Journal of Information Systems Engineering and Management
2025, 10(30s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ECIS: EEG Based Classifier for Inner Speech

Dr. Elizabeth Isaac[1], Leya Elizabeth Sunny[2], Rini T Paul[3], Rotney Roy Meckamalil[4], Rebeka Raju[5], Akash Ajayan[6]

[1 2 3 4 5 6] Department of computer science and engineeering, Mar Athanasius College of Engineering, India

[1]elizabethisaac@mace.ac.in, [2]leyabejoy81@gmail.com, [3]rinitpaul@mace.ac.in, [4]rotney.rotney@gmail.com,

[5]rebekaraju6@gmail.com, [6]akashajayan1123@gmail.com

ARTICLE INFO

Received: 29 Dec 2024

Revised: 12 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Introduction: Inner speech, the silent stream of thoughts, offers a novel means of communication and control, particularly for individuals with motor impairments. The proposed EEG-based Communication and Interaction System (ECIS) leverages transfer learning and signal processing techniques to classify inner speech cues, enabling hands-free interactions. This research explores the usability of inner speech in various domains, including healthcare, education, and entertainment, by analyzing its effectiveness in EEG-based command execution.

Objectives: This study aims to evaluate the feasibility of inner speech as a reliable communication modality using EEG signals. It seeks to classify directional cues (up, down, left, right) from inner, pronounced, and visualized speech while comparing their recognition accuracy. Additionally, the research examines the scalability of ECIS for more extensive datasets and its potential applications for assistive technologies.

Methods: EEG signals were processed through segmentation and transformed using Mean Phase Coherence (MPC) and Magnitude Squared Coherence (MSC) analysis. A transfer learning approach was applied to classify the selected directional cues, and performance was compared across inner, pronounced, and visualized speech conditions. The system was evaluated on a selected dataset, with accuracy comparisons made against existing approaches and simplified cue sets.

Results: The model demonstrated the highest classification accuracy for inner speech at 73.05%, outperforming both visualized and pronounced speech. Accuracy comparisons with previous studies using the same cues showed an improvement in recognition rates. Additionally, using a simpler set of cues resulted in variations in accuracy, highlighting the impact of cue complexity on model performance.

Conclusions: The findings confirm that inner speech is a viable modality for EEG-based communication, offering significant potential for hands-free interaction. ECIS provides a foundation for future research, particularly in developing scalable models with larger, high-volume datasets. This work paves the way for enhanced human-computer interaction, benefiting individuals with disabilities and expanding applications in various fields.

Keywords: EEG, Mean Phase Coherence, Magnitude Squared Coherence, Transfer learning

INTRODUCTION

Electroencephalogram monitoring measures the brain's electrical activity through a series of electrodes placed on the scalp. This method provides valuable insights into the functioning of the brain and has been used extensively in various fields such as medicine, psychology, and neuroscience (Maganti & Rutecki, 2013) (Meng, Zhang, Ma, Gao, & Kong, 2023) (O'Shea, Lightbody, Boylan, & Temko, 2017) (Ullah, Hussain, ul Haq Qazi, & Aboalsamh, 2018). EEG monitoring has been used for emotion classification, in the treatment of ADHD, understanding stress and

attention levels and other neurological phenomenon directly and indirectly. Inner speech is the thought process or speech inside the mind, when mentally articulating words or imagining actions. There are various attempts to classify inner speech using the EEG data obtained during the process (Gasparini, Cazzaniga, & Saibene, 2022) (Panachakel & Ganesan, 2021) to name a few. Some of these methods use similar methodologies used for classifying emotions from EEG data. By analyzing the patterns of electrical activity in the brain, it is possible to data on mental processes, including inner speech. The healthcare, education, and entertainment sectors stand to gain much from the smooth integration of inner speech categorization technologies. By enabling hands-free device operation and information retrieval, it creates opportunities for the creation of more inclusive and accessible systems that improve people's quality of life and contribute to societal well-being.

