2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME) | 978-1-6654-1955-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICBME51989.2020.9319439

Imagined Speech Decoding From EEG: The Winner of 3rd Iranian BCI Competition (iBCIC2020)

Nastaran Hamedi†
Department of Biomedical Engineering
K. N. Toosi University of Technology
Tehran, Iran
nas.hamedi@email.kntu.ac.ir

Susan Samiei†
Department of Biomedical Engineering
K. N. Toosi University of Technology
Tehran, Iran
s.samiei@email.kntu.ac.ir

Mehdi Delrobaei
Department of Biomedical Engineering
K. N. Toosi University of Technology
Tehran, Iran
delrobaei@kntu.ac.ir

Ali Khadem*
Department of Biomedical Engineering
K. N. Toosi University of Technology
Tehran, Iran
alikhadem@kntu.ac.ir

†indicates co-first authorship

Abstract— **Brain-computer interface (BCI) is defined as the combination of machine and brain signals to control a device or computer to improve the quality of life, e.g., for people with paralysis. In this paper, we focus on people with speech disorders and investigate the capability of electroencephalogram (EEG) signals to discriminate four classes, including the speech imagination of three Persian words corresponding to the English words "rock," "paper," and "scissors," in addition to the resting state. We used the data available from the 3rd Iranian BCI competition (iBCIC2020), acquired from 10 healthy participants in a randomized study. Initially, the mutual information (MI) was used to find the optimum frequency band. Then, features were extracted from the data using the Common Spatial Pattern (CSP) algorithm. Afterward, the most discriminative features were selected using the neighborhood component analysis (NCA). These features were then fed to a meta-classifier based on the stacking ensemble learning. The results show that working on an optimum frequency band will enhance the results compared with the fixed frequency band. It is also worth mentioning that the optimum frequency band is subject dependent; therefore, it is substantial to be selected accurately. Our method achieved an average classification accuracy of 51.90%±2.73 across all participants, which is promising compared with the results of previous studies in the field of imagined speech recognition in subject dependent BCI systems with randomized order of the stimuli.**

Keywords: ***Brain-computer interface (BCI); electroencephalogram (EEG); word imagery, common spatial patterns (CSP); neighborhood component analysis (NCA), stacking ensemble learning.***

I. Introduction

Speech is an essential requirement of everyday human life and the primary means of communication with the social community. Some mental disorders and brain injuries can cause partial or complete impairment in speech, permanently or temporarily. In such a situation, it would be precious to have a word recognition system predicting the word that the user wants to say but is unable to do. Moreover, for paralyzed people giving control commands to a system (e.g., a wheelchair) is more comfortable and secure by saying commands in mind. Inspired by these demands, imagined speech-based brain-computer interfaces (BCIs) was developed [1]. Imagined speech means a process in which the person tries to imagine speaking a word with no movement in articulatory muscle or producing audible sound [2]. Recent studies related to speech imagination have used various modalities to record brain signals. However, electroencephalogram (EEG)-based BCIs are more popular than other BCI systems due to their low cost, portability, and high temporal resolution. Some previous studies on the classification of word imagery using EEG are as follows.

In 2009, a study was conducted by the aim of examining the effect of unspoken word order on classification accuracy. In this research, researchers used a Hidden Markov Model (HMM) to classify EEG signals associated with five different words that were imagined by 21 healthy persons. The result showed that word order strongly affected the classification result [3]. Furthermore, a study was done on two words with eight healthy persons. Then feature vectors of their EEG signals extracted by common spatial patterns (CSP), were classified by support vector machine (SVM). The obtained classification accuracy was 67% [4]. After that, Rezazadeh *et al.* worked on the EEG signal of 12 healthy persons. They extracted the root-mean-square and standard deviation of the wavelet coefficients obtained from 4-level discrete wavelet decomposition as features and used a regularized neural network (RNN) to distinguish between two words. The achievement of this study showed that if all the signals recorded in just one session, it would have a better result compared with more than one session recordings [5]. Afterward, researchers conducted a word imagery experiment with three short words. The recorded data consisted of 22 blocks for each word. In one block, participants repeated one of the three words seven times in their minds. Later, by band-pass filtering the data to the frequency band of 8-70 Hz and extracting the Riemannian manifold features, followed by classification using a relevance vector machine, the average classification accuracy of 50.1% was obtained [6]. In 2019, Lee *et al.* increased the number of words to twelve. For each word, 22 blocks were captured, each block containing four trials of

978-1-6654-1955-0/20/$31.00 ©2020 IEEE

word imagination. CSP-based features were extracted from the trials filtered to the frequency band of 0.5-40 Hz, and classification was then performed using a random forest (RF) classifier. The average classification accuracy of 20.4% was achieved for 13 classes (including rest) across all participants [7].

