Decoding Neural Patterns for Naturalistic Speech Perception

Ashutosh Rai (2020BCS0020)
Roshin Nishad (2020BCS0019)
Pratik Raj (2020BCS0112)
Sai Teja (2020BCS0145)

Guided By,
Dr. Suchithra M S

Introduction

Indian Institute of Information Technology Kottayam

- This research represents a unique departure from conventional Brain-Computer Interface (BCI) studies.
- Traditional approaches focus on decoding neural signals during active speech or writing, this project focuses on **speech perception** rather than production.
- This research will investigate how the brain processes speech when individuals are **listeners rather than active participants**.
- This approach analyzes the complex neural patterns involved in understanding speech.

Problem Statement

The problem is to process brain activity data for decoding speech perception using deep learning methods.

This research also aims to increase our understanding of the brain's response to heard speech, with the objective of translating brain signals into intelligible speech output, while benefiting people with speech difficulties.

Literature Review

S no.	Author(s) and Publication Year	Study Title	Key Findings	Limitations	Relevance to Project
1.	Luo, S., Rabbani, Q. & Crone (2022)	Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication	The paper explores Brain-Computer Interfaces (BCIs) as Augmentative and Alternative Communication (AAC) devices, particularly for individuals with conditions like locked-in syndrome (LIS) where traditional communication is severely limited Technology: ECoG, DNN, ML	Long-term Electrocorticography signal stability for speech decoding has not yet been fully investigated and safety of long-term ECoG implants are in question.	The paper discusses various technologies and methods, such as various machine learning algorithms, for decoding spoken phonemes and synthesizing speech.
2.	Brumberg, J. S., Kennedy, P. R. & Guenther, F. H. (2009)	Artificial speech synthesizer control by brain-computer interface	The paper uses an intracortical microelectrode device with wireless transmission to map neural firing rates in the speech motor cortex to intended speech utterances. The formant frequencies predicted by the system are synthesized in real-time, providing instantaneous auditory feedback for closed-loop BCI control. Technology: Kalman Filter Algorithm	The study is limited to a single subject, which may not provide a comprehensive understanding of the system's applicability to a broader population.	The paper provides a robust framework for decoding neural signals to facilitate artificial speech production. This methodology could serve as a foundational reference for the decoding of neural patterns associated with speech perception
3.	Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G and Schultz T (2015)	Brain-to-text: decoding spoken phrases from phone representations in the brain.	Describes the design and development of a Brain-to-Text system that uses neural phone models and language information to decode spoken phrases from brain activity. Technology: ECoG, Neural Phone Models, Language Models	In this study, only one context-independent model is trained for each phone, i.e., without consideration of preceding or succeeding phones due to the limited amount of data	The approach introduced here may have important implications for the design of novel brain-computer interfaces, because it may eventually allow people to communicate solely based on brain signals associated with natural language function and with scalable vocabularies.

Literature Review

S no.	Author(s) and Publication Year	Study Title	Key Findings	Limitations	Relevance to Project
4.	Zheng, X., Chen, W., Li, M., Zhang, T., You, Y., & Jiang, Y. (2020)	Decoding human brain activity with deep learning	The study presents an LSTM-CNN deep learning model for EEG-based visual object classification, achieving a test accuracy of 94.4%. This architecture leverages LSTM for temporal EEG sequence analysis and CNN for spatial feature extraction. Technology: LSTM-CNN, EEG Analysis, Object Classification	The performance of the model is highly sensitive to the order of the LSTM and CNN layers, indicating a potential limitation in the flexibility of the model's architecture.	The paper's LSTM-CNN model for EEG data decoding offers a structural foundation for our project's deep neural network aimed at speech perception analysis and provides a framework relevant to our research objectives.
5.	Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D. & Mesgarani, N. (2019)	Towards reconstructing intelligible speech from the human auditory cortex	The paper discusses the use of a deep neural network architecture that consists of two stages: feature extraction and feature summation. This framework calculates a high-dimensional representation of the input, which is then used to regress the output of the model. Technology: Deep Neural Networks, Auditory Cortex Analysis, Vocoder Representation.	The study mentions the limited diversity of the neural responses in their recording, which limits the added information that is gained from additional electrodes.	The paper presents a deep neural network architecture that consists of two stages: feature extraction and feature summation. This framework can serve as a blueprint for the fully connected deep neural network employed in our research
6.	Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W. and Ball, T. (2017)	Deep learning with convolutional neural networks for EEG decoding and visualization.	The paper presents results showing that their deep learning model can effectively decode and visualize EEG data. They highlight the success of their model in understanding and representing brain signals. Technology: Deep Learning (CNNs), EEG Analysis.	The paper mentions that The flexibility of ConvNets (CNNs) may be a limitation in some brain-signal decoding scenarios.	The research underscores the efficacy of deep learning methodologies, particularly CNNs, in the intricate task of EEG decoding. Such an approach holds promise for the analysis of (iEEG) data, associated with speech perception decoding.

