

Sentence-Level Neural Language Decoding based on Speech Imagery from EEG Signals

Seong-Hyun Yu^{*1}, Hyeong-Yeong Park^{*2}, Seo-Jin Lee^{*3}, and Ji-Hoon Jeong^{**}

This research was supported by Chungbuk National University Korea National University Development Project (2022)

Abstract

This study focuses on the importance of communication in the healthcare field and the challenges faced by patients who are unable to speak due to medical conditions or treatments in expressing their needs and concerns. It highlights the potential for providing assistive solutions in such situations. To address this issue, we propose a novel approach using speech imagery of sentences, a technique where one imagines speaking without actually producing sound. The study collected electroencephalography(EEG) data from healthy participants and compared it with data collected during speech imagery generation. The study employed a dataset comprising four affirmative class sentences and four negative class sentences to conduct the experiment, utilizing two classifiers and two deep learning techniques for analysis. The results revealed that the classification accuracy for the affirmative class sentences was highest when employing regularized linear discriminant analysis(RLDA), while the classification accuracy for the negative class sentences was highest when using support vector machine(SVM). Although the study was conducted with a sample of healthy participants, it underscores the potential of speech imagery as a bidirectional communication modality for individuals who are unable to speak. Furthermore, this research represents a promising avenue for future investigations, focusing on decoding the intended messages of a select population with communication impairments.

요약

의료 분야와 같은 환경에서 직접적인 음성 발화가 아닌 발화 상상만으로 의사소통을 하는 상황을 가정하였을 때, 문장 단위의 발화 상상이 해석될 수 있는지에 대한 연구를 진행하였다. 피실험자들은 뇌전도 측정 기기(이하 EEG)를 착용한 후, 4개의 긍정문과 4개의 부정문 총 8문장에 대한 발화 상상을 진행했다. 기록된 EEG 데이터는 전처리하여 4개의 분류기(RLDA, SVM, LSTM, DNN)를 통해 분류하였다. 긍정문 4문장, 부정문 4문장 분류에서는 각각 RLDA (0.4145), SVM (0.3625) 이 가장 높은 분류 성능을 보였으며, 이는 4개 문장의 chance level인 0.2500 보다 모두 높은 값을 보여 발화 상상이 해석될 수 있음에 유의미한 수치를 보여준다. 이 연구는 문장 수준의 발화 상상에 대한 해석 가능성을 확인했으며, 추후 발화에 어려움이 있는 실제 환자들을 대상으로 추가 연구가 필요하다.

Keywords

brain-computer interface, BCI, electroencephalography, EEG, speech imagery, neural language, machine learning

* Dept. of Computer Science, Chungbuk National University
- ORCID¹: <https://orcid.org/0000-0003-3455-9753>
- ORCID²: <https://orcid.org/0000-0001-6943-4171>
- ORCID³: <https://orcid.org/0000-0003-3311-0890>
** Dept. of Computer Science & Research Institute for
Computer and Information Communication, Chungbuk National
University
- ORCID: <https://orcid.org/0000-0001-6940-2700>

• Received: Jun. 15, 2023, Revised: Jul. 07, 2023, Accepted: Jul. 10, 2023
• Corresponding Author : Ji-Hoon Jeong
Dept. of Computer Science, Chungbuk National University, 28644,
Chungdaero 1, Cheongju, Chungcheongbuk-do, Korea
Tel.: +82-43-261-2254, Email: jh.jeong@chungbuk.ac.kr

1. Introduction

Brain-computer interface(BCI) is a technology designed to enable direct communication between the human brain and external devices[1]. The BCI systems first acquire brain signals and then convert them into commands that control external devices such as prosthetic limbs or computer programs[2]. The development of BCIs aims to find new ways to support individuals with disabilities or impairments that limit their ability to communicate or interact with the world around them[3][4]. BCIs have shown great promise in helping people with conditions such as paralysis, locked-in syndrome, and motor neuron disease communicate and control their environments[5]. In particular, electroencephalography (EEG)-based BCI has recently received a lot of attention due to its advantage of being able to acquire brain signals non-invasively from the scalp[6].

Speech imagery is one of the promising paradigms for EEG-based BCI to allow individuals to communicate with others by modulating their brain activities based on the cognitive process involving mentally rehearsing speaking without producing any audible sounds[7]. Speech imagery has several advantages over other BCI paradigms, such as being suitable for people with motor impairments or speech disorders[8] and being a natural and intuitive way to communicate[9][10]. Moreover, it has the potential to generate continuous speech output, allowing for more fluid and flexible communication[11]. Previous studies have suggested that speech imagery can activate similar brain regions to actual speech production, making it a promising communication tool for patients with speech disorders. However, more research is needed to explore its potential applications. In order to achieve classification of a greater number of classes, it is imperative to pursue a multifaceted approach across various domains. For instance, the absence of established, formalized algorithms for speech imagery

classification necessitates the application of diverse algorithms, with special attention given to the frequency domain[12].

The main contributions of this study are as below. The primary objective of this study is to investigate the classification of language at the sentence level, as opposed to the word level. The utilization of EEG signals generated during speech imagery serves as the basis for accurately classifying sentences containing specific words. Previous research efforts in the field have predominantly concentrated on word-level language classification and vowel classification, yielding noteworthy advancements and significant findings. However, this study represents a notable contribution by providing empirical evidence and demonstrating the feasibility of effectively classifying language at the sentence level.

