JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Investigating EEG Spectral Signatures for Decoding Envisioned Speech Recognition

[1]Sujal, [2]Sandeep Kumar
[1,2]Electronics and Communication Engineering,
[1,2]National Institute of Technology Delhi
Delhi, India

Abstract : Automatic speech recognition (ASR) interfaces are becoming more and more common in daily life for electronic device control and interaction. Many users, including those with severe privacy requirements, paralysis, locked-in syndrome, speech disorders, and others, cannot use the interfaces that are now in use. In these situations, an interface that uses electroencephalogram (EEG) signals to identify imagined speech can be quite helpful. Numerous studies have been conducted in the past to address this issue. Nevertheless, some research has been done to pinpoint the EEG signal's frequency bands (δ, θ, α, β, and γ) that are important for the intended speech recognition. Thus, the goal of this work is to examine the role that various EEG frequency bands and signals originating from various brain lobes play in the recognition of imagined speech. A spatiotemporal deep learning architecture with a convolutional neural network (CNN) and long short-term memory (LSTM) has been used to extract various frequency bands from the signals received from various brain lobes. A publicly accessible dataset with three classification tasks—digital, character, and image—is used to assess the performance. For each of the three challenges, we achieved a classification accuracy of 85.713%, 87.276%, and 85.935%.

IndexTerms **- Convolutional Neural Network, Long short-term memory, ASR (Automatic speech recognition)**.

I. INTRODUCTION

EEG-based speech imaging recognition is an innovative approach that utilizes electroencephalography signals to capture and analyze neural activity related to speech production and perception [1]. The field of brain-computer interfacing is interested in the investigation of EEG frequency bands for imagined speech recognition. The goal of brain-computer interfaces is to improve brain-computer or brain-machine connection and communication. By using real-time digital signal processing techniques to convert brain inputs into actions, these interfaces avoid the conventional neuromuscular pathways.

Frequency analysis is one method for examining EEG signals in order to recognize speech. Based on signal frequencies ranging from 0.1 Hz to more than 100 Hz, these signals are commonly classified as delta, theta, alpha, beta, and gamma [2]. Through an examination of the trends and variances in these frequency ranges during speech-related brain activity, scientists can create models and algorithms for speech recognition. In order to convert the brain impulses connected to the imagined speech into understandable speech commands or text, these algorithms and models can subsequently be applied. Numerous EEG signal processing techniques, including wavelet transform, autoregressive approaches, rapid Fourier transform, eigenvector methods, and time-frequency distributions, can be used to achieve this strategy. Through the application of these techniques, pertinent information can be extracted from the EEG signals and utilized by classification algorithms to precisely identify and comprehend spoken commands. With the use of a communication technology called the Brain-Computer Interface (BCI), people may communicate with the outside world without depending on their peripheral nerves and muscles [3][4]. To improve the efficiency and precision of brain-computer interfaces in comprehending and translating the intents of people who are unable to communicate orally, the analysis of EEG frequency bands for envisioned speech recognition is being done. A thorough analysis and comparison of the benefits and drawbacks of various EEG signal processing techniques is necessary for the intended speech recognition application. Finding the best feature extraction technique based on performance and enhancing the general quality of life for people with speech impairments are the two main goals of this investigation. EEG frequency bands are essential for analyzing brain signals in the context of future brain-computer interfaces that aim to recognize speech. They make it possible to identify particular waveforms within various frequency ranges, which is useful for creating speech recognition models and algorithms. The brain impulses connected to the imagined speech can then be deciphered and translated into understandable speech commands or text using these algorithms and models. The need for organic methods of interacting with electronic devices has grown with the ubiquity of electronic devices. As a result, human-computer interfaces that rely on gestures and speech have significantly improved. Nonetheless, there are several circumstances when the application of these mediums is restricted. Speech impairments, conditions like locked-in syndrome, or circumstances where the person values their privacy above all else may be the cause of this. Imagined speech recognition systems based on non-invasive electroencephalography could be a fascinating substitute in certain situations. Many Brain Computer Interfaces have been

proposed over the years, including ones based on the concepts of text entry and motor movement imagination. However, the majority of these concentrate on a user who must concentrate on a specific image or letter for a considerable amount of time in order to pick it or visualize a specific movement in order to provide orders. Through practical brain–computer interfacing (BCI), a person can use neural signals that emerge from the cerebral cortex of the brain to communicate with surrounds or other equipment [5][6]. A more natural method of interacting with electronic gadgets can be offered to a user via a BCI framework that recognizes imagined speech [4].