Classifying inner speech from EEG data is relatively more difficult compared to using EEG data for identifying emotions. While emotion classification can be done to certain degree of accuracy by passing raw data to basic neural networks, similar methods might not yield good results when used for inner speech classification. It can be due to the overall lack of information in EEG that can be used for inner speech classification. There have also been differences on opinion on selection of channels, with increase in accuracy sometimes of increase in overall number of channels and sometimes on selection of a suitable subset of channels. Using techniques like SVM, XGBoost and even neural networks like LSTM and BiLSTM (Gasparini et al., 2022) might yield accuracy values that are borderline random. Various processing or transformation of the raw data can be done to produce more suitable data for training models. Transforming the data to the frequency domain is possible method that is shown to increase the accuracy of classifying other information using EEG data. More mathematical transformations including calculation of Mean Phase Coherence and Magnitude Squared Coherence (Panachakel & Ganesan, 2021) can be made and their properties can be made use of to increase the context of the input within the same input shape.

ECIS investigates the variation in the accuracy of thought under different conditions including inner speech where each participant imagines their voice giving direct commands to device, pronounced speech where the participant tries to pronounce the cues and visualized speech in which the participant mentally imagines moving a circle in the direction of the cue. Mean Phase Coherence and Magnitude Squared Coherence(citing transfer) and their inherent symmetry allows the construction of image counterparts of the EEG data which can be used for classification. "Thinking Out Loud" dataset (Nieto, Peterson, Rufiner, Kamienkowski, & Spies, 2022) used of this study contains the inner speech of 10 subjects through three sessions. It has 128-channel data with four directional classes of up, down, left and right. The dataset primarily gives importance to inner speech, with pronounced speech having lesser number of trials recorded. The variation of the performance degree of similarity of cues were also analyzed, supported by previous works employing similar methods (Panachakel & Ganesan, 2021). This will provide a foundation for developing systems to classify EEG data under larger set of cues, on the availability of higher volume datasets.

OBJECTIVES

The purpose of this study is to investigate whether inner speech may be used as a hands-free communication medium with EEG signals. The study uses preprocessing techniques like segmentation and coherence evaluation through Mean Phase Coherence (MPC) and Magnitude Squared Coherence (MSC) to classify directional cues (up, down, left, and right) from EEG signals associated with inner, pronounced, and visualised speech. To improve classification accuracy, a transfer learning strategy is used, and its effectiveness is evaluated under various speech situations. The study also looks at the impact of analysing differences in model accuracy using a more basic set of directional signals. The study also looks into how ECIS might be used in assistive technology, specifically in the fields of entertainment, education, and healthcare. Finally, the study aims to assess the scalability of the system for larger datasets and provide insights into improving EEG-based human-computer interaction for accessibility and usability.

METHODS

Training an advanced inner speech classification model is the goal, and it will be done using 'Thinking Out Loud' that was acquired by using 128 EEG channels.

DATASET EMPLOYED IN THE WORK

The dataset utilized in the study, known as "Thinking Out Loud" (Nieto et al., 2022) plays a crucial role in the creation of the EEG-based inner speech classification model. Using the BioSemi ActiveTwo high-resolution bio-potential measurement equipment, 128-channel EEG recordings were acquired for the dataset, which aims to capture the brain activity related to inner speech creation. In order to avoid actual physical articulation, participants were taught to mentally simulate pronouncing directional commands such as up, down, left and right. In order to gain important insight into the neurological correlates of cognitive processes without overt speech production, this imagery task sought to extract the brain activity patterns associated with inner speech production. Specific conditions were set to further analyse in detail how variations in the procedure applied during thought can affect the result of the classification.

To guarantee that participants consistently visualized the directional signals at predetermined intervals, timed beats and visual cues on a screen directed them throughout the data collection procedure for certain subset of trails. The consistency of the participants' cognitive activities was made possible by the standardizing the procedure, which improved the repeatability and dependability of the EEG data that was gathered. Additionally, Independent Component Analysis (ICA) was used to remove undesired artifacts, including noise, in order to improve the quality of the EEG recordings. Through careful preprocessing of the obtained EEG signals and removal of highly coherent parts, the dataset was fine-tuned to highlight the key brain patterns linked to inner speech, providing a strong basis for further analysis and classification tasks.

