Engineering Journal issn: 0125-8281

Article

Comparison of Thai and English Speaking Signals from Brain Using Deep Learning and EEG

Ekkarat Adsawinnawanawa[a], Soliman Faragy[b], Supakij Buasod[c], and Narongdech Keeratipranon[d,*]

Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand
E-mail: [a]Ekkaratadsawinnawanawa@gmail.com, [b]solimankhulna@gmail.com, [c]supakijbuasod@gmail.com, [d,*]narongdech.k@chula.ac.th (Corresponding author)

Abstract. This study investigates the decoding and comparison of brain signals associated with spoken Thai and English words using deep learning techniques and EEG equipment. In the field of Brain-Computer Interfaces (BCI), researchers have extensively explored methods to decode brain signals into text. Two primary approaches exist: invasive (e.g., ECoG) and non-invasive (e.g., EEG). Invasive methods require surgery and offer high-quality signals but carry infection risks. Conversely, non-invasive methods employ scalp electrodes, resulting in lower signal quality but greater practicality for daily use. The present research utilizes three datasets each for Thai and English to evaluate the effectiveness of EEG and compare the outcomes for both languages. The Thai word data consists of three sets: single words (หิว, ปวด, เจ็บ, หนาว, ร้อน), two-word phrases (หิวมาก, ปวดท้อง, เจ็บแขน, หนาวมาก, ร้อนมาก), and three-word sentences (ฉันหิวมาก, ฉันปวดท้อง, ฉันเจ็บแขน, ฉันหนาวมาก, ฉันร้อนมาก). The English word datasets correspond semantically to each Thai set. All results are tested and compared using two machine learning approaches: Multi-Layer Perceptron (MLP) with statistical features and Convolutional Neural Network (CNN) with stacked spectrogram features. The MLP achieved an overall accuracy of 98%, while the CNN achieved 64%.

Keywords: BCI, ECoG, EEG, MLP, CNN.

ENGINEERING JOURNAL Volume 29 Issue 4
Received 10 July 2024
Accepted 3 April 2025
Published 30 April 2025
Online at https://engj.org/
DOI:10.4186/ej.2025.29.4.53

1. Introduction

Brain-computer interface (BCI) [1] [2] technology aims to facilitate direct communication between humans and computers by measuring the brain's electrical activity. There are two primary methods for acquiring brain signals: invasive and non-invasive. The invasive method involves a surgical procedure, which, although it can provide high-quality signals, carries a risk of infection [3]. This approach is utilized in various applications, including epilepsy surgery monitoring [4], control devices [5], and decoding attempted handwriting [6]. In contrast, the non-invasive method involves placing sensors on the scalp, offering lower signal quality but greater practicality for daily use. This method is applied in controlling a robot arm [7], wheelchair [8] [9], drone [10], silent EEG-speech recognition [11], and typing using a steady-state visual evoked potential (SSVEP) keyboard [12].

In contemporary times, numerous individuals experience impaired communication abilities due to conditions such as strokes [13], locked-in syndrome [14], spinal cord injury [15], laryngeal cancer [16], and others.

The progress of BCI technology inspires many researchers to investigate its potential in aiding individuals with neurological disorders. However, substantial obstacles impede these endeavors. Obtaining high-quality brain signals often requires invasive techniques, involving placing sensors directly on the cortical surface, albeit with the drawback of requiring surgical procedures and posing infection risks. In contrast, non-invasive methods like EEG provide an alternative that avoids surgical intervention and infection risks, utilizing their ease of use and portability, albeit with lower signal resolution compared to invasive approaches [17].

In prior research [18], EEG equipment demonstrated sufficient brain signal quality for classifying Thai-speaking words. This study enhances Thai word complexity and contrasts it with English words of equivalent meaning, evaluating two methods: MLP employing statistical features and CNN using stacked-spectrogram features. The primary objective is to advance toward practical applications in patient care

This paper is organized into seven sections. The first section provides an introduction to the research. The second section reviews related works, describes the equipment setup, and outlines the feature extraction techniques. The third section elaborates on the proposed methodology. The fourth section presents the experimental results, followed by additional experimental details in the fifth section. The sixth section offers the conclusion, while the final section discusses potential directions for future work.

2. Related Works

2.1. Invasive and Non-invasive Method Explanation

Numerous techniques exist for acquiring brain signals to decode them into text, encompassing both invasive and non-invasive methods. Invasive techniques, such as electrocorticography (ECoG), involve implanting electrodes directly on the brain's surface. While offering high-resolution signals, these methods are highly surgical and carry significant risks. Non-invasive techniques, like electroencephalography (EEG), are more widely used due to their safety and portability. However, EEG signals have lower spatial resolution compared to invasive methods.

2.1.1. Invasive method

The primary advantage of this method is the acquisition of high-quality brain signals through the implantation of electrodes into the intracranial region, which must be performed by an expert. However, this approach carries risks associated with surgical procedures, such as infection and scarring. Long-term implantation of the electrodes is not feasible due to the increased risk, and subjects require follow-up even after electrode removal [19].

Various types of equipment are utilized in this method for decoding brain signals into text. For instance, ECoG can decode spoken phrases from phonemic representations in the brain with a 25% word error rate [20], and high-resolution equipment can enhance the quality of speech decoding [21]. Additionally, microelectrodes have achieved a speech decoding speed of 62 words per minute [22].

