A hybrid neural network based on the small-world network for inner speech recognition

Sichu Sun

University of Nottingham, Nottingham, United Kingdom

wyyahge@gmail.com

Abstract. This study is devoted to the identification of human inner speech using an electroencephalogram (EEG), where inner speech refers to an individual's subjective experience of language, disconnected from discernible audible articulation. The core aim of this system is the rapid and precise classification of signals via human inner speech, thus facilitating enhanced control and interaction functionalities. The research entails a comprehensive analysis of 10 volunteers' brain activity across 128 channels from OpenNeuro's Inner speech dataset. A hybrid neural network, which incorporates the small-world network structure, is employed to model neural activity within the brain. This approach outperforms random chance and aligns with current research expectations.

Keywords: EEG, Inner Speech Recognition, BCI, Small-world Network.

1. Introduction

A Brain-Computer Interface (BCI) is an integrated system of hardware and software, facilitating the control of computers or external devices through brain activity. This offers communication capabilities to individuals suffering from severe paralysis or locked-in syndrome [1]. Inner speech recognition pertains to the interpretation of one's internal language, enabling people to mentally visualize words without vocalizing them. These silent articulations can be captured and interpreted by computers, making it easier for individuals with disabilities to communicate their needs and give commands. In the long view, BCIs can forge a more seamless and efficient connection between humans and their surrounding objects.

Certain studies have achieved the classification of human brain signals using an LSTM-RNN algorithm based on wavelet scattering and deep learning, employing basic EEG headsets for signal collection and capture. The acquired commands control brain-driven wheelchairs [2], utilizing specific neural behavior for movement control without directly capturing or isolating human speech.

In terms of direct human inner speech classification, researchers have employed two-dimensional convolutional neural networks based on the EEGNet architecture to categorize brain signals while contemplating different words [3]. Further investigation has yielded promising results using Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) in inner speech classification.

Within the domain of inner speech classification, the utilization of Spiking Neural Networks (SNN) is relatively rare, possibly due to its inability to employ backpropagation effectively for supervised learning. SNNs have demonstrated commendable nonlinear processing capabilities in speech recognition, particularly under complex conditions, reinforcing the biological plausibility of SNN-based neural networks [4]. In alignment with SNN's conformity to brain physiological principles, this experiment employs a hybrid algorithm combining SNN and small-world network, constructing a mixed neural network. By using Convolution Neural Network(CNN) to process the spike sequences obtained from SNN, it explores the untapped potential of SNN in simulating brain-like neural activities for classification in this field.

2. Dataset

The inner speech dataset, constructed using electroencephalogram (EEG) technology, serves as a basis for brain-computer interfaces aimed at recognizing inner speech. Neural activity, as captured by EEG, offers a non-invasive approach with remarkable temporal resolution. To ensure the integrity of the experiment, ten healthy participants, all native Spanish speakers without any neurological impairments, were engaged in the creation of this dataset.

With a configuration that recorded 128 channels and 1154 samples of signals, the data underwent preprocessing using the MNE library within Python. As illustrated in Figure 1, Participants were tasked to respond to visual cues displayed on the computer screen, with each epoch spanned 4.5 seconds. A triangle pointing in various directions (up, down, left, right) would be displayed in the initial 0.5 seconds, followed by its disappearance after the first second. Participants would then mentally imagine speaking, cease brain activity at 3.5 seconds, and begin to relax [5].

The dataset's nature suggests that the data's accuracy and applicability are contingent on the subject's attention level, contributing to the experimental complexity. This method bears similarity to meditation, where participants must suppress complex brain activities, concentrating solely on visual clues to ascertain corresponding vocal neural signals. Research, including observations of quantitative EEG signal changes through brief meditation exercises among ASD patients [6], reveals that meditation causes consistent alterations in EEG signals, undoubtedly compounding the challenges in data collection.