II. BACKGROUND

A variety of methods exist for measuring brain activity, from non-invasive electroencephalograms (EEG) to invasive procedures like surgically implanted electrodes. In this work, we apply the EEG technique to gather non-invasive brain data. The simplicity of EEG and the fact that it causes little to no discomfort for the user have made it popular despite its low spatial resolution.

Attempts have been made to identify imagined speech using EEG at many levels, including word, syllable, and vowel imagination [7]. We first applied Discrete Wavelet Transform was used to extract features after the Signal to Noise Ratio was improved using the Common Average Reference technique. After that, the retrieved characteristics were fed into an Long Short Term Memory for classification [8]. Most of the currently envisioned speech recognition algorithms rely significantly on the meaningful handmade feature extraction, which necessitates deep domain expertise. As deep learning made progress, researchers shifted their focus to automatically extracting features from unprocessed EEG signals. Due to the way the signals were collected, discriminatory information regarding the classification task was contained in the features. By building a 2D mesh based on the spatial information obtained from the electrode distribution, a two-dimensional Convolutional Neural Network was used to process the EEG signals [9].

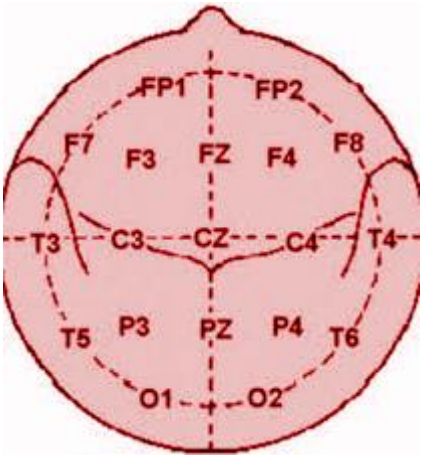
This paper examines the role that several EEG frequency bands (δ, θ, α, β, γ) and signals obtained from distinct brain lobes (frontal, temporal, occipital, and parietal) play in the recognition of imagined speech using EEG data. Feeding a one-dimensional Convolutional Neural Network - Long short-term memory-based architecture are these bandpass filtered (BPF) signals [5][8]. While the LSTM block is utilized to extract temporal information, the one-dimensional Convolutional Neural Network block is intended to extract spatial features from the signals [10]. Additionally, we pick particular brain lobes, feeding signals to the classifier just from those lobes in order to acquire the classification performance. We utilize a band reject filter (BRF) to remove the frequency bands θ and α from the classifier feed because we have seen that these bands contribute the least to the classification of imagined speech. It is demonstrated that this method outperforms using the complete raw EEG data in terms of classification accuracy and precision.

.

This is required to decrease implementation complexity, lower information processing costs, and eliminate the possibility of information compression. Through the examination of distinct frequency bands within EEG data, scientists are able to derive pertinent characteristics that facilitate the identification and comprehension of imagined speech. Many techniques, including temporal frequency distributions, rapid Fourier transform, eigenvector approaches, wavelet transform, and auto-regressive method, have been used in recent years to extract characteristics from EEG data [8].

With the use of these techniques, the EEG signal may be thoroughly analyzed, revealing information on the frequency composition and temporal dynamics of imagined speech. Furthermore, feature extraction from nonstationary EEG data is a particularly good application for time-frequency domain techniques like wavelet transform.