Secondly, the study demonstrates the capability of machine learning and deep learning techniques, such as regularized linear discriminant analysis(RLDA), support vector machines(SVM), long short-term memory(LSTM), and deep neural networks(DNN), to decode the speech imagery paradigm from EEG signals. Specifically, the research findings indicate the ability to accurately classify four types of sentences, including affirmative and negative statements, based on the patterns of brainwave signals generated during speech imagery. This signifies the successful decoding of language-related information from EEG signals and highlights the potential of these computational approaches in understanding and classifying sentence-level language representations. Overall, the paper's contributions suggest that EEG-based BCIs using speech imagery as a paradigm for communication could provide a practical and effective means for individuals with speech impairments to communicate effectively, including the ability to express negation in their communication.

II. Related Works

As mentions earlier in the Introduction section, BCI is a field of research and development focused on enabling direct communication between humans and external devices. A prominent example of this is the decoding of motor imagery to control exoskeletons[12]. The technology of detecting eye movements to facilitate communication by typing the intended message is a well-known early BCI communication tool[13].

Table 1 provides a comprehensive overview of research conducted on speech imagery decoding. Indeed, many studies on speech imagery decoding have utilized EEG(Electroencephalography) to capture the brain's electrical signals. EEG is a non-invasive neurophysiological measurement technique widely used in scientific research and clinical applications. As shown in Fig. 1., this process entails the placement of electrodes on the scalp to capture the brain's electrical activity, which offers valuable insights into temporal dynamics of brain function and cognition.

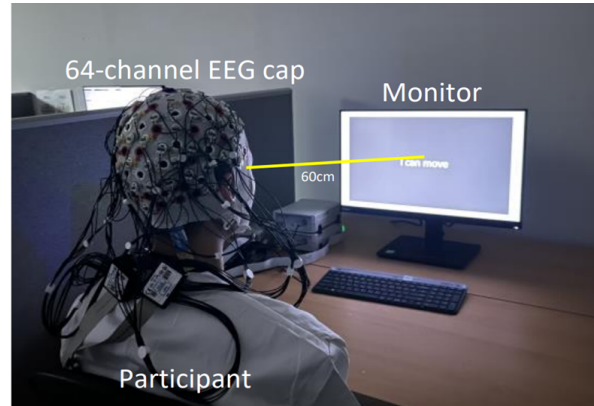


Fig. 1. Experimental protocol: The overall EEG recording experimental setup for acquiring brain activity and a snapshot of the actual task performed by the participants, which lasted for 3 seconds

In the early research on speech imagery, decoding of phoneme-level imagined speech was conducted, and analysis was also carried out to investigate which parts of the brain are activated in conjunction with a neuroscientific approach. Notably, advanced decoding algorithms have been applied to evaluate the neural correlates of vowels and consonants, a particularly popular experimental paradigm.

Table 1. Related studies for speech imagery decoding from neural activity

References	Method	Experiment type	Experiment paradigm
D'Zmura et al., 2009 [24]	EEG	Non-invasive	Imagined speech of two syllables spoken in one of three rhythms
Leuthardt et al., 2011 [25]	ECoG	Invasive	Overt and imagined phoneme articulation
Kim et al., 2013 [18]	EEG	Non-invasive	Speech of monosyllabic Korean words representing two categories of meaning (number and face)
Martin et al., 2014 [26]	ECoG	Invasive	Overt and covert reading of short stories
Iqbal et al., 2015 [27]	EEG	Non-invasive	Imagined speech of vowels /a/ and /u/, and no action
Iqbal et al., 2016 [28]	EEG	Non-invasive	Imagined speech of vowels /a/ and /u/, and no action
Yoshimura et al., 2016 [29]	EEG/fMRI	Non-invasive	Imagined speech production of Japanese vowels /a/ and /i/
Nguyen et al., 2017 [30]	EEG	Non-invasive	Imagined speech of short words, long words, and vowels
Hashim et al., 2018 [31]	EEG	Non-invasive	Imagined speech word production
Lee et al., 2018 [32]	EEG	Non-invasive	Imagined speech word production
Lee et al., 2021 [33]	EEG	Non-invasive	Imagined speech word production

However, in order to further advance it as a more natural communication tool, ongoing research is being conducted on decoding speech imagery at the word level. To decode speech imagery, participants need to engage in speech imagery while their EEG signals are recorded according to a specific paradigm. In the field of speech imagery research, the most validated paradigm is the endogenous paradigm[23][24]. In this paradigm, participants are instructed to generate speech imagery internally, without any external stimuli or cues. They imagine themselves speaking or uttering specific words or sentences in their minds, without actually producing any audible sounds.

There are still limitations in the meaningful speech imagery decoding research of many researchers. One limitation in current speech imagery research is the relatively low decoding accuracy and limited vocabulary size achieved in decoding imagined speech from brain signals. This can restrict the effectiveness and usability of speech imagery-based communication tools. Another limitation is the generalizability of decoding models across individuals. This individual variability poses challenges in developing robust and personalized speech imagery decoding systems. Therefore, to overcome the limitation of vocabulary size in speech imagery decoding as stated in the first limitation, we aim to propose a solution by implementing subject-dependent decoding of full-sentence speech imagery.

III. Sentence-Level Neural Language Decoding

3.1 EEG preprocessing

The raw EEG data were initially processed in MATLAB R2022b(MathWorks Inc., USA) using the EEGLAB toolbox. A Butterworth bandpass filter was applied to the continuous EEG data between 30 and 125 Hz to remove unwanted frequencies. To remove

artifacts from the EEG signal, we conducted independent component analysis(ICA)[25]. We carefully examined the ICA components and removed the ones associated with obvious artifacts. Subsequently, we applied baseline correction by subtracting the mean amplitude of the EEG signal in the prestimulus period from each time point. The processed EEG data were visually inspected to ensure it was free of artifacts.

