. Subathraa. "Chapter 217 Novel CNN Architecture for Human Action Recognition", Springer Science and Business Media LLC, 2021.

[45] Z. Huang, Y. Chen, M. Pan. "Time-frequency characterization of atrial fibrillation from surface ECG based on Hilbert-Huang transform", Journal of Medical Engineering & Technology, 2009.

[46] Li Hai, Xue Guo-qiang, Zhao Pan, Zhong Huasen, Muhammad Younis Khan. "The Hilbert– Huang Transform-Based Denoising Method for the TEM Response of a PRBS Source Signal", Pure and Applied Geophysics, 2016.

[47] R. B. Pachori, Time-Frequency Analysis Techniques and their Applications, 1st Edition. Boca Raton, FL: CRC Press, 2023.

[48] N. E. Huang, X. Chen, M.-T. LO, and Z. WU, “on Hilbert spectral representation: a true time-frequency representation for nonlinear and nonstationary data,” Advances in Adaptive Data Analysis, vol. 03, no. 01n02, pp. 63–93, Apr. 2011.

[49] N. E. Huang, Z. Shen, and S. R. Long, “A new view of nonlinear water waves: the Hilbert spectrum,” vol. 31, no. 1, pp. 417–457, Jan. 1999.

[50] N. E. Huang, X. CHEN, M.-T. LO, and Z. WU, “on Hilbert spectral representation: a true time-frequency representation for nonlinear and nonstationary data,” Advances in Adaptive Data Analysis, vol. 03, no. 01n02, pp. 63–93, Apr. 2011.

[51] Ciaran Cooney, Raffaella Folli, Damien Coyle. "Opportunities, pitfalls and trade-offs in designing protocols for measuring the neural correlates of speech", Neuroscience & Biobehavioral Reviews, 2022.

[52] DaSalla, C.S.; Kambara, H.; Sato, M.; Koike, Y. Single-trial classification of vowel speech imagery using common spatial patterns. Neural Netw. 2009, 22, 1334–1339.

[53] Brigham, K.; Kumar, B.V.K.V. Imagined Speech Classification with EEG Signals for Silent Communication: A Preliminary Investigation into Synthetic Telepathy. In Proceedings of the 2010 4th International Conference on Bioinformatics and Biomedical Engineering, Chengdu, China, 18–20 June 2010.

[54] Nrushingh Charan Mahapatra, Prachet Bhuyan. "Multiclass Classification of Imagined Speech Vowels and Words of Electroencephalography Signals Using Deep Learning", Advances in Human-Computer Interaction, 2022.

[55] Dong-Yeon Lee, Minji Lee, Seong-Whan Lee. "Classification of Imagined Speech Using Siamese Neural Network", 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020.

[56] Bogue, R. “Brain-computer interfaces: Control by thought”, Ind. Robot. Int. J. 2010.

[57] Cooney, C.; Folli, R.; Coyle, D. Mel Frequency Cepstral Coe_cients Enhance Imagined Speech Decoding Accuracy from EEG. In Proceedings of the 29th Irish Signals and Systems Conference (ISSC), Belfast, UK,21–22 June 2018.

[58] S. Dash, R. K. Tripathy, G. Panda and R. B. Pachori, "Automated Recognition of Imagined Commands from EEG Signals Using Multivariate Fast and Adaptive Empirical Mode Decomposition Based Method," in IEEE Sensors Letters, vol. 6, no. 2, pp. 1-4, Feb. 2022.

[59] C. Park, D. Looney, N. ur Rehman, A. Ahrabian and D. P. Mandic, "Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 21, no. 1, pp. 2013.

[60] P. Kumar and E. Scheme, "A Deep Spatio-Temporal Model for EEG-Based Imagined Speech Recognition," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021.

[61] A. L. Rusnac and O. Grigore, "Convolutional Neural Network applied in EEG imagined phoneme recognition system," 2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 2021.

[62] Yoshiyuki Shiraishi, Yoshinobu Kawahara, Okito Yamashita, Ryohei Fukuma et al. "Neural decoding of electrocorticographic signals using dynamic mode decomposition", Journal of Neural Engineering, 2020.

[63] D. -H. Lee, S. -J. Kim and K. -W. Lee, "Decoding High–level Imagined Speech using Attention–based Deep Neural Networks," 2022 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Korea, Republic of, 2022.

[64] Y. V. Varshney and A. Khan, “Imagined speech classification using six phonetically distributed words,” Frontiers, [Accessed: 27-Sep-2022].

[65] S. Dash, R. K. Tripathy, G. Panda, and R. B. Pachori, “Automated Recognition of Imagined Commands from EEG Signals using Multivariate Fast and Adaptive Empirical Mode Decomposition based Method,” IEEE Sensors Letters, pp. 1–1, 2022.