Delta δ (0.5-4 Hz): Delta waves are typically associated with deep sleep and unconscious processes. While they may not directly contribute to speech recognition, their presence or absence can indicate the overall state of arousal or attentiveness, which can influence speech processing. Theta θ (4-8 Hz): Theta waves are prominent during states of focused attention and memory retrieval. In the context of speech recognition, increases in theta power have been linked to enhanced processing of auditory stimuli and encoding of speech information. Therefore, analyzing theta activity can provide insights into the listener's engagement with speech stimuli. Alpha waves, or α waves (8–12 Hz), are noticeable when the brain is in a wakeful, relaxed state. Examples of these situations include meditation and closed eyelids without sleep. They also seem to be linked to higher levels of creativity and lower levels of arousal. Beta waves, or β waves, occur between 12 and 30 Hz and are indicative of cognitive activity, problem-solving, decision-making, or concentrated concentration. They are usually connected to mental activity and attentiveness. The fastest brainwaves, known as gamma waves γ (30-100 Hz), are linked to higher order cognitive processes including memory, perception, and awareness. Additionally, they have a role in functions including sensory binding and attentiveness.



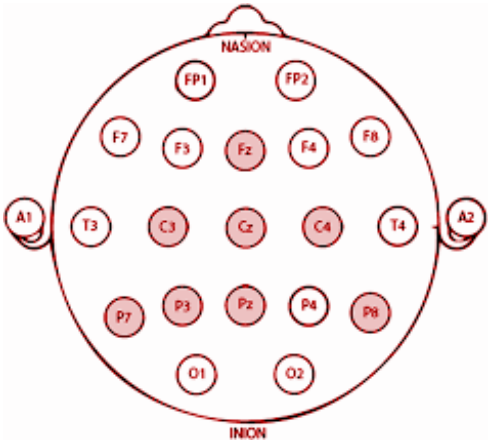


Fig. 1 and 2. Diagrams showing where the electrodes will be placed to gather data. The illustrations are based on [11] and [12] respectively.

III. METHODOLOGY - DATASET

Utilizing recordings from a participant between the ages of 25 to 45, we utilize an imagined speech dataset that is accessible to the public. There are 15 channels on the headset: C3, C4, F3, F4, F7, F8, FP1, FP2, T6, T5, P3, CZ, P4, O1, and O2. According to the International System of Electrode Placement, the electrodes were positioned at the subject's scalp. These electrodes were positioned above the ears, along with two additional reference electrodes, a driven right leg, and common mode sensation. Prior to being down sampled to 150 Hz, the signals were captured at 2.5K Hz sampling frequency.

The subject is shown an object on the screen. After that, the patient is told to close their eyes and visualize the thing they just saw for ten seconds while at rest. To make sure the subject goes back to rest before showing the following item, a 20-second pause is maintained before presenting the participant with the next image. In this way, the subjects were presented with three types of cues, and the matching EEG was recorded.

First, there are nine numbers in total. Secondly, there are ten uppercase English alphabets: A, B, C, D, E, V, W, X, Y, and Z. Items used in daily life are included in the last category: a table, a mango, a squirrel, a butterfly, juice, a pencil, a laptop, a helicopter, a kite, and carpet. 30 EEG recordings total—10 recordings for each category, or 10 classes total—have been gathered from the participants.

IV. METHODOLOGY - EXPERIMENTAL SETUP

Using the three subsets of the provided dataset—digits, characters, and images—classification was carried out. Ten seconds make up the whole EEG signal that was captured for a given class. For classification, the one-dimensional Convolution Neural Network Long Short-Term Network model receives the filtered signals independently [13]. We can identify 8 channels from the 14 electrodes where the EEG signal was captured, which relate to the frontal lobe, and 2 channels that correspond to the temporal, parietal, and occipital lobes.

EEG signals from each of these distinct lobes are also utilized independently for categorization, both in their raw form and by extracting information from various frequency bands.



Fig. 3. Envisioned speech recognition recording setup

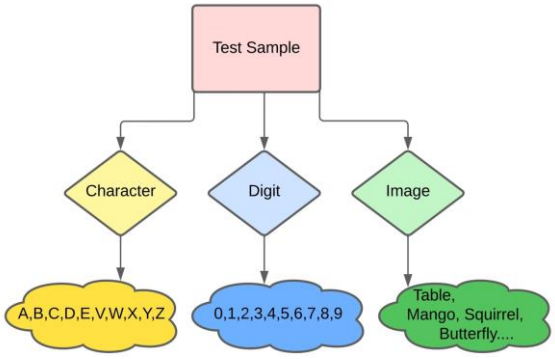


Fig 4. Tasks performed [14]