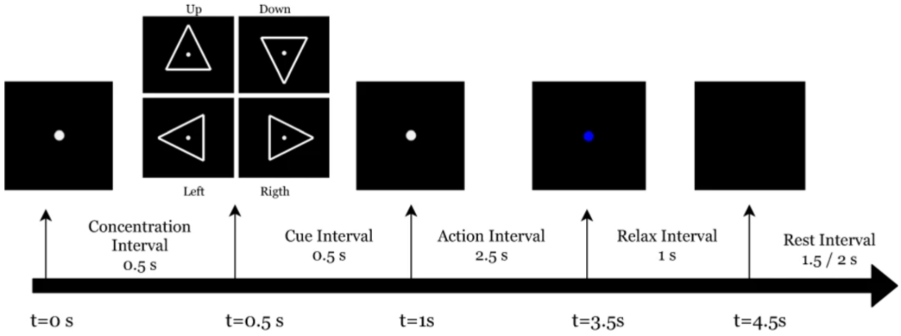

Figure 1. Trail Flow, The screen presented to the participants for each time interval [5].

3. Theoretical Model

3.1. Small-World Network

The small-world network, a structure conceived by Watts and Strogatz in 1998, illustrates a system's pronounced clustering. Dynamic systems that feature small-world coupling exhibit robust propagation capabilities, particularly in modeling infectious disease transmission [7]. Despite most nodes in the small-world model being unconnected, the intervening paths are remarkably short. This concept has implications for understanding the structure of the human brain. Characterized as a multifaceted network with diverse spatial-temporal scales, the brain is discreetly allocated within specific regions. The marked local clustering delineates clear functional divisions across different brain areas, thereby achieving an effective balance. Furthermore, the brain's evolutionary process enhances efficiency, striving to curtail information processing costs. The neuroanatomy of the brain has been confirmed to be replete with sparse and locally clustered neural connections, features that align with the high clustering and abbreviated path length intrinsic to small-world networks [8].

In light of these insights, the present research amalgamates the small world network with SNN for EEG signal analysis, and each neuron in the SNN is coupled with exact coordinate information. The ambition of this work is to emulate neuronal activity in the actual brain as faithfully as possible, thereby enhancing the recognition and categorization of the data.

3.2. Izhikevich model

The Izhikevich model is a computational framework introduced by Eugene M. Izhikevich in 2003 for characterizing neuronal dynamics [9]. Building on the principles of Hodgkin–Huxley-type neuronal models, it translates them into a two-dimensional system of ordinary differential equations, maintaining both the fidelity and precision of the original models. By achieving a harmonious balance between computational efficiency and modeling accuracy, the Izhikevich model is capable of capturing intricate spiking patterns of neurons through elementary mathematical representations. The model's ability to reproduce complex spiking and burst-firing behaviors in cortical neurons with merely four parameters enables the simulation of extensive neural networks, even on standard personal computers.

$$v' = 0.04v^2 + 5\mathrm{v} + 140 - \mathrm{u} + \mathrm{I} \tag{1}$$

$$u' = \mathrm{a}(bv - u) \tag{2}$$

$$\text{if } \mathrm{v} \geq 30\text{mV,\textbackslash then } \begin{cases} \mathrm{v} \leftarrow \mathrm{c} \\ \mathrm{u} \leftarrow \mathrm{u}+\mathrm{d} \end{cases} \tag{3}$$

In equation(1), v denotes the membrane potential, revealing how it evolves over time, u functions as its recovery variable, and “I” represents synaptic currents. Equation (2) employs parameter a to determine the recovery time scale for u and b indicates its sensitivity to v. As specified in equation(3), when v exceeds 30mV, the neuron discharges, leading to a reset of both v and u [9].

In this research, an SNN is architected utilizing the Izhikevich model, leveraging the Bindsnet library. Bindsnet, an open-source PyTorch-based framework, offers an expeditious simulation of spiking neural networks on both CPUs and GPUs. Achieving a remarkable classification accuracy of 95% on the MNIST dataset, Bindsnet's effectiveness in the development of SNNs is underscored [10]. The Izhikevich model's computational elegance, paired with its accurate rendering of diverse spiking phenomena, positions it as a powerful instrument for simulating neural activity within the brain.

4. Algorithm

4.1. Hybrid Neural Network

This study devises and implements a hybrid neural network, taking into account the high dimensionality and complexity of EEG signals. The need to analyze 128 channels with 1154 samples each demands an intricate spatiotemporal analysis. Furthermore, the susceptibility of brain signals to disturbances such as noise and electromagnetic interference, coupled with substantial individual variances, makes conventional linear analysis techniques inadequate for classifying EEG signals. To counter these challenges, the neural network integrates both SNN and CNN to enhance classification performance. Initially, SNN is employed to learn the dataset and generate a new spike sequence by monitoring neuronal activity. Subsequently, a CNN incorporating backpropagation is used to categorize the spike sequences generated by the learned SNN. Backpropagation computes synaptic updates by transmitting error signals through reverse connections, mirroring the brain's natural synaptic modifications for learning and improvement [11]. This innovative hybrid neural network not only builds a physiologically plausible brain-like network for unsupervised learning of genuine biological signals but also ensures the efficacy of supervised learning using backpropagation. The architecture of this hybrid network is depicted in Figure 2.

