

Mind to Message: Textual Decoding of EEG Patterns

Jeremy Farrugia

Supervisor: Prof. Alexiei Dingli

June 2025

*Submitted in partial fulfilment of the requirements
for the degree of Bachelor of Science in Information Technology (Honours)
(Artificial Intelligence).*



L-Università ta' Malta
Faculty of Information &
Communication Technology

Abstract

Communication is central to human interaction. However, millions of individuals with severe speech or motor impairments, including those with amyotrophic lateral sclerosis (ALS), locked-in syndrome (LIS), or nonverbal autism, face significant challenges in expressing themselves. Assistive technologies aim to bridge this communication gap, and brain-computer interfaces (BCIs) offer a promising approach by enabling users to convey information using neural signals alone.

Many modern BCIs use electroencephalography (EEG) to interpret brain signals, but often depend on stimulus-evoked responses or virtual keyboards, which can be slow and unintuitive. This project explores a more direct approach: converting expressive imagined speech (internal articulation of a word without actual vocalisation) into text, bypassing the need for external stimuli or physical interaction.

To address this, a custom EEG dataset was collected from 21 participants (15 male), each imagining 20 commonly used words selected for their utility. The model architecture used was EEG-Deformer, which includes a shallow convolutional encoder for low-level features and deeper transformer-based modules for temporal dynamics.

In addition to training a baseline model from scratch, this study presents the use of cross-task transfer learning for imagined speech decoding, a novel approach in this domain. The approach involved pretraining models on motor imagery and visual evoked potential datasets, followed by fine-tuning on imagined speech data.

The baseline model achieved a mean within-subject classification accuracy of 84.1%. Models pretrained on motor imagery and visual evoked potentials achieved 54.0% and 77.5% respectively. While pretrained models showed slightly faster convergence, they did not outperform the baseline, indicating limited knowledge transfer from the chosen tasks.

This work presents a domain-specific imagined speech BCI using a small, fixed vocabulary and EEG signals collected using a low-cost device. It highlights both the potential of imagined speech decoding for faster, more intuitive communication and the current limitations of cross-task transfer learning in this domain.

Acknowledgements

I would like to extend my sincere gratitude to the NGO Inspire for their invaluable support throughout this project, granting me access to professionals and resources that were instrumental to my work. I am particularly thankful to Dr. Stefanie Turk for her vital feedback and insights from a neurolinguistic perspective during the early stages of this research.

My deepest appreciation goes to my parents for their unwavering love, support, and motivation. Their affection and guidance have been a constant source of strength, helping me navigate challenges throughout my academic journey and beyond.

I would also like to extend my gratitude to my supervisor, Prof. Alexiei Dingli, whose expertise and guidance were indispensable to this process. His time and input were crucial in helping me realise my full potential.

Finally, I would like to thank my friends and peers for their companionship and support. Our discussions were not only intellectually stimulating, leading to greater insights during this course, but also provided the motivation to truly strive for excellence in my studies. Among them, I wish to specially remember our dear friend, Oleg, who sadly passed during our first year. Your gentle spirit and company were incredibly dear to me, and I will forever cherish our friendship. Your memory lives on through me and all our friends.

Contents

Abstract	i
Acknowledgements	ii
Contents	v
List of Figures	vii
List of Tables	viii
List of Abbreviations	xi
Glossary of Symbols	1
1 Introduction	1
1.1 Motivation	1
1.2 Problem Definition	1
1.3 Aims & Objectives	2
1.4 What makes this study stand out	3
1.5 Document Structure	4
2 Background and Literature Review	5
2.1 Background	5
2.1.1 Brain-Computer Interfaces	5
2.1.2 Neural Interfaces	5
2.1.3 Brain Signal Frequency Ranges	7
2.1.4 Challenges and Limitations in EEG-based BCIs	8
2.1.5 Vocabulary	8
2.2 Literature Review	9
2.2.1 BCI Spellers vs Neural Decoding	9
2.2.2 Related work	11
2.2.3 Overcoming Data Scarcity	16
2.3 Summary	18

3 Methodology	19
3.1 Datasets	19
3.1.1 Imagined Speech Dataset	19
3.1.2 Publicly Available Datasets	23
3.1.3 Data Preprocessing	24
3.2 Model Selection	26
3.3 Model Training	26
3.3.1 Pre-training	27
3.3.2 Within-subject training	27
3.3.3 Cross-subject training	28
3.4 Evaluation Strategy	28
3.5 Summary	29
4 Evaluation	30
4.1 Within-Subject Performance	30
4.2 Cross-Subject Performance	34
4.3 Summary	36
5 Conclusion	37
5.1 Revisiting the aims and objectives	37
5.2 Critiques and Limitations	38
5.3 Future Work	39
5.4 Availability of Data	40
5.5 Final Remarks	40
A Supplemental Literature	58
A.1 Convolutional Neural Networks	58
A.2 Transformers	59
A.3 Overcoming cross-subject variability	61
B Supplemental Results	63
B.1 Cross-Subject Confusion Matrices	63
B.2 Brain Activity	65
B.2.1 Topomaps	65
B.2.2 Time-Series Heatmaps	69
B.3 Saliency	72
B.3.1 Topomaps	72
B.3.2 Time-Series Saliency Maps	76
B.4 t-SNE Plots	76

C	FREC Approval	81
D	Data Management Plan	85
E	Information Letter	88
F	Consent Form	91

List of Figures

Figure 2.1 Neural Interface Placement.	6
Figure 2.2 EEG-Deformer Architecture.	15
Figure 2.3 HCT Block Architecture.	15
Figure 3.1 Methodology Overview.	19
Figure 3.2 Emotiv EPOC X Electrode Placement.	20
Figure 3.3 Data Collection Procedure.	21
Figure 4.1 Accuracy Per Participant (Within-Subject).	32
Figure 4.2 Baseline Confusion Matrix (Within-Subject).	33
Figure 4.3 MI-Pretrained Confusion Matrix (Within-Subject).	33
Figure 4.4 VEP-Pretrained Confusion Matrix (Within-Subject).	34
Figure 4.5 Accuracy Per Participant (Cross-Subject).	36
Figure A.1 Transformer Architecture.	60
Figure B.1 Baseline Confusion Matrix (Cross-Subject).	63
Figure B.2 MI-Pretrained Confusion Matrix (Cross-Subject).	64
Figure B.3 VEP-Pretrained Confusion Matrix (Cross-Subject).	64
Figure B.4 Brain Activity by subject (SI Dataset).	67
Figure B.5 Brain Activity by class (SI Dataset).	67
Figure B.6 Brain Activity by class (MI Dataset).	68
Figure B.7 Brain Activity by class (VEP Dataset).	69
Figure B.8 Time-series Brain Activity (SI Dataset).	70
Figure B.9 Time-series Brain Activity (MI Dataset).	71
Figure B.10 Time-series Brain Activity (VEP Dataset).	71
Figure B.11 Saliency by Class (Baseline).	73
Figure B.12 Saliency by Subject (Baseline).	74
Figure B.13 Saliency by Class (MI-Pretrained).	74
Figure B.14 Saliency by Subject (MI-Pretrained).	75
Figure B.15 Saliency by Class (VEP-Pretrained).	75
Figure B.16 Saliency by Subject (VEP-Pretrained).	76
Figure B.17 Time-series Saliency by Subject (Baseline).	77

Figure B.18 Time-series Saliency by Subject (MI-Pretrained).	77
Figure B.19 Time-series Saliency by Subject (VEP-Pretrained).	78
Figure B.20 t-SNE Visualisation of EEG Data Embeddings (Baseline).	79
Figure B.21 t-SNE Visualisation of EEG Data Embeddings (MI-Pretrained).	80
Figure B.22 t-SNE Visualisation of EEG Data Embeddings (VEP-Pretrained).	80

List of Tables

Table 1.1 Overview of Project Objectives. (Source: Self)	3
Table 3.1 Selected Vocabulary.	20
Table 3.2 Summary of used Datasets.	25
Table 3.3 Summary of Model Hyperparameters. (Source: Self)	26
Table 4.1 Cross-Subject Performance.	31
Table 4.2 Cross-Subject Performance.	35

List of Abbreviations

AAC Augmentative and alternative communication.

ALS Amyotrophic lateral sclerosis.

ASD Autism spectrum disorder.

BCI Brain-Computer Interface.

BiLSTM Bidirectional long short-term memory.

BLEU Bilingual evaluation understudy.

CNN Convolutional Neural Network.

CPM Characters per minute.

DANN Domain Adversarial Neural Network.

DIP Dense information purification.

ECoG Electrocorticography.

EEG Electroencephalography.

ELU Exponential linear unit.

ERN Error-related negativity responses.

ERP Event Related Potential.

FFN Feed forward network.

fMRI Functional Magnetic Resonance.

fNIRS Functional Near-Infrared Spectroscopy.

FTL Fine-grained temporal learning.

GAN Generative Adversarial Network.

GELU Gaussian error linear unit.

HCT Hierarchical coarse-to-fine transformer.

ICA Independent Components Analysis.

ITR Information transfer rate.

KMI Kinaesthetic motor imagery.

LIS Locked-in syndrome.

LLM Large language model.

LOSO Leave-one-subject-out.

MEG Magnetoencephalography.

MI Motor Imagery.

ML Machine learning.

MRCP Movement-related cortical potentials.

MTDANN Multi-source transfer learning method based on domain adversarial neural network.

NLP Natural language processing.

PDF Probability density function.

ReLU Rectified linear unit.

RNN Recurrent neural network.

ROUGE Recall-oriented understudy for gisting evaluation.

SAN Subject Adaptation Network.

SI Speech Imagery.

SMR Sensory-Motor Rhythms.

SNR Signal-to-noise ratio.

SSVEP Steady-state visually evoked potential.

UI User Interface.

VEP Visually Evoked Potential.

VI Visual imagery.

ViT Vision Transformer.

VMI Visual motor imagery.

WGAN Wasserstein Generative Adversarial Network.

WGAN-GP Wasserstein Generative Adversarial Network with gradient penalty.

WPM Words per minute.

1 Introduction

1.1 Motivation

Communication is a fundamental human ability, enabling social interaction, self-expression, and access to education and employment. However, millions of individuals worldwide face significant barriers to effective communication due to speech and motor impairments. Conditions such as ALS, LIS, and severe autism spectrum disorder (ASD) can greatly impact an individual's ability to interact with others.

ALS is a progressive neurodegenerative disease. A meta-analysis by Xu et al. estimated that ALS affects approximately 4.42 individuals per 100,000 people globally [1]. In France, Couratier et al. reported that approximately 6,000 individuals are currently living with ALS [2].

Locked-in syndrome is characterised by complete paralysis of voluntary muscles with the exception of those controlling eye movements, leaving individuals fully conscious yet unable to move or speak [3]. LIS is an extremely rare condition, with an estimated prevalence of fewer than 1 in 1,000,000 individuals according to Orphanet [4], and approximately 500 documented cases reported in France by Schnetzer et al. [5]. Even among those who experience partial recovery, long-term impairments are common [6].

Autism spectrum disorder, particularly in its severe forms, can also lead to significant communication challenges. Recent data reviewed by Bougeard et al. suggests that approximately 1 in 89 children in Europe are diagnosed with ASD [7]. While not all individuals with ASD experience speech impairments, a significant subset faces difficulties in verbal communication, necessitating alternative methods to express themselves.

These impairments can cause isolation, reduced quality of life, and loss of independence. Assistive technologies, particularly BCIs that translate brain activity into speech or text, offer a promising way to restore communication and support inclusion for severely affected individuals.

1.2 Problem Definition

Decoding brain activity to enable direct communication for non-verbal individuals is an active area of research within neuroscience and machine learning (ML). Current assistive technologies, such as eye-tracking systems or stimulus-driven BCIs, can be slow, or require external stimuli, limiting their practicality and ease of use [8, 9].

One promising direction in BCI research is the decoding of expressive imagined speech. Also referred to as inner speech [10]. Expressive imagined speech is the mental simulation of speaking a word or phrase without any external articulation or muscle movement. Unlike speech planning, which involves preparing linguistic content (determining *what* to say) and the necessary muscle movements (speech motor planning) [11], or articulation (vocally expressing the content through motor execution), expressive imagined speech reflects the internal experience of speaking without producing audible output.

1.3 Aims & Objectives

The primary aim of this research is to develop an ML-based system capable of decoding EEG signals recorded during expressive imagined speech into word-level text outputs for a closed vocabulary. This work seeks to enable direct and efficient communication for non-verbal individuals through a word-based BCI.

In doing so, the project contributes to the development of more practical BCI systems by investigating how models can overcome data scarcity and generalise to individuals unseen during training, reducing reliance on subject-specific calibration and increasing scalability. At the same time, it contributes to the scientific understanding of inner speech decoding by examining how models pre-trained on distinct cognitive processes, such as motor imagery (MI) and visually evoked potential (VEP), can be fine-tuned for imagined speech tasks. This generates insights into potential neural and representational alignments between these processes and imagined speech, and may illustrate the transferability of EEG-based features across cognitive domains.

To achieve these aims, the project is structured around a series of objectives as seen in Table 1.1.

Table 1.1 Overview of Project Objectives. (Source: Self)

AIM: Develop and Evaluate an Imagined Speech BCI			
No.	Objective	Research Question	Chapter/s
1	Data Collection and Preparation: Select an appropriate fixed vocabulary and design a procedure for collecting high-quality imagined speech data, then collect and preprocess the dataset.	How can vocabulary and experimental design be optimised to produce a clean, usable imagined speech dataset for deep learning?	2, 3
2	Dataset Selection: Identify and prepare existing EEG datasets for use in transfer learning.	Which publicly available EEG datasets are suitable for pretraining models for imagined speech decoding?	3
3	Model Development: Design and implement a deep learning model for imagined speech classification.	What model architecture is best suited for classifying imagined speech EEG signals into predefined word classes?	2, 3
4	Model Training: Pretrain models on selected datasets and fine-tune them on the imagined speech dataset. Train a baseline from scratch using imagined speech data.	How does pretraining on different EEG paradigms affect performance, and does fine-tuning improve accuracy over training from scratch?	3, 4
5	Evaluation and Validation: Evaluate the final system's performance, including its ability to generalise to unseen participants.	How well does the trained model generalise across participants, and what benefits does transfer learning provide?	4

1.4 What makes this study stand out

Unlike previous projects done at the University of Malta, such as the work done by Demicoli [12], and Saliba [13], which primarily utilised either convolutional neural network (CNN) or transformer architectures, this work explores a more powerful model architecture which combines both CNNs and transformers for high-quality feature

extraction and classification. Additionally, this project aims to use supplemental datasets and fine-tuning to addressing limitations in data scarcity. Furthermore, rather than evaluating performance solely on participants seen during training (within-subject evaluation), it also explores how well the model generalises to individuals unseen by the model during training (cross-subject evaluation). Lastly, by incorporating a more expansive vocabulary, this study enhances the practicality and usability of the model for communication tasks.

Previous studies typically employ transfer learning for *within-task* transfer, wherein a model trained on data from an individual or group of individuals performing a specific task, such as classifying right-fist and left-fist motor imagery, is adapted for a new individual performing the same task. In contrast, this study explores *cross-task* and *cross-domain* transfer by pre-training models on MI data from one set of participants and subsequently fine-tuning them on speech imagery (SI) data from a different participant group, as a means to overcome data scarcity.

1.5 Document Structure

To guide the reader through this study, the paper is structured as follows:

Background and Literature review: In the background section the fundamental technologies support this project such as BCI and EEG technology are explained. In the literature review we present an overview of related work, covering a variety of approaches such task-specific models, paradigm-agnostic models and approaches to overcoming data scarcity.

Methodology: In this chapter we explain the approach to data collection and selection of relevant datasets and their preprocessing. We then go over the selection of model architecture and its hyperparameters. We then explain the training process of all models. Finally we explain the planned approach to testing the models and mention the relevant metrics

Evaluation: The results are presented alongside relevant observations, discussing the strengths and weaknesses of the different models.

Conclusion: This work is concluded with a summarisation of the work done in this research, revisiting and reflecting on which aims and objects have been achieved. We also provide a discussion of the limitations inherent to this project, and expand on directions for future research. Finally, the paper concludes by discussing its potential impact and the broader implications of our findings.

2 Background and Literature Review

This chapter is divided into two main sections: background and literature review. The background section introduces the concept of BCIs, outlining their key components, typical use cases, and foundational concepts. The literature review then presents an overview of related work, highlighting current approaches to the challenges of building a BCI, with a focus on communication-oriented systems.

2.1 Background

2.1.1 Brain-Computer Interfaces

The process of collecting neural signals and classifying or decoding them involves the use of a BCI. A BCI is a combination of hardware and software acting as a communications system that allows an individual's neural activity to control an external device without needing any additional signals [14]. The main goal of BCIs is to allow individuals with limitations or disabilities to better interact with the world around them. BCIs have a rich history, their first roots can be traced back to the 1920s when Hans Berger first discovered EEG [15]. In the 1970s, Vidal continued Berger's research in the field and successfully detected brain events, now known as event related potentials (ERPs), in real-time using EEG [16]. Since then, BCI technology has greatly expanded, being used for a variety of applications, from autism detection to prosthetics [17–19].

2.1.2 Neural Interfaces

A BCI generally consists of a system to acquire brain signals (a neural interface), a system that processes and decodes the signals detected by the neural interface, and a control interface [20]. Signal acquisition can be performed through a variety of methods [21], all of which yield their own respective advantages and disadvantages. Neural interfaces can be split into two main groups: Invasive and non-invasive.

Invasive Neural Interfaces

Invasive approaches involve the placement of electrodes under the skull, requiring a surgical procedure to implant. These include:

Depth electrodes: which are implanted in the brain's white matter, measuring electrical activity in a local region. Depth electrodes permit a high level of sensitivity, up to the single-neuron level. However, accurate interpretation of

signals may require a larger context, due to the recorded signal potentially having propagated from another region in the brain [22].

Microelectrode arrays: which are implanted in the brain's cortex, providing high quality signal data gathered from large groups of neurons by recording the action potentials (spikes in activity) produced by neurons when communicating.

Electrocorticography (ECoG): which involves placing electrodes on the surface of the brain, producing data with a lower spatial resolution and signal-to-noise ratio than the other invasive approaches due to a larger volume of signals originating from different parts of the brain are recorded. ECoG has been shown to be sufficient for a variety of tasks, such as distinguishing between spoken words [23].

Although invasive approaches offer high levels of accuracy [24, 25], these interfaces come with significant risks whose longer term effects are not yet fully understood and require ongoing maintenance due to electrode degradation [22, 26–28], requiring further research [29]. The placement of these neural interfaces is illustrated in Figure 2.1.

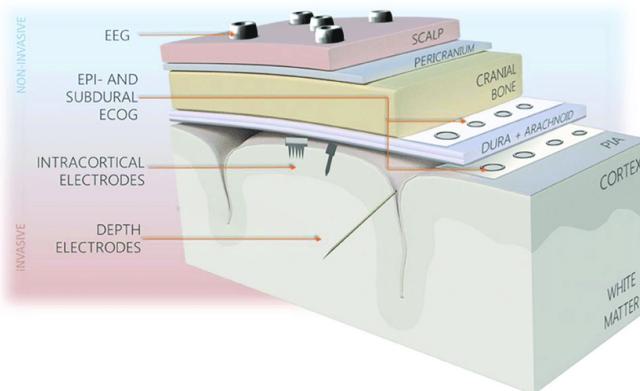


Figure 2.1 Neural Interface Placement. (Source: Szostak et al. [30])

Non-Invasive Neural Interfaces

Non-invasive interfaces involve significantly reduced risk [31], making them far more ideal until the long-term impacts of invasive approaches are better understood and mitigated. Multiple non-invasive modalities exist [21], such as:

Magnetoencephalography (MEG): which measures magnetic fields created by neural activity but requires highly specialised equipment which is very expensive and non-portable.

Functional magnetic resonance (fMRI): which measures blood oxygen levels of the brain, revealing haemodynamic response, also requiring specialised equipment which is non-portable. While fMRI has been shown to be viable for the creation of BCIs [32, 33], it may be more useful as a tool to identify potential areas of interest for implantable BCIs [34].

Functional near-infrared spectroscopy (fNIRS): which uses near-infrared light to similarly measure blood oxygen levels. fNIRS technology is both portable and wearable making it ideal for use as part of a BCI. However, haemodynamic response is an indirect measurement of neural activity, which occurs at a slight delay and allows a lower temporal resolution compared to other neural interfaces, making it less ideal for a real-time signal-to-text BCI. Despite its limitations, fNIRS has been proven to be a suitable option for use in BCIs [35].

EEG: which measures electrical activity across the scalp. EEG has worse signal-to-noise ratio than other neural interfaces and has a lower spatial resolution than fMRI. However, EEG devices are highly available, far cheaper and directly measure neural activity through electrical potentials, allowing a high temporal resolution, making it more ideal for an accessible real-time BCI. A significant shortcoming of EEGs is their susceptibility to noise, with subtle movements largely affected the collected data.

Comparison of Invasive and Non-Invasive approaches

The main distinction between invasive and non-invasive neural interfaces lies in the trade-off between signal quality and safety. Invasive methods offer higher signal quality due to direct neural contact [31], but require surgery, incurring risks [27, 28], additional costs, and lower user acceptance [36, 37]. In contrast, non-invasive approaches like EEG are safer, more accessible, and cost-effective, though they suffer from lower spatial resolution and greater susceptibility to noise [38].

2.1.3 Brain Signal Frequency Ranges

EEG signals are categorised into broadly defined groups depending on their frequency bands. The exact values of these frequency bands vary between literature [39–41], but can be broadly defined as follows: **Delta Waves** (0.5–4 Hz), **Theta Waves** (4–8 Hz), **Alpha Waves** (8–13 Hz), **Beta Waves** (13–30 Hz) and **Gamma Waves** (Over 30 Hz). While changes in the high gamma frequency range are associated with overt and covert speech, imagined speech and other aspects of language processing or production also cause changes in other lower frequency ranges [42]. The selection of

frequency range to maintain after filtering varies greatly across literature with the most common ranges being 8–20 Hz and 2–50 Hz [43]. Research suggests taking a broader range of frequencies may be more ideal as it allows extraction of a more robust set of features, better able to distinguish between different signal patterns. One study [44], obtained higher accuracy by using a high gamma 60–80 Hz range. However, other studies suggest that this higher gamma frequencies may be closely related to muscle movement artifacts when using an EEG [43]. Additionally, higher frequencies may be subject to more noise due to higher frequencies being less common, increasing the impact of outliers and leading to a lower overall signal-to-noise ratio.

2.1.4 Challenges and Limitations in EEG-based BCIs

EEG-based BCIs face several technical and practical challenges that constrain their performance and generalisability [38]. A fundamental issue is the low signal-to-noise ratio (SNR) in EEG recordings [45]. Neural signals of interest are often obscured by noise from muscle activity, eye movements, and external electrical sources, making extraction of the relevant signals difficult.

Inter-subject variability presents another significant limitation [46]. Differences in brain anatomy, cognitive strategies, and skull morphology lead to highly individualised neural responses. This makes it challenging to develop models that generalise across users. Additionally, intra-subject variability, caused by the non-stationary nature of brain signals due to factors such as fluctuations in attention, mental state, or fatigue, introduces inconsistency even within the same individual across time and different trials.

Systemic variability, such as inconsistencies in electrode placement, headset fit, hardware differences across recording sessions or sites [47], or the set-up/location itself [48], can further degrade data quality and reduce reproducibility, particularly in multi-user or multi-laboratory studies.

Finally, data scarcity remains a major barrier [38]. Collecting large, annotated EEG datasets is time-consuming and resource-intensive, especially for tasks like imagined speech, which are cognitively demanding and lack clear behavioural outputs. This scarcity limits the capacity to train robust ML models and further exacerbates issues of overfitting and poor generalisability.

2.1.5 Vocabulary

The choice of vocabulary for a closed-vocabulary communication device is crucial, as it directly impacts the system's ability to meet users' everyday communication needs. Research in both augmentative and alternative communication (AAC) systems and BCIs

highlight the importance of “core vocabulary” [10, 49]. Core vocabulary is a set of high-frequency, versatile words that enable broad communication across various contexts, identified through analyses of various communication samples [50]. Core vocabulary includes function words (e.g. “I”, “you”) and essential content words (e.g. “want”, “help”) that account for the majority of daily communication.

In contrast, “fringe vocabulary” consists of less common, more context-specific words (e.g. “strawberry”, “guitar”) [51]. These words are typically chosen based on individual user needs or specific topics to allow for specialised expression. While fringe vocabulary is essential for expressing specialised ideas or engaging in detailed discussions, it is typically less versatile and accounts for a smaller proportion of everyday communication. The balance between core and fringe vocabulary is an essential consideration in both AAC and BCI system design. For a general-purpose world-level BCI, prioritising core vocabulary is crucial, as it ensures broader applicability and usability across diverse users and scenarios.

2.2 Literature Review

2.2.1 BCI Spellers vs Neural Decoding

BCIs intended on facilitating speech can be broadly categorised into two different groups: Those aimed at direct translation, referred to as neural decoding, and those aimed at using an interface to allow individuals to produce speech on the character level, referred to as BCI spellers.

BCI Spellers

BCI spellers make use of a user interface (UI) which implements variations of what is referred to as a brain cursor - a visual interface controlled via signals gathered by the neural interface [52]. Three main paradigms exist within BCI spellers, P300 systems [53], steady-state visually evoked potential (SSVEP) systems, and BCI-MI systems [54].

P300 Systems P300 BCI spellers use a positive ERP which occurs roughly 300ms after a stimulus, most commonly a visual stimulus [55]. These spellers consist of letters arranged in a grid, users need to focus on their target letter as the columns and rows in the grid flash, once a column or row with the target letter is highlighted the brain produces the P300 signal, which is then detected by a BCI and interpreted accordingly. The row and column producing the largest P300 signal would be interpreted as containing the target character, and the character at their intersection would be selected. While P300 systems have fairly high accuracy and are more easily

generalisable than direct translation paradigms, they suffer from a low rate of information transfer [56, 57].

SSVEP SSVEP is another widely studied signal, it refers to a continuous electrical response generated in the brain when a person focuses on a visual stimulus that flickers or oscillates at a constant frequency. The original design of SSVEP spellers consist of an interface with a keyboard and four fields with arrows to move the keyboard cursor left and right, up and down, and a “Select” field to allow character selection. Each of these fields flicker at specific frequencies, causing a steady-state visual evoked response which the BCI uses as an input [58]. Compared to P300 systems, SSVEP systems have a greater information transfer rate (ITR) and lower SNR [59, 60]. Additionally, compared to other paradigms, SSVEP BCIs can be implemented with a much lower number of electrodes. Despite this, they still suffer from a low overall ITR.

MI-based systems MI-BCI systems make use of the motor imagery evoked by imagining the movement of a body part such as the arm. Characters in MI-BCI spellers are split into different wheels which may be selected through MI signals, after which the desired character may be chosen from the selected wheel. MI systems have the advantage of being independent from gaze, unlike SSVEP and traditional P300 approaches.

One study created a typing interface which allows users to input, cancel, delete, and confirm a message [61]. By stacking features extracted by a CNN and recurrent neural network (RNN) in parallel their BCI achieved theoretical speed of 6.67 characters per minute (CPM) at an accuracy of 95.53% and 94.27%. Another study achieved a typing rate of 6.6 words per minute (WPM) or 31.9 CPM with an accuracy of 90.36% by making use of emotion based word predictions [62]. This study also made use of a novel interface consisting of multiple character wheels, allowing users to navigate between wheels to access different characters at greater speeds.

While BCI spellers allow high levels of freedom with respect to vocabulary and achieve high levels of accuracy, individuals are greatly limited by the system’s low speed of communication when compared to overt speech, imposing major limitations on the users of these paradigms.