Table 1. Types of trials in the dataset

Trial Type	Description
Inner Speech	Each participant imagines their voice giving direct commands to device
Pronounced Speech	Each participant tries to pronounce the cue
Visualized Speech	Each participant mentally imagines moving a circle in the direction of the cue

FEATURE EXTRACTION

Since the final structure of the data that is used for classification is an image, it was preferred that all channels of the data be kept for maximum resolution in the image generated. The final image array is composed of two features:

1. Mean Phase Coherence

2. Magnitude Squared Coherence

Mean Phase Coherence

Mean Phase Coherence (MPC) is a measure used in EEG signal analysis to assess the synchronization or phase consistency between pairs of EEG channels, expressed as Equation (1). It quantifies the degree of phase locking or coordination of neural activity between different brain regions. MPC is particularly useful for studying functional connectivity and neural communication patterns during cognitive tasks such as inner speech processing.

$$MPC_{i,k} = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{-j(\Phi_i(n) - \Phi_k(n))} \right| \quad (1)$$

Magnitude-Squared Coherence

It measures the linear relationship between the channels in the spectral domain, using respective power spectral densities S(ω). It lies between 0 and 1, captured using Welch's periodogram with hamming window. The MSC between i and k channels are given by Equation (2).

$$MSC_{i,k}(\omega) = |S_{i,k}(\omega)|^2 / S_{i,i}(\omega) S_{k,k}(\omega) \quad (2)$$

FINAL FORMATTING OF DATA

Mean Phase Coherence (MPC) and Magnitude Squared Coherence (MSC), two retrieved characteristics, must be carefully transformed into a structured array that can be used as the classification model's input in the final formatting of the data in EEG signal processing for inner speech classification. By creating a thorough representation of the brain dynamics collected by MPC and MSC, this crucial step will improve the model's capacity to reliably and effectively classify inner voice commands. Researchers establish a unified representation of brain activity patterns by merging MPC and MSC values, which reflect neuronal synchronization and coherence. By utilizing the complimentary information offered by both characteristics, researchers are able to improve the input data for the classification model.

The MPC and MSC values are subjected to a band pass filter, which limits the data to particular frequency bands like the alpha, beta, and gamma ranges in order to assure data quality and concentrate on pertinent frequency ranges. In order to prepare the data for further processing, this filtering phase improves the neural information pertinent to inner speech classification. After that, the structured data is grouped into an array that resembles an image, with the MPC and MSC values placed in a certain way to maintain the spectral and spatial correlations between the EEG channels. A single array encapsulating the combined neuronal coherence and synchronization information in a format appropriate for deep learning-based classification models is constructed by superimposing the upper triangular portion of the MPC values and the lower triangular part of the MSC values. The complex brain dynamics involved in inner speech activities are captured by this image-like representation, which guarantees that the crucial information represented as the MPC and MSC features is retained. After formatting, it can be used to classify the features that were extracted using inner speech commands accurately. This method helps the model learn and differentiate between various directed signals with high accuracy.

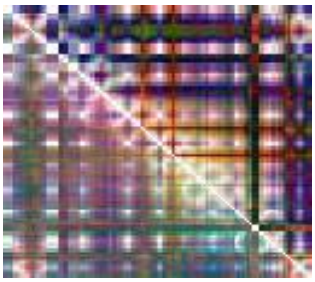

Figure 1. The final array created by superimposing the lower triangular part of the MSC and upper triangular part of the MPC.

METHOD USED TO INCREASE EFFECTIVE DATASET SIZE

In an effort to increase the volume of available samples, data augmentation strategies have been investigated in response to the problem of limited data in training datasets(Nieto et al., 2022), a number of data augmentation techniques have been investigated, including overlapping or sliding window approaches (O'Shea et al., 2017) (Kwak, M"uller, & Lee, 2017) (Ullah et al., 2018) (Majidov & Whangbo, 2019) and generative adversarial networks(GAN) (Luo & Lu, 2018) (Wei, Zou, Zhang, & Xu, 2019) (Chang & Jun, 2019). However, because of the restricted amount of data available, GANs are not thought to be the best solution for our particular issue. Also, the possibility of overfitting the model to the training data was identified when sliding windows were used.