Literature Review Summary

- The first three papers collectively investigate BCIs for Augmentative and Alternative Communication, employing technologies like ECoG, DNN, ML, and Kalman Filters to decode speech and synthesize communication for individuals with speech impairments. They address challenges in signal stability, safety of long-term implants, and data limitations while providing frameworks for speech production and phrase decoding from neural activity.

- The last three papers collectively advance deep learning in neural decoding, with LSTM-CNN models achieving high accuracy in EEG-based classification and DNNs effectively extracting and summarizing features from auditory cortex data. These approaches, employing technologies like LSTM, CNN, and vocoder analysis, provide a foundation for enhancing speech perception analysis in BCI applications.

Motivation

- Our motivation for this project largely aligns with the problem statement. This research serves as an exploration into a less-studied domain of passive speech perception.
- We're studying neural patterns linked to speech perception using iEEG data and an FC DNN, aiming to improve speech synthesis in Brain-Computer Interfaces.
- By translating neural signals from passive speech perception into clear speech output, we want to make communication more accessible and effective for those who need it.

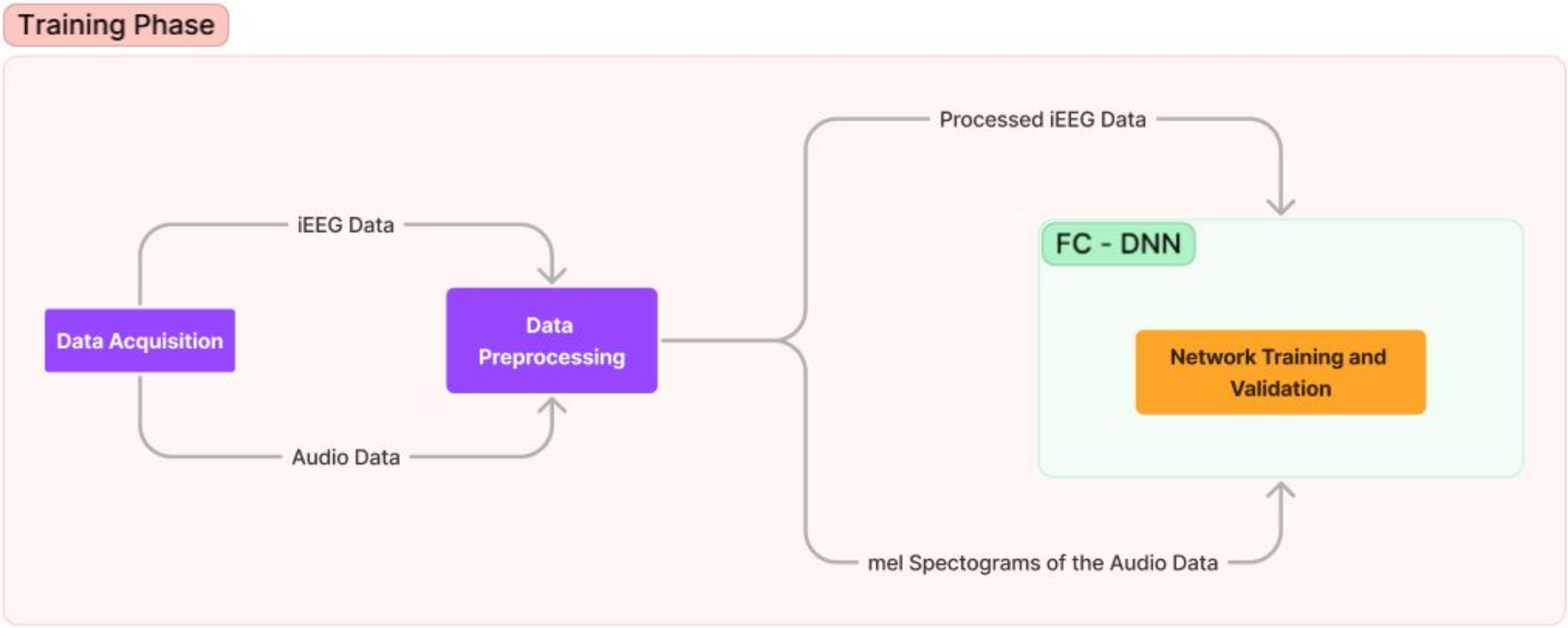
Workflow

The aim of the project is to decode neural patterns associated with speech perception through intracranial electroencephalography (iEEG) data, leveraging a rich and unique dataset

