Toward Restoring Communication in ALS: A Non-Invasive EEG Imagined-Speech Translation Model and Device

Alex Steiner[0009-0002-3436-0605]1

1 H-Farm, Treviso TV 3486, ITA  
alex.steiner@student.h-is.com

Abstract. *Background:* Amyotrophic lateral sclerosis (ALS) destroys motor function and speech, thereby leaving patients locked-in. Imaging-speech decoding non-invasive brain–computer interfaces (BCIs) using EEG (electroencephalogram) are promising, yet they suffer from low accuracy and bulky hardware. *Objective:* To develop and evaluate an entire non-invasive EEG system, a dry micro-needle headset, and a sequence-to-sequence model to decode imagined speech into open-vocabulary text for ALS communication restoration. *Methods:* *Hardware:* Created a 48-channel dry micro-needle array EEG set covering Broca's, Wernicke's, and frontal regions, minimized needle impedence, and 75% Au 25% Pt alloy for theretical ≈5 kΩ skin–electrode impedance. *Model:* A three-stage pipeline(1) per-region 1D convolutional attention mechanism, (2) transformer brain-region encoder with learned fusion, and (3) cross-modal BART decoder end-to-end fine-tuned on EEG–text pairs. *Results:* *Decoding:* Achieved 49.6% BLEU-1, 14.3% BLEU-4, and 36.7% ROUGE-L F1—outperforming closed-vocabulary baselines and comparable to spoken-speech EEG benchmarks. *Hardware:* Demonstrated viable micro-needle fabrication and PCB (printed circuit board) integration, with <5 kΩ per-channel impedance and seamless compatibility with ADS1299 ADC and STM32H743 MCU (microcontroller unit). *Conclusions:* This is the first non-invasive imagined-speech EEG-to-text system that combines high-density micro-needle hardware and a deep learning model achieving high benchmarks. A crucial step toward practical BCI communication for3 ALS, facilitating personal usage, real-world deployment, and giving speech back ones more to who lost it.

Keywords: Imagined Speech Translation, ALS, Machine Learning, Non-invasive BCI, Microneedle Array

1. Introduction
   1. Background And Motivation

Throughout the last centuries, each advancement came together with a development in communication, at first in …, however since the invention of the phone and SMS, recently in i

One of the most promising frontiers of human–computer interaction is the seamless integration of the physical world and digital interfaces. Advances in the field have yielded increasingly natural, context-aware systems. For example, dialogue agents such as ChatGPT and DeepSeek can now conduct human-like conversations, and voice-based assistants as Siri and Gemini are integral parts of our daily lives, relying on sophisticated speech-recognition deep learning models that accurately translate spoken words into text, enabling devices to understand and reply to queries more humanely than ever.

Although significant recent advances have been made in the field of assistive technology, all existing methods rely on the analysis of visible behavioral output and thus require users to produce specific activities or performances, whether these are spoken or motorial. While this is not a limitation for the majority of individuals, it is impossible when the control over certain muscles of the body is lost, such as in the case of inelastic motor paralysis caused by injury or neurodegenerative illness. Specifically, loss of speech can occur due to injury or disease of motor control, speech production, or language processing. Stroke, traumatic brain injury, and amyotrophic lateral sclerosis (ALS) are common conditions that affect patients so severely that they cannot communicate anything at all due to a complete loss of voluntary motor function. As such, assistive technology is either not available to them or is limited, even though it's such individuals who are more in need of the technology, and yet cannot take advantage of it.

To address these issues, through Brain Computer Interfaces (BCIs), new channels of direct communication between the brain and a given technological tool are been developed. As these proposals are not based on or require analysis of a given behavioral output; rather they create a bi-directional communication interface that reads the signals generated by the human brain and converts them into a desired cognitive task, allowing seamless interaction with assistive technology for those affected by motor or speech disabilities, ensuring communication directly with a computer (and thereby other individuals) without the need to move any body parts to type a message or produce speech out loud. The goal of this research is to provide a direct communication channel for connecting the human brain with a computer interface without the need for any behavioral output, and to do this without the use of invasive techniques to preserve the usability of the device. It is more than significant to us to enhance the quality of life of less fortunate people with injuries or disabilities; thus, our mission is to develop a high-performing, adaptive, and non-invasive system able to decode imagined speech to permit those people to talk and express their thoughts once more.

## 1.2. ALS And Communication Loss

ALS is the most common type of neurodegenerative disease, which affects motor neurons (neurons responsible for voluntary and involuntary muscle activation) both inside the cerebrum and spinal cord. As motor neurons degenerate and die, they stop sending messages to muscles, causing the muscles to weaken, start twitching, and atrophy (lose mass). In more advanced phases, ALS impairs voluntary movements, such as walking, talking, chewing, and other motor functions, as well as breathing. The weakness most commonly starts in the limb muscles, more often in the distal muscles than in the proximal muscles. In 25-30% of cases, ALS presents with bulbar onset characterized by dysarthria, dysphagia, dysphonia, or, more rarely, masseter weakness [1]. With a median survival of approximately three years after symptom onset, death is mostly attributed to respiratory failure. Approximately 50% of patients suffer from extra-motor manifestations to some degree in addition to their motor problems, including speech production, which is one of the first areas hit by the disease [1]. Approximately 25-30% of patients develop frontotemporal dementia in the first phase, a percentage that rises to 50-60% in later phases of the disease. Frontotemporal dementia is characterized by degeneration of the frontal and anterior temporal lobes and presents clinically as behavioral changes, impairment of executive functioning, and/or language impairment [2].

The epidemiology of ALS shows that the most common neuropathological signature is cytoplasmic aggregation of TDP-43, a protein encoded by TARDBP, found in more than 95% of ALS cases [2]. TDP-43 is an RNA and DNA-binding protein involved in multiple processes such as transcription, splicing, RNA maturation, and stress granule formation. Similar to other neurodegenerative diseases [2], ALS is caused by a combination of genetic and environmental factors, as well as aging-related dysfunction. Overall, even if there is a low correlation to which indicates a genetic inheritance of the disease, its appearance is still considered sporadic, meaning that the disorder seems to happen at random with no associated risk factors except for aging [2]. In addition, ALS is mainly a clinical diagnosis, and early uncertainty leads to a mean time from symptom onset to diagnosis of approximately one year. The updated research on the epidemiology of the global incidence of ALS is approximately 1–2.6 cases per 100,000 persons annually, with a prevalence of approximately 6 cases per 100,000 [2]. Furthermore, the mortality rates for males decreased over this period, whereas there was no change in the rate for females. Analysis of age-specific results showed a slight increase in the death rate for adults 20–49 years of age, but a decrease for those aged ≥ 65 years well as for males [2]. The symptomatology of this age group remains quite high, with 50-60% of patients developing dementia, with behavioral deficits, such as linguistic impediment (the focus of our research), and many other symptoms, after the first year of diagnosis.

Speech and communication support, therefore, constitute a very valuable tool, if not fundamental, for patients affected by ALS and for the supporting channels that are made more effective by a fluid and rapid communication. As of now, computer-based speech synthesizers mainly use eye-tracking devices that allow a person to use the Internet and type on custom screens to communicate, such as the *CallText 5010 DECtalk* text-to-speech synthesizer used by Stephen Hawking. Our research aims to simplify and make the integration more natural for the patients, building a direct and most important fast challenge of communication that is supported only by the detection and analysis of the EEG waves, removing the barrier constituted by the detection of any behavioral output, such as speech production or eye movements. As eye movement may also be affected by the disease.

## 1.3 Imagined Speech Decoding: Invasive And Non-Invasive

As already outlined, the scope of this work is to take advantage of non-intrusive methods of obtaining data, presenting a portable, safe, and morally viable solution that is capable of extracting the neural signals of silent speech in the absence of the risk or logistical disadvantage of implanted surgery. Although invasive measurements such as electrocorticography and intracortical microelectrodes consistently reach classification rates of 70 % or greater, they are doing so at the expense of the risk of surgery, cost, and user burden. In contrast, non-intrusive modalities—MEG, fMRI, fNIRS, and EEG—are capable of a different tradeoff of spatial definition, temporal acuity, and usability that makes them better situated for mass use in brain–computer interfaces.

The functional MRI produces exquisite spatial maps of activated parts of the brain but relies on hemodynamic signals involving delay of the order of seconds, which is thus bad for the capture of the millisecond‑spersed dynamics of imagined speech. The MEG, then, provides high temporal fidelity but is compelled to maintain a static, motionless set of sensors in a magnetically shielded room—constraints that significantly prohibit portability and practical‑world applicability. The functional near‑infrared spectroscopy (fNIRS), being more portable and user‑friendly, samples low blood‐oxygen‐change rates as well, and is beset with poor cortical penetration. Of the other non-invasive methods, the most convenient EEG modality is imagined-speech decoding. EEG systems capture the sampled activities of the neurons in millisecond time resolutions, and they record the rapid time course of internally self-generated speech accurately. Adjustments in amplifier building and transmission in radio waves have provided us with transportable, wearable headsets used in a litany of environments—that is, research laboratories to home living rooms—that accommodate long training sessions and high usage frequency with low infrastructure. In addition, the commensurability in the cost of the EEG devices makes entry less expensive, allowing for larger group sizes and algorithmic iterations more often.

2. Methods

2.1 Overview & Challenges

The primary concern of decoding internally imagined speech into text is the practical application of brain–computer interfaces. The rationale for that is, namely, the internal low signal‑to‑noise ratio of the EEG patterns generated internally—such that the speech‐correlate neural signals are not easily separable and reliably recoverable—and the absence of large, well‑annotated datasets spanning the whole range of imagined phonemes, prosody, and interindividual variability. Defeating the two challenges is required for the extraction of strong decoding schemes and the move toward reliable imagined‐speech communication systems.

Currently, studies done to classify imagined speech signals have shown that the main brain activations in the *Wernicke* and *Broca* areas, which are already known to be responsible for speech comprehension and reproduction, seem to be active during imagined speech processing. In addition to *Broca*’*s* and *Wernicke*’*s* areas, several structures of both the left (“verbal” hemisphere for right-handed people) and right hemispheres are involved in speech activity, both real verbalization and inner or covert speech [3]. These new findings suggest that the activation of brain areas does not vary consistently for spoken speech activities and imagined speech ones. The main difference, however, is believed to be reduced only to peculiarities of interaction between several brain areas. In both areas involved in speech verbalization, both inner and outer speech are inferior frontal and supramarginal gyri, areas of the white matter adjacent to the *left supramarginal gyrus*, which are part of the *dorsal attention network* that is believed to play a crucial role in the elaboration of speech content. This remains a crucial limitation because of the complexity in establishing the precise areas responsible for the elaboration of speech content [4].