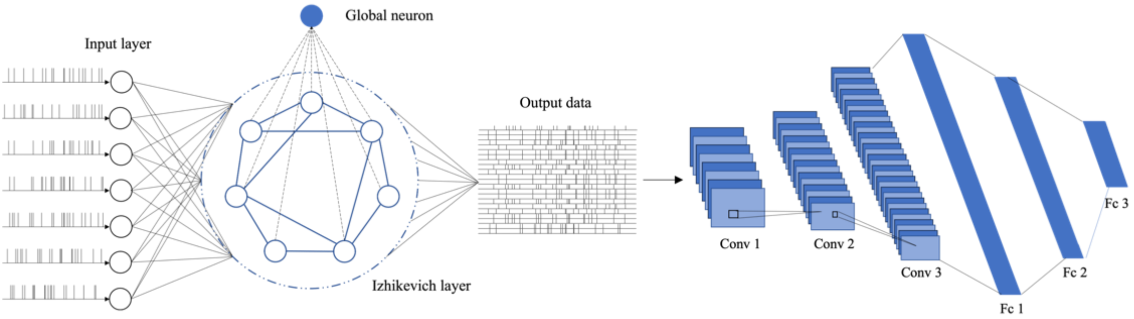

Figure 2. The structure of SWNet.

4.2. Spike Neural Network

In this experiment, the Spiking Neural Network (SNN) constitutes a three-tiered architecture, comprising an input layer, an Izhikevich layer, and a global neuron layer. The input layer, connected to the Izhikevich layer, learns from the input signals via the WeightDependentPostPre learning rule, facilitated by bindsnet. Within the Izhikevich layer, neurons interconnect following the structure of a small-world network. This uniquely crafted small-world network incorporates the three-dimensional coordinate details of High-density EEG electrode placement, reflecting the electrode layout (shown in Figure 3) used in the Inner speech dataset measurement device. This design culminates in a brain-mimicking neural network structure, depicted in Figure 4. Through Hebbian learning, neurons in the Izhikevich layer continually adjust the inter-neuronal weights, enabling unsupervised learning. The third layer, consisting of global neurons, interconnects with each neuron in the Izhikevich layer, overseeing the network's holistic behavior. Additionally, the network is equipped with monitors to capture neural activities, calculating and generating neural spikes of the simulated neurons. The pulse data detected by the global neurons are also integrated as offset into the subsequent classification process, enhancing the network's functionality.

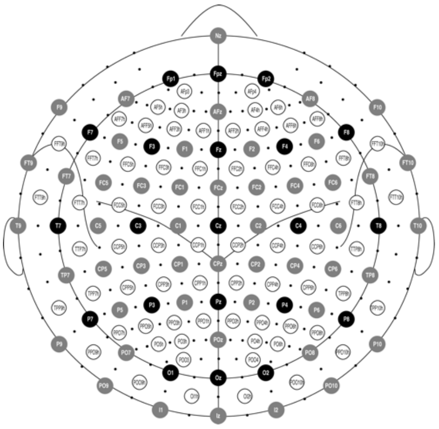

Figure 3. Electrode layout of the 128-channel EEG system [12].

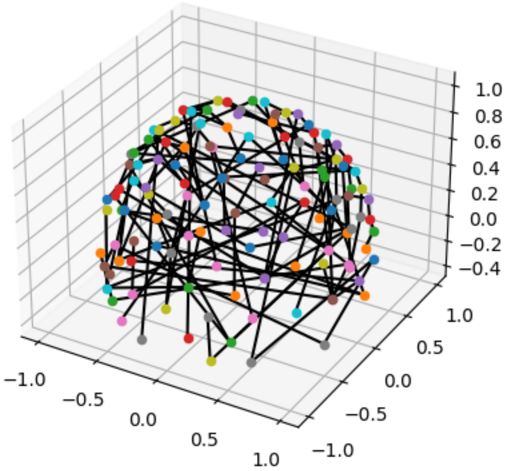

Figure 4. A 3D neural network connection schematic generated based on the electroencephalogram layout.

