2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS) | 978-1-6654-0170-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICHMS53169.2021.9582457

Inner Speech Classification using EEG Signals: A Deep Learning Approach

Bram van den Berg
Department of Cognitive Science and AI
Tilburg University
Tilburg, The Netherlands
b.vdnberg@tilburguniversity.edu

Sander van Donkelaar
Department of Cognitive Science and AI
Tilburg University
Tilburg, The Netherlands
s.n.vandonkelaar@tilburguniversity.edu

Maryam Alimardani
Department of Cognitive Science and AI
Tilburg University
Tilburg, The Netherlands
m.alimardani@tilburguniversity.edu

Abstract—**Brain computer interfaces (BCIs) provide a direct communication pathway between humans and computers. There are three major BCI paradigms that are commonly employed: motor-imagery (MI), event-related potential (ERP), and steady-state visually evoked potential (SSVEP). In our study, we sought to expand this by focusing on "Inner Speech" paradigm using EEG signals. Inner Speech refers to the internalized process of imagining one's own "voice". Using a 2D Convolutional Neural Network (CNN) based on the EEGNet architecture, we classified the EEG signals from eight subjects when they internally thought about four different words. Our results showed an average accuracy of 29.7% for word recognition, which is slightly above chance. We discuss the limitations and provide suggestions for future research.**

Keywords—Brain Computer Interface (BCI), EEG, Inner Speech Classification, Deep Learning, Convolutional Neural Network (CNN)

I. Introduction

Brain computer interfaces (BCIs) are systems that allow users to control or communicate with external devices through brain activity [1]. BCIs are mainly developed for individuals with severe disabilities to provide them a non-muscular communication channel with their surroundings by translating their intentions to digital commands for assistive applications such as speech synthesizer, wheelchair, or neural prostheses [2].

Several BCI paradigms have been developed in the past, allowing extraction of information from brain activity during a specified task (see Abiri et al. [3] for a review). For instance, motor imagery BCIs decode event-related desynchronization (ERD) in sensorimotor rhythm every time the user imagines a body part movement [4]. P300 spellers use the P300 event-related potential (ERP) in response to visual stimulation such as a digital keyboard and enable the user to spell words by selectively paying attention to the intended letters [5]. Steady State Visually Evoked Potential (SSVEP) is yet another BCI paradigm that monitors brain responses to visual stimulation at specific frequencies [6].

Although previous BCI paradigms haven proven to be useful developments, they each come with their own drawbacks [2, 3]. For instance, motor imagery BCIs require a long training phase until the users become able to voluntarily modulate their brain activity and gain control over the system [3]. Other BCIs such as the P300 and SSVEP paradigms depend on external stimulation and hence the user must continuously pay attention to the visual or auditory stimuli, which will not allow engagement in other activities [3]. Spelling words letter-by-letter or performing a movement imagination for every command that the user considers necessitates a great deal of mental effort [7]. Additionally, almost all BCI paradigms suffer from a low Information Transfer Rate (ITR) and they require calibration in order to tune the system for individual users [7]. Therefore, what current BCI systems lack is a fluid and intuitive paradigm that allows users communicate with the system in a natural way without excessive dependence on external stimulation and training [7, 8].

To meet this challenge, inner speech recognition has been proposed as an alternative communication paradigm for BCIs [8, 9]. Inner speech (also known as imagined, covert or internal speech) refers to the internalized process of speech, without articulating it [8]. It is usually associated with auditory imagery or an inner voice. There is evidence from past neuroscience research that inner speech engages brain regions that are commonly associated with language comprehension and production [10]. This includes temporal, frontal and sensorimotor areas predominantly in the left hemisphere of the brain [10, 11]. Therefore, by monitoring these brain areas, it is theoretically possible to develop an inner speech BCI that classifies neural representations of imagined words [11].

At the moment, most studies that research the classification of inner speech focus on more invasive methods such as electrocorticography (ECoG) [9] as they provide higher spatial resolution. Relatively little research is available that has attempted the classification of inner speech using EEG signals [8, 12]. It is important for a BCI application to be non-invasive, accessible and easy to implement so that it can be used by a large group of patients. Therefore, the current study focused on classifying inner speech from EEG data, as it is non-invasive, relatively cheap and more accessible to general users [8].