According to previous studies, the structure of the experimental paradigm is highly influential on the classification results. Recording data in block mode leads to a higher recognition rate than the random mode. This higher recognition rate is merely due to the temporal correlation created in each block, not the inherent differences between the imagination of different words [3]. Besides, a word imagery-based BCI system in the real-life application must recognize words that the user imagines randomly; accordingly, recording data in block mode does not seem appropriate for real-life applications. Furthermore, studies reviewed above have filtered the data of all participants to a fixed frequency band. Since the individual characteristics of EEG signals are not the same among participants, filtering the signal to the optimal frequency band for each participant can increase the classification accuracy [8]. Moreover, stacking ensemble learning has rarely been used as a classifier in previous studies, and due to its discrimination power, we expect it to result in higher classification accuracy than traditional single classifiers.

In our study, we used optimum frequency band filtering, CSP filters, neighborhood component analysis (NCA), and stacking ensemble learning to classify a four-class word imagery problem, including three words plus resting state recorded in a random mode.

This paper is organized as follows. In section II, the EEG dataset used in the study is introduced. Afterward, in section III, we will propose our method. In section IV, the results obtained using our proposed method are presented. Finally, in section V, the results will be discussed, the paper will be concluded, and some future works will be proposed.

II. Dataset Description

In this work, we used the 3rd Iranian BCI competition (iBCIC2020) dataset, provided by the National Brain Mapping Laboratory (NBML), Tehran, Iran. The data is freely accessible for download [9].

The data was recorded from ten healthy native Persian participants (five women and five men; mean age: 31). The four classes in the dataset were speech imagination of three Persian words (/sæŋ/, /kɑːqæz/, and /qeɪtʃi/) corresponding to the English words "rock" (class 1), "paper" (class 2), "scissors" (class 3) respectively, and resting state (class 4). Twenty-five trials per class were recorded from each participant. Fig. 1 depicts the structure of a single trial of word imagery. First, a fixation cross was shown in the center of the monitor for 2 seconds. Participants were asked to relax their mind and wait for the visual cue during this time. Then, a random visual cue (rock, paper, or scissors) appeared. After 2 seconds, the visual cue disappeared, and the word "Go" appeared for 0.5 seconds. The participants only had 3 seconds to pronounce the word corresponding to the cue once in their mind. Afterward, the screen turned black for 2 seconds as a break. 3-second intervals containing word imagination (blue box) were extracted from trials. We used a total of 100 epochs (the length of each epoch was 3 seconds) to design and evaluate our model.

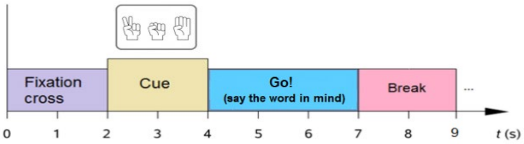

Figure 1. Structure of a single trial [9]. A 3-second interval containing word imagination (blue box) was used to train and evaluate the model.

Data was acquired using 64 EEG electrodes according to the 10-10 system and with a sampling frequency of 2400 Hz. Fpz position and right mastoid were chosen as ground and reference, respectively. A notch filter for suppressing 50 Hz power-line noise was enabled. Furthermore, the EEG signals were filtered by a band-pass filter (1-130Hz). Fig. 2 illustrates the location of EEG electrodes.

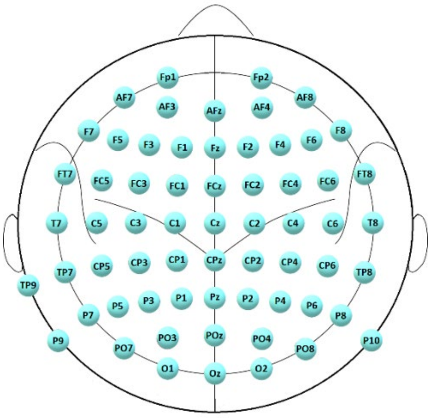

Figure 2. The location of EEG electrodes [10].

III. Proposed Method

Fig 3. Shows a summary of the proposed algorithm, which includes five steps: preprocessing (downsampling and filtering), feature extraction (CSP-based features), feature selection (NCA), classification (stacking ensemble learning), and evaluation (10-fold cross-validation). More details of each step are described below.