2.1.2. Non-invasive method

The non-invasive method is straightforward and user-friendly, as it does not require surgical operations, carries a lower risk of infection, and needs no follow-up procedures. This translates to a more comfortable experience for participants and increased accessibility compared to invasive methods. However, a trade-off exists: non-invasive methods may have inherent delays in signal processing and potentially lower quality data compared to their invasive counterparts.

The equipment used in non-invasive speech decoding varies widely depending on the specific application. For instance, electromyography (EMG) can be used to decode imagined and intended spoken speech [23]. Functional near-infrared spectroscopy (fNIRS) focuses on identifying perceived speech [24] or recent studies have explored various methodologies to optimize fNIRS-based Brain-Computer Interface (BCI) models [25]. Meanwhile, electroencephalography (EEG) has shown promise in classifying silent speech of up to 9 words using a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [26]. For a more in-depth exploration of EEG-based imagined speech decoding, a comprehensive review is available [27].

2.2. Setting EEG Equipment

Our approach uses EEG equipment because it is easy to use, doesn't require surgical operations, is lower cost, and portable. All of these reasons will allow people to use it in a wide range once our work has more maturity. Although it's easy to use, there still need to be a few setups before using it as a brain signal acquisition. The EEG equipment setup can be divided into four steps. EEG equipment comes in various forms, such as headbands, caps, and even earbuds, each designed for different use cases. For example, an EEG headband can be used for daily stress monitoring, while a more comprehensive EEG cap might be utilized in a clinical setting for detailed brain activity analysis.

2.2.1. Connect the EEG equipment

The first step in preparing the EEG equipment is to connect all components. This involves linking the EEG headset, which sits comfortably on the scalp and captures brainwave activity, to an amplifier. The amplifier strengthens the weak electrical signals produced by the brain, making them easier to record and analyze. Once amplified, the signal is then connected to the computer via the provided line. It is important to ensure that the EEG headset is fully charged prior to use, as a low battery can lead to dropped signals and unreliable data. Avoid using the EEG headset while it's charging to prevent electrical noise from interfering with the delicate brainwave recordings. By following these steps carefully, you can ensure that your EEG equipment is properly set up to capture high-quality data for your experiment.

2.2.2. Wearing the EEG headset

To ensure accurate and reliable electroencephalography (EEG) data collection, proper placement of the EEG headset is crucial. This is achieved by positioning the headset on the scalp according to the international 10-20 system. This standardized system relies on anatomical landmarks to guide electrode placement. In the case of EEG headset application, two specific reference points are used to create a cross-line measurement. The first line extends from the Nasion, the indentation at the bridge of the nose, to the Inion, the bony bump at the back of the head. The second line is perpendicular to the first and spans from the Right Pre-Auricular Point, located in front of the earlobe, to the Left Pre-Auricular Point. The central electrode (Cz) should be positioned at the intersection of these two lines. This meticulous positioning ensures consistent electrode locations across participants, enabling reliable comparisons of EEG data. A visual representation of the correct headset placement is provided in Fig. 1.

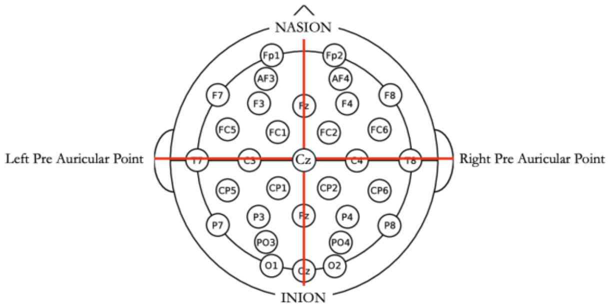

Fig. 1. The position of the EEG headset.

2.2.3. Application of conductive gel to electrode sensors

Each electrode sensor was meticulously filled with conductive gel using a syringe to ensure optimal scalp electroencephalogram (EEG) signal acquisition. The selection of the gel was crucial, as its conductivity played a vital role in facilitating the transmission of the brain's electrical activity. This low-impedance medium allowed for accurate measurement of the subtle voltage fluctuations generated by neural processes, ultimately providing a clear picture of brain function. As depicted in Fig. 2, the gel was applied directly to the scalp at designated electrode placement sites.

Fig. 2. Demonstration of Gel putting.

2.2.4. Impedance validation

Prior to data acquisition, impedance validation is a crucial step to ensure effective gel application and high-quality electrophysiological recordings. This process verifies proper electrical contact between the electrodes and the scalp, minimizing artifacts and maximizing signal fidelity. All channels must exhibit impedance values below a specified threshold, typically around 20 kilo-ohms. Higher impedance values can hinder data collection by the electrode sensor due to increased resistance to current flow, leading to attenuated signals and reduced signal-to-noise ratio. In the worst-case scenario, an infinite impedance value (INF) indicates a complete disconnect between the electrode and the scalp, rendering that channel unusable for data acquisition. This highlights the importance of meticulous electrode preparation techniques and proper gel application to achieve optimal low impedance values.