[66] Juan A. Ramirez-Quintana, Jose M. Macias-Macias, Graciela Ramirez-Alonso, Mario I. Chacon-Murguia, Luis F. Corral-Martinez. "A novel Deep Capsule Neural Network for Vowel Imagery patterns from EEG signals", Biomedical Signal Processing and Control,2023.

[67] Shaswati Dash, Rajesh Kumar Tripathy, Dinesh Kumar Dash, Ganapati Panda, Ram Bilas Pachori. "Multiscale Domain Gradient Boosting Models for the Automated Recognition of Imagined Vowels Using Multichannel EEG Signals", IEEE Sensors Letters, 2022.

[68] M. R. Asghari Bejestani, G. R. Mohammad Khani, V. R. Nafisi, and F. Darakeh, “EEG-Based Multiword Imagined Speech Classification for Persian Words,” BioMed Research International, vol. 2022.

[69] Fu Li, Weibing Chao, Yang Li, Boxun Fu, Youshuo Ji, Hao Wu, Guangming Shi. "Decoding imagined speech from EEG signals using hybrid-scale spatial-temporal dilated convolution network", Journal of Neural Engineering, 2021.

[70] Diego Lopez-Bernal, David Balderas, Pedro Ponce, Arturo Molina. "A State-of-the-Art Review of EEG-Based Imagined Speech Decoding", Frontiers in Human Neuroscience, 2022.

[71] Dipti Pawar, Sudhir Dhage. "Imagined Speech Classification using EEG based Brain- Computer Interface", 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022.

[72] Zhao, S.; Rudzicz, F. Classifying phonological categories in imagined and articulated speech. In Proceedingsof the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane,Australia, 19–24 April 2015.

[73] Alexander E. Hramov, Vladimir A. Maksimenko, Alexander N. Pisarchik. "Physical principles of brain-computer interfaces and their applications for rehabilitation, robotics and control of human brain states", Physics Reports, 2021.

[74] P.P. Mini, Tessamma Thomas, R. Gopikakumari. "EEG based direct speech BCI system using a fusion of SMRT and MFCC/LPCC features with ANN classifier", Biomedical Signal Processing and Control, 2021.

[75] O. Karabiber Cura and A. Akan, “Analysis of epileptic EEG signals by using dynamic mode decomposition and spectrum,” Biocybern. Biomed. Eng., vol. 41, no. 1, pp. 28–44, 2021.

[76] Y. Pei et al., "A Tensor-Based Frequency Features Combination Method for Brain–Computer Interfaces," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 465-475, 2022.

[77] Mohamad Amin Bakhshali, Morteza Khademi, Abbas Ebrahimi-Moghadam. "Investigating the neural correlates of imagined speech: An EEG-based connectivity analysis", Digital Signal Processing, 2022.

[78] MUKUL, Manoj Kumar, and Fumitoshi MATSUNO. "Feature Extraction from Subband Brain Signals and Its Classification", SICE Journal of Control Measurement and System Integration, 2011.

[79] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, “A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition,” IEEE Sensors Journal, vol. 21, no. 3, pp. 3579–3591, Feb. 2021.

[80] G. A. Pressel Coretto, I. E. Gareis, and H. L. Rufiner, “Open access database of EEG signals recorded during imagined speech,” in 12th International Symposium on Medical Information Processing and Analysis, 2017.

[81] Puce, A.; Hämäläinen, M.S. A review of issues related to data acquisition and analysis in EEG/MEG studies.Brain Sci. 2017.

[82] A. O. Barja, “Alternative Form of Ordinary Differential Equation of Electroencephalography Signals During an Epileptic Seizure,” Malaysian Journal of Fundamental and Applied Sciences, vol. 17, no. 2, pp. 109–113, Apr. 2021.

[83] T. Baba, “Time-Frequency Analysis Using Short Time Fourier Transform,” The Open Acoustics Journal, vol. 5, no. 1, pp. 32–38, Aug. 2012.

[84] S. M. Qaisar, L. Fesquet, and M. Renaudin, “An Adaptive Resolution Computationally Efficient Short-Time Fourier Transform,” Research Letters in Signal Processing, vol. 2008, pp. 1–5, 2008.

[85] L.-C. Shi, Y. Li, R. Sun, and B.-L. Lu, “A Sparse Common Spatial Pattern Algorithm for Brain-Computer Interface,” pp. 725–733, Nov. 2011.

[86] D. Cheng, S. Zhang, Z. Deng, Y. Zhu, and M. Zong, “kNN Algorithm with Data-Driven k Value,” Advanced Data Mining and Applications, pp. 499–512, 2014.

[87] P. Cunningham and S. J. Delany, “k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples),” arXiv:2004.04523 [cs, stat], Apr. 2020.