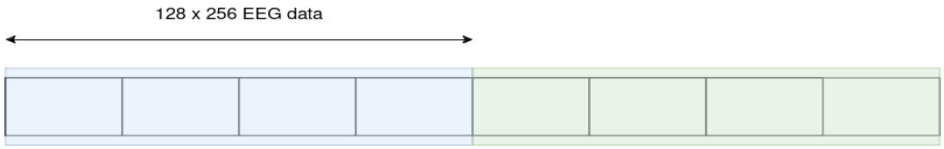

Figure 2. Data augmentation with non-overlapping windows.

DETAILS OF THE CLASSIFIER

A ResNet101 neural network, with its weights trained on the ImageNet dataset is used for transfer learning. For the purpose of this classification, the preexisting layers were frozen and to be used only for feature extraction. Additional set of two convolution layers along with pooling were added for the purpose of classification and dense layers to generate a classifier for two classes. Multiple iterations of the model were made based on different set of cues.

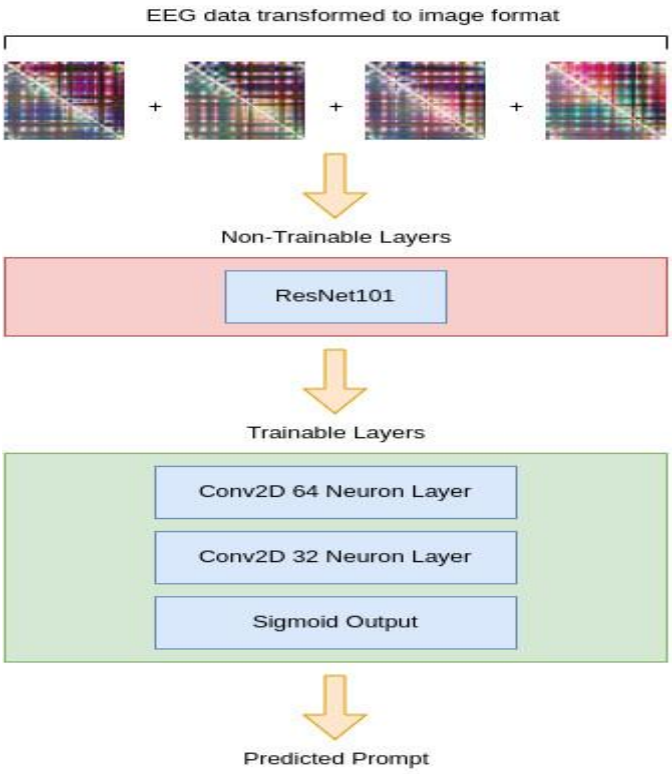

Figure 3. Description of the layers of the model.

A learning rate of 1e −4 using the Adam optimizer is used for training the model. A 10 fold cross-validation over 20 epochs is carried out. The accuracy metric was measured as the average of the accuracy values during each crossvalidations and other metrics and convolution matrix was created using separate data kept for evaluation which was in the ratio of 1:18 to the overall dataset. The smaller amount of data for evaluation is due to the overall decrease in training data size on allotting more data for evaluation. Furthermore, the model was run on AWS Sagemaker ml.g4dn.xlarge notebook instances.

EXPERIMENTS

Multiple experiments were conducted using the available EEG data. Initially, the raw EEG data was used for the purpose of classification. The data, fed to a 1D model of the specifications listed in the previous section. The raw data was segmented using the MNE(Gramfort et al., 2013) library, in the useful region from raw data i.e., 1.5.-3.5 sec was segmented. Since the dataset consists of visualized, inner and pronounced speech, for the purpose of the experiment, only the inner speech data was considered. Further, to use the model with reasonable accuracy and EEG data being complex, to reduce the overall ambiguity in classifying the data, only the up and down prompts were used. This method was mostly used to infer how much the CNN model could infer from the raw data and to essentially give a base to how much relatively the other models are performing. The second experiment involved processing the EEG data into image-like arrays with the superimposed Mean Phase Coherence and Magnitude-squared coherence as mentioned in the methodology. This, along with the reasonable performing ResNet101

model with weights trained from ImageNet with additional convolution layers for classification. Finally, the data after segmentation was used as increasing the amount of data is essential to increase the performance of this model. The third experiment involved making making multiple comparisons on (up, down) and (left, right) set of cues and the fourth experiment involved comparisons on how the performance of the model is different on inner speech, visualized speech and pronounced speech.