Neural Decoding

Neural decoding involves BCIs which translate brain signals directly into text or text fragments. Neural decoding can be achieved through an open vocabulary system, which allows for the interpretation and generation or processing of any possible word

rather than being restricted to a limited, predetermined list, or a closed-vocabulary which consists of a fixed predetermined list of curated words. Neural decoding generally relies on using expressive imagined speech.

Different studies took different approaches such as using syllables as prompts [63], breaking down words into phonemes [49], and directly translating into full words [64, 65]. While breaking down words into phonemes or vowels may potentially allow for a more expensive vocabulary with less data, errors in phoneme decoding can compound, making accurate word-level reconstruction difficult, while requiring greater mental effort from the user to produce clear, distinguishable signals for each phoneme. This can also lead to fatigue, especially through extended usage.

2.2.2 Related work

In recent years, a number of studies on signal- to-text BCIs have focused on the use of CNNs [66–68] and transformers [69, 70].

Brain2Char Sun et al. [66] used data from four participants who were prompted to read sentences on a screen while their speech and ECoG data were synchronously recorded. The researchers proposed Brain2Char, a neural speech recognition framework with modular architecture composed of three parts: an encoder, a decoder and a latent representation regulariser. Unlike models that decode neural signals into word sequences, Brain2Char focused on decoding ECoG data into character sequences, using sensorimotor signals for fine-grained temporal information. The encoder passes 3D ECoG signals into a set of inception layers (a special type of convolutional layer), extracting spatial-temporal features. These layers are followed by two bidirectional long short-term memory (BiLSTM) layers which create the latent feature representation of the data. To guarantee meaningful representations in feature space the regularisation module performs simple feed-forward transformations to account for the variance in neural signals. Finally, the decoder employs dilated CNNs to convert these latent representations into character sequences, yielding the model's output.

While the model achieved impressive word error rates (10.6%, 8.5%, and 7%) across vocabularies of up to 1900 words, Brain2Char's reliance on invasive ECoG data greatly reduces its accessibility and appeal, with a majority of impaired individuals stating a preference of non-invasive BCIs [36, 37, 71]. Furthermore, the low participant count and lack of testing on out-of-dataset individuals raises concerns about its generalisability and applicability to a larger number of individuals, the main limitation in real-world deployment for BCIs.

DeWave Duan et al. [72] proposed DeWave, an EEG-to-text translation framework that integrates deep learning with large language models (LLM) to achieve open-vocabulary decoding. The architecture consists of three main components: a quantised variational encoder, a discrete codex representation module, and an LLM-based decoder. First, the encoder processes raw EEG signals into vectorised representations, which are then transformed into discrete codex entries. This quantisation step addresses the high variability and noise in EEG signals by aligning them to a shared feature space. The discrete codex entries are subsequently fed into a pre-trained language model, such as BART, to generate textual outputs. By evaluating the distances between EEG features and codex entries, the model minimises the impact of signal variance, ensuring more robust and consistent outputs. DeWave eliminates the need for event markers (such as eye fixation points) typically required for segmentation, allowing it to operate on continuous EEG data.

DeWave was mainly evaluated using bilingual evaluation understudy (BLEU) and recall-oriented understudy for gisting evaluation (ROUGE) scores, which measure translation accuracy. The model demonstrated significant performance improvements on the ZuCo and ZuCo 2.0 datasets [73, 74], achieving higher BLEU-1 and ROUGE-F scores compared to the baseline. However, its dependence on data from natural reading tasks raises concerns about its applicability to imagined speech, its stated target domain, which while similar, consists of distinct neural activity. Furthermore, DeWave's performance evaluation relies on teacher forcing, which raises concerns about its viability in real-world applications, due to a strong positive bias in the reported results, not reflecting real-world use; an issue commonly observed in a number of related studies [75]. Additionally, while the generalisability of DeWave seems promising, the model's performance falls far below what impaired individuals consider acceptable [36, 37, 71], with BLEU-{1, 2, 3, 4} scores of 41.35, 24.15, 13.92, and 8.22 respectively. An additional concern with the results is that BLEU does not account for semantic coherence [76], meaning high scores can still arise from outputs containing unrelated or nonsensical content, a reality demonstrated within the research done by Duan et al., with the non-overlap words being mostly incoherent.

Paradigm Agnostic EEG Models

A number of researchers have attempted to create 'task-agnostic'/'paradigm agnostic' models, capable of handling EEG data from a variety of domains, applicable for different tasks, rather than being made with a specific modality or paradigm in mind (e.g. a P300 communication BCI).

EEGNet Lawhern et al. [77] introduced ‘EEGNet’, a CNN for EEG-based BCIs. EEGNet was evaluated across 4 different paradigms: P300 visual-evoked potentials, error-related negativity responses (ERN), movement-related cortical potentials (MRCP), and sensory-motor rhythms (SMR). EEGNet’s primary goal was to address the limitation of traditional BCI approaches that often required paradigm-specific signal processing, feature extraction, and classification methods. EEGNet’s architecture incorporates depthwise and separable convolutions, adapted from computer vision, to encapsulate well-known EEG feature extraction concepts such as optimal spatial filtering and filter-bank construction. This reduces the need for requires subject-matter expertise and a priori knowledge about the expected EEG signal, making the creation of EEG-based BCIs quicker and more accessible.

The architecture begins with a temporal convolution layer that applies 2D filters to extract band-pass frequency representations. This is followed by a depthwise convolution layer that learns spatial filters while minimising parameter count. Batch normalisation regularises the data and an exponential linear unit (ELU) activation introduce non-linearity, after which the output is average pooled. The next block, a separable convolution (depthwise followed by pointwise convolutions), further reduces parameters and separates within and across-feature relationships. The final steps involve average pooling, flattening, and a softmax classifier.

EEGNet demonstrated competitive performance across all paradigms. In within-subject testing, it roughly matched baseline performance for the P300 and SMR paradigms, and outperformed them in the ERN and MRCP paradigms. However, in cross-subject testing, EEGNet’s performance declined, performing similarly to the baselines across all tasks, with slightly reduced performance to the best baseline in the ERN and MRCP paradigms.

EEG Conformer Song et al. [78] proposed the use of an ‘EEG Conformer’ - a convolutional transformer model for decoding EEG data. The architecture aims to combine the strengths of convolutional layers for low-level temporal and spatial feature extraction, with self-attention mechanisms to model long-range dependencies. The model was proposed to address the limitations of task-dependent feature engineering, in line with the motivations behind EEGNet [77], but with an increased ability to capture global relationships in time series data.

EEG Conformer consists of 3 main modules: a convolution module, self-attention and fully-connected classifier. The model first applies temporal convolution using a kernel size of $(1, 25)$ and a stride of $(1, 1)$, focusing on extracting features across the time dimension. This is followed by spatial convolution using kernels of size $(ch, 1)$, with a stride of $(1, 1)$, where ch is the number of electrode channels within the EEG data, allowing the model to learn spatial patterns across

electrodes. Batch normalisation and an ELU function are applied, regularising the data and introducing non-linearity. To reduce temporal resolution, average pooling is performed with a kernel size of $(1, 75)$ and a stride of $(1, 15)$, smoothing features in the temporal domain and minimising learnable parameters. The resulting feature maps are reshaped into token representations, which are passed into a multi-head self-attention module repeated N times, where N is a tunable hyperparameter. Finally, the output is classified via two fully connected layers and a softmax function.

Song et al. evaluated the model using three datasets: the BCI competition IV datasets 2a and 2b [79], which are widely used MI datasets, and the SEED dataset [80], which focuses on emotion recognition from EEG signals. To augment the data, the authors applied signal segmentation and reconstruction in the time domain prior to training. EEG Conformer outperformed its baselines, including EEGNet, achieving higher average accuracies across participants. However, its evaluation was limited to within-subject experiments, and no cross-subject testing was performed, raising questions about its generalisability. Additionally, unlike EEGNet, the model was primarily tested on MI datasets, with less emphasis on ERP-based paradigms, making its performance across different BCI tasks less certain.

EEG-Deformer Ding et al. [81] proposed ‘EEG-Deformer’, a dense convolutional transformer for BCIs. EEG-Deformer builds on the work done by Song et al. [78], with the goal of better capturing coarse-to-fine temporal dynamics of EEG signals through the use of two novel components: a hierarchical coarse-to-fine transformer (HCT) block, and a dense information purification (DIP) module. Rather than using a CNN as the main body of the model, EEG-Deformer uses a CNN as an adaptive feature encoder to adaptively preprocess the EEG data and a transformer to learn long-range temporal relationships in the data.

EEG-Deformer consists of 3 components, as seen in Figure 2.2: a CNN based shallow feature encoder, a HCT composed of D cascading hierarchical coarse-to-fine blocks, where D is a tunable hyperparameter, and a DIP module for latent representations.

The shallow feature encoder consists of a two layer CNN to capture spatial and temporal information, and a tokeniser. The model first applies temporal convolution using a kernel size of $(1, 0.1 \cdot f_s)$, where f_s is the EEG’s sampling frequency. This is inspired by the brain’s microstates which last roughly 100 ms. This is followed by spatial convolution using kernels of size $(ch, 1)$, where ch is the number of electrode channels within the EEG data. Batch normalisation and an ELU function are applied, followed by max-pooling in the temporal domain with a kernel of size $(1, 2)$ and no overlap. The features are then rearranged to serve as tokens and combined with a learnable position encoding. The shallow features are then passed through the HCT, as seen in Figure 2.3,

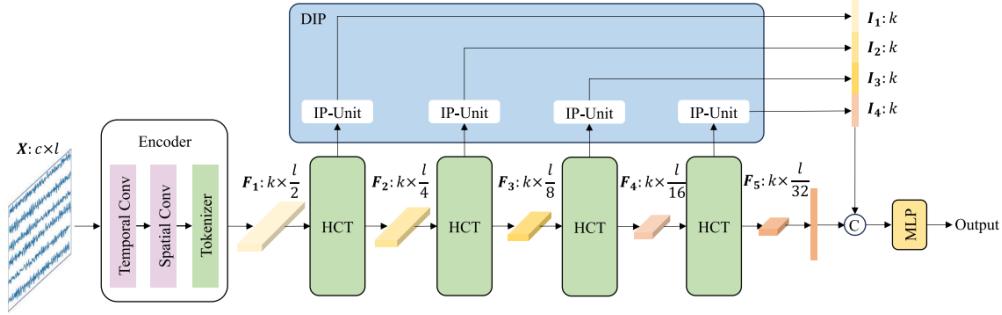


Figure 2.2 EEG-Deformer Architecture.(Source: Ding et al. [81])

which consists of parallel transformer-based and CNN-based branches.

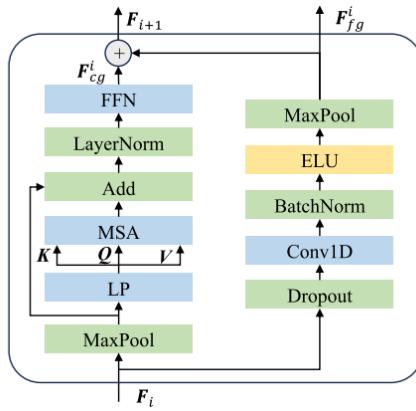


Figure 2.3 HCT Block Architecture. (Source: Ding et al. [81])

In the transformer branch, data is first max-pooled using a kernel of size $(1, 2)$ and a stride of $(1, 2)$ to reduce the size of the input. The features are then linearly projected and multi-head self attention is applied to extract the attention across the temporal tokens. A residual connection from the max-pooling is added to the output of the attention module, and layer normalisation is performed. The data is then passed to the feed forward network (FFN) which consists of two layers, the first with a gaussian error linear unit (GELU) activation function for linearity, followed by a dropout layer, and the second a linear layer, also followed by dropout. The CNN branch referred to as a fine-grained temporal learning (FTL) module, uses 1D convolution across the time dimension whose size is a tunable hyperparameter. The output of the 1D convolution is batch normalised and ELU is applied. After max-pooling with a kernel of size $(1, 2)$ and a stride of $(1, 2)$, the outputs of both branches are added to fuse the features. Each HCT block is connected to an information purification unit using the output of the block's FTL module, denoted by \mathbf{F}_{fg}^i . The information purification unit encodes frequency information in the EEG signals by calculating the signal's power as seen in Equation 2.1, where I_i is the i^{th} learned hidden representation, l_t^i is the length of the EEG data across the time dimension in the i^{th} HCT block, and $f_{fg}^{i,j}$ is one learned

representation from the j^{th} CNN kernel in the FTL module of the i^{th} HCT block.

$$I_i = IP_{power} (\mathbf{F}_{fg}^i) = \left\{ \log \left(\frac{1}{l_t^i} \sum (f_{fg}^{i,j})^2 \right) : f_{fg}^{i,j} \in \mathbf{F}_{fg}^i \right\} \quad (2.1)$$

After passing through all D HCT blocks, the output from each information purification unit is concatenated together, which is then also concatenated with the flattened output of the final HCT block. The data is then passed through a single layer using the softmax function for classification.

Ding et al. evaluated the model using 3 EEG datasets: a cognitive attention dataset [82], a driving fatigue dataset [83], and a cognitive workload dataset [84]. The authors pre-processed the dataset using a band-pass filter (differing for each dataset), independent components analysis (ICA) for artifact removal, and segmented the data using a 4 second sliding window with 50% overlap. The model's performance was compared against 8 different baselines, notably including the EEGNet and EEG Conformer models. The model was evaluated using cross-subject evaluation, using a leave-one-subject-out (LOSO) approach for each participant.

To assess how EEG-Deformer compares with transformer-based architectures, the baselines also included a purely transformer-based model: an adapted Vision Transformer (ViT) [85], which treats the EEG signal as a sequence of non-overlapping temporal patches. This ViT variant lacks any convolutional components and applies global self-attention directly to these temporal segments. EEG-Deformer outperformed the baseline models on each dataset, achieving accuracies of 82.72%, 79.32% and 73.18% respectively. The ViT model performed notably worse than approaches which incorporated CNNs, such as EEGNet and EEG Conformer. However, unlike EEGNet and EEG Conformer, EEG-Deformer was not tested using any MI data. Furthermore, the datasets used are not related to traditional assistive BCI paradigms such as P300 or SSVEP, raising concerns about its utility in these domains.

2.2.3 Overcoming Data Scarcity

Data scarcity is a significant challenge in designing high quality EEG-to-text BCIs. To address this limitation, researchers have explored various strategies to increase the volume and diversity of training data, while maintaining and promoting the most generalisable features.

Data Augmentation

One prominent approach was applying data augmentation, a method to increase the diversity and quantity of available data through systematic modifications. EEG data can be augmented using a range of techniques, including signal augmentation techniques

[86, 87]. The impact of these techniques has been evaluated across different applications, such as sleep stage classification and motor imagery tasks. While data augmentation has shown measurable improvements in model performance for both tasks, its effectiveness appears to be greater in motor imagery related tasks, likely due to the distinct characteristics and signal patterns associated with this activity. Certain techniques may harm the learnability of the data due to decreasing the already low SNR, leading researchers opt for techniques such as signal segmentation and reconstruction [78]. A simple approach to augmenting EEG data without harming its SNR is segmenting the data using a sliding window with overlap, and has been applied in a number of studies [81, 88, 89].

GANs

Another approach to augmenting datasets is synthesising new data using deep learning techniques such as autoencoders and generative adversarial networks (GAN). These methods have demonstrated significant potential for improving the performance of EEG-related models across various tasks [90, 91]. In [92], Bhat and Hortal made use of a Wasserstein generative adversarial network with gradient penalty (WGAN-GP) to augment data for use in a four-class emotion classification task. Wasserstein generative adversarial network (WGAN) models make use of Wasserstein distance, a metric that provides a more stable measure distance between probability distributions, to help avoid mode collapse (a common failure in GANs where the generator learns to produce limited or identical outputs). A WGAN consists of two modules: a generator, which aims to produce synthetic data samples that resemble the true data distribution, and a critic (replacing the discriminator in traditional GANs), which evaluates how closely generated samples match real data. The inclusion of synthetic data generated by the GAN improved classification accuracy, with reported gains ranging from 0.3% for one class to a significant 17.5% for another class when using neural networks [92]. Beyond data augmentation, GANs have also been applied to reduce noise in EEG signals, further enhancing model performance and robustness [93]. Despite their advantages, GANs may not be able to produce high quality generalisable outputs if sufficient is not available, reducing their utility for smaller more complex datasets [94].

Transfer Learning

Collecting high-quality data during imagined speech presents significant challenges and often requires training participants to produce reliable signals. To address data scarcity, researchers have leveraged transfer learning, a machine learning technique where knowledge gained from training on one task or dataset is reused to improve performance on a different but related task [95].

Following this approach, Lee et al. collected data from both covert (imagined) speech and overt (spoken) speech, noting that signals from overt speech are typically clearer and more consistent [96]. The authors pre-trained a model using EEG signals from overt speech and then applied transfer learning to fine-tune the model on imagined speech data. While transfer learning did not lead to significant improvements, the transferred models performed comparably to those trained solely on imagined speech, suggesting potential shared features between the two modalities.

Building on this idea, Liu et al. proposed a method to enhance transfer learning when working with heterogeneous datasets, addressing challenges such as differences in stimuli, equipment, and participant characteristics [97]. The authors introduced multi-source transfer learning method based on domain adversarial neural network (MTDANN), a model combining a domain adversarial neural network (DANN) (a type of GAN) with transfer learning principles. MTDANN incorporated a feature extractor based on the EEGNet architecture [77], alongside a classifier and a domain discriminator designed to reduce the domain shift between source and target data. The model was evaluated on three motor imagery (MI) datasets from the BCI Competition III and IV [79], consistently outperforming EEGNet and other baselines, demonstrating its effectiveness. An unexplored avenue is the use of transfer across different domains, such as between visual imagery (VI) and SI modalities, which may offer great potential in overcoming data scarcity for future models.

2.3 Summary

In this chapter, we presented background information on BCI technology, outlining its key components and applications. We also summarised the relevant brain signal frequency bands, discussing their interpretations and identifying the most suitable ranges for imagined speech EEG systems. The background section concluded with a brief account for selecting an appropriate vocabulary in communication BCIs. The literature review examined current approaches to communication BCIs, contrasting direct decoding methods with BCI speller systems. It also explored related work from both paradigm-specific and paradigm-agnostic perspectives, concluding with a review of techniques to address data scarcity, one of the challenges this research seeks to address.

3 Methodology

This chapter is organised into four main sections. First, the datasets used in this research are introduced, along with the rationale behind their selection. The data collection procedure and setup for the Imagined Speech dataset are also described, followed by a discussion of the preprocessing techniques and their justifications. The next section covers the model architecture and hyperparameter tuning process. This is followed by a detailed explanation of the training methodology. Finally, the evaluation strategy is presented. A visual overview of the complete methodology is shown in Figure 3.1.

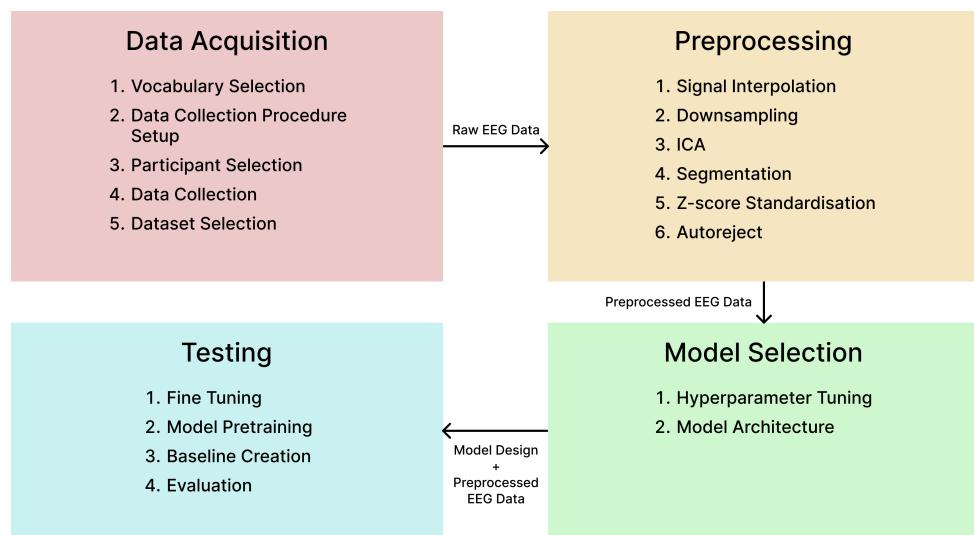


Figure 3.1 Methodology Overview. (Source: Self)

3.1 Datasets

3.1.1 Imagined Speech Dataset

Vocabulary

Before data collection, a fixed vocabulary of 20 words was selected to simplify the process and maximise the number of samples per word. As discussed in Section 2.1.5, the vocabulary needed to maximise both utility and expressiveness, allowing users of the BCI to communicate effectively despite a limited word set.

To meet these criteria, the selected words were drawn from the Universal Core Vocabulary [98], developed by the Center for Literacy and Disability Studies at the University of North Carolina. This widely adopted set provides high-frequency, functional vocabulary designed for users of AAC systems and is appropriate for diverse

contexts and age groups, making it a strong foundation for a communication-focused BCI.

From this source, a 20-word subset, as seen in Table 3.1, was chosen to further prioritise frequency and practical utility, while ensuring a diverse mix of word types such as pronouns, question words, and interjections. This helps maximise the communicative power of the system. The selected words are also all phonologically distinct, maximising class separation, potentially improving performance.

I	More	What	Who	Help
You	Want	When	On	Stop
Yes	Do	Where	In	Give
No	Put	Why	Get	Come

Table 3.1 Selected Vocabulary. (Source: Self)

Data Collection

Since this research prioritises utility and accessibility, the chosen EEG device had to reflect these goals. For this reason, a commercially available EEG device requiring minimal set-up was selected. Data collection was therefore performed using the Emotiv EPOC X EEG headset, a device with 14 EEG channels and an additional 4 reference channels, with electrodes placed in accordance with the 10-20 system [99]. The Emotiv EPOC X's electrode placement is shown using a 10-20 layout in Figure 3.2. The use of Emotiv devices in research has also been validated, indicating their viability [100].

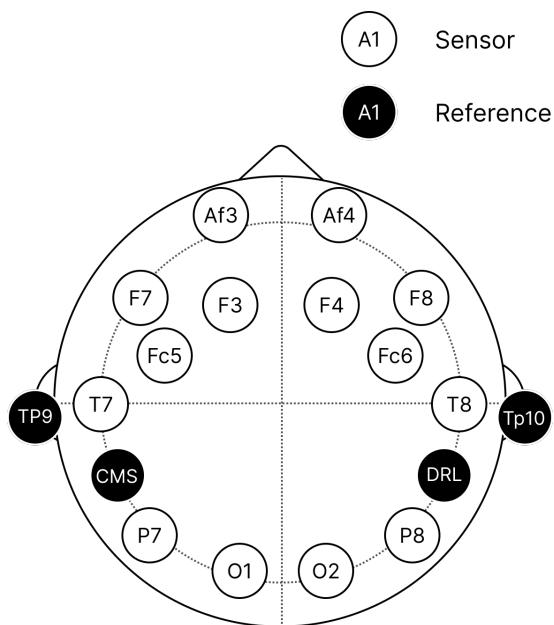


Figure 3.2 Emotiv EPOC X Electrode Placement. (Source: Self)

The data collection procedure was designed using PsychoPy [101, 102], an open source software package built in python, for designing and running experiments in psychology, neuroscience, and psychophysics.

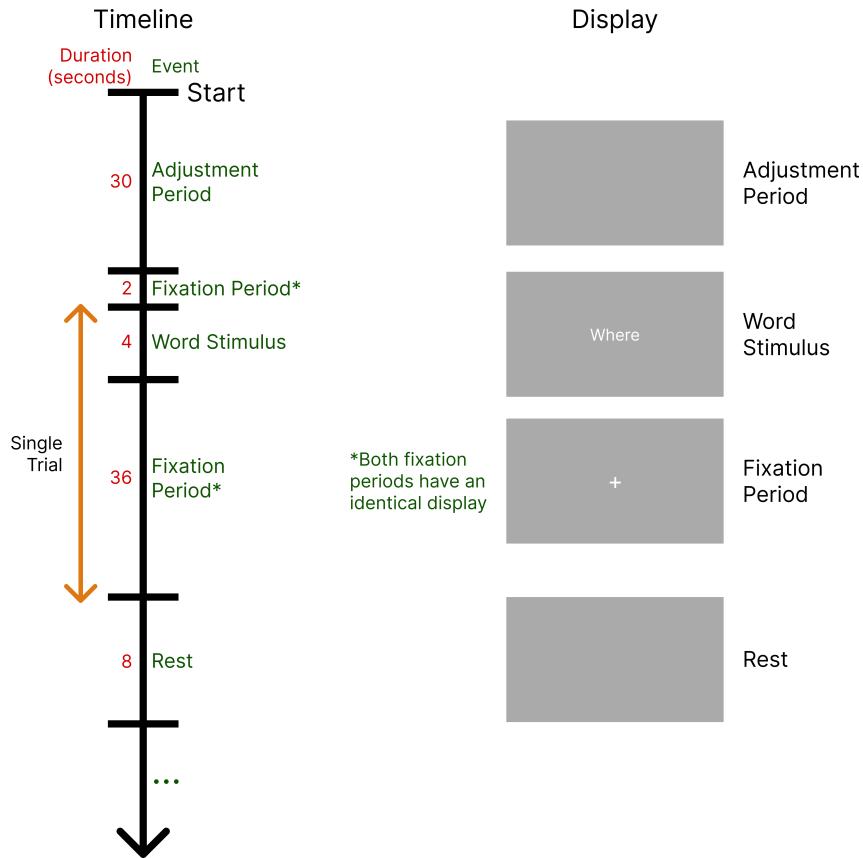


Figure 3.3 Timeline of the data collection procedure, and corresponding visual stimuli.
(Source: Self)

The EEG device was placed on each participant's head to record their brain activity during expressive imagined speech, sampled at 256 Hz. The timeline of the data collection procedure is illustrated in Figure 3.3 and described in detail below.

Preparation Participants were seated in a quiet room in front of a screen that displayed visual prompts. The procedure was explained to ensure participant understanding and cooperation, facilitating the collection of high-quality data. Once participants indicated comfort, the EEG device was positioned and adjusted to ensure optimal signal quality. A brief acclimatisation period followed, during which participants became familiar with the device. EEG recording began once this setup phase was complete.

Procedure Only visual stimuli were used in the procedure to avoid contamination of the EEG signal due audio stimuli [43]. Participants were instructed to remain as still as

possible, limiting movement to designated rest periods. They were also asked to keep their eyes open, allowing natural blinking, while focusing on the centre of the screen. This reduced both eye movement artifacts and the occurrence of alpha waves, which are neural oscillations that become more prominent when individuals close their eyes and can negatively impact the SNR in the EEG data [103]. During the trial period (stimulus presentation and post-stimulus fixation) participants were instructed to repeat the provided word mentally, with no articulation.

The full procedure consisted of the following stages:

1. **Adjustment Period** (30 seconds): Participants took time to relax, adjust their posture, and prepare for the task.
2. **Initial Fixation** (2 seconds): A fixation cross (+) appeared, signalling that a stimulus was to be shown shortly and prompting the participant to focus.
3. **Stimulus Presentation** (4 seconds): After the initial fixation, a word (e.g. *Where*) replaced the fixation cross. Participants began silently repeating the word internally. A marker was placed in the EEG stream to denote stimulus onset for later segmentation.
4. **Post-Stimulus Fixation** (36 seconds): After stimulus presentation, the fixation cross (+) reappeared, replacing the word. Participants continued repeating the word mentally while fixating to minimise eye movement artifacts and any visual stimuli.
5. **Rest Period** (8 seconds): After the post-stimulus fixation, the screen went blank, signalling a rest phase during which participants could adjust their posture.
6. **Repetition of Trials**: Steps 2–5 were repeated until each word in the vocabulary was seen **once**. This resulted in 40 seconds of continuous EEG data (4 seconds with the word, 36 seconds post-stimulus) per word. The full procedure lasted approximately 17 minutes. Each participant was shown the words in a random order to ensure that any bias relating to the ordering of words is mitigated.