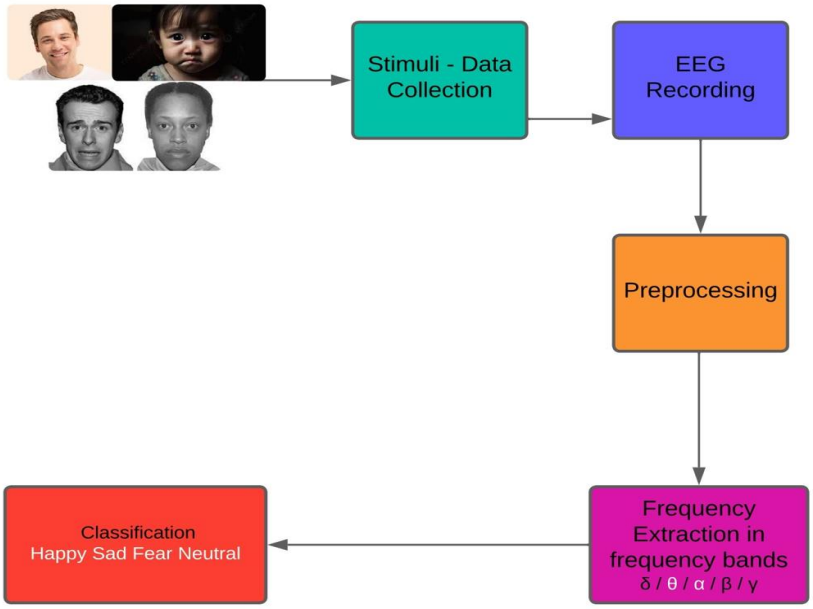


Fig 5. Flow Diagram [15]

V. RESULTS AND DISCUSSION

The performance of EEG data collected from various brain lobes and bandpass filtered for various EEG frequency bands is listed in Table I.

Regarding the various frequency bands, it is evident that the low frequency δ band has the most amount of information, succeeded by the γ and β bands. When it comes to performance, the θ and α bands fall well short of the other frequency bands. Therefore, we see an improvement in classification performance when the signal is bandpass filtered and the θ and α frequency bands are eliminated. This fits in well with the research on various frequency bands and how frequently they occur in relation to the work at hand. Additionally, it is evident that the frontal lobe houses the majority of information, which explains why signals evaluated only in this region have a noticeably high accuracy.

TABLE 1 - RECOGNITION ACCURACIES FOR DIGIT TASK

Part | Band	δ	θ	α	β	γ	δ + β + γ	All
Frontal	0.6929	0.3281	0.265	0.5825	0.6317	**0.74**	0.7244
Temporal	0.2019	0.1148	0.1178	0.2825	0.3834	0.4694	**0.4982**
Occipital	0.1248	0.102	0.103	**0.275**	0.2207	0.1266	0.2213
Parietal	0.1426	0.1054	0.1128	**0.3498**	**0.3438**	0.2916	0.299
All	0.8433	0.501	0.5359	0.6769	0.7471	**0.8591**	**0.8584**

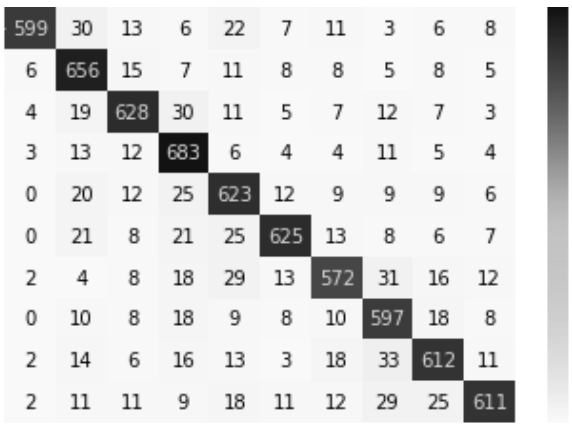


Fig. 6. Confusion matrix for the digit task.

