SIDDAGANGA INSTITUTE OF TECHNOLOGY, TUMAKURU-572103

(An Autonomous Institute under Visvesvaraya Technological University, Belagavi)

Project Report on

"EEG-Based Speech Recognition"

Submitted in partial fulfillment of the requirement for the completion of VII semester of

BACHELOR OF ENGINEERING

in

ELECTRONICS & COMMUNICATION ENGINEERING

Submitted by

Aditya Keshav Harikantra	1SI21EC004
Harsha M	1SI21EC037
Rahul Jain S V	1SI21EC074
Rohith Ingaleshwar	1SI21EC078

under the guidance of

Dr.Veena Karjigi

Associate Professor

Department of E&CE

SIT, Tumakuru-03

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

2024-25

SIDDAGANGA INSTITUTE OF TECHNOLOGY, TUMAKURU-572103

(An Autonomous Institute under Visvesvaraya Technological University, Belagavi)

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

CERTIFICATE

Certified that the project work entitled "EEG-Based Speech Recognition" is a bonafide work carried out by Aditya Keshav Harikantra (1SI21EC004), Harsha M (1SI21EC037), Rahul Jain SV (1SI21EC074) and Rohith Ingaleshwar (1SI21EC078) in partial fulfillment for the completion of VII semester of Bachelor of Engineering in Electronics & Communication Engineering from Siddaganga Institute of Technology, an autonomous institute under Visvesvaraya Technological University, Belagavi during the academic year 2024-25. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The Project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering degree.

Dr.Veena Karjigi	Head of the Department	Principal
Associate Professor	Dept. of E&CE	SIT,Tumakuru-03
Dept. of E&CE	SIT,Tumakuru-03	
SIT,Tumakuru-03		

External viva:

Names of the Examiners **Signature with date**

1.

2.

ACKNOWLEDGEMENT

We offer pranams at the lotus feet of **His** Holiness, **Dr. Sree Sree Sivakumara Swamigalu**, Founder President and **His** Holiness, **Sree Sree Siddalinga Swamigalu**, President, Sree Siddaganga Education Society, Sree Siddaganga Math for bestowing upon their blessings.

We deem it as a privilege to thank late **Dr. M N Channabasappa**, Director, SIT, Tumakuru, **Dr. Shivakumaraiah**, CEO, SIT, Tumakuru, and **Dr. S V Dinesh**, Principal, SIT, Tumakuru for fostering an excellent academic environment in this institution, which made this endeavor fruitful.

We would like to express our sincere gratitude to **Dr. K V Suresh**, Professor and Head, Department of E&CE, SIT, Tumakuru for his encouragement and valuable suggestions.

We thank our guide **Dr.Veena Karjigi**, Associate Professor, Department of Electronics & Communication Engineering, SIT, Tumakuru for the valuable guidance, advice and encouragement.

Aditya Keshav Harikantra	(1SI21EC004)
Harsha M	(1SI21EC037)
Rahul Jain S V	(1SI21EC074)
Rohith Ingaleshwar	(1SI21EC078)

Course Outcomes

After successful completion of major project, graduates will be able to:

CO1: To identify a problem through literature survey and knowledge of contemporary engineering technology.

CO2: To consolidate the literature search to identify issues/gaps and formulate the engineering problem.

CO3: To prepare project schedule for the identified design methodology and engage in budget analysis, and share responsibility for every member in the team.

CO4: To provide sustainable engineering solution considering health, safety, legal, cultural issues and also demonstrate concern for environment.

CO5: To identify and apply the mathematical concepts, science concepts, engineering and management concepts necessary to implement the identified engineering problem.

CO6: To select the engineering tools/components required to implement the proposed solution for the identified engineering problem.

CO7: To analyze, design, and implement optimal design solution, interpret results of experiments and draw valid conclusion.

CO8: To demonstrate effective written communication through the project report, the one-page poster presentation, and preparation of the video about the project and the four page IEEE/Springer/paper format of the work.

CO9: To engage in effective oral communication through power point presentation and demonstration of the project work.

CO10: To demonstrate compliance to the prescribed standards/safety norms and abide by the norms of professional ethics.

CO11: To perform in the team, contribute to the team and mentor/lead the team.

CO-PO Mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2
CO-1											3		3
CO-2		3										3	
CO-3										3			3
CO-4						3							3
CO-5	3	3										3	
CO-6					3						3		3
CO-7			3	3								3	
CO-8										3			3
CO-9										3			3
CO-10								3					3
CO-11									3				3
Average	3	3	3	3	3	3	3	3	3	3	3		

Attainment level:- 1: Slight(low) 2: Moderate(medium) 3: Substantial(high)

POs: PO1: Engineering knowledge, PO2: Problem analysis, PO3: Design of solutions, PO4: Conduct investigations of complex problems, PO5: Engineering tool usage, PO6: Engineer and the world, PO7: Ethics, PO8: Individual and collaborative work, PO9: Communication, PO10: Project management and finance, PO11: Life-long learning.

Abstract

Brain-Computer Interface (BCI) technology has advanced significantly, offering innovative ways to interface with external devices. This project focuses on non-invasive Electroencephalogram (EEG)-based word recognition, where individuals mentally rehearse words without any physical expression. While invasive BCIs achieve high accuracy, their use is limited to severe cases due to the risks associated with surgical procedures. The project seeks to develop a more accessible, non-invasive approach to recognizing imagined speech, with potential applications in assistive technologies and human-computer interaction. By enabling thought-based communication, this research aims to improve quality of life and enhance independence for individuals with neuro-degenerative conditions. Furthermore, the potential for secure, silent communication in defense operations and innovative human-computer interaction underscores the broader significance of this technology.

The raw EEG signals collected during imagined speech undergo extensive pre-processing to isolate neural activity. Filtering techniques are employed to remove noise caused by eye blinks, muscle movements and electrical interference, as well as to isolate relevant frequency bands associated with cognitive processing. Additionally, Independent Component Analysis (ICA) is applied to identify and remove artifacts, ensuring the data's integrity. After pre-processing, key features related to brain activity and word processing are extracted. These features are then classified into their corresponding imagined words, demonstrating the system's ability to decode thought-based communication.

The classification process employed two separate models, Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks. The SVM model was trained and tested on data from all participants collectively, achieving an accuracy of 51% before applying the ICA and 69% after applying ICA-based pre-processing, highlighting the significance of artifact removal in enhancing performance. In the LSTM model, training and testing was conducted individually for each participant, utilizing the temporal dependencies in raw EEG data and achieving a maximum accuracy of 62.50% and a minimum accuracy of 8.33%.

Keywords: Electroencephalogram, Brain-Computer Interface, Independent Component Analysis, Support Vector Machine, Long Short-Term Memory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Communication plays a crucial role in everyday life, serving as the thread that weaves together daily interactions and experiences. A key technological goal is to create a connected environment where people can seamlessly link their physical actions to the virtual world, blending real-life activities with digital experiences. Technologies like speech-based tools, such as Google Voice Search and Siri, have become integral to daily routines, allowing users to communicate with electronic devices. These systems rely on speech recognition algorithms that convert spoken words into text.

However, there is a significant population for whom interacting with speech-based technologies is challenging due to health conditions such as paralysis. In such cases, despite being physically immobilized, if their cognitive abilities remain intact, harnessing the power of imagined speech recognition could become a transformative clinical tool. This would enable individuals with severe paralysis to communicate effectively and interact with their surroundings, offering them a lifeline. As a result, BCI are being developed to support direct communication for individuals with conditions like locked-in syndrome or paralysis.

The interest in imagined speech arises from the invention of EEG as a tool for synthetic telepathy. Synthetic telepathy refers to the transmission of thoughts or information directly from one brain to another or to a computer, without the need for verbal or physical communication. This concept relies on advanced technology like BCIs, neural implants and transmit it to another individual or system. Synthetic telepathy has potential applications not only in enhancing human-computer interaction but also in areas like military operations or covert communications.

While various forms of communication are used to exchange information, imagined speech holds a particularly important role. Imagined speech refers to the mental process of formulating words without physically engaging the articulatory system. Brain signals reflect how the brain handles tasks, regulates behavior and processes information from other parts of the body, such as internal organs or sense organs. If the brain signals of someone imagining speech can be decoded to identify the words they intend to speak, it could be a groundbreaking advancement in helping individuals with disabilities, such as those with locked-in syndrome, communicate effectively. Decoding imagined speech could significantly improve the quality of life for these individuals, enabling them to express their thoughts, needs, and emotions, participate in social interactions and engage in everyday activities.

Recent research on BCIs, which utilize brain waves to control machines, is focused on creating intuitive systems that leverage imagination. These systems could allow both patients and healthy individuals to communicate without the limitations of traditional communication methods like event-related potentials (ERPs) or motor imagery. While these paradigms have improved communication, they face limitations in terms of stimulus control and practicality for a wide range of communication scenarios. To implement BCI-based communication systems, both invasive and non-invasive methods are used to capture brain activity via electrodes. Invasive methods, which require surgery to implant electrodes into the brain tissue, are primarily used in clinical settings due to their high spatial and temporal resolution. These methods often involve techniques like electrocorticography (ECoG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI). However, non-invasive methods, particularly EEG are preferred because they do not require surgery, are less expensive and offer good temporal resolution, making them more broadly applicable.

The pre-processing of raw EEG data is a critical step in ensuring the quality and accuracy of BCI systems. Filtering techniques and ICA are fundamental to enhancing signal quality by removing artifacts and isolating relevant neural activity. A well-designed filter is essential to focus on the brain's electrical signals while rejecting noise. Typically, band-pass filters are employed to isolate key frequency bands, such as alpha (8-13 Hz), beta (13-30 Hz) and theta (4-8 Hz), which are linked to cognitive processes. These filters allow brainwave activity to pass through while attenuating noise from sources like eye blinks, muscle movements and electrical interference. ICA, on the other hand, is invaluable in separating mixed EEG signals into statistically independent components. This technique helps identify and isolate non-brain activity, such as eye blinks or muscle contractions, from actual brain signals. By decomposing the EEG data, ICA enables the removal of artifacts while preserving neural activity. This separation significantly enhances the signal-to-noise ratio, yielding cleaner and more reliable data for subsequent analysis.

For classifying and decoding the imagined speech from the preprocessed EEG signals, machine learning algorithms like SVM and LSTM networks are employed. SVM is used to classify the extracted features into their corresponding imagined words, taking advantage of its effectiveness in high-dimensional spaces. LSTM networks, which are designed to capture temporal dependencies, are used to handle the sequential nature of brain activity, enhancing the recognition of patterns over time. The combination of filtering, ICA, SVM, and LSTM enables a robust system for accurately decoding thought-based communication in non-invasive BCI applications.

1.1 Motivation

The development of EEG-based word recognition is at the forefront of cutting-edge research, representing a significant leap forward in the fields of neuroscience and human-computer interaction. This technology stands as a testament to the potential of harnessing brain activity to enable direct thought-based communication. For individuals with Amyotrophic lateral sclerosis(ALS) and other neuro-degenerative disorders, where traditional means of communication are compromised, EEG-based systems provide an extraordinary opportunity to restore their ability to interact with the world. By decoding brain signals into text or speech, these individuals can regain autonomy and express themselves in ways that were once unimaginable, vastly improving their quality of life.

Beyond healthcare, EEG-based word recognition has far-reaching applications in the defense sector. The ability to communicate covertly, without speaking or physically interacting with devices, offers new possibilities for secure communication during high-risk military operations. This method reduces the risk of interception and enhances operational efficiency in environments where traditional communication methods are limited.

Moreover, in the realm of human-computer interaction, this technology offers an innovative way to interface with digital devices. Whether for assistive purposes, like empowering those with physical disabilities or for everyday uses in gaming, virtual reality and immersive experiences, thought-based control systems have the potential to revolutionize how we interact with technology. By eliminating the need for physical input devices like keyboards or touchscreens, EEG-based systems open up new possibilities for efficiency, creativity and convenience.

As a cutting-edge area of research, this technology also holds promise for broader applications in education, entertainment and even the development of smart environments. Its potential to reshape industries and create more inclusive, intuitive ways of living and working makes EEG-based word recognition one of the most exciting frontiers of modern science.

