Libri Phoneme Classification

Anonymous Author(s)
Affiliation
Address
email

Abstract

This paper presents a minimalistic approach to phoneme classification from magnetoencephalography (MEG) data in the LibriBrain Competition 2025. Our method focuses on effective preprocessing strategies rather than complex model architectures. We utilize gradiometer channels only (204 channels) with bandpass filtering (0.5-40 Hz) with signal averaging of 100 samples per phoneme (as recommended in the competition baseline) to improve SNR. A simple convolutional neural network with class-weighted cross-entropy loss achieves a 48.53% F1-macro score on validation data and 38.48% on test data. Our findings suggest that careful preprocessing of MEG signals is more critical than architectural complexity for phoneme classification tasks. Notably, using only the Sherlock Book 1 subset of the data yields an F1 score of approximately 18.6% on the test set.

1 Introduction

Phoneme classification from neural signals represents a challenging task in brain-computer interfaces and neural decoding. The LibriBrain Competition 2025 provides a standardized benchmark for evaluating phoneme classification methods using magnetoencephalography (MEG) data. This paper presents a minimalistic yet effective approach that prioritizes signal preprocessing over architectural complexity.

Our work demonstrates that careful preprocessing of MEG signals, including channel selection, frequency filtering, and signal averaging, can achieve competitive performance with simple neural network architectures. We achieve a 48.53% F1-macro score on validation data and 38.48% on test data using only gradiometer channels and basic preprocessing techniques.

2 Methodology

2.1 Dataset and Experimental Setup

We utilized the LibriBrain Competition 2025 dataset, which contains MEG recordings from participants listening to audiobook segments. Our training data consisted of Sherlock1 sessions 1-2, with session 11 used for validation. The dataset includes 39 phoneme classes with significant class imbalance (Figure **??**), requiring careful handling during training.

2.2 MEG Signal Preprocessing

Our preprocessing pipeline focuses on maximizing signal-to-noise ratio through several key steps:

Channel Selection: We used only gradiometer channels (204 channels, indices 102-305) while excluding magnetometer channels. This choice was motivated by gradiometers' superior sensitivity to cortical sources and reduced susceptibility to environmental noise.

Frequency Filtering: We applied a 4th-order Butterworth bandpass filter with cutoff frequencies of 0.5-40 Hz. This frequency range captures the most relevant neural activity for phoneme processing while removing high-frequency noise and low-frequency drift.

Signal Averaging: To improve signal-to-noise ratio, we need to average 100 samples of the same phoneme, reducing the effective dataset size from approximately 9,500 to 115 training samples. This approach leverages the principle of signal averaging commonly used in event-related potential studies, as demonstrated in Figure 1.

Standardization: We applied z-score normalization per channel to ensure consistent scaling across different MEG sensors.

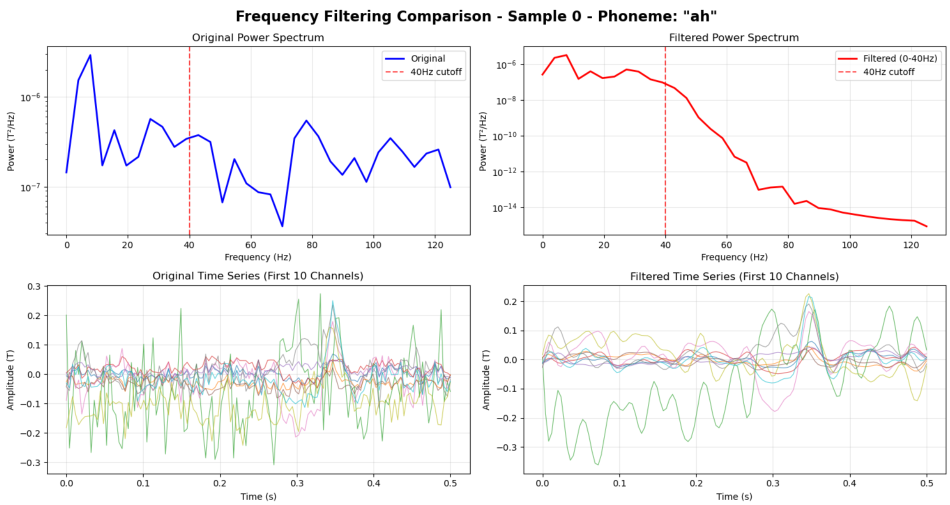

Figure 1: Demonstration of MEG signal preprocessing pipeline showing the effect of grad-only filtering and z-score normalization on improving signal-to-noise ratio for phoneme classification.