Tables II and III display the outcomes of tests conducted on the dataset's picture and character tasks. The patterns seen in these two activities are comparable to those seen in the task of classifying numbers.

TABLE 2 - RECOGNITION ACCURACIES FOR CHARACTER TASK

Part | Band	δ	θ	α	β	γ	δ + β + γ	All
Frontal	0.6976	0.344	0.39	0.568	0.6434	**0.7875**	0.752
Temporal	0.1557	0.1034	0.1079	0.315	0.393	**0.522**	**0.5221**
Occipital	0.139	0.1085	0.1263	0.302	**0.346**	0.1867	0.2525
Parietal	0.138	0.109	0.114	0.319	**0.3275**	0.2447	0.289
All	0.851	0.541	0.131	0.6875	0.72	0.8729	**0.8738**

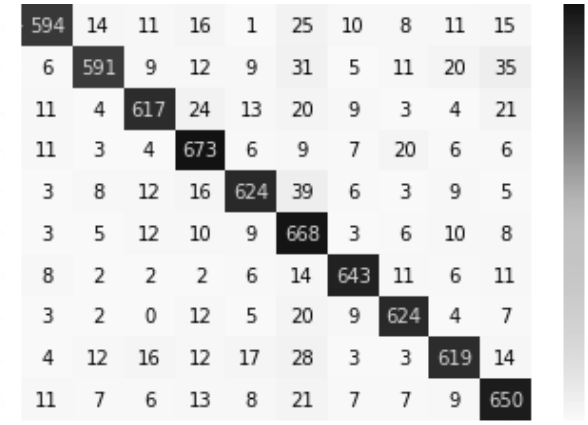


Fig. 7. Confusion matrix for the character task.

TABLE 3 - RECOGNITION ACCURACIES FOR IMAGE TASK

Part | Band	δ	θ	α	β	γ	δ + β + γ	All
Frontal	0.7155	0.202	0.3759	0.5635	0.6236	**0.804**	0.7596
Temporal	0.1904	0.102	0.1299	0.3145	0.373	**0.5155**	0.4968
Occipital	0.132	0.1156	0.1094	0.1907	**0.3504**	0.208	0.2159
Parietal	0.1465	0.0961	0.2007	**0.3368**	0.3146	**0.3363**	0.3237
All	0.8395	0.4963	0.5904	0.658	0.7265	**0.8754**	0.87

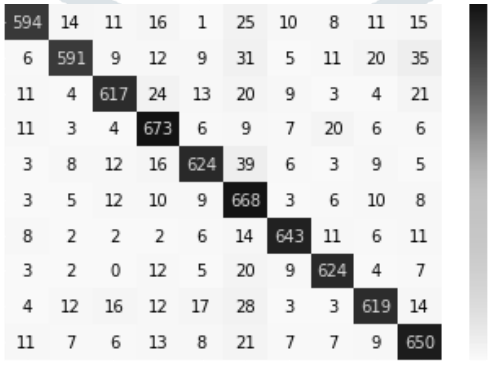


Fig. 8. Confusion matrix for the image task.

Real-world Brain Computer Interface applications have increased in tandem with the quick development of affordable EEG measuring equipment.

For persons with vocal problems or diseases like locked-in syndrome, it's crucial to be able to identify imagined speech. Furthermore, it is crucial for those with strong privacy demands, such as those who frequent public spaces where voice is not the most effective form of communication.

Compared to other motor activities like grip and lift, movement, etc., there has been less research done in this topic despite the enormous potential use cases. Speech is a natural medium for both human-to-human communication and connection with other technological devices, therefore it seems sense that a system that decodes imagined speech would be extremely important.

The majority of research on imagined speech recognition has been conducted using conventional signal processing techniques, such as machine learning classifiers like SVM, Random Forest k-NN, and Logistic Regression in conjunction with feature extraction techniques like common spatial patterns (CSP), auto-regression coefficients, and low-level descriptors like energy and entropy.

Utilizing architectures like CNN, LSTM, GRU, and others to understand the intricacies in EEG signals has been a focal point of more recent efforts that aim to capitalize on the advances in Deep Learning.

