LARGE EEG FOUNDATION MODEL LEARNS INFORMATIVE LOW-FREQUENCY REPRESENTATIONS FROM INTRACRANIAL BRAIN SIGNALS

Qiang Sun[*], Liuyin Yang[*], Eva Calvo Merino, & Marc M. Van Hulle
Laboratory for Neuro- and Psychophysiology, Department of Neurosciences
KU Leuven, Belgium
{qiang.sun,liuyin.yang,eva.calvomerino,marc.vanhulle}@kuleuven.be

ABSTRACT

Foundation models (FMs) for Electroencephalography (EEG) time-series have recently emerged as powerful tools, leveraging large-scale non-invasive datasets to learn robust, universal spatiotemporal neural representations. The question is whether these models can bridge the gap towards invasive recording modalities, such as Electrocorticography (ECoG), which is hailed for providing superior signal-to-noise ratios but with limited spatial coverage and small patient cohorts. In this work, we investigate the transferability of EEG-FMs to ECoG-based decoding tasks. We propose an ECoG-to-EEG channel projection module and a lightweight adaptation strategy to effectively update the pretrained FM backbone. Benchmarking results demonstrate that our adapted EEG-FM outperforms conventional ECoG decoders in extracting finger movement-related information from low-frequency sampled ECoG signals (i.e., a sampling rate of 128 Hz). These findings establish a new paradigm for intracranial brain-computer interfaces, suggesting that cross-modality knowledge can be leveraged to improve neural decoding of finger movements.

Track: Research

1 INTRODUCTION

Brain signals acquired through multiple synchronized sensors represent high-dimensional multivariate time-series characterized by complex spatiotemporal dynamics. These signals offer immense potential for automated sleep monitoring (Mousavi et al., 2019), clinical diagnostics (Tveit et al., 2023; Mlinarič et al., 2026), and the development of Brain-Computer Interfaces (BCIs) (Benabid et al., 2019; Ofner et al., 2019; Metzger et al., 2023). Electroencephalography (EEG) and electrocorticography (ECoG) are two widely utilized modalities for recording neural activity across distinct spatial scales.

The non-invasive nature of Electroencephalography (EEG) has enabled the modeling of massive, diverse neural datasets, catalyzing a shift toward EEG Foundation Models (EEG-FMs) (Yao et al., 2025). By leveraging self-supervised objectives on vast unlabelled repositories, these models have begun to transcend the limitations of task-specific architectures by capturing robust, generalizable neurophysiological signatures across a broad spectrum of downstream clinical and cognitive tasks (Yang et al., 2023; Jiang et al., 2024; Wang et al., 2024). While EEG records macro-scale brain dynamics, intracranial ECoG offers a high-fidelity mesoscale alternative characterized by superior signal-to-noise ratios (SNR) and precise spatial localization (Schalk & Leuthardt, 2011). However, the clinical necessity of invasive surgery and the high variability in electrode grid configurations result in significant data sparsity and structural heterogeneity. These challenges have hindered the development of a unified ECoG foundation model.

Can the representational performance of EEG-FMs be effectively transferred to ECoG-based decoding tasks? Despite the shared neural origins of EEG and ECoG, the feasibility of

[*]Equal contribution and shared correspondence.

cross-modality transfer between non-invasive FMs and invasive recordings remains unexplored (see Related Works Appendix A.1). **In this work, we present the first systematic investigation into repurposing EEG-FMs for ECoG applications.**

We leverage ST-EEGFormer (Yang et al., 2026), the winning architecture of the NeurIPS 2025 EEG Foundation Challenge regression track[1]. With a highly scalable design reaching 328.4M parameters, this model provides a robust framework for investigating cross-modality transfer. To bridge the ECoG-EEG modality gap, we propose a learnable Coordinate-Latent Projector (CLP) that maps ECoG electrode coordinates into the FM's latent channel embedding space. Downstream adaptation uses Low-Rank Adaptation (LoRA) (Hu et al., 2021). We evaluate this framework on predicting finger movement trajectories from ECoG recordings (Kubanek et al., 2009). Given that high-frequency component is unavailable in EEG, this study focuses on the capability of EEG-FMs to extract low-frequency representations from 128 Hz-sampled ECoG signals. We compare our adapted EEG-FMs to conventional ECoG decoders (Figure 1). The key findings are as below:

- The adapted EEG-FM significantly outperforms conventional ECoG decoders in decoding finger movements from low-frequency ECoG signals.
- Applying LoRA-based adaptation to population-level data significantly enhances decoding performance.
- Scaling the FM backbone does not yield transfer gains, likely because of data scarcity of the downstream ECoG tasks and adaptation strategy.