Figure 2.Human brain circling Broca’s and Wernicke’s areas. Source: Wikimedia Commons.pasted-image.tiff

A comparative analysis showed that the recognition accuracy of electrographic patterns recorded from neocortical structures and associated with internal speech presented a BLUE-1 score of only 0.3–0.35, whereas that of real speech achieved 0.7-0.75. The latter may be due to the presence of muscle artifacts in the recorded activity associated with speaking and specific to each word, and the effects associated with the processing of reverse afferentation, including those from muscle, tendon, and joint receptors activated during speech execution [4]. Furthermore, these different studies based on different approaches obtained quite different results, even though very few approaches reached more than 0.4, whereas with spoken speech, the level of accuracy also reached 0.85-0.9 [5]. This is because of many factors, including the low-architecture ML models and training used in these investigations. Nevertheless, this presents a significant challenge, as patterns of brain activity used for classification are believed to be associated with speech activity itself, but not with its content, particularly with individual words, which are currently described in sufficient detail. Largely because when studying the mechanisms and electrographic manifestations of speech activity, researchers often limit themselves to individual neocortical areas, in particular the sensorimotor cortex. Thus, without proper muscle activation that comes with speech performance, the detection of imagined speech brain signals becomes significantly harder.

An additional challenge is added upon with the lack of a proper English database for labelling imagined speech signals, and hence training a supervised sequence-to-sequence model. Currently, only three major datasets for the designated task are available: *“Thinking Out Loud”, “Imagined Speech”,* and *Chisco* [6][7][22]. The most relevant one, *“Thinking Out Loud,”* has several limitations: the study was conducted in Spanish; additionally, the database was designed to classify only four words: *arriba* (up), *abajo* (down), *derecha* (right), and *izquierda* (left). The experimental setup included three conditions: inner/pronounced speech and visualized condition. During the imagined speech condition, participants were asked to imagine their voice, repeating the corresponding word. In the pronounced speech condition, they were asked to repeat the words aloud in response to visual prompts. The visualized condition involved participants focusing on mentally moving a circle displayed in the center of a screen in the direction indicated by a visual cue. We emphasize that the dataset's size is not sufficient to develop the prototype model we're planning. Moreover, Spanish is not ideal for building this dataset because of its longer words with more syllables compared to English. Additionally, just four words are far from enough to cover the variety needed to translate entire phrases. Therefore, a key step in our project is creating an actual dataset for training the model. We will detail our plan for this in the section on the experimental setup. Nevertheless, for the current state of the research, the model and training presented in the upcoming sections were done on the Chisco dataset, the largest imagined speech dataset available, with 5 patients recording EEG close to 30,000 imagined speech sentences, though in Mandarin Chinese [22].

2.2 Experimental Protocol & Dataset Collection

### 2.2.1 Experimental Protocol

As we have outlined in the previous section, the lack of a proper well well-structured existing English database requires us to develop a new reliable dataset ourselves. In this section, we are going to discuss and propose the experimental setup that is going to be employed for the construction of a dataset.

Starting from the dataset size, to obtain a sufficient amount of data to train the model, the number of samples that we need to collect is from a minimum of 10,000 to a desirable amount of 50,000 sample imagined speech sentences in English. Each patient volunteer will have to sign a consent agreement in which they agree to undergo the standard procedure and are aware of the possible risks involved, complying with the ethical regulations. After signing, the subject will be wearing the EEG headset (the specifics of the equipment and its construction are described in the upcoming sections). The patient will be located in a semi-dark room, without distractions that could interfere with the experimentation, thereby avoiding the unnecessary activation of other areas of the brain, which would add noise to the signal. He/She will sit in an armchair in front of a screen where there will be a sentence shown at a time (black font color and white background for contrast and clarity). The researcher will be sitting outside the room, out of the subject's field of view, not to interfere with the experimentation [8].

The subject will be shown one sentence at a time, for which he will have 10 seconds to memorize, after the screen is turned off and the subject is asked to imagine as if the sentence were being pronounced aloud, though without actually saying it, to record imagined speech. Considering 20 seconds per sentence recording. The whole procedure is directed by a light on the bottom of the screen indicating the task (*red for memorizing; green for imagining aloud).* Avoiding interference with the EEG measurements, as well as potentially the decoding of the researcher's voice, could activate the same regions involved in imagined speech production and hence add additional noise to the signal.

### 2.2.2 Dataset Collection

Another relevant challenge that we want to address in this paper is the structure of the datasets, based on pairing EEG waves with sentences corresponding to the imagined speech. Few datasets of this structure are available today following this metric, particularly lacking an extensive English-based dataset, as to our best current knowledge, the only fulfilling option—though in Mandarin Chinese— is the Chisco dataset [22]. This choice was motivated as, from a computational point of view, considering the problem as a stream of EEG data fed to a deep learning model in time sections, determining the probabilistically most likely token or word in a sentence, inspired by the training of Large Language Models *(LLMs)*.

Nevertheless, this raises a consistent number of challenges, regarding the number of sentences necessary to collect to create ensure proper communication and stable training, avoiding overfitting. We must thereby consider that modern English—following the Oxford English Dictionary—consist of approximatlynumber 170,000 words, with an average lenght of eight words per sentence, taking into consideration grammatic and semantic constrains, the number of possible phrases is overwhelmingly large.

As such, the necessety of constraining the dataset to a range of sentences that the model will be trained on, that ensures proper communication range for the application of the assitive technology, though levereging next token prediciton, meaning that the model generalizes and predicts well even on unseen sentences, thereby removing the necessety of recording slight variations of a given sentence *(e.g. “Hey, how are you?”* And *“Hello, how are you?”)*. This approach ensure that the model is able to decode general conversations without the limitation of a specific sector of semantic as a result of tokenization and sentence embedding. Nevertheless, the number of words used in the sentences is a precompiled list of 1000, that can be found here. Generally including predicates, nouns, and adjectives.

As an overall evaluation of this section, we must estimate the extent of the resources that will be required in order to complete this sampling phase. As exposed in the experimental setup, the sampling process of one sentence takse 20 seconds per subject, in addition to that, in order to have enough data to train the model, covering diverse patiens, we need at least twelve subjects—composed of three ages group *(teenagers, adults, and elderly)* in which the two sexes *(male and female),* and the predominant brain hemispheres *(left and right)* are repressented. As such the total recording time for all subjects is approximately 65 hours.

2.3 EEG Hardware & Micro-Needle Design

### 2.3.1 Signal Acquisition

Electroencephalography (EEG), due to its high temporal resolution, low cost, and non-invasive nature, makes it the best approach for recording brain activity for most researchers. However, when applied to imagined speech, the low signal-to-noise ratio (SNR), due to the absence of muscular activations usually involved in spoken speech operations, and its non-invasive nature, results in higher impedances, making the training on such data more complex, as the model struggles learning semantically. Imagined speech EEG signals are harder to decode than spoken speech ones, also due to the distinction between the background brain activity, as the areas in which the thought process stems from are not fully comprehended yet.

As such, classical machine learning methods that have proven to be successful in the recognition of motor imagery tasks have not obtained good performance when applied to imagined speech recognition. Nevertheless, deep learning models have proven to be much more effective in recognizing imagined speech EEG signals. Thus, our approach will discuss the usage of similar and improved models.

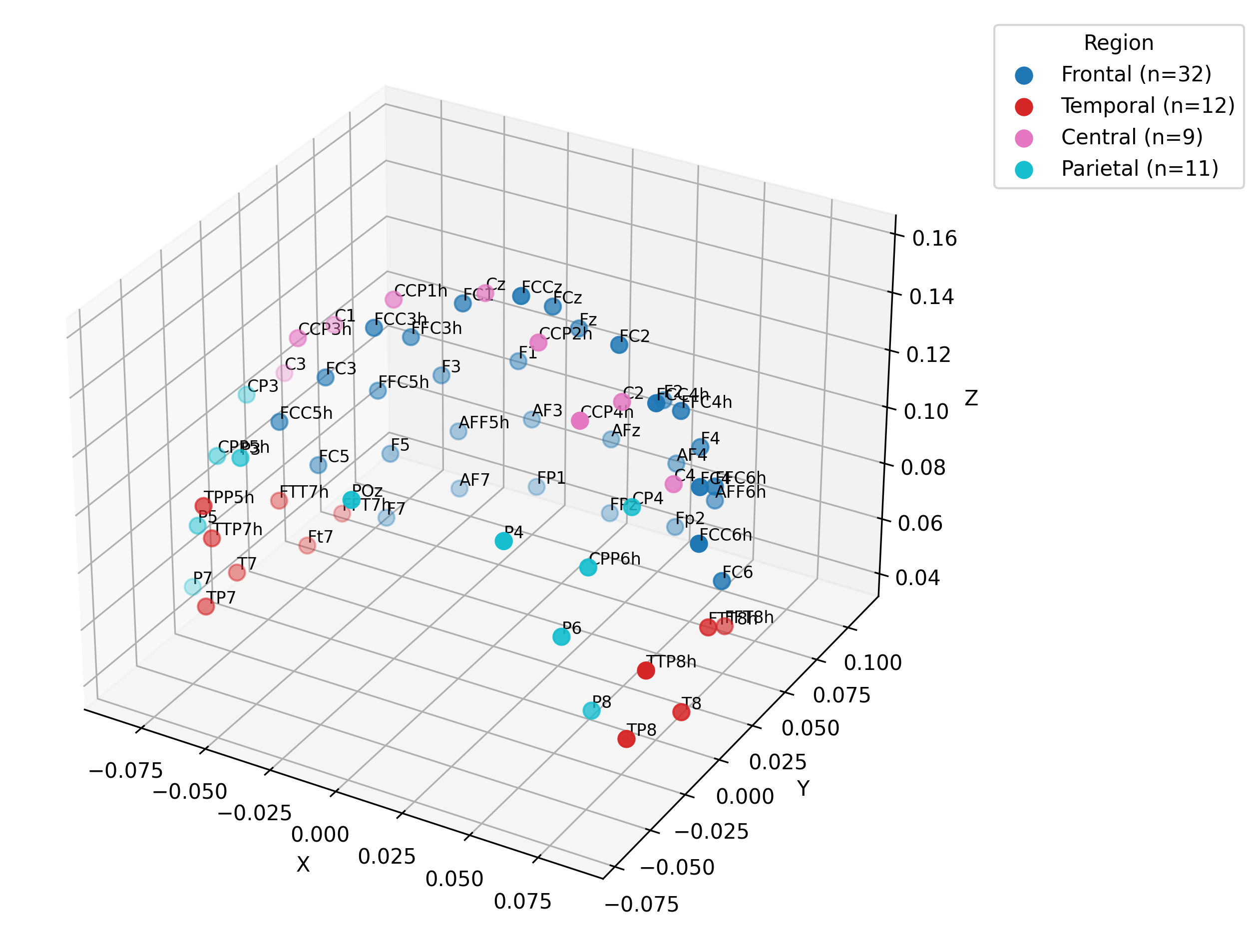
Due to the end of the lack of certainty regarding the specific areas where speech is created, a crucial point to address is the placement of the electrodes on the scalp. Placing them on the entire surface of the scalp would only decrease the user friendliness of the EEG and, in all likelihood, would not bring a substantial advantage in terms of captured signals, which would be significant for our research, as the deep learning model would struggle learning with large amounts of channels. As such, the placement of the electrodes needs to follow a qualitative measure, while following fundamental parameters. As a result, we decided to set a threshold of 48 electrodes that will be placed in specific areas of the scalp that we will analyze, as indicated in *Table 1* below. Nevertheless, still favoring in terms of quantity the preferred areas such as Broca’s and Wernicke’s, as outlined in section 2.1.