Data was collected voluntarily from 21 participants (15 male), all of whom were proficient English speakers and used English as their primary language. Among them, one participant was left-handed. The age of participants ranged from 20 to 66 years, with a mean age of 42.7 and a standard deviation of 17.3. The EEG data collection protocol was approved by the Ethics Committee of the Faculty of Information and Communication Technology at the University of Malta. All participants were sufficiently informed and provided informed consent prior to data collection. Further details are provided in Appendix F.

3.1.2 Publicly Available Datasets

Two publicly available EEG datasets were selected for model pre-training: the BCI2000 Motor Movement/Imagery Dataset [104, 105], hosted on PhysioNet [106], and the EEG Dataset for Natural Image Recognition through Visual Stimuli [107], available on Mendeley.

These datasets were chosen to represent distinct cognitive task domains, MI and VEP which are well-established paradigms in EEG research and known to elicit rich, distinct neural responses [108, 109]. Additionally, gathering consistent data for both MI and VEP is easier than collecting SI data. MI tasks involve participants imagining specific movements, such as hand or foot movements, which reliably activate motor-related brain regions. Furthermore, MI is a common and well studied paradigm, with a larger number of available datasets. Similarly, VEPs are elicited through visual stimuli like flashing lights or images, and result in more consistent, EEG responses. In contrast, speech imagery involves complex cognitive processes without overt behaviour, making it challenging to ensure participant compliance and consistency.

Beyond their physiological relevance, these specific datasets were prioritised due to their technical compatibility with the constraints of the experimental setup, particularly regarding sampling rate, EEG channel configuration, and consistent trial structure. Although neither dataset directly relates to SI, their distinct task domains and technical suitability support the research objective of evaluating transfer learning from unrelated yet common cognitive tasks. Further details are provided in the following sections, with a summary provided in Table 3.2.

BCI2000 EEG Motor Movement/Imagery Dataset

The BCI2000 EEG Motor Movement/Imagery dataset [104, 105] is a MI dataset collected from 109 volunteers, resulting in approximately 1,500 recordings, each lasting 1–2 minutes. Data was recorded using the BCI2000 system with 64 EEG channels at a sampling rate of 160 Hz. The data collection procedure involved four tasks:

1. A target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then the subject relaxes. (Motor movement)
2. A target appears on either the left or the right side of the screen. The subject imagines opening and closing the corresponding fist until the target disappears. Then the subject relaxes. (Motor Imagery)
3. A target appears on either the top or the bottom of the screen. The subject opens and closes either both fists (if the target is on top) or both feet (if the target is on

the bottom) until the target disappears. Then the subject relaxes. (Motor movement)

4. A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then the subject relaxes. (Motor Imagery)

The dataset includes five classes: rest (neutral state), clench left fist, clench right fist, clench both fists, and clench both feet. Each task was repeated over several trials lasting around 4 seconds each. The dataset was selected for its large participant pool, offering the potential for learning more generalisable features. The rest class was excluded in this research due to its tendency to harm model learning, by leading to overfitting and reduced performance.

Dataset for natural image recognition through Visual Stimuli

The EEG Dataset for Natural Image Recognition through Visual Stimuli [107], is a VEP-based dataset collected from 32 participants. EEG recordings were acquired using the Emotiv EPOC X headset at a sampling rate of 128 Hz, the same equipment used to collect the imagined speech dataset. Prior to data collection, participants were screened using the Vividness of Visual Imagery Questionnaire (VVIQ), which assessed their ability to visualise mental imagery. Participants who scored below a predefined threshold were excluded to ensure high data quality.

During the recording sessions, participants viewed a static image displayed on a white background for 60 seconds. Images belonged to one of four object classes: Apple, Car, Flower, or Human Face. Each participant viewed images from only two of the four classes. EEG responses to the visual stimuli were recorded to capture evoked potentials associated with object recognition.

3.1.3 Data Preprocessing

Preprocessing procedures were tailored to each dataset to account for differences in format, channel configuration, and signal quality. However, a number of common steps were applied across all datasets to ensure compatibility with the EEG-Deformer architecture.

For the Imagined Speech Dataset, preprocessing began with correcting mislabelled or unmarked segments identified during data collection. Channels showing persistent signal loss or poor connectivity were flagged for later interpolation. Similarly, in all datasets, any segments or channels marked as problematic were interpolated.

Dataset	Participants	Sampling Rate (Hz)	EEG Channels	Num Classes	Task
Imagined Speech Dataset (Ours)	21	256	14	20	Imagined Speech (SI)
BCI2000 EEG Motor Movement/ Imagery Dataset	109	160	64	5 (4 used)	Motor Imagery (MI)
Dataset for natural image recognition through Visual Stimuli	32	128	14	4	Visually Evoked Potentials (VEP)

Table 3.2 Summary of used Datasets. (Source: Self)

The BCI2000 dataset, which includes a larger set of EEG channels than the EEG device used in this study, was downsampled to include only electrodes common to both systems. Channel names were standardised to ensure consistency during feature extraction and downstream processing.

All datasets were then resampled to a common sampling rate of 128 Hz to match the requirement of fixed-sized inputs. A band-pass filter (2–50 Hz), as discussed in Section 2.1.3, was applied to isolate relevant EEG frequencies while removing low-frequency drift and high-frequency noise. ICA was performed to reduce artifacts, particularly those related to eye movement and muscle activity.

Since the Imagined Speech and Visual Stimuli datasets contained trials longer than 4 seconds, a sliding window segmentation strategy was employed using a 4-second window with a 2-second step. This approach, described in Section 2.2.3 and adopted from the EEG-Deformer paper [81], increases the number of training samples while preserving temporal structure.

To further mitigate motion-related artifacts, the Autoreject algorithm [110] was used for automated artifact rejection and reconstruction. Parameters were set to interpolate up to one bad channel per segment and required an 85% consensus threshold to retain segments. This ensured high data retention while discarding segments with irrecoverable noise. This follows the approach used in similar research [111].

Finally, z-score standardisation was applied on a per-participant basis to reduce inter-subject variability. As shown in Equation 3.1, the standardised data $StandardisedData_S$ for participant S is calculated by subtracting the mean μ_s and dividing by the standard deviation σ_s of their EEG signals:

$$StandardisedData_S = \frac{x_S - \mu_S}{\sigma_S} \quad (3.1)$$

	Baseline Model	Pre-Trained Model
Learning Rate	5e-3	5e-3
Batch Size	32	32
Dropout	0.25	0.25
Loss Function	Cross-Entropy	Cross-Entropy
Regularisation	Label Smoothing $\epsilon = 0.1$	Label Smoothing $\epsilon = 0.1$
	On Plateau	On Plateau
Lr Decay	Patience = 10 Factor = 0.65	Patience = 10 Factor = 0.65
Temporal Kernel Size	15	11
Number of Kernels	128	128
Deformer Depth	8	6
Transformer Heads	12	12
FFN neurons	64	24
Transformer Head Size	40	64

Table 3.3 Summary of Model Hyperparameters. (Source: Self)

3.2 Model Selection

The EEG-Deformer architecture [81] was used in this experiment due to its open-source availability and paradigm-agnostic design, which allows it to generalise across the chosen datasets. A detailed description of the model and comparison to other models is provided in Section 2.2.2. Hyperparameters were optimised using Optuna [112], an automated hyperparameter tuning framework. Optimisation was conducted on subsets of all datasets for key hyperparameters. A summary of the selected hyperparameters and their values is provided in Table 3.3.

3.3 Model Training

Model training was divided into two stages to evaluate both within-subject and cross-subject performance, further detailed in Section 3.4. A baseline model was trained from scratch on the Imagined Speech dataset twice: once to evaluate within-subject performance, and the second time to evaluate cross-subject performance. In addition, two pre-trained models were developed: one using the BCI2000 Motor Imagery dataset and the other using the Visual Stimuli dataset. Each pre-trained model was instantiated once and then fine-tuned separately for within-subject and cross-subject evaluations.

3.3.1 Pre-training

Both pre-training datasets were split into 80% training and 20% validation sets. Models were trained for up to 1000 epochs with early stopping triggered if validation loss did not improve over 25 consecutive epochs. A label smoothing value of $\epsilon = 0.1$ was applied to reduce overfitting and improve generalisation. Additionally, if validation loss stagnated for 10 epochs, the learning rate was reduced by 35% to allow the model to learn finer details. Label smoothing is a regularisation technique that replaces the hard target labels with a softened version to prevent the model from becoming overconfident [113]. Instead of assigning a probability of 1 to the correct class and 0 to others, a small value ϵ is distributed among all classes, as demonstrated in Equations 3.2 and 3.3.

$$y_{smooth} = (1 - \epsilon) \text{ for the true class} \quad (3.2)$$

$$y_{smooth} = \frac{\epsilon}{k - 1} \text{ for all other classes} \quad (3.3)$$

Where k is the number of classes in the dataset.

3.3.2 Within-subject training

For within-subject evaluation, models were trained on the Imagined Speech dataset using 5-fold cross-validation, increasing confidence in the reported results. Each fold served once as a test set, while the remaining four were merged and further split 85:15 into training and validation sets. This ensured each model was exposed to all 21 participants, evaluating its ability to generalise within known subjects. Training followed a similar set-up to pre-training in terms of early stopping, learning rate reduction and label smoothing.

A baseline model was trained from scratch on this dataset, while the pre-trained models were fine-tuned for comparison. Due to differing output dimensions across datasets, the pre-trained models had their classification heads replaced with a new head containing 20 outputs. During fine-tuning, a learning rate of 5e-3 was applied to the new head and 6.25e-4 to the remaining layers.

To ensure fair data representation, both subject and class distributions were preserved proportionally in all training, validation, and test splits. For example, 85% of data from a specific subject-class combination (e.g., subject 4, word “Where”) was guaranteed in the training set if an 85:15 split was used.

3.3.3 Cross-subject training

To assess cross-subject generalisation, a LOSO strategy was used: models were trained on 20 participants and tested on the held-out participant. Within the 20 training participants, an 80:20 split was used for training and validation.

As with the within-subject setup and model-pretraining, all models were trained for up to 1000 epochs with early stopping, label smoothing of $\epsilon = 0.1$, and adaptive learning rate reduction. For further analysis, 5-fold cross-validation was also applied to the training set for a subset of participants, to provide further insights into the model's performance.

Again, pre-trained models were fine-tuned with replaced classification heads (20 outputs), using learning rates of 5e-3 and 6.25e-4 for the head and remaining layers, respectively. Baseline models were trained from scratch for direct comparison.

3.4 Evaluation Strategy

Once the models were trained, they were evaluated using two performance in two different scenarios, as mentioned in Section 3.3:

- **Within-subject performance:** This evaluates how well the model learns to recognise brain activity patterns from individuals it has seen during training. It involves training on data from **all** participants and testing on held-out data from the **same** participants. Results are further validated through 5-fold cross-validation.
- **Cross-subject performance:** This assesses the model's ability to generalise to unseen individuals by using a LOSO strategy. The model is trained on data from **all but one** participant and tested on the **excluded** participant.

Since the task is framed as a multi-class classification problem, two evaluation metrics are used: accuracy and macro-averaged F1-score (F1-macro). Accuracy, defined in Equation 3.4, provides an overall measure of correct predictions, while F1-macro (Equation 3.8) offers a more balanced assessment in cases of class imbalance by giving equal weight to each class.

F1-macro is calculated as the mean F1-score across all classes. The F1-score (Equation 3.7) is the harmonic mean of precision and recall. Precision (Equation 3.5) calculates the proportion of correct predictions within a predicted class, while recall (Equation 3.6) measures how well the model identifies all instances of a class.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.6)$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.7)$$

$$\text{F1-macro} = \frac{1}{K} \sum_{i=1}^K \text{F1}_i \quad (3.8)$$

In the equations:

TP (True Positives) refers to the number of instances correctly classified as belonging to a specific class.

FP (False Positives) refers to the number of instances incorrectly classified as that class.

FN (False Negatives) refers to the number of instances that belong to the class but were misclassified as something else.

TN (True Negatives) refers to the number of instances correctly identified as not belonging to the class in question

K refers to the number of classes, in this case the 20 words shown in Table 3.1.

3.5 Summary

In this chapter we outlined the acquisition and preparation of data for model training, including the data collection procedure, dataset selection, and preprocessing methods. We then described the model architecture, the selected hyperparameters, and the training process. Finally, the evaluation methodology was presented, along with the metrics used to assess model performance across within-subject and cross-subject scenarios.

4 Evaluation

The primary aim of this research was to develop an ML-based system capable of decoding EEG signals captured during expressive imagined speech, with a focus on within-subject performance. In addition to this, the research pursued two secondary aims:

- Developing a model that generalises to individuals outside the training set (cross-subject performance).
- Investigating the effectiveness of cross-task transfer learning as a strategy to overcome data scarcity, by evaluating improvements over a baseline in both within-subject and cross-subject contexts.

This chapter evaluates the performance of the trained models, discusses the resulting findings, and determines the extent to which the stated aims have been achieved.

4.1 Within-Subject Performance

When evaluating within-subject performance, the baseline model achieved the highest mean accuracy of 84.1%, with a corresponding F1-macro score, converging after an average of 295 epochs. In contrast, the pre-trained models underperformed: the MI-pretrained model reached an accuracy of 54.0%, while the VEP-pretrained model achieved a more competitive 77.5%. The F1-macro scores closely followed the accuracy values, suggesting relatively uniform performance across classes.

Notably, both pre-trained models converged more quickly than the baseline. The MI-pretrained model converged in an average of 271 epochs, and the VEP-pretrained model in 259 epochs, on average reducing training time by 24 and 36 epochs, respectively. The confidence intervals for convergence epochs were estimated using a two-tailed *t*-distribution-based method, appropriate for small sample sizes ($n = 5$), but inherently more sensitive to variance. As such, the limited number of training runs may reduce the reliability of these estimates. Moreover, the time saved through faster convergence was outweighed by the additional cost of pre-training, and the observed reduction in accuracy may ultimately outweigh any practical benefits.

To further quantify the performance differences between the models, model predictions were bootstrapped 5000 times to estimate confidence intervals for the drop in accuracy between each pre-trained model and the baseline. This provides an estimate of the distribution of performance differences. The resulting 95% confidence intervals did not include zero, indicating that the decreases in accuracy for both the MI-

Metric	Baseline	MI Pretrained	VEP Pretrained
Mean Accuracy	0.841	0.540	0.775
Confidence Interval (Accuracy)	0.833–0.849	0.529–0.550	0.766–0.784
Mean decrease in accuracy from baseline	N/A	0.302	0.066
Confidence Interval (decrease from baseline)	N/A	0.287–0.315	0.054–0.078
Mean F1-Macro	0.841	0.539	0.775
Confidence Interval (F1-Macro)	0.833–0.849	0.528–0.550	0.766–0.784
Mean Epochs	295	271	259
Confidence Interval (Epochs)	259–331	222–321	232–287

Table 4.1 Within-Subject Performance. (Source: Self)

and VEP-pretrained models were statistically significant. The full results are summarised in Table 4.1. Confidence intervals for the accuracy and the f1-macro were similarly calculated by bootstrapping for 5000 times.

All models demonstrate performance significantly exceeding chance levels (5%). However, it is important to assess success based on the preferences of potential users of communication BCI systems. A survey conducted with individuals who have spinal cord injuries revealed that over half of the participants would be willing to use a BCI if the model achieved an accuracy of at least 70%, and 77% of respondents would be willing if the accuracy was at least 80% [36]. The reported results fulfil the primary aim of this research for both the baseline and VEP-pretrained models, as their confidence intervals exceed this threshold. In contrast, a similar survey of individuals with ALS found that only 49% of participants were willing to use a model that achieved at least 80% accuracy [37]. This suggests that, while the baseline model may suffice for some users, further improvements are necessary to meet the needs of all potential users.

The relatively strong performance of the VEP-pretrained model, which approached that of the baseline, may suggest that more transferable features exist between VEP and SI tasks than between MI and SI tasks. While speech imagery includes speech motor planning [11], which involves activity in the motor and premotor cortical regions, similar to MI, the expected alignment was not observed. In contrast, although occipital activation is typically less pronounced during SI [11], distancing it from VEP at a neuroanatomical level, many participants reported that the prolonged repetition of words during the data evoked mental imagery. This effect may have been exacerbated by the relatively long trial durations, which could allow attention to drift toward visual representations of the imagined words. Given that VI uses similar neural substrates to those engaged during VEP [114], this could account for the improved performance of the VEP-pretrained model.

However, it is important to interpret this result with caution. Both the VEP

dataset and the imagined speech dataset were recorded using similar EEG devices, which may reduce variability and inadvertently increase compatibility between the source and target domains. This device-level consistency could partially explain the improved performance, independent of true task-level feature transfer. Further research is needed to disambiguate these effects and to assess transferability across a broader range of source tasks.

Beyond general task misalignment, the drop in performance when pretraining may be attributed to inter-subject variability, due to the simple approach taken to fine-tuning, as opposed to more sophisticated techniques such as those described in Section 2.2.3. The reduction may also be due to the neurological differences between the tasks, as discussed above and in Section 2.1.4. SI is incredibly complex, involving both motor planning, and more complex patterns, such as a feedback loop between the auditory and motor cortices [115], and greatly differs to both MI and VEP, both in the type of signal and its localisation in the brain.

Overall, the results provide limited evidence for the effectiveness of cross-task transfer learning in a within-subject setting. While one of the secondary aims was to investigate whether pretraining on related tasks could help overcome data scarcity, the observed decrease in accuracy relative to the baseline suggests that, in this case, pretraining did not lead to improved performance for within-subject use.

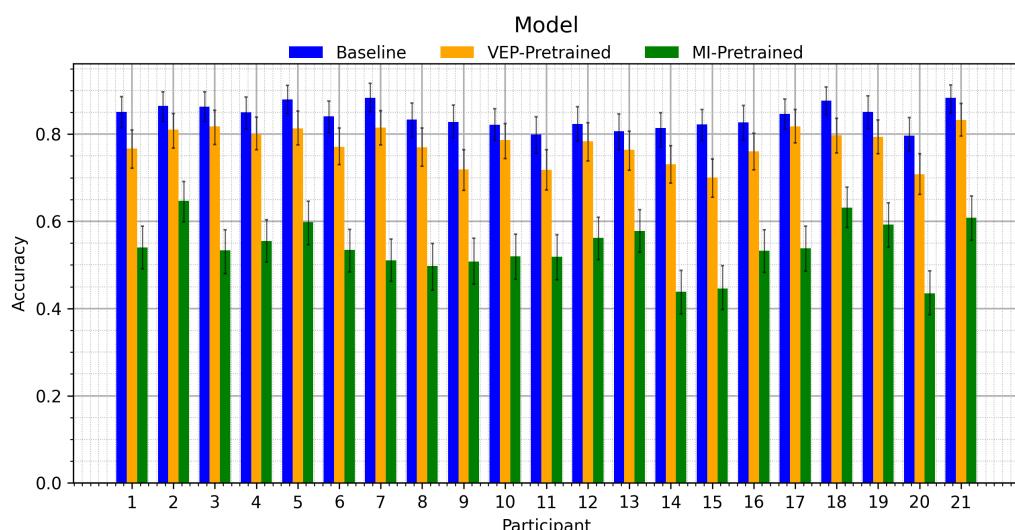


Figure 4.1 Accuracy Per Participant (Within-Subject). Error bars represent a confidence interval of 95%. (Source: Self)

To further assess the model's performance, accuracy was evaluated for each participant, as seen in Figure 4.1. The baseline model demonstrated stable performance across all participants, with the lowest accuracy being 79.8% (participant 20) and the highest being 88.5% (participant 7). This consistent performance increases confidence in the model's ability to generalise across different subjects in a

within-subject context. Similarly, the VEP-pretrained model exhibited stable performance, with minimum and maximum accuracies of 70.0% (participant 15) and 83.3% (participant 21), respectively. In contrast, the MI-pretrained model showed greater variance, with its minimum accuracy at 43.5% (participant 20) and its maximum at 64.6% (participant 2).

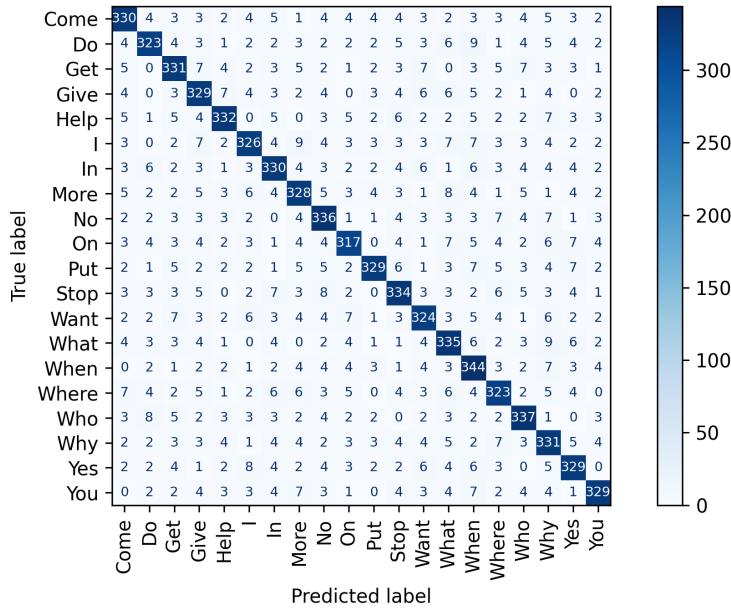


Figure 4.2 Baseline Confusion Matrix (Within-Subject). (Source: Self)

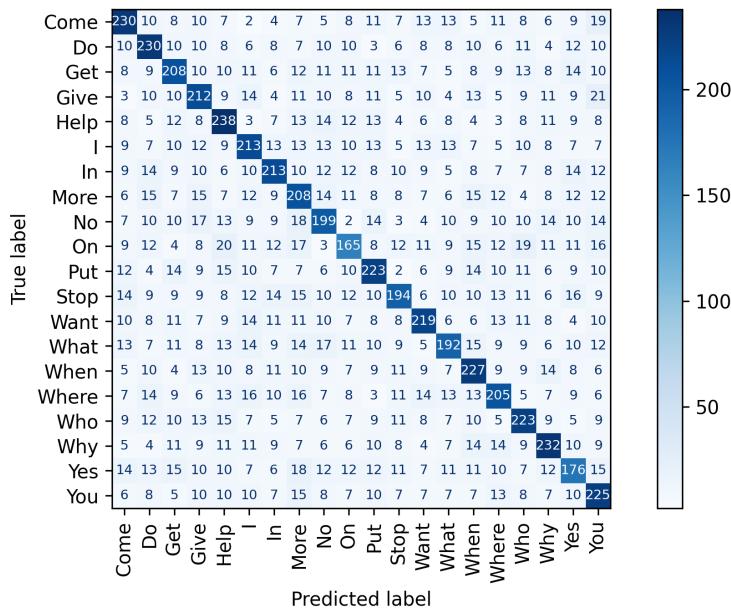


Figure 4.3 MI-Pretrained Confusion Matrix (Within-Subject). (Source: Self)

To better understand the model's performance across different classes, confusion matrices are presented in Figures 4.2–4.4. The confusion matrices reveal that the models perform relatively uniformly across the set of target words. The word

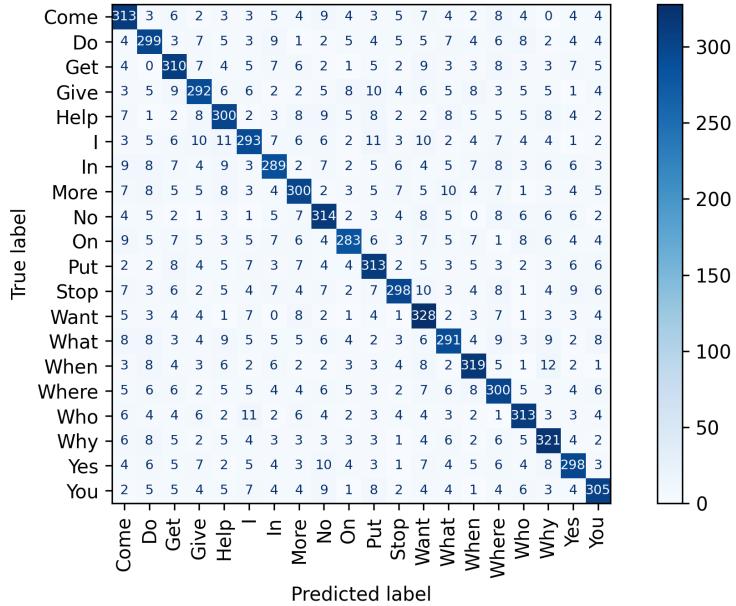


Figure 4.4 VEP-Pretrained Confusion Matrix (Within-Subject). (Source: Self)

'On' consistently shows the lowest accuracy across all models, while the best-performing words vary by model: 'When' for the baseline, 'Help' for the MI-pretrained model, and 'Want' for the VEP-pretrained model. Additionally, the misclassifications appear to be evenly distributed rather than concentrated among specific word pairs, suggesting that the models are generally effective at distinguishing between the different classes. Notably the only two phonologically similar words 'do' and 'who' show no indication of being confused with one another.

4.2 Cross-Subject Performance

When evaluating cross-subject performance, the baseline model achieved the highest mean accuracy of 5.4%, with a corresponding F1-macro score, and converged after an average of 307 training epochs. This performance is markedly lower than within-subject results, suggesting the model is likely overfitting to individual participants. The pre-trained models performed slightly worse: the MI-pretrained model reached a mean accuracy of 5.0%, and the VEP-pretrained model achieved 4.8%. As with the within-subject evaluation, F1-macro scores closely mirrored accuracy values, indicating a relatively uniform distribution of performance across all classes. Importantly, this evaluation was conducted using a strict LOSO approach, ensuring that the test participant's data was entirely excluded from training and validation, thereby preventing any data contamination and enabling a clean assessment of generalisation.

The MI-pretrained model's performance aligns exactly with the theoretical chance level of 5%, while the VEP-pretrained model falls slightly below it. Although the

Metric	Baseline	MI Pretrained	VEP Pretrained
Mean Accuracy	0.054	0.05	0.048
Confidence Interval (Accuracy)	0.049–0.059	0.046–0.055	0.043–0.052
Mean change in accuracy over chance level	0.004	0.0	-0.002
Confidence Interval (Change over chance level)	-0.001–0.009	-0.004–0.005	-0.007–0.002
Mean F1-Macro	0.054	0.05	0.047
Confidence Interval (F1-Macro)	0.049–0.059	0.045–0.055	0.042–0.052
Mean Epochs	307	262	272
Confidence Interval (Epochs)	291–323	247–277	258–286

Table 4.2 Cross-Subject Performance. (Source: Self)

baseline model exceeds chance by 0.4% on average, this difference is not statistically significant, as the lower bound of its confidence interval includes chance level performance. These results suggest that none of the models generalise effectively in a cross-subject context. A summary of these findings is provided in Table 4.2.

The results indicate that the secondary aim, developing a model capable of generalising to participants not seen during training, was not achieved. Model performance in the cross-subject setting remained far below levels typically deemed acceptable by potential BCI users and did not significantly exceed chance level. Furthermore, pretraining with external datasets did not result in any measurable improvement in this context.