3.2 Experimental protocols

Six healthy male individuals between the ages of 20 and 25 participated in the study after receiving an explanation of the experimental paradigm and protocol. Prior to the experiment, written informed consent was obtained from all participants. They provided written informed consent, as per the Declaration of Helsinki guidelines. The Institutional Review Board of Korea University approved all the experimental protocols under the KUIRB-2019-0143-01.

Standard EEG data preprocessing techniques were used, including filtering, artifact rejection, and feature extraction. The accuracy of the speech imagery decoding was evaluated using several performance metrics, including classification accuracy.

The EEG signals were recorded using a 64-channel EEG cap and BrainVision(BrainProduct GmbH, Germany) Recorder software, with a sampling rate of 1,000 Hz and an impedance below 10 Ω for all channels. The participants were seated in a soundproof and dimly lit room and instructed to keep their eyes closed and remain relaxed for 10 minutes before EEG data was recorded. They performed EEG recordings following the paradigm as depicted in Fig. 2.

Fig. 1. and Fig. 2. show detailed information about the experimental environment and the experimental paradigm, respectively. After ensuring that the participants have taken sufficient rest, a 3-second visual cue is presented, followed by 1 second of fixation before performing the imagined speech task for 3 seconds.

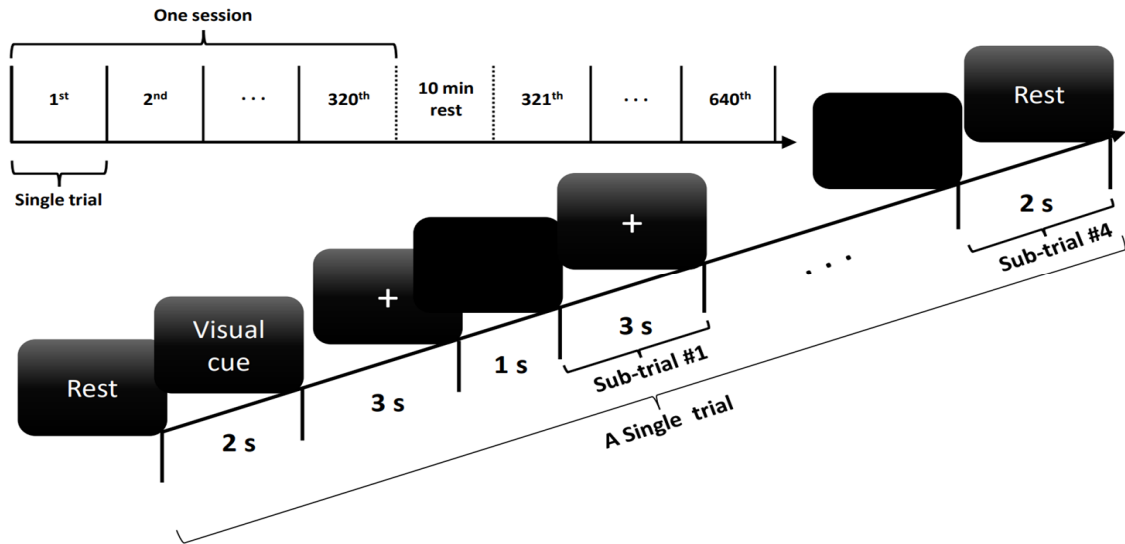


Fig. 2. Experimental paradigm of imagined speech. The figure presents one single trial of the experiment. Visual cues were provided for imagined speech. Four sub-trial sessions were conducted per class to output up to 10 visual cues per class and overall the experimental paradigm consisted of two sessions consisting of a total of 320 trials per session

3.3 Data analysis

The preprocessed EEG data were segmented into epochs of 3 seconds duration related to stimulus onset.

The experiment involved three phases - training, calibration, and testing, where participants were instructed to imagine speaking affirmative sentences (e.g., “I am thirsty”, “I can move”, “I am comfortable”, “I am digestible”) and negative sentences (e.g., “I am not thirsty”, “I can not move”, “I am uncomfortable”, “I am indigestible”). In the calibration phase, a classifier was built for affirmative and negative speech imagery using different sentences. In the testing phase, new sentences were given to decode the intended speech imagery using RLDA[26], SVM[27], LSTM[28], and DNN[29] and classification accuracy was used to evaluate the performance.

RLDA: A model that has been developed as a variant of LDA for pattern recognition and classification problems. LDA seeks to identify a separating hyperplane that maximizes the ratio of between-class scatter to within-class scatter. However, the estimation of these quantities in LDA can become unstable when dealing with high-dimensional data. To

address this issue, RLDA introduces regularization methods to control the complexity of the model parameters. RLDA can improve the stability and generalization performance of LDA, especially in high-dimensional data.

SVM: In EEG signal analysis, both temporal and spatial characteristics need to be considered. Therefore, SVM can extract the nonlinear characteristics of EEG signals by using a nonlinear kernel function.

LSTM: LSTM has been widely used in the analysis of EEG signals. In EEG signal analysis, LSTM has been applied to a variety of tasks, including the classification of EEG signals for the diagnosis of neurological disorders, seizure detection, and brain-computer interfaces. By modeling the temporal dynamics of EEG signals using LSTM, it is possible to capture complex relationships and patterns that may not be evident through traditional signal processing techniques. In this study, we employed an LSTM layer consisting of 32 neurons, incorporating 0.2 dropout and 0.001 regularizers in order to mitigate overfitting[30]. A total of 83.34% of the data was utilized for training with 10-fold cross-validation, while the remaining 16.66% was reserved for testing the model's performance.

