Investigation on Effect of Speech Imagery EEG Data Augmentation with Actual Speech

Jaehoon Choi
School of Computing
KAIST
Daejeon, South Korea
basedseal@gmail.com

Netiwit Kaongoen
School of Computing
KAIST
Daejeon, South Korea
ghiejo10jo@kaist.ac.kr

Sungho Jo
School of Computing
KAIST
Daejeon, South Korea
shjo@kaist.ac.kr

2022 10th International Winter Conference on Brain-Computer Interface (BCI) | 978-1-6654-1337-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/BCI53720.2022.9735108

***Abstract*—Speech imagery based brain-computer interfaces (BCI) have been researched as a potentially powerful method to provide communication and control without explicit bodily movements. Compared to other BCI systems, speech imagery based BCIs can employ mental tasks directly related to control tasks, making them more intuitive for use in daily-life. However, training data is difficult to collect in sufficient size and number in terms of time and user comfort; preparation time to wear electroencephalogram (EEG) measurement devices are lengthy and the devices themselves are uncomfortable to use, making it difficult for subjects to undergo long training sessions without feeling fatigued. To overcome this problem, we suggest a novel data augmentation method using actual speech data. Based on the similarity between the tasks imagined speech and actual speech, we augment EEG training data of imagined speech with EEG data of actual speech and show significant improvements for two out of three subjects.**

***Index Terms*—Brain-computer interface (BCI), Speech-imagery, Actual speech, data augmentation**

I. Introduction

Brain-computer interfaces (BCI) have been proposed as alternative mode of control, especially for patients suffering from locked-in syndrome (LIS) as BCIs do not require bodily movements to provide control and instead work by identifying specific brain patterns that are then mapped to desired functions [1], [2]. BCIs are commonly categorized into two groups: active and reactive BCIs. Reactive BCIs refer to systems that learns brain activities that emerge in response to a fixed stimulus [3], [4]. While easy to train, reactive BCIs are limited by the requirement of a stimulus, making it tiring to use for long time [5]. Active BCIs, on the other hand, are interfaces that detect brain patterns that occurs from an user imagining an action. This can be moving a body part, a specific image, or speaking out a word [6]–[9]. Compared to reactive BCIs, active BCIs are less restrictive, but harder for the user to train [5].

Speech imagery is an active BCI that measures electroencephalogram (EEG) while the subject imagines a word without actually speaking out or moving articulators [9]. Speech imagery based BCIs have several distinct advantages over other types of BCIs. First, speech imagery tasks are more friendly to the users as it is very similar to speaking inside one's head, a common activity in daily life. Unlike other active BCIs like motor imagery or visual imagery, users do not have to spend time training to make sure they are performing BCI task correctly. Another benefit is the flexibility of task design. As speech imagery is based on speech, any task can be mapped to a word with related meaning, making the system much more intuitive (e.g. word "power" for turning a television on or word "forward" to move a wheelchair ahead).

Due to these advantages, several studies regarding speech imagery have been carried out using both electrocorticography (ECoG) and EEG [9]–[13]. In a previous research, we investigated the efficiency of using ear-EEG device to measure speech imagery as a starting point for a BCI system that can be used in daily life [12].

One difficulty that plagues BCI systems is data collection. In order to collect EEG, users often have to stay still wearing uncomfortable EEG measurement device, making it difficult to carry out long experimental sessions. Due to this, the size of collected data per subject is much smaller compared to other fields. Due to this, various augmentation methods have been employed for BCI systems [14]–[17]. However, due to temporal nature of speech imagery features [12], these may not be suitable when applied to speech imagery. In this paper, we suggest that EEG collected during actual speech can be used in tandem with speech imagery data as a novel form of data augmentation specific for speech imagery. Actual speech data is easier to label compared to imagery, and a system for acquiring actual speech imagery during daily-life with auto-labelling based on audio recording may be an effective way to collect large amount of EEG data. Using actual speech EEG data collected along with speech imagery EEG data, we compare classification performance when using only speech imagery data with when using training data augmented with actual speech data for three subjects and show that two of the subjects show significant improvements.