A. Preprocessing

The data of each participant was downsampled to 240 Hz and then filtered using a 3rd order Butterworth filter with the optimum frequency range. The mutual information (MI) was used to select the optimum frequency range of each participant [8]. The process of selecting the optimum frequency band is shown in Fig. 4. First, the training data is filtered using a filter bank (4-7 Hz, 5-8 Hz, …, 37-40 Hz). Next, the extracted features from filtered data are assessed by their MI with true class labels. If the value of MI corresponding to a frequency band is higher than the average of all MI values, the band is selected. The selected frequency ranges are merged if they are continuous. Finally, the most extensive frequency range is selected as the optimum frequency band. Afterward, the downsampled training and test data are re-filtered using a band-pass filter with the optimum frequency range.

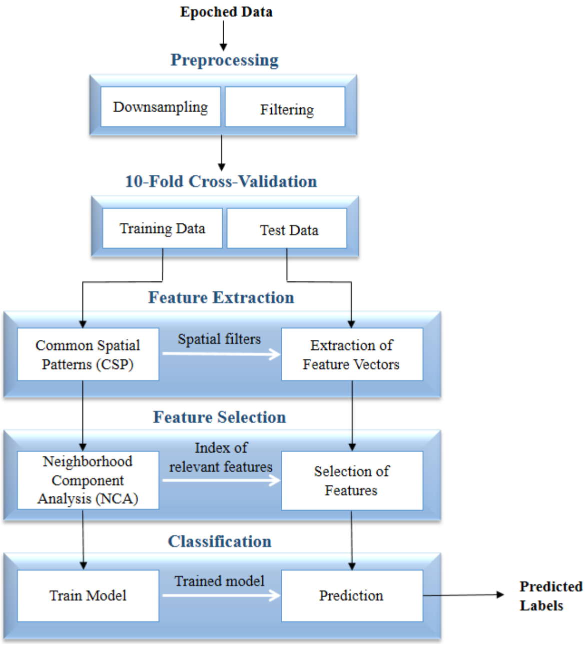

Figure 3. Flow chart of the proposed method

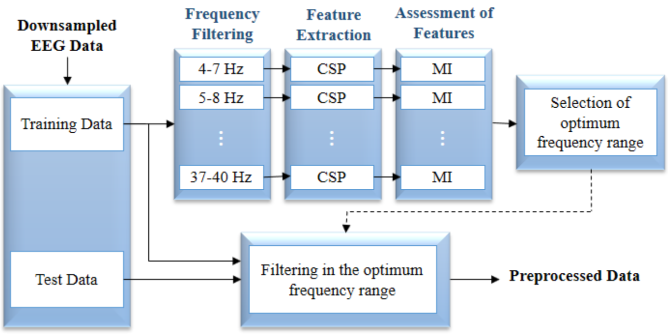

Figure 4. The block diagram of the optimum frequency band selection method

B. *Feature extraction*

In this study, common spatial filters (CSP) were used as a supervised algorithm for channel reduction and feature extraction. The principle idea behind these filters is to project the data into a low-dimensional spatial subspace where there is more distinction between classes [11]. This can be achieved by minimizing the cost function defined as in (1). In this equation, R_1 and R_2 are the normalized autocorrelation matrices of classes one and two, respectively. Also, w is a matrix containing all the spatial filters that can be obtained by solving this optimization problem. The projection matrix W will be composed by selecting a certain number of filters from the matrix w. Therefore, EEG signals based on (2) will be projected into a new spatial space (the new dimensional is determined to be p = 4 in our study), and the useful features will be extracted as in (2). E and Z denote each epoch and its projection in new spatial space, respectively. Finally, (3) calculates the variance of each projected epoch as feature vectors.

$$J(w) = \frac{w_T R_1 w}{w_T R_2 w} \qquad (1)$$

$$Z = WE \qquad (2)$$

$$f_p = \log\left(\frac{\mathrm{var}(Z_p)}{\sum_{i=1}^{4} \mathrm{var}(Z_i)}\right) \qquad p = 1, 2, 3, 4 \qquad (3)$$

To generalize the algorithm to multiclass paradigms, we used one versus one (OVO) and one versus all (OVA) approaches. Hence, 40 features were obtained for each epoch.

C. *Feature selection*

To eliminate redundant and irrelevant information, we used the NCA algorithm. As a nonparametric and supervised feature selection method, this technique aims to obtain a weight vector corresponding to the importance of features by minimizing the mean leave-one-out classification error across the training data with a regularization term. Cross-validation can be used to find the optimum value of the regularization parameter. Features with weights greater than the defined threshold are selected as the final features. For further details on this method, please refer to [12].