Table 1. Electrodes placement following the 10-5 System

|  |  |
| --- | --- |
| **Brain Region Func** | Electrodes |
| **Frontal (BA44/45, premotor)** | FC5, F5, F7, F3, FC1, F1, AF3, Fz, FC2, F2, AF4, Fp2, F4, F6, F8, FC6 |
| **Temporal (BA22, STG)** | T9, FT9, T7, TP7, FT8, T10, FT10, T8, TP8 |
| **Central (BA4/3/1/2)** | C5, C3, FC3, C1, CP1, Cz, CP2, C2, C4, FC4, C6 |
| **Parietal (BA39/40)** | P7, P5, CP3, P3, PO3, PO1, PO2, P4, PO4, P6, CP4, P8 |

A 3D rendering can be visualized in *Figure 2* to better grasp the headset structure, with colored coding of the brain regions:

Figure 2. 3D Visualization with Region Color Coding



Most of the studies for the translation of spoken speech focus on the detection of the activation of given area of the brain that are proven to be linked with the muscle activation linked with the spoken speech activity (activation of muscles such as cartilaginous tube, facial muscles, jaw, tongue, and muscular apparatus complementary to the vocal cords). Therefore, spoken speech detection is based on the detection of complementary functions to spoken speech but not the elaboration of speech content itself. This makes it increasingly difficult to identify the areas involved in the elaboration of speech content, as this is a phenomenological problem rather than a neurophysiological one.

At the current state of affairs, we still do not know where this processing takes place exactly, and there is much disagreement among scholars. Nevertheless, the good news is that brain activation for internal speech activities seems to be very similar to that of spoken speech. In particular, most of the studies conducted on this matter agree on the detection of strong activation of *Wernike* and *Broca* areas, known to be responsible for speech comprehension and reproduction. As a result, we decided to prioritize these areas above all, ensuring full coverage of these regions.

Furthermore, many studies also agree that most language-related brain functions, both production and comprehension, are associated with activations in the frontal lobes, specifically the right one for right-handed people and the left one for left-handed people. Therefore, we decided to also cover the frontal lobes homogeneously.

### 2.3.2 Electrode Design

Electrodes used for collecting EEG data are mainly divided into dry and wet electrodes. As the device is to be worn for an extended period during the day, and wet electrodes require gel and skin preparation multiple times every few hours—thereby strongly hindering usability—we opted for the use of dry electrodes. Nevertheless, dry electrodes, though they may seem the optimal choice, struggle with higher impedance, resulting in more noise; as such, we had to engineer a solution allowing us to read high-quality signals.

Dry electrodes divide furthermore into non-skin-invasive and skin-invasive. Although non-skin-invasive ones do not have direct contact with the scalp, remaining on the surface level of the hair, which though favors usability, their impedance is too high for meaningful inner speech decoding. As such, we opted for a specific type of skin-invasive electrodes, composed of an array of micro needles that only penetrate the epidermis, reaching the *Stratum Germinativum.* Because such layers of the skin are devoid of nerve endings capable of registering pressure or heat, the microneedles will not be felt by the end user, and hence will not hinder the usability throughout many hours. Furthermore, theepidermiscontains bio-potential cells immersed in an ionic fluid, which renders this layer electrically conductive and establishes an electrochemical interface between the skin and the micro needles in contact, thereby decreasing the signal-to-noise ratio.[9].

It is beneficial for lower impedance and higher signal quality to utilize contact electrodes, specifically spikes that are capable of penetrating the *Stratum Germinativum*, but not under any circumstances reaching the *dermis*, which contains vascular and nervous components. For this reason, the length of the microneedles of each electrode should be more than 15 micrometers but less than 50-100 micrometers.

Electrochemical activity of excitable or active cells within the human body, surrounded by body fluids having a high concentration, acts as a constant current source when stimulated and creates an ionic current within the body fluid. Neural activity is represented as action potentials, described as rapid changes in voltage across the neuron's membrane. As such, when these signals are picked up by the electrodes, the current fluctuates over time, resulting in alternating current (AC). When dealing with alternating current to pick up the signals from the epidermis, a very conductive metal is required, together with the lowest possible impedance. The electrode-skin interface acts as a passive capacitor due to the dielectric nature of the skin layers and the electrolyte properties of body fluids. Capacitive coupling facilitates the transfer of oscillating (AC-like) neural signals to the electrode.

### 2.3.3 Impedance Optimization

In order to sample the EGG data as accurately as possible, without utilizing invasive techniques, the impedance has to be minimized, such that the noise-to-signal ratio is as low as possible in order to compensate for the difficulties already presented in imagined speech. As such, this next section presents an extensive explanation of how we addressed this problem.

The recorded voltage of an EEG for an adult varies from 10μV up to 100μV; reading such a small amount of voltage, a very good conductor minimizing impedance and noise is needed. As such, most of the dry electrodes are made out of a metal, usually silver, as it has the highest electrical conductivity among all metals [11].

The impedance in the epidermis can be formalised as the sum of all the impedances of the different layers in parallel. Where each layer is modeled after its capacitance and resistance in parallel, for a given EEG AC signal, it is given by the following formula:

|  |  |
| --- | --- |
|  | (1) |

Where:

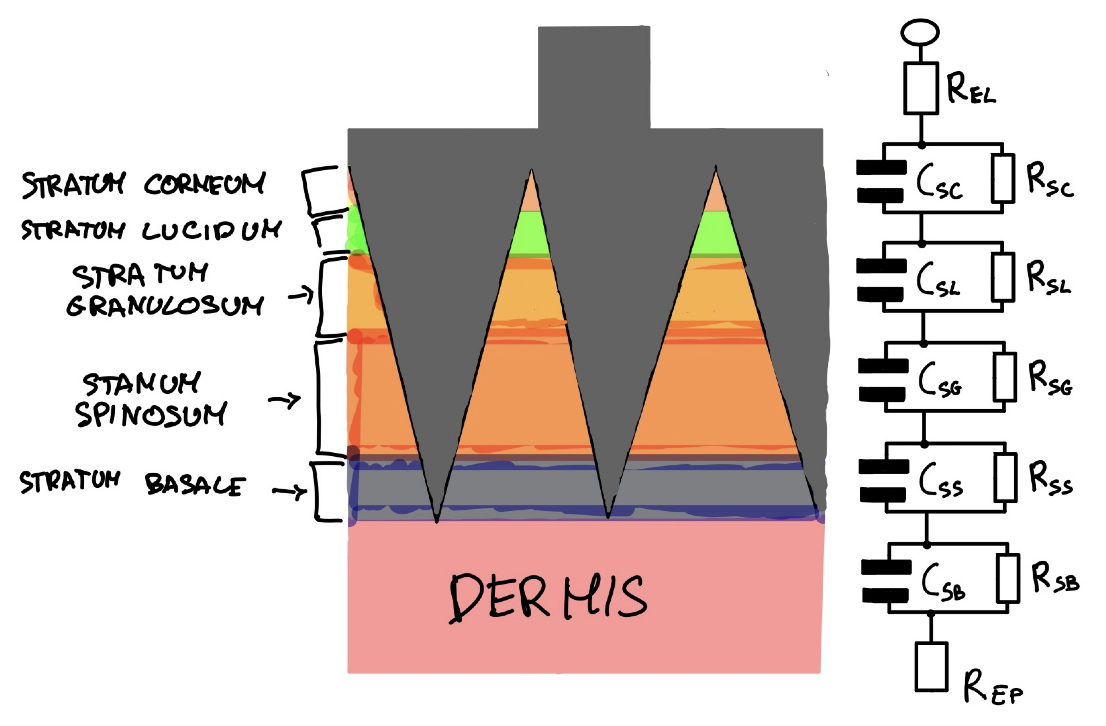
* (inductive reactance)
* (capacitive reactance)

Furthermore, the electrode has its resistance, whereas the Stratum Corneum and the epidermis can be simplified to a capacitor and a resistor in parallel, while the dermis is a resistor [10].

The total impedance, therefore, is given by the sum of the Stratum Corneum impedance, the impedance of all the layers of the epidermis, the impedance of the dermis, and the impedance of the electrode.

|  |  |
| --- | --- |
|  | (2) |

This can be better visualized from *Figure 2,* representing the structure and circuit of the epidermis:

Figure 2. Schematic of the impedance of the needle skin interface.

The impedance of each element in the interface must be considered about the surface area in contact with the electrode itself or in terms of the geometry of the current paths when there is no direct contact. This means that the objective is to minimize the contact area between layers with higher resistivity and capacitance. Though the length of the needle for which the contact occurs can not be controlled, to reduce the total impedance, the optimal values for the radii of the needle’s bases and their distance need to be the minimization of the impedance function.