The drop in accuracy from the within-subject to the cross-subject setting suggests several potential causes, the most likely being model overfitting to subject-specific features in the training data. Performance degradation in cross-subject decoding is not uncommon in EEG research, and prior work has explored techniques such as transfer learning to address inter-subject variability (see Section 2.2.3). In this study, the decline may also have been influenced by the experimental setup, including relatively long trial durations, which could introduce greater variability due to participant engagement and mental state.

Another contributing factor may be variability in electrode positioning. The EEG device used in this study featured a fixed-layout design rather than an individually fitted cap. As a result, electrode placement likely varied across participants in relation to underlying cortical structures. While within-subject models benefit from consistent placement for a single individual, cross-subject models rely on inter-participant spatial consistency. Previous studies have shown that even with standardised caps, electrode positions can differ significantly [116, 117], potentially leading to differences in the recorded neural signals and impairing model generalisation [118].

To further assess model performance, accuracy was evaluated individually for

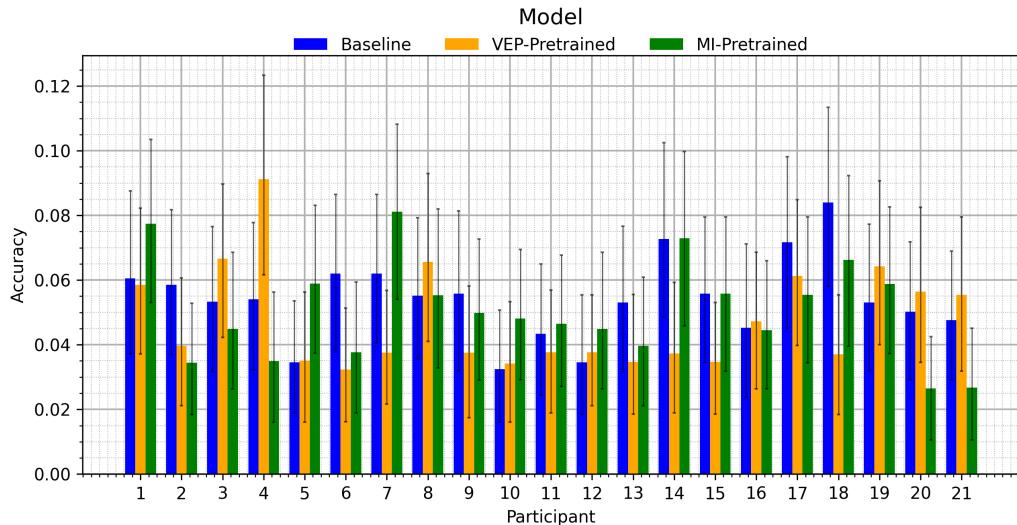


Figure 4.5 Accuracy Per Participant (Cross-Subject). Error bars represent a confidence interval of 95%. (Source: Self)

each participant, as seen in Figure 4.5. In contrast to within-subject results, cross-subject accuracy exhibited relatively higher variability across participants for all models. The baseline model produced participant-specific accuracies ranging from 3.2% (participant 10) to 8.4% (participant 18). The VEP-pretrained model showed a comparable minimum accuracy of 3.2% (participant 6), with the highest overall accuracy of 9.1% (participant 4). The MI-pretrained model yielded the lowest overall accuracy of 2.7% (participant 21) and a maximum of 8.1 (participant 7). Although these results highlight participant-level fluctuations, the uniformly low accuracies across all models limit the extent to which meaningful subject-specific patterns or conclusions can be drawn.

4.3 Summary

In this chapter we presented and analysed the experimental results. While the primary aim of achieving reliable within-subject performance was met, the secondary aim, generalising to unseen participants, was not achieved. Additionally, transfer learning did not yield improvements in either within-subject or cross-subject settings. Participant-level analyses further illustrated consistent within-subject performance and increased variability in the cross-subject context.

5 Conclusion

5.1 Revisiting the aims and objectives

The primary aim of this research was to develop a ML-based system capable of decoding EEG signals recorded during expressive imagined speech. Two secondary aims were also defined: to enable the model to generalise to participants unseen during training, and to investigate whether cross-task transfer learning could improve performance in both within-subject and cross-subject scenarios. To address these aims, the following objectives were defined at the outset of the project:

1. **Data Collection and Preparation:** A 20-word vocabulary was selected for expressive utility and practical use, and best practices from literature were employed to design a high-quality experimental protocol. EEG data were successfully collected from 21 distinct participants.
2. **Dataset Selection:** Two publicly available EEG datasets (MI and VEP) were identified and prepared for pretraining, chosen for their distinction from imagined speech.
3. **Model Development:** The EEG-Deformer model was selected for its paradigm agnostic design, ideal for cross-task transfer learning. It combines shallow convolutional feature extraction with HCT blocks to model long-range temporal dependencies.
4. **Model Training:** Models were pretrained on the selected MI and VEP datasets, and subsequently fine-tuned on the collected imagined speech dataset to assess cross-task transfer potential. A baseline model was also trained from scratch on the same imagined speech data for direct comparison.
5. **Evaluation and Validation:** Model performance was comprehensively evaluated in both within-subject and cross-subject scenarios. The pretrained models were compared against the baseline, revealing no significant improvement through transfer learning.

This research successfully met its primary aim, with the baseline model achieving 84.1% within-subject accuracy for imagined speech decoding. However, secondary aims remained unfulfilled: the model failed to generalise to unseen participants, performing near chance in cross-subject evaluation, and transfer learning did not yield improvements. Nevertheless, the results suggest task alignment influences transfer effectiveness, exemplified by the VEP-pretrained model's 77.5%

within-subject accuracy, in contrast to the MI-pretrained model's 54.0%. These findings highlight avenues for future exploration of task similarity and transfer learning.

5.2 Critiques and Limitations

Despite the contributions of this research, several limitations constrained its scope and potential impact.

Data Collection and Experimental Design

Data collection was limited in both scale and depth. The participant pool comprised only 21 individuals, potentially restricting model generalisation and obscuring broader trends in neural representations of language. Each word had a maximum of 19 samples per participant, with some discarded due to unrecoverable noise. The use of single trials per word likely reduced the robustness of learned features, while longer trial durations may have increased cognitive fatigue and introduced additional variability.

Device Constraints and Inter-Subject Variability

The study employed a 14-channel consumer-grade EEG device with fixed electrode positioning. While chosen for accessibility, this device inherently contributed to a lower SNR and limited spatial resolution compared to research-grade systems. Its fixed design also led to significant variations in electrode placement relative to underlying cortical structures across participants. These factors collectively exacerbated challenges in collecting high-quality, consistent neural data. Furthermore, these limitations significantly contributed to a higher inter-subject variability, which alongside the limited participant pool, reduced the model's ability to learn generalised features and hence the observed poor cross-subject performance.

Limited Vocabulary

The use of a closed vocabulary, particularly one limited to only 20 words, greatly reduces the applicability of this BCI for broader communicative use and the restoration of full speech for impaired users. While appropriate for an initial proof-of-concept, this constraint means the system cannot generalise to arbitrary words or phrases, a critical requirement for a truly fully functional communication BCI.

Dataset Selection and Model Training

Challenges also arose in dataset selection and model training. The use of a similar device for the VEP and SI datasets may have introduced superficial alignment not reflective of true task similarity. Compatibility issues and time constraints restricted the inclusion of other SI datasets, limiting external validation and broader comparisons. Furthermore, the deliberately simple approach to model fine-tuning might have been insufficient for mitigating inter-subject variability and domain mismatch, suggesting a need for more advanced transfer learning strategies. Finally, the system's reliance on 4-second input segments deviates from real-time BCI requirements, highlighting the need for more time-efficient decoding.

In summary, while this research explored a promising direction through cross-task transfer learning, it also faced significant challenges that limited its ability to fully demonstrate that potential. These limitations offer important guidance for future work, which will be covered in the next section.

5.3 Future Work

In light of the discussed challenges and limitations, there are several avenues for future research that could build on this work.

First, improvements in data collection would likely benefit both model performance and scientific insight. Increasing the participant pool could help models capture more generalisable features, thus improving cross-subject performance. Additionally, collecting more data per participant, such as through shorter, more frequently repeated trials, potentially across multiple sessions, may reduce trial-specific overfitting, intra-trial drift and improve model performance. While the use of a consumer-grade EEG device supported the goal of accessibility, replicating this study with a research-grade EEG system may offer a lower SNR and enable finer-grained spatial analyses and deeper insights.

Exploring a broader range of pretraining datasets may also help identify which tasks yield transferable features to imagined speech decoding. In particular, the use of publicly available speech imagery datasets could serve as a valuable same-task pretraining baseline, enabling future work to compare cross-task transfer learning against task-aligned pretraining approaches.

Among the various directions for future work, more sophisticated transfer learning methods, particularly domain-adversarial approaches such as those implemented in MTDANN [97], appear especially promising. These methods directly address inter-subject variability and potentially cross-task variability, both major limitations identified in this study.

More broadly, the current within-subject accuracy of 84% remains below the performance expectations of many potential communication-oriented BCI users. Achieving accuracies closer to 90% remains an important target for practical deployment. Similarly important is narrowing the gap in cross-subject performance, which may require not only larger and more diverse datasets but also methods that explicitly target inter-subject variability. In addition, moving toward more real-time decoding is imperative for enabling responsive, interactive systems. Finally, expanding the vocabulary beyond 20 words would significantly improve the expressive capacity of such models, bringing them closer to real-world communication needs.

5.4 Availability of Data

In the spirit of reproducibility and open-source technology, the code used in this project and the collected dataset are made publicly available on GitHub.

Code: <https://github.com/JeremyFarrugia/FYP>

Dataset: <https://github.com/JeremyFarrugia/Speech-Imagery-Dataset>

5.5 Final Remarks

Numerous studies have attempted to address data scarcity in EEG research through techniques such as cross-subject and cross-dataset transfer learning. However, to our knowledge, this is the first study to explore cross-task transfer learning in the context of EEG, specifically for speech imagery decoding. In doing so, this work introduces a novel direction for future research that may help address the challenge of limited task-specific data, particularly in domains where data collection is inherently difficult, such as imagined speech.

This thesis achieved 84.1% accuracy in decoding a 20-word imagined speech vocabulary in within-subject context, validating both the data collection methodology and the EEG-Deformer's applicability to SI tasks. Moreover, our findings suggest that some tasks are more naturally aligned than others in a transfer learning setting, which may inform future model pretraining strategies.

While there remain substantial challenges, particularly in achieving generalisable performance across individuals, this research lays a foundation for further investigations into more efficient, scalable, and versatile EEG-based BCIs. With continued progress, such systems may one day become powerful tools for restoring communication in individuals with severe speech or motor impairments, including those with conditions such as LIS, ALS or ASD.

References

- [1] L. Xu *et al.*, "Global variation in prevalence and incidence of amyotrophic lateral sclerosis: A systematic review and meta-analysis," *Journal of Neurology*, vol. 267, no. 4, pp. 944–953, 2020. DOI: 10.1007/s00415-019-09652-y.
- [2] P. Couratier, P. Corcia, G. Lautrette, M. Nicol, P.-M. Preux, and B. Marin, "Epidemiology of amyotrophic lateral sclerosis: A review of literature," *Revue Neurologique*, vol. 172, no. 1, pp. 37–45, 2016, *Neuroepidemiology*, ISSN: 0035-3787. DOI: <https://doi.org/10.1016/j.neurol.2015.11.002>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0035378715009285>.
- [3] M. L. Hordila, C. García-Bravo, D. Palacios-Ceña, and J. Pérez-Corrales, "Locked-in syndrome: A qualitative study of a life story," *Brain and Behavior*, vol. 14, no. 8, e3495, 2024. DOI: <https://doi.org/10.1002/brb3.3495>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/brb3.3495>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/brb3.3495>.
- [4] Orphanet, *Locked-in syndrome*, <https://www.orpha.net/en/disease/detail/2406>, Accessed: 2025-05-06, 2024.
- [5] L. Schnetzer, M. McCoy, J. Bergmann, A. Kunz, S. Leis, and E. Trinka, "Locked-in syndrome revisited," *Therapeutic Advances in Neurological Disorders*, vol. 16, p. 17562864231160873, 2023. DOI: 10.1177/17562864231160873. [Online]. Available: <https://doi.org/10.1177/17562864231160873>.
- [6] T. Halan, J. F. Ortiz, D. Reddy, A. Altamimi, A. O. Ajibowo, and S. P. Fabara, "Locked-in syndrome: A systematic review of long-term management and prognosis," *Cureus*, vol. 13, no. 7, 2021.
- [7] C. Bougeard, F. Picarel-Blanchot, R. Schmid, R. Campbell, and J. Buitelaar, "Prevalence of autism spectrum disorder and co-morbidities in children and adolescents: A systematic literature review," *Frontiers in psychiatry*, vol. 12, p. 744709, 2021. DOI: 10.3389/fpsyt.2021.744709. [Online]. Available: <https://doi.org/10.3389/fpsyt.2021.744709>.
- [8] A. Fischer-Janzen, T. M. Wendt, and K. Van Laerhoven, "A scoping review of gaze and eye tracking-based control methods for assistive robotic arms," *Frontiers in Robotics and AI*, vol. Volume 11 - 2024, 2024, ISSN: 2296-9144. DOI: 10.3389/frobt.2024.1326670. [Online]. Available:

- [https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2024.1326670.](https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2024.1326670)
- [9] M. F. Mridha, S. C. Das, M. M. Kabir, A. A. Lima, M. R. Islam, and Y. Watanobe, "Brain-computer interface: Advancement and challenges," *Sensors (Basel, Switzerland)*, vol. 21, no. 17, p. 5746, 2021. DOI: 10.3390/s21175746.
 - [10] N. Nieto, V. Peterson, H. L. Rufiner, J. E. Kamienkowski, and R. Spies, "Thinking out loud, an open-access eeg-based bci dataset for inner speech recognition," *Scientific Data*, vol. 9, no. 1, p. 52, 2022. DOI: <https://doi.org/10.1038/s41597-022-01147-2>.
 - [11] A. B. Silva, J. R. Liu, L. Zhao, D. F. Levy, T. L. Scott, and E. F. Chang, "A neurosurgical functional dissection of the middle precentral gyrus during speech production," *Journal of Neuroscience*, vol. 42, no. 45, pp. 8416–8426, 2022. DOI: 10.1523/JNEUROSCI.1614–22.2022.
 - [12] K. Demicoli, *Brain-to-text*, Bachelor's dissertation, B.Sc. IT (Hons)(Melit.), Msida, Malta, 2024. [Online]. Available: <https://www.um.edu.mt/library/oar/handle/123456789/127989>.
 - [13] D. Saliba, "Neuro-linguistic transcription: Text generation from eeg patterns," M.S. thesis, University of Malta, Msida, Malta, 2024.
 - [14] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012, ISSN: 1424-8220. DOI: 10.3390/s120201211. [Online]. Available: <https://www.mdpi.com/1424-8220/12/2/1211>.
 - [15] M. Tudor, L. Tudor, and K. I. Tudor, "Hans berger (1873-1941)–the history of electroencephalography," *Acta medica Croatica: casopis Hrvatske akademije medicinskih znanosti*, vol. 59, no. 4, pp. 307–313, 2005.
 - [16] J. Vidal, "Real-time detection of brain events in eeg," *Proceedings of the IEEE*, vol. 65, no. 5, pp. 633–641, 1977. DOI: 10.1109/PROC.1977.10542.
 - [17] M. Baygin *et al.*, "Automated asd detection using hybrid deep lightweight features extracted from eeg signals," *Computers in Biology and Medicine*, vol. 134, p. 104 548, 2021. DOI: <https://doi.org/10.1016/j.compbiomed.2021.104548>.
 - [18] N. Robinson, R. Mane, T. Chouhan, and C. Guan, "Emerging trends in bci-robotics for motor control and rehabilitation," *Current Opinion in Biomedical Engineering*, vol. 20, p. 100 354, 2021, ISSN: 2468-4511. DOI: <https://doi.org/10.1016/j.cobme.2021.100354>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2468451121000945>.

- [19] W. Mao, H. Fathurrahman, Y. Lee, and T. Chang, "Eeg dataset classification using cnn method," in *Journal of physics: conference series*, IOP Publishing, vol. 1456, 2020, p. 012017. DOI: 10.1088/1742-6596/1456/1/012017.
- [20] W. Pang, C. Yuan, Y. Zheng, T. Zhong, and P. Lai, "Applications over the horizon – advancements and challenges in brain-computer interfaces," *The Innovation Life*, vol. 2, no. 1, p. 100058, 2024, ISSN: 2959-8761. DOI: 10.59717/j.xinn-life.2024.100058. [Online]. Available: <https://www.the-innovation.org/life/article/id/65ecdc06bf979918c9d84fd4>.
- [21] M. J. Vansteensel and B. Jarosiewicz, "Chapter 7 - brain-computer interfaces for communicationacurrent affiliation: Neuropace, inc., mountain view, ca, united states," in *Brain-Computer Interfaces*, ser. Handbook of Clinical Neurology, N. F. Ramsey and J. del R. Millán, Eds., vol. 168, Elsevier, 2020, pp. 67–85. DOI: <https://doi.org/10.1016/B978-0-444-63934-9.00007-X>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978044463934900007X>.
- [22] W. J. Marks and K. D. Laxer, "Chapter 7 - invasive clinical neurophysiology in epilepsy and movement disorders," in *Aminoff's Electrodiagnosis in Clinical Neurology (Sixth Edition)*, M. J. Aminoff, Ed., Sixth Edition, London: W.B. Saunders, 2012, pp. 165–185, ISBN: 978-1-4557-0308-1. DOI: <https://doi.org/10.1016/B978-1-4557-0308-1.00007-8>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9781455703081000078>.
- [23] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering*, vol. 7, no. 5, p. 056007, Sep. 2010. DOI: 10.1088/1741-2560/7/5/056007. [Online]. Available: <https://dx.doi.org/10.1088/1741-2560/7/5/056007>.
- [24] A. Wilsch and J. Obleser, "What works in auditory working memory? a neural oscillations perspective," *Brain Research*, vol. 1640, pp. 193–207, 2016, Auditory Working Memory, ISSN: 0006-8993. DOI: <https://doi.org/10.1016/j.brainres.2015.10.054>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006899315008410>.
- [25] Y. Wang, X. Yang, X. Zhang, Y. Wang, and P. Weihua, "Implantable intracortical microelectrodes: Reviewing the present with a focus on the future," *Microsystems & Nanoengineering*, vol. 9, 2023, ISSN: 2055-7434. DOI:

- <https://doi.org/10.1038/s41378-022-00451-6>. [Online]. Available: <https://doi.org/10.1038/s41378-022-00451-6>.
- [26] D. Edell, V. Toi, V. McNeil, and L. Clark, "Factors influencing the biocompatibility of insertable silicon microshafts in cerebral cortex," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 6, pp. 635–643, 1992. DOI: 10.1109/10.141202.
 - [27] L. J. Szymanski *et al.*, "Neuropathological effects of chronically implanted, intracortical microelectrodes in a tetraplegic patient," *Journal of Neural Engineering*, vol. 18, no. 4, 0460b9, 2021. DOI: 10.1088/1741-2552/ac127e.
 - [28] S. C. Colachis, C. F. Dunlap, N. V. Annetta, S. M. Tamrakar, M. A. Bockbrader, and D. A. Friedenberg, "Long-term intracortical microelectrode array performance in a human: A 5 year retrospective analysis," *Journal of Neural Engineering*, vol. 18, no. 4, p. 0460d7, 2021. DOI: 10.1088/1741-2552/ac1add.
 - [29] G.-T. Go, Y. Lee, D.-G. Seo, and T.-W. Lee, "Organic neuroelectronics: From neural interfaces to neuroprosthetics," *Advanced Materials*, vol. 34, no. 45, p. 2201864, 2022. DOI: <https://doi.org/10.1002/adma.202201864>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.202201864>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202201864>.
 - [30] K. M. Szostak, L. Grand, and T. G. Constandinou, "Neural Interfaces for Intracortical Recording: Requirements, Fabrication Methods, and Characteristics," English, *Frontiers in Neuroscience*, vol. 11, Dec. 2017, Publisher: Frontiers, ISSN: 1662-453X. DOI: 10.3389/fnins.2017.00665. [Online]. Available: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2017.00665/full> (visited on 05/13/2025).
 - [31] E. C. Leuthardt, D. W. Moran, and T. R. Mullen, "Defining surgical terminology and risk for brain computer interface technologies," *Frontiers in Neuroscience*, vol. 15, p. 599549, 2021. DOI: <https://doi.org/10.3389/fnins.2021.599549>.
 - [32] M. G. Bleichner, J. M. Jansma, J. Sellmeijer, M. Raemaekers, and N. F. Ramsey, "Give me a sign: Decoding complex coordinated hand movements using high-field fmri," *Brain topography*, vol. 27, pp. 248–257, 2014. DOI: 10.1007/s10548-013-0322-x.
 - [33] M. Bleichner, J. Jansma, E. Salari, Z. Freudenburg, M. Raemaekers, and N. Ramsey, "Classification of mouth movements using 7 t fmri," *Journal of neural engineering*, vol. 12, no. 6, p. 066026, 2015. DOI: 10.1088/1741-2560/12/6/066026.

- [34] N. Ramsey, M. van de Heuvel, K. Kho, and F. Leijten, "Towards human bci applications based on cognitive brain systems: An investigation of neural signals recorded from the dorsolateral prefrontal cortex," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 214–217, 2006. DOI: 10.1109/TNSRE.2006.875582.
- [35] N. Naseer and K.-S. Hong, "Fnirs-based brain-computer interfaces: A review," *Frontiers in Human Neuroscience*, vol. 9, Jan. 2015. DOI: 10.3389/fnhum.2015.00003.
- [36] J. E. Huggins, A. A. Moinuddin, A. E. Chiodo, and P. A. Wren, "What would brain-computer interface users want: Opinions and priorities of potential users with spinal cord injury," *Archives of physical medicine and rehabilitation*, vol. 96, no. 3, S38–S45, 2015. DOI: <https://doi.org/10.1016/j.apmr.2014.05.028>.
- [37] J. E. Huggins, P. A. Wren, and K. L. Gruis, "What would brain-computer interface users want? opinions and priorities of potential users with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 12, no. 5, pp. 318–324, 2011. DOI: <https://doi.org/10.3109/17482968.2011.572978>.
- [38] S. A. Murad and N. Rahimi, "Unveiling thoughts: A review of advancements in eeg brain signal decoding into text," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, no. 1, pp. 61–76, 2025. DOI: 10.1109/TCDS.2024.3462452.
- [39] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.
- [40] M. Li and B.-L. Lu, "Emotion classification based on gamma-band eeg," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1223–1226. DOI: 10.1109/IEMBS.2009.5334139.
- [41] T. Harmony, "The functional significance of delta oscillations in cognitive processing," *Frontiers in Integrative Neuroscience*, vol. 7, 2013, ISSN: 1662-5145. DOI: 10.3389/fnint.2013.00083. [Online]. Available: <https://www.frontiersin.org/journals/integrative-neuroscience/articles/10.3389/fnint.2013.00083>.
- [42] 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, 2022, ISSN: 1662-5161. DOI: 10.3389/fnhum.2022.867281. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2022.867281>.

- [43] J. T. Panachakel and A. G. Ramakrishnan, "Decoding covert speech from eeg-a comprehensive review," *Frontiers in Neuroscience*, vol. 15, 2021, ISSN: 1662-453X. DOI: 10.3389/fnins.2021.642251. [Online]. Available: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2021.642251>.
- [44] K. Koizumi, K. Ueda, and M. Nakao, "Development of a cognitive brain-machine interface based on a visual imagery method," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1062–1065. DOI: 10.1109/EMBC.2018.8512520.
- [45] T. Radüntz, "Signal quality evaluation of emerging eeg devices," *Frontiers in Physiology*, vol. Volume 9 - 2018, 2018, ISSN: 1664-042X. DOI: 10.3389/fphys.2018.00098. [Online]. Available: <https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2018.00098>.
- [46] S. Saha and M. Baumert, "Intra- and inter-subject variability in eeg-based sensorimotor brain computer interface: A review," *Frontiers in Computational Neuroscience*, vol. Volume 13 - 2019, 2020, ISSN: 1662-5188. DOI: 10.3389/fncom.2019.00087. [Online]. Available: <https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2019.00087>.
- [47] F. Farzan *et al.*, "Standardization of electroencephalography for multi-site, multi-platform and multi-investigator studies: Insights from the canadian biomarker integration network in depression," *Scientific Reports*, vol. 7, no. 1, p. 7473, Aug. 2017, ISSN: 2045-2322. DOI: 10.1038/s41598-017-07613-x. [Online]. Available: <https://doi.org/10.1038/s41598-017-07613-x>.
- [48] K. B. Mikkelsen, Y. R. Tabar, C. B. Christensen, and P. Kidmose, "Eegs vary less between lab and home locations than they do between people," *Frontiers in Computational Neuroscience*, vol. Volume 15 - 2021, 2021, ISSN: 1662-5188. DOI: 10.3389/fncom.2021.565244. [Online]. Available: <https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2021.565244>.
- [49] D. A. Moses *et al.*, "Neuroprosthesis for decoding speech in a paralyzed person with anarthria," *New England Journal of Medicine*, vol. 385, no. 3, pp. 217–227, 2021. DOI: 10.1056/NEJMoa2027540.
- [50] E. H. Hiebert, "The core vocabulary: The foundation of proficient comprehension," *The Reading Teacher*, vol. 73, no. 6, pp. 757–768, 2020. DOI: <https://doi.org/10.1002/trtr.1894>. eprint: <https://ila.onlinelibrary.wiley.com/doi/pdf/10.1002/trtr.1894>.