[88] R. Pati, A. K. Pujari, P. Gahan, and V. Kumar, “Independent Component Analysis: A Review, with Emphasis on Commonly used Algorithms and Contrast Function,” Computación y Sistemas, vol. 25, no. 1, Feb. 2021.

[89] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” Neural Networks, vol. 13, no. 45, pp. 411–430, 2000.

[90] R. Li, D. Yang, F. Fang, K.-S. Hong, A. L. Reiss, and Y. Zhang, “Concurrent fNIRS and EEG for Brain Function Investigation: A Systematic, Methodology-Focused Review,” Sensors, vol. 22, no. 15, p. 5865, Aug. 2022.

[91] A. Rezazadeh Sereshkeh, R. Yousefi, A. T. Wong, F. Rudzicz, and T. Chau, “Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech,” Brain-Computer Interfaces, vol. 6, no. 4, pp. 128–140, Oct. 2019.

[92] D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, “A State-of-the-Art Review of EEG-Based Imagined Speech Decoding,” Frontiers in Human Neuroscience, vol. 16, Apr. 2022.

[93] Y. Höller et al., “Reliability of EEG Measures of Interaction: A Paradigm Shift Is Needed to Fight the Reproducibility Crisis,” Frontiers in Human Neuroscience, vol. 11, Aug. 2017.

[94] P. Agarwal and S. Kumar, “EEG-based imagined words classification using Hilbert transform and deep networks,” May 2023.

[95] B. V. Berg, S. Donkelaar and M. Alimardani, "Inner Speech Classification using EEG Signals: A Deep Learning Approach," 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), Magdeburg, Germany, 2021.

[96] S. J. Albert and J. Kesselring, “Neurorehabilitation of stroke,” Journal of Neurology, vol. 259, no. 5, pp. 817–832, Oct. 2011.

[98] Y. V. Varshney and A. Khan, “Imagined Speech Classification Using Six Phonetically Distributed Words,” Frontiers in Signal Processing, vol. 2, Mar. 2022.

[99] J. Gilles, "Empirical Wavelet Transform," in IEEE Transactions on Signal Processing, vol. 61, no. 16, pp. 3999-4010, Aug.15, 2013.

[100] A. S. Al-Fahoum and A. A. Al-Fraihat, “Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains,” ISRN Neuroscience, Feb. 13, 2014.

[101] T. Proix et al., “Imagined speech can be decoded from low- and cross-frequency intracranial EEG features,” Nature Communications, vol. 13, no. 1, Jan. 2022.

[102] S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, “Transfer learning in imagined speech EEG-based BCIs,” Biomedical Signal Processing and Control, vol. 50, pp. 151–157, Apr. 2019.

[103] R. M. Parry et al., “k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction,” The Pharmacogenomics Journal, vol. 10, no. 4, pp. 292–309, Jul. 2010

[104] R. Kronland-Martinet, “The Wavelet Transform for Analysis, Synthesis, and Processing of Speech and Music Sounds,” Computer Music Journal, vol. 12, no. 4, p. 11, 1988.

[105] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, “A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition,” IEEE Sensors Journal, vol. 21, no. 3, pp. 3579–3591, Feb. 2021.

[106] T. Warbrick, “Simultaneous EEG-fMRI: What Have We Learned and What Does the Future Hold?” Sensors, vol. 22, no. 6, p. 2262, Mar. 2022.

[103] J. Semmlow, “Chapter 3 - Fourier Transform: Introduction,” ScienceDirect, Jan. 01, 2012.

[104] B. Aiazzi, S. Baronti, and M. Selva, “2 - Image fusion through multiresolution oversampled decompositions,” ScienceDirect, Jan. 01, 2008.

[105] J. T. Panachakel, A. G. Ramakrishnan, and T. V. Ananthapadmanabha, “A Novel Deep Learning Architecture for Decoding Imagined Speech from EEG,” arXiv.org, Mar. 18, 2020.

[106] J. T. Panachakel, A. G. Ramakrishnan and T. V. Ananthapadmanabha, "Decoding Imagined Speech using Wavelet Features and Deep Neural Networks," 2019 IEEE 16th India Council International Conference (INDICON), Rajkot, India, 2019.

[107] J. A. Ramirez-Quintana, J. M. Macias-Macias, G. Ramirez-Alonso, M. I. Chacon-Murguia, and L. F. Corral-Martinez, “A novel Deep Capsule Neural Network for Vowel Imagery patterns from EEG signals,” Biomedical Signal Processing and Control, vol. 81, p. 104500, Mar. 2023.

[108] C. Brunner, A. Delorme, and S. Makeig, “Eeglab – an Open Source Matlab Toolbox for Electrophysiological Research,” Biomedical Engineering / Biomedizinische Technik, Jan. 2013.

[109] Y. Paul, R. A. Jaswal and S. Kajal, "Classification of EEG Based Imagine Speech Using Time Domain Features," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp. 2921-2924.