2.3. Feature Extraction

Feature extraction is a crucial process in data analysis, serving as a bridge between raw data and a model's ability to learn and differentiate patterns. Objectively, it seeks to extract relevant and informative features from the data, effectively transforming it into a format that a classification model can readily understand and utilize. This process plays a critical role in model learning, as the quality of the extracted features directly affects the model's performance. On one hand, overly complex feature extraction can lead to overfitting, where the model memorizes the training data too closely and fails to generalize to unseen examples. Conversely, overly simplistic feature extraction can result in underfitting, where the model struggles to learn the underlying relationships within the data and achieves subpar accuracy. Striking the right balance between complexity and information capture is essential for optimal model performance.

In this study, we leverage two distinct feature extraction methods: statistical features and stacked spectrograms. This combined approach aims to capture complementary information from the data, providing a richer representation for the classification model and potentially improving its ability to differentiate between samples.

2.3.1. Statistic feature

This feature includes 5 statistical functions: **minimum (Min), maximum (Max), mean, standard variation (SD), and variant (Var).** These functions calculated all data over the speaking periods of each selected EEG channel. The definition of each function is described in Table 1.

Table 1. Statistic function definition.

Function	Definition
Min	Minimum value of data
Max	Maximum value of data
Mean	Represent overall value over the speaking period
Var	Variance is the average of the squared differences from the mean.
SD	Standard deviation is the square root of the variance

2.3.2. Stacked spectrogram feature

The Spectrogram converts voltage magnitude data, captured during speech intervals, into a visual representation of the frequency spectrum. This spectral representation is a cornerstone feature in various disciplines due to its ability to reveal hidden information within a signal. For instance, spectrograms are widely used in audio processing to visualize and analyze the components of music or speech. In speech analysis, spectrograms can be used to identify phonemes, the basic units of sound in a language. Similarly, spectrograms play a crucial role in biomedical research fields like brain-computer interface development, where they can be used to understand brain activity patterns.

In standard practice, spectrograms typically process data from a single channel, providing a time-frequency representation of a single signal. In contrast, our research involves data from 14 channels, capturing information from multiple sources simultaneously. To account for this multi-channel data, we necessitate the concatenation of spectrograms from each channel into a row vector. Unlike traditional methods that convert each spectrogram into a separate, our technique directly concatenates the spectrogram values into a single vector. This approach offers advantages in terms of computational efficiency and memory usage. Figure 3 provides a visual representation of this Stacked Spectrogram approach.

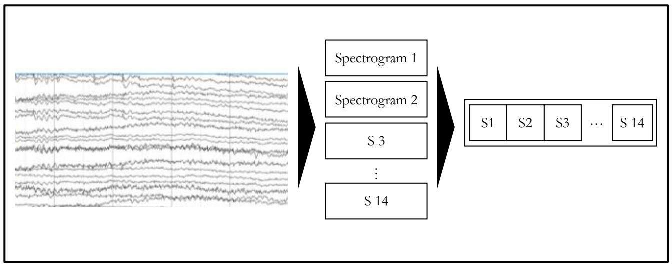

Fig. 3. The demonstration of Stacked Spectrogram feature.

3. Proposed Method

The proposed method involves six sequential processes illustrated in Fig. 4. Firstly, Data Acquisition involves acquiring brain signal data using a 32-channel EEGOmylab system from Ant Neuro. Details regarding the specific model, sampling rate, and electrode placement can be found in Table 2. This initial stage is crucial, as the quality of the acquired data directly impacts the effectiveness of subsequent processing steps. Following data acquisition, the EEG Channel Selection process identifies the most informative channels for speech classification. This selection can be based on various criteria, such as signal-to-noise ratio or prior knowledge about brain regions involved in speech production.

Next, the Signal Down Sampling process reduces the data sampling rate to a more manageable level for computational efficiency. However, this down-sampling must be carefully chosen to retain the essential information for speech classification. Here, a discussion on the chosen down-sampling method and its rationale can be included. The following step, Speaking Period Detection, segments the continuous EEG data into speech and non-speech intervals. This segmentation allows the model to focus on the relevant brain activity during speech production.

Extracted features from the segmented data form the basis for Feature Extraction. These features capture the temporal and spectral characteristics of the brain signals associated with speech. The selection of appropriate features is critical for accurate word classification. Finally, the Word Classification process utilizes a machine learning model to classify the extracted features into the different word.

Table 2. EEG Equipment Information.

Device	Information
Brand	Ant-neuro
Device	EEGOmylab
Channel	32 Channels
Electrode placement	10-20 system
Sampling rate	16384
Electric Conductor	Gel based

3.1. Data Acquisition

The dataset encompasses 30 words (28 records per word) of brain signals. These words were carefully chosen to represent a balanced mix of Thai and English, with 15 words from each language. The Thai words were further subcategorized into three sets based on their length: one set containing single words, another containing two-word phrases, and a final set containing three-word phrases. Mimicking this structure, the English words were also divided into three corresponding sets with the same intended meaning. This approach allows for an investigation into the influence of both language and word complexity on brain activity.

3.2. EEG Channel Selection

The EEG headset typically covers the entire head, allowing researchers to capture a comprehensive picture of brain activity across various lobes and regions. This comprehensive data can be valuable for understanding large-scale brain networks. However, for specific applications like speech production research, focusing on relevant brain areas can offer several advantages. By selectively choosing channels from the entire set, researchers can reduce the complexity of the data. This streamlining process not only simplifies analysis but can also potentially improve the accuracy of models trained on the data. Fewer channels lead to a lower dimensionality of the data, which can make it easier for machine learning algorithms to identify the underlying patterns relevant to speech production.

