GCN-LSTM for EEG Classification based on Unspoken Speech of Bilinguals

Chengfang Li, Liang Song*
Academy for Engineering & Technology, Fudan Univeirsity, Shanghai, China
Email:cfli20@fudan.edu.cn; songl@fudan.edu.cn

***Abstract*—Research on the bilingual mechanism of the brain is significant for bilingual learning and human cognition. The classification of EEG signals of bilinguals under different languages is an important precursor for studying bilingual mechanisms. Compared to functional Magnetic Resonance Imaging (fMRI), Electroencephalogram (EEG) has a higher temporal resolution and can capture the changes in neural signals at the millisecond level. Event-related potentials (ERPs) can only be detected in specific areas. They are also susceptible to the mental and attentional influences of the subjects. This has serious limitations for further exploration of bilingual brain regions and information flow. Therefore, research on the bilingual mechanisms of the brain using EEG signals is urgently needed. This study collects an unspoken speech EEG signals dataset from six subjects. Furthermore, propose a novel Graph Convolution Network-Long Short Term Memory (GCN-LSTM) for classification. The highest classification accuracy of a single subject is 92.86%, and the average accuracy is 83.81%. It proves there are differences in EEG signals when bilinguals silently read different languages. This study helps to investigate the bilingual mechanism and location further using EEG.**

Keywords*—*electroencephalogram, bilingual mechanisms, unspoken speech, graph convolution network, long short-term memory

I. Introduction

Bilingualism is a neural mechanism unique to the human brain. Under normal circumstances, humans can not only master their native language but also learn two or more languages and switch freely between them [1]. Comprehensively exploring the bilingual mechanism of the brain can reveal the nature of language and thinking, which have significance for researching brain function, bilingual learning, and improving human cognition [2], [3]. However, due to the limitations of algorithms, experimental equipment, and computing tools, early research used functional Magnetic Resonance Imaging (fMRI) [4] to explore the bilingual mechanism. Due to its high spatial resolution, many studies [5], [6] have used fMRI to verify the differences between brain regions under different languages on bilinguals. Due to the low temporal resolution of fMRI, it is not possible to capture the rapidly changing neural activity. Event-related potentials (ERPs) [7], [8] are the main methods for studying bilingual mechanisms in recent years. However, the conditions for realizing ERPs are very demanding, mainly because they require the initiative of the subjects, and their psychology and attention influence them. In addition, ERPs have strict requirements on the signal acquisition position. More importantly, there is a tighter locking relationship between the latency of ERPs and the stimulus, which appears almost immediately or instantaneously within a certain period when the stimulus is given. This has serious limitations for fully exploring the localization of functional brain areas and the transmission of neural signals in bilinguals when they use different languages. Compared with fMRI, EEG has a higher temporal resolution, capturing signal changes at the millisecond [9]. Presently, EEG has been widely used in Brain-computer Interface (BCI) research. In studies of EEG using neural networks, EEG signal characteristics over the total period were considered, compensating for the shortcomings of ERPs that only capture signal changes within transients. Therefore, EEG is more useful for further investigation of bilingual mechanisms. However, few studies focused on classifying EEG signals under different languages in bilinguals.

In an applied BCI study, Balaji et al. [10] collected EEG signals from subjects whose native language was Hindi and English was the second language while reading English ('Yes'/'No') and Hindi ('Haan'/'Na') silently. The highest accuracy was 92.18% for the classification of both languages using Artificial Neural Networks (ANN). To be mentioned, this study only focused on one word, where the neural activity time of the brain is relatively short and not fully active. To the best of our knowledge, there is still a gap in research on the differences in EEG signals of bilinguals in their corresponding language environments. Therefore, in the present study, we collected a dataset of EEG signals from bilinguals silently reading different languages as a more sophisticated benchmark for bilingual classification.