- [Online]. Available:
<https://ila.onlinelibrary.wiley.com/doi/abs/10.1002/trtr.1894>.
- [51] D. Fuller and L. Lloyd, *Principles and Practices in Augmentative and Alternative Communication*. Taylor & Francis, 2024. DOI:
<https://doi.org/10.4324/9781003525905>.
- [52] L. R. Hochberg *et al.*, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, Jul. 2006, ISSN: 1476-4687. DOI: 10.1038/nature04970. [Online]. Available:
<https://doi.org/10.1038/nature04970>.
- [53] L. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988, ISSN: 0013-4694. DOI: [https://doi.org/10.1016/0013-4694\(88\)90149-6](https://doi.org/10.1016/0013-4694(88)90149-6). [Online]. Available:
<https://www.sciencedirect.com/science/article/pii/0013469488901496>.
- [54] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, "A note on brain actuated spelling with the berlin brain-computer interface," in *Universal Access in Human-Computer Interaction. Ambient Interaction: 4th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2007 Held as Part of HCI International 2007 Beijing, China, July 22-27, 2007 Proceedings, Part II 4*, Springer, 2007, pp. 759–768. DOI:
10.1007/978-3-540-73281-5_83.
- [55] J. Pan, X. Chen, N. Ban, J. He, J. Chen, and H. Huang, "Advances in p300 brain-computer interface spellers: Toward paradigm design and performance evaluation," *Frontiers in Human Neuroscience*, vol. 16, 2022, ISSN: 1662-5161. DOI: 10.3389/fnhum.2022.1077717. [Online]. Available:
<https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2022.1077717>.
- [56] A. Bilal Aygün and A. Reşit Kavsaoglu, "An innovative p300 speller brain-computer interface design: Easy screen," *Biomedical Signal Processing and Control*, vol. 75, p. 103 593, 2022, ISSN: 1746-8094. DOI:
<https://doi.org/10.1016/j.bspc.2022.103593>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S174680942200115X>.
- [57] V. H. G. G. d. Novais and A. G. Leal, "A systematic review of the efficiency of p300-based digital spellers," in *Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence*, ser. ICAAI '24, Association for Computing Machinery, 2025, pp. 141–147, ISBN: 9798400718014. DOI:

- 10.1145/3704137.3704196. [Online]. Available: <https://doi.org/10.1145/3704137.3704196>.
- [58] O. Maslova *et al.*, "Non-invasive eeg-based bci spellers from the beginning to today: A mini-review," *Frontiers in Human Neuroscience*, vol. 17, 2023, ISSN: 1662-5161. DOI: 10.3389/fnhum.2023.1216648. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2023.1216648>.
- [59] Z. Al-Qaysi *et al.*, "A comprehensive review of deep learning power in steady-state visual evoked potentials," *Neural Computing and Applications*, vol. 36, no. 27, pp. 16 683–16 706, 2024. DOI: 10.1007/s00521-024-10143-z.
- [60] X. Li *et al.*, "Evaluation of an online ssvep-bci with fast system setup," *Journal of Neurorestoratology*, vol. 12, no. 2, p. 100122, 2024, ISSN: 2324-2426. DOI: <https://doi.org/10.1016/j.jnrt.2024.100122>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2324242624000299>.
- [61] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, "Converting your thoughts to texts: Enabling brain typing via deep feature learning of eeg signals," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2018, pp. 1–10. DOI: 10.1109/PERCOM.2018.8444575.
- [62] A. Shahid, I. Raza, and S. A. Hussain, "Emowrite: A sentiment analysis-based thought to text conversion," *arXiv preprint arXiv:2103.02238*, 2021. DOI: <https://doi.org/10.48550/arXiv.2103.02238>.
- [63] B. Min, J. Kim, H.-j. Park, and B. Lee, "Vowel imagery decoding toward silent speech bci using extreme learning machine with electroencephalogram," *BioMed Research International*, vol. 2016, no. 1, p. 2 618 265, 2016. DOI: <https://doi.org/10.1155/2016/2618265>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2016/2618265>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2016/2618265>.
- [64] N. Hashim, A. Ali, and W.-N. Mohd-Isa, "Word-based classification of imagined speech using eeg," in *Computational Science and Technology: 4th ICCST 2017, Kuala Lumpur, Malaysia, 29–30 November, 2017*, Springer, 2018, pp. 195–204.
- [65] H. Liu, D. Hajialigol, B. Antony, A. Han, and X. Wang, "Eeg2text: Open vocabulary eeg-to-text translation with multi-view transformer," in *2024 IEEE International Conference on Big Data (BigData)*, Los Alamitos, CA, USA: IEEE Computer Society, Dec. 2024, pp. 1824–1833. DOI:

- 10.1109/BigData62323.2024.10825980. [Online]. Available: [10.1109/BigData62323.2024.10825980](https://doi.org/10.1109/BigData62323.2024.10825980).
- [66] P. Sun, G. K. Anumanchipalli, and E. F. Chang, "Brain2char: A deep architecture for decoding text from brain recordings," *Journal of Neural Engineering*, vol. 17, no. 6, p. 066 015, Dec. 2020. DOI: 10.1088/1741-2552/abc742. [Online]. Available: <https://dx.doi.org/10.1088/1741-2552/abc742>.
- [67] A. Srivastava and T. Shinde, "Think2type: Thoughts to text using eeg waves," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 06, pp. 2278–018, 2020. DOI: 10.17577/IJERTV9IS060431.
- [68] F. Mattioli, C. Porcaro, and G. Baldassarre, "A 1d cnn for high accuracy classification and transfer learning in motor imagery eeg-based brain-computer interface," *Journal of Neural Engineering*, vol. 18, no. 6, p. 066 053, 2022. DOI: 10.1088/1741-2552/ac4430.
- [69] Y.-E. Lee and S.-H. Lee, "Eeg-transformer: Self-attention from transformer architecture for decoding eeg of imagined speech," in *2022 10th International Winter Conference on Brain-Computer Interface (BCI)*, 2022, pp. 1–4. DOI: 10.1109/BCI53720.2022.9735124.
- [70] B. Abibullaev, A. Keutayeva, and A. Zollanvari, "Deep learning in eeg-based bcis: A comprehensive review of transformer models, advantages, challenges, and applications," *IEEE Access*, vol. 11, pp. 127 271–127 301, 2023. DOI: 10.1109/ACCESS.2023.3329678.
- [71] M. P. Branco *et al.*, "Brain-computer interfaces for communication: Preferences of individuals with locked-in syndrome," *Neurorehabilitation and neural repair*, vol. 35, no. 3, pp. 267–279, 2021. DOI: 10.1177/1545968321989331.
- [72] Y. Duan, C. Chau, Z. Wang, Y.-K. Wang, and C.-t. Lin, "Dewave: Discrete encoding of eeg waves for eeg to text translation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [73] N. Hollenstein, J. Rotsztejn, M. Troendle, A. Pedroni, C. Zhang, and N. Langer, "Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading," *Scientific data*, vol. 5, no. 1, pp. 1–13, 2018. DOI: 10.1038/sdata.2018.291.
- [74] N. Hollenstein, M. Troendle, C. Zhang, and N. Langer, "Zuco 2.0: A dataset of physiological recordings during natural reading and annotation," eng, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, N. Calzolari *et al.*, Eds., Marseille, France: European Language Resources Association, May 2020, pp. 138–146, ISBN: 979-10-95546-34-4. [Online]. Available: <https://aclanthology.org/2020.lrec-1.18/>.

- [75] H. Jo, Y. Yang, J. Han, Y. Duan, H. Xiong, and W. H. Lee, "Are eeg-to-text models working?" *arXiv preprint arXiv:2405.06459*, 2024. DOI: <https://doi.org/10.48550/arXiv.2405.06459>.
- [76] B. T. M. Wong and C. Kit, "Extending machine translation evaluation metrics with lexical cohesion to document level," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, J. Tsujii, J. Henderson, and M. Paşca, Eds., Jeju Island, Korea: Association for Computational Linguistics, Jul. 2012, pp. 1060–1068. [Online]. Available: <https://aclanthology.org/D12-1097/>.
- [77] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional neural network for eeg-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, Jul. 2018, ISSN: 1741-2552. DOI: 10.1088/1741-2552/aace8c. [Online]. Available: <http://dx.doi.org/10.1088/1741-2552/aace8c>.
- [78] Y. Song, Q. Zheng, B. Liu, and X. Gao, "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023. DOI: 10.1109/TNSRE.2022.3230250.
- [79] A. Schögl, C. Brunner, R. Leeb, G. Pfurtscheller, and G. Müller-Putz, *Bci competition iv datasets*, <https://www.bbci.de/competition/iv/>, Accessed: 2025-04-21, 2008.
- [80] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015. DOI: 10.1109/TAMD.2015.2431497.
- [81] Y. Ding *et al.*, "Eeg-deformer: A dense convolutional transformer for brain-computer interfaces," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 3, pp. 1909–1918, 2025. DOI: 10.1109/JBHI.2024.3504604.
- [82] J. Shin, A. von Lühmann, D.-W. Kim, J. Mehnert, H.-J. Hwang, and K.-R. Müller, "Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset," *Scientific Data*, vol. 5, no. 1, p. 180003, Feb. 2018, ISSN: 2052-4463. DOI: 10.1038/sdata.2018.3. [Online]. Available: <https://doi.org/10.1038/sdata.2018.3>.
- [83] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Scientific data*, vol. 6, no. 1, pp. 1–8, 2019. DOI: 10.1038/s41597-019-0027-4.

- [84] I. Zyma *et al.*, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, 2019, ISSN: 2306-5729. DOI: 10.3390/data4010014. [Online]. Available: <https://www.mdpi.com/2306-5729/4/1/14>.
- [85] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. DOI: <https://doi.org/10.48550/arXiv.2010.11929>.
- [86] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on eeg: A systematic comparison," *Journal of Neural Engineering*, vol. 19, no. 6, p. 066020, Nov. 2022. DOI: 10.1088/1741-2552/aca220. [Online]. Available: <https://dx.doi.org/10.1088/1741-2552/aca220>.
- [87] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with c-lstm," *Journal of Neural Engineering*, vol. 17, no. 1, p. 016041, Jan. 2020. DOI: 10.1088/1741-2552/ab57c0. [Online]. Available: <https://dx.doi.org/10.1088/1741-2552/ab57c0>.
- [88] L. Cao *et al.*, "A novel deep learning method based on an overlapping time window strategy for brain-computer interface-based stroke rehabilitation," *Brain Sciences*, vol. 12, no. 11, 2022, ISSN: 2076-3425. DOI: 10.3390/brainsci12111502. [Online]. Available: <https://www.mdpi.com/2076-3425/12/11/1502>.
- [89] M. Lee, H.-G. Kwak, H.-J. Kim, D.-O. Won, and S.-W. Lee, "Seriessleepnet: An eeg time series model with partial data augmentation for automatic sleep stage scoring," *Frontiers in Physiology*, vol. Volume 14 - 2023, 2023, ISSN: 1664-042X. DOI: 10.3389/fphys.2023.1188678. [Online]. Available: <https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2023.1188678>.
- [90] C. He, J. Liu, Y. Zhu, and W. Du, "Data augmentation for deep neural networks model in eeg classification task: A review," *Frontiers in Human Neuroscience*, vol. Volume 15 - 2021, 2021, ISSN: 1662-5161. DOI: 10.3389/fnhum.2021.765525. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2021.765525>.
- [91] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in eeg analysis: An overview," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, p. 40, 2023. DOI: 10.1186/s12984-023-01169-w.

- [92] S. Bhat and E. Hortal, "Gan-based data augmentation for improving the classification of eeg signals," in *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*, ser. PETRA '21, Corfu, Greece: Association for Computing Machinery, 2021, pp. 453–458, ISBN: 9781450387927. DOI: 10.1145/3453892.3461338. [Online]. Available: <https://doi.org/10.1145/3453892.3461338>.
- [93] G. Siddhad, M. Iwamura, and P. P. Roy, *Enhancing eeg signal-based emotion recognition with synthetic data: Diffusion model approach*, 2024. arXiv: 2401.16878 [cs.HC]. [Online]. Available: <https://arxiv.org/abs/2401.16878>.
- [94] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. [Online]. Available: <https://doi.org/10.1186/s40537-019-0197-0>.
- [95] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in eeg signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2020.09.017>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220314223>.
- [96] S.-H. Lee, Y.-E. Lee, S. Kim, B.-K. Ko, and S.-W. Lee, "Towards neural decoding of imagined speech based on spoken speech," in *2023 11th International Winter Conference on Brain-Computer Interface (BCI)*, 2023, pp. 1–4. DOI: 10.1109/BCI57258.2023.10078707.
- [97] D. Liu, J. Zhang, H. Wu, S. Liu, and J. Long, "Multi-source transfer learning for eeg classification based on domain adversarial neural network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 218–228, 2023. DOI: 10.1109/TNSRE.2022.3219418.
- [98] K. Erickson, L. Geist, P. Hatch, and N. Quick, "The universal core vocabulary: Technical report," Center for Literacy and Disability Studies, Tech. Rep., 2019, Version date: December 20, 2019.
- [99] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and clinical neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987. DOI: 10.1016/0013-4694(87)90206-9.
- [100] D. S. Benitez, S. Toscano, and A. Silva, "On the use of the emotiv epoch neuroheadset as a low cost alternative for eeg signal acquisition," in *2016 IEEE Colombian Conference on Communications and Computing (COLCOM)*, 2016, pp. 1–6. DOI: 10.1109/ColComCon.2016.7516380.

- [101] PsychoPy Community, *Psychopy: Open-source software for psychology experiments*, <https://www.psychopy.org/>, Accessed: 2025-04-27, 2024.
- [102] J. Peirce *et al.*, "PsychoPy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019, ISSN: 1554-3528. DOI: 10.3758/s13428-018-01193-y. [Online]. Available: <https://doi.org/10.3758/s13428-018-01193-y>.
- [103] W. Hohaia, B. W. Saurels, A. Johnston, K. Yarrow, and D. H. Arnold, "Occipital alpha-band brain waves when the eyes are closed are shaped by ongoing visual processes," *Scientific reports*, vol. 12, no. 1, p. 1194, 2022. DOI: 10.1038/s41598-022-05289-6.
- [104] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: A general-purpose brain-computer interface (bci) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004. DOI: 10.1109/TBME.2004.827072.
- [105] A. L. Goldberger *et al.*, *Eeg motor movement/imagery dataset*, <https://www.physionet.org/content/eegmmidb/1.0.0/>, Accessed: 2025-04-27, 2000.
- [106] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, e215–e220, 2000, [Online]. DOI: 10.1161/01.cir.101.23.e215.
- [107] S. Anwar, N. Tiwari, and V. Bhattacharjee, *Eeg dataset for natural image recognition through visual stimuli*, Mendeley Data, Version 2, 2025. DOI: 10.17632/g9shp2gxhy.2.
- [108] S. Tobimatsu and G. G. Celesia, "Studies of human visual pathophysiology with visual evoked potentials," *Clinical Neurophysiology*, vol. 117, no. 7, pp. 1414–1433, 2006, ISSN: 1388-2457. DOI: <https://doi.org/10.1016/j.clinph.2006.01.004>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1388245706000344>.
- [109] J. Decety, "Do imagined and executed actions share the same neural substrate?" *Cognitive Brain Research*, vol. 3, no. 2, pp. 87–93, 1996, Mental representations of motor acts, ISSN: 0926-6410. DOI: [https://doi.org/10.1016/0926-6410\(95\)00033-X](https://doi.org/10.1016/0926-6410(95)00033-X). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/092664109500033X>.
- [110] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort, "Autoreject: Automated artifact rejection for meg and eeg data," *NeuroImage*, vol. 159, pp. 417–429, 2017. DOI: 10.1016/j.neuroimage.2017.06.030.

- [111] Z. Zhang *et al.*, "Chisco: An EEG-based BCI dataset for decoding of imagined speech," *Scientific Data*, vol. 11, no. 1, p. 1265, Nov. 2024, ISSN: 2052-4463. DOI: 10.1038/s41597-024-04114-1. [Online]. Available: <https://doi.org/10.1038/s41597-024-04114-1>.
- [112] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19, Anchorage, AK, USA: Association for Computing Machinery, 2019, pp. 2623–2631, ISBN: 9781450362016. DOI: 10.1145/3292500.3330701. [Online]. Available: <https://doi.org/10.1145/3292500.3330701>.
- [113] G. Pereyra, G. Tucker, J. Chorowski, Ł. Kaiser, and G. Hinton, *Regularizing neural networks by penalizing confident output distributions*, 2017. arXiv: 1701.06548 [cs.NE]. [Online]. Available: <https://arxiv.org/abs/1701.06548>.
- [114] N. Dijkstra, S. E. Bosch, and M. A. van Gerven, "Shared neural mechanisms of visual perception and imagery," *Trends in Cognitive Sciences*, vol. 23, no. 5, pp. 423–434, 2019, ISSN: 1364-6613. DOI: <https://doi.org/10.1016/j.tics.2019.02.004>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364661319300592>.
- [115] J. Orpella, F. Mantegna, M. F. Assaneo, and D. Poeppel, "Decoding imagined speech reveals speech planning and production mechanisms," *bioRxiv*, 2022. DOI: 10.1101/2022.05.30.494046. eprint: <https://www.biorxiv.org/content/early/2022/09/27/2022.05.30.494046.full.pdf>. [Online]. Available: <https://www.biorxiv.org/content/early/2022/09/27/2022.05.30.494046>.
- [116] C. L. Scrivener and A. T. Reader, "Variability of eeg electrode positions and their underlying brain regions: Visualizing gel artifacts from a simultaneous eeg-fmri dataset," *Brain and Behavior*, vol. 12, no. 2, e2476, 2022. DOI: [https://doi.org/10.1002\(brb3.2476](https://doi.org/10.1002(brb3.2476). eprint: [https://onlinelibrary.wiley.com/doi/pdf/10.1002\(brb3.2476](https://onlinelibrary.wiley.com/doi/pdf/10.1002(brb3.2476). [Online]. Available: [https://onlinelibrary.wiley.com/doi/abs/10.1002\(brb3.2476](https://onlinelibrary.wiley.com/doi/abs/10.1002(brb3.2476).
- [117] S. R. Atcherson, H. J. Gould, M. A. Pousson, and T. M. Prout, "Variability of Electrode Positions Using Electrode Caps," *Brain Topography*, vol. 20, no. 2, pp. 105–111, Dec. 2007, ISSN: 1573-6792. DOI: 10.1007/s10548-007-0036-z. [Online]. Available: <https://doi.org/10.1007/s10548-007-0036-z>.

- [118] Y. Gamano, Y. Saito, K. Takamori, and K.-I. Morishige, "Effects of eeg electrode positional deviations for classification accuracy on different days," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan: IEEE Press, 2018, pp. 2277–2282. DOI: 10.1109/SMC.2018.00391. [Online]. Available: <https://doi.org/10.1109/SMC.2018.00391>.
- [119] 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. 6999–7019, 2022. DOI: 10.1109/TNNLS.2021.3084827.
- [120] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, 2022, ISSN: 2666-6510. DOI: <https://doi.org/10.1016/j.aiopen.2022.10.001>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666651022000146>.
- [121] A. Vaswani, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.
- [122] S. Latif, A. Zaidi, H. Cuayahuitl, F. Shamshad, M. Shoukat, and J. Qadir, *Transformers in speech processing: A survey*, 2023. arXiv: 2303.11607 [cs.CL]. [Online]. Available: <https://arxiv.org/abs/2303.11607>.
- [123] J. Sun, J. Xie, and H. Zhou, "Eeg classification with transformer-based models," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 2021, pp. 92–93. DOI: 10.1109/LifeTech52111.2021.9391844.
- [124] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *Bert: Pre-training of deep bidirectional transformers for language understanding*, 2019. arXiv: 1810.04805 [cs.CL]. [Online]. Available: <https://arxiv.org/abs/1810.04805>.
- [125] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI*, 2018.
- [126] Y. Ming *et al.*, "Subject adaptation network for eeg data analysis," *Applied Soft Computing*, vol. 84, p. 105 689, 2019, ISSN: 1568-4946. DOI: <https://doi.org/10.1016/j.asoc.2019.105689>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494619304703>.
- [127] C. Cano-Melle, E. Villar-Rodríguez, M. Baena-Pérez, M. A. Parcet, and C. Avila, "Effects of lateralization of language on cognition among left-handers," *Neurobiology of Language*, vol. 6, Apr. 2025, ISSN: 2641-4368. DOI: 10.1162/nol_a_00165.
- [128] A. K. Lindell, "In your right mind: Right hemisphere contributions to language processing and production," *Neuropsychology Review*, vol. 16, no. 3, pp. 131–148, 2006. DOI: 10.1007/s11065-006-9011-9.

- [129] Y. J. Yang, E. J. Jeon, J. S. Kim, and C. K. Chung, "Characterization of kinesthetic motor imagery compared with visual motor imageries," *Scientific Reports*, vol. 11, no. 1, p. 3751, 2021. DOI: 10.1038/s41598-021-82241-0.
- [130] C. M. Peters, M. W. Scott, R. Jin, M. Ma, S. N. Kraeutner, and N. J. Hodges, "Evidence for the dependence of visual and kinesthetic motor imagery on isolated visual and motor practice," *Consciousness and Cognition*, vol. 127, p. 103802, 2025, ISSN: 1053-8100. DOI: <https://doi.org/10.1016/j.concog.2024.103802>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053810024001697>.
- [131] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review of Neuroscience*, vol. 27, no. Volume 27, 2004, pp. 649–677, 2004, ISSN: 1545-4126. DOI: <https://doi.org/10.1146/annurev.neuro.27.070203.144220>. [Online]. Available: <https://www.annualreviews.org/content/journals/10.1146/annurev.neuro.27.070203.144220>.
- [132] E. P. Simoncelli, "Vision and the statistics of the visual environment," *Current Opinion in Neurobiology*, vol. 13, no. 2, pp. 144–149, 2003, ISSN: 0959-4388. DOI: [https://doi.org/10.1016/S0959-4388\(03\)00047-3](https://doi.org/10.1016/S0959-4388(03)00047-3). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959438803000473>.
- [133] L. Z. and, "Theoretical understanding of the early visual processes by data compression and data selection," *Network: Computation in Neural Systems*, vol. 17, no. 4, pp. 301–334, 2006, PMID: 17283516. DOI: 10.1080/09548980600931995. eprint: <https://doi.org/10.1080/09548980600931995>. [Online]. Available: <https://doi.org/10.1080/09548980600931995>.
- [134] J. V. Haxby, E. A. Hoffman, and M. Gobbini, "The distributed human neural system for face perception," *Trends in Cognitive Sciences*, vol. 4, no. 6, pp. 223–233, 2000, ISSN: 1364-6613. DOI: [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364661300014820>.
- [135] D. D. Lundstrom, T. Huang, and M. Razaviyayn, "A rigorous study of integrated gradients method and extensions to internal neuron attributions," in *Proceedings of the 39th International Conference on Machine Learning*, K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, Eds., ser. Proceedings of Machine Learning Research, vol. 162, PMLR, Jul. 2022,

REFERENCES

- pp. 14 485–14 508. [Online]. Available:
<https://proceedings.mlr.press/v162/lundstrom22a.html>.
- [136] L. van der Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: <http://jmlr.org/papers/v9/vandermaaten08a.html>.

Appendix A Supplemental Literature

This appendix provides foundational background and supplemental context to support the work presented in the main body of the thesis. It is divided into three sections: an overview of CNNs, and Transformer architectures, as well as a discussion of a strategy for overcoming cross-subject variability in EEG decoding. The first two sections focus on the fundamental principles and relevance of CNNs and Transformers in processing neural data, particularly EEG signals. The final section examines a domain adaptation approach, the Subject Adaptation Network (SAN), which aims to improve model generalisability by mitigating subject-specific differences in brain activity.

A.1 Convolutional Neural Networks

CNNs are a specialised class of neural networks designed to process multi-dimensional data, particularly two-dimensional inputs such as images [119]. They utilise convolutional layers that apply learnable filters (also known as kernels) across the input data. These filters are trained to detect hierarchical features, beginning with simple patterns such as edges or textures in the early layers, and progressing to more complex structures in deeper layers. The filters operate through a sliding window mechanism, producing outputs known as feature maps.

Following each convolutional layer, non-linear activation functions, most commonly the rectified linear unit (ReLU), are applied. These non-linearities are important, as they enable the network to learn complex, non-linear relationships within the data. Without them, stacking multiple convolutional layers would yield only a linear transformation, significantly limiting the model's expressive capacity.

Pooling layers are commonly interleaved with convolutional layers to downsample the feature maps. This downsampling reduces the spatial dimensions of the data, thereby lowering the number of parameters and hence, the model's computational complexity. Additionally, these layers enhance the robustness of the learned features to minor variations in the input and help mitigate overfitting. Pooling, like convolution, operates via a sliding window and can be implemented using different strategies, such as max-pooling, which selects the maximum value within each window to represent the selected region in the resultant feature map, or average-pooling, which computes the window's mean instead.

The architecture of CNNs makes them particularly effective at capturing spatial dependencies within data, an ability that is crucial when the spatial structure encodes meaningful information. This characteristic extends well to EEG data, where the spatial relationships between electrodes placed on the scalp carry essential physiological

signals. Moreover, local temporal patterns, interpretable as spatial relationships across the time domain, also contribute valuable information. CNNs can be adapted to process EEG signals by treating the multi-channel time series as a two-dimensional spatio-temporal matrix, where one dimension represents channels and the other represents time.

A.2 Transformers

Transformers, often referred to as attention-based neural networks or attention models, have rapidly emerged as the dominant architecture for handling sequential data [120]. Originally conceived in the field of natural language processing (NLP) to address the challenges of text processing [121], their self-attention mechanism have since proven highly adaptable, leading to their widespread adoption across diverse domains such as computer vision [85], speech recognition [122], and even decoding of EEG signals [69, 123].

Unlike traditional RNNs, which process sequences step-by-step, transformers use a self-attention mechanism to process entire sequences in parallel. This mechanism dynamically weighs the relevance of each element within the sequence in relation to all others, allowing the model to efficiently capture both local and global dependencies. This parallel processing and inherent ability to model long-range dependencies also effectively mitigates common issues found in RNNs, such as vanishing or exploding gradients.

The Transformer architecture, as seen in Figure A.1, is built around an encoder-decoder framework, originally proposed for sequence-to-sequence tasks like machine translation [121]. However, many modern implementations use only one of these components depending on the task: encoder-only models (such as BERT [124]) are suited for understanding tasks such as classification or question answering, while decoder-only models (such as GPT [125]) are optimised for generative tasks like language modelling.

Each encoder and decoder block is composed of a stack of identical layers, where each layer contains two key sub-components: a multi-head self-attention mechanism, and a position-wise feed-forward neural network. To facilitate the training of deep models and preserve information flow across layers, each sub-layer is surrounded by residual (skip) connections and followed by layer normalisation.

One unique challenge with transformer models arises from their lack of inherent sequential structure. Unlike RNNs, which maintain a natural temporal order through their recurrent connections, transformers process all inputs simultaneously. To compensate for this, positional encodings are added to the input embeddings, providing

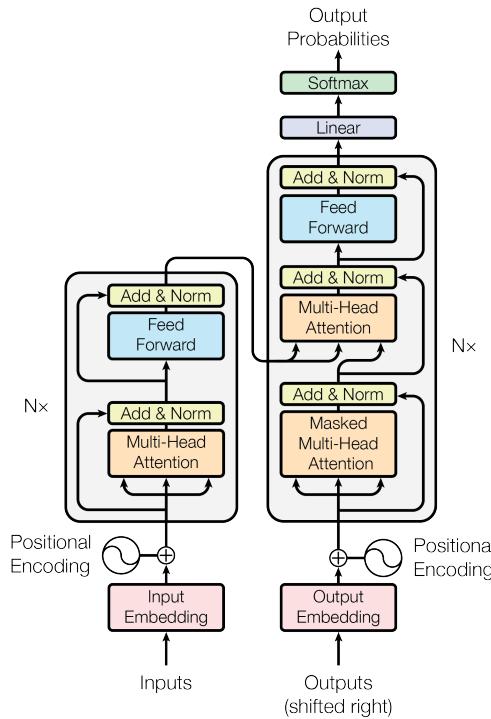


Figure A.1 Transformer Architecture. (Source: Vaswani et al. [121])

the model with information about the position of each element in the sequence.

At the heart of the Transformer architecture lies the attention mechanism, which enables the model to dynamically focus on different parts of the input sequence based on contextual relevance. This mechanism computes a weighted sum of value vectors, where the weights are derived from the similarity between query and key vectors, all of which are learned representations of the input. For each token in the sequence, self-attention allows it to attend to every other token, capturing complex relationships regardless of distance. This capability is essential for modelling long-range dependencies, a task that traditional RNNs struggle with.

Transformers further enhance this mechanism through multi-head attention, which allows the model to learn multiple perspectives of relevance in parallel. Each attention head independently computes attention using separate sets of learned weights, and the results are concatenated and projected to form the final output of the layer. This improves the model's capacity to understand different types of dependencies within the sequence.

However, a key limitation of the original self-attention mechanism is its quadratic computational complexity with respect to sequence length. This makes the model computationally expensive for very long inputs, prompting ongoing research into more efficient attention mechanisms that maintain performance while improving scalability.

Transformers are increasingly used in the field of brain signal decoding for their

ability to extract meaningful features from complex, multi-dimensional neural signals. Typically, raw signal inputs are first passed through a CNN to downsample the data across the temporal dimension, creating embeddings which are fed into the transformer’s attention layers [69, 78, 81]. The transformer’s capacity to model long-range dependencies and capture contextual relationships across time and channels makes it particularly well-suited for EEG decoding, where relevant neural patterns may span across distant time points or emerge from distributed brain regions.