In this study, following previous research on speech production [28], only 14 channels were chosen from a possible 32. These channels were specifically selected to target brain regions known to be crucial for speech production processes and to avoid artifacts from other areas e.g. eye artifacts. The specific positions of these 14 channels are visualized in Fig. 5.

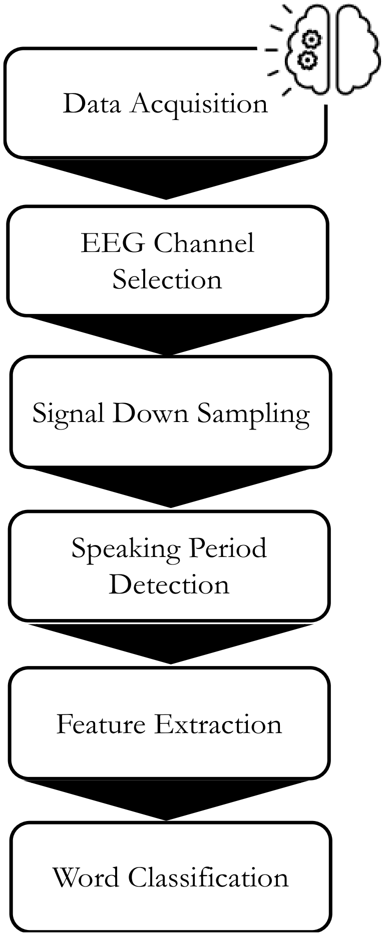

Fig. 4. Overview of the Proposed Method.

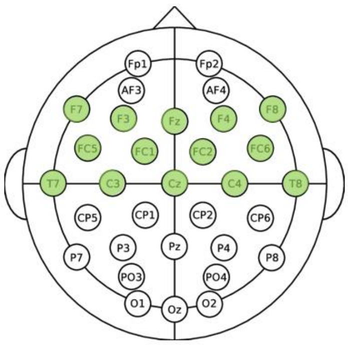

Fig. 5. The Position of the Selected EEG channel. The green highlight is the selected channels. It covers all the motor system areas.

DOI:10.4186/ej.2025.29.4.53

Table 3. Explanation of data detail

Data	Detail
Thai	**Set 1, 1 word**
	หิว, ปวด, เจ็บ, หนาว, ร้อน
	Set 2, 2 words
	หิวมาก, ปวดท้อง, เจ็บแขน, หนาวมาก, ร้อนมาก
	Set 3, 3 words
	ฉันหิวมาก, ฉันปวดท้อง, ฉันเจ็บแขน, ฉันหนาวมาก, ฉันร้อนมาก
English	**Set 1**
	Hungry, Ache, Pain, Cold, Hot
	Set 2
	Starve, Stomach ache, Arm Pain, Very Cold, Very Hot
	Set 3
	I am Starving, I have a stomach ache, I have Arm pains, I feel very cold, I feel very hot

3.3. Signal Down Sampling

The data was obtained at a sampling rate of 16,384 Hz, which translates to 16,384 data points captured per second. This high sampling rate captures intricate details of the signal but also demands substantial memory resources for storage and processing. To mitigate potential out-of-memory problems and enhance processing speed, a common practice known as downsampling was implemented. This technique reduces the number of data points collected over a specific time period. In this study, the sampling rate was reduced from 16,384 Hz to 256 Hz, resulting in a 64-fold decrease in data volume. This significant reduction allowed for more efficient storage and processing while still retaining the essential information within the signal for further processing.

3.4. Speaking Period Detection

Our current investigation focuses on word classification tasks within spoken language data. To achieve this, we have opted for a manual approach to segmenting speech periods. This involves the implementation of timestamps at the onset and offset of each speaking turn. Subsequently, during data utilization, the speech segments will be extracted by applying a time window of 3 seconds before and after the designated timestamps. This results in a final segment length of 6 seconds for each speaking period.

3.5. Feature Extraction

In this study, we utilized both statistical and stacked-spectrogram features with our models. The statistical features, including minimum, maximum, mean, variance, and standard deviation, succinctly summarize the data over time in an interpretable manner. These features, while straightforward, effectively capture key characteristics and demonstrate resilience to artifacts due to their summarizing nature. Additionally, stacked spectrograms were incorporated to represent the time-frequency spectrum, offering the model a more comprehensive and detailed understanding of the data, despite the increased complexity.

3.6. Word Classification

This study employs two deep learning models implemented within the Keras framework. The first model utilizes a Multi-Layer Perceptron (MLP) architecture trained on handcrafted statistical features. The second model leverages a Convolutional Neural Network (CNN) architecture that processes stacked spectrograms.

4. Experiment

The experiment was divided into three main groups: 1) Testing Thai words, 2) Testing English words, and 3) Testing mixed Thai and English. The first group assessed the classifier's performance on Thai words by testing and comparing each Thai set, along with cross-comparisons between sets. This methodology allowed for a comprehensive evaluation of the classifier's ability to differentiate between various Thai words with potentially subtle semantic differences. The second group employed the same methodology but with English words semantically equivalent to the Thai words. This comparison aimed to identify potential language-specific effects on the classifier's performance. Finally, the third group evaluated the classifier's ability to handle mixed Thai and English input, simulating a more realistic scenario where users might switch languages mid-thought.