DNN: Our adopted deep neural network(DNN) architecture employs the backpropagation algorithm with Adam optimization for training, and is a four-layered multi-layer perceptron[31][32]. The first layer functions as the input layer, whereas the second and third layers are each comprised of two hidden layers. Specifically, each hidden layer includes a fully connected(FC) layer, a batch normalization layer, and a leaky-RELU non-linear activation function. The output layer comprises a FC layer(dense layer) and a softmax function, which is utilized as the loss function during model training.

IV. Experimental Results

We conducted classification experiments using conventional classifiers to distinguish between eight classes of speech-related signals in two conditions (affirmative and negative). Due to BCI illiteracy, decoding was not possible for 2 out of 6 subjects, and they were therefore excluded from the results[33].

Firstly, we assessed whether classification between sentences corresponding to the affirmative and negative classes and a rest period state(Rest class) was feasible to measure sentence-level decoding performance. The average classification accuracies between each class and the Rest class, as shown in Tables 2 and 3, were 0.7079 (RLDA), 0.6875 (SVM), 0.7053 (LSTM), 0.6293 (DNN) for the affirmative class and 0.7191 (RLDA), 0.7258 (SVM), 0.7031 (LSTM), 0.6205 (DNN) for the negative class. These accuracies surpass the chance level (0.5000), indicating significant performance. Tables 4 and 5 present the classification accuracies for each subject using a classifier for the four classes corresponding to the two conditions, which were 0.4145 (RLDA), 0.3958 (SVM), 0.3073 (LSTM), 0.3990 (DNN) for the affirmative class and 0.3505 (RLDA), 0.3625 (SVM), 0.3035 (LSTM), 0.3340 (DNN) for the negative class.

The overall average accuracy of the affirmative class was slightly higher than that of the negative

Table 2. Binary classification accuracy of each affirmative sentence and the “Rest” state

Task vs. Rest	RLDA	SVM	LSTM	DNN
I am thirsty	0.7500	0.6720	0.7290	0.6565
I can move	0.6745	0.5935	0.6700	0.5795
I am comfortable	0.6410	0.7035	0.6875	0.6090
I am digestible	0.7660	0.7810	0.7345	0.6720
Average	0.7079 (±0.0518)	0.6875 (±0.0672)	0.7053 (±0.027)	0.6293 (±0.0369)

Table 3. Binary classification accuracy of each negative sentence and the “Rest” state

Task vs. Rest	RLDA	SVM	LSTM	DNN
I am not thirsty	0.6875	0.7020	0.7345	0.6410
I can not move	0.7345	0.6870	0.6875	0.6065
I am uncomfortable	0.7510	0.7810	0.7185	0.6250
I am indigestible	0.7035	0.7330	0.6720	0.6095
Average	0.7191 (±0.0249)	0.7258 (±0.0359)	0.7031 (±0.0247)	0.6205 (±0.0138)

Table 4. Accuracy of 4-affirmative sentence classification

Subjects	RLDA	SVM	LSTM	DNN
Subject 1	0.3920	0.3840	0.3190	0.3210
Subject 2	0.4220	0.3750	0.2790	0.3530
Subject 3	0.4380	0.4220	0.3350	0.5310
Subject 4	0.4060	0.4020	0.2960	0.3910
Average	0.4145 (±0.0172)	0.3958 (±0.0180)	0.3073 (±0.0214)	0.3990 (±0.0801)

Table 5. Accuracy of 4-negative sentence classification

Subjects	RLDA	SVM	LSTM	DNN
Subject 1	0.3560	0.4040	0.3140	0.3400
Subject 2	0.3280	0.2970	0.2880	0.3120
Subject 3	0.3270	0.350	0.2780	0.3620
Subject 4	0.3910	0.3900	0.3340	0.3220
Average	0.3505 (±0.0261)	0.3625 (±0.0412)	0.3035 (±0.0220)	0.3340 (±0.0190)

class, but we found no correlation between the accuracy of the two classes and the result value.

Among the classifiers, RLDA performed the best for the affirmative class, while SVM performed the best for the negative class. However, the LSTM classifier exhibits the lowest classification accuracy in both the affirmative and negative classes.

Fig. 3. is the grand-average confusion matrix for the affirmative sentence class across 4 subjects. The x-axis represents the predicted values, while the y-axis represents the actual values. The RLDA classifier was used for this analysis. The true positive rates(TPR) for all 4 classes surpassed the chance level (>0.25). Notably, the sentence ‘I am digestible’ exhibited the highest precision, with a precision value of 0.4855, indicating the most accurate classification performance within in this class. and Fig. 4. show the grand-average confusion matrix for the negative sentence class, utilizing the SVM classifier. The TPR for all 4 negative sentences were higher than the chance level (>0.25). The sentence ‘I am indigestible’ achieved the highest precision, with a precision value of 0.5210, signifying the most accurate classification performance within this class. These matrices of Fig. 3. and Fig. 4. were generated by a standard 10-fold cross-validation method. In our study, we visualized the brain activity regions using topoplots, which displayed the EEG signals as spatial distributions. Fig. 5. and Fig. 6. show the topoplots of participant #2. It was observed that stimulation occurred in Brodmann areas 44 and 45[33].

The left hemisphere for both the affirmative and negative classes. In the negative class, clear stimulation in the left parietal lobe's Brodmann areas 44 and 45 was observed between 2.4 and 3.0 seconds, while in the affirmative class, stimulation mainly occurred in the left parietal lobe at early times (0.6 seconds - 1.8 seconds), and in the negative class, it occurred at later times (1.8 seconds - 3.0 seconds).