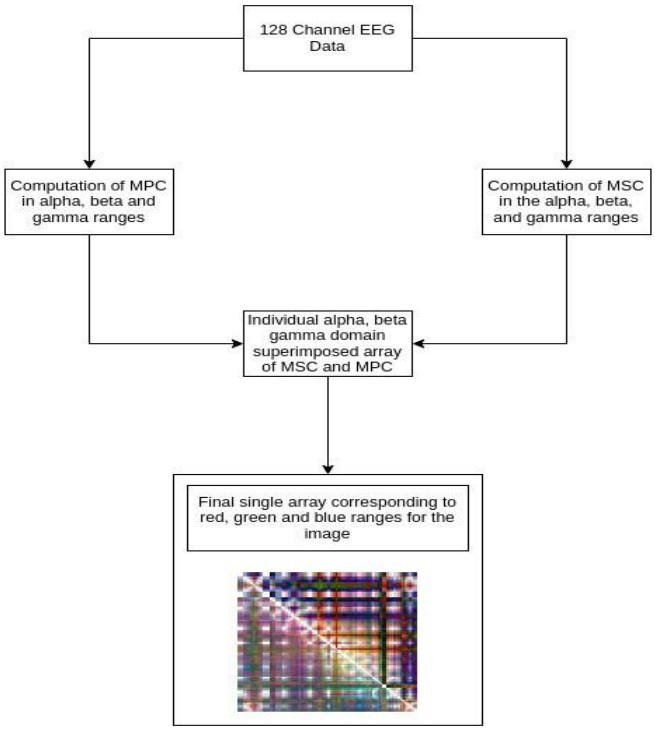

Figure 4. The process of creating the image for the classifier. The end image is of 128 x 128 x 3 dimensions.

RESULTS

On comparing the performance of the model on pronounced, visualized and inner speech, the model performed best when inner speech was used. The relative drop in performance in pronounced speech can be related to the comparatively less amount of trials of pronounced speech that was recorded.

Table 3. Metrics for inner speech considering a small subset of data from all subjects for evaluation

Prompt	TP	FP	TN	FN	Precision	Recall
Up	69	7	48	0	90.79	1
Down	48	0	69	7	1	87.27

Table 4. Metrics for visualized speech considering a small subset of data from all subjects for evaluation

Prompt	TP	FP	TN	FN	Precision	Recall
Up	63	24	38	1	98.44	72.41
Down	38	1	63	24	97.44	37.62

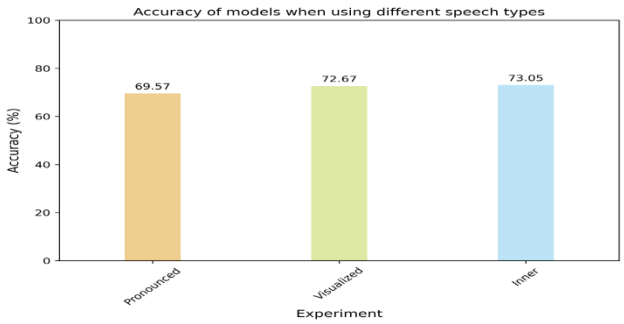

Figure 5. Accuracy of models when using different speech types.