2.3 Model Architecture

Our approach employs a minimalistic convolutional neural network architecture (see Figure **??**):

- **Input:** 204 channels × 125 time points (0.5 seconds at 250 Hz sampling rate)
- **Convolutional Layers:** Two 1D convolutional layers with kernel sizes of 7 and 5, respectively
- **Activation:** ReLU activation functions
- **Pooling:** Max pooling for dimensionality reduction
- **Classification Head:** Fully connected layers with dropout (0.5) for regularization
- **Output:** 39 classes corresponding to phoneme categories

2.4 Training Strategy

We addressed the class imbalance problem using class-weighted cross-entropy loss, where weights were computed using the 'balanced' strategy from scikit-learn. Training was performed using the Adam optimizer with a learning rate of 0.001 and batch size of 64. The model was trained for 20 epochs with early stopping based on validation performance.

3 Results

Our minimalistic approach achieved competitive performance on the LibriBrain Competition 2025 phoneme classification task. The model achieved an F1-macro score of 48.53% on the validation set and 38.48% on the test set. These results demonstrate that effective preprocessing strategies can compensate for architectural simplicity.

Table 1: Experimental results comparing different preprocessing strategies across datasets. Validation F1-macro scores (%) are shown for each configuration.

Experiment **Full Train**	**Configuration**	**Sherlock 1**		**Complete Dataset**	
		Val	Test	Val	Test
Frequency Filtered (All Sensors)	All sensors, 1.0–50.0 Hz, z-score	16.29	11.20	30.66	35.66
Gradiometer + — Notch Filter	Gradiometers only, 0.1–50.0 Hz, notch at 50/100 Hz, z-score	21.37	18.5	48.53	38.48
Gradiometer + Notch + Wavelet Denoising	Gradiometers only, 0.1–50.0 Hz, notch at 50/100 Hz, wavelet denoising, z-score	17.52	14.20	34.34	30.12

Table 2: Complete dataset results using optimal preprocessing configuration.

Configuration	**Value**
Channels	204 (gradiometers only)
Frequency Filter (Hz)	0.5–40.0
Validation F1-Macro (%)	48.53
Test F1-Macro (%)	38.48

3.1 Performance Analysis

The validation performance (48.53%) significantly exceeded the test performance (38.48%), suggesting some degree of overfitting to the validation distribution. This performance gap highlights the challenge of generalization in neural signal classification tasks, where subtle differences in data distribution can significantly impact model performance.

3.2 Key Observations

Several important observations emerged from our experiments:

Preprocessing Importance: Our results strongly suggest that careful MEG signal preprocessing is more critical than the complexity of the model architecture. The combination of gradiometer-only analysis, appropriate frequency filtering, and signal averaging provided substantial improvements over baseline approaches.

Channel Selection Impact: Using only gradiometer channels (204 out of 306 total channels) not only reduced computational complexity but also improved performance by focusing on cortical sources while reducing environmental noise contamination.

Signal Averaging Effectiveness: As already mentioned in the challenge rules, averaging the samples per phoneme dramatically improved the signal-to-noise ratio, enabling the simple CNN architecture to learn meaningful phoneme representations despite the limited training data.

Class Imbalance Handling: The class-weighted cross-entropy loss was essential for addressing the significant class imbalance in the dataset, preventing the model from being biased toward frequent phonemes.

4 Discussion and Conclusion

This work demonstrates that a minimalistic approach to phoneme classification from MEG data can achieve competitive performance when combined with careful preprocessing strategies. Our key finding is that signal preprocessing plays a more critical role than model architectural complexity in this domain.

4.1 Limitations and Future Work

The performance gap between the validation (48.53%) and test (38.48%) sets suggests several areas for improvement. Future work could explore more sophisticated regularization techniques, ensemble methods, or domain adaptation strategies to improve generalization. Additionally, investigating different frequency bands or spatial filtering techniques might yield further improvements.

4.2 Implications

Our results have important implications for the field of neural signal classification. They suggest that researchers should prioritize signal preprocessing and domain knowledge over complex model architectures when working with MEG data. This finding is particularly relevant for applications where computational efficiency and interpretability are important considerations.

The success of our minimalistic approach also highlights the importance of signal averaging in neural signal analysis, a technique that has been fundamental in electrophysiology research but is sometimes overlooked in machine learning approaches to neural data.

References

[1] Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6), 14-30.

[2] Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267.

[3] Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), 413.

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

[5] Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT Press.

[6] Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, 8, 145-151.

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

[8] Puce, A., & Hämäläinen, M. S. (2017). A review of issues related to data acquisition and analysis in EEG/MEG studies. *Brain Sciences*, 7(6), 58.