II. Methods

A. Data Collection Setup

Data was collected using BrainVision actiCHamp amplifier with BrainVision Recorder software at sampling rate of 500Hz using 32 electrodes arranged as shown in figure 1. Electrodes

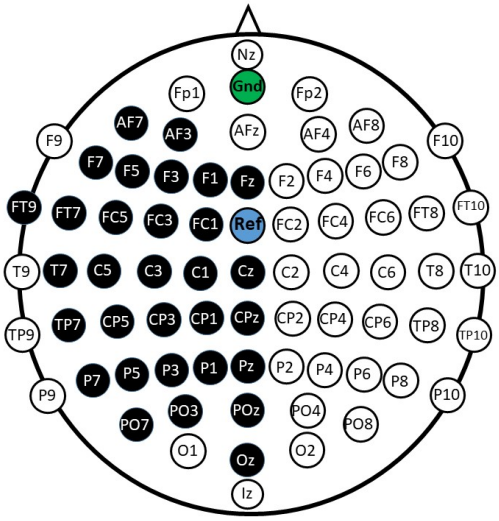

Fig. 2. Electrode layout for data collection

were mainly arranged around the left cerebral hemisphere due to regions related to speech production and comprehension located in this area. Fpz and FCz were chosen as ground and reference channels respectively. All experiments were carried out in an air-conditioned sound-proof room with instructions given to subjects to minimize movement during data recording.

B. Experimental Subjects

Three male subjects, aged between 25 and 30 participated in the experiment. All subjects were healthy with normal or corrected-to-normal vision. All subjects were fluent in English and had some prior experience using brain computer interfaces.

C. Experimental Protocol

Speech imagery for four different words (”left”, ”right”, ”go back”, ”forward”) and rest class were collected following experimental protocol based on our previous study [12]. One experimental trial for a class is shown in figure 2. First, an audio cue of the class spoken in a female American accent is given for two seconds to start. Then, a fixation cross is shown for one second. After that, a circular cue is given for two seconds, during which subjects were instructed to speak out the given word. Subjects are given one second to rest before starting speech imagery. A loading bar is shown during which subjects are instructed to imagine speaking out the target class for two seconds. This is repeated five times in a row, with one second of rest in between. After this, a resting period of 3.5 seconds is given before the next trial.

An experimental block consists on five trials for each class. The order of the trials in a block is randomized. One experimental session contained ten experimental blocks, resulting in 50 trials in total. Each subject carried out six sessions over three different days, two sessions per day. Subjects were given sufficient time to rest in between the sessions.

D. Data Processing

For preprocessing, we apply a notch filter at 60Hz to remove line noise. We then apply a bandpass filter with cutoff frequencies at 0.1 and 100Hz. Both speech imagery data and actual speech data are acquired by selecting two seconds of epoch starting from when the visual cue for actual speech and speech imagery are shown.

We use two different features for speech imagery classification. First is common spatial pattern (CSP) [18], which is frequently used to identify useful spatial features. CSP features are acquired using spatial filters obtained by optimizing the following equation, where w is the filter to be obtained.

$$\mathrm{J(w)} = \frac{w^T X_1 X_1^T w}{w^T X_2 X_2^T w} = \frac{w^T C_1^T w}{w^T C_2^T w}$$

X_1 represents EEG signals corresponding to resting state and X_2 signals corresponding to a speech imagery class. C_1 and C_2 similarly represents spatial covariance matrix for resting and speech imagery class respectively, assuming zero mean for EEG signals.