Table 5. Metrics for pronounced speech considering a small subset of data from all subjects for evaluation

Prompt	TP	FP	TN	FN	Precision	Recall
Up	26	0	34	2	94.44	1
Down	34	2	26	0	1	94.44

DISCUSSION

When using the left and right cues instead of up and down, the model performed slightly better although the verbal similarity is higher for the latter case. It could be due to the channel wise correlation to the hemispheres of the brain. Further comparison with works using long words(independent and cooperate) and long-short words(in and cooperate) shows how the similarity between words affects the ability of the model for distinguish the cues(Panachakel & Ganesan, 2021) as in Fig.10. The proposed methods were used on the partial up and down subset of the data. The accuracy obtained initially bordered the probability of random guess at 50%. Later, further fine-tuning of the model for the first experiment allowed further increase in accuracy to 58%. The model peaked at this amount number, which can be due to the considerably low amount of data available for classification. However, using the ResNet + Transfer Learning model allowed accuracy values up to 78% with is reasonably good when compared to previous works using the same dataset. After segmentation, the performance of the model dropped to 73%. But it can be attributed to better adjustment of the model to unseen data and the ability of the model to use 1 second EEG data instead of 2 seconds thereby increasing the practical usability of the model for real-world use cases. It is to be noted that the current work uses only two of the cues and the work used for comparison used four classes, thus the baseline accuracy has been shifted to adjust for the difference in classes(Gasparini et al., 2022).

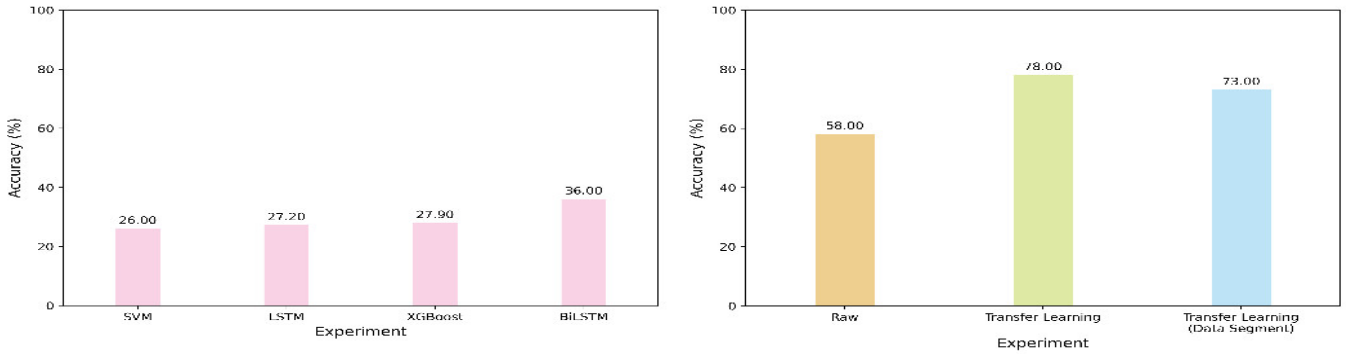

Figure 6. Accuracy comparison of the models (separate graphs are provided to represent the change in baseline).

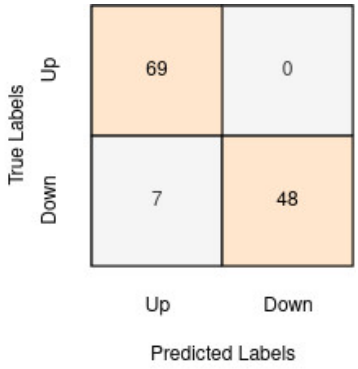

Figure 7. Confusion matrix for (Up, Down) cues for inner speech.

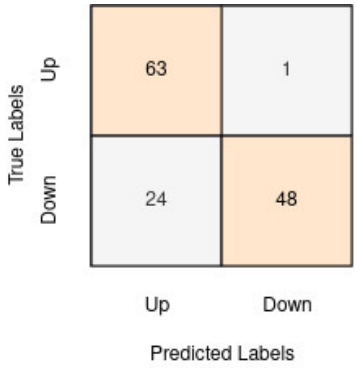

Figure 8. Confusion matrix for (Up, Down) cues for visualized speech.

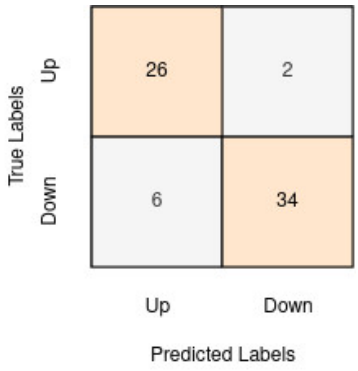

Figure 9. Confusion matrix for (Up, Down) cues for pronounced speech.