D. *Classification*

In order to distinguish between four different classes, we used stacking ensemble learning. It is a general ensemble learning method that combines different base classifiers to make a more accurate meta-classifier. In our study, the meta-classifier was made up of six base classifiers: Decision Tree, K-nearest neighbor (KNN), SVM with linear and Gaussian kernels, linear discriminant analysis (LDA), and Naive Bayes. In this case, the final accuracy is better than using a fixed base-classifier. Moreover, we used 10-fold cross-validation (CV) to evaluate our results. In this way, we considered 10% of all data in each fold as test data and the remaining as training data. The final performance of the method is calculated based on the average performance of 10 folds.

IV. Results

The proposed method was evaluated on a four-class word imagery data using 10-fold CV. The classification performance measures including, the accuracy of each class, the multiclass accuracy, and Cohen's kappa coefficient for each participant, as well as the total average accuracy among all participants, were calculated and provided in Table I. The optimum frequency band (in Hz) of each participant is also reported in the last column of the table. Our method attained an average accuracy of 51.90%±2.73 for four classes across all participants. The accuracy of each class for all participants is higher than the chance level of 25%. The best performance in each column is highlighted in bold.

V. Discussion and Conclusion

In this study, we proposed an approach to achieve a speech imagery-based BCI system of high classification performance. The results of the proposed method are reported in Table I. Due to the structural differences of the experimental paradigm between studies in the field of imagined speech, our results cannot be directly compared with the results of previous works. In most past studies, data recording has been done in block mode, meaning that in each block, the participant continuously repeats a word, or the corresponding cue of a word is displayed consecutively. Although this type of experimental paradigm leads to higher classification accuracy, it does not seem appropriate for real-life applications. As far as we know, there have been few word imagery EEG studies on randomized order of the stimuli, and their average recognition rates were 19.48% (five-class, chance level of 20%) [3] and 18.58% (six-class, chance level of 16.67%) [13]. In this study, we obtained average classification accuracy of 51.90% (four-class including resting state, chance level of 25%). Thus, the proposed method seems to be a promising step toward solving the open problem of imagined speech recognition in participant dependent BCI systems with randomized order of the stimuli.

Table I. Classification performance measures (%) of 10-fold cross-validation.

S. No	C_1	C_2	C_3	C_4	Multi-class	Kappa	Band (Hz)
1	36	40	48	76	50	33.33	3-15
2	40	**52**	40	**92**	**56**	**41.33**	9-22
3	40	36	36	76	47	29.33	4-16
4	**48**	40	36	84	52	36.00	9-19
5	**48**	36	44	80	52	36.00	11-25
6	44	40	40	84	52	36.00	5-18
7	36	48	40	72	49	32.00	8-23
8	44	44	56	72	54	38.67	6-19
9	32	44	44	88	52	36.00	10-25
10	44	40	**60**	76	55	40.00	12-25
Mean	41.20	42.00	44.40	80.00	51.90	35.87	-
SD	5.35	5.08	8.10	6.80	2.73	3.64	-

Besides, CSP is an algorithm susceptible to noise and artifacts, so removing noise and artifacts that are not related to the desired task can enhance the classification accuracy. Since the effect of a mental task on individuals is different, the optimum frequency range is entirely subject dependent. To analyze the effect of filter range on classification results, we performed classification on filtered data of five frequency bands, including Delta (1-4 Hz), Alpha (8-15 Hz), Delta to Alpha (1-15 Hz), Delta to Beta (1-40 Hz), and optimum frequency band obtained in section III. Fig. 5 shows the classification accuracies for five frequency ranges. According to Fig. 5, the individualized optimum frequency band outperformed other fixed frequency bands in all participants. Therefore, it can be concluded that selecting an individualized frequency range for each participant can achieve a higher classification accuracy compared with using a fixed frequency range for all participants.

On the other hand, the performance of our method and the number of extracted features lead us to the question of whether feature selection is necessary or not. The classification accuracies in two conditions comprise using no feature selection, and using the NCA feature selection method are illustrated in Fig. 6. In order to statically compare these two conditions, first the normality of the data was confirmed using the Shapiro-Wilk test (p≥0.05). Then, a paired-samples t-test was performed to compare classification accuracies obtained in feature selection (Mean=51.90, SD=2.73) and no feature selection (Mean=46.10, SD=3.76) conditions. There was a significant improvement in the classification accuracy in the NCA feature selection condition compared with the no feature selection condition (t(9)= -8.74, p-value=0.000011).