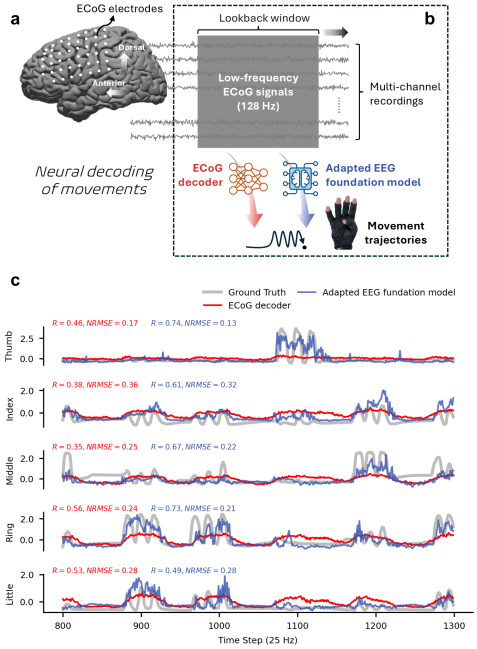

Figure 1: Movement decoding from intracranial brain signals. (a) ECoG electrodes (1 cm center-to-center spacing) are placed over the sensorimotor cortex. (b) We decode finger movement trajectories from a lookback window of low-frequency ECoG signals and compare adapted EEG foundation models against diverse ECoG decoders. (c) Exemplary trajectories (from **Subject cc**) show that our adapted EEG fundation model outperforms the ECoG decoder in predicting flexion dynamics.

2 METHODS

2.1 TASK FORMULATION

Given a preprocessed ECoG segment $\mathbf{X}_t \in \mathbb{R}^{C\times T}$, where C denotes the channel count and $T = 128$ represents a 1-second lookback window at 128 Hz. The neural decoder $f(\cdot)$ maps the 1-second sliding window of neural features $\mathbf{X}_{t-1:t}$ to the instantaneous finger positions $\mathbf{Y}_t \in \mathbb{R}^{5\times 1}$, enabling continuous decoding of hand kinematics. We used the *fingerflex* dataset from the Stanford Human ECoG Library (Miller, 2019), which contains ECoG recordings and synchronized digital glove data from nine epilepsy patients performing self-paced, single finger flexion-extensions. During preprocessing, ECoG signals were downsampled from 1,000 Hz to 128 Hz. The finger trajectories were sampled at 25 Hz, which serves as the model's prediction frequency. Further details on the dataset and preprocessing pipeline are available in Appendix A.2.

2.2 COORDINATE-LATENT PROJECTOR

To bridge the gap between variable ECoG electrode configurations and fixed EEG latent spaces, we propose the Coordinate-Latent Projector (CLP). This module maps 3D MNI coordinates into a continuous embedding space through spatial affinity and residual refinement (Figure 2).

[1]EEG Challenge (2025): https://eeg2025.github.io

Coordinate Encoding and Spatial Projection Given C ECoG channels with coordinates $P \in \mathbb{R}^{C\times 3}$, we first extract a shared spatial representation $H \in \mathbb{R}^{C\times 128}$ using a two-layer MLP with GELU activation. To align these channels with a pre-trained EEG latent space, a linear projection transforms H into an affinity matrix $\mathbf{W} \in \mathbb{R}^{C\times M}$, representing the spatial correspondence between C ECoG channels and M EEG latent anchors. The base ECoG embeddings $\mathbf{E}_{\text{base}} \in \mathbb{R}^{C\times D}$ are then computed as:

$$\mathbf{E}_{\text{base}} = \mathbf{W}\mathbf{E}_{\text{EEG}} \tag{1}$$

where $\mathbf{E}_{\text{EEG}} \in \mathbb{R}^{M\times D}$ denotes the frozen EEG-FM positional embeddings.

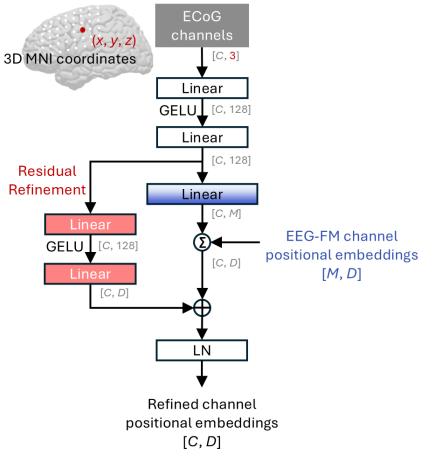

Figure 2: Coordinate-Latent Projector.

Residual Refinement and Normalization While $\mathbf{E}_{\text{base}}$ captures global spatial alignment, it may overlook modality-specific ECoG characteristics. We introduce a residual path $\mathcal{G}_\theta$ that branches from the shared representation H to learn fine-grained adjustments. This path consists of a MLP (Linear-GELU-Linear) that projects H directly to the embedding dimension D. The final refined embeddings $\mathbf{E}_{\text{refined}}$ are obtained by aggregating the spatial projection and the residual refinement, followed by Layer Normalization (LN):

$$\mathbf{E}_{\text{refined}} = \text{LN}(\mathbf{E}_{\text{base}} + \mathcal{G}_\theta(H)) \tag{2}$$

To ensure training stability, the final layer of $\mathcal{G}_\theta$ is initialized to zero, allowing the module to initially function as a pure spatial interpolator before gradually adapting to ECoG-specific features.

