Speech Imagery Decoding for Emotional Expression by Exploring Visual Perception from EEG Signals

Hyeong-Yeong Park
Dept. of Computer Science,
Chungbuk National University,
Cheongju, Republic of Korea
hyeong.y.park@chungbuk.ac.kr

Seong-Hyun Yu
Dept. of Computer Science,
Chungbuk National University,
Cheongju, Republic of Korea
sh.yu@chungbuk.ac.kr

Seo-Jin Lee
Dept. of Computer Science,
Chungbuk National University,
Cheongju, Republic of Korea
sj.lee@chungbuk.ac.kr

Ji-Hoon Jeong*
Dept. of Computer Science,
Chungbuk National University,
Cheongju, Republic of Korea
jh.jeong@chungbuk.ac.kr

***Abstract*—This study investigates the impact of visual perception, specifically emojis, on speech imagery decoding in brain-computer interfaces for emotional representation. Nine participants took part in two sessions: a speech imagery session and speech imagery with visual perception session. The results show that deep learning, specifically EEGNet, achieved the highest accuracy in the speech imagery session. Interestingly, the inclusion of visual perception was found to reduce the user's focus on expressing emotions through speech imagery. These findings support the notion that speech imagery decoding provides a more intuitive approach to natural emotional expression and user intention communication.**

***Index Terms*—Brain-computer interface, electroencephalogram, speech imagery, emotional expression**

I. Introduction

In recent years, brain-computer interfaces (BCIs) have gained prominence as a means of human-machine communication. Speech imagery, specifically in relation to emotions, has emerged as a promising modality within the field of BCIs [1], [2]. This study aims to explore the connection between speech imagery, emotion recognition, and visual perception in BCI technology, with the goal of advancing intuitive communication tools for individuals with motor impairments or speech disorders.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00252624).

*J.-H. Jeong is the corresponding author.

II. Materials and Methods

A. Experiments

Nine healthy subjects (7 males, 2 females), aged 23 on average, participated in the experiment. The study protocol and paradigm were explained to the participants, who provided written informed consent in accordance with the Declaration of Helsinki (approved by KUIRB-2019-0143-01).

The experiment consisted of two sessions: Session I and Session II. In Session I, subjects engaged in speech imagery for a given word without the use of emoji images. In session II, subjects performed speech imagery for the given word while presented with emoji images for visual perception.

TABLE I
BINARY CLASSIFICATION RESULTS FOR EACH CLASS IN EACH SESSION, CATEGORIZED BY MODEL

Classes	Session	ML/DL Models		
		CSP+SVM	CSP+LDA	EEGNet
'Happy'	I	0.6522	0.6468	0.8778
	II	0.6111	0.6000	0.7694
'Hurt'	I	0.6702	0.6479	0.8653
	II	0.6113	0.5760	0.7611
'Angry'	I	0.6827	0.6567	0.8653
	II	0.6142	0.6010	0.8069
'Sad'	I	0.7043	0.6889	0.8708
	II	0.6069	0.5891	0.8042
Average	I	0.6774 ($\pm$0.0190)	0.6600 ($\pm$0.0171)	**0.8715** ($\pm$0.0044)
	II	0.6109 ($\pm$0.0026)	0.5915 ($\pm$0.0101)	**0.7854** ($\pm$0.0204)

B. Data Analysis and Evaluation

The EEG data were pre-processed using band-pass filtering in the frequency range of 30-130 Hz, specifically targeting the Beta and Gamma bands. To decode speech imagery related to emotional expression, machine learning models such as SVM and RLDA were used. Feature extraction was performed using the CSP algorithm. EEGNet, a deep learning technique that combines components of a convolutional neural network, was also used to capture temporal and spatial information in EEG signals. Model performance was validated using a 10-fold cross-validation method. Binary classification tasks were performed based on a 3-second rest state and 2-second segments representing each emotion class ('Happy', 'Hurt', 'Angry', and 'Sad'). Four binary classifications were conducted in each session for evaluation purposes.

III. RESULTS AND DISCUSSION

Table 1 shows that EEGNet achieved higher accuracy than machine learning methods in all sessions and emotion classes. In session I, EEGNet had the highest accuracy of 0.871, while LDA had the lowest accuracy of 0.6600. In session II, EEGNet achieved an accuracy of 0.7854, while LDA had an accuracy of 0.5915. Contrary to expectations, visual perception, represented by emojis, appeared to hinder speech imagery decoding performance.

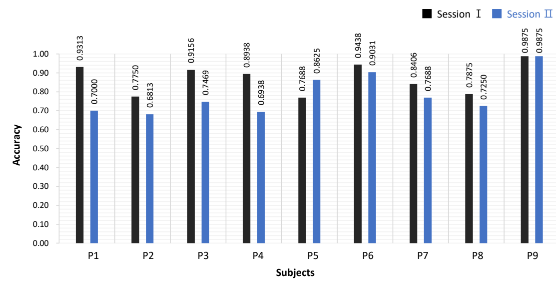

Fig. 1. Graph representing the average binary classification results for each emotion using EEGNet across the subjects.

Fig. 1 displays the average classification results for each emotion using EEGNet. Most of the participants performed better in session I, except for subject P5. Subject P9 achieved the highest performance in both sessions with a score of 0.9875. Subject P2 had the lowest performance with accuracies of 0.7750 and 0.6813 in the two sessions, respectively. However, these accuracies were still significantly above the chance level. Subject P1 showed the largest discrepancy in performance between sessions, with a difference of 0.2313, indicating a negative impact of visual perception on decoding. On the contrary, subject P5 performed better in session II with a difference of 0.0937.

IV. CONCLUSION

The study highlights the effectiveness of deep learning techniques in decoding speech imagery for emotions, outperforming machine learning methods. Inclusion of visual perception using emoji imagery was found to improve classification performance for some participants, while its impact varied between individuals. These findings contribute to our understanding of the relationship between speech imagery, emotion, and visual perception, with potential implications for practical BCI.

REFERENCES

[1] 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 Trans. Cybern.*, 2022.

[2] D.-Y. Lee, M. Lee, and S.-W. Lee, "Decoding imagined speech based on deep metric learning for intuitive BCI communication," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 29, pp. 1363–1374, 2021.