1.2 Objective of the project

The key objectives of the project are:

- To design a filter for pre-processing raw EEG data.
- To implement effective artifact removal techniques.
- To extract features from processed EEG data and predict the imagined word.

1.3 Organisation of the report

The project report is divided into several chapters. The first Chapter discusses about the motivation and the main objectives for the development of the project. Second Chapter tells about the literature review of the papers related to brain-computer interfaces for imagined speech, artifact removal techniques and classification methods for imagined word recognition. The third chapter provides an overview of the system and details the dataset used in the project. The fourth chapter focuses on the preprocessing of raw EEG data, including analysis, splitting and filtering. The fifth chapter discusses artifact removal using ICA, explaining its working and the decomposition of EEG signals. The sixth chapter describes the classification of imagined words using SVM, comparing the model's performance with and without ICA. The seventh chapter explores imagined word recognition using LSTM networks. The discussion and conclusion is presented in Chapter eight.

Chapter 2

Literature Survey

This chapter includes the various literature survey carried out related to the project.

2.1 Brain-Computer Interfaces for Imagined Speech

Brain-Computer Interfaces (BCIs) for imagined speech offer a transformative communication solution for individuals unable to produce speech due to conditions like locked-in syndrome or practical limitations in noisy environments. By decoding neural signals associated with internally generated speech, these systems provide an intuitive alternative to audible communication, allowing users to envision words or phrases without sound. Research has evaluated various brain imaging techniques for recognizing speech from neural signals. While metabolic-based methods like Functional Near Infrared Spectroscopy (fNIRS) and Functional Magnetic Resonance Imaging (fMRI) are valuable for understanding speech-related neural mechanisms, their low temporal resolution limits their suitability for real-time applications. In contrast, electrophysiological signals, such as those captured by EEG and electrocorticography, offer the high temporal resolution necessary for effective Automatic Speech Recognition (ASR) in imagined speech BCIs [1].

Deep learning approaches have further pushed the boundaries of imagined speech decoding. Hierarchical deep neural networks have been employed to classify speech tokens and phonological categories by leveraging spatial and temporal features, achieving state-of-the-art accuracy in tasks such as binary phonological classification and token identification [2]. Additionally, integrating wavelet-based feature extraction with deep neural networks has improved recognition rates, highlighting the importance of combining time-frequency analysis with advanced classifiers [3].

Imagined speech has also been explored alongside visual imagery, expanding the scope of BCIs as scalable and intuitive paradigms for multi-class communication systems. These studies revealed that high-frequency EEG bands, along with cortical regions such as Broca's and Wernicke's areas, play a critical role in decoding performance, even as the number of communication classes increases [4]. Furthermore, multimodal datasets that integrate EEG with facial and audio modalities have shown remarkable accuracy in classifying phonological categories and distinguishing activity states like resting, stimulus response, and active thinking. These datasets enable the development of silent-speech BCIs, with reported classification accuracies exceeding 90% for consonants and 95% for activity states [5].

The collective advancements in decoding imagined speech highlight the growing potential of BCIs in bridging communication gaps, providing a voice to mute individuals, and enhancing accessibility in diverse real-world applications.

2.2 Artifact Removal Techniques in EEG Signal Processing

Preprocessing methods aim to enhance the signal-to-noise ratio of EEG recordings by Reducing external disturbances. Bandpass filters are frequently used to eliminate noise outside the typical EEG frequency range of 0.5 Hz to 60 Hz. Butterworth filters effectively reduce power line interference at 50/60 Hz, ensuring cleaner recordings. ICA is another widely adopted statistical technique that separates EEG signals into independent components, allowing for the identification and removal of artifacts such as eye blinks and muscle movements. Studies Utilizing spatial patterns from electrodes, like Brodmann areas, highlight ICA's effectiveness in isolating unwanted signals. Blind Source Separation (BSS) extends the capabilities of ICA by incorporating algorithms such as Principal Component Analysis (PCA) [6].

ICA a blind source separation technique, was applied to isolate and eliminate residual electrooculography (EOG) artifacts such as eye movements and muscle contractions. This process involved manually inspecting ICA components to identify and exclude artifact related signals before reconstructing the time series using artifact free components. Raw EEG data were bandpass filtered (0.5–100 Hz) to focus on significant frequency bands and baseline-corrected by subtracting the mean amplitude of the pre stimulus interval. Trials with extreme signal amplitudes exceeding ±100 µV were automatically excluded. These preprocessing steps ensured cleaner data for downstream analysis, enhancing the reliability of feature extraction and classification [7].

EEG artifact removal includes techniques such as regression-based methods, which utilize reference channels and subtract artifacts. Wavelet transform approaches, which decompose signals into multi resolution components for selective noise suppression. Ongoing Progress include automated machine learning models and hybrid methods combining traditional and modern approaches for enhanced artifact removal efficiency. These advancements continue to drive improvements in the preprocessing pipeline for EEG-based applications like brain-computer interfaces and cognitive research [8].

Artifact removal is an integral component of EEG signal preprocessing, focused on eliminating ocular and muscular artifacts using ICA in the EEGLAB toolbox. This approach decomposes the EEG data into independent components, which were manually inspected to identify and remove artifact related components before signal reconstruction. Signal amplitudes exceeding a threshold of ± 80 µV were also discarded to exclude extreme noise. The preprocessing pipeline also included bandpass filtering in the 0.5–57 Hz frequency range using a 5th-order Butterworth filter, targeting the frequency bands relevant to cognitive processes. Non-parametric cluster based permutation tests were utilized for feature validation, ensuring the extracted components were statistically significant and artifact free. This method accounted for spatial adjacency and controlled false positives in feature selection [9].

Removing artifacts is a crucial step in EEG signal processing to ensure clean and reliable data. Filtering methods, like bandpass filters (0.5–100 Hz), help focus on important brain activity while removing irrelevant frequencies and notch filters tackle power line noise. Spatial filters, such as common average referencing (CAR), reduce noise that's common across all electrodes. Other techniques like ICA, Wavelet transforms are another approach, breaking down signals into different levels to target and remove short-term artifacts. By combining these methods, researchers create effective preprocessing pipelines that prepare EEG data for accurate analysis and interpretation [10].

2.3 Classification Techniques in Imagined Word Recognition

Recent advancements in BCIs have explored various methods for imagined speech classification, aiming to enable non-invasive communication systems for individuals with motor impairments. One study investigated covert speech tasks using EEG signals and regularized multilayer perceptron (MLP) neural networks. Discrete wavelet transform (DWT) was used for feature extraction, achieving classification accuracies of 75.7% for imagined speech versus rest, 63.2% for imagined speech versus other tasks, and 54.1% for ternary classification. The MLP models outperformed traditional classifiers, demonstrating robustness and reliability across inter-session data [11].

Extending this line of research, the classification of unspoken speech in bilingual individuals was analyzed using artificial neural networks (ANN). EEG signals targeting Broca's and Wernicke's areas were processed, with features from alpha, beta, and gamma bands reduced using PCA. The ANN classifier achieved high accuracies in decision classification, language identification, and bilingual word classification, showcasing the potential of multilingual BCIs to address challenges such as cross-language interference [12].

Efforts have also been made to leverage low-cost EEG devices like the Emotiv Epoc for imagined speech recognition. Using MFCC for feature extraction and k-NN for classification, data from four participants yielded an average accuracy of 58%, with the cosine distance metric achieving 63%. While the performance was modest, the findings highlight the feasibility of affordable EEG devices for practical BCI applications, with future work focused on addressing subject variability and expanding vocabularies [13]. Further exploration of EEG signals for speech classification utilized the FEIS dataset to distinguish 16 English phonemes. Features such as amplitude spectrum and relative phase, extracted via Short-Time Fourier Transform (STFT), significantly enhanced classification performance. With SVM and neural networks, a 375 ms analysis segment produced optimal results, demonstrating the promise of EEG-based phoneme recognition while underscoring the need for larger datasets and improved models [14].

To improve feature extraction methods, combining DWT and maximum linear cross-correlation (MaxLCor) with an SVM classifier has been shown to enhance accuracy. This approach outperformed other classifiers like k-NN, Naive Bayes, and LDA by effectively handling high-dimensional data. Despite its success, further optimization is needed to make the system suitable for real-world applications [15].

The integration of spectral methods and deep learning has advanced imagined speech recognition. A study involving 15 subjects imagining 15 words processed EEG data into time-frequency representations (TFR) using smoothed pseudo-Wigner–Ville distribution (SPWVD). Classification using convolutional neural networks (CNNs) highlighted the gamma band as particularly effective, reinforcing its importance in non-invasive communication systems. However, the study emphasizes the necessity of diverse datasets for broader applicability [16].

2.4 Summary of Literature Survey

The reviewed studies emphasize advancements in EEG-based imagined speech recognition through various approaches. Preprocessing techniques, including bandpass filtering, ICA using EEGLAB, and wavelet transforms, are crucial for enhancing signal quality by removing noise and artifacts like eye blinks and muscle movements. Classification methods utilize machine learning and deep learning algorithms such as SVM, MLP, and CNNs, integrated with feature extraction techniques like wavelets and PCA to decode imagined speech effectively. From this literature survey, bandpass filters, ICA using EEGLAB, and SVM are taken into consideration, as they demonstrate robust performance and are well suited for EEG-based imagined speech recognition tasks.

Chapter 3

System Overview

EEG-Based Speech Recognition employs BCI technology to decode imagined speech from EEG signals, analyzing patterns in the KaraOne dataset to identify imagined words.

3.1 Imagined Speech Recognition

EEG-Based Speech Recognition involves the discrimination between a fixed set of imagined words from the EEG captured during covert speech. Fig 3.1 illustrates the key steps involved in the Imagined Speech Recognition (ISR) system, starting with the collection of EEG data from the KaraOne dataset. The process includes visualizing and segmenting the data, applying filtering and artifact removal techniques, and extracting relevant features. These steps prepare the data for input into machine learning models which are used to predict the imagined speech. The flowchart provides a clear visual representation of how data flows through the system, from initial processing to final prediction.

Figure 3.1: Block diagram of Imagined Speech Recognition

The workflow for Imagined Speech Recognition using the KaraOne dataset involves several key steps to preprocess and analyze EEG data. Initially, data is sourced from the KaraOne dataset, which contains EEG recordings of imagined speech tasks. After loading the dataset, the data is visualized to ensure proper loading and understanding. The continuous EEG data is then segmented into epochs, corresponding to specific thinking or imagination tasks, with irrelevant segments excluded. For noise reduction, low-pass and high-pass filters are applied to retain only the frequency band of interest (2–45 Hz), removing high-frequency noise and baseline drift. Artifacts such as muscle activity and eye blinks can be isolated and removed using appropriate signal processing techniques. The next step involves feature extraction, where power spectral density and connectivity metrics are computed from the cleaned EEG data. The features extracted from the EEG data are utilized to train and test models that capture signal patterns, while the raw EEG data is used to develop models focusing on temporal dependencies. These models predict the imagined word based on the EEG data, providing an output corresponding to the subject's task.

3.2 Dataset Description

The KARA ONE Database is used for the research work, integrating three modalities (EEG, face tracking and audio) across imagined and spoken phonemic and single-word prompts. The EEG data was utilized for the requirement in this work. Fourteen participants were involved in the dataset collection process. A 64-channel Neuroscan Quick-cap was used to collect EEG signals, following the 10- 20 electrode placement system in which the electrode locations as depicted in the Fig 3.2.