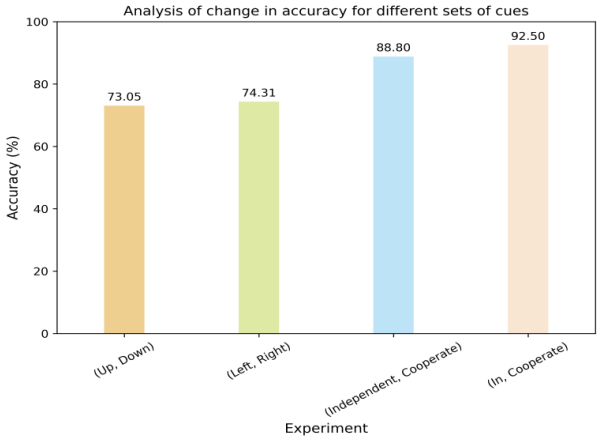

Figure 10. Analysis of change in accuracy for different sets of cues.

CONCLUSION

The paper uses a ResNet+Transfer Learning model which uses Mean Phase Coherence and Magnitude Squared Coherence for transforming the EEG data into image-like format for classifying various types of imagined cues. On the available set of data, data segmentation methods were used to increase the quantity of data available for training the model. On comparison with previous works, the model performs reasonably well. The performance of the model could be further improved using larger datasets On increasing the number of classes, similar methodology can be used for creating mobility devices like wheel-chairs for people. It could also be used in the gaming industry for providing more realistic experiences when paired with VR devices.

REFRENCES

[1] Chang, S., & Jun, H. (2019). Hybrid deep-learning model to recognise emotional responses of users towards architectural design alternatives. Journal of Asian Architecture and Building Engineering, 18(5), 381–391.

[2] Gasparini, F., Cazzaniga, E., & Saibene, A. (2022). Inner speech recognition through electroencephalographic signals. Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., . . . Hamalainen, M. (2013). Meg and eeg data analysis with mne-python. Frontiers in Neuroscience, 7.

[3] Kwak, N.-S., M"uller, K.-R., & Lee, S.-W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PLoS One, 12(2), e0172578.

[4] Luo, Y., & Lu, B.-L. (2018). Eeg data augmentation for emotion recognition using a conditional wasserstein gan. In 2018 40th annual international conference of the ieee engineering in medicine and biology society (embc) (p. 2535-2538)

[5] Maganti, R. K., & Rutecki, P. (2013, Jun). Eeg and epilepsy monitoring. Continuum (Minneapolis, Minn.), 19(3 Epilepsy), 598–622.
[6] Majidov, I., & Whangbo, T. (2019). Efficient classification of motor imagery electroencephalography signals using deep learning methods. Sensors, 19(7).
[7] Meng, M., Zhang, Y., Ma, Y., Gao, Y., & Kong, W. (2023, feb). Eeg-based emotion recognition with cascaded convolutional recurrent neural networks. , 26(2), 783-795.
[8] Nieto, N., Peterson, V., Rufiner, H. L., Kamienkowski, J. E., & Spies, R. (2022). Thinking out loud, an open-access eegbased bci dataset for inner speech recognition. Scientific Data, 9(1), 52.
[9] O'Shea, A., Lightbody, G., Boylan, G., & Temko, A. (2017). Neonatal seizure detection using convolutional neural networks. In 2017 ieee 27th international workshop on machine learning for signal processing (mlsp) (p. 1-6).
[10] Panachakel, J. T., & Ganesan, R. A. (2021). Decoding imagined speech from eeg using transfer learning. IEEE Access, 9, 135371-135383.
[11] Ullah, I., Hussain, M., ul Haq Qazi, E., & Aboalsamh, H. (2018). An automated system for epilepsy detection using eeg brain signals based on deep learning approach. Expert Systems with Applications, 107, 61- 71.
[12] Wei, Z., Zou, J., Zhang, J., & Xu, J. (2019). Automatic epileptic eeg detection using convolutional neural network with improvements in time-domain. Biomedical Signal Processing and Control, 53, 101551