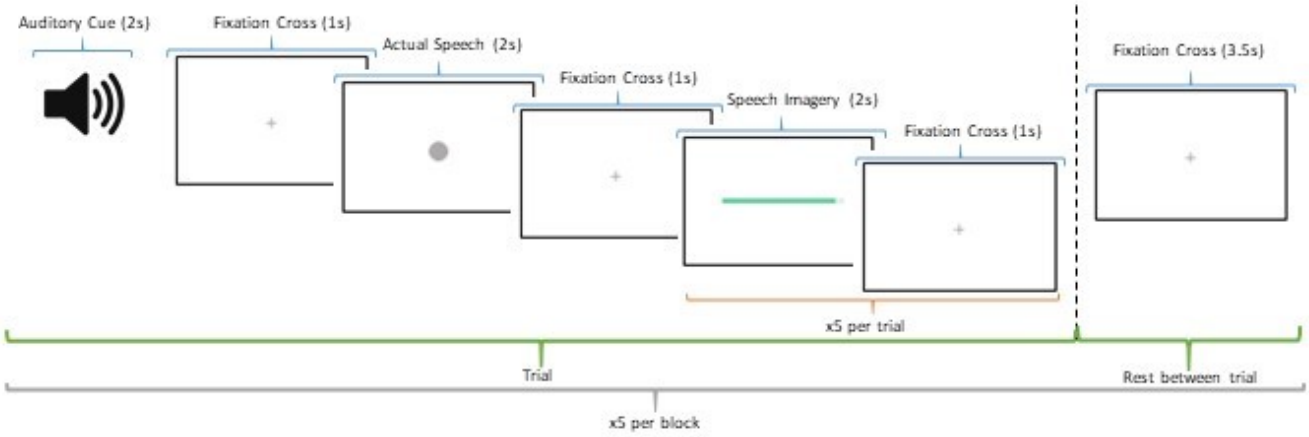

Fig. 1. Overview of Single Trial

The second type of features extracted are Riemannian tangent space vectors [19]. We first calculate covariance matrix for each EEG epoch. Then, a Riemannian tangent space vector is calculated from each covariance matrix using the following equation:

$$\mathrm{v_i} = \mathrm{upper}(\mathrm{C_R^{-\frac{1}{2}}} \log_{C_R}(C_i) C_R^{-\frac{1}{2}})$$

where $\mathrm{C_i}$ is the covariance matrix, $\mathrm{C_R}$ the Riemannian mean of covariance matrix and $\mathrm{v_i}$ the Riemannian tangent space vector.

We carry out classification using a support vector machine with linear kernel for five classes (four speech imagery classes and rest) using two different features and compare their performances. We then carry out classification using two different types of training data. For the first case, we carry out 5-fold cross validation using only speech imagery data. For the other case, we first split the speech imagery data into train and test set as in previous 5-fold cross validation. Before training the model, we augment the training data set with actual speech data, and carry out evaluation on the test data set, which contains only speech imagery data.

III. Results and Discussion

Table 1 shows the classification results when applying different feature extraction methods to only speech imagery data and when training data is augmented with actual speech data.

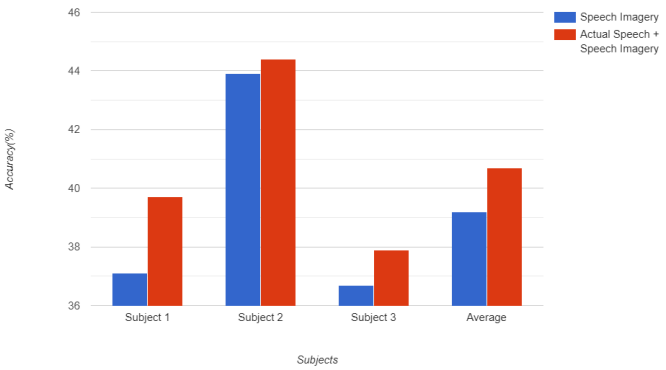

Fig. 3. Accuracy using tangent space vectors with only speech imagery and with both speech imagery and actual speech

When using only speech imagery as training data, Subject 3 shows the best average performance using CSP features at 30.6% and Subject 2 shows the best average performance using tangent space vectors with accuracy of 43.9%. All three subjects show significantly improved performance when using tangent space vectors at 0.05 confidence level.

When using training data augmented with actual speech EEG, Subject 3 likewise shows the best average performance using CSP features at 29.5% and Subject 2 using tangent space vectors with accuracy of 44.4%. All three subjects again show significantly improved performance with tangent space vectors as features at 0.05 confidence level.

When comparing the effect of data augmentation, with CSP features, only Subject 1 shows slight improvement in performance, with all three subjects showing no significant differences. However, when using tangent space vectors, all three subjects show improvements in both average and maximum accuracy. Subject 1 and Subject 3 show significant increase in accuracy at 0.05 confidence level.

To examine the effect of augmentation, we examined the confusion matrix for subject 1, who showed the most significant improvement when augmentation method was applied. This is shown in figure 4. All four speech imagery classes showed improvements in classification with actual speech augmentation, with greatest improvement shown in "left" class. Although rest class showed lower performance after augmentation, the true positive rate instead increased with augmentation applied. An interesting point to note is that "left" class is also the class with the lowest accuracy before augmentation; in general, classes with lower accuracy showed greater increase in performance, suggesting our proposed method improves performances for classes with lower accuracy more effectively.