2.3 Cross-Subject Adaptation via Parameter-Efficient Fine-Tuning

To overcome the inherent data scarcity in individual ECoG recordings, we aggregate training data across all subjects (denoted as population-level training) to provide a robust dataset for adapting the pre-trained ST-EEGFormer. To leverage the representational power of the model, we adopt a parameter-efficient fine-tuning approach using LoRA. We integrate LoRA adapters into the final $n = 4$ Transformer encoder layers, targeting the Query, Key, Value (Q, K, V), and projection matrices within the multi-head self-attention modules. During the adaptation phase, the pre-trained ST-EEGFormer backbone remains frozen, with the exception of the LoRA layers. The CLP module, LoRA adapters, and the downstream regression head are optimized jointly. Detailed hyperparameter configurations are provided in Appendix A.5.

2.4 Evaluation

We benchmarked the adapted EEG-FM against 23 conventional ECoG decoders. These baselines were constructed by pairing three feature types—physiological, time-frequency, and raw time-series—with an array of model architectures. Detailed specifications for feature extraction and model implementations are provided in Appendicies A.3 and A.4, respectively. All decoders were trained on the initial 400 seconds (or the first two-thirds) of the ECoG recordings and synchronized finger trajectories, with the remaining data reserved for testing. Performance was evaluated using the Pearson correlation coefficient (*R*) to assess linear tracking and the normalized Root Mean Square Error (*NRMSE*) to quantify prediction error across all five fingers. Statistical significance of decoder comparison was assessed using the Wilcoxon signed-rank test with a significance threshold of $\alpha = 0.05$. P-values were adjusted for multiple comparisons using the False Discovery Rate (FDR) correction where applicable. Significance levels are denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and $n.s.$ for non-significant results.

3 Results and Discussion

The adapted EEG-FM (ST-EEGFormer-s) achieved the highest overall performance, yielding a Pearson correlation (*R*) of $0.4000_{\pm 0.1398}$ (Table 1). This result significantly outperformed the top-

Table 1: Comparison of decoders. Metrics are $mean_{\pm std}$. Red: Top ECoG decoder; Blue: Top EEG-FM; **Bold**: Overall best. Per-finger results are provided in Appendix A.6.

Feature Type	Model	$R \uparrow$	$NRMSE \downarrow$
Physiological (LMP, Band Power)	PLS	$0.2520_{\pm 0.1073}$	$0.1743_{\pm 0.0333}$
	NPLS	$0.1981_{\pm 0.0989}$	$0.2407_{\pm 0.1193}$
	HOPLS	$0.1821_{\pm 0.1215}$	$0.1694_{\pm 0.0351}$
	RF	$0.3056_{\pm 0.1279}$	$0.1632_{\pm 0.0326}$
	MLP	$0.2722_{\pm 0.1205}$	$0.1649_{\pm 0.0308}$
	LSTM	$0.2922_{\pm 0.1367}$	$0.1667_{\pm 0.0318}$
	CNN-LSTM	$0.2011_{\pm 0.1169}$	$0.1774_{\pm 0.0401}$
	WaT	$0.2465_{\pm 0.1190}$	$0.1836_{\pm 0.0353}$
	WaTFi	$0.2746_{\pm 0.1372}$	$0.1729_{\pm 0.0332}$
	WaTEi	$0.2453_{\pm 0.1130}$	$0.1782_{\pm 0.0345}$
Time-Frequency (5-60 Hz)	PLS	$0.2673_{\pm 0.1194}$	$0.1690_{\pm 0.0318}$
	NPLS	$0.2258_{\pm 0.1338}$	$0.1922_{\pm 0.0531}$
	HOPLS	$0.2098_{\pm 0.1258}$	$0.1698_{\pm 0.0360}$
	RF	$0.3137_{\pm 0.1171}$	$0.1634_{\pm 0.0319}$
	MLP	$0.2594_{\pm 0.1386}$	$0.1646_{\pm 0.0301}$
	LSTM	$0.2880_{\pm 0.1379}$	$0.1642_{\pm 0.0297}$
	CNN-LSTM	$0.1918_{\pm 0.1134}$	$0.1642_{\pm 0.0401}$
	WaT	$0.2415_{\pm 0.1211}$	$0.1808_{\pm 0.0376}$
	WaTFi	$0.2555_{\pm 0.1112}$	$0.1722_{\pm 0.0344}$
	WaTEi	$0.2578_{\pm 0.1270}$	$0.1735_{\pm 0.0313}$
Raw Time-Series (128 Hz)	EEGNet	$0.2210_{\pm 0.1341}$	$0.1696_{\pm 0.0339}$
	HiLoFuseNet	$0.3061_{\pm 0.1865}$	$0.1697_{\pm 0.0368}$
	Conformer	$0.1700_{\pm 0.1363}$	$0.1756_{\pm 0.0356}$
	ST-EEGFormer-s	$\mathbf{0.4000_{\pm 0.1398}}$	$\mathbf{0.1608_{\pm 0.0304}}$
	ST-EEGFormer-b	$0.3931_{\pm 0.1619}$	$0.1634_{\pm 0.0339}$
	ST-EEGFormer-l	$0.2865_{\pm 0.1400}$	$0.1760_{\pm 0.0367}$