This study aims at investigating to what extent inner speech can be classified using EEG data. We employed a recently published dataset that includes EEG recordings

978-1-6654-0170-8/21/$31.00 ©2021 IEEE

during inner speech of four imagined words [12]. Given the complexity of EEG signals and the potential loss of information in traditional machine learning approaches, we employed a convolutional neural network (CNN) model for our multi-class classification task. CNN is a subclass of deep learning architectures that can learn spatial and temporal representations in the EEG data. Its major advantage is that it can automatically extract task-related features from raw signals in an end-to-end fashion [13]. This not only removes the need for time-consuming preprocessing and feature engineering, but also could potentially lead to superior performance as compared to traditional machine learning methods [14, 15].

II. Methods

A. Dataset Description

The dataset employed by this study is a publicly available dataset shared by Nieto et al. (2021) [12]. The dataset consists of EEG signals from 8 participants recorded by 134 channels distributed all over the scalp according to the 'ABC' layout of BioSemi system [12]. The participants were instructed to produce inner speech for four words; 'up', 'down', 'left', and 'right' based on a visual cue they saw in each trial. The cue was given in form of an arrow on a computer screen that rotated to the corresponding directions. This was repeated 220 times for each participant. However, since some participants reported fatigue, the final amount of trials included in the dataset for each participant differed slightly. The total number of trials was 2236 with equal number of trials per class for all participants.

The EEG signals included event markers and were already preprocessed. The preprocessing included a band pass filter between 0.5-100 Hz, a notch filter at 50 Hz, artifact rejection using Independent Component Analysis (ICA) and downsampling to 254 Hz. Since the neural correlates of inner speech processing are reported to be mainly present in the left hemisphere (see Introduction), we used a subset of 26 channels distributed throughout this area (see Fig 1).

B. CNN Classification Model

After channel selection, each EEG segment corresponded to a matrix with a dimension of [26 × 1153], where 26 indicates the number of channels and 1153 is the number of samples in the EEG segment. These matrices were stacked into a tensor of dimension [Trials × 30 × 1153]. After importing the data, it was scaled using the RobustScaler of SKlearn. The model was then trained on raw EEG signals and no features were further extracted. This would allow the deep learning model to automatically extract the necessary features.

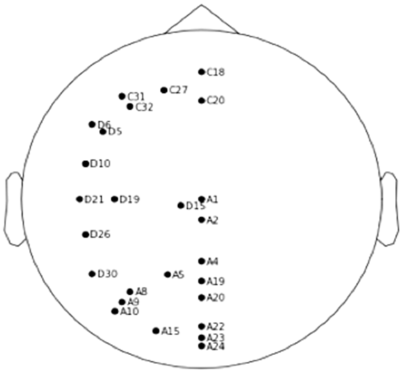

Fig. 1 An overview of the selected channels for inner speech classification in the current study.

EEGNet [16] was chosen as the architecture for our CNN model. EEGnet is a relatively small convolutional neural network designed for various EEG-related classification tasks, and is able to capture common temporal and spatial EEG features through its convolutional filters. We used an implementation of EEGNet that is available on GitHub [17].

C. Training and Evaluation

The model was trained and then tested in Google Colab using K-fold cross validation. K-fold cross validation is a technique in which the full dataset is shuffled and partitioned into k equal-size data folds. Out of these folds, one fold is used to test the model and the remaining folds are used to train the model. This is repeated for all folds and the obtained performance metrics are averaged over all folds, leading to a more reliable estimate. In this study, the whole dataset was split into four folds. Moreover, the model was trained and tested on each subject individually. In order to get a complete overview of the model performance, accuracy, precision, recall and F1-score are reported.

IV. Results

Table 1 summarizes the EEGNet model performance for each subject. The inner speech was predicted with an average accuracy of 29.67%. Therefore, the model predicted inner speech slightly above chance. The precision, recall and F1 scores indicated similar values. The highest accuracy score was reported for Sub 8, with an average accuracy of 34.50%, and the lowest scoring participant was Sub 4, with an average accuracy of 23.75%.

TABLE I. CNN performance for each subject on the 4-class Inner speech classification task.