IV. Conclusion and Future Works

In this paper, we collected actual speech data together with speech imagery data (four speech imagery words + rest class) using a 32 EEG channel system arranged around the left cerebral hemisphere from three subjects. We classified the collected data using two different features, CSP and tangent space vectors, under two different conditions: 1)using only speech imagery data as training data 2)using actual speech data together with speech imagery data during training. Our results show significant improvement when using both actual

TABLE I
Comparison of mean ± std and min ÷ max of the accuracy (%) of all subjects for different methods

	Speech Imagery Only		Speech Imagery with Actual Speech	
	CSP+SVM	TS+SVM	CSP+SVM	TS+SVM
Subject 1	24.3±4.2 18.8÷30.0	37.1±7.2 30.8÷50.4	25.8±5.4 17.2÷32.0	39.7±7.4 31.6÷51.4
Subject 2	27.5±2.8 24.8÷31.2	43.9±3.7 39.6÷48.0	27.0±3.3 21.6÷30.4	44.4±5.3 37.2÷50.4
Subject 3	30.6±7.1 20.8÷40.8	36.7±7.8 21.6÷42.4	29.5±6.9 18.4÷37.6	37.9±7.8 22.8÷44.4

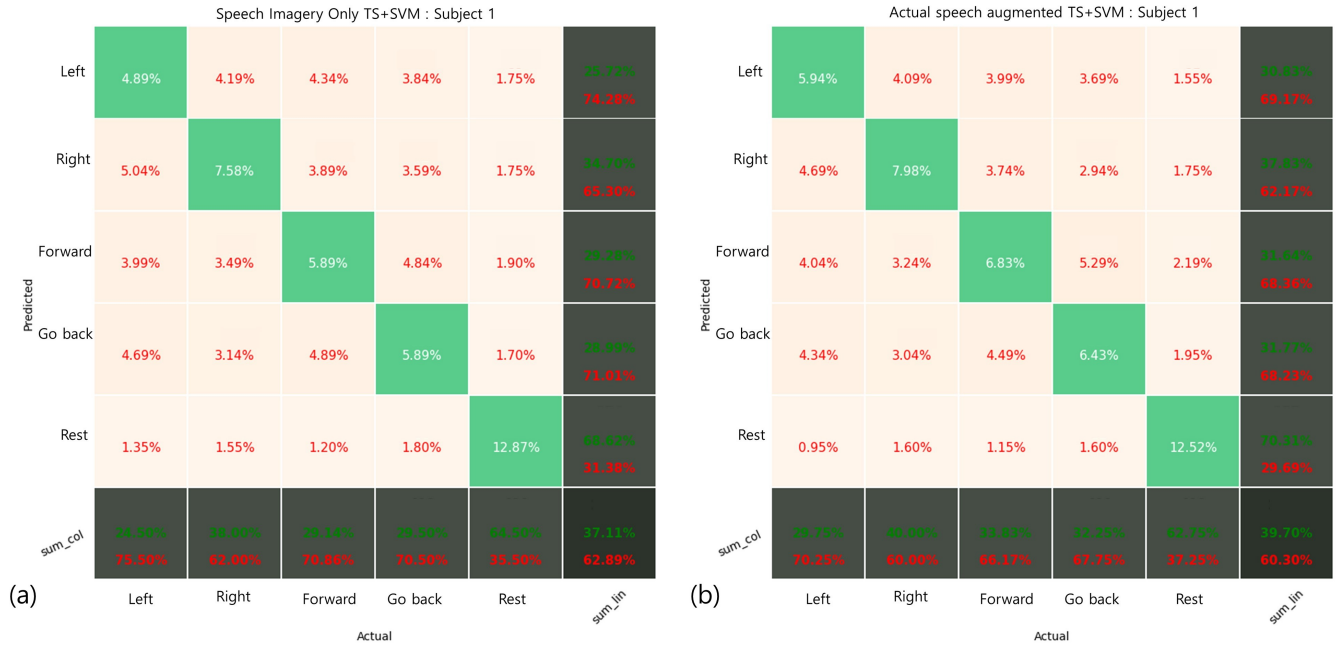

Fig. 4. Confusion matrix of Subject 1 using tangent space vectors with (a) only speech imagery and with (b) both speech imagery and actual speech

speech and speech imagery data for training for two subjects with tangent space vector features, suggesting that using actual speech data along with speech imagery may be an effective method to augment training data. Since the number of subjects recruited for this work is not enough to make a strong conclusion, future work will involve recruiting more participants. We will also carry out more detailed comparison with other augmentation methods, as well as different features extraction and classification algorithms. Finally, we plan to design a auto-labelling actual speech EEG collection system to collect data for augmentation easily while carrying out daily life activities.