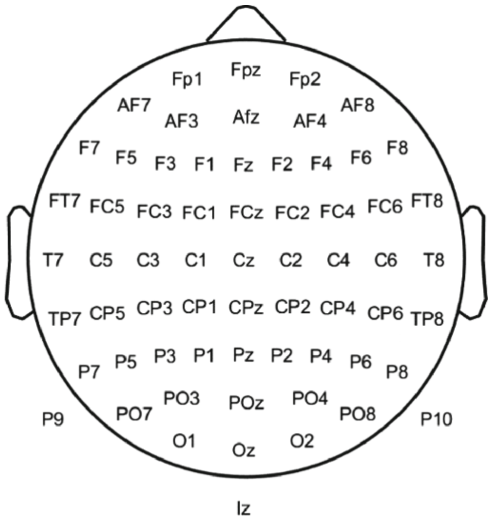


Figure 3.2: Electrode placement locations [17]

The EEG signal consists of continuous repetition of 7 phonemic/syllabic prompts (/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) and 4 words derived from Kent's list of phonetically similar pairs (pat, pot, knew, and gnaw). Database comprising of EEG data corresponding to 11 prompts. The experimental procedure taken for each prompt is comprised of four main states as shown in Fig 3.3. Firstly, a 5-second rest state directed the participant to relax and clear their mind. Following this, a stimulus state commenced, displaying prompt text on the screen accompanied by its auditory counterpart played through computer speakers. Subsequently, a 2-second period allowed participants to position their articulators in preparation for pronouncing the prompt. This was succeeded by a 5 sec-

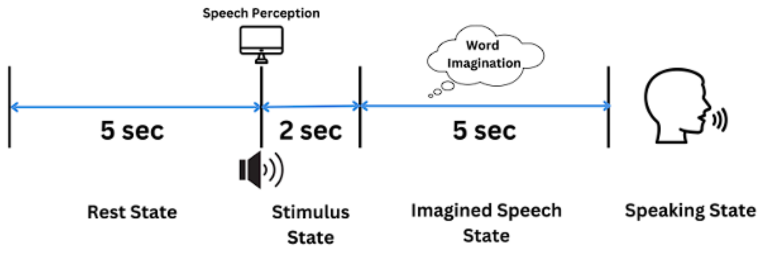


Figure 3.3: Schematic sequence of the experimental paradigm

ond imagined speech state, during which participants mentally articulated the prompt without physical movement. Finally, the speaking state involved participants audibly voicing the prompt while the kinect sensor captured both audio and facial features. The Database consists of 14 participants data. Each 11 prompts was repeated 12 times for 12 participants and 15 times for 2 participants, amounting to 132 trials and 165 trials in total for the participant who repeats the prompt 12 times and 15 times respectively. Initially, the phonemic/syllabic prompts were displayed, followed by the presentation of the 4 'Kent' words. Within each of these sections, the trials were randomly rearranged. The sampling rate of EEG signal is 1000 Hz. EEG signals were segmented into 132 epochs and 165 epochs for 12 and 15 times repetition of words respectively, which represents significant part of a continuous EEG signal.

Chapter 4

Pre-processing of Raw EEG

EEG signals are inherently sensitive to noise arising from both internal physiological processes and external environmental factors. This makes meticulous pre-processing an essential step for ensuring reliable analysis. The pre-processing workflow in this study is based on the Harvard HAPPE (Harvard Automated Processing Pipeline for EEG) methodology, which has been specifically tailored to the KARA One dataset to account for its unique characteristics and experimental setup.

4.1 Raw EEG Data Analysis

The raw EEG data contains the unprocessed electrical activity of the brain, but as shown in Fig 4.1 it is overwhelmed by several types of noise. The most prominent issue is low-frequency drift and high-frequency noise. Low- frequency drift typically caused by slow physiological processes like skin potentials or eye movements. These drifts can significantly distort the signal, particularly at very low frequencies. High-frequency noise, often originating from muscle activity or electrical interference from surrounding equipment.

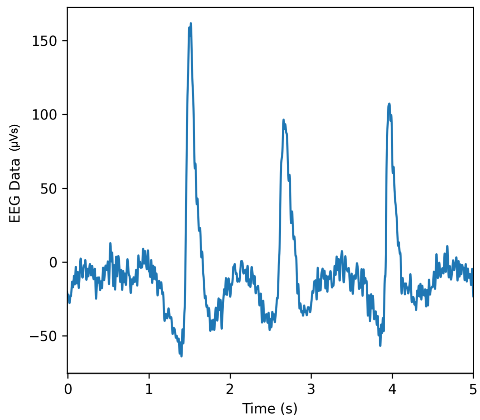


Figure 4.1: Participant MM05 - Trial 124: Raw Data Plot

When analyzed through its power spectral density (PSD), the raw EEG signal further reveals the extent of these issues. The low-frequency range (2 Hz) shows considerable power, reflecting baseline shifts, while the high-frequency range (45 Hz) indicates noise contributions from muscle activity and power-line interference. The power distribution in the PSD plot highlights how the neural signal is obscured by these unwanted components, making it challenging to identify key brainwave patterns, such as alpha, beta or theta waves. Therefore, raw EEG data, while rich in potential information, is unsuitable for further analysis without pre-processing.

4.2 Data Splitting

The data splitting process plays a crucial role in structuring EEG signals into distinct phases to facilitate effective analysis. The recorded EEG data is segmented based on specific time intervals representing different cognitive states. The process begins with a 5-second rest state, where participants remain idle to establish a baseline of brain activity. This is followed by a 2-second stimulus state, during which participants are presented with an auditory or visual cue to perceive a word. Subsequently, the 5-second imagined speech state is introduced, during which participants internally simulate or think about the given word without any physical expression. Finally, there is a 3-second speaking state, where participants articulate the imagined word aloud.

The focus is on extracting EEG signals specifically from the imagined speech state. Signals from this phase are carefully divided into smaller epochs corresponding to individual trials. Each trial contains neural activity associated with internally processing the prompted word. The segmentation process ensures that only the data corresponding to this state is retained for further pre-processing and analysis, while data from other states is excluded to maintain relevance and reduce noise. By isolating the imagined speech state, the process ensures a cleaner and more targeted dataset, paving the way for accurate decoding and classification in subsequent stages.

The data from multiple participants is systematically organized into folders, each containing their respective EEG recordings and related metadata. The raw EEG recordings are loaded and divided into segments based on predefined time intervals, stored separately for each participant. These intervals specify the start and end points for the rest and imagined speech phases, ensuring that the segmented data aligns accurately with the experimental conditions. For each participant, the data is separated into two main categories: one corresponding to brain activity during the imagined speech phase and the other representing baseline activity during the rest phase.

In addition to segmentation, metadata such as the specific prompts provided to participants during the experiment and the electrode configurations used in data collection are retrieved and organized. This metadata is crucial for understanding the context of the segmented signals and ensuring compatibility for subsequent analysis. The data is sampled at a rate of 1024Hz, providing high temporal resolution for precise signal analysis. Once the data is segmented and enriched with metadata, it is saved in a structured format for each participant, ensuring that all relevant information is preserved for future processing.

4.3 Filtering

In the pre-processing stage of EEG signal analysis, the design and implementation of filters are critical for isolating relevant neural activity while eliminating noise and artifacts. The filtering process in this project utilized Finite Impulse Response (FIR) filters, which were specifically designed to retain frequency components within the range of 2 Hz to 45 Hz. These FIR filters were implemented using the windowed time-domain design method, ensuring a stable and precise filtering process with minimal distortion of the signal as shown in Fig 4.2.

The low-pass filter was designed to eliminate high-frequency noise above 45 Hz, including muscle artifacts and electrical interference. It employs a Hamming window with a passband ripple of 0.0194 and stopband attenuation of 53 dB, ensuring smooth attenuation of frequencies beyond 45 Hz while preserving cognitive brainwave bands such as theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). The high-pass filter removes low-frequency components below 2 Hz, associated with baseline drifts due to physiological processes like skin potentials or eye movements. It also uses a Hamming window with the same specifications to effectively attenuate slow drifts without distorting higher-frequency EEG signals. The specifications for both filters are provided in Table 4.1.

Both the high-pass and low-pass filters were designed as one-pass, zero-phase, and non-causal filters. The zero-phase design ensures that the filtering process does not introduce phase distortion, which is critical for maintaining the temporal structure of the EEG signal. Additionally, the one-pass design applies the filter only once, further preserving the signal's integrity. The removal of frequencies below 2 Hz and above 45 Hz during EEG pre-processing is a vital step in ensuring the accuracy and reliability of neural data

Parameter	Low-Pass Filter	High-Pass Filter
Frequency Range	$\leq$ 45 Hz	$\geq$ 2 Hz
Window Type	Hamming	Hamming
Passband Ripple	0.0194	0.0194
Stopband Attenuation	53 dB	53 dB
Cutoff Frequency (-6 dB)	50.62 Hz	1 Hz
Transition Bandwidth	11.25 Hz	2 Hz
Filter Length	295 samples	1651 samples
Filter Order	294	1650
Filter Duration	0.295 s	1.651 s
Purpose	Removes high-frequency noise	Removes low-frequency drifts

Table 4.1: Specifications of the Low-Pass and High-Pass Filters

analysis as shown in Fig 4.3. Frequencies below 2 Hz primarily consist of low-frequency drifts and other non-neural artifacts . These include baseline drifts, which result from slow changes in the electrical properties of the skin or variations in electrode impedance, and slow eye movement artifacts caused by blinking or gradual ocular movements. Such components are irrelevant to cognitive processes and can dominate the lower end of the power spectrum, masking the neural signals of interest. By applying a high-pass filter with a cutoff frequency of 2 Hz, these slow, non-neural components are attenuated, allowing for a cleaner representation of the relevant brain activity as shown in Fig 4.4.

The Power Spectral Density analysis of EEG data provides critical insights into the distribution of power across different frequencies, helping to distinguish meaningful neural activity from noise. The PSD of raw EEG data and filtered EEG data were analyzed to evaluate the impact of filtering on signal clarity and interpretability.

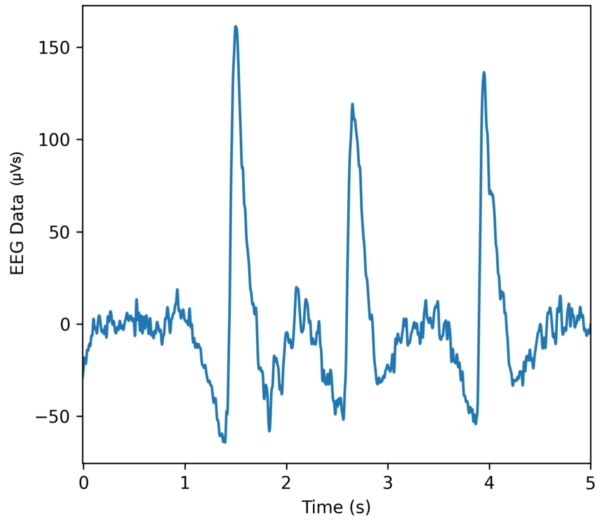
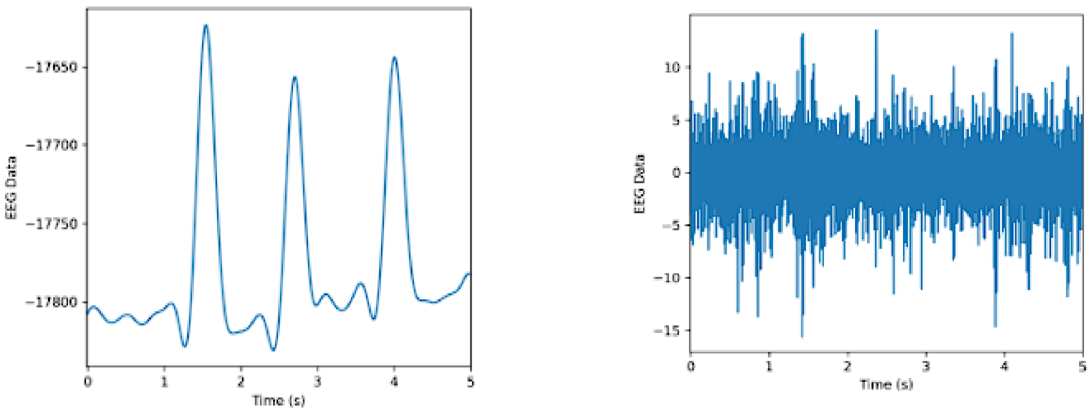