With the development of deep learning, Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) have boosted the performance of EEG signal classification [11], [12]. EEG signals are typically non-Euclidean structured data with high latitude and disorder. These methods may not be able to perform complex EEG signal classification tasks because RNNs and CNNs cannot capture the spatial connections between collected channels [2], [13]. To address this problem, we propose a novel Graph Convolution Network-Long Short-Term Memory (GCN-LSTM) to capture the spatial and temporal associations between channels in raw data. Graph Convolution Networks (GCN) can capture the dynamic characteristics of graph structure and effectively overcome individual differences. To capture the temporal relationship, the Long Short Term Memory (LSTM) [14] after GCN-LSTM. Furthermore, the Pearson correlation coefficient (PCC) is used to select the EEG channels with more neural activity in bilingualism. It helps reduce the amount of calculation and achieves better performance. We classified the EEG signals of

*Corresponding author. This research is funded by the China Mobile Research Fund of Chinese Ministry of Education (Grant No. KEH2310029).

each bilingual silently reading different languages. The highest classification accuracy of one subject is 92.86%, and the average classification accuracy is 89.46%, which outperforms other popular methods for EEG signal classification. The main contributions are summarized as follows:

- We designed an experiment and collected a dataset of unspoken speech EEG signals in different languages by 6 bilingual subjects.
- We propose a novel GCN-LSTM for complex EEG signals classification.
- On our dataset, the proposed model achieves superior performance to EEGNet [15], DeepCovNet [16], and ShallowCovNet [17].

II. DATASET

A. Experiment Setup

A 64-channel wet electrode wireless device is used for EEG collection. The sampling frequency is $1000Hz$. A wireless amplifier on the acquisition device amplifies the acquired EEG data by 2×10^4 times. The 64-channel includes 59 brain channels and ECG, HEOR, HEOL, VEOU, and VEOL channels. The corresponding brain position of each channel complies with the 10–20 system electrode placement method [18]. Before the experiment, inject conductive paste into each channel electrode to ensure all the impedance of channels is less than $20\mu V$. Six volunteers participated in our experiment (23.52 ± 2.6 years, average age 21, 3 females and 3 males). They are all right-handed and have no other diseases. Their native language is Chinese, and their second language is English. They start learning English around age ten.

Before the experiment, we inform them of the experimental procedure in detail and obtained the signed consent form. The experimental procedure is conducted under the ethical standards encoded in the latest Declaration of Helsinki. We use paper reading materials instead of screens to reduce the screen lighting influence. The English material has 300 to 450 words, and the words Chinese material are between 600 and 1000. The article contains interesting short stories, dialogues, and descriptions of scenery and things. Articles are all daily words and no uncommon words. Before the start of the experiment, we play 20 seconds of Chinese or English audio to make the subjects into the corresponding language environment. Throughout the experiment, we only need to ensure that subjects read the corresponding language materials silently and did not care about the deviation caused by reading speed. In the corresponding language environment, the neural mechanism of the brain processing the related language will not change due to reading speed. The equipment will automatically mark each stage according to the program to facilitate subsequent cutting. Due to the different reading speeds, the end time of the experiment varies, and the end mark of the experiment is made manually by the experiment monitor. Each subject conducts three experiments except subject 6. Subject 6 conducts forty experiments, which ensures the diversity of data, and we are constantly enriching the data. We guarantee subjects sufficient rest in the middle of each group of experiments. Psychologists and language experts evaluate the experiment and the reading materials. The experimental flow diagram is shown in Fig. 1.

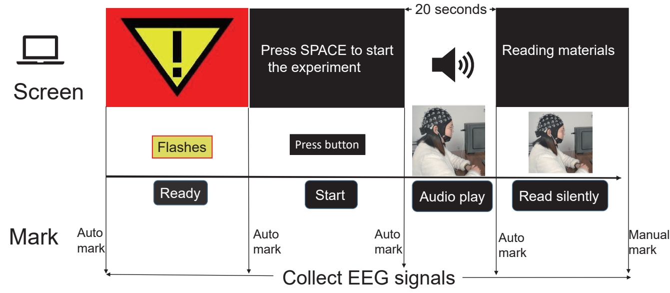

Fig. 1. Experiment setup. The screen will prompt the experiment process. Subjects wear headphones when listening to the audio and read the paper version of the text silently without looking at the screen.

B. Data Format

Each subject collects 3 to 5 groups of English and Chinese unspoken speech EEG signals, respectively. Each group collects data.bdf, evt.bdf, and instruction files. The data.bdf includes the EEG data of the whole experiment process. The evt.bdf includes marks and corresponding data points. The instruction file is in json format and includes experiment time, channel name, and device Information. We crop out the unspoken speech EEG signals every 2 seconds. Therefore, each subject has Read Chinese Silently (RCS) and English Silently (RES) datasets. The data in subjects 1 to 6 are 133, 74, 90, 131, 138, and 614. Using the same network to classify data of different magnitudes is also challenging. We divided the dataset of each subject into the training and test in the ratio of 80% : 20%.