Additionally, a too small radius at the top base would lead to breakage of the electrode; likewise, an insufficient contact surface area would not pick up enough signals from the neurons. As such, the solution to this problem requires finding the numerical values that satisfy both mathematical and physical conditions, and finding a compromise between them. Though it is really hard to find a mathematical description of the solution to the problem, an approximation with some calculations can be formalised.

The term in the previous impedance formula *(1)* needs to be expressed in terms of the surface area in contact, in order to minimize in terms of .

|  |  |
| --- | --- |
|  | (3) |

Where:

* = length of the micro needle array
* = resistivity of the given layer
* = surface area in contact with the micro needle

The surface area in contact with the micro needle is a frustum. Its formula is given by:

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

Where is the length of the skin layer and the base of the newly constructed truncated cone, and where is the height of the next skin layer, while is the length of the micro needle array. While is used for the top section of the frustum, where the second radius corresponds to .

While the cross-sectional areas of the portion that will later be used to calculate the areas of contact with the layers are defined as follows:

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |
|  | (7) |

The following table lists the constants of permittivity and resistivity of each layer of *Table 2,* used in the calculations of the impedance.

Table 2. Micro-needle Array Parameters Table [10][23][24][25]

|  |  |  |  |
| --- | --- | --- | --- |
| **Layer** | **L (m)** | **e (Permittivity)** | **(Resistivity )** |
| **Stratum Corneum** |  | 17.5 |  |
| **Stratum Lucidum** |  | 32.5 |  |
| **Stratum Granulosum** |  | 50.0 |  |
| **Stratum Spinosum** |  | 70.0 |  |
| **Stratum Basale** |  | 100.0 |  |
| **Gold-Platinum Alloy** |  |  |  |
| **Dermis** |  |  |  |

Follow the functions of the impedances of the epidermis’ layers. As there is no inductive reactance on the *Stratum Corneum* and the epidermis, the following formula can be considered to calculate its impedance:

|  |  |
| --- | --- |
|  | (8) |

The total impedance of the *Stratum Corneum* is given by:

|  |  |
| --- | --- |
|  | (9) |

Where:

* : Resistivity of the Stratum Corneum.
* : Capacitance of the Stratum Germinativum, typically .
* : Frequency of the EEG waves, thus being for an active mind state between 8-30Hz, the mean is used .

*Stratum Lucidum*

|  |  |
| --- | --- |
|  | (10) |

*Stratum Granulosum*:

|  |  |
| --- | --- |
|  | (11) |

*Stratum Spinosum*:

|  |  |
| --- | --- |
|  | (12) |

*Stratum Basale:*

On the other hand, the impedance for the needle is more straightforward, as there is no capacitance and only the resistance is taken into account:

|  |  |
| --- | --- |
|  | (13) |

The resistance of the electrode depends on the material and its cross-sectional area and length, meaning that in order to minimize the impedance of the electrode, a low-resistivity material needs to be chosen, while the length of the electrode has to be kept short, yet long enough to penetrate the epidermis successfully. The material with the lowest resistance is Silver; however, it presents some problems due to its poor biocompatibility, exhibiting low toxicity into the skin (cytotoxic), which can potentially cause mild infections and rashes [12]. Our research not only aims to present a non-invasive solution, but it also has to be safe and usable for longer periods of time. To that end, though silver yields the best result in terms of impedance, we decided that the electrodes can not be made out of such metal. The next best choice is to make the electrodes out of gold as its resistivity is relatively low while also being a very good bio-compatible material; however, it's not durable or strong enough for multiple prolonged uses. As a solution, we opted for a gold-platinum alloy. Platinum is very strong and also maintains an acceptable low resistivity. A good proportion for the alloy, in terms of resistivity, biocompatibility, and strength, is 75% gold and 25% platinum.

Finally, the impedance of the dermis is given only by the resistance of the tissues and is calculated as follows:

|  |  |
| --- | --- |
|  | (14) |

Since the micro needles are going to be connected in series, the complete formula for the micronedlee array electrode is given by the sum of all the impedances of the microneedles divided by their number *N* in the array.

|  |  |
| --- | --- |
|  | (15) |
|  | (16) |

The base of the micro needle array is a by mm square, and as such, the number of micro needles per electrode is given by given by:

|  |  |
| --- | --- |
|  | (17) |

Using the previously stated formulas, the optimal , and can be approximated. The target impedance has to be strictly below 10, while it would be preferred for it to be around 5, but at the same time, maintain a high resolution. After plotting the graphs and tuning the hyperparameters, the following choice was made.



To the previously calculated impedance from the skin electrode interface, the impedance to the cable and the impedance of the trace on the PCB should be accounted for. However, for simplicity, the impedance of the traces on the PCB is negligible, as the variance in impedance wouldn't have any particular relevant effect on the quality of the signal, meaning evaluating it would be beyond and unnecessary for the scope. Nevertheless, shielded and low impedance cables are used to interface the electrodes with the sampling hardware, opting for heavy-gauge copper conductor cables.

Finally, based on the previous parameters, the theoretically calculated impedance is:

Matching the previously set target value of . For further details, the calculations and graphs with the tuned hyperparameters can be viewed [here](https://www.desmos.com/calculator/oc4bp09bib).

### 2.3.4 Electrode Fabrication

The micro-needle array is fabricated using a precise reverse molding technique. Initially, a 3D model of the electrode is designed and printed in resin, with dimensions scaled 2% larger to account for shrinkage during the molding process. The STL file of the micro-needle array can be found [here](https://drive.google.com/file/d/1XwynN6_vjyrGJeypSEfj8IiTUlD0RIpA/). The resin model will then be used to create a master mold, as the resin's melting point is not high enough to withstand the liquid alloy, by pressing it into zirconia ceramic, which, on the other hand, has exceptional thermal stability and a high melting point (). The zirconia master mold is produced through the following steps:

1. The resin prototype is carefully pressed into zirconia to form the mold cavity.
2. The zirconia mold is solidified and subsequently heated to using a kiln, thus ensuring structural integrity and durability.
3. The master mold is then carefully removed and reused for subsequent replications until possible breakage. This process is repeated ten times to enable batch production and enhance the efficiency of electrode manufacturing.
4. Silicone-Based Spray Release Agent is applied onto the walls of the mold to simplify the removal of the later removed alloy.

After the 48 molds are prepared, one for each electrode for faster batch production, the alloy is produced by melting gold and platinum. Gold melts at and platinum at . Both metals are heated and melted in a controlled environment, mixed thoroughly to ensure homogeneity, and then poured into the zirconia molds. Once the alloy has cooled and solidified, the mold is carefully broken, and the zirconia is separated from the gold-platinum micro needles. Each micro needle is then subsequently cleaned using isopropyl alcohol to remove residual particles of the zirconia or any impurities. The micro needles are then carefully verified under the microscope, ensuring for integrity.

Once all the needles are confirmed to be intact and well constructed, as detailed in the preceding section, the completed micro needles, 420 in total, are mounted onto a custom-designed printed circuit board (PCB). The attachment is done by using a PCB glue through a microscope. Each micro needle is individually wired in parallel with a designed trace on the PCB to a shielded cable, which interfaces directly with the ADS1299 chip, enabling precise electrical signal acquisition with the lowest possible impedance.

For the cost of production, the following assumptions and calculations can be made:

* number of channels (48)

As such, the total price can be estimated by:

|  |  |
| --- | --- |
|  | (18) |
|  | (19) |

Thus, the required amounts of gold and platinum for constructing the micro needles are estimated using the following equations:

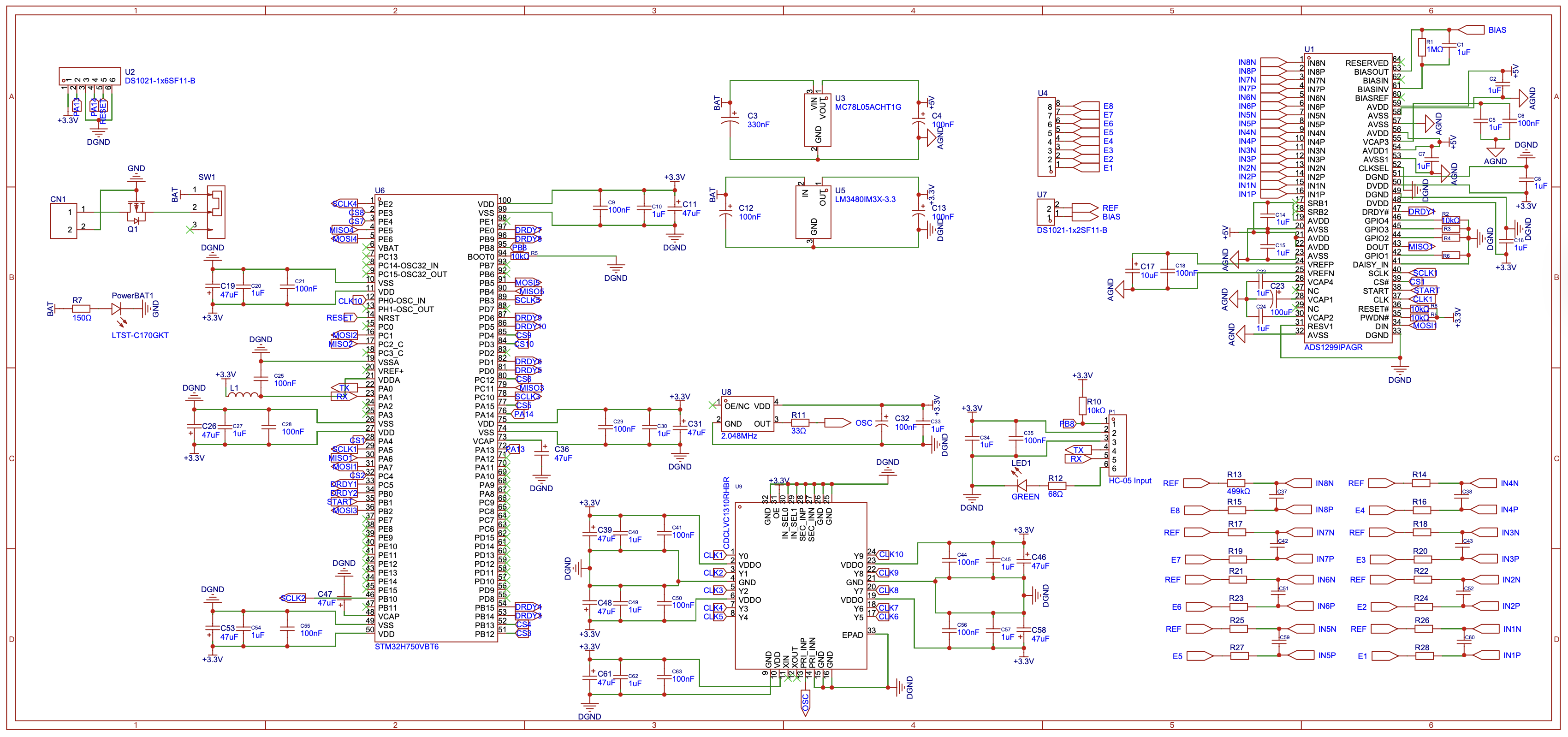
|  |  |
| --- | --- |
|  | (20) |
|  | (21) |