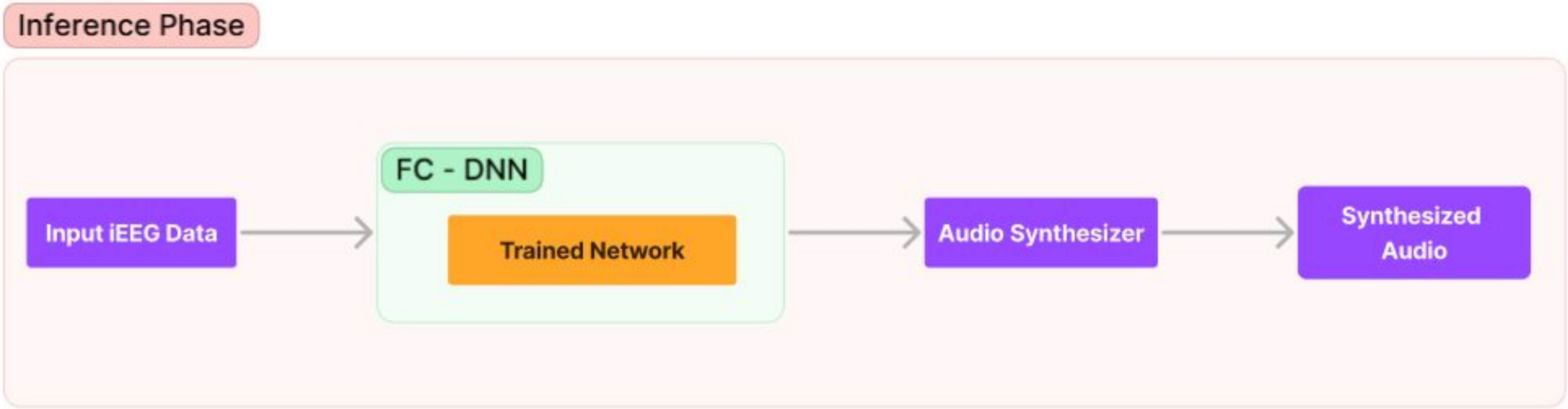
The key steps of the implementation:

- Dataset includes iEEG and fMRI data from 51 subjects watching a movie.
- Preprocessing: noise removal, detrending, 70-170 Hz bandpass filtering.
- Extracts High Gamma features from iEEG, using Hilbert transform for time series.
- Audio synchronized with iEEG via time-shifting, resampling, Mel spectrogram conversion.
- FC-DNN maps iEEG High Gamma features to audio Mel spectrograms, aiming to reduce error.
- Model's audio predictions from iEEG are compared with original for evaluation.