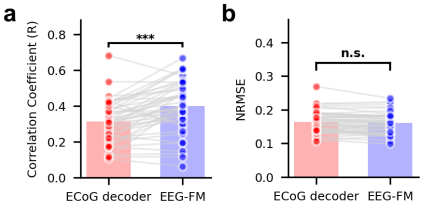

Figure 3: Comparison of the top EEG-FM and the top ECoG decoder across (a) *R* and (b) *NRMSE*. Data points ($N = 45$) represent individual fingers across all 9 subjects.

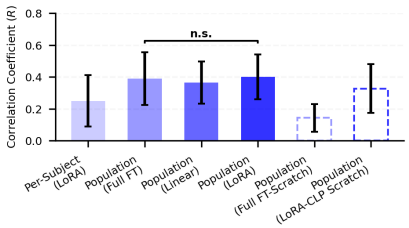

Figure 4: Comparison of EEG-FM adaptation strategies. Only non-significant ($n.s.$) pairs are explicitly labeled.

performing conventional ECoG decoder (Random Forest using Time-Frequency features, $R = 0.3137_{\pm 0.1171}$; Figure 3a). Although the adapted EEG-FM also achieved the lowest error ($NRMSE = 0.1608_{\pm 0.0304}$), the performance gain was not statistically significant compared to the Random Forest (RF) model using physiological features ($NRMSE = 0.1632_{\pm 0.0326}$; Figure 3b). Qualitative visualization (Figure 1c) highlights that conventional ECoG decoders (red, RF with Time-Frequency features) typically track only the coarse, low-frequency trends of finger flexion. In contrast, the adapted EEG-FM (blue, ST-EEGFormer-s) demonstrates a remarkable ability to predict intricate finger flexion dynamics (detailed trajectory predictions are provided in Appendix A.7).

Interestingly, the smallest variant (ST-EEGFormer-s) consistently outperformed its larger counterparts, including the Base and Large versions (Table 1). There are two primary potential explanations for these findings. First, the downstream adaptation employed LoRA parameters optimized specifically for the Small version. Applying these same configurations to the Base and Large models may have constrained their respective representational capacities. Second, given the relatively constrained scale of the *fingerflex* dataset, more compact FM backbones appear to be less susceptible to overfitting, thereby facilitating more robust cross-modal generalization.

Regarding downstream adaptation strategies (Figure 4), population-level methods–including full fine-tuning (Full FT), linear probing (Linear), and LoRA–significantly outperformed per-subject adaptation. This disparity suggests that leveraging task-relevant data across subjects is essential for stabilizing cross-modality transfer from EEG to ECoG. Furthermore, control experiments utilizing randomly initialized EEG-FM positional embeddings (LoRA-CLP Scratch) and trainable parameters (Full FT-Scratch) yielded markedly inferior decoding performance under identical training protocols. This performance gap underscores the critical role of pretrained representations in providing a robust prior for intracranial decoding. Among various population-level methods, LoRA adaptation achieved the highest mean correlation coefficient, though it did not significantly surpass full fine-tuning ($p > 0.05$, denoted by n.s.). Notably, the relatively poor performance of linear probing indicates that merely adjusting the CLP and regression head is insufficient. Updating the internal weights of the pretrained EEG-FM is critical for bridging the domain gap between EEG and ECoG.

Finally, visualization of the ECoG-to-EEG spatial projections (weighting vectors from the affinity matrix $\mathbf{W}$) reveals that the learned spatial weights exhibit a distributed global attention pattern rather than a strictly localized anatomical alignment (Figure A.3).

4 CONCLUSION

We demonstrate for the first time that EEG foundation models can be adapted for predicting continuous finger movements from intracranial ECoG signals, and even outperform conventional ECoG decoders using low-frequency features. Future research will aim to generalize these findings across a broader range of behavioral tasks and evaluate the model's performance with high-density brain recordings and high-frequency information to further enhance decoding performance.