Each data set was divided into Train/Test/Validate splits, with a 70%/20%/10% proportion, respectively, ensuring a fair representation of each class within each split. This division helps to prevent overfitting and allows for robust evaluation of the model's generalization capabilities. This study conducted two methods: first, using a Multi-Layer Perceptron (MLP) with the extracted statistical features, and second, using a Convolutional Neural Network (CNN) with stacked spectrograms as input. The results for all groups, employing both MLP and CNN architectures, are presented in Tables 4, 5, and 6.

Table 4. Thai Word Testing Result

Data Set	MLP-Stat	CNN-SS
TH1	0.96	0.96
TH2	0.82	1.00
TH3	0.86	0.82
TH1+TH2	0.95	0.46
TH1+TH3	0.93	0.59
TH2+TH3	0.91	0.30
TH1+TH2+TH3	0.99	0.65

Table 4 presents the performance comparison of MLP-Stat, utilizing statical features, and CNN-SS, employing stacked spectrograms, evaluated on the Thai word dataset. Both models achieve notable accuracy across TH1, TH2, and TH3. However, CNN-SS exhibits decreased accuracy, particularly on the mixed Set, unlike MLP-Stat, which demonstrates consistently high performance across all datasets. These results indicate that CNN-SS may encounter challenges with increased dataset complexity in Thai, potentially affecting its ability to effectively differentiate samples under more complex features and data conditions.

Table 5. English Word Testing Result.

Data Set	MLP-Stat	CNN-SS
EN1	1.00	1.00
EN2	0.96	1.00
EN3	0.86	0.89
EN1+ EN2	0.96	1.00
EN1+ EN3	0.96	0.82
EN2+ EN3	0.95	1.00
EN1+ EN2+ EN3	1.00	0.71

The findings from Table 5 indicate promising results for English word testing, demonstrating the effectiveness of both MLP-Stat and CNN-SS methods. Both models consistently outperform all datasets, with CNN-SS generally showing superior performance compared to MLP-Stat across most datasets, albeit showing a decline in performance on the final dataset (Mixed English words), which is the most challenging in the table. Nevertheless, MLP-Stat maintains consistent high performance throughout.

Table 6. Mixed Thai and English word testing result.

Data Set	MLP-Stat	CNN-SS
TH1+ EN1	0.96	0.98
TH2+ EN2	1.0	0.98
TH3+ EN3	0.93	0.66
TH1+ EN2	0.89	0.93
TH1+EN3	0.98	0.98
TH2+EN1	0.93	0.71
TH2+EN3	0.98	0.91
TH3+EN1	0.98	0.59
TH3+EN2	0.93	0.79
Mixed All TH + EN	0.98	0.64

Table 6 presents the results of the most challenging experiment, which evaluated the classifier's performance on a mixed dataset of Thai and English words. The data from each language set (TH1, TH2, TH3 for Thai and EN1, EN2, EN3 for English) were combined and cross-validated to assess the model's ability to differentiate between languages and across varying word complexities within each language. Notably, the Multi-Layer Perceptron (MLP) utilizing statistical features (MLP-Stat) maintained consistent high accuracy across all datasets. While the Convolutional Neural Network employing stacked spectrograms (CNN-SS) achieved slightly better accuracy in some cases, the difference was not statistically significant. Furthermore, the performance of CNN-SS generally declined with more complex Thai datasets, indicating potential limitations in capturing relevant features for mixed-language classification tasks with varying word complexity. This decline may be attributed to artifacts in the signal, as the duration and complexity of the spoken words increase.

MLP-Stat outperformed in the most complex dataset (Mixed of all Thai and English words), achieving 98% of accuracy while CNN-SS achieved 64%. The detail of the MLP-Stat model is shown below.

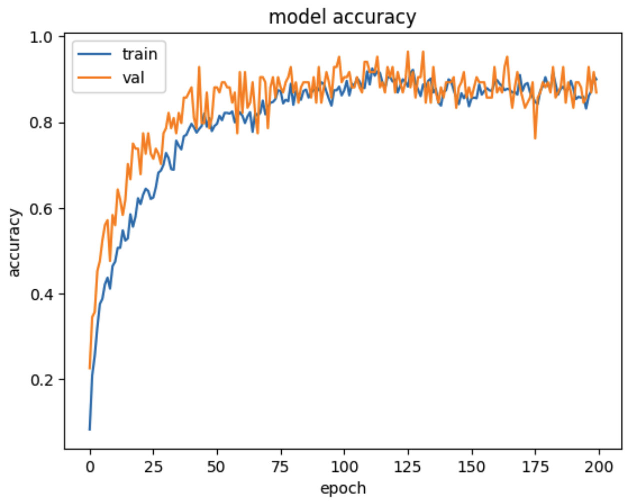

Fig. 6. Visualize the accuracy of training/validation in each epoch.