A.3 Overcoming cross-subject variability

As discussed in Section 2.1.4, variability is a significant challenge in EEG data analysis, leading to inconsistencies that hinder broader model generalisation. To address this, Ming et al. proposed a SAN [126], drawing inspiration from GANs. The core idea is to find a mapping function that transforms raw EEG data into a feature space where the probability density functions (PDFs) from different subjects are better aligned. While directly optimising this alignment is computationally intractable due to the high-dimensional nature of EEG data, the SAN circumvents this by enforcing alignment with a pre-designed, artificial distribution in a lower-dimensional space. This adversarial approach allows the network to learn a transformation that brings disparate distributions into a coherent one.

The SAN’s architecture conceptually mirrors a GAN, comprising a generative network and a discriminative network. The generative component is further divided into an adaptor network (A) and a mapper network (M), with the discriminator denoted by D. Unlike a typical GAN, the SAN’s generator receives real data samples (not random noise) to align existing distributions, and its discriminator is trained on samples from the designed artificial distribution, rather than real-world samples. This competitive training process ensures that the adaptor learns to project the high-dimensional EEG data into an embedding space where inter-subject variance is reduced, making the representations more suitable for subsequent tasks like classification. The artificial target distribution itself is crafted by estimating statistical properties, such as modality and relative cluster size, from the original sample data, enabling the network to align with a statistically representative, yet controlled, target.

Using the SAN for EEG data processing typically involves two stages. The first stage focuses on training the adaptor, mapper, and discriminator in an adversarial fashion until the discriminator can no longer reliably distinguish between inputs from the generator and the artificial distribution. This signifies that the EEG data has been effectively adapted. In the second stage, the trained adaptor (and potentially the mapper) acts as a crucial pre-processing step, providing more consistent and

generalised representations to downstream components. For instance, these adapted representations can be fed into a classifier for improved cross-subject performance.

To demonstrate the SAN's practical efficacy, Ming et al. evaluated the model using different EEG datasets, namely a driving EEG dataset and a VEP dataset. By aligning distributions in an intermediate latent space, the SAN produced adapted representations that, when fed into a separate classifier, yielded comparable or slightly better results than the selected baseline models. Notably, the SAN outperformed EEGNet, which reached an accuracy of 80.5%, by achieving an accuracy of 81.5%. While this highlights the SAN's potential to enhance generalisation performance by addressing both intra-subject and cross-subject variability, further testing is required to validate the observed improvement over a larger variety of EEG datasets and paradigms.

To evaluate the practical efficacy of the SAN model, Ming et al. tested it on two distinct EEG datasets: a driving-related EEG dataset and a VEP dataset. The SAN aligned data distributions in a shared latent space, producing domain-adapted representations that were subsequently classified using a separate classifier. This approach resulted in performance that was comparable to, or slightly better than, that of baseline models. Notably, SAN outperformed EEGNet, improving classification accuracy from 80.5% to 81.5%. While these results suggest that SAN addresses both intra-subject and cross-subject variability, enhancing model generalisation, further validation across a broader range of EEG datasets and paradigms is necessary to confirm its robustness.

Appendix B Supplemental Results

This appendix presents additional analyses and visualisations that provide deeper insight into the findings of this project. It includes cross-subject confusion matrices, brain activity patterns (via topomaps and time-series heatmaps), saliency visualisations, and t-SNE plots. Together, these results offer a more detailed look at the neurological patterns observed, as well as the implications of model performance and interpretability in the context of imagined speech decoding.

B.1 Cross-Subject Confusion Matrices

Due to space limitations, the confusion matrices are not included in the main body of this thesis. Instead, they are presented here in Figures B.1–B.3.

As with the within-subject evaluations, the cross-subject confusion matrices indicate that model performance is relatively consistent across the different target words, although this uniformity reflects consistently poor performance.. This further supports the conclusion that the models operate at/close to chance level, as no clear or systematic patterns emerge in the predictions. Notably, the word ‘On’ exhibits the poorest performance, mirroring the within-subject results. Additionally, It was the least frequently predicted word across all evaluated models.

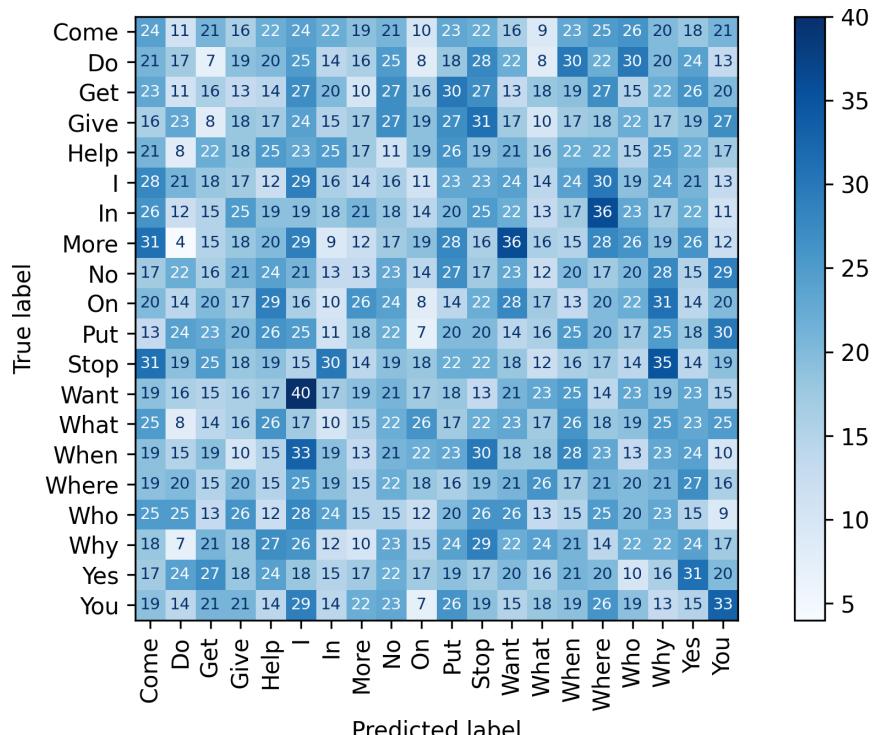


Figure B.1 Baseline Confusion Matrix (Cross-Subject). (Source: Self)

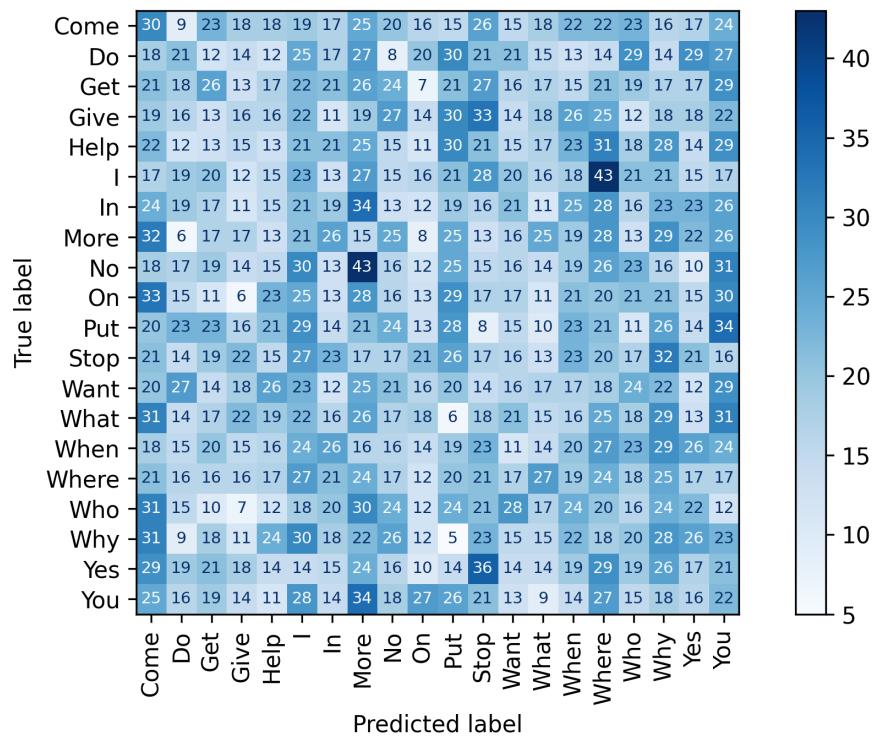


Figure B.2 MI-Pretrained Confusion Matrix (Cross-Subject). (Source: Self)

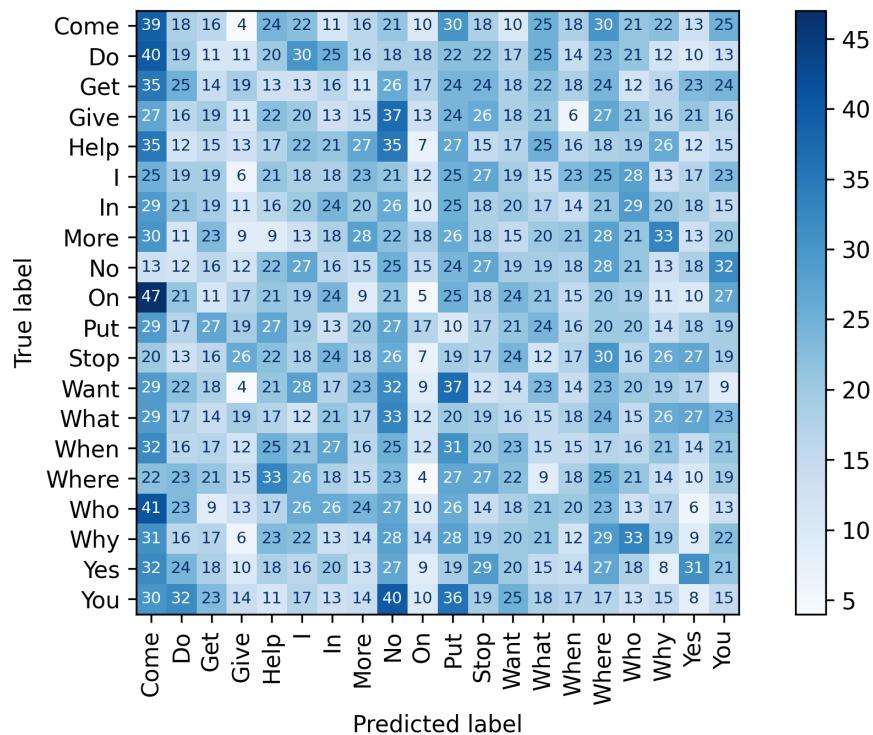


Figure B.3 VEP-Pretrained Confusion Matrix (Cross-Subject). (Source: Self)

B.2 Brain Activity

To further explore the differences across the used datasets, we present visualisations of participants' brain activity recorded during data collection. In these figures, yellow indicates regions of higher activity, while dark blue represents lower activity.

The visualisations are divided into two main types: topographic maps (topomaps) and time-series plots. Topomaps illustrate the spatial distribution of brain activity across the scalp and are generated by averaging the signal over time. This produces a single, representative image that captures the overall spatial pattern of neural activation during the recording session. Each topomap corresponds to a specific stimulus, enabling comparisons of spatial activity patterns between classes. Additionally, topomaps are included for each subject in the SI dataset to emphasise the variability in brain activity patterns across individuals.

Topomaps are commonly used in EEG research to provide an intuitive, spatial representation of neural activity. They are generated by interpolating voltage values from electrodes placed on the scalp, resulting in a smooth visual depiction of brain activity distribution. This allows researchers to observe which regions of the brain are more or less active in response to different stimuli or tasks, facilitating spatial comparisons across conditions or subjects.

Time-series plots, on the other hand, display the EEG signal as it evolves over time for each channel. These plots allow for the examination of temporal dynamics and signal characteristics such as event-related potentials or oscillatory patterns.

B.2.1 Topomaps

During imagined speech, primary regions of neural activity are typically expected near the T7 electrode, which lies over the left temporal lobe (approximately the centre-left in the provided topographic diagrams), as well as in the frontal cortex (toward the front but not the very front of the scalp). These regions correspond to Wernicke's area and Broca's area, respectively, which are widely recognised for their roles in speech comprehension and production. As discussed in Section 4.1, imagined speech also engages motor-related areas, which supports the expectation of activity in the frontal cortex. Additionally, as indicated in the same section, speech involves a feedback loop between auditory (temporal lobe) and motor (frontal cortex) regions of the brain, further reinforcing this expectation.

However, the data presented in Figures B.4 and B.5 diverge from these expectations. Notably, some subjects exhibit increased activity in the occipital lobe (located at the back of the head), with Subject 14 serving as a clear example. In addition, Subject 14 demonstrates stronger activation in the right hemisphere of the

brain, in contrast to the expected dominance of the left hemisphere for language processing.

While it is not unprecedented for Broca's and Wernicke's areas to be lateralised to the right hemisphere, this is relatively uncommon. For reference, atypical right-hemispheric dominance for language is found in approximately 4–6% of right-handed individuals and up to 24% of left-handed individuals [127].

The frequency of right-lateralised activity observed in this dataset exceeds the expected distribution, suggesting that additional factors may be influencing the prevalence of right-hemisphere activation. The right hemisphere is typically considered less efficient in linguistic processing (for individuals with a left language dominance) and is primarily associated with features such as prosody [128]. However, it has also been shown to support nonpropositional speech, which involves automatic, context-bound utterances such as counting, reciting days of the week, or habitual expressions.

It is possible that the structure of the data collection protocol, particularly the prolonged duration of mental word repetition, inadvertently shifted the task toward a nonpropositional form of speech. Repeating a single word continuously for 40 seconds may have engaged mechanisms more akin to automatic or ritualistic speech (e.g., mantra-like repetition), rather than spontaneous, propositional language. This shift may partially account for the increased right-hemispheric activity observed in some participants.

As mentioned in Section 4.1, a number of participants reported having mental imagery evoked during the prolonged mental repetition of the word. This may be the major contributing reason for the increased activity in the occipital region as it associated with VEP and VI [114]. An interesting study may be exploring the prevalence of this mental imagery across the population, and understanding what the contributing factors may be.

Figure B.4 further illustrates the level of cross-subject variability in brain activity. The regions of dominant activation vary significantly across individuals, though some degree of frontal activity is consistently observed across the majority of participants. This aligns partially with our original expectation. The average topomap across all subjects, also shown in Figure B.4, highlights areas of overlapping activity and provides a broader view of common activation patterns.

Figure B.5 reveals the variation in brain activity patterns associated with the imagined speech of different words. These distinct neural activation profiles increase inter-class separability, potentially enabling machine learning models to achieve improved classification performance.

Figure B.6 presents the brain activity averaged across participants from the MI dataset, depicted as topomaps. These visualisations illustrate that even highly similar tasks, such as clenching the right fist versus clenching the left fist, elicit distinct neural

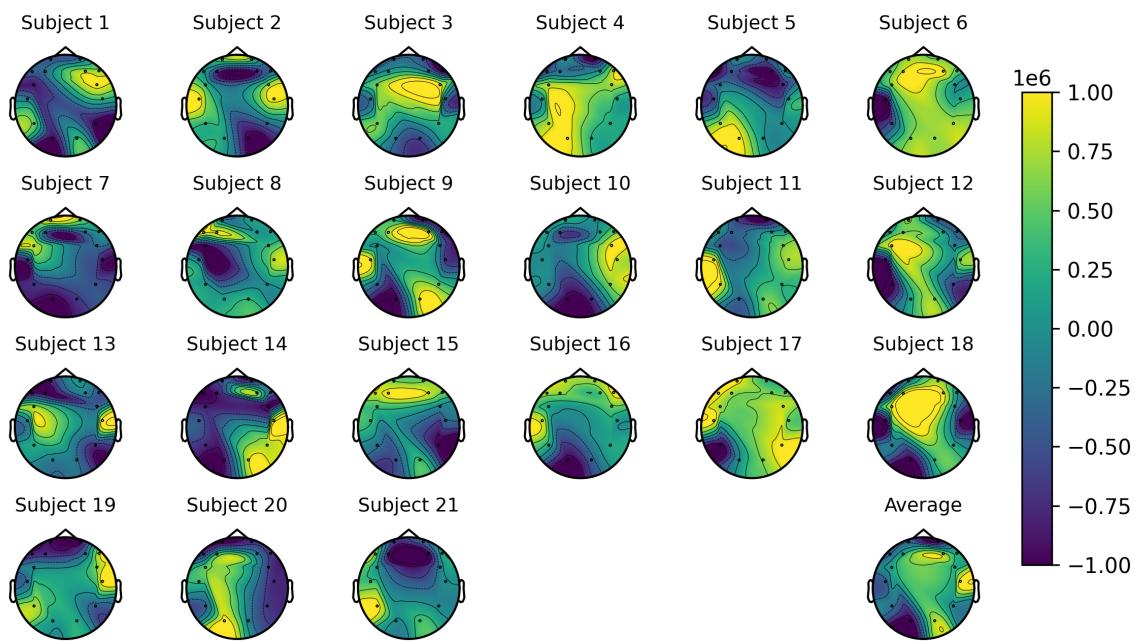


Figure B.4 Brain Activity by subject (SI Dataset). (Source: Self)

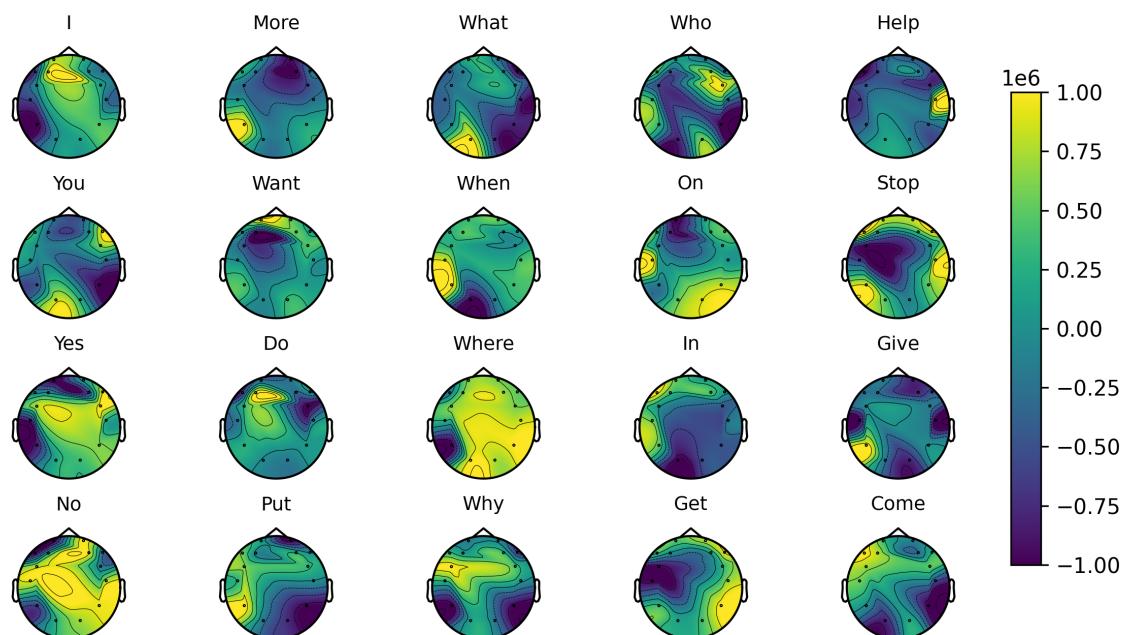


Figure B.5 Brain Activity by class (SI Dataset). (Source: Self)

signatures, particularly in their spatial distribution. As anticipated, the predominant activity is observed in the frontal cortex, consistent with its role in motor planning and execution.

However, notable activity is also present in the occipital region. This observation could initially be attributed to the visual nature of the prompts used in this study. Yet, the relative absence of occipital activation for data gathered during ‘both feet’ MI may indicate a deeper reason. MI is broadly categorized into two main types: visual motor imagery (VMI) and kinaesthetic motor imagery (KMI) [129]. VMI involves mentally visualising the movement, such as visually imagining closing one’s right fist. Conversely, KMI focuses on the proprioceptive sensations associated with the movement, like imagining the feeling of muscles straining while clenching a fist. Due to its inherent visual component, VMI elicits neural activity patterns that overlap with those during visual observation of movement, including increased activation in the occipital cortex. In contrast, KMI is more directly linked to the neural pathways involved in motor execution, with activity predominantly concentrated in the frontal cortex rather than the occipital region.

Considering that individuals can flexibly employ different MI strategies depending on the specific task [130], it is plausible that the observed occipital activity stems from participants engaging in VMI for certain tasks. The variability in MI strategy, and the inherent distinction between VMI and KMI, could therefore account for the differential occipital activation patterns across various MI prompts in this study.

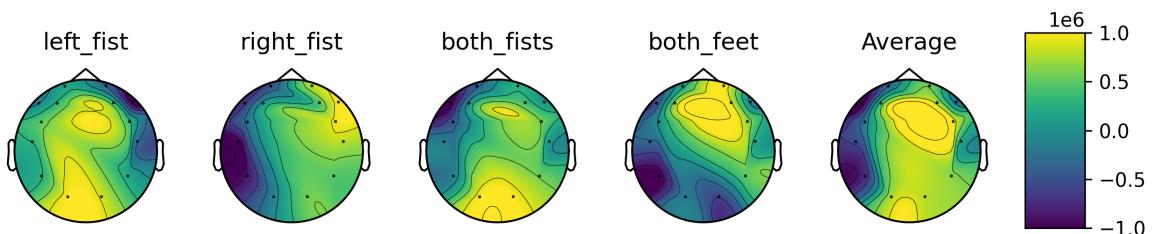


Figure B.6 Brain Activity by class (MI Dataset). (Source: Self)

Figure B.7 presents the brain activity averaged across participants from the VEP dataset, displayed as topomaps for each stimulus class. The overall average topomap largely aligns with expectations for a VEP, demonstrating a concentration of activity in the occipital region, an area generally associated with the processing of visual information.

However, investigating brain activity across the different stimulus categories reveals significant variations. The reduced activity observed in the occipital regions for ‘Flower’ and ‘Apple’ stimuli can be ascribed to their relative visual simplicity. The human visual cortex processes information in a hierarchical manner, with increasingly

specialised computations occurring along the occipito-temporal ('what' pathway) and occipito-parietal ('where/how' pathway) streams [131]. For simple and highly familiar objects like flowers or apples, the initial visual features may be processed efficiently and rapidly in early occipital areas [132, 133]. This efficient processing could result in highly transient occipital activity that is less prominently represented in time-averaged topomaps, compared to stimuli requiring more sustained early visual engagement.

On the other hand, 'Face' stimuli elicit a more distributed pattern of activity. This is consistent with the brain's specialised and complex interconnected networks for face processing, given their significant social and social importance [134]. Beyond primary visual cortex, this network involves regions such as the fusiform face area and occipital face area for identity, and areas in the superior temporal sulcus (STS) and parietal regions associated with changeable aspects like eye gaze and emotional expression. The engagement of these broader, higher-order visual and cognitive networks contributes to the overall scalp distribution. Furthermore, the apparent central localisation of activity for faces might be influenced by factors inherent to EEG data, such as volume conduction, where electrical activity spreads across the scalp, and the spatial interpolation techniques used to generate topomaps, which can sometimes centralise the projection of distributed or deeper sources.

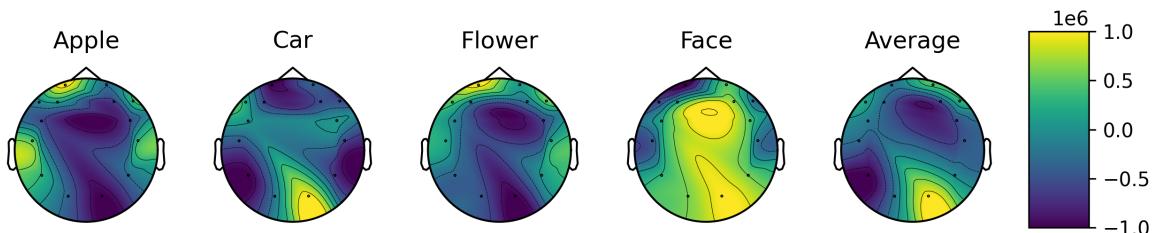


Figure B.7 Brain Activity by class (VEP Dataset). (Source: Self)

Notably, the average topomap of the SI dataset exhibits areas of overlap with both the VEP and MI datasets, particularly within the occipital and frontal cortical regions. These shared patterns suggest some common neural engagement across tasks. However, despite this partial convergence, the overall brain activity profiles in the VEP and MI datasets remain largely distinct from those observed in the SI data, reflecting the fundamental differences in the underlying cognitive processes.

B.2.2 Time-Series Heatmaps

A time-series visualisation of brain activity from the SI dataset is presented in Figure B.8. This heatmap represents the processed EEG signal values across all channels over a 4-second epoch, offering insight into the temporal dynamics and amplitude fluctuations of the neural signals.

The visualisation reveals significant variability in signal amplitude, with alternating periods of heightened activity (indicated in yellow) and reduced activity (dark blue) occurring across different channels and time intervals. Notably, increased activity is periodically visible around the FC6 electrode (and to a lesser extent the FC5 electrode), suggesting that participants may have been mentally repeating the presented word at a relatively consistent pace. The somewhat jagged appearance of the heatmap is likely attributable to the lower participant count ($n=21$), as averaging across a larger sample size typically produces a smoother signal representation.

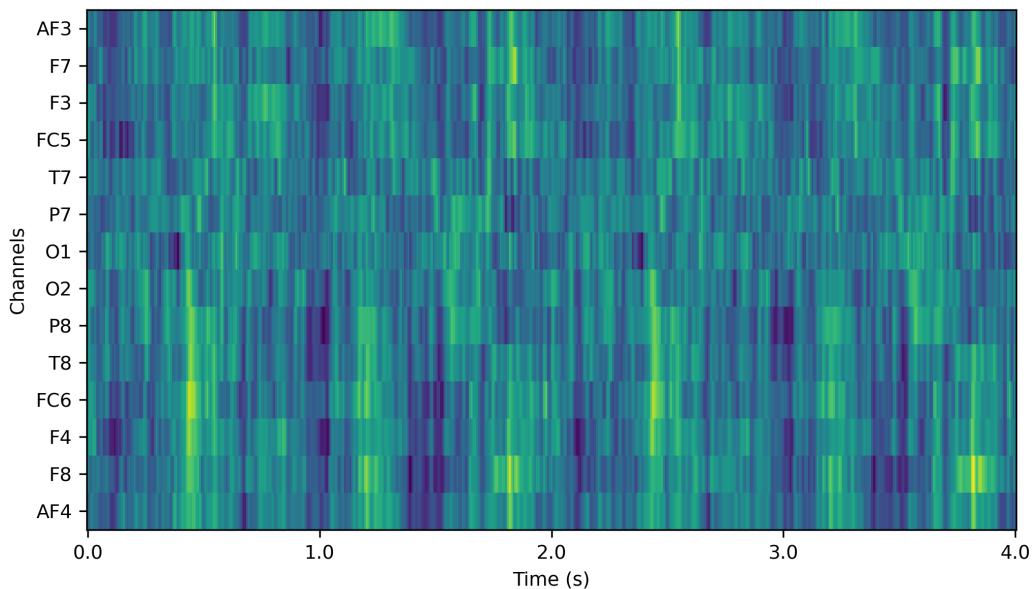


Figure B.8 Time-series Brain Activity (SI Dataset). (Source: Self)

In contrast, the corresponding time-series heatmap from the MI dataset, as seen in Figure B.9, displays a significantly smoother pattern, which can be attributed to its much larger participant pool ($n=109$). A distinct surge in activity is observed at the onset of the trial, likely corresponding to the initiation of motor imagery, followed by more consistent attenuated signal patterns over time.