5 ACKNOWLEDGMENTS

QS is supported by China Scholarship Council (No. 202206050022). LY is supported by the Belgian Fund for Scientific Research—Flanders (1S65622N). ECM is supported by the Belgian Fund for Scientific Research—Flanders (1102925N). MMVH is supported by research grants received from Horizon Europe's Marie Sklodowska-Curie Action (grant agreement No. 101118964), Horizon 2020 research and innovation programme under grant agreement No. 857375, the special research fund of the KU Leuven (C24/18/098), the Belgian Fund for Scientific Research – Flanders (G0A4321N, G0C1522N, G031426N), and the Hercules Foundation (AKUL 043). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

REFERENCES

Alim Louis Benabid, Thomas Costecalde, Andrey Eliseyev, Guillaume Charvet, Alexandre Verney, Serpil Karakas, Michael Foerster, Aurélien Lambert, Boris Morinière, Neil Abroug, et al. An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration. *The Lancet Neurology*, 18(12):1112–1122, 2019.

Andrey Eliseyev and Tetiana Aksenova. Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. *PloS one*, 11(5):e0154878, 2016.

Axel Faes, Flavio Camarrone, and Marc M Van Hulle. Single finger trajectory prediction from intracranial brain activity using block-term tensor regression with fast and automatic component extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7):8897–8908, 2022.

Quentin Ferdinand, Rémi Souriau, Lucas Struber, Henri Lorach, Philippe Ciuciu, Marina Reyboz, and Tetiana Aksenova. Ecog-based movement classification and limbs 3d translation prediction: a deep learning study. In *2025 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10. IEEE, 2025.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL `https://arxiv.org/abs/2106.09685`.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. *arXiv preprint arXiv:2405.18765*, 2024.

JOJWGSJ Kubanek, Kai J Miller, Jeffrey G Ojemann, Jonathan R Wolpaw, and Gerwin Schalk. Decoding flexion of individual fingers using electrocorticographic signals in humans. *Journal of neural engineering*, 6(6):066001, 2009.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Zhanhui Lin, Xinyu Jiang, Chenyun Dai, and Fumin Jia. Towards real time efficient and robust ecog decoding for mobile brain-computer interface. *Journal of Neural Engineering*, 2025.

Henri Lorach, Andrea Galvez, Valeria Spagnolo, Felix Martel, Serpil Karakas, Nadine Intering, Molywan Vat, Olivier Faivre, Cathal Harte, Salif Komi, et al. Walking naturally after spinal cord injury using a brain–spine interface. *Nature*, 618(7963):126–133, 2023.

Eva Calvo Merino, A Faes, and MM Van Hulle. The role of distinct ecog frequency features in decoding finger movement. *Journal of Neural Engineering*, 20(6):066014, 2023.

Sean L Metzger, Kaylo T Littlejohn, Alexander B Silva, David A Moses, Margaret P Seaton, Ran Wang, Maximilian E Dougherty, Jessie R Liu, Peter Wu, Michael A Berger, et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 620(7976):1037–1046, 2023.

Kai J Miller. A library of human electrocorticographic data and analyses. *Nature human behaviour*, 3(11):1225–1235, 2019.

Kai J Miller, Dora Hermes, Christopher J Honey, Adam O Hebb, Nick F Ramsey, Robert T Knight, Jeffrey G Ojemann, and Eberhard E Fetz. Human motor cortical activity is selectively phase-entrained on underlying rhythms. 2012.

Tjaša Mlinarič, Laure Spruyt, Elvira Khachatryan, Benjamin Wittevrongel, Mariska Reinartz, Koen Van Laere, Patrick Dupont, Rik Vandenberghe, and Marc M Van Hulle. Early aperiodic EEG changes in preclinical and prodromal alzheimer's disease. *Alzheimer's Research & Therapy*, 2026.

Sajad Mousavi, Fatemeh Afghah, and U Rajendra Acharya. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PloS one*, 14(5):e0216456, 2019.

Patrick Ofner, Andreas Schwarz, Joana Pereira, Daniela Wyss, Renate Wildburger, and Gernot R Müller-Putz. Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal cord injury. *Scientific reports*, 9(1):7134, 2019.

Steven M Peterson, Zoe Steine-Hanson, Nathan Davis, Rajesh PN Rao, and Bingni W Brunton. Generalized neural decoders for transfer learning across participants and recording modalities. *Journal of Neural Engineering*, 18(2):026014, 2021.

Gerwin Schalk and Eric C Leuthardt. Brain-computer interfaces using electrocorticographic signals. *IEEE reviews in biomedical engineering*, 4:140–154, 2011.

Maciej Śliwowski, Matthieu Martin, Antoine Souloumiac, Pierre Blanchart, and Tetiana Aksenova. Deep learning for ECoG brain-computer interface: end-to-end vs. hand-crafted features. In *International Conference of the Italian Association for Artificial Intelligence*, pp. 358–373. Springer, 2022.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

Qiang Sun, Eva Calvo Merino, Bob Van Dyck, Yuan Yang, Jiayuan He, and Marc M Van Hulle. Spectro-temporal fusion of high-gamma and low-frequency ecog signals for intracranial finger movement decoding. *Authorea Preprints*, 2025.