Subject	Accuracy	Precision	Recall	F1-score
Sub 1	30.00%	31.09%	30.00%	30.25%
Sub 2	30.41%	30.57%	30.42%	30.44%
Sub 3	30.56%	30.54%	30.56%	30.43%
Sub 4	23.75%	23.36%	23.36%	23.46%
Sub 5	30.00%	30.10%	30.10%	29.89%
Sub 6	27.78%	27.60%	27.60%	27.55%
Sub 7	30.41%	30.21%	30.21%	30.21%
Sub 8	**34.50%**	34.66%	34.50%	34.57%
Average	29.67%	29.76%	29.68%	29.61%

V. Discussion

The results of our study indicated that we were able to classify inner speech with an average accuracy of 29.67% per participant. While the accuracy scores of our model were rather low, they were still slightly above chance. This suggests that inner speech classification using EEG signals is possible and further improvements of the model and training data is required to achieve better performance. This would ultimately facilitate development of a BCI paradigm that enables communication in a more fluid and natural manner than the currently available paradigms provide.

Past studies had already reported some results regarding classification of inner speech signals using EEG [8], however these studies differ with regard to the choice of prompts for inner speech task (e.g. vowels vs. words), channel selection, EEG processing and classification approaches [8]. For instance, in the study by Sereshkeh et al. [18], regularized neural networks were used to classify the mental repetition of the words 'yes' and 'no', versus an unconstrained rest phase. This study had successful results with an average accuracy of 75.7% when classifying covert speech versus the rest trials. When considering all three classes, the study yielded an average accuracy of 54.1%. All classifications were conducted using EEG features that were extracted via a discrete Wavelet transform [18].

In another paper by Cooney et al. [19], a CNN model was used together with Transfer Learning to classify between five vowels. The reported overall accuracy was 36%, which was higher than chance level (20%). The results of their study indicated that transfer learning could significantly improve classification accuracy. A later study by Cooney and colleagues [20] investigated how hyper-parameters affect accuracies in classifying inner speech. The study compared multiple well-known CNN architectures for EEG classification, including EEGNet, Shallow CNN and Deep CNN. They found that EEGNet yielded the best performance, with an average classification accuracy of around 30% for the classification of five words.

A potential reason for superiority of EEGNet performance in the study of Cooney et al. [19] could be the application of nested cross validation (nCV), which allowed for hyperparameter optimization during the training. Authors argue that BCI classifiers can benefit from nCV, as these classifiers are usually trained on small datasets, which can lead to overfitting when deep learning models are applied. Therefore, future research should consider hyperparameter optimization as an important step in the model training particularly in complex tasks such as inner speech classification [20].

The current study entailed some limitations. First, our CNN model was trained with a small dataset including only a few participants and limited number of trials. This could have constrained the model performance, as deep learning methods are extremely data hungry. Additionally, the choice of words (movement direction) and visual cues for the inner speech task could have elicited brain activity related to functions other than the intended task e.g. motor imagery and visual processing. Previous study using surface electromyography (EMG) signals suggests that inner speech production is not a process merely regulated by abstraction of linguistic representations, but that it might incorporate simulation of motor system depending on the task [21].

Finally, the current study focused on a user-dependent classification in which the model was only trained and tested on one subject at a time. This requires calibration of the BCI system for every individual, inhibiting the practicality of the system [22]. Recent studies suggest transfer learning as an approach to mitigate the two challenges inherent to EEG classification; data scarcity and individual differences [23]. The advantage of transfer learning models is that they increase learning speed and cross-subject generalization, which are desired qualities in inner speech classification and BCI systems in general [19, 23].

In sum, our study provided preliminary results for decoding of inner speech using EEG signals. Although our results did not reveal significant performance by the CNN model, they did support the possibility of inner speech classification using a non-invasive and accessible brain imaging technique [8]. Future works should employ larger datasets or alternatively use data augmentation techniques that would improve generalizability and robustness of EEG classifiers [24]. This study only relied on a deep learning approach for automatic feature recognition from a subset of EEG channels located in the left hemisphere of the brain, however, future research can further explore the efficacy of traditional machine learning methods that are trained on hand-crafted features obtained from all brain regions. This will provide insight into the cognitive and neural mechanisms that govern inner speech production. Furthermore, future research should explore the possibility of a user-independent classifier that is able to discriminate inner speech signals in a calibration-free setting [22].