4.3. Convolution Neural Network

In this study, the designed CNN network classifies structures arising from the co-occurrence functions produced by SNN. It encompasses three convolutional layers, having 8, 16, and 32 filters, respectively. Each convolutional layer is followed by a batch normalization layer, stabilizing learning and minimizing training time. Post convolution, the network employs three sizes of fully connected layers, supplemented with two dropout layers after the first two, utilizing regularization to mitigate overfitting. The ReLU activation function is deployed in every convolutional and fully connected layer, aside from the output layer, infusing non-linearity to aid the network in deciphering complex patterns within the co-occurrence matrix. The Adam is chosen as the optimization method, which is both computationally efficient and straightforward to implement. A hallmark of Adam is its employment of the first and second moment estimates of the gradient, facilitating an adaptive learning rate for individual parameters [13]. In the training process, cross-entropy loss computation is executed in each batch, with weight updates through backpropagation.

5. Results

5.1. Data Preprocessing

Prior to initiating the experiment, it is essential to preprocess the EEG data supplied by the Inner Speech Dataset. This dataset encompasses data from ten distinct volunteers, with each set detailing information across three sessions. Specifically, these datasets record the temporal voltage variations across 128 electrodes. For methodology, the Bens Spiker Algorithm (BSA) is employed to transduce EEG signals into a pulse sequence. BSA evaluates two distinct errors at every instance, as represented by Equation (4) and Equation (5). In these equations, τ denotes time and h represents the filtering function. An analysis of the first error vis-à-vis the differential between the second error and the threshold determines the subsequent steps. Should the first error be inferior to this differential, a pulse is generated and subsequently subtracted from the input using a filter. In the absence of this condition, no action ensues.[14] The experiment avails of the FIR reconstruction filter, establishing an optimal threshold at 0.9550. Within the dataset, the numbers 31-34 denote the Spanish terms "arriba", "abajo", "derecha", and "izquierda". During preprocessing, these denominations are transmuted into a numeric array spanning 0-4 and are cataloged under 'input_label'.

$$\Sigma_{\mathrm{k}=0}^{\mathrm{M}}\mathrm{abs}\big(\mathrm{s}(\mathrm{k}+\tau)-\mathrm{h}(\mathrm{k})\big) \tag{4}$$

$$\Sigma_{k=0}^{M} \mathrm{abs}\big(s(k+\tau)\big) \tag{5}$$

5.2. Evaluating Methodologies and Outcomes

In this study, an SNN grounded on the Small-World Network paradigm is leveraged to emulate the collective activation patterns observed among diverse brain neurons. Input data stems from preprocessed pulse sequences, enabling the scrutiny of interactivity characteristics amongst the 128 simulated neurons. This study introduces and contrasts three distinct methodologies for harnessing the nuances of EEG signal attributes. In the initial methodology, the SNN is operationalized, documenting spike timings for 128 neurons situated within the Izhikevich layer via a spike monitor. The algorithm subsequently generates a co-occurrence function matrix by analyzing pulse time distributions, aiming to encapsulate the concurrent excitation patterns of various neurons. The secondary technique incorporates spatial weighting to the aforementioned co-occurrence function, determining inter-neuronal distances and yielding a corresponding matrix. Elevated spatial weightings proportionally enhance co-occurrence frequencies. Conversely, the tertiary approach bypasses intermediary processes, opting for direct utilization of spike timing data from neurons. This method modulates simulated pulse data across diverse epochs, contingent upon the global neuronal initial excitation's temporal ratio.

The classification precision associated with the initial two methodologies registers below random benchmarks, inhibiting efficacious classification. While the third methodology surpasses this random threshold, its classification efficacy remains suboptimal. For enhanced clarity, outcomes from all three methodologies were visualized. Pertinently, the initial methodology indicates minimal divergences in co-occurrence matrix distributions across varied labels but reveals a general trend of convergence (as seen in Figure 5). The subsequent methodology exhibits pronounced susceptibility to spatial weightings (evident in Figure 6). Unfortunately, both methods fall short of delivering distinct classificatory features, culminating in subpar classification outcomes. Significantly, the 128 neurons roughly demarcate 128 cerebral regions. Empirical evidence corroborates the notion of concentrated brain region involvement during pronunciation, underscoring the potential limitations of emulating co-activation patterns for inner speech classification tasks. In stark contrast, the final methodology, which embarks on a direct temporal analysis of neuronal activations (illustrated in Figure 7), demonstrates commendable efficacy.