The time-series heatmap for the VEP dataset, as seen in Figure B.10, shares visual similarities with the SI dataset, exhibiting a jagged structure and a clearer periodic pattern. One particularly interesting observation is the apparent lateral progression of activity across the scalp during the task. The activity appears to transition sequentially from the right temporal region to the right parietal, then to the occipital lobe, continuing on to the left parietal and left temporal regions, before reversing direction. This pattern aligns with known characteristics of visual processing, which progresses hierarchically through occipital, parietal, and temporal areas, reflecting the dynamic and distributed nature of VEP processing.

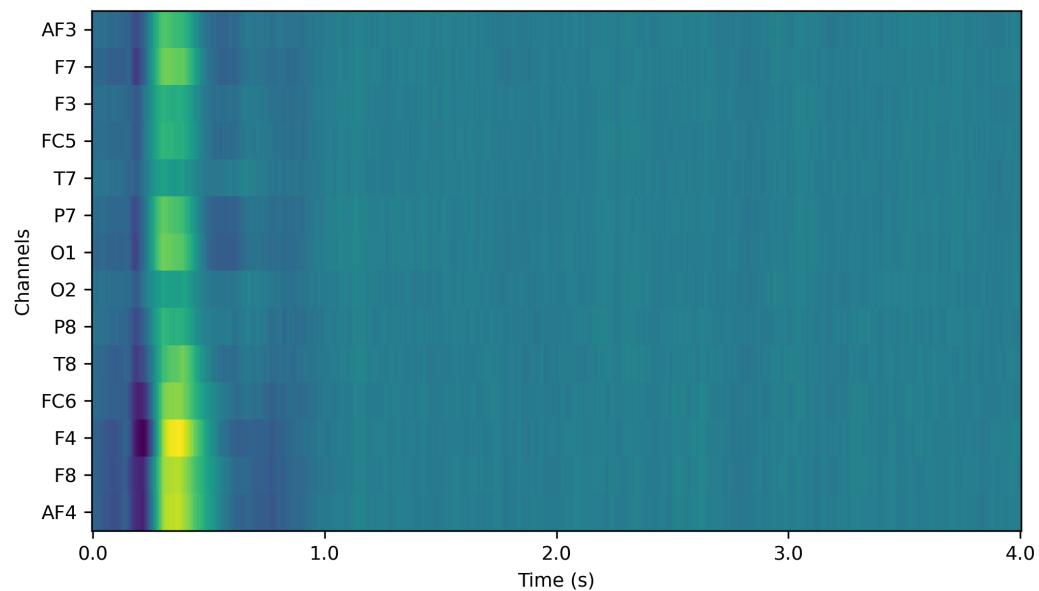


Figure B.9 Time-series Brain Activity (MI Dataset). (Source: Self)

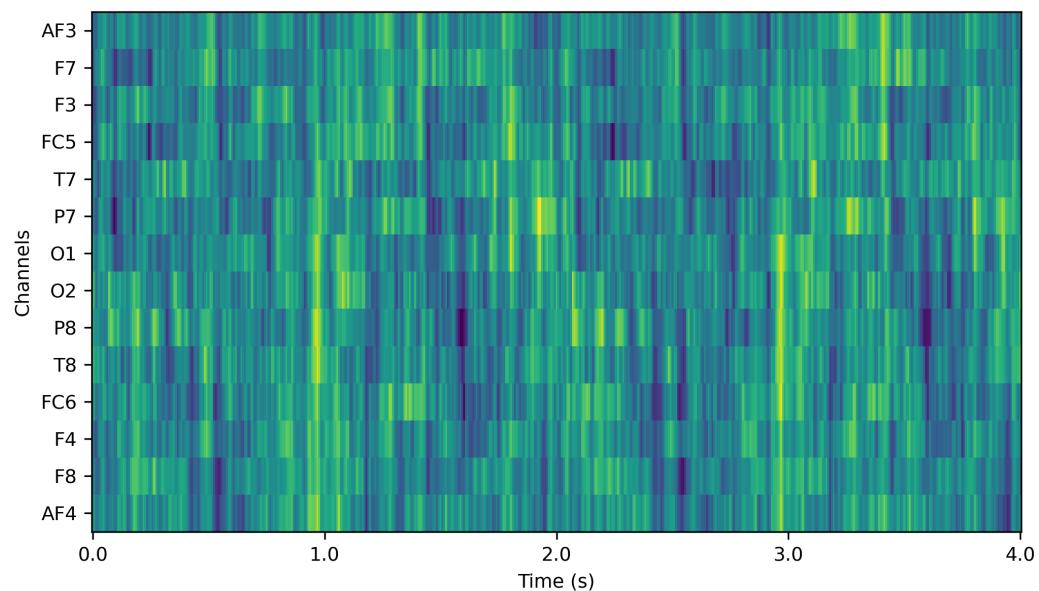


Figure B.10 Time-series Brain Activity (VEP Dataset). (Source: Self)

B.3 Saliency

To enhance the interpretability of our model’s classifications, we provide saliency maps which visually highlight the specific input features most influential in its decision-making process. Essentially, saliency quantifies the level of importance attributed to different parts of the input data for a given prediction. These maps are generated on a per-subject and per-class basis, allowing us to analyse how the model’s focus shifts across varying contexts. Furthermore, by generating time-series saliency maps, we gain insights into the model’s responsiveness to both spatial characteristics and temporal evolution within the data. Similar to the brain activity visualisations, a yellow colour indicates more importance, while a dark blue colour indicates the least amount of importance.

To generate these saliency maps, we used the Integrated Gradients method, implemented using the Captum library. Unlike some other gradient-based techniques that can suffer from saturation or noisy attributions, Integrated Gradients provides a more stable approach [135]. It works by attributing importance to each input feature by accumulating gradients along a linear path from a baseline (or reference) input to the actual input. This integration process ensures that the attributions are complete and sensitive to the entire range of the input signal, capturing both local and global influences on the model’s output.

Specifically, for each subject and class, we computed the Integrated Gradients with respect to the input features that led to the model’s classification. The magnitude of these integrated gradients then forms the basis of our saliency scores, with higher values indicating a greater contribution to the final prediction.

B.3.1 Topomaps

Contrary to the distinct neural signatures observed in Section B.2, the per-class saliency topomaps for the baseline model, as seen in Figure B.11, reveal a consistent concentration of saliency in the left occipital region (O1), with residual saliency around the left temporal and parietal areas (T7 and P7). While the model’s focus on the left temporal region somewhat aligns with initial expectations regarding its role in the task, the intense emphasis on seemingly irrelevant areas, particularly the left occipital region (which, on average, exhibited the least activity in the data), is highly concerning. This consistent hyper-focus across all different classes strongly suggests that the model may have overfit to the provided data, learning false correlations or noise patterns rather than the true underlying signal.

The model’s relatively high classification accuracy of 84.1% on the test set, despite this apparent overfitting, can likely be attributed to the specific structuring of

the trials. Since data for each class was collected from a single continuous trial and subsequently subdivided into smaller sub-trials, any noise present within a given continuous trial would be expected to be highly consistent across its derived sub-trials. This consistency in noise patterns across the subdivided data may have allowed a sufficiently large model to ‘memorise’ these unique noise signatures for each word and individual, leading to high performance on the dataset even without learning generalisable neural features.

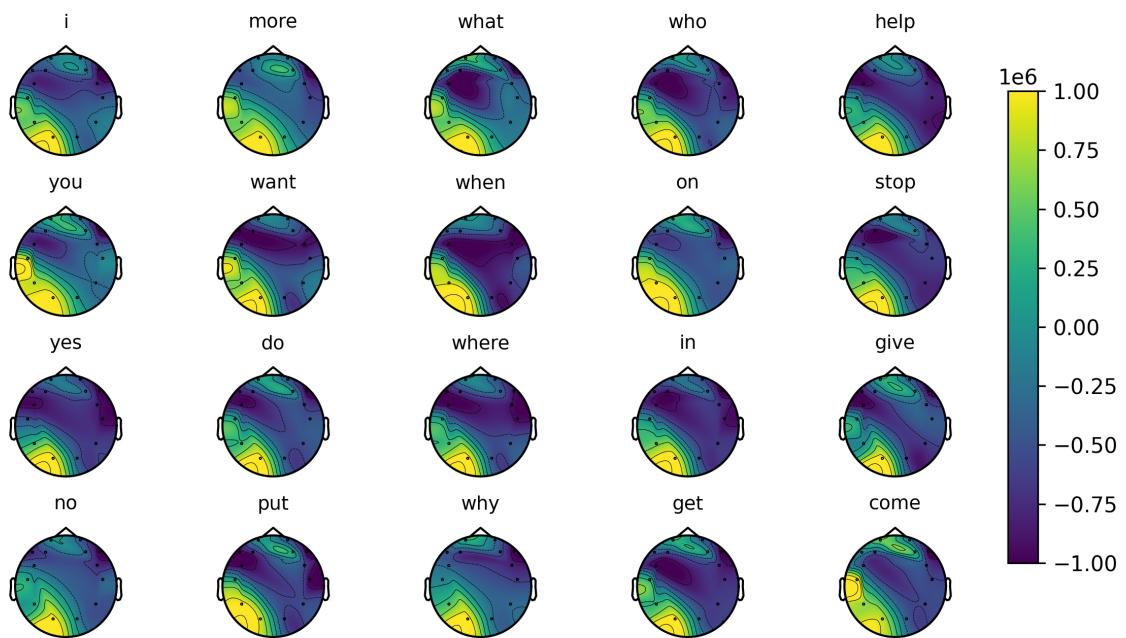


Figure B.11 Saliency by Class (Baseline). (Source: Self)

The per-subject saliency maps, as seen in Figure B.12, further reinforce this concern. While a consistent pattern across subjects would typically be desirable as an indicator of generalisable learning, in this context of potential overfitting, it instead suggests that the model is consistently latching onto the same problematic features across different participants. An intriguing exception is Subject 19, whose saliency pattern is notably more distributed across the topomap.

A similar pattern of hyper-localised saliency persists in the models pretrained on MI data and subsequently fine tuned. As shown in Figure B.13, this model also exhibits an intense, localised focus, though the specific focal points differ: primarily the frontal region towards the right forehead electrode (AF4), the left frontal lobe (F7), and with less pronounced saliency in the right occipital region (O2). This pattern of hyper-focused saliency continues to be observed in the per-subject saliency map, as seen in Figure B.14, and across the VEP-pretrained model saliency maps, as seen in Figures B.15, and B.16. Consistent with the baseline model, Subject 19 again presents a more broadly distributed saliency pattern across all these models.

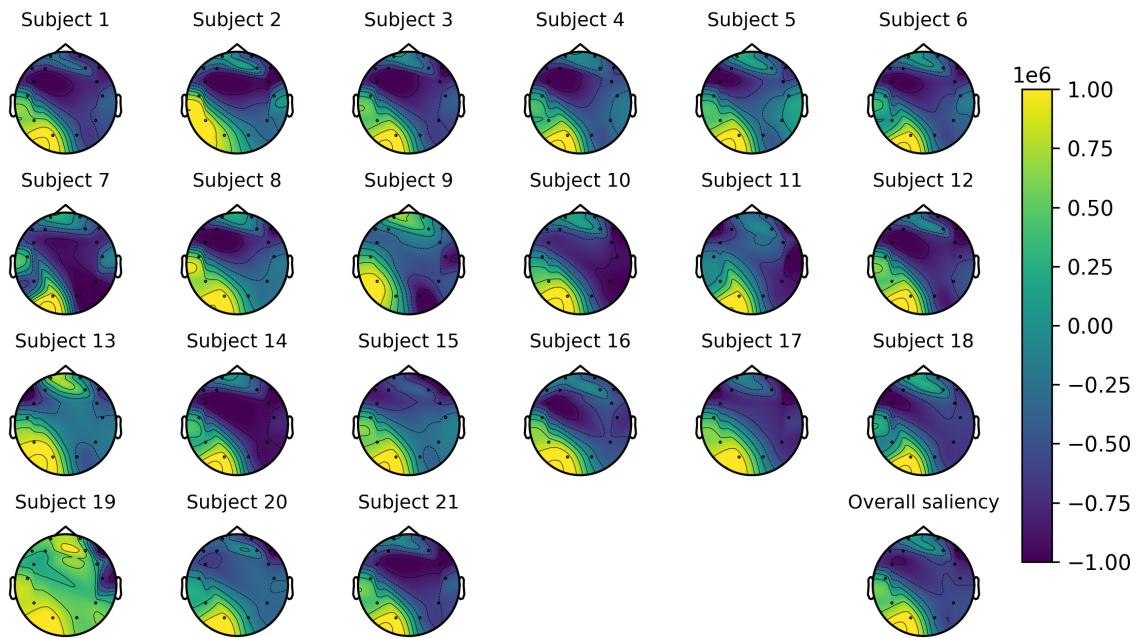


Figure B.12 Saliency by Subject (Baseline). (Source: Self)

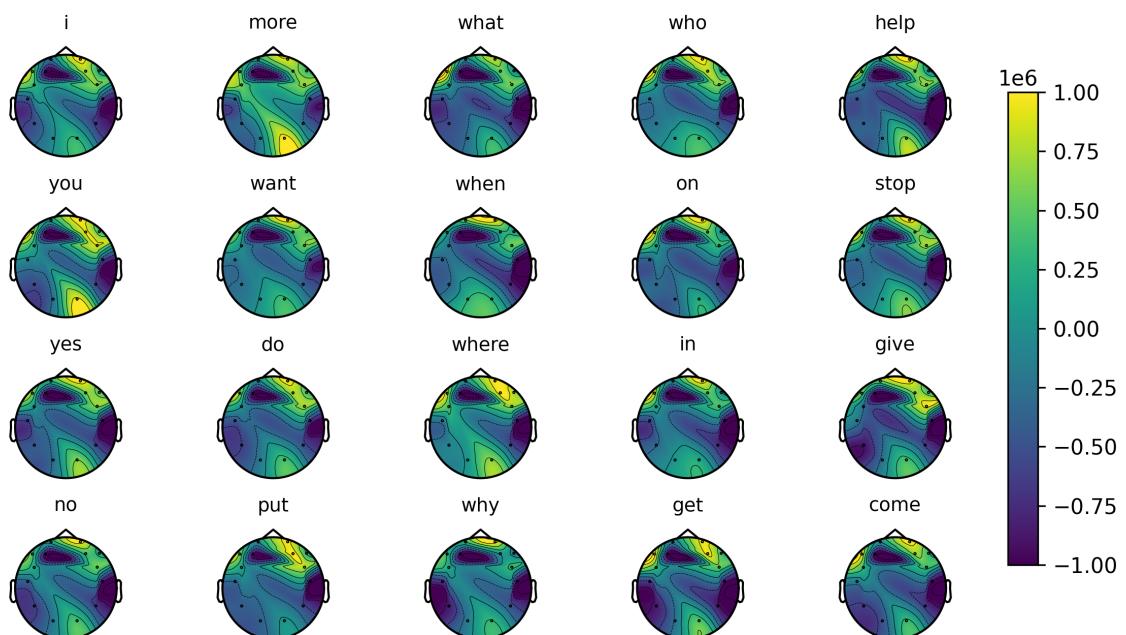


Figure B.13 Saliency by Class (MI-Pretrained). (Source: Self)

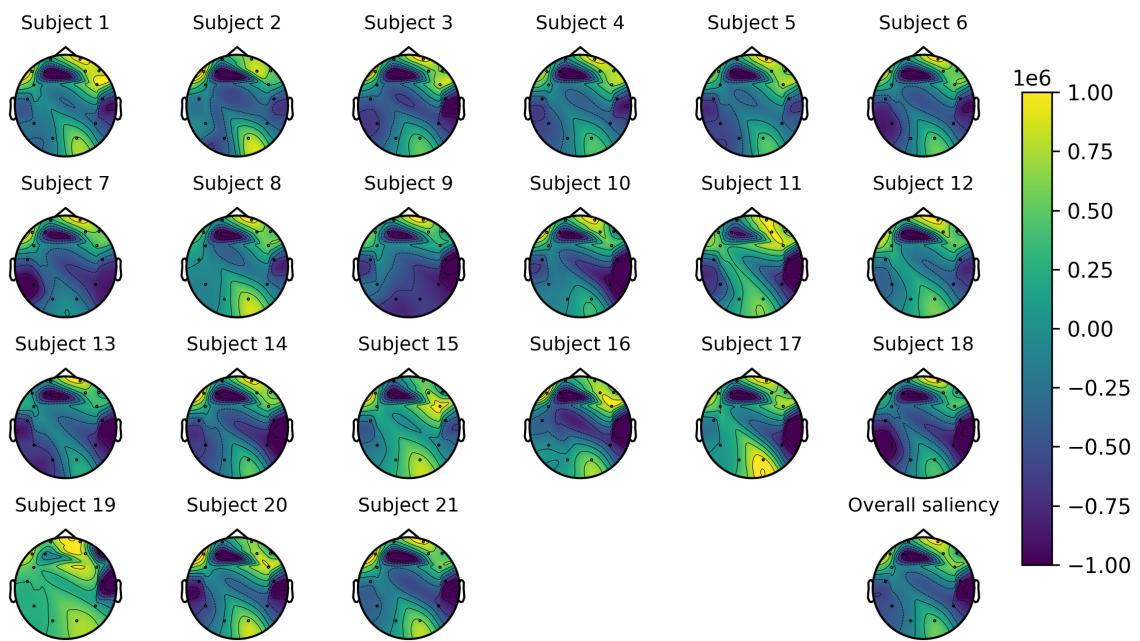


Figure B.14 Saliency by Subject (MI-Pretrained). (Source: Self)

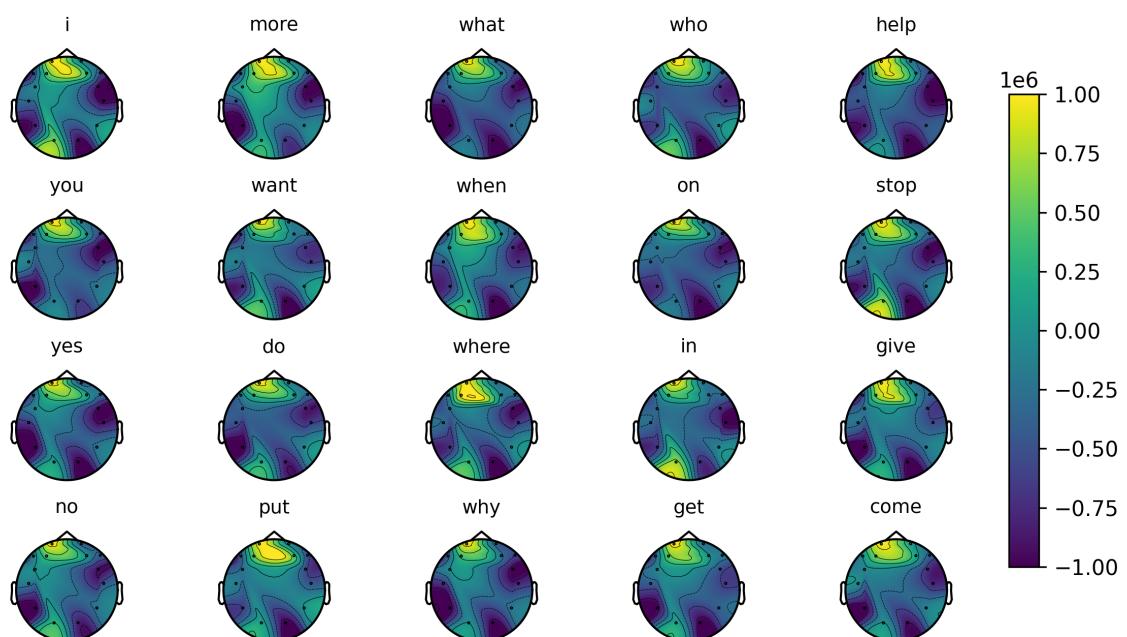


Figure B.15 Saliency by Class (VEP-Pretrained). (Source: Self)

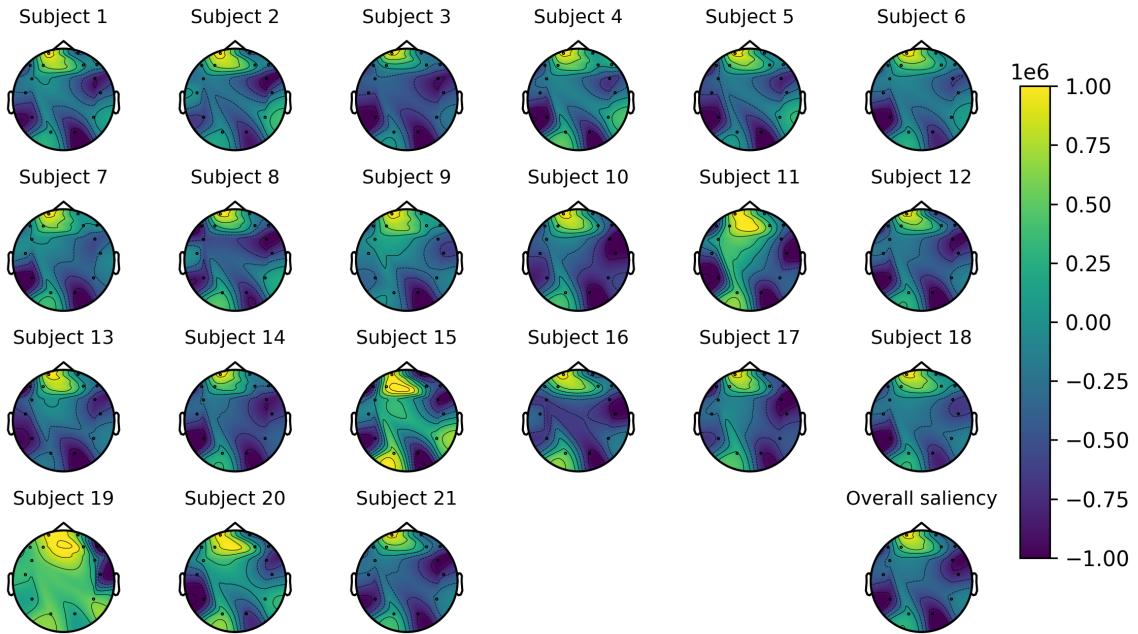


Figure B.16 Saliency by Subject (VEP-Pretrained). (Source: Self)

B.3.2 Time-Series Saliency Maps

The time-series saliency maps shown in Figures B.17, B.18, and B.19, further strengthen the hypothesis of overfitting. Instead of highlighting temporal segments specifically correlating with imagined speech, saliency is consistently and uniformly distributed throughout the entire time period, with only a slight reduction at the edges. This suggests the models are not selectively focusing on key temporal events, but rather exploiting consistent, potentially non-specific, patterns present across the entire duration of the trials.

B.4 t-SNE Plots

To further assess the quality of our model’s learned representations and the separability of different classes, we also generated t-SNE plots. t-SNE (t-Distributed Stochastic Neighbour Embedding) is a non-linear dimensionality reduction technique used for visualising high-dimensional data by mapping it to a lower-dimensional space (typically two or three dimensions) [136]. Its core function is to preserve the relationships between closely related data points, ensuring that points which are similar in the original high-dimensional space remain close together in the low-dimensional visualisation.

In our application, these plots visualise the high-dimensional features extracted by our model, projecting them into a two-dimensional space. This allows for a

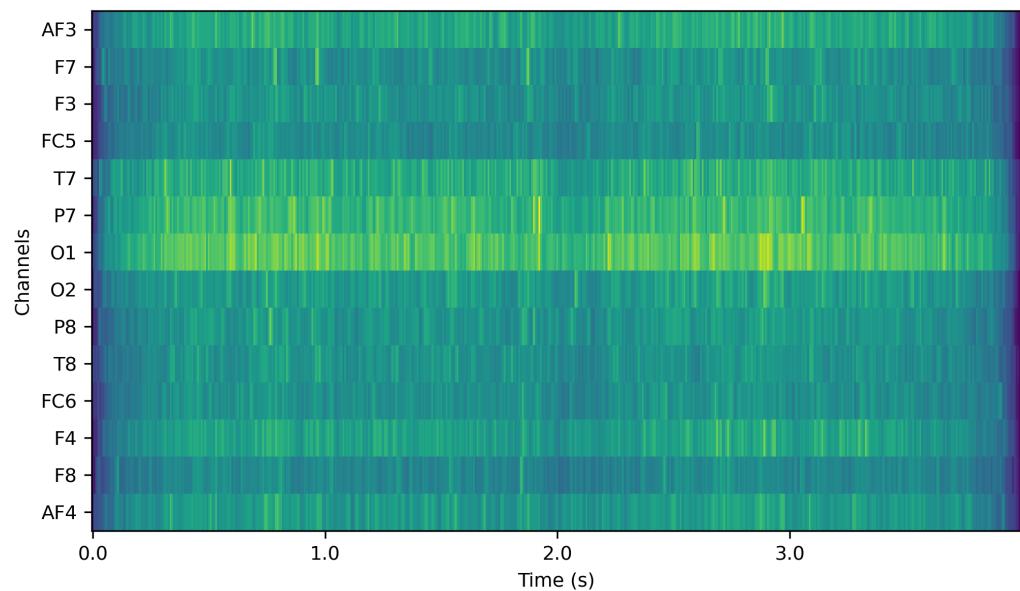


Figure B.17 Time-series Saliency by Subject (Baseline). (Source: Self)

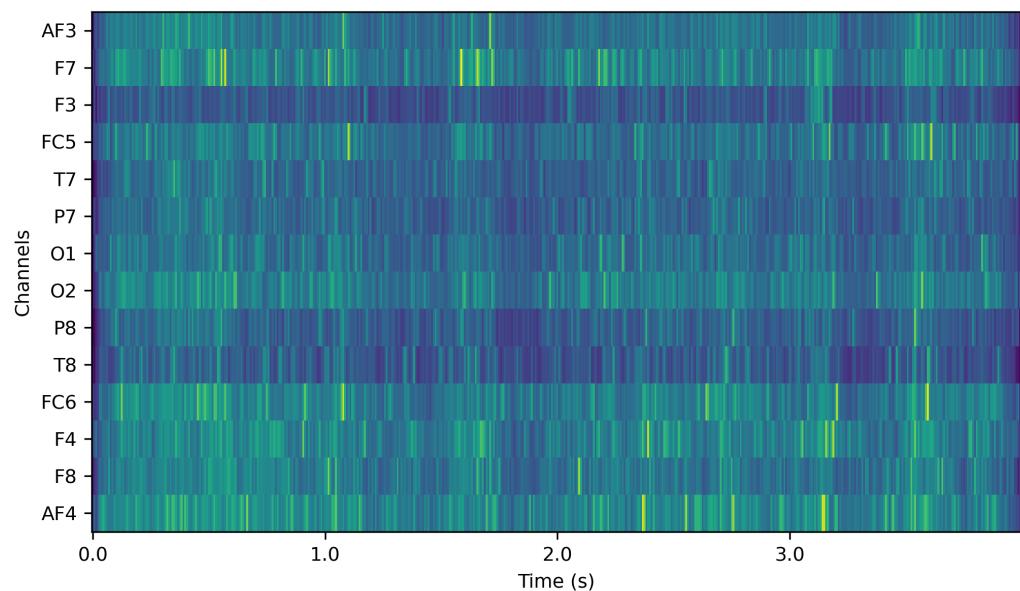


Figure B.18 Time-series Saliency by Subject (MI-Pretrained). (Source: Self)

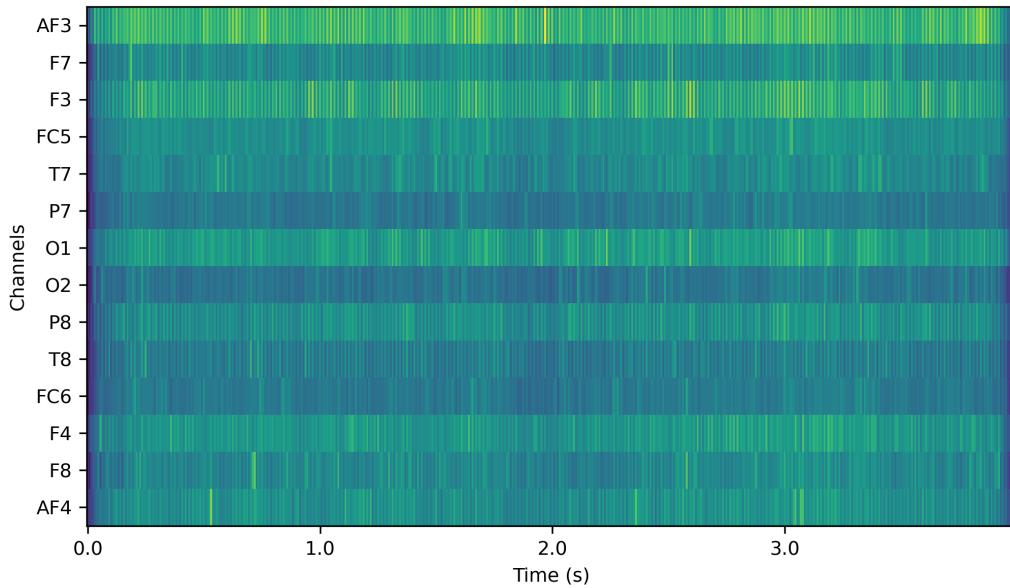


Figure B.19 Time-series Saliency by Subject (VEP-Pretrained). (Source: Self)

qualitative inspection of how well the model’s internal representations distinguish between different classes. By examining the clustering of data points, we can ascertain if the model is forming distinct and well-separated groups for each predicted class. Well-defined clusters for each class would indicate that the model has successfully learned discriminative features, while overlapping or scattered points might suggest challenges in learning clear class boundaries or the presence of intertwined patterns. Additionally, for the pretrained models we provide t-SNE maps from before and after fine tuning to demonstrate the transferability of class features between datasets and the effects of fine tuning.