In the affirmative class, the stimulation patterns typically appeared in a small area near the left hemisphere, while in the negative class, the patterns tended to occur in larger areas and often in two or more spatial regions. Additionally, the patterns in the negative class were more dispersed. These differences

may stem from distinct neurobiological mechanisms underlying the affirmative and negative classes.

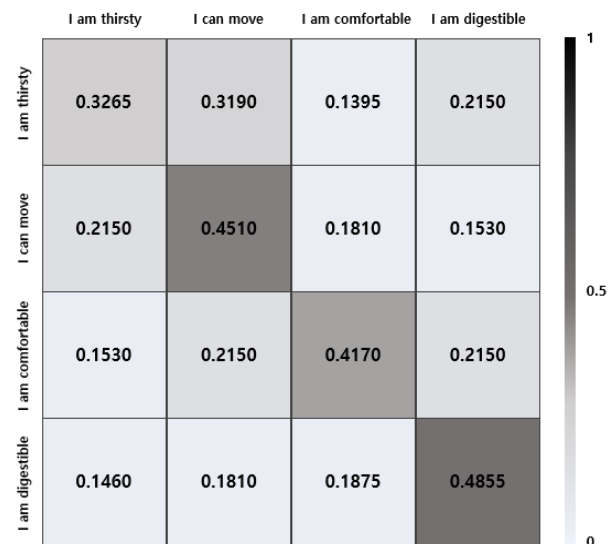


Fig. 3. Grand-average confusion matrix for the affirmative sentence class was obtained using the RLDA classifier. The class ‘I am digestible’ exhibited the highest accuracy of 0.4855 among the classes

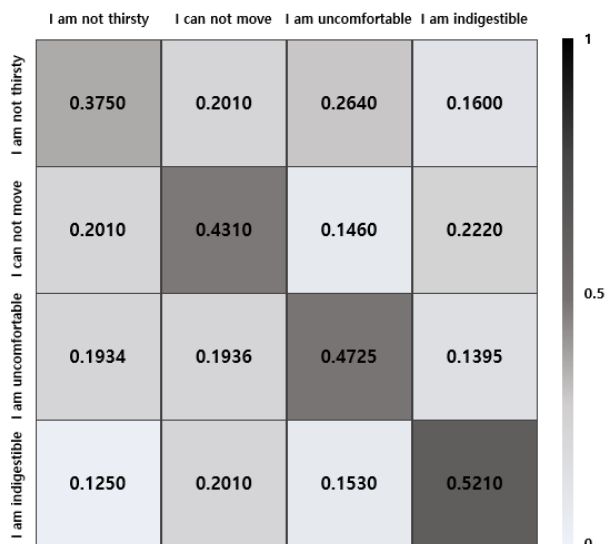


Fig. 4. Grand-average confusion matrix for the negative sentences class was represented. Although SVM showed higher average accuracy compared to RLDA, the RLDA classifier was chosen due to its superior performance in most cases. Specifically, the class ‘I am indigestible’ exhibited the highest accuracy of 0.521

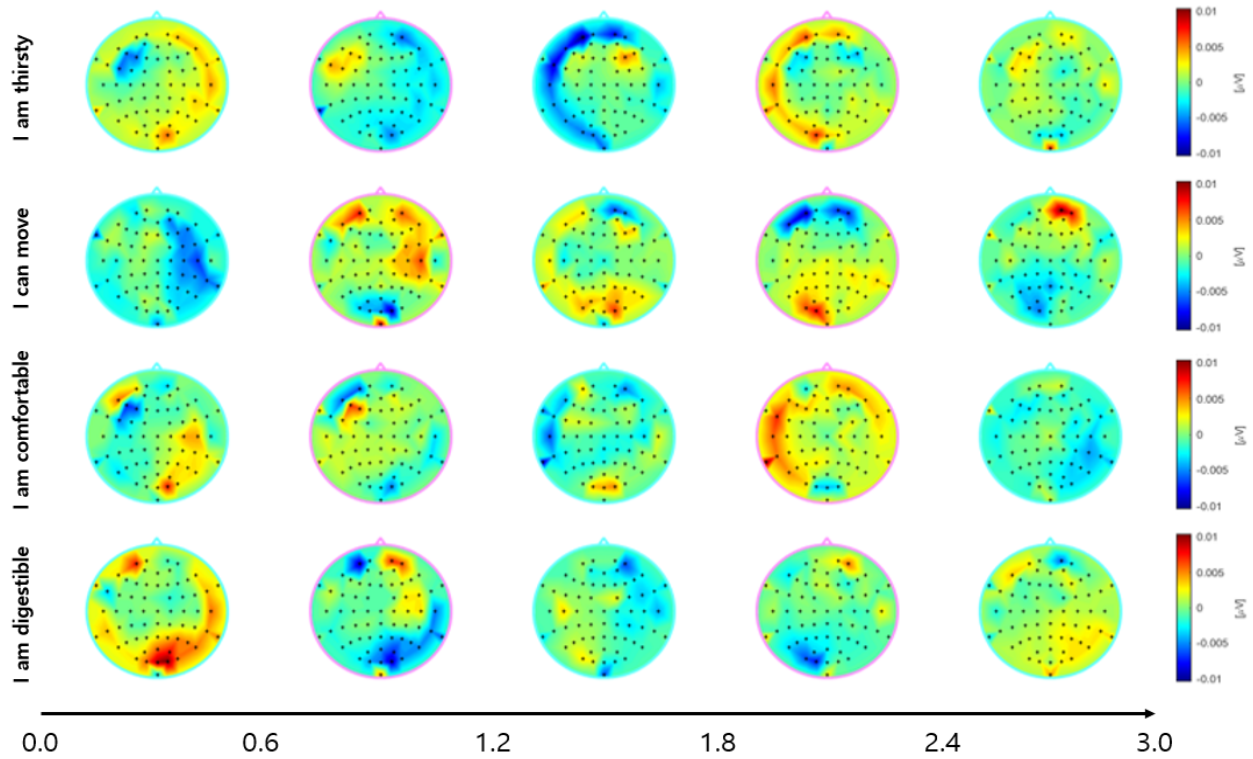


Fig. 5. EEG topography from participant#2 for the affirmative sentence class, are divided into five equal parts at 0.6 seconds intervals from 0 to 3 seconds. The patterns indicate that stimulation in the left parietal lobe occurred mainly at early times (0.6 seconds - 1.8 seconds) for the affirmative class