Jesper Tveit, Harald Aurlien, Sergey Plis, Vince D Calhoun, William O Tatum, Donald L Schomer, Vibeke Arntsen, Fieke Cox, Firas Fahoum, William B Gallentine, et al. Automated interpretation of clinical electroencephalograms using artificial intelligence. *JAMA neurology*, 80(8):805–812, 2023.

Ksenia Volkova, Mikhail A Lebedev, Alexander Kaplan, and Alexei Ossadtchi. Decoding movement from electrocorticographic activity: a review. *Frontiers in neuroinformatics*, 13:74, 2019.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. *arXiv preprint arXiv:2412.07236*, 2024.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36:78240–78260, 2023.

Liuyin Yang, Qiang Sun, Ang Li, and Marc M. Van Hulle. Are EEG foundation models worth it? comparative evaluation with traditional decoders in diverse BCI tasks. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=5Xwm8e6vbh.

Lin Yao, Bingzhao Zhu, and Mahsa Shoaran. Fast and accurate decoding of finger movements from ECoG through riemannian features and modern machine learning techniques. *Journal of Neural Engineering*, 19(1):016037, 2022.

Yuxuan Yao, Hongbo Wang, Li Chen, Yiheng Peng, and Jingjing Luo. Foundation models for EEG decoding: current progress and prospective research. *Journal of Neural Engineering*, 22(6): 061002, 2025.

Qibin Zhao, Cesar F Caiafa, Danilo P Mandic, Zenas C Chao, Yasuo Nagasaka, Naotaka Fujii, Liqing Zhang, and Andrzej Cichocki. Higher order partial least squares (hopls): A generalized multilinear regression method. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1660–1673, 2012.

A Appendix

A.1 Related Works

While EEG-FMs have yet to be applied to ECoG-based tasks, Peterson et al. (2021) have explored the cross-modality generalization of supervised neural decoders between EEG and ECoG. Their approach involves mapping electrode positions to localized, predefined brain regions to account for variations in electrode placement across subjects and modalities. Notably, they demonstrated that fine-tuning a pre-trained decoder with data from the unseen modality allows performance to approach that of decoders specifically tailored for that modality (Peterson et al., 2021). However, these findings are restricted to binary movement detection, which fails to capture the complex behaviors that ECoG signals are capable of decoding. To date, ECoG research has primarily relied on hand-crafted features—such as physiological features (Faes et al., 2022; Yao et al., 2022) or time-frequency features (Lorach et al., 2023; Ferdinand et al., 2025)—which have frequently outperformed end-to-end deep learning models in decoding performance (Śliwowski et al., 2022). Unlike EEG, ECoG offers broader bandwidth; in particular, the high-gamma band (60–200 Hz) provides the most critical information for decoding (Volkova et al., 2019; Merino et al., 2023).

A.2 Dataset and preprocessing details

The Stanford *fingerflex* dataset (Miller, 2019) contains recordings from nine epilepsy patients implanted with subdural ECoG grids (4 mm diameter, 1 cm pitch) across various cortical regions[2]. Participants performed self-paced flexion-extension of individual fingers in response to visual cues. Each finger was tested in 30 interleaved trials, consisting of 2 seconds of movement (typically 2–5 flexions) followed by 2 seconds of rest. ECoG signals (band-pass filtered at 0.3–200 Hz) and finger trajectories (recorded via a digital glove) were sampled at 1000 Hz and 25 Hz, respectively.

Continuous ECoG signals were first processed with a one-pass, 3rd-order Butterworth IIR band-stop filter to eliminate 60 Hz power line interference and its harmonics (120 Hz, 180 Hz). Following visual inspection to exclude abnormal channels (see Table A.1), the signals underwent Common Average Referencing (CAR). All preprocessing steps were strictly causal to ensure compatibility with real-time applications. To align the modalities, we compensated for a 40 ms instrumental latency between the ECoG signals and glove trajectories. Data were partitioned chronologically in line with (Lin et al., 2025; Sun et al., 2025): for recordings exceeding 600 seconds, the first 400 seconds served as the training set; otherwise, a 2:1 training-to-testing split was applied (see Table A.1).

Training and testing samples were generated using a sliding window approach; a 1-second lookback window preceded the movement trajectory to maintain causality, with a 40 ms stride to match the 25 Hz glove sampling rate. Each 1-second ECoG segment was downsampled to 128 Hz following anti-aliasing filtration to align with the input of the ST-EEGFormer.