Figure B.20 presents the t-SNE visualisation of the full dataset after processing through the baseline version of our EEG Deformer model. The plot distinctly illustrates significant class separation, with each word forming its own well-defined and separated cluster. Within these clusters, we observe varying degrees of distinguishability: tightly grouped points representing ‘easy’ classifications, less distinct but still clustered ‘medium’ points, and more scattered ‘hard’ points that the model struggles to classify.

Consistent with the discussion on saliency in Section B.3, this t-SNE plot further highlights potential overfitting. The data points appear to roughly segregate into two groups: those for which the model has overfit, forming highly distinct and separate clusters; and those which the model struggles to classify, appearing as more scattered points that likely do not share the specific ‘noise’ characteristics the model exploited for overfitting the first group.

Similar observations regarding class separation and potential overfitting can be

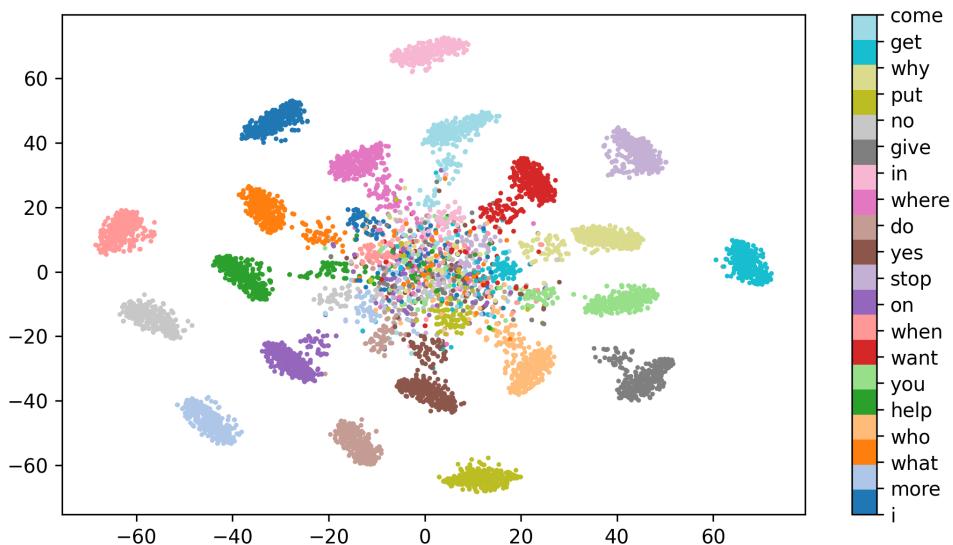


Figure B.20 t-SNE Visualisation of EEG Data Embeddings (Baseline). (Source: Self)

drawn for the MI pretrained model after fine-tuning, as seen in Figure B.21b. Notably, this model's t-SNE plot appears to distinguish between 'easy' and 'hard' data points, with the elimination of the presence of a 'medium' category. We also observe instances of poorly processed data points, where outliers from one class appear interspersed within the clusters of another.

Significantly, the t-SNE visualisation before fine-tuning, as seen in Figure B.21a, reveals a complete inability of the MI-pretrained model to discriminate between classes, resulting in a single, indistinct cluster containing points from all classes. This contrast suggests a lack of immediate transferable features between the MI and SI domains, implying that pretraining an SI model on MI data, at least with the datasets used in this project, may not be a suitable strategy.

Similar observations regarding class separation and the presence of 'easy,' 'medium,' and 'hard' groups can be drawn for the VEP-pretrained model after fine tuning, as seen in Figure B.22b. Notably, unlike the MI-pretrained model, the VEP-pretrained model appears to achieve a more proper distinction of different classes within its respective clusters, suggesting better separation of internal representations for the target classes.

A key distinguishing feature between the MI- and VEP-pretrained models lies in their t-SNE visualisations before fine-tuning, as seen in Figure B.22a. While the MI-pretrained model presented as a single, undifferentiated cluster, the VEP-pretrained model exhibits one dominant large cluster containing data from all classes, along with several smaller, yet mixed, clusters. Although these smaller clusters are composed of data from various classes, indicating limited direct utility for the target

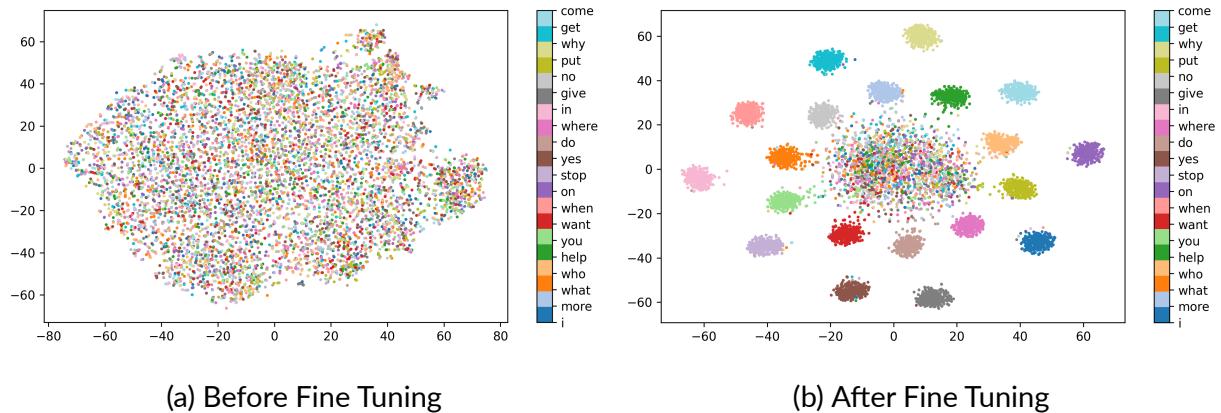


Figure B.21 t-SNE Visualisation of EEG Data Embeddings (MI-Pretrained): (a) Before Fine Tuning, and (b) After Fine Tuning. (Source: Self)

domain, the fact that a few are dominated by a small number of specific classes suggests a subtle, albeit limited, level of positive knowledge transfer between the VEP and VEP datasets prior to fine-tuning.

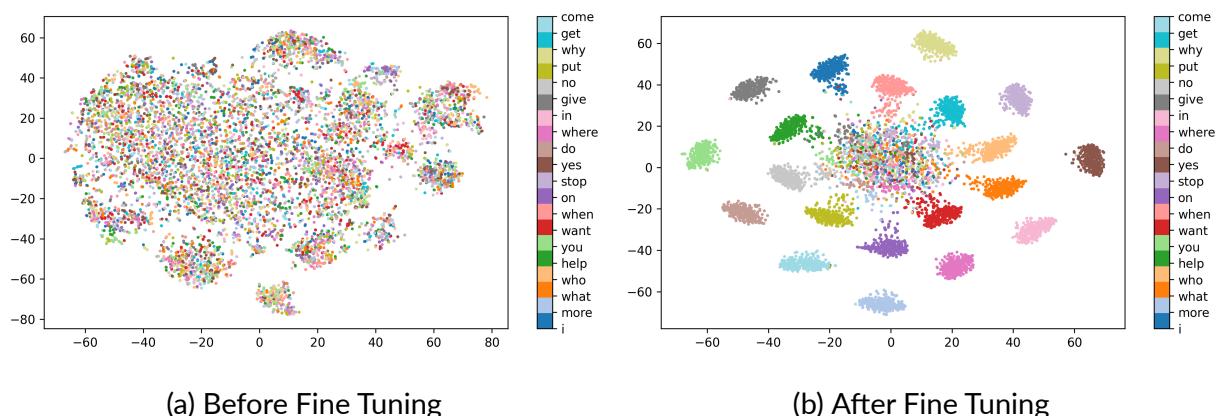


Figure B.22 t-SNE Visualisation of EEG Data Embeddings (VEP-Pretrained): (a) Before Fine Tuning, and (b) After Fine Tuning. (Source: Self)

Appendix C FREC Approval

This appendix contains the Faculty Research Ethics and Data Protection approval form submitted as part of the ethics review process. It includes the completed application form describing the research project, along with the official approval.

Research Ethics and Data Protection Form

University of Malta staff, students, or anyone else planning to carry out research under the auspices of the University, must complete this form. The UM may also consider requests for ethics and data protection review by External Applicants.

Ahead of completing this online form, please read carefully the University of Malta [Research Code of Practice](#) and the University of Malta [Research Ethics Review Procedures](#). Any breach of the Research Code of Practice or untruthful replies in this form will be considered a serious disciplinary matter. It is advisable to download a full digital version of the form to familiarise yourself with its contents (<https://www.um.edu.mt/research/ethics/resources/umdocuments/>). You are also advised to refer to the FAQs (<https://www.um.edu.mt/research/ethics/faqs>).

Part 1: Applicant and Project Details

Applicant Details

Name: Jeremy

Surname: Farrugia

Email: jeremy.farrugia.20@um.edu.mt

Applicant Status: Student

Please indicate if you form part of a Faculty, Institute, School or Centre: * Faculty of Information & Communication Technology

Department: * Department of Artificial Intelligence

Principal Supervisor's Name: * Prof. Alexiei Dingli

Principal Supervisor's Email: * alexiei.dingli@um.edu.mt

Co-Supervisor's Name:

Study Unit Code: * ICT3909

Course Title: * Final Year Project in Artificial Intelligence

Student Number: * 303902L

Project Details

Title of Research Project: * Mind to Message: Textual Decoding of EEG Patterns

Project description, including research question/statement and method, in brief: *

A number of individuals suffer from impairments limiting their ability to communicate with others. With communication being such an important part of our lives, these individuals are significantly disadvantaged when compared to their peers. This project aims to develop a Brain Computer Interface (BCI) capable of translating EEG patterns into text through the use of machine learning techniques applied to data gathered by commercially available EEG bands. Additionally, part of the aim is to create a generalisable BCI which requires minimal calibration when used by individuals not part of the original training set. This project will require an in depth look at current literature covering EEG-based BCIs and current methods used to improve the generalisability of BCIs. Data will be collected from a number of participants during imagined speech. This data will then need to be cleaned and pre-processed before being used to train any model.

A model such as a transformer will be developed to decode the EEG signals. The model will be evaluated on its accuracy and generalisability.

By the end of this project a BCI that can accurately translate EEG signals into text for an individual, regardless of whether they were part of the training set should be produced, opening up new possibilities for communication.

Will project involve collection of primary data from human participants? Yes / Unsure

Explain primary data collection from human participants:

a. Salient participant characteristics (e.g. min-max participants, age, sex, other): *

Healthy Adults, Roughly 20 participants are to be included in the data collection procedure.

b. How will they be recruited (e.g. sampled, selected, contacted, etc.): *

Convenience sampling

c. What they will be required to do and for how long: *

Participants will be asked to be seated down and wear an Emotiv Epoc X wireless EEG headset. As they are wearing this headset they will be provided visual prompts containing text which they will have to repeat to themselves mentally, imagining saying the word/phrase in the prompt. This process will be repeated for 20 different words with short breaks between each word to maintain focus. This will take approximately 20 minutes to complete.

d. If inducements/rewards/compensation are offered: *

N/A

e. How participants/society may benefit: *

Participants may not directly benefit but, the research may benefit others by advancing knowledge in brain-computer interface technology, which could improve communication tools for individuals with speech or motor impairments

f. Is the participant's identity recorded at any stage of the research (e.g. in consent forms, records, publications): *

Identity will only be recorded through consent forms.

g. The manner in which you will manage and store the data and records (including consent forms): *

All raw data that contains potentially identifiable information will be encrypted and stored offline on a secure, external hard drive or flash drive. Any other data relevant to the research will be stored on my password protected laptop with backups taken on external hard drives to ensure no loss of data.

All digital data will be encrypted using secure encryption protocols to prevent unauthorised access. Identifiable data will be stored offline, minimising risk of unauthorised access. Only authorised personnel will have the encryption keys and access to the data.

Will project involve collection of primary data from animals (live non-human vertebrates/cephalopods, their foetuses and larvae, their tissue/samples and/or dead specimens of these animals)?

No

Part 2: Self Assessment and Relevant Details

Human Participants

1. Risk of harm to participants: No / N.A.

2. Physical intervention: No / N.A.

3. Vulnerable participants: No / N.A.

4. Identifiable participants: No / N.A.

5. Special Categories of Personal Data (SCPD): Yes / Unsure

The research involves the collection of a form of biometric data; EEG data, which in some research has been shown to be sufficient to identify an individual.

6. Human tissue/samples: No / N.A.

7. Withheld info assent/consent: No / N.A.

8. 'opt-out' recruitment: No / N.A.

9. Deception in data generation: No / N.A.

10. Incidental findings: No / N.A.

Unpublished secondary data

11. Human: No / N.A.

12. Animal: No / N.A.

13. No written permission: No / N.A.

Animals

14. Live animals, lasting harm: No / N.A.

15. Live animals, harm: No / N.A.

16. Source of dead animals, illegal: No / N.A.

General Considerations

17. Cooperating institution: No / N.A.

18. Risk to researcher/s: No / N.A.

19. Risk to environment: No / N.A.

20. Commercial sensitivity: No / N.A.

Other Potential Risks

21. Other potential risks: No / N.A.

22. Official statement: Do you require an official statement from the F/REC that this submission has abided by the UM's REDP procedures?

No / N.A.

Part 3: Submission

Which F/REC are you submitting to? * Faculty of Information & Communication Technology

Attachments:

- [merged_output.pdf](#) (size: 2.6 MB, uploaded on: 06/02/2025 15:27:18)
- [Information_Letter.docx.pdf](#) (size: 117.8 KB, uploaded on: 06/02/2025 15:27:18)
- [Data_Management_Plan.pdf](#) (size: 84.8 KB, uploaded on: 06/02/2025 15:27:19)
- [Consent_Form.docx.pdf](#) (size: 123.5 KB, uploaded on: 06/02/2025 15:27:19)

- Information and/or recruitment letter*
- Consent forms (adult participants)*
- Consent forms for legally responsible parents/guardians, in case of minors and/or adults unable to give consent*
- Assent forms in case of minors and/or adults unable to give consent*
- Data collection tools (interview questions, questionnaire etc.)
- Data Management Plan
- Data controller permission in case of use of unpublished secondary data
- Licence/permission to use research tools (e.g. constructs/tests)
- Any permits required for import or export of materials or data
- Letter granting institutional approval for access to participants
- Institutional approval for access to data
- Letter granting institutional approval from person directly responsible for participants
- Other

Please feel free to add a cover note or any remarks to F/REC

UREC-DP2412011ICT - ICT-2024-00272

The "merged_output.pdf" file contains all the 3 documents with changes tracked, whereas the individual files are provided as they would be seen by participants.

For convenience the addressed points are provided:

Regarding the provided feedback points I have done the following:

1.1 - The approximate number of participants expected to be engage for this study is 20.

2.1 - The disposal of identifiable data has been clarified, data will be fully anonymised and identifiable information destroyed upon the completion of the research process. This process expected to be completed by the end of May this year and data deleted within the following month. This has been clarified in the data management plan and information letter.

2.2 - The primary researcher's (my) daytime phone number has been added to both the information and consent forms

2.3 & 2.4 - Point 7 has been removed and the treatment of data clarified in both the consent form and the information letter.

2.5 - Storage of data has been clarified in both the information letter and the consent form. Data will be stored on an external device such as a hard drive or flash drive accessible only by the primary researcher (myself).

Declarations: *

- I hereby confirm having read the University of Malta Research Code of Practice and the University of Malta Research Ethics Review Procedures.
- I hereby confirm that the answers to the questions above reflect the contents of the research proposal and that the information provided above is truthful.
- I hereby give consent to the University Research Ethics Committee to process my personal data for the purpose of evaluating my request, audit and other matters related to this application. I understand that I have a right of access to my personal data and to obtain the rectification, erasure or restriction of processing in accordance with data protection law and in particular the General Data Protection Regulation (EU 2016/679, repealing Directive 95/46/EC) and national legislation that implements and further specifies the relevant provisions of said Regulation.

Applicant Signature: * Jeremy Farrugia

Date of Submission: * 06/02/2025

If applicable: Date collection start date

Administration

REDP Application ID ICT-2024-00272

Current Status Approved

If a submitted application needs to be amended, it can be withdrawn, edited, and resubmitted, and it will retain the same reference number. There is no need to submit a new application.

Appendix D Data Management Plan

This appendix includes the data management plan that was submitted to the Ethics Committee of the Faculty of ICT at the University of Malta as part of the ethical review process. The document outlines how research data would be collected, stored, protected, and managed in accordance with institutional and legal data protection guidelines.

Data Management Plan

Mind to Message: Textual Decoding of EEG Patterns

Data Collection and Types of Data

- **Data Types:** The study will collect EEG data and any metadata relevant to the participant demographics (e.g., age, gender).
- **Data Collection Methods:** Data will be gathered through EEG recordings where participants will be asked to imagine saying words or phrases in a controlled setting, any metadata will be gathered through questions before the procedure.

Data Confidentiality and Anonymisation

- **Confidentiality Measures:** All collected data will be treated with strict confidentiality to protect participants' privacy.
- **Anonymisation Procedures:** After data collection (during the research process), all identifying information (e.g., names, contact details) will be replaced with unique participant codes. EEG data and any metadata will be linked to these anonymous codes rather than personal identifiers. Once the research process is complete, any identifying information (such as consent forms) will be destroyed and data fully anonymised
- **De-Identification:** To further anonymise, any sensitive metadata that could link data back to individuals will be removed or generalised (e.g., grouping ages rather than specific birth years).

Data Storage

- **Raw Identifiable Data:** All raw data that contains potentially identifiable information will be encrypted and stored offline on a secure, external hard drive or flash drive.
- **Research Data:** Any other data will be stored on the primary researcher's password protected laptop with backups taken on external hard drives to ensure no loss of data.
- **Access Controls:** Only the primary researcher (myself) and my academic supervisor will have access to the encrypted files. In rare cases where necessary, examiners may be granted temporary access to anonymised data.

Data Access and Sharing

- **Access During Research:** Only the primary researcher and supervisor will have access to the raw, identifiable data. In exceptional cases, anonymised data may be shared with examiners upon request.
- **Data Sharing Post-Analysis:** After analysis, fully anonymised and aggregated, data may be included in publications or as part of the final thesis to promote transparency and reproducibility, or to allow use in future research. No personally identifying information will be included in any shared or published materials.

Data Security Measures

- **Encryption:** All digital data will be encrypted using secure encryption protocols to prevent unauthorised access.
- **Offline Storage:** Identifiable data will be stored offline, minimising risk of unauthorised access.
- **Access Restriction:** Only authorised personnel (myself, my supervisor, and in some cases, examiners) will have the encryption keys and access to the data.

Data Retention and Disposal

- **Identifiable Data Disposal:** Raw identifiable data will be securely deleted or destroyed after it is no longer necessary for the project, which will be done through secure deletion methods for digital data and shredding for hard copies. The research process is expected to be completed by the end of May this year and any identifiable data will be disposed of within the following month.
- **Anonymised Data Storage:** Anonymised data may be retained for educational or research purposes, ensuring that no personally identifiable information is included.

FREC ETHICS FORM CODE

ICT-2024-00272

Appendix E Information Letter

This appendix contains the information letter that was provided to all potential participants prior to obtaining their consent. The letter outlined the purpose of the study, what participation involved, and the voluntary nature of participation, allowing individuals to make an informed decision about taking part.

Providing this letter was part of the ethical protocol approved by the Ethics Committee of the Faculty of ICT at the University of Malta, and served to ensure that participants were fully informed prior to consenting



Information letter

13/02/2025

Dear Participant

My name is Jeremy Farrugia and I am a student at the University of Malta, presently reading for a Bachelors of Science in Information Technology (Artificial Intelligence). I am conducting a research study for my thesis titled Mind to Message: Textual Decoding of EEG Patterns; this is being supervised by Prof. Alexiei Dingli. This letter is an invitation to participate in this study. Below you will find information about the study and about what your involvement would entail, should you decide to take part.

The aim of my study is to develop and improve brain-computer interface technology that can interpret expressive imagined speech from EEG data. Your participation in this study would help contribute to a better understanding of how brain activity patterns correspond to specific thoughts and intentions, potentially leading to new communication methods for individuals with speech or motor impairments.

Should you choose to participate, you will be asked to sit down and wear an Emotiv Epoc X wireless EEG headset. As you are wearing this headset you will be provided visual prompts containing text which you will have to repeat to yourself mentally, imagining saying the word/phrase in the prompt (expressive imagined speech). This process will be repeated for 20 different words with short breaks between each word to maintain focus. This will take you approximately 20 minutes to complete.

All data collected for this study will be treated with strict confidentiality and will be pseudonymised, and once the research process is complete fully anonymised to ensure that individual participants cannot be identified. Only the primary researcher (myself) and my academic supervisor will have access to the raw data during the research process. After analysis, anonymised data may be shared in publications or as part of the final thesis to support transparency and reproducibility, but no personally identifying information will be included. Raw identifiable data will be encrypted and stored offline on an external hard drive or flash drive, accessible only to the primary researcher (myself). Any material in hard copy form will be placed in a locked cupboard. Only my supervisor and myself (and in exceptional cases, examiners) will have access to this data.

The findings which emerge from this research may be published (e.g., in a dissertation, academic journals) and/or presented (e.g., during conferences, meetings). Your name (or any other identifying information) will not appear when the findings are reported.

Participation in this study is entirely voluntary; in other words, you are free to accept or refuse to participate, without needing to give a reason. You are also free to withdraw from the study at any time, without needing to provide any explanation and without any negative repercussions for you. Should you choose to withdraw, any data collected from your interview will be erased as long as this is technically possible (for example, before it is anonymised or published), unless erasure of data would render impossible or seriously impair achievement of the research objectives, in which case it shall be retained in an anonymised form.

Any identifying information (such as consent forms) will be destroyed once the research process is complete and all pseudonymised data will be fully anonymised. The research process is expected to be complete by the end of May this year and data will be fully anonymised following this.

If you choose to participate, please note that there are no direct benefits to you. Your participation does not entail any known or anticipated risks.

Please note also that, as a participant, you have the right under the General Data Protection Regulation (GDPR) and national legislation to access, rectify and where applicable ask for the data concerning you to be erased.

All data collected will be stored in an anonymised form on completion of the study, allowing use in future research.

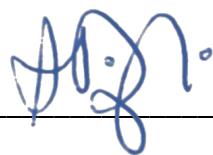
A copy of this information sheet is being provided for you to keep and for future reference.

Thank you for your time and consideration. Should you have any questions or concerns, you may contact the primary researcher (myself) or my supervisor on the details provided below.

Yours Sincerely,
Jeremy Farrugia



Jeremy Farrugia
jeremy.farrugia.20@um.edu.mt
99109983



Prof. Alexiei Dingli
alexiei.dingli@um.edu.mt

Appendix F Consent Form

This appendix includes the blank consent form that was provided to all participants prior to their involvement in the study. The form outlines the study's purpose, procedures, and participants' rights, including the right to withdraw at any time without consequence.

The research adhered to established ethical guidelines, and informed consent was obtained from all participants in accordance with the approval granted by the Ethics Committee of the Faculty of ICT at the University of Malta.

Participants were initially approached on a convenience basis and provided with the information letter (see Appendix E), which outlined the study and invited them to indicate their interest in taking part. Those who responded positively were then given an overview of the data collection procedure and the consent form was subsequently provided.

Prior to data collection, participants were shown the EEG device and given a detailed explanation of the procedure. They were reminded of their right to withdraw at any time, including just before the session began, ensuring that participation remained entirely voluntary throughout.



Participant's Consent Form

Mind to Message: Textual Decoding of EEG Patterns

I, the undersigned, give my consent to take part in the study conducted by Jeremy Farrugia. This consent form specifies the terms of my participation in this research study.

1. I have been given written and/or verbal information about the purpose of the study; I have had the opportunity to ask questions and any questions that I had were answered fully and to my satisfaction.
2. I also understand that I am free to accept to participate, or to refuse or stop participation at any time without giving any reason and without any penalty. Should I choose to participate, I may choose to decline to answer any questions asked. In the event that I choose to withdraw from the study, any data collected from me will be erased as long as this is technically possible (for example, before it is anonymised or published), unless erasure of data would render impossible or seriously impair achievement of the research objectives, in which case it shall be retained in an anonymised form.
3. I understand that I have been invited to participate in the collection of EEG data during imagined speech in which the researcher will gather brain signal data during expressive imagined speech to investigate the efficacy of a new approach to EEG to text transcription. I am aware that the collection process of the EEG data will take approximately 20 minutes. I understand that the collection of EEG data is to be conducted in a place and at a time that is convenient for me.
4. I understand that my participation does not entail any known or anticipated risks.
5. I understand that there are no direct benefits to me from participating in this study. I also understand that this research may benefit others by advancing knowledge in brain-computer interface technology, which could improve communication tools for individuals with speech or motor impairments. This work could lead to new methods for helping people who cannot speak or move independently communicate more effectively and interact with their surroundings.
6. I understand that, under the General Data Protection Regulation (GDPR) and national legislation, I have the right to access, rectify, and where applicable, ask for the data concerning me to be erased.
7. I am aware that my identity and personal information will not be revealed in any publications, reports or presentations arising from this research.
8. I have been provided with a copy of the information letter and understand that I will also be given a copy of this consent form.
9. I am aware that my data will be pseudonymised, i.e., my identity will not be noted anywhere, but instead, a code will be assigned. The codes that link my data to my identity will be stored securely and separately from the data. Raw identifiable data will

be encrypted and stored offline on an external hard drive or flash drive, accessible only to the primary researcher (myself). Any material in hard copy form will be placed in a locked cupboard. Only my supervisor and myself (and in exceptional cases, examiners) will have access to this data.

I have read and understood the above statements and agree to participate in this study.

Name of participant: _____

Signature: _____

Date: _____



Jeremy Farrugia
jeremy.farrugia.20@um.edu.mt
99109983



Prof. Alexiei Dingli
alexiei.dingli@um.edu.mt