VI. Conclusion

This study aimed to investigate the feasibility of inner speech classification of raw EEG signals using a deep learning approach. We employed a CNN model (EEGNet) on a recent EEG dataset to classify four classes of inner speech. Results showed that our model was able to predict internal word imagery with an average accuracy of 29.67%. Our results are consistent with the current literature and provide insight into potential applications of deep learning models for inner speech BCI paradigm. Future work can further expand the model performance by taking the limitations of the current study into account.

References

[1] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology, 113*(6), 767-791.

[2] Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. Sensors, 12(2), 1211-1279.

[3] Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering, 16*(1), 011001.

[4] Pfurtscheller, G., Brunner, C., Schlögl, A., & Da Silva, F. L. (2006). Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage, 31*(1), 153-159.

[5] Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering, 5*, 14.

[6] Zhu, D., Bieger, J., Garcia Molina, G., & Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Computational Intelligence and Neuroscience, 2010*.

[7] Rashid, M., Sulaiman, N., Majeed, A. P. A., Musa, R. M., Nasir, A. F. A., Bari, B. S., & Khatun, S. (2020). Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Frontiers in Neurorobotics, 14*, 25.

[8] Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding Covert Speech From EEG-A Comprehensive Review. *Frontiers in Neuroscience, 15*, 392.

[9] Martin, S., Iturrate, I., Millán, J. D. R., Knight, R. T., & Pasley, B. N. (2018). Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis. *Frontiers in Neuroscience, 12*, 422.

[10] Amit, E., Hoeflin, C., Hamzah, N., & Fedorenko, E. (2017). An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fMRI. *NeuroImage, 152*, 619-627.

[11] Bocquelet, F., Hueber, T., Girin, L., Chabardès, S., & Yvert, B. (2016). Key considerations in designing a speech brain-computer interface. *Journal of Physiology-Paris, 110*(4), 392-401.

[12] Nieto, N., Peterson, V., Rufiner, H. L., Kamienkowski, J., & Spies, R. (2021). " Thinking out loud": an open-access EEG-based BCI dataset for inner speech recognition. *bioRxiv*.

[13] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering, 16*(5), 051001.

[14] Alimardani, M., & Kaba, M. (2021, May). Deep Learning for Neuromarketing; Classification of User Preference using EEG Signals. In *12th Augmented Human International Conference* (pp. 1-7).

[15] Tibrewal, N., Leeuwis, N., & Alimardani, M. (2021). The Promise of Deep Learning for BCIs: Classification of Motor Imagery EEG using Convolutional Neural Network. *bioRxiv*.

[16] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering, 15*(5), 056013.

[17] https://github.com/vlawhern/arl-eegmodels

[18] Sereshkeh, A. R., Trott, R., Bricout, A., & Chau, T. (2017). EEG classification of covert speech using regularized neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*(12), 2292-2300.

[19] Cooney, C., Folli, R., & Coyle, D. (2019, October). Optimizing layers improves CNN generalization and transfer learning for imagined speech decoding from EEG. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 1311-1316). IEEE.

[20] Cooney, C., Korik, A., Folli, R., & Coyle, D. (2020) Evaluation of Hyperparameter Optimization in Machine and Deep Learning Methods for Decoding Imagined Speech EEG. *Sensors, 20*(16), 4629.

[21] Nalborczyk, L., Grandchamp, R., Koster, E. H., Perrone-Bertolotti, M., & Loevenbruck, H. (2020). Can we decode phonetic features in inner speech using surface electromyography?. *PloS one, 15*(5), e0233282.

[22] Kwon, O. Y., Lee, M. H., Guan, C., & Lee, S. W. (2019). Subject-independent brain–computer interfaces based on deep convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems, 31*(10), 3839-3852.

[23] Wan, Z., Yang, R., Huang, M., Zeng, N., & Liu, X. (2021). A review on transfer learning in EEG signal analysis. Neurocomputing, 421, 1-14.

[24] Lashgari, E., Liang, D., & Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods*, 108885.