III. METHOD

A. EEG Graph Construction

EEG with high latitude and strong sequential nature is typical of non-Euclidean structured data. The graph structure can better represent non-Euclidean structured data [19]. In the graph construction $G(N, E)$, N is the set of nodes, E is the set of edges, the weight set $E_s = \{E_{ij}|(i,j) \in N\}$, E_{ij} is the weight of edge between node i and j. We construct the EEG data into a graph structure to catch the spatial connection and represent the characteristics of the data well. Neural networks use end-to-end learning methods to learn complex features from data automatically. Craik [20] et al. suggest that machine learning encourages researchers to feed raw signal values directly into neural networks without manually designing features, which may help deep learning capture raw EEG data features directly. Therefore, we construct the raw data into a graph structure to capture the features between the data. We take the 59 channels as the nodes of the graph. The weight of the edge between nodes is the Pearson Correlation Coefficient (PCC) value, and the value follows the formula 2. When the PCC value is greater than 0, the weight on edge is the PCC value. When less than or equal to 0, take 0. The weight of the graph enhances the relation of channels. PCC is calculated as follows:

$$\rho(X,\ Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \quad (1)$$

where cov is the covariance, and σ_X, σ_Y are the standard deviation of the two variables.

$$w(i,\ j) = \begin{cases} \rho(X,\ Y), & \rho(X,\ Y)>0, \\ 0, & \rho(X,\ Y) \leq 0. \end{cases} \quad (2)$$

The constructd data includes a feature matrix and an adjacency matrix. There are N channels and T time points, the characteristic matrix $F \in \mathbb{R}^{N\times T}$, and the responding adjacency matrix $A \in \mathbb{R}^{N\times N}$. In the research, N is 59 and T is 2000.

B. GCN-LSTM

Extending neural networks to graph structures can better capture the spatial relationships between data [21]. Spectral perspective is one of the streams of the Graph Convolution Network, which considers the locality of graph convolution in the form of spectral analysis. The other is the spatial perspective, which directly applies convolution filters to graph nodes and their neighbors. This work follows the spirit of the spatial perspective, we construct a graph structure for each data. Each data corresponds to its unique CNN filter, effectively overcoming the individual difference in the EEG signal. The adjacency matrix is symmetric in an undirected weighted graph, and the principal diagonal is zero. The adjacency matrix L introduces the degree matrix and regularizes, $L = I_n - D^{-1/2}AD^{-1/2}$, where, $I_n - D^{-1/2}AD^{-1/2} \approx D^{-1/2}\tilde{A}D^{-1/2}$. The feature matrix as a filter and apply it to all nodes. The size of the convolution kernel is $N \times F$. The receptive field is the whole node. Therefore, the GCN-LSTM layer is:

$$f_{(out)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}FW(i,h)\right), \quad (3)$$

where σ is the activation function, W is the training weight matrix, i is the input dimension, and h is the output dimension. A more detailed explanation of this formulation can be found in [22]. Each feature matrix matches its corresponding adjacency matrix. Therefore, the convolution kernel corresponding to each data is unique. To fully capture the spatial connection between the data, we take the output of the previous layer as the feature matrix of the next layer. The transfer relationship between the GCN layer and the layer is as follows:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}^{(d)}\tilde{D}^{-\frac{1}{2}}H^{(l)}W(i,h)\right), \quad (4)$$

where $H^{(l)} \in R^{N\times h}$ is the output of the previous layer as the characteristic matrix of this layer. $\tilde{A}^{(d)}$ is the d-th adjacency matrix that corresponding to the d-th characteristic matrix.