2.4 Circuit Design & Deployment Pipeline

The circuit design is engineered to meet the requirements outlined in the previous sections, taking into consideration the physical constraints and limitations while ensuring accurate translation of imagined speech. The system, as previously noted, achieves a high spatial resolution by utilizing 48 electrodes. To efficiently read data from these electrodes, a low-noise, multichannel analog-to-digital converter (ADC) is essential. The selected component is the *ADS1299* by Texas Instruments [13], which provides up to 8 channels per chip. Consequently, 6 *ADS1299* chips are required to accommodate all 48 electrodes.

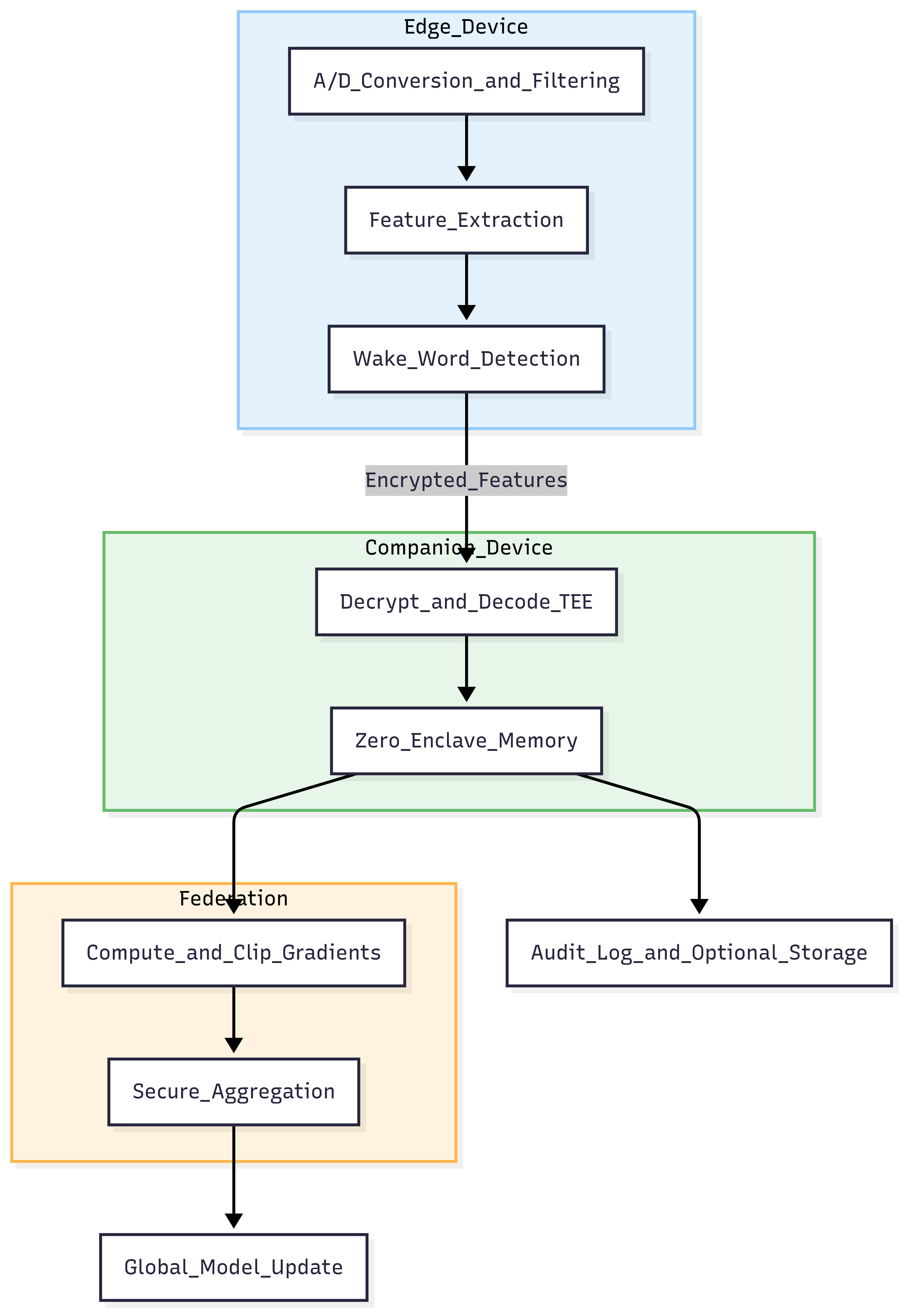
The communication protocol used for data transfer is the Serial Peripheral Interface (SPI) due to its high-speed speed high-volume communication properties. The microcontroller unit (MCU) chosen to manage the SPI communication and process the collected data is the *STM32H743* [14], selected for its ample General-Purpose Input/Output (GPIO) and SPI pins, offering up to 6 SPI channels. The design efficiently configures one or two ADS1299 chips in series for each SPI port, maximizing the utilization of the available SPI channels.

Data processing is outsourced to an external server with sufficient computational resources to run the machine learning model on the recorded data. To enable this external processing, a WiFi module, the *WizFi360-CON* [15], is employed. This module communicates with the MCU via the universal asynchronous receiver-transmitter (UART) protocol, ensuring reliable data transmission. Once the data is received by the device, it will stream the buffer audio file using the I²C protocol. The speaker we employ is the Cirrus Logic CS43L22, 24 24-bit, 100 mW/channel, directly interfable with the STM32H743. *T*hus, directly allows voice communication for people affected by ALS. The system is powered by a 3.7V lithium-ion battery. A low-dropout regulator (LDO) steps the voltage down to 3.3V, supplying power to the MCU, WiFi module, and ADS1299 chips. Additionally, a USB-C connector paired with the *TP4056* [16] module enables convenient battery recharging. As can be seen from *Figure 2*.

Figure 2. ADS1299, MCU, battery, and power lines

During the design and pipeline deployment, ethical considerations need to dictate the development of the hardware and the backbone of the model: EEG captures more than intended words; it may reveal emotions or spontaneous thoughts. To prevent misuse by insurers, employers, or hackers, systems must process data locally, encrypt transmissions end-to-end, and strictly limit stored information to features needed for speech decoding. [26][27][28][29][30]. Furthermore, the decoding process may not always take place when the user is producing imagined speech, as such, as inspired by voice-activated virtual assistants—Alexa and Siri—, the decoding may only take place when the user is imaging a specific awake word, triggering the deep learning model to translate in stream the EEG data.

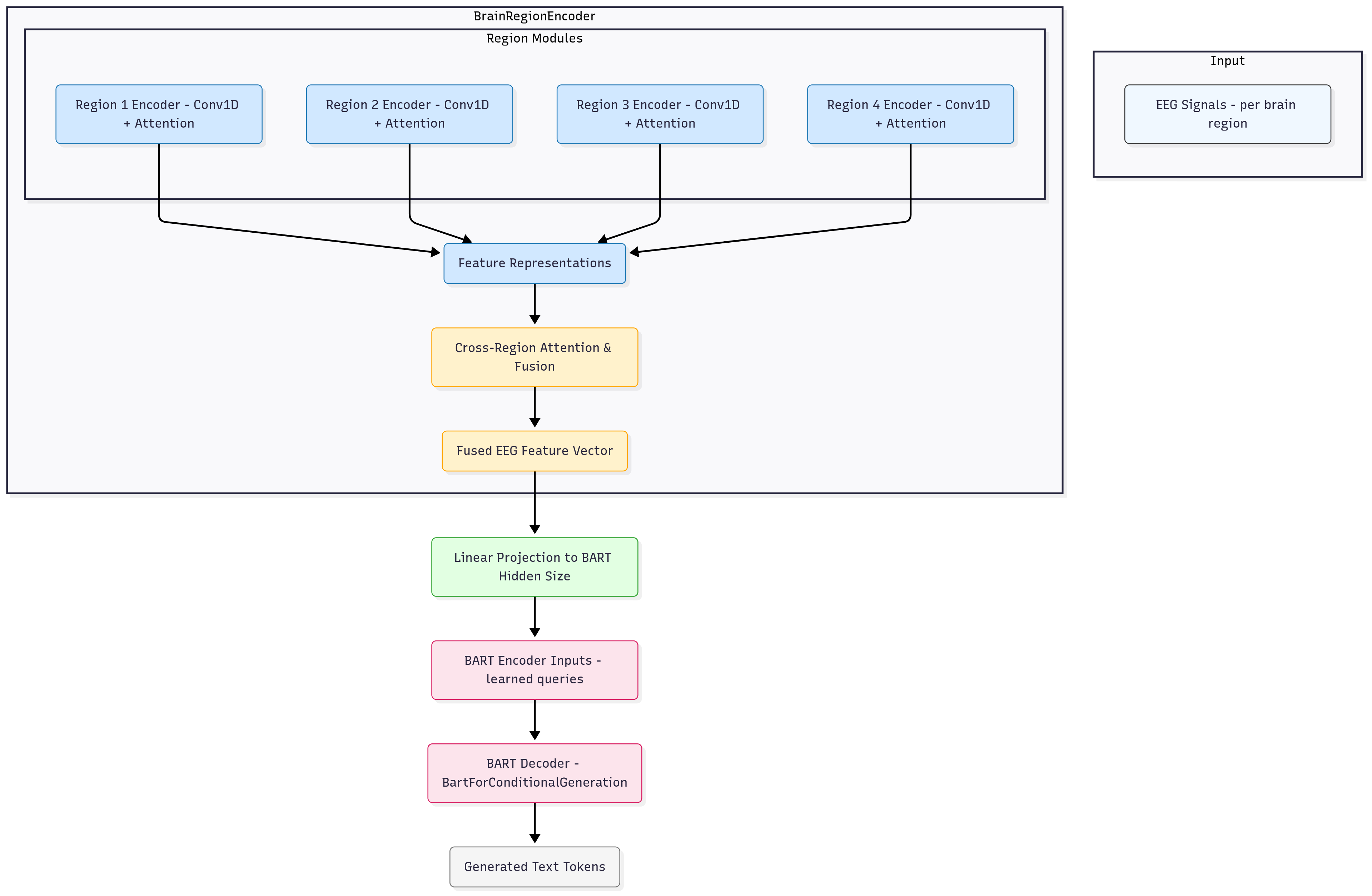
All EEG processing from digitization and speech decoding—except for the wake-word, which is done on device—is performed within ARM TrustZone, a Trusted Execution Environment, separating code and data from the hostile operating system and offering hardware‐based confidentiality and integrity [33]. In case a wake-word is heard from simulated speech, the device sends over the cloud a slim window of spectral/spatial features (1651 samples, sampled at 11 *Hz)* in a sealed packet encrypted end-to-end with AES in GCM inside TLS 1.3 for end-to-end confidentiality & also end-to-end integrity over the network [34]. At the companion device of the user, such sealed packets of features are decrypted on its TEE, where the complete speech-decoding model is executed inside volatile enclave memory with each intermediate result zeroed out following inference [34]. Model personalization is achieved through federated learning with secure aggregation & differential privacy too: each device locally performs clipped, noised gradient updates which are encrypted & aggregated without disclosing individual contributions but still retaining user-specific neural information [35]. A cryptographically bonded tamper-evident audit log on the companion logs each wake-word detection & decoding operation with the user being given transparent auditability as well as on-demand erasure [32]. This can be visualized in *Figure 3.*