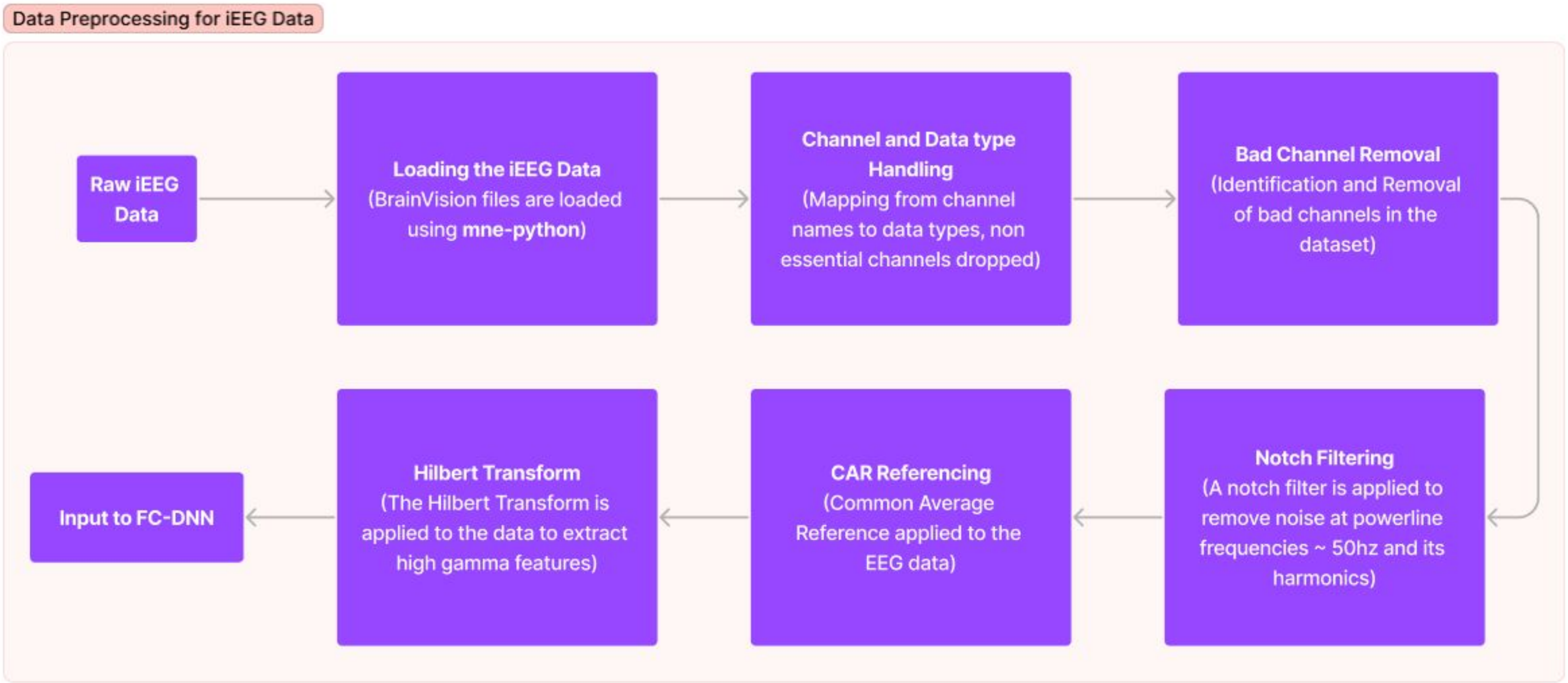
Workflow Diagram



Workflow Diagram

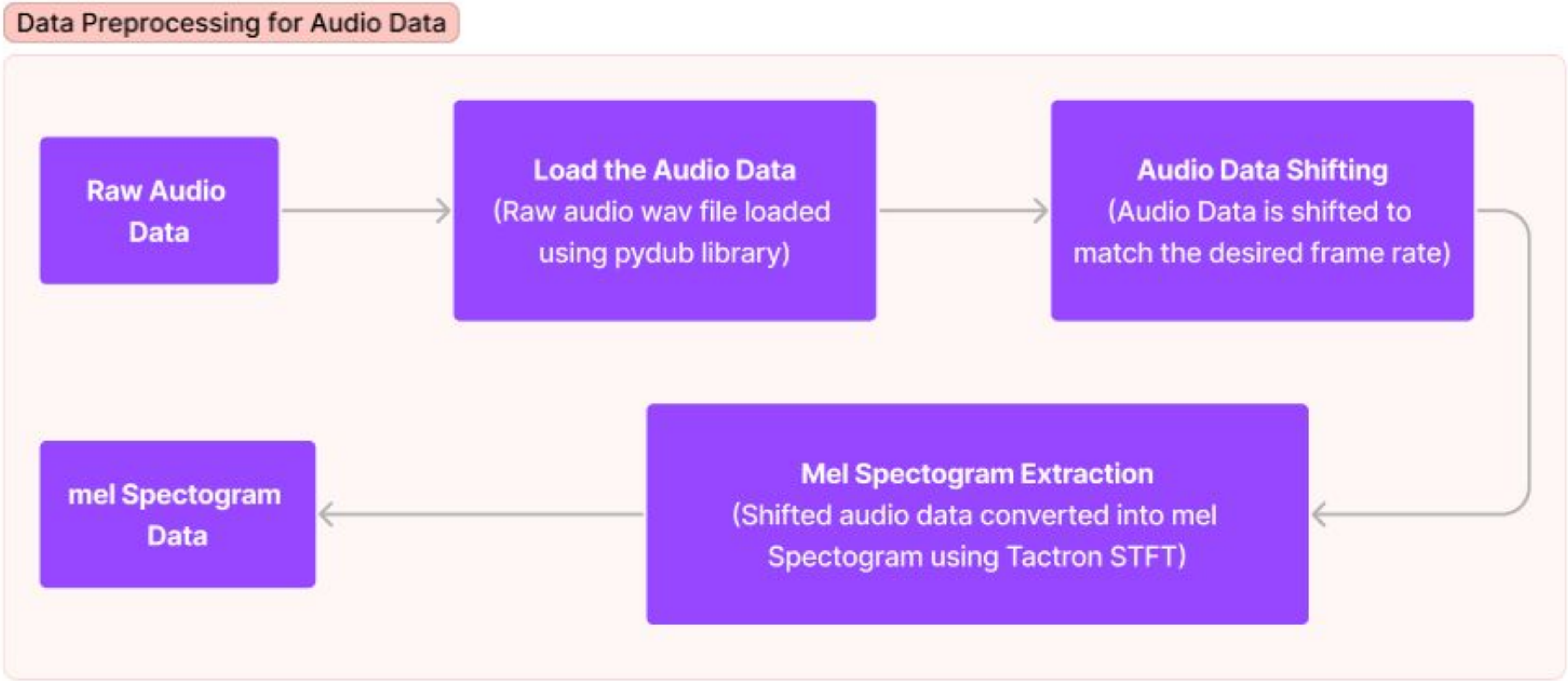


Workflow Diagram



Workflow Diagram

Indian Institute of Information Technology Kottayam



Experimental Results (Setup)

Dataset Overview:

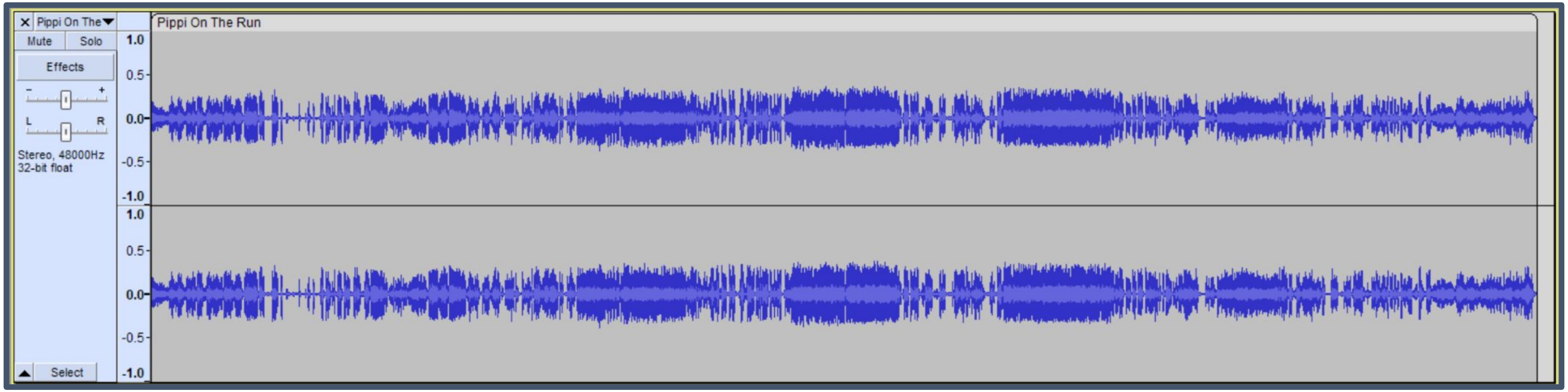
Our dataset is a valuable and extensive resource of intracranial electroencephalography (iEEG) brain activity data. This dataset offers a unique combination of iEEG and functional magnetic resonance imaging (fMRI) recordings acquired through a audiovisual movie stimulus.

- stimuli
- sub-01
  - sub-01_sessions.json
  - sub-01_sessions.tsv
  - ses-iemu
    - sub-01_ses-iemu_scans.tsv
    - ieeg
      - sub-01_ses-iemu_acq-clinical_coordsystem.json
      - sub-01_ses-iemu_acq-clinical_electrodes.tsv
      - sub-01_ses-iemu_acq-render_photo_seeg_bottom.jpg
      - sub-01_ses-iemu_acq-render_photo_seeg_left.jpg
      - sub-01_ses-iemu_acq-render_photo_seeg_right.jpg
      - sub-01_ses-iemu_acq-render_photo_seeg_top.jpg