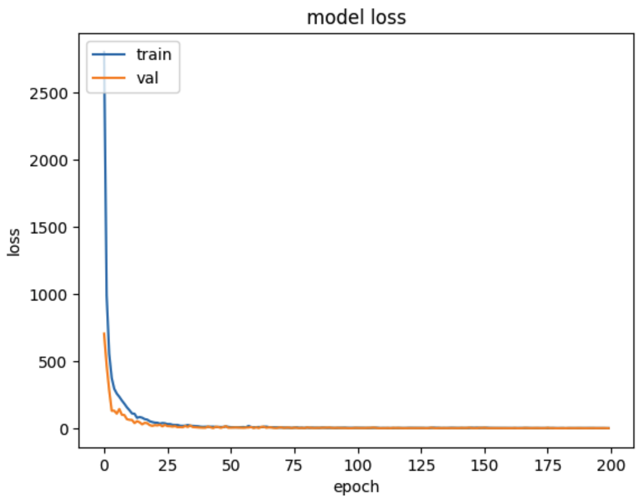

Fig. 7. Visualize the loss convergence while training the model.

Figure 6 and Fig. 7 illustrate the convergence of the training process. The model initially learns and converges rapidly within the first 40 epochs, then gradually slows down, achieving its optimal state at the 126th epoch.

The optimal model was evaluated using the dataset, and the accuracy details for the Mixed All TH + EN dataset are presented in Table 7. P is precision, R is Recall and F1 is F1-score. This dataset is the most complex among all experiments, comprising a mixture of Thai and English words. The training data consists of approximately 28 samples, the validation data around 3 samples, and the testing data between 5 and 6 samples. The data split was performed using the train_test_split() function from the scikit-learn python library. When the number of samples for training, validation, or testing is not an integer, it is rounded up or down according to the library's logic.

5. Additional Experiment

Statistical features serve as an effective method for simplifying raw EEG data. However, this is not the case for stacked spectrogram features, which exhibit higher complexity. To enhance the quality of these features, an artifact removal process can be employed. In this study, Independent Component Analysis (ICA) was applied in conjunction with Kurtosis to mitigate artifacts in EEG signals. Specifically, ICA was utilized to separate mixed sources within the raw EEG data, while Kurtosis was used to assess the tailedness of the signal. By integrating these techniques, independent components were effectively isolated, and signal tailedness was reduced, as it was assumed to be indicative of artifacts.

The impact of artifact removal was evaluated using the Mixed Thai and English Word dataset, with results presented in Table 8. The models MLP-Stat (ICA) and CNN-SS (ICA) represent the proposed approaches incorporating ICA. Notably, the accuracy of CNN-SS (Non-ICA) improved substantially following artifact removal (CNN-SS (ICA)), indicating that CNN-SS (Non-ICA) is highly sensitive to artifacts. In contrast, statistical features, which are inherently simplified, demonstrated a decline in accuracy after artifact removal. This reduction may be attributed to ICA inadvertently discarding essential signal components along with artifacts.

6. Conclusion

This study investigates the challenges associated with using EEG equipment for brain signal acquisition, with a particular focus on signal resolution and quality in the context of decoding spoken language into text. The experiment evaluated two classification methods, MLP-Stat (Non-ICA) and CNN-SS (Non-ICA), using EEG signals corresponding to Thai and English words with three levels of linguistic complexity. The findings reveal that MLP-Stat consistently outperforms CNN-SS (Non-ICA) across most datasets. Notably, the accuracy of CNN-SS (Non-ICA) declines as data complexity increases, likely due to the over-detailed nature of stacked-spectrogram feature extraction and the presence of artifacts that introduce noise into the model. However, applying Independent Component Analysis (ICA) to CNN-SS (ICA) resulted in a 33% improvement in accuracy. In contrast, MLP-Stat demonstrated robust performance without ICA, but its accuracy decreased when ICA was applied. This decline may be attributed to the ICA process removing potentially informative components, coupled with the inherent data simplification of the statistical feature extraction method, leading to an excessive reduction in information content.

Table 7. Testing Result of MLP-Stat on Mixed All TH + ENG.

	P	R	F1	#
ปวด (Ache) TH	1.00	1.00	1.00	6
หนาว (Cold) TH	1.00	0.83	0.91	6
ร้อน (Hot) TH	0.86	1.00	0.92	6
หิว (Hungry) TH	1.00	1.00	1.00	5
เจ็บ (pain) TH	1.00	1.00	1.00	6
ปวดท้อง (stomach ache) TH	1.00	1.00	1.00	6
หนาวมาก (very_cold) TH	1.00	0.80	0.89	5
ร้อนมาก (very_hot) TH	0.86	1.00	0.92	6
หิวมาก (Starve) TH	1.00	1.00	1.00	5
เจ็บแขน (Arm Pain) TH	1.00	1.00	1.00	6
ฉันปวดท้อง (I have stomach ache) TH	0.86	1.00	0.92	6
ฉันหนาวมาก (I feel very cold) TH	1.00	1.00	1.00	5
ฉันร้อนมาก (I feel very hot) TH	1.00	1.00	1.00	6
ฉันหิวมาก (I am very hungry) TH	1.00	0.83	0.91	6
ฉันเจ็บแขน (I have arm pain) TH	1.00	1.00	1.00	5
Ache_EN	1.00	1.00	1.00	5
Cold_EN	1.00	1.00	1.00	5
Hot_EN	1.00	1.00	1.00	6
Hungry_EN	1.00	1.00	1.00	5
Pain_EN	1.00	1.00	1.00	6
Stomach_Ache_EN	1.00	1.00	1.00	6
Very_Cold_EN	1.00	1.00	1.00	5
Very_Hot_EN	1.00	1.00	1.00	6
Starve_EN	1.00	1.00	1.00	6
Arm_Pain_EN	1.00	1.00	1.00	6
I_Have_a_Stomach_Ache_EN	1.00	1.00	1.00	5
I_Feel_Very_Cold_EN	1.00	0.83	0.91	6
I_Feel_Very_Hot_EN	0.86	1.00	0.92	6
I_Am_Starving_EN	1.00	1.00	1.00	5
I_Have_Arm_Pain_EN	1.00	1.00	1.00	5
accuracy			0.98	
macro avg	0.98	0.98	0.98	
weighted avg	0.98	0.98	0.98	