In our work, we use three-lay GCN to classify. Each network layer has a training weights matrix, and the number of hidden layers differs. The activation function of the first and third layers is Relu, and the second layer is Tanh. We use three-lay LSTM after the GCN network to better capture temporal features. Then, the network uses the fully connected layer and softmax to classify the EEG signals in Chinese and English. The inputs of GCN-LSTM are each feature matrix and its corresponding adjacency matrix, constantly feeding back and training the network in the convolution process to achieve the dynamic balance of the network and capture the dynamic spatial characteristics between channels. The model is achieved by PyTorch framework using python language. We also use early stopping, dropout, and regularization to prevent overfitting. The loss function is cross-entropy, the learning rate is 0.001, uses the Adam optimizer, and the decay is 0.0001. The three hidden layers of GCN-LSTM are 132, 125, and 32, and the dropout is 0.05. The LSTM has three layers, and the hidden layer size is 20. Cross-entropy assesses the gap between the predicted value and the true value. The cross-entropy formula is as follows:

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\cdot\log(q(x_i)), \quad (5)$$

where $p(x_i)$ represents the true distribution of the sample, and $q(x_i)$ represents the distribution predicted by the model.

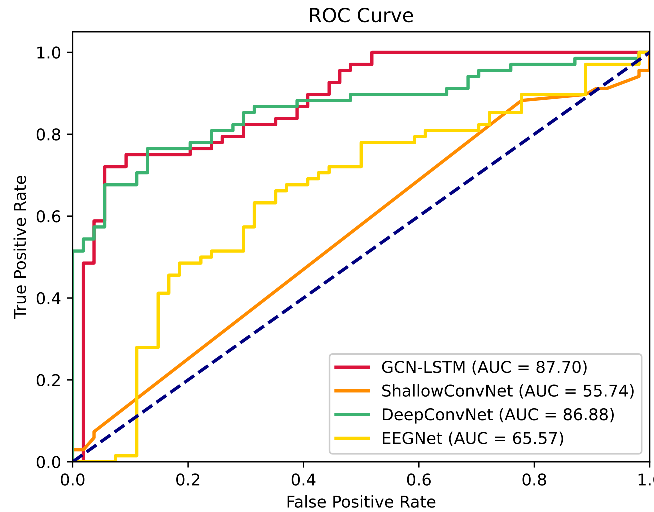

Fig. 2. The ROC curves of GCN-LSTM and other contrast networks.

IV. Results and Discussion

We use 10-fold cross-validation and classify the data using three comparison networks, EEGNet, DeepCovNet, and ShallowCovNet.. The results are shown in Table I. S1–S6 represents subjects 1–6. We use the same GCN-LSTM to classify EEG bilingual data of different orders of magnitude. The results show that our proposed network performs better than other comparison networks in terms of our data. We show the ROC curves of GCN-LSTM and contrast networks for the classification of S6 in Fig. 2.

To verify that the GCN-LSTM captures the dynamic features in the neural signal, we output the third intermediate layer of S1 in RES during the training of the GCN-LSTM, as

TABLE I. The classification results (%) using all channels

	EEGNet	DeepCovNet	ShallowCovNet	GCN-LSTM
S1	53.85	80.76	58.83	**88.46**
S2	57.14	85.71	57.14	**92.86**
S3	58.82	58.82	58.82	**76.47**
S4	56.12	64.10	56.12	**72.17**
S5	59.26	81.48	59.26	**85.19**
S6	65.57	86.88	55.74	**87.70**

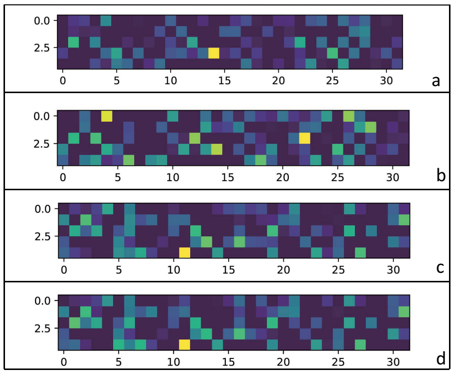

Fig. 3. **a** and **b** are the outputs of the two data randomly selected in the first epoch of subject S1 in RES. **c** and **d** are the outputs of the two data randomly selected in the 50-th epoch of subject S1 in RES.

shown in Fig.3. Comparing the first and the 50th epoch output, and found that after multiple epochs, the EEG data under the same task tended to be consistent, proving that the GCN-LSTM captured the dynamic neural characteristics of EEG signals. In the output of the first epoch, the data has certain rules, but the standard features are not prominent. After 50 epochs, the data of the same label remains dynamic and consistent.