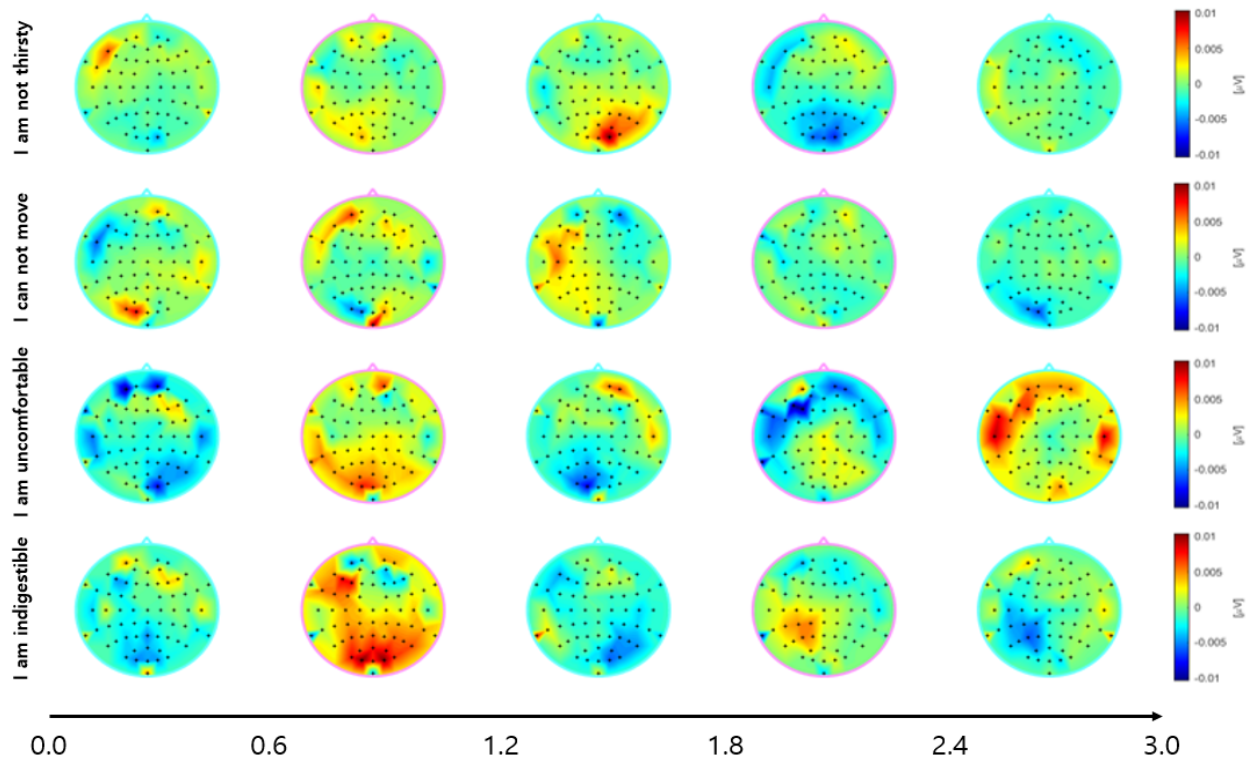


Fig. 6. EEG topography from participant#2 for the negative class. The patterns display the occurrence of stimulation in the left hemisphere's Brodmann areas 44(Borca's area) and 45. Clear stimulation in the left parietal lobe's Brodmann areas 44 and 45 is observed between 2.4 and 3.0 seconds

V. Discussion and Conclusion

The findings of this study suggest the potential for developing alternative communication methods to address the needs of individuals who are unable to communicate verbally. If future research confirms the distinguishability and reliability of affirmative and negative speech imagery patterns, it could provide a foundation for the development of non-invasive communication technologies based on brain-computer interfaces.

These technologies have the potential to greatly enhance the quality of life for individuals facing communication challenges due to neurological disorders, traumatic brain injury, or stroke, among other conditions.

However, it is important to acknowledge certain limitations and areas for improvement in our study. Firstly, to increase the generalizability of the experimental results, it would be beneficial to expand the participant pool and include a larger sample size. Additionally, future research should consider including actual patients who are unable to communicate verbally, in order to better align with the intended target population.

Overall, the classification of speech imagery holds promise in the field of communication and related disciplines. To fully realize the potential of this technology, further research is needed to address the aforementioned considerations and explore its broader applications in communication and related fields.

This paper demonstrated an extension of the preliminary study on decoding speech imagery from EEG signals, with the aim of building a high-performance decoding model specialized for sentence decoding. Therefore, the study focused on confirming the feasibility of decoding long sentences using speech imagery based on machine learning and deep learning techniques.