Table 8. MLP-Stat-ICA and CNN-SS-ICA Mixed Thai And English Word (**Mixed All TH + EN)** Accuracy Result.

	Non-ICA	**ICA**
MLP-Stat	<u>0.98</u>	0.90
CNN-SS	0.64	0.97

7. Future Work

At this preliminary stage of the experiment, several key questions remain to be explored. One crucial aspect is understanding the precise contribution of Independent Component Analysis (ICA) to our models. Additionally, it is necessary to determine whether a single model can effectively classify brain signals across different individuals, which requires data from a diverse group of participants for validation. Furthermore, developing practical solutions for real-world applications is essential for the advancement of this research

The ultimate aim of this research is to assist individuals suffering from conditions that impair their ability to move or speak, such as Locked-in syndrome, in regaining their ability to communicate. This could also benefit elderly individuals who have lost the ability to speak, allowing them to communicate with their families and those around them in the final 10-20 years of their lives. Beyond these specific cases, the broader implications of this research could revolutionize how we understand and interact with the human brain, potentially leading to new therapies and interventions for a variety of neurological conditions.

Acknowledgement

We extend our heartfelt gratitude to the Cognitive Fitness Center, King Chulalongkorn Memorial Hospital, the Thai Red Cross Society for their generous provision of EEG equipment and invaluable suggestions.

This research project is supported by grants for development of new faculty staff, Ratchadaphiseksomphot Fund, Chulalongkorn University.

References

[1] P. V. Dhole, D. G. Chaudhary, V. D. Dhangar, S. D. Shejul, and B. W. Gawali, "A review on EEG-based brain-computer interfaces for imagined speech classification," *Educational Administration: Theory and Practice*, vol. 30, no. 5, pp. 6019–6036, 2024.

[2] R. Ramadan, S. Refat, M. Elshahed, and R. Ali, "Basics of brain computer interface," in *Brain-Computer Interfaces: Current Trends and Applications*. Springer, 2015, pp. 31-50.

[3] R. Arya, F. Mangano, P. Horn, K. Holland, D. Rose, and T. Glauser, "Adverse events related to extraoperative invasive EEG monitoring with subdural grid electrodes: A systematic review and meta-analysis," *Epilepsia*, vol. 54, no. 5, pp. 828-839, 2013.

[4] T. Maehara, "Epilepsy surgery monitoring intraoperative electrocorticography," *No Shinkei Geka*, pp. 517-524, 2023.

[5] E. Leuthardt, G. Schalk, J. Wolpaw, J. Ojemann, and D. Moran, "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 1, pp. 63-71. 2004.

[6] F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, and K. V. Shenoy, "High-performance brain-to-text communication via handwriting," *Nature*, vol. 593, pp. 249–254, 2021.

[7] R. Bousseta, I. Ouakouak, M. Gharbi, and F. Regragui, "EEG based brain computer interface for controlling a robot arm movement through thought," IRBM, vol. 39, no. 2, pp.129-135, 2018.

[8] A. Dev, Md A. Rahman, and N. Mamun, "Design of an EEG-based brain controlled wheelchair for quadriplegic patients," in *IEEE 3rd International Conference for Convergence in Technology (I2CT)*, 2018.

[9] E. V. Krishna Rao, N. Yaswanthkumar Reddy, B. V. S. S. Greeshma, and Y. S. S. Vardhan Reddy, "EEG based smart wheelchair for disabled persons using non-invasive BCI," in *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, Greater Noida, India, 2022, pp. 440-446.

[10] A. Abdulwahhab, I. Myderrizi, and M. Mahmood, "Drone movement control by electroencephalography signals based on BCI system," *Advances in Electrical and Electronic Engineering*, pp. 216-224, 2022.

[11] D. Vorontsova, I. Menshikov, A. Zubov, K. Orlov, P. Rikunov, E. Zvereva, L. Flitman, A. Lanikin, A. Sokolova, S. Markov, and A. Bernadotte, "Silent EEG-speech recognition using convolutional and recurrent neural network with 85% accuracy of 9 words classification," *Sensors*, vol. 21, no. 20, p. 6744, 2021.

[12] X. Chen, B. Liu, Y. Wang, and X. Gao, "A spectrally-dense encoding method for designing a high-speed SSVEP-BCI with 120 stimuli," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2764-2772, 2022.

[13] A. J. Arumairaj, S. Agarwal, R. Borkar, and H. Park, "Habtes I. hemidiaphragmatic paralysis post stroke leading to hypercapneic respiratory failure," *Cureus*, vol. 13, no. 2, 2021.

[14] 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.