V. CONCLUSION

This study fills the blank of the research on the difference in EEG signals in a pure language environment. We collected EEG signals from six subjects in corresponding language environments to prove it and proposed a novel GCN-LSTM to classify them. The novel network is based on network topology, which uses the abstract human brain structure and its response mechanism to external stimuli to perform nonlinear relation expression and logical operation on complex information. We use the same GCN-LSTM to achieve better classification results for collecting data. The classification accuracy of our proposed network is higher than other comparison networks on all the collected datasets. This also demonstrates the effectiveness of our proposed network in the classification of bilingual EEG signals. In future research, we will extend our dataset to other languages(second languages), such as German, French, etc. We will seek more effective methods to combine EEG and neural signals, which have significance for studying the activation and differences of channels under neural activity. This study provides a basis for further exploring the differences in neural activities under different languages.

REFERENCES

[1] A. Rodriguez-Fornells, M. Rotte, H.-J. Heinze, T. Nösselt, and T. F. Münte, "Brain potential and functional mri evidence for how to handle two languages with one brain," *Nature*, vol. 415, no. 6875, pp. 1026–1029, 2002.

[2] Y. Liu, J. Liu, X. Zhu, D. Wei, X. Huang, and L. Song, "Learning task-specific representation for video anomaly detection with spatial-temporal attention," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 2190–2194.

[3] L. Song, X. Hu, G. Zhang, P. Spachos, K. Plataniotis, and H. Wu, "Networking systems of ai: on the convergence of computing and communications," *IEEE Internet of Things Journal*, 2022.

[4] A. E. Hernandez, A. Martinez, and K. Kohnert, "In search of the language switch: An fmri study of picture naming in spanish–english bilinguals," *Brain and language*, vol. 73, no. 3, pp. 421–431, 2000.

[5] K. H. Kim, N. R. Relkin, K.-M. Lee, and J. Hirsch, "Distinct cortical areas associated with native and second languages," *Nature*, vol. 388, no. 6638, pp. 171–174, 1997.

[6] J. Yang, X. Li, Q. Li, X. Xiao, D. Tong, Q. Wu, and J. Wu, "Phonological processing of japanese kanji and chinese characters in bilingual japanese: An fmri study," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, 2017, pp. 1098–1103.

[7] E. M. Moreno, A. Rodríguez-Fornells, and M. Laine, "Event-related potentials (erps) in the study of bilingual language processing," *Journal of Neurolinguistics*, vol. 21, no. 6, pp. 477–508, 2008.

[8] A. Angrilli, C. Dobel, B. Rockstroh, L. Stegagno, and T. Elbert, "Eeg brain mapping of phonological and semantic tasks in italian and german languages," *Clinical Neurophysiology*, vol. 111, no. 4, pp. 706–716, 2000.

[9] M. X. Cohen, "Where does eeg come from and what does it mean?" *Trends in neurosciences*, vol. 40, no. 4, pp. 208–218, 2017.

[10] A. Balaji, A. Haldar, K. Patil, T. S. Ruthvik, C. Valliappan, M. Jartarkar, and V. Baths, "Eeg-based classification of bilingual unspoken speech using ann," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 1022–1025.

[11] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik, "Speech synthesis using eeg," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1235–1238.

[12] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep learning the eeg manifold for phonological categorization from active thoughts," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2762–2766.

[13] J. Liu, Y. Liu, D. Li, H. Wang, X. Huang, and L. Song, "Dsdcla: Driving style detection via hybrid cnn-lstm with multi-level attention fusion," *Applied Intelligence*, Feb. 2023.

[14] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[16] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[17] S.-J. Kim, D.-H. Lee, and S.-W. Lee, "Rethinking cnn architecture for enhancing decoding performance of motor imagery-based eeg signals," *IEEE Access*, vol. 10, pp. 96 984–96 996, 2022.

[18] R. W. Homan, "The 10-20 electrode system and cerebral location," *American Journal of EEG Technology*, vol. 28, no. 4, pp. 269–279, 1988.

[19] B. A. Miller, N. T. Bliss, and P. J. Wolfe, "Toward signal processing theory for graphs and non-euclidean data," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 5414–5417.

[20] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.

[21] Y. Liu, Z. Guo, J. Liu, C. Li, and L. Song, "Osin: Object-centric scene inference network for unsupervised video anomaly detection," *IEEE Signal Processing Letters*, 2023.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.