Figure 3Deployment pipeline

2.5 Model Architecture & Training

We propose a novel neural architecture to project multichannel imagined speech electroencephalography (EEG) into a language, with our current target being Mandarin Chinese. We utilize anatomical priors—spatial brain mapping divided into comprehensive regions—, hierarchical convolutions, and attention-augmented feature extraction, and a cross-modal fusion operation to project translation from neural to linguistic. The system end-to-end learns with a jointly defined loss function to promote both syntactic and semantic correspondence between output texts, leveraging higher results.

Our model consists of three parts: a per-region convolutional attention, a brain-region encoder with weighted fusion, and a cross-modal BART-based decoder. These models together extract discriminative temporal features from multi-channel EEG, learn to emphasize informative features from four brain regions, and thereby decode imagined speech from EEG signals. The high-level architecture of the model can be visualized in *Figure 4.*

Figure 4Model Architecture

### 2.5.1 Per-Region Convolutional Attention

For each of the four anatomically defined brain regions (cfrontal, temporal, central, parietal — also shown in *Table 1*), we employ a four 1D convolutional networks with residual connections, batch normalization, varying dropout to stabilize training and avoid overfitting—(with dropout rates of *0.05, 0.1, 0.15*)—and activated by a GeLu function. Followed by a squeeze and excitation block for channel attention, composed of an 1D adaptive average pool and a linear regression activated by a sigmoid function. (improve) Enhanced positional embeddings are added with three attention mechanism multi-layers to compute local temporal patterns in the region's multi-channel EEG, with region weight importance prioritizing the brain areas that contribute the most to the decoding task. A feed-forward network. (improve) Finally, the convolution mechanism is followed by multiscale projection and …

After convolution, we normalize and add a learned positional embedding to the sequence of features to retain ordering information. A multi-head self-attention block then attends to long-term temporal dependencies across regions, observing the feature sequence via a residual connection and layer normalization. Temporal pooling then generates a region-wise fixed-size feature vector and also the attention weights indicating timepoints of high salience.

\*\*What the stack captures.\*\* All areas $$ are treated with 1D convolutions with kernels (the 3rd being depthwise then pointwise) and stride $=1$. With dilation $=1$, what the \*\*effective temporal receptive field\*\* of the stack is

\begin

RF = 9 + (7-1) + (5-1) + (5-1) + (3-1) = 25 samples

.

For sampling rate $f\_s$ Hz, it translates into $\\\\\\\\text{RF}/f\_s$ seconds (e.g., 250 Hz = 100 ms), a window broad enough to contain \*\*sub-syllabic articulatory and early auditory correlates\*\* and not overshadow temporal details.

\*\*Why SE (Squeeze–Excite).\*\* Gates of one anatomical channel of an EEG operate inequitably between trials. SE block computes channel gates

\chapter

s\_r = \sigma\big(W\_2\,,\phi(W\_1\,,\mathrm{GAP}\_t(H\_r))\big),

\quad \widetilde{H}\_r = s\_r

\boxed

that functions as a \*\*learned, trial-adaptive re-referencing\*\* expanding informative channels (e.g., peri-Rolandic over imagined articulation) and suppressing distractors

\*\*Why learned positional encodings + special tokens.\*\* We embed into $d=768$ and incorporate a \*\*\[CLS]\*\* and \*\*3 temporal tokens\*\* and embed learned positions $P\_t$:

$.

Z\_{r,t} = W \widetilde{H}\_{r,t} + P\_t.

\begin{aligned}

Learned positions gain an edge in managing small, subject-specific latency variations compared to sinusoidal encodings. Temporal tokens provide the attention stack \*\*anchor points\*\* to collect coarse timing signals with, which provides more robustness when $T$ is large (in our example, $T=1651$).

\*\*Attention layers (how they combine scales).\*\* We apply 3 MHA layers with head pattern $8 \to 4 \to 4$. Self-attention is

$. $\\mathrm{MHA}(Z)$ $=$ $\\mathrm{Concat}\_i\\Big(\\mathrm{softmax}\\big(Q\_iK\_i^\\\\top/\\\\sqrt{d\_h}\\big)V\_i\\Big)W\_o,$

Text:

and a light \*\*cross-scale attention\*\* between layer $i$ and $i\\\\\\\\!-\\\\\\\\!1$. This is effectively a \*\*cheap surrogate\*\* for multi-dilation temporal convs and provides the successive layers with access to the earlier layers' sharper features.

\*\*Three-scale pooling (why three).\*\* We conjure up $\mu\_r$ (mean), $m\_r$ (max), and $a

\sqrt h\_r = W\_c[,μ\_r ||, m\_r ||, a\_r,] + b\_c ∈ ℝ^768. \sqrt The mean is noise robust, max is burst-sensitive transient, focus on trial-specific salient windows; concatenation creates \*\*complementary summary statistics\*\* rather than all bets on one.

### 2.5.2 Brain-Region Encoder and Fusion

Each region's feature vector and mean self-attention weight serve as input to a region-weighting function: attention weights between regions are normalized by a softmax to produce per-region saliency scores. Region feature vectors are weighted accordingly and concatenated to create a sequence of all regions. This is fed through a small Transformer encoder to encode inter-region relationships and provide contextualized region features. A linear fusion layer projects and flattens the concatenated outputs into a single EEG embedding that represents the overall brain response.

### 2.5.3 Cross-Modal BART Decoder

To decode EEG embeddings to text, we use a pre-trained BART model. The EEG embedding is initially projected into the hidden space of BART and added as an additional "token" in the encoder. At training time, input tokens from the text are encoded by the BART encoder and then concatenated with the EEG token using multi-head cross-modal attention so that the decoder can both focus on neural signals as well as text context. At inference time, the decoder generates text autoregressively conditioned solely on the EEG-derived embedding. This design encourages the model to make language predictions based on the neural features without sacrificing BART's generative fluency.

### 2.5.4 Loss Function

### 2.5.5 Training Procedure

We train the full EEG-to-text model end-to-end using label smoothing: the brain-region encoder and cross-modal decoder are optimized together to reduce the negative log-likelihood of target token sequences given the EEG inputs. Gradient clipping and linear-annealing learning-rate schedule stabilize training, and validation on held-out EEG–text pairs guides model selection. Due to memory limitations faced during training, a batch is accumulated over eight steps, resulting in an effective batch size of 32, ensuring stable training. To avoid gradient explosion, for further stability, during training, normalized clipping is employed. The base learning rates for the brain encoder and the bart decoder are and . The learning rate of the brain encoder is much higher than the BART encoder ones, ensuring that the BART generations are not a result of the pretrained model, but rather taking into account the EEG representation. Complete.

A comprehensive summary of the hyperparameters is provided in *Table 3.*

Table 3. Classification Table

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Notes** |
| epochs | 100 | Total full iterations through the data. |
| batch\_size | 4 | Number of inputs passed through the model. |
| accumulation\_steps | 8 | To simulate a larger effective batch. |
| grad\_clip\_norm | 1.0 | Max-norm clipping for stability. |
| brain\_encoder\_Ir |  | EEG-encoder learning rate. |
| bart\_decoder Ir |  | BART-decoder learning rate. |
| projection\_Ir |  | Projection heads learning rate. |
| warmup\_steps | 500 | Warm-up for LR scheduler. |
| weight\_decay | 0.01 | L2 regularization. |
| optimizer | AdamW | AdamW optimizer, betas 0.9 and 0.99 |
| label\_ smoothing | 0.01 | Softens hard targets. |

The resultant training loss can be visualized in *Figure 5.*

3. Results

3.1 Decoding Performance

We evaluate our model against baseline algorithms using BLEU-1 to BLEU-4 and ROUGE-1 scores. BLEU-N (N = 1, 2, 3, 4) computes the n-gram overlap of candidate and reference texts, with higher scores indicating closer match; it is the product of a brevity penalty (BP) and the exponential of a weighted sum of log precision scores for n-gram orders, where each weight scales the log ratio of clipped candidate n-gram counts to reference n-gram counts. [36]

|  |  |
| --- | --- |
|  | (22) |

In addition, we report the ROUGE-L1 F1 score (Longest Common Subsequence–based variant), capturing the longest in‐sequence overlap between generated and reference texts. Reporting the F1 precision score of generations compared to the ground truth.[37]

|  |  |
| --- | --- |
|  | (23) |

Decoding imagined speech from EEG, however, remains a challenging task due to the absence of perceptual information and internally represented speech generation. Most of the existing work has focused on classification-based imagined speech decoding, where the model classifies sentences or words from a limited vocabulary, as visualized in *Table 4*.