      - sub-01_ses-iemu_task-film_acq-clinical_run-1_channels.tsv
      - sub-01_ses-iemu_task-film_acq-clinical_run-1_ieeg.eeg
      - sub-01_ses-iemu_task-film_acq-clinical_run-1_ieeg.json
      - sub-01_ses-iemu_task-film_acq-clinical_run-1_ieeg.vhdr
      - sub-01_ses-iemu_task-film_acq-clinical_run-1_ieeg.vmrk
      - sub-01_ses-iemu_task-film_run-1_events.tsv
      - sub-01_ses-iemu_task-rest_acq-clinical_run-1_channels.tsv
      - sub-01_ses-iemu_task-rest_acq-clinical_run-1_ieeg.eeg
      - sub-01_ses-iemu_task-rest_acq-clinical_run-1_ieeg.json
      - sub-01_ses-iemu_task-rest_acq-clinical_run-1_ieeg.vhdr
      - sub-01_ses-iemu_task-rest_acq-clinical_run-1_ieeg.vmrk
      - sub-01_ses-iemu_task-rest_run-1_events.tsv

Experimental Results (Setup)

Data Collection:

The dataset was obtained through a movie-watching experiment, the participants were engaged in watching a 6.5-minute short movie composed of segments from **Pippi on the Run** film. The movie featured 13 interleaved blocks of speech and music, each lasting 30 seconds.