Figure 4.2: Processed EEG Plot After Filtering



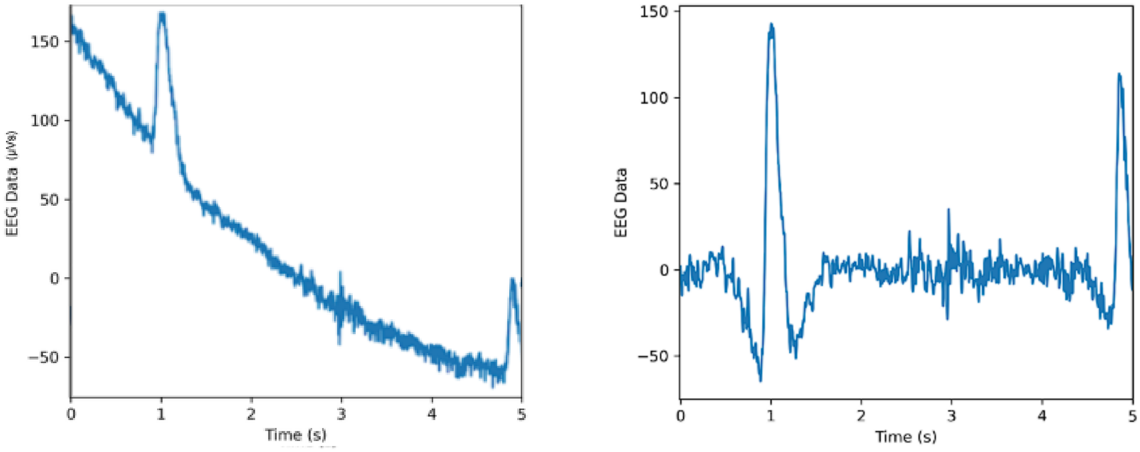
(a) <2Hz Frequency

(b) >45Hz Frequency

Figure 4.3: Filtered Plot: For Participant MM05 - Trial 124

The PSD of the raw EEG data exhibits a significant concentration of power at low frequencies, particularly below 2 Hz. This low-frequency power is typically caused by baseline drifts arising from slow physiological processes such as skin potentials, respiration, and slow eye movements. Additionally, the raw EEG spectrum shows elevated power at higher frequencies, which reflects noise contributions from muscle activity and electrical interference, including power line artifacts. The broad and diffuse nature of the raw PSD lacks distinct peaks in frequency ranges associated with cognitive brain activity, making it challenging to directly interpret the signal as shown in Fig 4.5.

After applying the filter, the PSD of the filtered EEG data demonstrates a significant



(a) Slow drift (b) Filtered plot

Figure 4.4: Signal drift below 2 Hz before and after filtering.

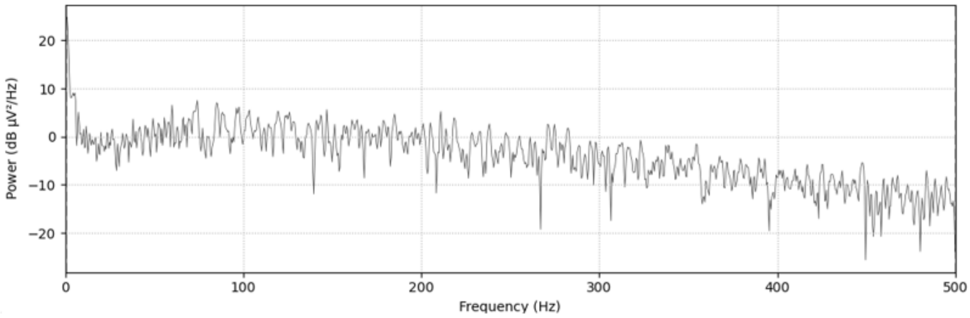


Figure 4.5: PSD before filtering

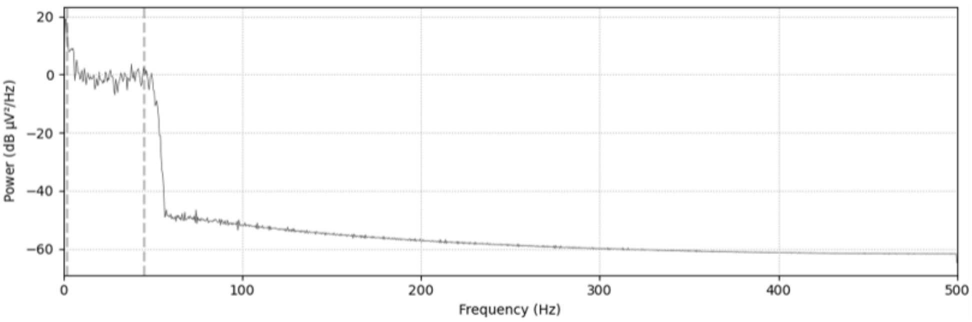


Figure 4.6: PSD after filtering

improvement in signal clarity. Low-frequency drifts and high-frequency noise have been effectively attenuated, resulting in a cleaner and more focused power distribution. In contrast to the raw data, the filtered PSD reveals distinct peaks within the theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands as shown in Fig 4.6.

The comparison between raw and filtered PSDs highlights the importance of pre-processing in EEG analysis. The raw EEG PSD is dominated by noise and irrelevant components, masking the neural activity of interest. However, filtered PSD effectively isolates key cognitive brainwave bands, ensuring that the neural signal can be accurately interpreted and analyzed. The power peaks within these bands clearly reflect the underlying brain activity, free from interference caused by physiological or environmental noise.

Chapter 5

Artifact Removal Using Independent Component Analysis

Independent Component Analysis (ICA) is a computational technique used to separate a multivariate signal into statistically independent components. It is widely applied in fields like neuroscience, telecommunications, and biomedical engineering, particularly for analyzing complex signals such as EEG data. The primary goal of ICA is to identify and separate mixed signals into their original, independent sources, which are often hidden in the observed data. In EEG signal analysis, ICA is particularly valuable for isolating neural activity from non-neural artifacts such as eye blinks, muscle movements, and electrical interference. These artifacts are often mixed with brain signals, making it challenging to accurately analyze the neural activity of interest.

5.1 The working of ICA

ICA assumes that the EEG data is a linear combination of independent source signals, each corresponding to distinct neural or non-neural activities. It further relies on the assumption that the source signals are statistically independent and non-Gaussian in nature, which is typically valid for EEG signals. Preprocessing of EEG data is a crucial first step in preparing it for ICA decomposition. This involves removing the mean of the signals and normalizing their variance, ensuring that the data is centered and whitened. Such preprocessing is essential for ICA algorithms to perform effectively. The ICA algorithm then decomposes the mixed signals into independent components by maximizing their statistical independence, utilizing methods like FastICA or RunIca that optimize a contrast function to measure independence. After decomposition, the resulting components are visually inspected using topographical maps and temporal plots. Components that represent artifacts such as eye blinks, muscle movements, and electrical noise are identified based on their spatial and temporal characteristics. These artifact components are then removed, and the remaining components are recombined to reconstruct the cleaned

EEG signals, which effectively enhances the signal quality by retaining only the neural signals. FastICA and runICA are two widely used ICA algorithms in EEG data processing, each with distinct strengths and implementation techniques. runICA, which is often used in EEG analysis through the EEGLAB toolbox, is designed to effectively separate independent sources by maximizing the non-Gaussianity of the components. It employs a symmetric approach, meaning that it does not rely on an initial guess of the components and adjusts iteratively to converge to a solution that isolates independent sources from the observed mixed signals. This iterative process uses gradient descent to optimize the transformation matrix until the separation reaches its maximum potential. The symmetric approach in runICA makes it particularly stable for handling EEG data, where noise and artifacts are common, providing robustness in scenarios where there are significant fluctuations or artifacts.

FastICA is a more general-purpose ICA algorithm that is widely recognized for its computational efficiency and speed, making it suitable for processing large EEG datasets in real-time applications. FastICA works by maximizing non-Gaussianity using higher-order statistics, such as kurtosis or negentropy, to distinguish independent components from the mixed signals. Before applying the fixed-point iteration process, FastICA centers and whitens the data to remove dependencies between the signals and ensure they are decorrelated. The algorithm then iteratively adjusts the separation matrix using a fixed-point iteration to maximize the contrast function and extract the independent components. This process makes FastICA faster than runICA, particularly for large datasets, as it requires fewer computational resources and time to converge.

While both algorithms are effective in separating brain activity from artifacts, runICA tends to perform better in complex or noisy environments due to its more optimization technique and the stability of the symmetric approach. However, FastICA's speed makes it ideal for applications that require quick processing, such as real-time EEG monitoring, brain-computer interfaces or large-scale studies where computational efficiency is critical. The decision between runICA and FastICA depends on the specific requirements of the EEG analysis, including dataset size, the level of noise and the need for real-time processing. Both methods significantly enhance EEG signal quality by removing unwanted artifacts, enabling more accurate analysis of the brain's neural activity.

5.2 Decomposition of EEG signals using ICA

The ICA performed on EEG data, decomposed the signal into components. Each represented by a scalp topography illustrating the spatial distribution of activity. These topographies in the Fig 5.1 use a heatmap color scheme to depict the amplitude and polarity of each component's activity across the scalp. The colors in the maps are as follows: red represents regions of maximum positive amplitude, indicating areas of strong activation, while blue indicates regions of maximum negative amplitude, corresponding to deactivation or inverse polarity. Yellow or orange marks transitional zones between positive amplitudes and the baseline, and green represents neutral or baseline activity. The smoothness and distribution of these colors provide insights into the spatial characteristics of each component.

Beneath each topography, classification labels such as "Muscle", "Brain", "Eye" or "Other" identify the likely origin of the component. These labels, combined with the accompanying percentage values, indicate the strength of association with the labeled source. The percentage values below each map indicate the variance explained by the component. High percentages for brain-related components suggest a strong neural signal, whereas components with high muscle or eye contributions typically represent artifacts. For example, components labeled "Brain" with contributions exceeding 90% (e.g., "Brain: 96.9%") likely correspond to meaningful neural activity, while those with high muscle-related contributions (e.g., "Muscle: 98.6%") are indicative of noise from muscle artifacts.

This classification and visualization aid in identifying components to retain or exclude during pre-processing. Brain-related components with smooth, centralized patterns and gradual color transitions are typically retained for further analysis, while components dominated by sharp red/blue contrasts from muscle or eye artifacts are excluded. Components categorized as "Other" with undefined spatial characteristics require additional scrutiny to determine their relevance or removal.

ICA algorithm uses the power spectrum to classify independent components by identifying the dominant frequency ranges and their associated activity. Peaks in specific frequency bands help determine the probabilities of labels such as "Brain", "Muscle", "Eye" or "Line Noise".For Brain Activity, the power spectrum typically shows peaks in the 10 Hz range (Alpha band) and moderate activity around 20 Hz (Beta band) as mentioned in