[15] E. Trevlaki, E. Panagiotopoulou, E. Zarra, A. Xristara-Papadopoulou, E. Trevlakis, and A.

DOI:10.4186/ej.2025.29.4.53

Chalkia, "Cervical spinal cord injuries: A comprehensive review of pathophysiology and quality of life," *International Journal of Physical Education Sports and Health*, pp. 221-227, 2023.

[16] E. Tan, M. D. Mody, and N. F. Saba, "Systemic therapy in non-conventional cancers of the larynx," *Oral Oncology*, vol. 82, pp.61-68, 2018.

[17] D. Steyrl, R. Kobler, and G. Müller-Putz, "On similarities and differences of invasive and non-invasive electrical brain signals in brain-computer interfacing," *Journal of Biomedical Science and Engineering*, vol. 9, no. 8, pp. 393-398, 2016.

[18] E. Adsawinnawanawa, S. Faragy, S. Buasod, R. Pruengkarn, and N. Keeratipranon, "Decode brain signal into thai word using EEG and L-SVM," in *2023 18th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, IEEE, 2023.

[19] 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, 2021, Art. no. 599549.

[20] C. Herff, D. Heger, A. De Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, "Brain-to-text: Decoding spoken phrases from phone representations in the brain," *Frontiers in Neuroscience*, vol. 9, 2015, Art. no. 217.

[21] S. Duraivel, S. Rahimpour, C. H. Chiang, M. Trumpis, C. Wang, K. Barth, S. C. Harward, S. P. Lad, A. H. Friedman, D. G. Southwell, S. R. Sinha, J. Viventi, and G. B. Cogan, "High-resolution neural recordings improve the accuracy of speech decoding," *Nature Communications*, vol. 14, no. 1, p. 6938, 2023.

[22] F. R. Willett, E. M. Kunz, C. Fan, D. T. Avansino, G. H. Wilson, E. Y. Choi, F. Kamdar, M. F. Glasser, L. R. Hochberg, S. Druckmann, K. V. Shenoy, and J. M. Henderson, "A high-performance speech neuroprosthesis," *Nature*, vol. 620, no. 7976, pp. 1031-1036, 2023.

[23] D. Dash, P. Ferrari, and J. Wang, "Decoding imagined and spoken phrases from non-invasive neural (MEG) signals," *Frontiers in Neuroscience*, vol. 14, p. 290, 2020.

[24] Y. Liu, and H. Ayaz, "Speech recognition via fNIRS based brain signals," *Frontiers in Neuroscience*, vol. 12, p. 695, 2018.

[25] J. H. Ong and K. S. Chia, "Developing an optimal brain computer interface model using functional near infrared spectroscopy: A Review," *Engineering Journal*, vol. 27, no. 10, pp. 21-31, 2023.

[26] D. Vorontsova, I. Menshikov, A. Zubov, K. Orlov, P. Rikunov, E. Zvereva, L. Flitman, A. Lanikin, A. Sokolova, S. Markov, and A. Bernadotte, "Silent EEG-speech recognition using convolutional and recurrent neural network with 85% accuracy of 9 words classification," *Sensors (Basel)*, vol. 21, no. 20, p. 6744, 2021.

[27] 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, p. 867281, 2022.

[28] A. Flinker, A. Korzeniewska, A. Y. Shestyuk, P. J. Franaszczuk, N. F. Dronkers, R. T. Knight, and N. E. Crone, "Redefining the role of Broca's area in speech," in *Proc Natl Acad Sci USA*, 2015.

DOI:10.4186/ej.2025.29.4.53

Ekkarat Adsawinnawanawa was born in Bangkok, Thailand, in 1993. He received the B.E. degree in computer engineering from the Dhurakij Pundit University, Bangkok, Thailand, in 2016, and the M.E. in computer engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 2019.

He has been PhD student in Computer Engineering, at Chulalongkorn University since 2022. He's interested in Artificial Intelligence, Brain Computer Interface, Natural Language Processing, and Optimization Problems for Fintech industry.

MD Soliman Faragy was born in Bagerhat, Bangladesh in 1988. He received the Bachelor of science degree in computer engineering from Metropolitan University, Sylhet, Bangladesh, in 2014.

He has been M.Sc. Engineering student in Computer Engineering, at Chulalongkorn University since 2022. He's interested in Artificial Intelligence, Brain Computer Interface, Modern Software Engineering and Algorithm for Fintech industry.

Supakij Buasod was born in Chiang Mai, Thailand in september 1999. After completing his pre-engineering studies at the College of Integrated Science and Technology, RMUTL, he enrolled at King Mongkut's University of Technology Thonburi, where he earned a B.Eng in computer engineering. Since 2022, he has had an M.Eng. degree in computer engineering from Chulalongkorn University. His areas of interest are artificial intelligence, microcontrollers, and embedded systems.

Narongdech Keeratipranon was born in Bangkok, Thailand, in 1981. He received the B.E. degree in computer engineering from the Chulalongkorn University, Bangkok, Thailand, in 2000, and the MEng.Sc. in computer and communication engineering and Ph.D. degrees from Queensland University of Technology, Brisbane, Australia, in 2003 and 2009, respectively.

He has worked in the Department of Computer Engineering, Chulalongkorn University since 2022. His current research interests include robotic, embedded system, and computational intelligence.