Experimental Results (Setup)

The experiment includes two modalities: **iEEG** and **fMRI**. For our experiment, we are only interested in the iEEG data.

Data Format:

The iEEG recordings in this dataset are stored in the widely used BrainVision format. The BrainVision format consists of three key file types, each serving a specific purpose. The following are the aforementioned file types:

1. **.eeg** (EEG Data File) file to access the raw EEG data
2. **.vmrk** (Marker File) file for event synchronization
3. **.vhdr** (Header File) file for important session information

Experimental Results (Output)

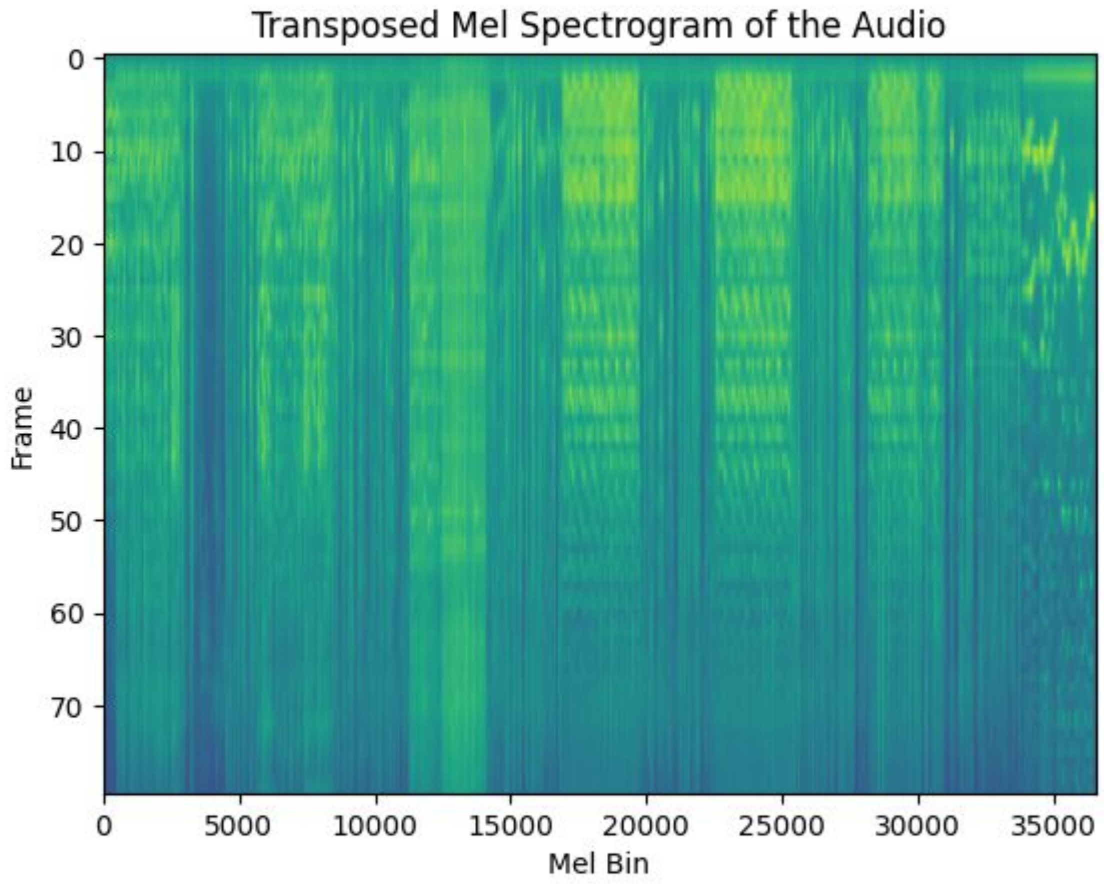


Fig 1.1: The Transposed Mel Spectrogram of our Audio (WAV File)

Experimental Results (Output)