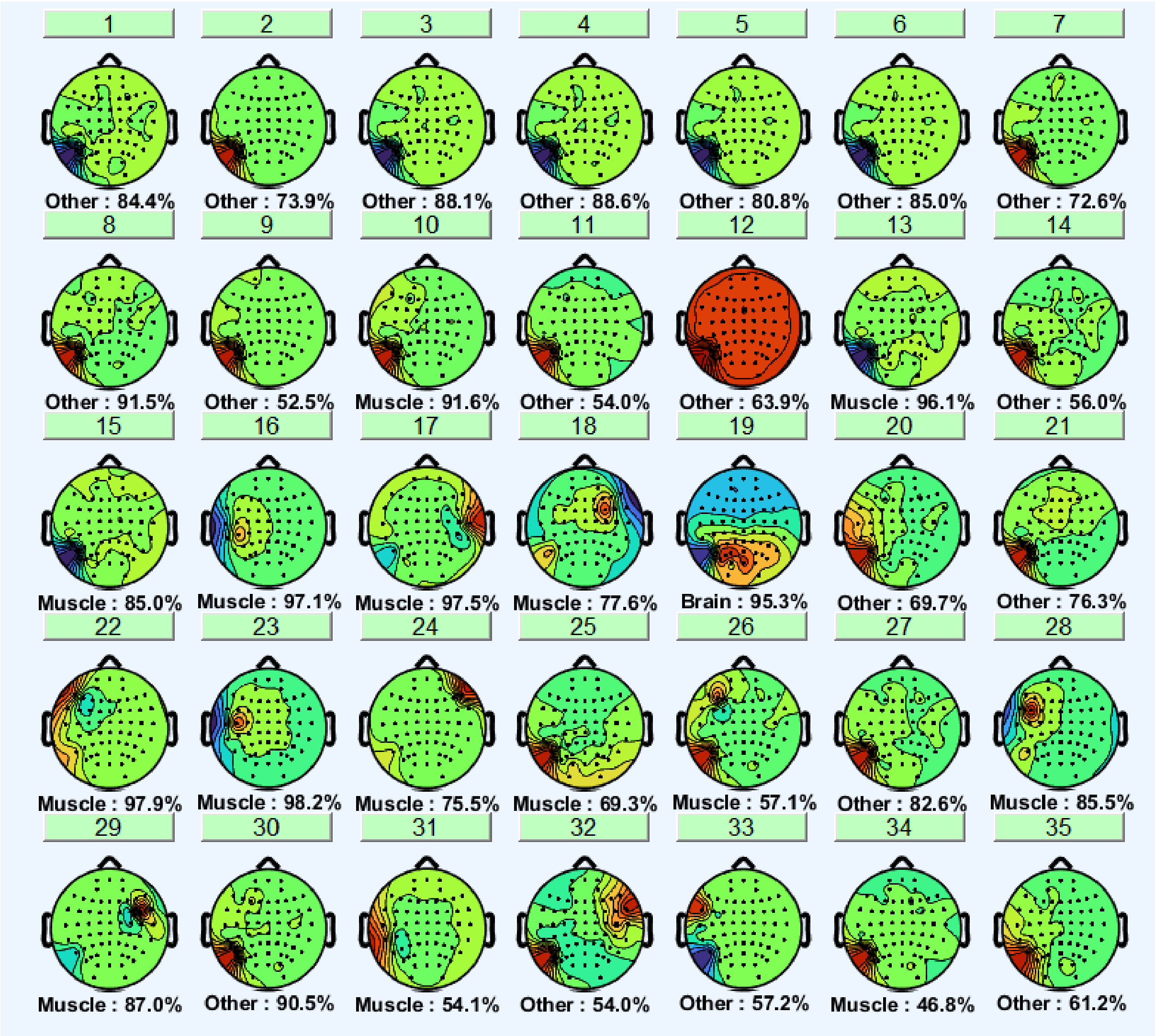


Figure 5.1: ICA Component Topographies

Table 5.1. Brain-related components usually exhibit very low power in high frequencies, as these frequencies are less representative of neural activity. In contrast, Muscle Activity becomes significant starting at 20 Hz and reaches its peak in the 30–50 Hz range, reflecting the high-frequency nature of muscle artifacts. For Eye Activity, the power spectrum is dominated by low frequencies, representing artifacts caused by eye movements or blinks. These components show minimal power in higher frequencies, distinguishing them from other artifacts. Line Noise appears as a narrow peak at >50 Hz, caused by electrical interference. This peak is distinct and can be easily identified in the power spectrum due to its sharp and isolated nature.

Frequency Range (Hz)	Source	Description
0–4	Eye/Heart (Artifact)	Associated with eye movements (slow blinks) or cardiac signals.
4–8	Brain (Theta Waves)	Theta waves, typically related to drowsiness, meditation.
8–12	Brain (Alpha Waves)	Alpha waves, prominent in relaxed, awake states.
12–30	Brain (Beta Waves) / Muscle	Beta waves are associated with active thinking, focus. Muscle artifacts also contribute in this range.
30–50	Muscle (Artifact)	Muscle activity (EMG), such as jaw clenching or head movements, typically appears in this range.
>50	Line Noise / Artifacts	Electrical noise from the recording equipment or external sources.

Table 5.1: Characteristics of Frequency Bands in EEG Data

5.3 Individual component analysis

By analyzing the frequency ranges of the power spectrum, the ICA algorithm determines the likelihood of each label. In the Fig 5.2 illustrates a detailed analysis of a single Independent Component, presenting multiple visualizations to evaluate its characteristics.The scalp topography shows a characteristic dipolar pattern, reflecting the spatial distribution of the component's activity across the scalp. The scrolling IC activity waveform provides a time-domain representation, showcasing rhythmic oscillations typical of neural signals. The continuous data heatmap visualizes the component's activity over time across various scalp channels.

5.3.1 Brain Component

The Independent Component IC19 is associated with brain-related activity, as identified by the ICLabel classification, which assigns 95.3% of its probability to the "Brain" category. The topographic map displays a dipolar pattern consistent with neural activity, with strong amplitude variations in specific regions, indicating the component's neural origin. The scrolling IC activity waveform shows rhythmic oscillations within a range of approximately ±4 µV, characteristic of neural signals. The continuous data heatmap reveals elevated RMS values across scalp channels. The power spectrum exhibits a prominent peak around 10 Hz in the alpha band, with power rapidly decreasing at higher frequencies and minimal activity beyond 20 Hz. This frequency-domain profile aligns with the characteristics of neural components, confirming IC19 as a brain-related component suitable for further analysis.

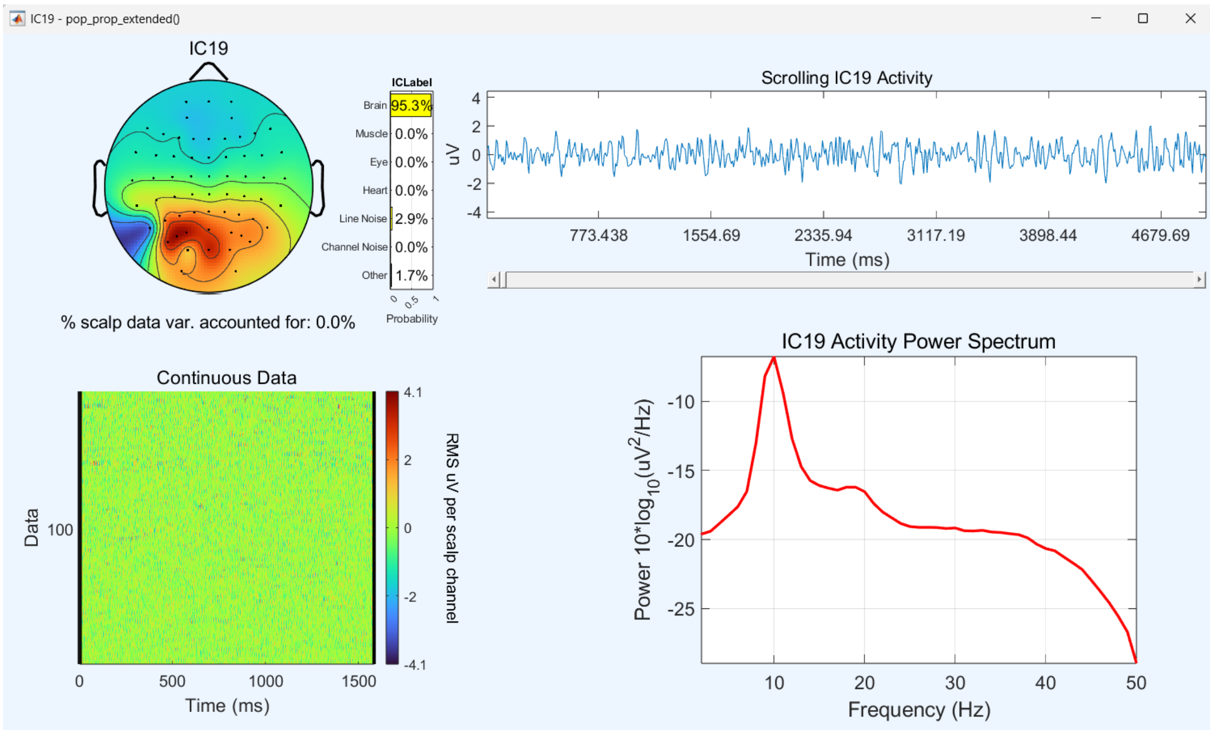


Figure 5.2: Detailed Analysis of ICA brain Component

5.3.2 Muscle Component

The Independent component IC15 primarily associated with muscle activity, as indicated by the ICLabel classification, which assigns 85% of its probability to muscle-related contributions. The topographic map in Fig 5.3 shows concentrated activity in the temporal region, suggesting that this component likely stems from muscle artifacts, such as facial or jaw movements. The scrolling IC activity waveform reveals high-amplitude fluctuations, characteristic of muscle activity. The power spectrum displays peaks in the higher frequency range (above 20 Hz), a typical feature of muscle artifacts. The component IC15 represents a significant source of muscle noise and should be removed to enhance the quality of the EEG signal.

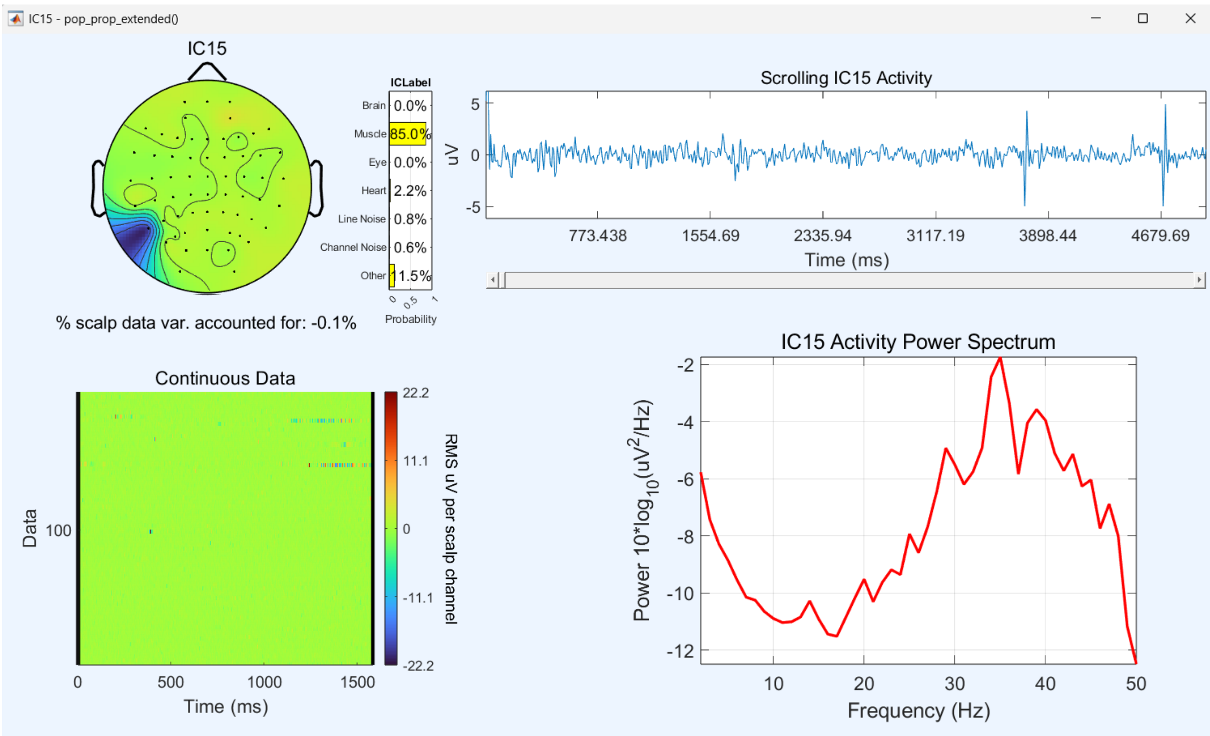


Figure 5.3: Detailed Analysis of ICA muscle Component

5.3.3 Eye Component

The Independent component IC41 associated with eye-related activity, as identified by the ICLabel classification, which assigns 76% of its probability to the "Eye" category. The topographic map in Fig 5.4 displays a pattern consistent with ocular artifacts, with strong activity in the frontal regions, likely reflecting blink or eye movement artifacts. The scrolling IC activity waveform shows relatively low-frequency fluctuations, typical of eye-related activity. The continuous data heatmap highlights regions of elevated RMS values in frontal channels, supporting the classification. The power spectrum exhibits prominent low-frequency peaks, further corroborating the presence of ocular artifacts. IC41 is identified as an eye-related component and should be removed during pre-processing to reduce contamination in the EEG data.