VI. CONCLUSION

The purpose of this work is to investigate the role of signals from various EEG frequency bands (δ, θ, α, β, γ) and brain lobes (frontal, occipital, parietal, and temporal) for imagined speech recognition. To apply imagined speech and object identification to a one-dimensional Convolutional Neural Network Long Short-term memory based deep learning architecture, we filter the EEG signal into matching frequency bands using parallel bandpass filters. In the one-dimensional Convolution Neural Network block, spatial characteristics are retrieved from the raw EEG signals, and temporal information is modeled over the features by the Long Short-Term Memory block. As a result, we used band reject filter as a preprocessing step for the EEG data before applying them to the imagined speech classification. We find that the frequency bands θ and α do not significantly contribute to imagined speech identification.

REFERENCES

[1] T. K. Bera, "A Review on The Medical Applications of Electroencephalography (EEG)," 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICBSII51839.2021.9445153.

[2] J. Satheesh Kumar, P. Bhuvaneswari, Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study, Procedia Engineering, Volume 38, 2012, Pages 2525-2536, ISSN 1877-7058, https://doi.org/10.1016/j.proeng.2012.06.298.

[3] Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. IEEE Trans. Rehabil. Eng. 8, 164–173. doi: 10.1109/tre.2000.847807.

[4] Kumar, S., Shekhar, S. & Agarwal, P. Cognitive capability identification in performing mental tasks using EEG-based coherence. Int J Syst Assur Eng Manag 14, 334–342 (2023). https://doi.org/10.1007/s13198-022-01799-8.

[5] P. Agarwal and S. Kumar "Electroencephalography-based imagined speech recognition using deep long short-term memory network " ETRI Journal 44.4 (2022): pp. 672-685. doi: 10.4218/etrij.2021-0118.

[6] Xu, X., Lee, D., Drougard, N. *et al.* Signature methods for brain-computer interfaces. *Sci Rep* 13, 21367 (2023). https://doi.org/10.1038/s41598-023-41326-8.

[7] N. Hashim, A. Ali, and W.-N. Mohd-Isa, "Word-based classification of imagined speech using eeg," in International Conference on Computa- tional Science and Technology. Springer, 2017, pp. 195–204.

[8] Agarwal P, Kumar S. Electroencephalography based imagined alphabets classification using spatial and time-domain features. Int J Imaging Syst Technol. 2022; 32(1): 111-122. doi:10.1002/ima.22655.

[9] Lei, D., Dong, C., Guo, H. *et al.* A fused multi-subfrequency bands and CBAM SSVEP-BCI classification method based on convolutional neural network. *Sci Rep* 14, 8616 (2024). https://doi.org/10.1038/s41598-024-59348-1.

[10] Aellen, F. M., Goktepe-Kavis, P., Apostolopoulos, S., and Tzovara, A. (2021). Convolutional neural networks for decoding electroencephalography responses and visualizing trial by trial changes in discriminant features. *J. Neurosci. Methods* 364, 109367. doi: 10.1016/j.jneumeth.2021.109367.

[11] Khan, J., Bhatti, M.H., Khan, U.G. et al. Multiclass EEG motor-imagery classification with sub-band common spatial patterns. J Wireless Com Network 2019, 174 (2019). https://doi.org/10.1186/s13638-019-1497-y.

[12] Behbahani, Soroor & Motie Nasrabadi, Ali. (2010). Hypnosis analysis by means of Similarity Index method. 2010 17th Iranian Conference of Biomedical Engineering, ICBME 2010 - Proceedings. 1-5. 10.1109/ICBME.2010.5704927.

[13] P. Kumar and E. Scheme, "A deep spatio-temporal model for eeg-based imagined speech recognition," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 995–999.

[14] Kumar, P., Saini, R., Roy, P.P. et al. Envisioned speech recognition using EEG sensors. Pers Ubiquit Comput 22, 185–199 (2018). https://doi.org/10.1007/s00779-017-1083-4.

[15] Shen, Fangyao & Peng, Yong & Kong, Wanzeng & Dai, Guojun. (2021). Multi-Scale Frequency Bands Ensemble Learning for EEG-Based Emotion Recognition. Sensors. 21. 1262. 10.3390/s21041262.