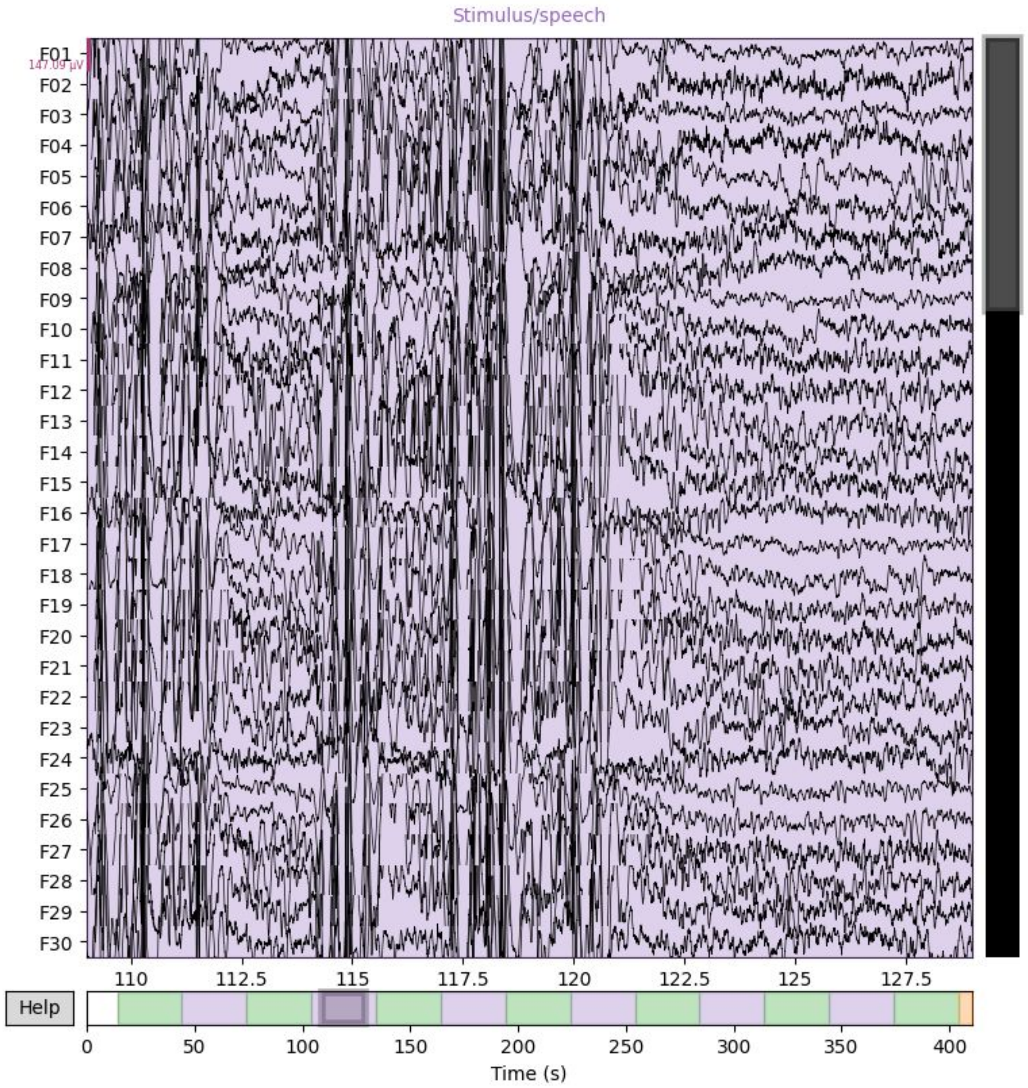


Fig 1.2: EEG data from 30 channels over a 20-second duration

Experimental Results (Output)