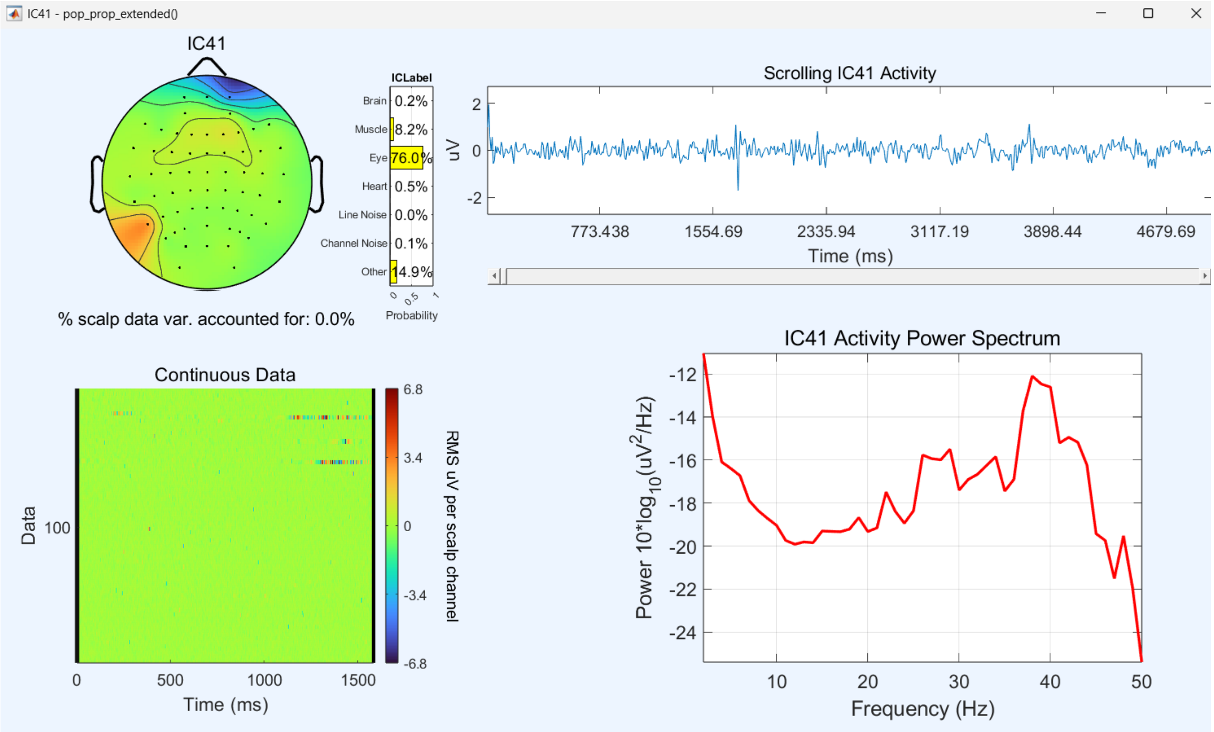


Figure 5.4: Detailed Analysis of ICA Eye Component

5.3.4 Noise Component

The Independent component IC58, in Fig 5.5 dominated by non-specific noise, with 85.5% of its probability assigned to the "Other" category by ICLabel. The topographic map shows diffuse activity across the scalp, lacking distinct patterns, while the scrolling IC activity waveform displays low-amplitude signals without notable features. The continuous data heatmap reveals uniformly low RMS values across channels, and the power spectrum lacks prominent peaks. These findings confirm IC58 as random noise, which can be excluded during pre-processing.

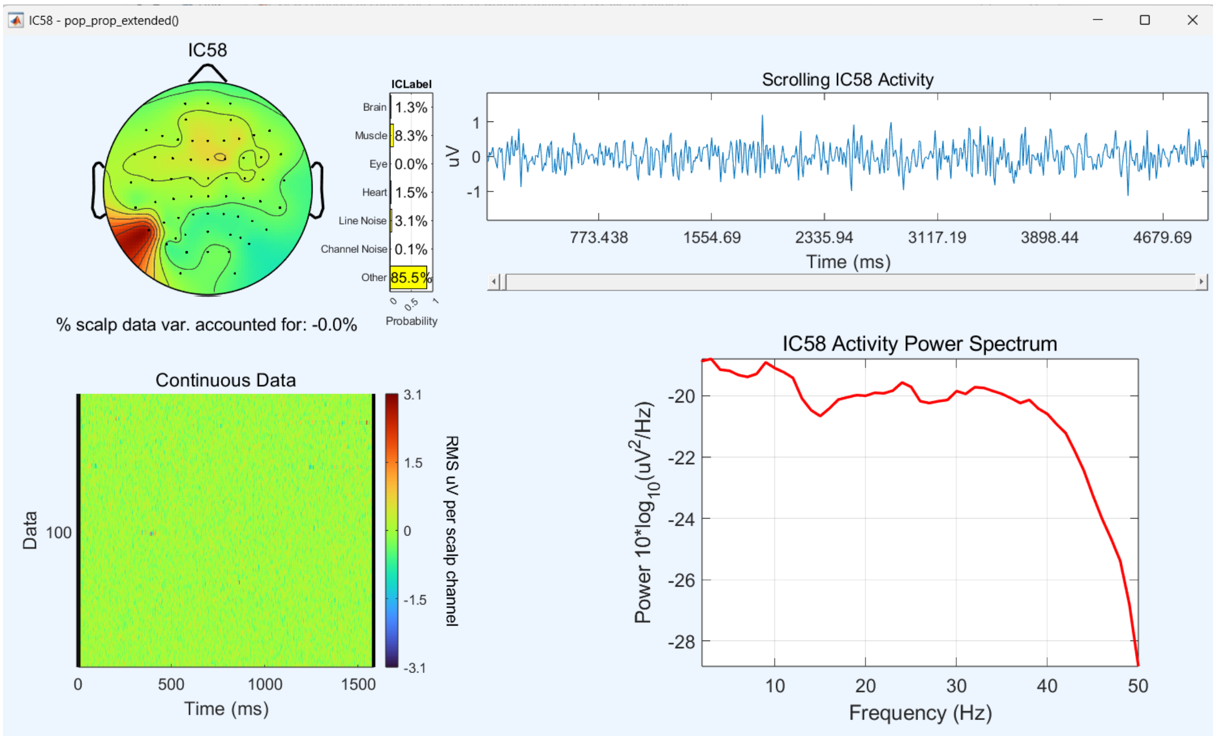


Figure 5.5: Detailed Analysis of ICA noise Component

5.4 Reconstruction of EEG signals

The ICA algorithm combines spectral features, spatial scalp distributions, and temporal dynamics to accurately classify components, effectively distinguishing brain activity from non-neural artifacts such as muscle movements, eye blinks, and electrical noise in EEG data. Identified artifact components are removed, and the remaining components are recombined to reconstruct the cleaned EEG signals. Fig 5.6 represents raw EEG data, which includes noise and artifacts such as eye blinks, muscle movements, or environmental interference. Fig 5.7 shows ICA-processed and reconstructed EEG data. The x-axis represents time, and the y-axis corresponds to EEG channels. ICA has been applied to isolate and remove these artifacts. As a result, the processed data appears smoother, with reduced noise and fewer large spikes compared to the raw data. The ICA processing enhances the clarity of neural signals, making the patterns more consistent and less distorted.

This improvement highlights the effectiveness of pre-processing steps, such as ICA, in enhancing data quality for further analysis and classification. The cleaned and artifact-free EEG data is then used in the subsequent phase, where it is provided as input to a

SVM model. In this phase, features such as spectral and spatial properties are extracted from the data to prepare it for classification.

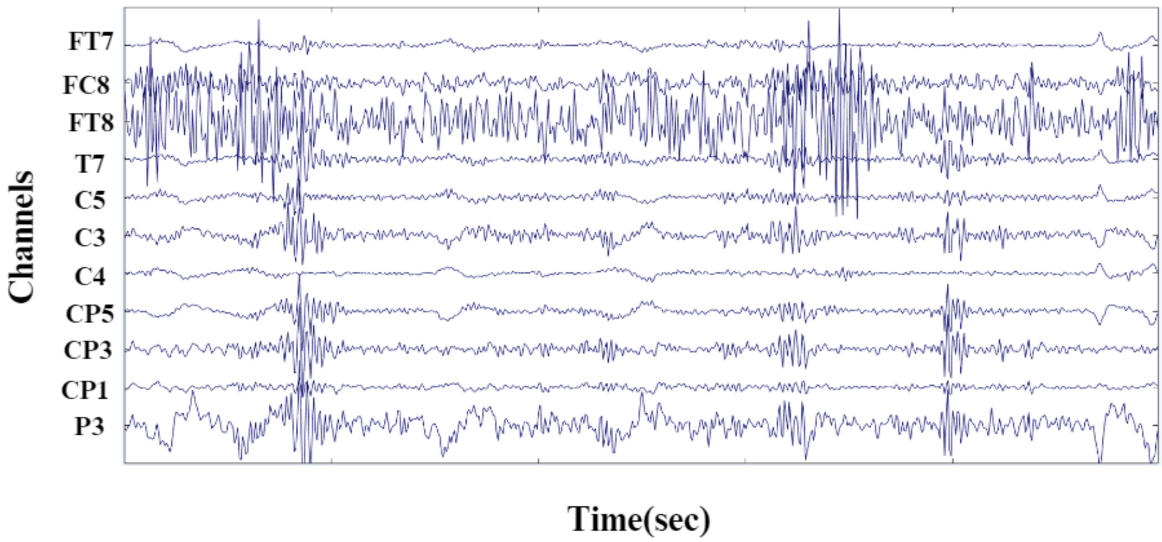


Figure 5.6: EEG Signals Before Rejection and Reconstruction

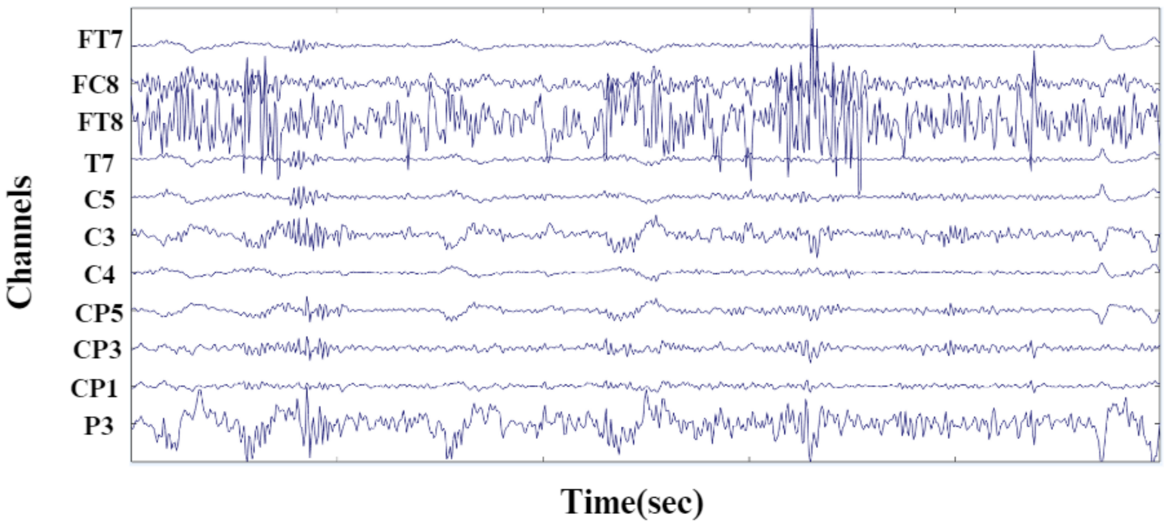


Figure 5.7: EEG Signals After Rejection and Reconstruction

Chapter 6

Classification of Imagined Words Using Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification tasks, particularly when the dataset is high-dimensional or involves complex boundaries. The primary goal of SVM is to find an optimal hyperplane that separates data points of different classes with maximum margin. For non-linearly separable data, SVM employs kernel functions, such as linear, polynomial, and radial basis function (RBF) kernels, to project the data into a higher-dimensional space where separation is feasible. The classification process using SVM for EEG signals involves preprocessing the raw EEG signals using techniques like ICA to remove artifacts, or directly using unprocessed signals (No-ICA). Features such as Mean, Energy, Curve Length and High-Frequency Energy (EHF) etc. are extracted from the signals to serve as inputs for the classifier. The SVM classifier is then trained using labeled data, optimizing the hyperplane for separation. Finally, the trained model is evaluated using metrics like accuracy, precision, recall, and F1-score to assess its performance.

