2024 CMBEC46
Toronto, ON
May 28-30, 2024
CMBES/SCGB

Synergizing Spectrotemporal Dynamics and Filterbank Common Spatial Pattern for Decoding Imagined Speech from EEG

Fatemeh Bagheri *,1,2 and Behrad TaghiBeyglou *,3,4

[1] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada
[2] North York General Hospital, Toronto, Canada
[3] Institute of Biomedical Engineering, University of Toronto, Toronto, Canada
[4] KITE Research Institute, Toronto Rehabilitation Institute – University Health Network, Toronto, Canada
* The authors contributed equally.

I. Introduction

Imagined speech refers to a cognitive process wherein an individual endeavours to mentally simulate the act of uttering a word, devoid of any physical movement within the articulatory musculature or the production of audible sound. Recent studies have revealed that the imagined speech can be decoded through electroencephalogram (EEG) recordings and accordingly provide the fundamentals that can be used for developing low-cost, portable, and accurate brain-computer interface (BCI) systems.

II. Materials and Methods

A. Dataset

EEG signals from 15 healthy right-handed individuals (5 females and 10 males) with an average age of 31 years, were acquired and then released in the 3rd Iranian BCI competition (iBCIC2020)[1]. The signals were sampled at 2400 Hz and captured from 64 channels configured according to 10-20 system EEG placement. The data were collected from the participants while they were comfortably sitting on a chair with a display screen placed half a meter away. The task the person must perform during the experiment is to imagine speaking a word silently. To do this, images of a hand in various rock, paper, and scissors gestures as well as No-action are shown to and asked from the participant. Therefore, in this study, we have four different classes spread through 60 trials of each of 15 participants, and each trial contains 9 seconds of recording.

B. Pre-processing

The recordings are first downsampled by a factor of 10, and then a linear-phase finite impulse response (FIR) filter with a passband range of 4 to 32 Hz is applied to them. For each trial, the first average of 250 ms of recording (from onset) is used as the baseline and removed from the whole trial. To eliminate the electrooculogram (EOG) artifact, independent component analysis (ICA) is applied and the sources that have summed correlation with three frontal-lobe channels are removed through the back-projection step. Likewise, for electromyogram (EMG) artifacts, the sources with a ratio of power within the range of 25 to 30 Hz over the power within the range of 20 to 25 Hz greater than 0.5 are eliminated through the back-projection step. Finally, the average of all channels is used as a reference to eliminate background activity. Each trial is processed independently to prevent biased pre-processing.

[1] https://nbml.ir/en/scientific-tournament/102640

C. Model development

In this study, we conducted feature extraction by capturing a diverse range of characteristics, which were subsequently merged to develop a classifier. The initial group of features comprised two filters derived from the filter bank common spatial pattern (FBCSP) [1], spanning 7 equidistant frequency bands ranging from 4 to 32 Hz. The second set of features corresponded to the coefficients of a 5th-order autoregressive (AR) model applied to channels C1, C2, C3, C4, CP3, CP4, P3, P4, FC3, and FC4. In the third group, we extracted the second norm of the 3rd level discrete wavelet transform (DWT) from the aforementioned channels. Lastly, the fourth group involved extracting the overall energy encompassed in the power spectral density (PSD) of the aforementioned channels. Thereafter, the NSGA-II evolutionary algorithm with an objective function of $(1-accuracy)$ was employed to reduce the number of features. Finally, we developed the Naive Bayes classifier in `one vs. rest` scheme and reported the results based on 5-fold cross-validation per participant.

III. Results

The average accuracy and Cohen's kappa on all participants is $44.37\pm5.04\%$ and 0.42 ± 0.07, respectively.

IV. Discussion

Our introduced pipeline introduced a simple approach for classifying three different imagined words and No-task EEG recordings and showed a comparatively acceptable result compared to the previous study [2] (average accuracy of 51.2% and kappa of 0.36), which illustrates the efficacy of employing more complex feature selection alongside simpler features.

References

1. K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008, pp. 2390–2397.
2. N. Hamedi, S. Samiei, M. Delrobaei, and A. Khadem, "Imagined speech decoding from eeg: The winner of 3rd iranian bci competition (ibcic2020)," in *2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)*. IEEE, 2020, pp. 101–105.