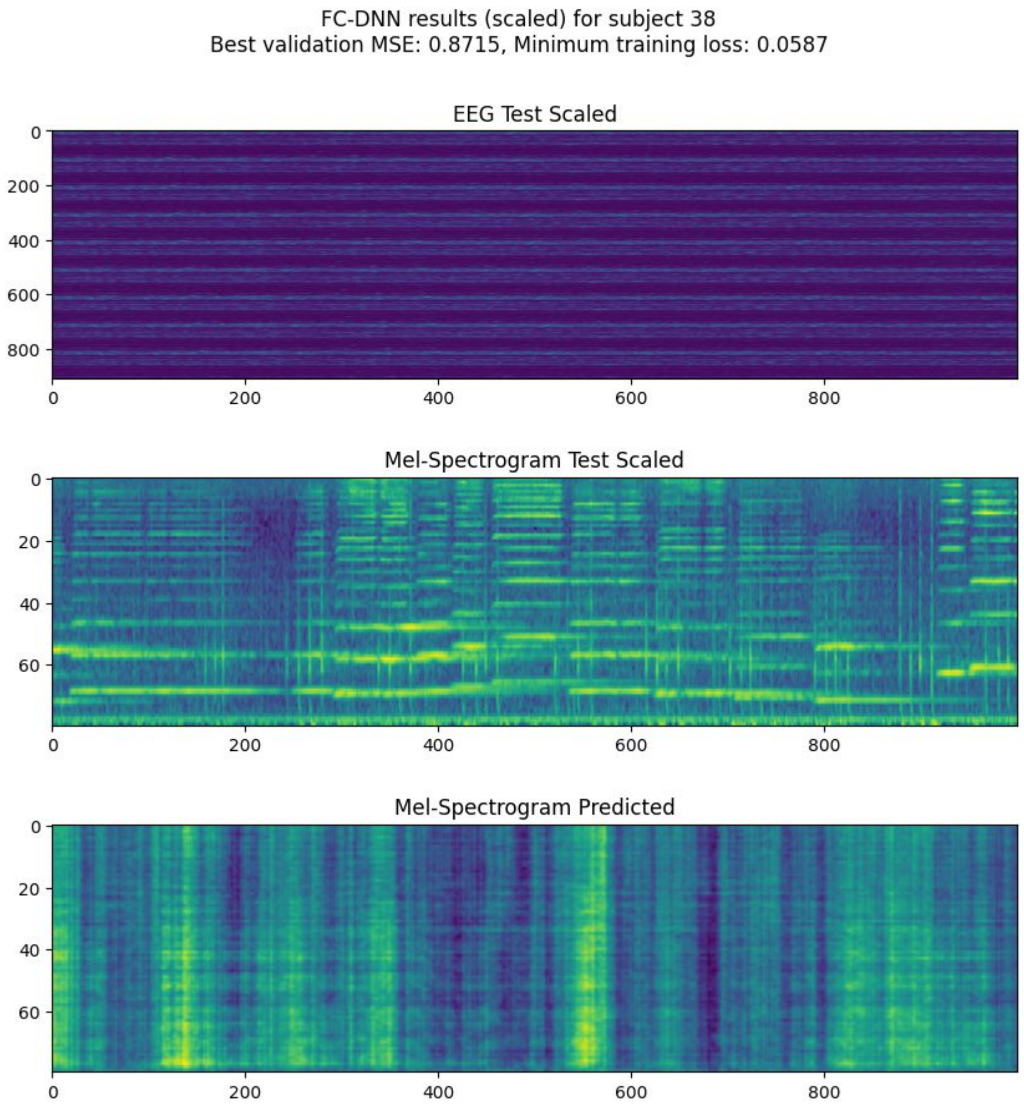


Fig 1.3: In this figure, the EEG Test Scaled is the input data for the model, The Mel-Spectrogram Test Scale is the ground truth, and the Mel-Spectrogram Predicted is the prediction.

Experimental Results (Output)

```
SSIM: 0.004424715235700905
MSE: 0.7671618073918052
PSNR: 1.1511302634295026
Cosine Similarity: 0.010131542568530438
```

SSIM - Structured Similarity Index
PSNR - Peak Signal to Noise Ratio

Fig 1.4: The metrics portrays that the model isn't performing well, however the main goal was to get the implementation going. We will be implementing a similar approach with a 2D-CNN in our future work, which would produce a better performance

DEMONSTRATION

Significance

- The significance of this research remains consistent with the problem statement, which is to enhance natural speech synthesis, resulting in more realistic speech output.

- This research could help people with speech disabilities communicate better and offers new understanding of how the brain processes language.

Conclusion

- In summary, this project uses a Fully Connected Deep Neural Network to decode brain activity during speech perception.

- We trained the network with intracranial EEG data and audio features, but the current model's predictions of audio from brain activity are not accurate. We're working on improving its accuracy,

- As we move forward, we also plan to adapt these methods to use 2D Convolutional Neural Networks, aiming to enhance Brain-Computer Interface systems for speech processing.

References

1. Luo, S., Rabbani, Q. & Crone, N.E. Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication. Neurotherapeutics 19, 263–273 (2022).

2. Brumberg, J. S., Kennedy, P. R. & Guenther, F. H. Artificial speech synthesizer control by brain–computer interface. In Tenth Annual Conference of the International Speech Communication Association (2009).

3. Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G and Schultz T. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front. Neurosci. 9, 217 (2015).

References

4. Zheng, X., Chen, W., Li, M., Zhang, T., You, Y., & Jiang, Y. (2020). Decoding human brain activity with deep learning. Biomedical Signal Processing and Control, 56, 101730.

5. Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D. & Mesgarani, N. Towards reconstructing intelligible speech from the human auditory cortex. Sci. Rep. 9, 1–12 (2019).

6. Schirrmeister, R. T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M. et al. Deep learning with convolutional neural networks for eeg decoding and visualization. Hum. Brain Mapp. 38, 5391–5420 (2017).