6.1 Training and Testing Data for the SVM Model

The dataset for training and testing the SVM model consists of EEG recordings of imagined speech, with a total of 696 trials, of which 209 trials are used for testing and the remaining 487 trials are used for training to ensure proper validation. Preprocessing involves applying a bandpass filter between 2 Hz and 45 Hz to eliminate low-frequency drifts and high-frequency noise, preserving the frequency range associated with neural activity. The analysis is conducted using two approaches: one without ICA and another with ICA. In the first approach, features such as mean, energy, and curve length are directly extracted from the raw EEG signals without applying artifact removal. In contrast, the second approach applies ICA to remove artifacts such as eye blinks and muscle movements, resulting in cleaner EEG data for feature extraction. The SVM model is trained using a Radial Basis Function (RBF) kernel, with hyperparameters tuned to optimize performance. Both approaches are evaluated using metrics such as precision, recall, and F1-score, with results including confusion matrices and other visual outputs analyzed in detail. This comprehensive process ensures the SVM model is effectively optimized for classifying imagined speech, while the comparison between ICA and non-ICA approaches offers valuable insights into the impact of preprocessing techniques on model performance.

6.2 Features Used in the Classification

The features extracted from EEG signals for SVM classification are designed to capture a comprehensive representation of the signal's characteristics. Each feature provides specific insights into different aspects of the signal, allowing the model to effectively classify imagined speech. These features are computed from 19 windows per trial, ensuring a granular analysis of the temporal dynamics of the EEG signals. The extracted features are organized into feature vectors of size 8×64, representing 8 features computed across 64 channels. These feature vectors serve as the input for the SVM classifier, ensuring that the spatial and temporal information of the EEG signals is fully leveraged. A detailed explanation of the features is as follows:

- **Mean**: The mean is the average value of the signal within a window, providing an overall trend of the signal in that segment.

$$\text{Mean} = \frac{1}{N} \sum_{k=1}^{N} x_k \tag{6.1}$$

where x_k represents the value at index k within the window W_n.

- **dMean (Difference Mean)**: The difference mean measures the rate of change of the mean across consecutive windows by capturing the variation in the signal trend over time.

$$d\text{Mean} = \frac{1}{N-1} \sum_{k=2}^{N} |x_k - x_{k-1}| \tag{6.2}$$

- **Energy**: Energy represents the total power or intensity of the signal within the window, computed as the sum of the squares of the signal amplitudes.

$$\text{Energy} = \sum_{k=1}^{N} x_k^2 \tag{6.3}$$

- **Absmin (Absolute Minimum)**: The absolute minimum feature identifies the lowest magnitude of the signal within the window without considering its sign.

$$\text{Absmin} = \min_{k=1}^{N} |x_k| \tag{6.4}$$

- **EHF (Energy in High Frequency Band)**: The total power in the high-frequency components of the signal, which is often analyzed to identify specific neural or physiological activities.

$$\text{EHF} = \sum_{k=\text{high frequency indices}} x_k^2 \tag{6.5}$$

where the high frequency indices represent the frequencies that are classified as high frequency.

- **Curve Length**: Curve length represents the cumulative sum of the absolute differences between consecutive signal values, reflecting the signal's complexity or irregularity.

$$\text{CurveLength} = \sum_{k=2}^{N} |x_k - x_{k-1}| \tag{6.6}$$

- **Absolute Mean**: The absolute mean is the average of the absolute values of the signal, representing the mean magnitude of the signal irrespective of its sign.

$$\text{Absmean} = \frac{1}{N} \sum_{k=1}^{N} |x_k| \tag{6.7}$$

- **Max-Min**: This feature measures the amplitude range of the signal by calculating the difference between the maximum and minimum values in the window.

$$\text{Max-Min} = \max_{k=1}^{N} x_k - \min_{k=1}^{N} x_k \tag{6.8}$$

6.3 SVM Model Performance without ICA

The No-ICA approach uses pre-processed EEG data without artifact removal, simplifying classification but introducing noise and artifact challenges. These distortions affect feature reliability and SVM classifier accuracy, serving as a baseline to assess the impact of preprocessing methods like ICA.

Table 6.1 presents the performance metrics of the SVM classification model on noisy EEG data. The model achieved an overall accuracy of 51%, indicating moderate predictive capability. For the class “gnaw,” the precision is 50%, recall is 58%, and the F1-score is 0.54, based on 53 actual instances. The class “knew” shows a precision of 56%, recall of 47%, and an F1-score of 0.51, with 58 actual instances. For “pot,” the model has a precision of 50%, recall of 57%, and an F1-score of 0.53, derived from 51 instances. The class “pat” exhibits a precision of 56%, recall of 49%, and an F1-score of 0.52, with 47 actual instances. These results reflect the model’s varying performance across the classes, impacted by the noise and artifacts in the EEG data.

Class	**Precision**	**Recall**	**F1-Score**	**Number of Samples**
gnaw	0.50	0.58	0.54	53
knew	0.56	0.47	0.51	58
pot	0.50	0.57	0.53	51
pat	0.56	0.49	0.52	47
Accuracy	0.51			209

Table 6.1: Classification Report of the SVM Model without ICA.

Table 6.2 presents the confusion matrix, offering a detailed view of the model’s predictions under the No-ICA approach. The diagonal elements represent correctly classified instances, such as 31 correctly predicted samples for the class “gnaw.” However, significant misclassifications are observed in the off-diagonal entries, such as 13 samples of “knew” misclassified as “gnaw” and similar errors for “pot” and “pat.” These misclassifications highlight the impact of noise on the classification process, which impairs the model’s ability to distinguish between classes. This analysis underscores the importance of artifact removal techniques like ICA to improve classification accuracy and reduce misclassification rates.

6.4 SVM Model Performance with ICA

Pre-processing and ICA techniques are essential for enhancing the quality of features derived from EEG signals. These methods effectively remove noise and artifacts from raw EEG data, making the extracted features more meaningful and representative of the underlying neural activity. By isolating cleaner signal components, ICA reduces variability caused by noise, thereby improving the reliability and precision of machine learning

	Predicted gnaw	**Predicted knew**	**Predicted pot**	**Predicted pat**
Actual gnaw	31	8	10	4
Actual knew	13	27	9	9
Actual pot	11	6	29	5
Actual pat	7	7	10	23

Table 6.2: Confusion Matrix for the Dataset Without ICA

models like the SVM classifier. This process provides a robust foundation for achieving accurate and consistent classification outcomes.

The classification results highlight the effectiveness of ICA in improving model performance. The SVM classifier achieved an overall accuracy of 69%, as shown in Table 6.3. Metrics like precision, recall, and F1-score were calculated for each class, providing a detailed evaluation of the classifier's ability to distinguish between the four categories. For instance, the 'pat' class recorded the highest F1-score of 0.79, while the 'pot' class showed slightly lower performance with an F1-score of 0.67. This consistent performance across classes underscores the value of ICA in feature refinement, leading to more reliable predictions.

Class	**Precision**	**Recall**	**F1-Score**	**Number of Samples**
gnaw	0.74	0.70	0.72	53
knew	0.68	0.69	0.68	58
pot	0.69	0.65	0.67	51
pat	0.75	0.83	0.79	47
Accuracy	0.69			209

Table 6.3: Classification Report of the SVM Model with ICA.

The confusion matrix in Table 6.4 provides additional insights into the classifier's performance. It summarizes the number of correctly and incorrectly classified instances for each class. For example, the 'gnaw' class achieved 37 correct predictions out of 53, while the 'knew' class had 40 correct predictions out of 58. Some misclassifications occurred, such as 'pot' being confused with 'knew' in 11 instances. Despite these errors, the overall results indicate that the ICA-based pre-processing effectively minimized confusion among classes, ensuring greater confidence and stability in the predictions.

	Predicted gnaw	**Predicted knew**	**Predicted pot**	**Predicted pat**
Actual gnaw	37	5	7	4
Actual knew	6	40	7	5
Actual pot	3	11	33	4
Actual pat	4	3	1	39

Table 6.4: Confusion Matrix for the Dataset With ICA

The EEG-based imagined speech classification achieved an overall accuracy of 71%, evaluated on data collected from 14 subjects and split into training and testing sets, with the test set comprising 209 trials. The trials corresponded to four prompts: "gnaw", "knew", "pot", and "pat". Each trial's EEG signals were processed to extract features such as Mean, Energy, Absmin, EHF, and others across specific frequency bands. These features were normalized and fed into an SVM model with an RBF kernel ($C = 10$). The model performed best on the "pat" prompt, with 83% recall, while performance for other prompts ranged from 65-70% recall.

The confusion matrix revealed misclassifications, particularly between "knew" and "pot", highlighting overlapping EEG feature representations. Overall, the results demonstrate the SVM's capability to generalize across subjects and trials, while pointing to areas for refinement, such as improved feature extraction and noise reduction.

Chapter 7

Imagined Word Recognition Using Long Short-Term Memory Network

Long Short-Term Memory (LSTM) is an advanced type of Recurrent Neural Network (RNN) designed to overcome challenges such as vanishing or exploding gradients that limit the effectiveness of traditional RNNs, particularly with long sequences. LSTMs address this by introducing memory cells and specialized gating mechanisms: the forget gate, input gate, and output gate. These gates regulate how information is retained, updated, or discarded across time steps, enabling the model to capture both short-term and long-term dependencies in sequential data effectively.

An LSTM Network extends the basic LSTM unit by incorporating additional layers, such as dense layers and dropout layers, to form a comprehensive network architecture. This allows the LSTM to learn intricate patterns in temporal data, making it particularly suitable for applications like EEG signal classification, where capturing temporal dependencies is essential. The model used raw EEG data as input, without any pre-processing or artifact removal, enabling it to process the data in its original form and extract patterns directly from the sequential signals.

7.1 Architecture of LSTM

The architecture of an LSTM consists of memory cells and three types of gates: the forget gate, input gate, and output gate. These gates work collaboratively to regulate the flow of information. The forget gate determines which parts of the memory should be discarded, enabling the network to eliminate irrelevant or outdated information. The input gate identifies new information to be stored in the memory, while the output gate decides which portion of the memory should contribute to the output of the current timestep. By integrating these mechanisms, LSTMs are capable of preserving context and dependencies across extended sequences, making them ideal for applications such as time-series forecasting, natural language processing, and activity recognition.