In future work, there are several implications for future research and clinical practice stemming from

our study. For example, future research could investigate the applicability of our method to different populations with communication impairments, such as patients with neurological disorders or brain injuries. Additionally, integrating EEG-based speech imagery decoding with other neurorehabilitation interventions, like transcranial magnetic stimulation or cognitive training, may enhance the efficacy of these interventions and promote functional recovery in individuals with communication impairments. To improve the classification accuracy of speech imagery, it is necessary to calibrate EEG with cross-session data and develop a subject-independent classification model capable of achieving consistent accuracy. These efforts will help to address the challenge of achieving consistent and accurate classification of speech imagery using EEG. In addition, by applying advanced deep learning techniques, the proposed approach seeks to develop an accurate EEG-based BCI that can effectively decode speech imagery for sentences to individuals with speech impairments. The development of non-invasive and effective assistive technologies for communication is crucial in improving the quality of life and social participation of individuals, and this proposed method provides a promising approach to achieving this goal. The study highlights the potential for further research in this area and underscores the importance of continuing to develop and improve EEG-based brain-computer interfaces for bidirectional communication rehabilitation.

References

- [1] J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-biLSTM network using EEG signals", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 28, No. 5, pp. 1226-1238, May 2020. <https://doi.org/10.1109/tnsre.2020.2981659>.

- [2] J.-H. Jeong, N.-S. Kwak, C. Guan, and S.-W. Lee, "Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 28, No. 3, pp. 687-698, Mar. 2020. <https://doi.org/10.1109/tnsre.2020.2966826>.
- [3] E. M. Holz, L. Botrel, T. Kaufmann, and A. Kübler, "Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study", *Archives of Physical Medicine and Rehabilitation*, Vol. 96, No. 3, pp. S16-26, Mar. 2015. <https://doi.org/10.1016/j.apmr.2014.03.035>.
- [4] J.-H. Jeong, K.-T. Kim, D.-J. Kim, and S.-W. Lee, "Decoding of multi-directional reaching movements for EEG-based robot arm control", 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, pp. 511-514, Oct. 2018. <https://doi.org/10.1109/smc.2018.00096>.
- [5] I. Käthner, S. Halder, and A. Kübler, "Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: a case study with a participant in the locked-in state", *Journal of NeuroEngineering and Rehabilitation*, Vol. 12, No. 1, pp. 1-13, Sep. 2015. <https://doi.org/10.1186/s12984-015-0071-z>.
- [6] J.-H. Jeong, J.-H. Cho, B.-H. Lee, and S.-W. Lee, "Real-time deep neurolinguistic learning enhances noninvasive neural language decoding for brain-machine interaction", *IEEE Transactions on Cybernetics*, pp. 1-14, Oct. 2022. <https://doi.org/10.1109/tcyb.2022.3211694>.
- [7] S.-H. Lee, M. Lee, J.-H. Jeong, and S.-W. Lee, "Towards an EEG-based Intuitive BCI Communication System Using Imagined Speech and Visual Imagery", 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, pp. 4409-4414, Oct. 2019. <https://doi.org/10.1109/SMC.2019.8914645>.
- [8] F. Nijboer, N. Birbaumer, and A. Kübler, "The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis—a longitudinal study", *Frontiers in Neuroscience*, Vol. 4, p. 55, Jul. 2010. <https://doi.org/10.3389/fnins.2010.00055>.
- [9] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns", *Neural Networks*, Vol. 22, No. 9, pp. 1334-1339, Nov. 2009. <https://doi.org/10.1016/j.neunet.2009.05.008>.
- [10] S.-H. Lee, M. Lee, J.-H. Jeong, and S.-W. Lee, "Towards an EEG-based intuitive BCI communication system using imagined speech and visual imagery", 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, pp. 4409-4414, Oct. 2019. <https://doi.org/10.1109/smc.2019.8914645>.
- [11] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control", *Clinical Neurophysiology*, Vol. 113, No. 6, pp. 767-791, Jun. 2002. [https://doi.org/10.1016/S1388-2457\(02\)00057-3](https://doi.org/10.1016/S1388-2457(02)00057-3).
- [12] G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, and J. C. Kline, "Silent speech recognition as an alternative communication device for persons with laryngectomy", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, No. 12, pp. 2386-2398, Dec. 2017. <https://doi.org/10.1109/taslp.2017.2740000>.
- [13] J. Choi, K.-T. Kim, J.-H. Jeong, L. Kim, S.-J. Lee, and H. Kim, "Developing a motor imagery-based real-time asynchronous hybrid BCI controller for a lower-limb exoskeleton", *Sensors*, Vol. 20, No. 24, Dec. 2020. <https://doi.org/10.3390/s20247309>.
- [14] M. D'Zmura, S. Deng, T. Lappas, S. G. Thorpe, and R. Srinivasan, "Toward EEG sensing of