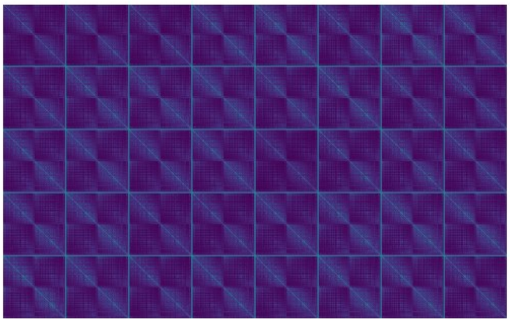

Figure 5. Partial visualization results of the co-occurrence matrix data.

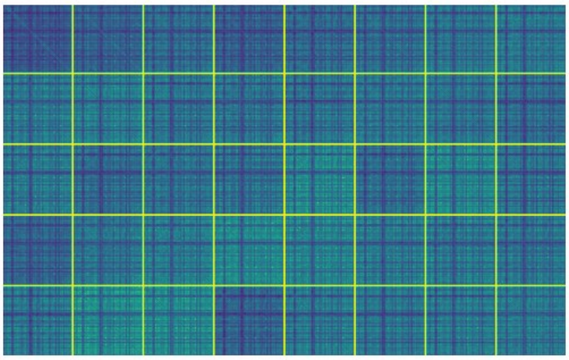

Figure 6. Partial visualization results of the co-occurrence matrix data processed with spatial weighting.

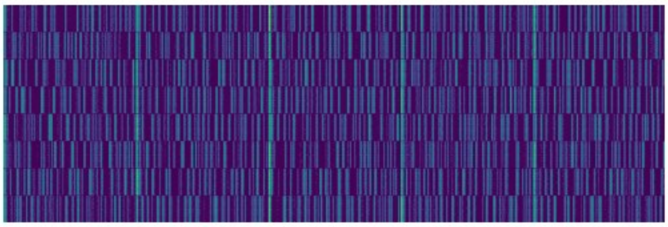

Figure 7. Neural spike data visualization.

5.3. Classification Performance and Analysis

Table 1 presents the performance of SWNet across various subjects. While the inner speech recognition demonstrated an average accuracy of 27.66% across all subjects, there were distinct performances for individual subjects. For instance, Sub 7 exhibited the lowest performance, with an accuracy rate of merely 23.33%. On the contrary, Sub 8 showcased the best performance, reaching an accuracy of 31.50%. For the evaluation of Precision, Recall and F1-score, the average values achieved were 39.33%, 38.90%, and 36.74% respectively.

Table 1. SWNet performance for each subject.

Subject	Accuracy	Precision	Recall	F1-score
1	29.67	43.33	36.39	38.86
2	29.44	38.43	35.65	34.17
3	25.67	30.22	33.06	30.45
4	26.94	44.10	41.32	41.95
5	26.39	40.45	42.18	39.48
6	28.70	40.90	46.67	42.16
7	23.33	39.12	35.88	31.60
8	31.50	46.03	46.33	44.42
9	27.22	33.96	42.74	36.75
10	27.78	36.72	28.75	27.60
Average	27.66	39.33	38.90	36.74

Table 2 offers a comparison of performance across different networks when utilizing all channel data for classification, drawing data from the Imagined Speech dataset.

Table 2. Comparison of SWNet, LSTM and EEGNet.

Classifier	Accuracy
SWNet	27.66
LSTM	27.20[15]
EEGNet	24.90[16]

The outcomes indicate that even if the model's holistic accuracy remains suboptimal, it exceeds mere probabilistic outcomes, highlighting the feasibility and potential of employing hybrid neural networks for EEG signal classification. Historical studies reveal that EEG signals can manifest commendable accuracy in binary classification tasks. For example, research by Li Wang et al. indicated an average accuracy of 66.87% for the imagined speech pertaining to the Chinese characters "left" and "one" [17]. Furthermore, Sereshkeh et al. achieved an accuracy of 54.1% when classifying three states: "yes", "no", and "rest" [18]. It's evident that as the number of categories escalates, the challenge of classifying EEG signals also intensifies.