A.3 Feature extraction

Physiological features Each ECoG segment was decomposed into six frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta1 (12–24 Hz), beta2 (24–34 Hz), and gamma1 (34–60 Hz) using a 4th-order Butterworth IIR filter. For each band, we calculate the log-transformed average power within moving sub-windows to reduce the sampling rate to 16 Hz. Additionally, we extract the Local Motor Potential (LMP) by applying a uniform moving average filter (window sized 8, at a 128 Hz sampling rate) to the raw signal, capturing the low-frequency fluctuations. In the end, the 1-second ECoG segment results in a feature tensor sized $C \times 16 \times 7$.

[2]Ethics statement: All patients participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12193). All patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate. These data originally appeared in the manuscript "Human Motor Cortical Activity Is Selectively Phase- Entrained on Underlying Rhythms" published in PLoS Computational Biology in 2012 (Miller et al., 2012)

Table A.1: Summary of ECoG Datasets Across Subjects

Subject	Channels	Train Duration (s)	Test Duration (s)
1 (bp)	46	400	≈ 210
2 (cc)	63	400	≈ 210
3 (ht)	64	400	≈ 210
4 (jc)	47	≈ 353	≈ 176
5 (jp)	55[a]	≈ 310	≈ 155
6 (mv)	42[b]	≈ 118	≈ 59
7 (wc)	64	400	≈ 210
8 (wm)	38	≈ 296	≈ 148
9 (zt)	61	400	≈ 210

[a] Excludes channels 31, 32, and 37. [b] Excludes channel 40.

Time-Frequency Features Each ECoG segment was transformed by the Morlet Wavelet. We defined a bank of 10 frequencies linearly spaced between 5 Hz and 60 Hz. The number of cycles for the mother wavelet is scaled relative to the frequency to optimize the trade-off between temporal and spectral resolution. The resulting scalograms are downsampled to 16 Hz via mean aggregation and log-transformed ($10 \cdot \log_{10}$) to stabilize variance. In the end, the 1-second ECoG segment results in a feature tensor sized $C \times 16 \times 10$.

Raw Time-Series The raw ECoG time-series (sized $C \times 128$) is preserved without manual feature engineering.

A.4 Details of ECoG decoder implementation

The ECoG regression model selection spanned linear estimators to tailored neural architectures, including: (1) Linear & Ensemble Models: Partial Least Squares (PLS), N-Way PLS (NPLS) (Eliseyev & Aksenova, 2016), Higher Order PLS (HOPLS) (Zhao et al., 2012), and Random Forest (RF) (Lin et al., 2025). (2) Deep Learning Architectures: Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), EEGNet (Lawhern et al., 2018), CNN-LSTM (Lin et al., 2025), and HiLoFuseNet (Sun et al., 2025). (3) Transformers: Conformer (Song et al., 2022) and Wavelet Transformers (WaT, WaTFi, WaTEi) (Ferdinand et al., 2025).

Given a raw ECoG segment sampled at 128 Hz with dimensions $C \times T$, we perform feature extraction to generate a spectral-temporal tensor of size $C \times T' \times F'$. Specifically, for physiological features, the reduced temporal dimension is $T' = 16$ with $F' = 7$ frequency components. For time-frequency features, the spectral dimension is increased to $F' = 10$. Detailed input formats and hyperparameter configurations for each ECoG decoder are provided in Table A.2.

For neural network training, we utilize the Adam optimizer with a constant learning rate of 10^{-3} and a mean squared error (MSE) loss function. The training process is governed by an early stopping criterion with a patience of 10 epochs, monitored on a validation set comprising the final 10% of the training data (partitioned chronologically). Training is conducted for a maximum of 100 epochs with a batch size of 128. To ensure reproducibility across all experiments, the random seed is fixed to 42.

A.5 Details of downstream adaptation strategy

The pre-trained ST-EEGFormer has three parameter-scaled configurations: small (32.7M), base (110.9M), and large (328.4M). To evaluate the transferability of the pre-trained ST-EEGFormer, we conducted a systematic investigation of various downstream adaptation strategies using the compact variant, ST-EEGFormer-s. These strategies range from per-subject LoRA adaptation to population-level linear probing, LoRA adaptation, and full-parameter fine-tuning.

Furthermore, to quantify the specific contribution of EEG-FM pre-training, we included two ablation baselines where the FM weights were randomly initialized: (i) a full fine-tuning from scratch variant to assess the gain over a non-pre-trained backbone, and (ii) a CLP-ablation variant to isolate the impact of the pre-trained EEG-channel positional embeddings. Detailed trainable parameter counts and architectural configurations for each strategy are summarized in Table A.3.