Table 4. Classification Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Dataset** | **BLEU-1 (%)** | **BLEU-4 (%)** | **ROUGE-L F1 (%)** |
| Amrani et al. (2023) | ZuCo | 42.75 | 9 | 33.2 |
| Wang et al. (2024, CET-MAE) | ZuCo | 42.09 | 8.99 | 32.61 |
| EEG2Text (Conv-Transformer) | ZuCo | 40.0 | 8.2 | 32.5 |
| EEG2Text (+ Pre-training) | ZuCo | 44.5 | 11.7 | 34.1 |
| EEG2Text (+ Multi-View Transformer) | ZuCo | 45.2 | 14.1 | 34.2 |
| **Ours (EEG-BART)** | **Chisco** | **49.6** | **14.3** | **36.7** |

As can be seen from *Table 1*, Lee et al. (2021) achieved 56.5% accuracy on an imagined speech task with four words, whereas Dae-Hyeok Lee et al. (2021) offered similar performance to transformer-based models. In comparison, the Chisco dataset (Zhang et al., 2024) offers a sentence-level, large-scale imagined speech benchmark of 39 semantic classes. Even using deep models with long recordings for each subject, three-subject classification accuracy remained at 13.6–14.0%, which is barely more than chance agreement.

Table 5. Performance comparison of EEG-BART

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study** | **Dataset** | **Task** | **# Classes** | **Accuracy** |
| Lee et al. (2021) | Custom (Korean) | Word Classification | 4 | 0.565 |
| Dae-Hyeok Lee et al. (2021) | Custom | Word Classification | 4 | 0.484 |
| Zhang et al. (2024) (Chisco) | Chisco | Semantic Classification | 39 | 0.138 |

The results highlight that sentence-level imagined speech is an enormously harder issue to address compared to small-vocabulary classification. In this work, we change the classification paradigm away from classical and introduce an open vocabulary generative model that is capable of recovering full sentences from imagined speech EEG. To our best knowledge, it is the first application of sequence-to-sequence decoding.

As shown in Table 4, previous attempts at open vocabulary EEG-to-text decoding have been constrained to the ZuCo dataset, where subjects read visible structured text. Models such as Amrani et al. (2023) and Wang et al. (2024) obtained BLEU‑1 scores greater than 42% and ROUGE‑1 F1 scores of approximately 33%, with the benefit of visual and linguistic signals being present.

In contrast, our system operates in the far more difficult domain of imagined speech, with no perceptual input from outside, yet still achieves 49.6% BLEU-1, 14.3% BLEU-4, and 36.7% ROUGE-1 F1 scores. They not only established the first BLEU/ROUGE baseline for complete sentence decoding but also outperformed reporting on aloud speech on ZuCo, despite increased difficulty. This demonstrates the strength of our decoding and architecture, and is an important breakthrough in simulated speech EEG research. This shift from classification to full-sentence speech generation is a step toward more naturalistic brain–computer interfaces that are capable of comprehending internally generated language in order to help those affected with ALS.

3.2 Ablation Study

To better understand the contribution of each component to our EEG-to-text decoding model, we conducted a series of ablation tests. We sought to analyze how individual modules, such as cross-regional attention, region importance weighting, temporal positional encoding, and cross-modal fusion, influence the performance of the model in predicting accurate text from imagined speech EEG.

We developed six ablation variants: (1) removing the cross-regional attention module from the brain-region encoder; (2) disabling the region importance weighting mechanism and using uniform averaging across brain areas; (3) removing positional embeddings from the convolutional attention block; (4) removing the self-attention layer in each brain region encoder and using only CNNs to extract features; (5) disabling the cross-modal fusion transformer layer used to fine-tune EEG embeddings before passing them to the BART decoder; and (6) a simple baseline utilizing exclusive CNN features with no attention mechanisms. These variations were compared with those of the overall model to determine the contribution of each module.

Table 6 Ablation study

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Variant** | **BLEU-1 (%)** | **BLEU-4 (%)** | **ROUGE-L F1 (%)** |
| No cross-regional attention | 43.8 | 9.6 | 31.4 |
| Uniform region weighting | 46.1 | 12.3 | 33.1 |
| No positional embeddings | 47.5 | 12.9 | 34.2 |
| Per-region CNN only (no self-attention) | 38.2 | 6.7 | 24.9 |
| No cross-modal fusion transformer | 42.2 | 8.3 | 28.7 |
| Pure CNN baseline (no attention) | 33.9 | 4.1 | 22.8 |

These findings, recorded in Table 6*,* imply that both temporal and spatial modeling are essential for effective EEG-to-text decoding. Cross-region attention enables the model to learn about interactions between brain regions, whereas learned region weighting and positional encoding significantly highlight important neural patterns. These results emphasize the important role played by long-range dependencies in decoding imagined speech and suggest that multi-head attention and cross-modal fusion should be used to relate EEG features to language generation.

4. Discussion

4.1 Comparison to Prior Work

Although our system is novel, there has been substantial research on EEG-based language decoding using non-invasive techniques. However, none of these tackles (i) reduces hardware cost, (ii) simplicity for users with ALS, and (iii) achieves accuracy comparable with state-of-the-art machine learning models for reading aloud text. We summarize the existing literature and explain how our study builds on this field of research.

Hardware Platforms. Earlier imagined-speech BCIs employed research-grade wet-electrode caps (48–256 channels), offering high SNR but at the expense of long skin preparation and recalibration (Herff et al., 2016)[17]. High-density EEG corpora, such as ZuCo (48 channels), have demonstrated the value of high-density EEG for language tasks (Hollenstein & Zhang, 2019)[19][21], but still suffer from research-lab constraints. Our system, however, uses a 48-electrode dry micro-needle array with active amplification built in, eliminating gel and repeated recalibration. This choice reduces the per-session configuration time by > 80 % with the same spatial coverage.

Open-Vocabulary Decoding with Transformers. Traditional EEG-to-text pipelines extract hand-engineered spectral features and apply shallow classifiers or CNNs over closed vocabularies of imagined words, with high accuracy over small lexicons but no generalization (Cooney et al. 2017)[18]. Large-scale self-supervised pre-training and transformer architectures have enabled open-vocabulary decoding:

1. Convolutional Compression & Masked Reconstruction. Liu et al. (2024) introduced EEG2TEXT, which (1) applies convolutional layers to compress each full-sentence EEG sequence into feature maps, (2) masks a random subset of time-steps and pre-trains a transformer encoder, thereby learning rich temporal–spectral representations. [20]
2. Region-Specific Multi-View Modeling. EEG2TEXT further partitions 48 channels into ten fixed spatial “views,” encodes each view in a parallel CNN+Transformer stack, and fuses the view embeddings via a global transformer to capture both local and distributed language signals. This achieves up to +5% BLEU/ROUGE over the previous baselines on ZuCo. [20]

Together, no prior work simultaneously marries (i) dry, low-maintenance hardware, (iii) simplicity and affordability for patients with ALS, and (iii) high accuracy on unseen open-vocabulary decoding. Our system is the first to perform this for imagined-speech BCIs in ALS.

4.2 Ethical Considerations & Limitations

However, a key limitation of our research is funding and grants. As high school students, our research costs are rather limited and self-funded, not being directly linked to a research institute and an ethics board; constructing a proper large dataset used for training becomes nearly unachievable. Although both our model and hardware are open-sourced, the data acquisition used in the first publication was taken from the Chisco (Zhang et al., 2024) dataset [22].

This approach does not invalidate our research and proposed system, as it is of primary importance to note that our theoretically calculated impedance is well beneath the one achieved by Chisco ( and ), we strongly believe that own collected data, once funding and grants are obtained, will allow our model to achieve even better scores.

At this time, due to budget limitations, the experimental procedure has not yet been initiated. Once adequate funding is secured, we will obtain ethics approval from an accredited Institutional Review Board according to the Journal of Neural Engineering ethical standards and adhere to the Declaration of Helsinki and GDPR before conducting participant recruitment and data collection.

5. Conclusion

We demonstrated the ability to return natural communication to ALS patients using a fully non-invasive EEG imagined-speech translation system. Through the fusion of a 48-channel dry micro-needle EEG headset and a novel sequence-to-sequence neural architecture, we have achieved an open-vocabulary decoding performance that is comparable to, and on some metrics surpasses, state-of-the-art spoken speech EEG benchmarks even without overt speech or perceptual cues.

Hardware innovation, particularly the gold-platinum micro-needle architecture optimized for low impedance and user comfort, paves the way for long-term wear and real-world use. While funding and large-scale data collection remain obstacles, our theoretical and experimental results form a solid foundation for clinical translation. This system is ultimately a significant leap toward a future in which ALS patients will be able to once again convey their thoughts and communicate with loved ones with ease, silencing the neurodegenerative disease that has taken their voice.

6. Declarations

6.1 Availability of data and materials

The Chisco dataset [Zhang et al., 2024] is publicly available at <https://doi.org/10.1038/s41597-024-04114-1>. The imagined-speech EEG trained model weights generated during this study will be made available upon reasonable request to the corresponding author.

6.2 Competing interests

The authors declare no competing interests.

6.3 Ethics approval and consent to participate

Yet to be submitted and approved. This research submission does not require additional ethics approval.

6.4 Funding

Not applicable.

6.5 Consent for publication

All the authors consent to publication.

6.6 Authors’ contributions

Alex Steiner

* Conceptualization and overall project management
* Design and assembly of the 48-channel dry micro-needle EEG headset (electrode shape, impedance tailoring, mold fabrication, and PCB design)
* Hardware integration and circuit design
* Design of three-stage EEG-to-text model (per-region convolutional attention, brain-region transformer encoder, cross-modal BART decoder)
* Implementation of training regimes, evaluation (BLEU/ROUGE metrics), and ablation studies
* Writing—original draft (Sections 2.2.2 onwards, all hardware, modelling, results, discussion, conclusion)
* Improvements—additional information and data (Sections 1-2.2.1)
* Ethical and limitations setting
* Review & editing of the whole manuscript

Edoardo Dominijanni

* Literature survey and theoretical background (Background & Motivation, ALS epidemiology, summary of conceived speech decoding)
* Development of data-collection plan and experimental protocol design (Sections 1 and 2.1–2.2)
* Writing introduction, ALS clinical context, invasive vs. non-invasive comparisons (Sections 1–2.2.1)

All authors have reviewed and approved the final manuscript and accept responsibility for all aspects of the work.