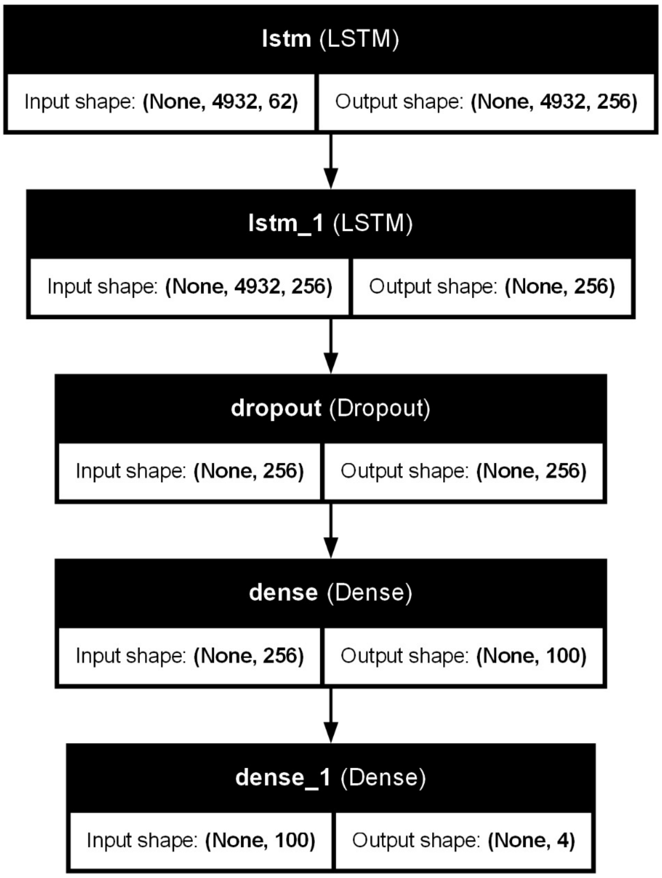


Figure 7.1: Layers of trained LSTM Model

7.2 LSTM Implementation

The LSTM model was implemented to classify sequences into one of four classes, leveraging its ability to extract and process temporal dependencies. The architecture begins with the first LSTM layer, which accepts input sequences of shape (None, 4932, 62), indicating that each sequence comprises 4932 timesteps, with 62 features per timestep. This layer processes the input to produce an output of shape (None, 4932, 256), representing a sequence of feature vectors where each timestep is condensed into a 256-dimensional representation. This transformation allows the model to capture short-term and long-term dependencies within the sequence effectively.

The second LSTM layer takes the output of the first LSTM layer and further refines it by producing a condensed representation of shape (None, 256). Unlike the first layer, this layer outputs only the final state, summarizing the most significant features extracted from the entire sequence. This process reduces the dimensionality of the data while retaining the essential temporal features required for classification.

To prevent overfitting, a dropout layer is included after the second LSTM layer. Dropout regularization is applied by randomly setting a fraction of the layer's activations to zero during training, ensuring that the model generalizes well to unseen data. The output of the dropout layer is then passed to two dense (fully connected) layers. The first dense layer reduces the dimensionality from 256 to 100, focusing on the most critical features learned by the LSTM layers. The second dense layer, which serves as the output layer, maps the 100-dimensional feature vector to 4 dimensions, corresponding to the four classes in the classification task as shown in Fig 7.1.

The LSTM model was chosen for its ability to effectively capture both short-term and long-term dependencies in sequential data, making it ideal for EEG signal classification. Its memory cells and gating mechanisms allow the model to handle temporal relationships, which are crucial for interpreting EEG signals over time. LSTMs are well-suited for real-time processing as they can process data sequentially, making predictions at each time step without the need for complete datasets upfront. This capability ensures efficient and accurate classification of imagined words in real-time scenarios.

7.3 EEG Signal Analysis with LSTM Networks

The analysis begins by emphasizing that the model was trained and tested individually for each participant to accommodate the unique characteristics of their EEG data. The raw EEG signals were directly fed into the model. For 12 participants (IDs: MM08 to MM21), a total of 48 trials were conducted per participant, with 38 trials used for training the model and the remaining 10 trials for testing. For the other 2 participants (IDs: MM05 and P02), 60 trials were conducted, of which 48 trials were allocated to training and 12 to testing. This participant-wise training approach aimed to better capture individual EEG patterns and improve classification accuracy.

The performance results, as summarized in Table 7.1, highlight the variability in accuracy rates achieved by the model across participants. MM15 recorded the highest accuracy of 62.50%, suggesting that their EEG data contained distinct and recognizable patterns that the model could effectively learn and classify. Conversely, P02 recorded the lowest accuracy of 8.33%, indicating significant challenges in generalizing the model to this participant's data, possibly due to overlapping features, noise, or less distinct patterns in their raw EEG signals.

Participants ID	Accuracy (%)
MM05	16.67
MM08	25.00
MM09	54.17
MM10	54.17
MM11	41.67
MM12	33.33
MM14	29.17
MM15	62.50
MM16	20.83
MM18	41.67
MM19	33.33
MM20	25.00
MM21	20.83
P02	8.33

Table 7.1: Participants IDs and their corresponding accuracy values

Several participants achieved relatively high accuracies, such as MM09 (54.17%) and MM10 (54.17%), demonstrating the model's effectiveness for certain individuals. However, lower accuracies were observed for participants like MM05 (16.67%), MM16 (20.83%), and MM21 (20.83%), reflecting challenges related to the quality of raw EEG data, inconsistencies, or insufficiently distinguishing features.

The variability in performance underscores the challenges of using raw EEG signals for classification and highlights the differences in individual EEG characteristics. These results emphasize the importance of exploring more advanced modeling techniques, such as pre-processing of raw EEG data or feature extraction, to address participant-specific variations and improve overall classification accuracy. Factors like noise levels, overlapping features between classes, and inter-individual differences in EEG patterns likely contributed to the observed range of accuracies.

Chapter 8

Conclusion

The development of an EEG-based speech recognition system represents a significant advancement in BCI technology. This approach effectively addressed the challenge of decoding imagined speech using non-invasive EEG signals, showcasing its potential for applications in assistive communication technologies. Through advanced pre-processing, artifact removal, and classification techniques, a robust framework for interpreting neural signals was established.

EEG-based imagined speech recognition was implemented using signal processing and machine learning techniques. FIR filters were designed for noise removal, and ICA was applied to eliminate artifacts like eye blinks and muscle movements, which enhanced signal quality. Features were extracted from the EEG signals and used for classification. SVM was employed, and its accuracy improved after applying ICA. LSTM networks, suitable for real-time processing due to their ability to handle sequential data, were trained and tested individually for each participant.

8.1 Future Scope

Future improvements can further enhance the system's capabilities. Incorporating a wider range of features and involving a more diverse participants will improve the adaptability and generalization of the system. Expanding the application to assistive technologies offers a promising opportunity to address communication challenges for individuals with impairments, creating a profound societal impact.

Bibliography

[1] C. Herff and T. Schultz. “Automatic speech recognition from neural signals: a focused review”. Frontiers in Human Neuroscience, vol. 10, Art no. 429, 2016.

[2] P. Saha, A. Mageed and S. Fels, “Towards Imagined Speech Recognition With Hierarchical Deep Learning”. Proc. Interspeech International Conference, Austria, 2019

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

[4] S. H. Lee, M. Lee and S. W. Lee, “Neural Decoding of Imagined Speech and Visual Imagery as Intuitive Paradigms for BCI Communication”. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 12, pp. 2647-2659, Dec. 2020.

[5] S. Zhao and F. Rudzicz. “Classifying phonological categories in imagined and articulated speech”. Proc. International Conference on Acoustics, Speech and Signal Processing, Brisbane Australia, 2015.

[6] D. Rojas, O. Ramos, and J. Saby. “Recognition of Spanish Vowels through Imagined Speech by using Spectral Analysis and SVM”. Journal of Information Hiding and Multimedia Signal Processing, vol. 7, pp. 889-897, 2016.

[7] M. N. Qureshi, B. Min, H. -j. Park, D. Cho, W. Choi and B. Lee, “Multiclass Classification of Word Imagination Speech With Hybrid Connectivity Features”. IEEE Transactions on Biomedical Engineering, vol. 65, no. 10, pp. 2168-2177, Oct. 2018.

[8] N. Nieto, V. Peterson, H.L. Rufiner, et al. “Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition”. Sci Data 9, Art no.52, 2022.

[9] M.A. Bakhshali, M. Khademi, A.E. Moghadam, S. Moghimi. “EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density”. Proc. Biomedical Signal Processing and Control, May. 2020.

[10] 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, Art no. 867281, 2022.

[11] A. R. Sereshkeh, R. Trott, A. Bricout and T. Chau, "EEG Classification of Covert Speech Using Regularized Neural Networks". IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 25, no. 12, pp. 2292-2300, Dec. 2017.

[12] A. Balaji et al., "EEG-based classification of bilingual unspoken speech using ANN". Proc. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017.

[13] N. Hashim, A. Ali and W. Noorshahida. "Word-Based Classification of Imagined Speech Using EEG". Proc. Computational Science and Technology, Springer, Singapore, 2018.

[14] R. Sakai, A. Kai and S. Nakagawa, "Classification of Imagined and Heard Speech Using Amplitude Spectrum and Relative Phase of EEG". Proc. 3rd Global Conference on Life Sciences and Technologies (LifeTech), Nara, Japan, 2021.

[15] D. Pawar and S. Dhage, "Imagined Speech Classification using EEG based Brain-Computer Interface". Proc. 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 2022.

[16] A. Kamble, P. Ghare, V. Kumar, A. Kothari and A. G. Keskar. "Spectral Analysis of EEG Signals for Automatic Imagined Speech Recognition". Transactions on Instrumentation and Measurement, vol. 72, Art no. 4009409, 2023.

[17] University of Toronto, Department of Computer Science, The KARA ONE Database: Phonological Categories in imagined and articulated speech, accessed on 08 November 2024, https://www.cs.toronto.edu/ complingweb/data/karaOne/karaOne.html

Self Assessment of the Project

Level	Poor (1)	Good (2)	Excellent (3)

Program Outcomes (PO) and Program Specific Outcomes (PSO)

	PO/PSO	**Contribution from the project**	**Level**
PO1	**Engineering Knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization respectively to develop solutions for complex engineering problems.	Exploring imagined speech processing through signal analysis and machine learning to classify EEG signals.	3
PO2	**Problem Analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.	Analyze EEG signals to extract features and scores, enabling classification for imagined speech recognition.	3
PO3	**Design/Development of Solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for public health and safety, whole-life cost, net zero carbon, culture, society, and environment as required.	Artifacts are removed, and the model is trained and tested, enabling its potential use in various medical and other applications.	2

PO4	**Conduct Investigations of Complex Problems:** Conduct investigations of complex engineering problems using research-based knowledge, including design of experiments, modeling, analysis & interpretation of data to provide valid conclusions.	Understanding the concepts of Machine Learning, artifact removal methods, feature extraction, the channels in EEG signal and SVM and LSTM algorithms.	1
PO5	**Engineering Tool Usage:** Create, select, and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modeling recognizing their limitations to solve complex engineering problems.	The project was carried out using modern tools such as MATLAB and EEGLAB for artifact removal, along with MNE-Python for EEG preprocessing and classification.	3
PO6	**The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.	The EEG-based speech recognition project enabled thought-based communication, enhancing human-computer interaction and addressing ethical responsibilities in engineering practice.	3
PO7	**Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.	Followed ethical guidelines in design and implementation, following professional standards and ensuring secure and dependable performance with a plagiarism check done for the report.	3

PO8	**Individual and Collaborative Teamwork:** Function effectively as an individual and as a member or leader in diverse/multi-disciplinary teams.	Worked collaboratively as a team, dividing responsibilities and ensuring effective coordination to achieve the project goals within the defined timeline.	3
PO9	**Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language and learning differences.	The report demonstrates clear communication through structured documentation, effective visuals and societal needs, ensuring technical concepts are accessible to both engineering and non-technical communities.	3
PO10	**Project Management and Finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team and to manage projects and in multidisciplinary environments.	Demonstrate efficient project planning, resource allocation and budget management to develop a cost-effective, flexible and customizable system with high functionality.	3
PO11	**Life-Long Learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.	Artifact removal techniques and machine learning algorithms used in the project can be further extended to other applications.	3

PSO1	The ability to analyse and design systems in the areas related to microelectronics, Communication, Signal Processing and embedded systems for solving real world problems (Professional Skills).	EEG signals are analyzed for imagined speech recognition using artifact removal, feature extraction, and classification with SVM and LSTM, demonstrating potential applications in medical and other fields.	3
PSO2	The ability to identify problems in the areas of communication and embedded systems and provide efficient solutions using modern tools/algorithms individually or working in a team (Problem solving Skills).	MATLAB, Python, EEGLAB, were used to develop the EEG-based speech recognition. Professional ethics were followed in the project development process and documentation. All members in the team contributed to the project.	3

Sustainable Development Goals (SDG)

Sustainable Development Goals Addressed

Goal	**Level**
Good Health and Well-being	2
Gender Equality	1
Decent Work and Economic Growth	3
Industry, Innovation, and Infrastructure	3
Reduced Inequalities	1
Partnerships for the Goals	2