- imagined speech", *Human-Computer Interaction. New Trends*, Vol. 5610, pp. 40-48, 2009. https://doi.org/10.1007/978-3-642-02574-7_5.
- [15] E. C. Leuthardt, C. Gaona, M. Sharma, N. Szrama, J. Roland, Z. Freudenberg, J. Solis, J. Breshears, and G. Schalk, "Using the electrocorticographic speech network to control a brain-computer interface in humans", *Journal of Neural Engineering*, Vol. 8, No. 3, Apr. 2011. <https://doi.org/10.1088/1741-2560/8/3/036004>.
- [16] S. Martin, P. Brunner, C. Holdgraf, H.-J. Heinze, N. E. Crone, J. Rieger, G. Schalk, R. T. Knight, and B. N. Pasley, "Decoding spectrotemporal features of overt and covert speech from the human cortex", *Front. Neuroeng.*, Vol. 7, May 2014. <https://doi.org/10.3389/fneng.2014.00014>.
- [17] S. Iqbal, Y. U. Khan, and O. Farooq, "EEG based classification of imagined vowel sounds", 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 1591-1594, May 2015.
- [18] S. Iqbal, P. P. M. Shanir, Y. U. Khan, and O. Farooq, "Time domain analysis of EEG to classify imagined speech", *Proc. of the Second International Conference on Computer and Communication Technologies*, Vol. 380, pp. 793-800, Sep. 2015. https://doi.org/10.1007/978-81-322-2523-2_77.
- [19] N. Yoshimura, A. Nishimoto, A. N. Belkacem, D. Shin, H. Kambara, T. Hanakawa, and Y. Koike, "Decoding of covert vowel articulation using electroencephalography cortical currents", *Frontiers in Neuroscience*, Vol. 10, May 2016. <https://doi.org/10.3389/fnins.2016.00175>.
- [20] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features", *Journal of Neural Engineering*, Vol. 15, No. 1, 2018. <https://doi.org/10.1088/1741-2552/aa8235>.
- [21] N. Hashim, A. Ali, and W.-N. MohdIsa, "Word-based classification of imagined speech using EEG", *ICCST 2017: Computational Science and Technology*, Kuala Lumpur, Malaysia, Vol. 488, pp. 195-204, Feb. 2018. https://doi.org/10.1007/978-981-10-8276-4_19.
- [22] D.-Y. Lee, M. Lee, and S.-W. Lee, "Decoding imagined speech based on deep metric learning for intuitive BCI communication", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 29, pp. 1363-1374, Jul. 2021. <https://doi.org/10.1109/tnsre.2021.3096874>.
- [23] D.-H. Lee, J.-H. Jeong, H.-J. Ahn, and S.-W. Lee, "Design of an EEG-based drone swarm control system using endogenous BCI paradigms", 2021 9th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea, pp. 1-5, Feb. 2021. <https://doi.org/10.1109/bci51272.2021.9385356>.
- [24] C.-H. Han, Y.-W. Kim, D.-Y. Kim, S.-H. Kim, Z. Nenadic, and C.-H. Im, "Electroencephalography-based endogenous brain-computer interface for online communication with a completely locked-in patient", *Journal of NeuroEngineering and Rehabilitation*, No. 18, pp. 1-13, Jan. 2019. <https://doi.org/10.1186/s12984-019-0493-0>.
- [25] H.-y. Li, Q.-h. Zhao, G.-l. Ren, and B.-j. Xiao, "Speech Enhancement Algorithm Based on Independent Component Analysis", 2009 Fifth International Conference on Natural Computation, Tianjian, China, pp. 598-602, Aug. 2009. <https://doi.org/10.1109/ICNC.2009.76>.
- [26] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp. 929-941, Jun. 2005. <https://doi.org/10.1109/TPAMI.2005.110>.
- [27] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, Vol. 20, pp. 273-297, Sep. 1995. <https://doi.org/10.1007/bf00994018>.

- [28] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, Nov. 1997. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, Vol. 521, pp. 436-444, May 2015. <https://doi.org/10.1038/nature14539>.
- [30] G. Pereyra, G. Tucker, J. Chorowski, Ł. Kaiser, and G. Hinton, "Regularizing neural networks by penalizing confident output distributions", *arXiv:1701.06548*, Jan. 2017. <https://doi.org/10.48550/arXiv.1701.06548>.
- [31] L. Deng and D. Yu, "Deep learning: methods and applications", *Foundations and Trends® in Signal Processing*, Vol. 7, No. 3-4, pp. 197-387, Jun. 2014. <https://doi.org/10.1561/20000000039>.
- [32] J. Zhang, C. Li, and T. Zhang, "Application of back-propagation neural network in the post-blast re-entry time prediction", *Knowledge*, Vol. 3, No. 2, pp. 128-148, Mar. 2023. <https://doi.org/10.3390/knowledge3020010>.
- [33] D. S. Weisberg, F. C. Keil, J. Goodstein, E. Rawson, and J. R. Gray, "The Seductive Allure of Neuroscience Explanations", *Journal of Cognitive Neuroscience*, Vol. 20, No. 3, pp. 470-477, Mar. 2008. <https://doi.org/10.1162/jocn.2008.20040>.

Authors

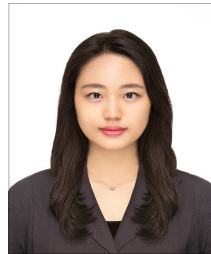
Seong-Hyun Yu



2017 ~ 2023 : B.S. degree in
School of Computer Science,
Chungbuk National University
2023 ~ present : Integrated
M.S.&Ph.D. degree in Dept.
Computer Science, Chungbuk
National University

Research interests : Machine Learning, Brain-machine
Interface, and Artificial Intelligence

Hyeong-Yeong Park



2018 ~ present : B.S. degree in
School of Computer Science,
Chungbuk National University
Research interests : Machine
Learning, Brain-machine
Interface, and Artificial
Intelligence

Seo-Jin Lee



2019 ~ present : B.S. degree in
School of Computer Science,
Chungbuk National University
Research interests : Machine
Learning, Artificial Intelligence,
Computer Vision

Ji-Hoon Jeong



2009 ~ 2015 : B.S. degree in
Department of Computer
Science & Dept. Brain and
Cognitive Sciences, Korea
University
2015 ~ 2021 : Ph.D. degree in
Dept. Brain and Cognitive

Engineering, Korea University
2022 ~ present : Assistant Professor, School of
Computer Science, Chungbuk National University
Research interests : Machine Learning, Brain-machine
Interface, and Artificial Intelligence