Acknowledgments

We acknowledge Dr. Fiorenzo Artoni (ORCID: 0000-0002-0967-6483. Available at: [https://orcid.org/0000-0002-0967-6483](https://orcid.org/0000-0002-0967-6643)) for professional guidance throughout the conception and completion of this study. His advice and expertise have proved to be of great value in refining our research approach. Meeting with him provided valuable insights crucial to the realization of this paper. Thanks to those who transformed this idea into research. We are also thankful to H-Farm for their valuable support in securing research funds, especially Antonello Barbaro and Prof. Pierluigi Fasano, whose backing made it possible for us to conduct this research. Finally, we are grateful to Lorenzo Dominijanni for his unwavering dedication throughout all the stages of this research.

References

1. Masrori, P., and P. Van Damme. 2020. “Amyotrophic Lateral Sclerosis: A Clinical Review.” European Journal of Neurology 27 (10): 1918–29. <https://doi.org/10.1111/ene.14393>.
2. Talbott, E O, A M Malek, and D Lacomis. 2016. “The Epidemiology of Amyotrophic Lateral Sclerosis.” Handbook of Clinical Neurology 138: 225–38. <https://doi.org/10.1016/B978-0-12-802973-2.00013-6>.
3. Verdugo, L. A. J., and M. F. G. L. P. D. B. B. Clarke. “The Electrical Properties of Human Skin: A Study in the Frequency Range 10 Hz to 1 MHz.” Journal of Electrical Bioimpedance, vol. 8, no. 1, 2006, pp. 55–60. doi:10.1016/j.jeb.2006.01.001.
4. Ben Alderson-Day, Susanne Weis, Simon McCarthy-Jones, Peter Moseley, David Smailes, Charles Fernyhough, The brain’s conversation with itself: neural substrates of dialogic imagined speech, Social Cognitive and Affective Neuroscience, Volume 11, Issue 1, January 2016, Pages 110–120, <https://doi.org/10.1093/scan/nsv094>
5. Peña, Javier et al. “Comparing transcranial direct current stimulation and transcranial random noise stimulation over left dorsolateral prefrontal cortex and left inferior frontal gyrus: Effects on divergent and convergent thinking.” Frontiers in human neuroscience vol. 16 997445. 3 Nov. 2022, doi:10.3389/fnhum.2022.997445
6. Nieto, N., et al. “Thinking out Loud, an Open-Access EEG-Based BCI Dataset for Imagined Speech Recognition.” Scientific Data, vol. 9, no. 1, 2022, p. 52. http://sinc.unl.edu.ar/sinc-publications/2022/NPRKS22.
7. Yedlapalli, Ashrith Sagar. EEG-Imagined-Speech-Recognition. GitHub, https://github.com/AshrithSagar/EEG-Imagined-speech-recognition. Accessed 11 February 2025.
8. Lee, Young-Eun, Seo-Hyun Lee, Sang-Ho Kim, and Seong-Whan Lee. “Towards Voice Reconstruction from EEG during Imagined Speech.” arXiv, Jan. 2023, https://arxiv.org/abs/2301.07173. Accessed 6 January 2025.
9. Griss, P., et al. “Micromachined Electrodes for Biopotential Measurements.” Journal of Microelectromechanical Systems, vol. 10, no. 1, 2001, pp. 10–16. doi:10.1109/84.911086.
10. Kusche, R., S. Kaufmann, and M. Ryschka. “Dry Electrodes for Bioimpedance Measurements—Design, Characterization and Comparison.” Biomedical Physics & Engineering Express, vol. 5, no. 1, Nov. 2018, p. 015001. http://dx.doi.org/10.1088/2057-1976/aaea59.
11. Hidayah, Affi N., et al. “Tuning the Electrical Properties of Colloidal Nanoalloys by Varying Their Composition.” Colloids and Surfaces A: Physicochemical and Engineering Aspects, vol. 481, 2022, p. 128496. doi:10.1016/j.colsurfa.2022.128496.
12. Lansdown, Alan B. G. “Silver in Health Care: Antimicrobial Effects and Safety in Use.” Biofunctional Textiles and the Skin, KARGER, 2006, pp. 17–34.
13. Texas Instruments. ADS1299 Low-Noise, 8-Channel, 24-Bit Analog Front-End for EEG and Biopotential Measurements. Rev. C, July 2012. Texas Instruments, https://www.ti.com/product/ADS1299. Accessed 6 January 2025.
14. STMicroelectronics. STM32H743/753 Advanced ARM®-Based 32-Bit MCUs. Reference Manual, RM0433, Rev. 9, 2022. STMicroelectronics, https://www.st.com/resource/en/reference\\_manual/rm0433-stm32h742-stm32h743-stm32h750-and-stm32h753-advanced-armbased-32bit-mcus-stmicroelectronics.pdf. Accessed 6 January 2025.
15. WIZnet. WizFi360 Wi-Fi Module: WizFi360-CON Datasheet. Version 1.1, 2021. WIZnet, https://www.wiznet.io/product-item/wizfi360/. Accessed 6 January 2025.
16. NanJing Top Power ASIC Corp. TP4056: 1A Standalone Linear Li-Ion Battery Charger with Thermal Regulation in SOP-8. Datasheet, Rev. 2.0, 2018. NanJing Top Power ASIC Corp., https://datasheetspdf.com/pdf-file/1305198/TP4056/1 Accessed 6 January 2025.
17. Herff, Christian, and Tanja Schultz. “Automatic Speech Recognition from Neural Signals: A Focused Review.” Frontiers in Neuroscience, vol. 10, 2016, article 429, doi:10.3389/fnins.2016.00429.
18. Sereshkeh, Alborz R., Robert Trott, Aurélien Bricout, and Tom Chau. “EEG Classification of Covert Speech Using Regularized Neural Networks.” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, Dec. 2017, pp. 2292–2300, doi:10.1109/TASLP.2017.2758148.
19. Hollenstein, Nora, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. “ZuCo, a Simultaneous EEG and Eye-Tracking Resource for Natural Sentence Reading.” Scientific Data, vol. 5, no. 1, Dec. 2018, article 180291, doi:10.1038/sdata.2018.291.
20. Liu, Hanwen, Daniel Hajialigol, Benny Antony, Aiguo Han, and Xuan Wang. “EEG2TEXT: Open Vocabulary EEG-to-Text Decoding with EEG Pre-Training and Multi-View Transformer.” arXiv, 3 May 2024, arXiv:2405.02165.
21. Hollenstein, N., Rotsztejn, J., Troendle, M. et al. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. Sci Data 5, 180291 (2018). <https://doi.org/10.1038/sdata.2018.291>.
22. Zhang, Z., Ding, X., Bao, Y. et al. Chisco: An EEG-based BCI dataset for decoding of imagined speech. Sci Data 11, 1265 (2024). <https://doi.org/10.1038/s41597-024-04114-1>
23. Van der Goot, P. M. P., and C. A. J. P. C. de Goot. “Skin Impedance and Resistivity in Bioelectric Measurements.” Bioelectrochemistry, vol. 76, no. 2, 2009, pp. 149–156.
24. de Souza, S. J. B., et al. “Measurement of Skin Impedance and Resistivity in Various Conditions.” Biomedical Engineering Online, vol. 10, no. 1, 2001, pp. 45–53.
25. L. M. A. F. Costa et al. Changes in skin electrical properties with hydration. Skin Research and Technology, 24(4):567–573, 2018. Notes variations in epidermal resistivity with hydration levels.
26. U.S. Department of Health and Human Services. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Public Law 104-191, Government Printing Office, 1996.
27. Federal Trade Commission. “Privacy and Security.” *FTC*, www.ftc.gov/tips-advice/business-centerprivacy-and-security. Accessed 12 July 2025.
28. European Parliament and Council. Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, 27 Apr. 2016.
29. International Human Rights Advocacy. “Neurorights.” 2023.
30. Global Privacy Assembly. “Resolution on Neuro-specific Data Protection.” Global Privacy Assembly, 2024.
31. OpenStax College. *1605 Brocas and Wernickes Areas‑02 [Diagram of the human brain circling Broca’s and Wernicke’s areas, with arrows denoting the arcuate fasciculus].* *Anatomy & Physiology*, Connexions Web site, 19 June 2013, cnx.org/content/col11496/1.6/. *Wikimedia Commons*, https://upload.wikimedia.org/wikipedia/commons/2/20/1605\_Brocas\_and\_Wernickes\_Areas-02.jpg. Creative Commons Attribution‑ShareAlike 4.0 International. Accessed 30 July 2025.
32. Crosby, Steven A. “Efficient Data Structures for Tamper-Evident Logging.” *Proceedings of the IEEE Annual Computer Security Applications Conference*, IEEE, 2009.
33. Matetić, Siniša, et al. “DELEGATEE: Brokered Delegation Using Trusted Execution Environments.” *Proceedings of the 27th USENIX Security Symposium*, USENIX Association, Aug. 2018, pp. 1391–1408.
34. Rescorla, Eric. “The Transport Layer Security (TLS) Protocol Version 1.3.” *RFC 8446*, Internet Engineering Task Force, Aug. 2018, https://www.rfc-editor.org/info/rfc8446.
35. Stevens, Timothy. “Efficient Differentially Private Secure Aggregation for Federated Learning via Hardness of Learning with Errors.” *arXiv*, Dec. 2021, arxiv.org/abs/2112.06872.
36. “Evaluate Models  |  Cloud Translation  |  Google Cloud.” *Google*, Google, cloud.google.com/translate/docs/advanced/automl evaluate#:~:text=The%20BLEU%20score%20is%20a,of%20high%20quality%20reference%20translations. Accessed 6 Aug. 2025.
37. Lin, Chin-Yew. “ROUGE: A Package for Automatic Evaluation of Summaries.” \*Text Summarization Branches Out\*, Barcelona: Association for Computational Linguistics, July 2004, pp. 74–81. [https://www.aclweb.org/anthology/W04-1013](https://www.aclweb.org/anthology/W04-1013).