The inherent challenge in EEG classification revolves around pinpointing specific brain regions activated during language functions. The EEG used in this study, which measures electrical signals across the entire cerebral cortex, is confined to 128 channels, thereby limiting its capacity to flawlessly record intricate brain activities. When expanding the dimensions of the classification tasks, the precision demanded from the data escalates. However, the limited availability of datasets in the domain of EEG-based inner speech classification often leaves experimental needs unmet.

The study predominantly zeroes in on the categorization of imagined speech, deliberately steering clear of deciphering subjects' linguistic intentions by examining unrelated neural activities. The non-invasive nature of EEG enables it to effortlessly capture cerebral signals. Consequently, BCIs implemented using this methodology can seamlessly discern user intentions during routine activities.

6. Conclusion

The present research unveils a composite neural network model, integrating both SNN and CNN functionalities, anchored on the Small-World Network paradigm, tailored for the inner speech quartile classification task. While an average accuracy rate of 27.66% outperforms mere chance, considerable enhancements remain to be addressed. EEG signals are subject to individual variances, however, this study has yet to delve deep into a focused analysis and adaptation of these distinct traits. The analysis encompasses data from all channels, without spefically focusing on brain regions dedicated to language comprehension and production. As future research progresses, there's potential to incorporate mechanisms adept at discerning unique EEG signal patterns from a diverse participant pool, accentuating individual variances to bolster accuracy metrics. Moreover, ensuing efforts will pivot towards employing localized EEG signals with pronounced attributes for a refined classification endeavor, minimizing potential interference from extraneous neural activities.

References

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

[2] Abdulghani, M. M., Walters, W. L., & Abed, K. H. (2022, December). EEG Classifier Using Wavelet Scattering Transform-Based Features and Deep Learning for Wheelchair Steering. In *Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence—Artificial Intelligence (CSCI'22–AI), IEEE Conference Publishing Services (CPS), Las Vegas, NV, USA* (pp. 14-16).

[3] van den Berg, B., van Donkelaar, S., & Alimardani, M. (2021, September). Inner speech classification using EEG signals: A deep learning approach. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)* (pp. 1-4). IEEE.

[4] Loiselle, S., Rouat, J., Pressnitzer, D., & Thorpe, S. (2005, July). Exploration of rank order coding with spiking neural networks for speech recognition. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* (Vol. 4, pp. 2076-2080). IEEE.

[5] Nieto, N., Peterson, V., Rufiner, H. L., Kamienkowski, J. E., & Spies, R. (2022). Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. *Scientific Data*, *9*(1), 52.

[6] Susam, B. T., Riek, N. T., Beck, K., Eldeeb, S., Hudac, C. M., Gable, P. A., ... & Mazefsky, C. (2022). Quantitative eeg changes in youth with asd following brief mindfulness meditation exercise. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 2395-2405.

[7] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440-442.

[8] Bassett, D. S., & Bullmore, E. D. (2006). Small-world brain networks. *The neuroscientist*, *12*(6), 512-523.

[9] Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on neural networks*, *14*(6), 1569-1572.

[10] Hazan, H., Saunders, D. J., Khan, H., Patel, D., Sanghavi, D. T., Siegelmann, H. T., & Kozma, R. (2018). Bindsnet: A machine learning-oriented spiking neural networks library in python. *Frontiers in neuroinformatics*, *12*, 89.

[11] Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020). Backpropagation and the brain. Nature Reviews Neuroscience, 21(6), 335-346.

[12] [12]Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. Clinical neurophysiology, 112(4), 713-719.

[13] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[14] Schrauwen, B., & Van Campenhout, J. (2003, July). BSA, a fast and accurate spike train encoding scheme. In Proceedings of the International Joint Conference on Neural Networks, 2003. (Vol. 4, pp. 2825-2830). IEEE.

[15] Gasparini, F., Cazzaniga, E., & Saibene, A. (2022). Inner speech recognition through electroencephalographic signals. arXiv preprint arXiv:2210.06472.

[16] 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.

[17] Wang, L., Zhang, X., Zhong, X., & Zhang, Y. (2013). Analysis and classification of speech imagery EEG for BCI. Biomedical signal processing and control, 8(6), 901-908.

[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.