Although our classification accuracy is considerably higher than the chance level (25%), such performance is not good enough for real-life BCI systems. Kaiser *et al.* reported that the patterns of cortical activity of the brain change during long-term training in motor imagery tasks [14]. Thus, a long-term feedback-based training may also increase the classification accuracy of a BCI system based on imagined speech. Furthermore, to verify the reliability of the proposed algorithm, more words and more trials of each class are required.

In summary, our results show that selecting an informative subset of CSP-based features extracted from the signal filtered to the individualized optimum frequency band, and the use of stacking ensemble learning as a classification method, can improve the classification accuracy of an imagined speech-based task with randomized order of words.

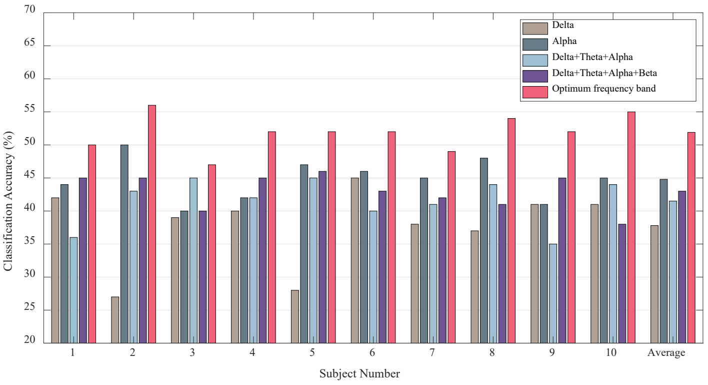

Figure 5. The effect of filter range on classification accuracy

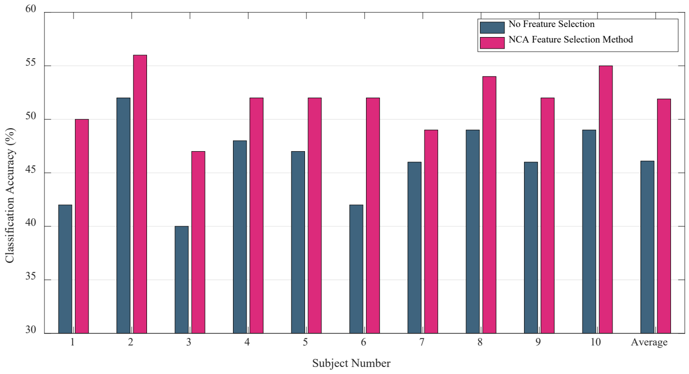

Figure 6. The effect of feature selection on classification accuracy

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain–computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and G. García-Aguilar, “Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification,” *Expert Syst. Appl.*, vol. 59, pp. 1–12, 2016.

[3] A. Porbadnigk, M. Wester, and T. S. Jan-p Calliess, “EEG-based

speech recognition impact of temporal effects," 2009.

[4] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis and classification of speech imagery EEG for BCI," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 901–908, 2013.

[5] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Eeg classification of covert speech using regularized neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 12, pp. 2292–2300, 2017.

[6] C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features," *J. Neural Eng.*, vol. 15, no. 1, p. 16002, 2017.

[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," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 4409–4414.

[8] L. Wang, X. Zhang, X. F. Zhong, and Z. W. Fan, "Selecting Filter Range of Hybrid Brain-Computer Interfaces by Mutual Information," in *Advanced Materials Research*, 2014, vol. 981, pp. 171–174.

[9] "3rd Iranian BCI Competition (iBCIC2020)." [Online]. Available: http://nbml.ir/FA/scientific-tournament/102640.

[10] W. Ibl, V. Dyke, and T. Ibl, "9 Applications of Interactive Boundary Layer Models," vol. 112, pp. 713–719, 2001.

[11] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *2005 IEEE engineering in medicine and biology 27th annual conference*, 2006, pp. 5392–5395.

[12] W. Yang, K. Wang, and W. Zuo, "Neighborhood Component Feature Selection for High-Dimensional Data.," *JCP*, vol. 7, no. 1, pp. 161–168, 2012.

[13] G. A. Pressel Coretto, I. E. Gareis, and H. L. Rufiner, "Open access database of EEG signals recorded during imagined speech," *12th Int. Symp. Med. Inf. Process. Anal.*, vol. 10160, p. 1016002, 2017.

[14] V. Kaiser *et al.*, "Cortical effects of user training in a motor imagery based brain-computer interface measured by fNIRS and EEG," *Neuroimage*, vol. 85, pp. 432–444, 2014.