V. Acknowledgement

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) under Grant 2017-0-00432, and the Defense Challengeable Future Technology Program of Agency for Defense Development, Republic of Korea.

References

[1] Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain–computer interface using electrocorticographic signals in humans. Journal of neural engineering. 2004 Jun 14;1(2):63.

[2] Pandarinath C, Nuyujukian P, Blabe CH, Sorice BL, Saab J, Willett FR, Hochberg LR, Shenoy KV, Henderson JM. High performance communication by people with paralysis using an intracortical brain-computer interface. Elife. 2017 Feb 21;6:e18554.

[3] Kaongoen N, Yu M, Jo S. Two-factor authentication system using p300 response to a sequence of human photographs. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017 Oct 4;50(3):1178-85.

[4] Luo A, Sullivan TJ. A user-friendly SSVEP-based brain–computer interface using a time-domain classifier. Journal of neural engineering. 2010 Mar 23;7(2):026010.

[5] Douibi K, Le Bars S, Lemontey A, Nag L, Balp R, Breda G. Toward EEG-based BCI applications for industry 4.0: challenges and possible applications. Frontiers in Human Neuroscience. 2021:456.

[6] Choi JW, Kim BH, Huh S, Jo S. Observing actions through immersive virtual reality enhances motor imagery training. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020 May 28;28(7):1614-22.

[7] Choi JW, Huh S, Jo S. Improving performance in motor imagery BCI-based control applications via virtually embodied feedback. Computers in Biology and Medicine. 2020 Dec 1;127:104079.

[8] Kosmyna N, Lindgren JT, Lécuyer A. Attending to visual stimuli versus performing visual imagery as a control strategy for EEG-based brain-computer interfaces. Scientific reports. 2018 Sep 5;8(1):1-4.

[9] Wang L, Zhang X, Zhong X, Zhang Y. Analysis and classification of speech imagery EEG for BCI. Biomedical signal processing and control. 2013 Nov 1;8(6):901-8.

[10] Nguyen CH, Karavas GK, Artemiadis P. Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features. Journal of neural engineering. 2017 Dec 1;15(1):016002.

[11] DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. Neural networks. 2009 Nov 1;22(9):1334-9.

[12] Kaongoen N, Choi J, Jo S. Speech-imagery-based brain–computer interface system using ear-EEG. Journal of neural engineering. 2021 Feb 23;18(1):016023.

[13] Martin S, Brunner P, Iturrate I, Millán JD, Schalk G, Knight RT, Pasley BN. Word pair classification during imagined speech using direct brain recordings. Scientific reports. 2016 May 11;6(1):1-2.

[14] Gubert PH, Costa MH, Silva CD, Trofino-Neto A. The performance impact of data augmentation in CSP-based motor-imagery systems for

BCI applications. Biomedical Signal Processing and Control. 2020 Sep 1;62:102152.

[15] Lee T, Kim M, Kim SP. Data augmentation effects using borderline-SMOTE on classification of a P300-based BCI. In2020 8th International Winter Conference on Brain-Computer Interface (BCI) 2020 Feb 26 (pp. 1-4). IEEE.

[16] Nagasawa T, Sato T, Nambu I, Wada Y. Improving fNIRS-BCI accuracy using GAN-based data augmentation. In2019 9th International IEEE/EMBS Conference on Neural Engineering (NER) 2019 Mar 20 (pp. 1208-1211). IEEE.

[17] Kalunga E, Chevallier S, Barthélemy Q. Data augmentation in Riemannian space for brain-computer interfaces. InSTAMLINS 2015 Jun 10.

[18] Ang KK, Chin ZY, Zhang H, Guan C. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence) 2008 Jun 1 (pp. 2390-2397). IEEE.

[19] Gaur P, Pachori RB, Wang H, Prasad G. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. Expert Systems with Applications. 2018 Apr 1;95:201-11.