Table A.2: Hyperparameter Configurations of ECoG Decoders

Model	Input Format	Configuration Details
PLS	$C \times T' \times F'$	Latent components: $[1, 30]$.
NPLS	$C \times T' \times F'$	Latent components: $[1, 30]$.
HOPLS	$C \times T' \times F'$	Latent components: $[1, 30]$; Decomposition rank: $[1, 20]$.
RF	$C \times T' \times F'$	$n_{estimators} = 200$; $max\_depth \in \{10, 20, \text{None}\}$; $min\_samples\_leaf \in \{2, 5, 10\}$; $max\_features =$ 'sqrt'.
MLP	$C \times T' \times F'$	3-layer MLP; $d_{hidden} = 256$; $p_{dropout} = 0.5$.
LSTM	$T' \times (C \cdot F')$	1-layer LSTM ($d_h = 256$); 2-layer MLP ($d_h = 128$); $p_{dropout} = 0.5$.
CNN-LSTM	$C \times T' \times F'$	Architecture follows (Lin et al., 2025); $p_{dropout} = 0.5$.
WaT	$T' \times (C \cdot F')$	Architecture follows (Ferdinand et al., 2025); 2-layer Transformer; $n_{head} = 4$; $d_{model} = 32$.
WaTFi	$T' \times C$	Architecture follows (Ferdinand et al., 2025); Frequency-independent; 2-layer Transformer; $n_{head} = 4$; $d_{model} = 16$.
WaTEi	$T' \times F'$	Architecture follows (Ferdinand et al., 2025); Electrode-independent; 2-layer Transformer; $n_{head} = 4$; $d_{model} = 16$.
EEGNet	$C \times T$	Architecture follows (Lawhern et al., 2018); $p_{dropout} = 0.5$.
HiLoFuseNet	$C \times T$	Architecture follows (Sun et al., 2025). Only keeping the low-frequency branch.
Conformer	$C \times T$	Architecture follows (Song et al., 2022); $L = 4$; $n_{head} = 4$.

To further train those foundation models, we utilize the AdamW optimizer with a 10-epoch warm-up period followed by cosine annealing, minimizing the mean squared error (MSE) loss. The training process is governed by an early stopping criterion with a patience of 30 epochs, monitored on a validation set comprising the final 10% of the training data (partitioned chronologically). All experiments maintain a batch size of 128 and a fixed random seed of 42 for reproducibility. For linear probing, training runs for 200 epochs with a learning rate of 3×10^{-3}. For LoRA fine-tuning (rank $r = 4$, dropout 0.2, $\alpha = 16$), we train for 100 epochs using a learning rate of 3×10^{-3} for small and base models, while large models use 3×10^{-4} to ensure stability. Finally, full fine-tuning is conducted for 100 epochs with a learning rate of 3×10^{-4}.

Table A.3: Downstream Adaptation Strategies and Parameter Details

Strategy	Cross-Subj.	Params	Description
Per-Subject (LoRA)	✗	0.179M	Updates CLP, regression head, and LoRA adapters in the last 4 encoder layers.
Population (Full FT)	✓	25.938M	Full fine-tuning of all parameters (backbone, CLP, and regression head).
Population (Linear)	✓	0.121M	Trains CLP and regression head only; backbone remains frozen.
Population (LoRA)	✓	0.179M	Updates CLP, regression head, and LoRA adapters in the last 4 encoder layers.
Ablation Variants			
Population (Full FT–Scratch)	✓	25.938M	Full fine-tuning of all components using random initialization.
Population (LoRA–CLP Scratch)	✓	0.179M	Re-initializes CLP positional embeddings randomly; updates CLP, head, and LoRA adapters.

A.6 Detailed finger movement decoding results

Detailed per-finger decoding results (R) for the five representative decoders—RF (Time-Frequency), HiLoFuseNet, and the three ST-EEGFormer variants—are summarized in Table A.4. While the adapted ST-EEGFormer-s achieved the highest overall performance ($R = 0.3889_{\pm 0.1500}$), individual subject analysis reveals nuanced results. For instance, HiLoFuseNet and RF outperformed the EEG-FM in specific subjects, such as Subjects 3, 4 (HiLoFuseNet) and Subject 5, 8 (RF), respectively.

A.7 Visualization of predicted movement trajectories

Representative trajectory predictions for **Subject cc** (the top-performing subject according to Table A.4) are provided in Figure A.1, extending the exemplary results shown in Figure 1c. To further validate the superiority of the adapted EEG-FM (ST-EEGFormer-s), an additional baseline, HiLo-FuseNet, is included for a more comprehensive comparison. These findings provide complementary evidence to the primary results, demonstrating that neural-network-based decoders—specifically

Table A.4: Regression performance (R) comparison of different decoders (mean±std)

Subject	Finger	RF	HiLoFuseNet	ST-EEGFormer-l	ST-EEGFormer-b	ST-EEGFormer-s
1 (bp)	Thumb	0.3321	**0.5651**	0.4602	0.5028	0.3831
	Index	0.2079	0.5284	0.3651	**0.5373**	0.4687
	Middle	0.1122	0.4090	**0.4699**	0.4612	0.4093
	Ring	0.3673	0.5740	0.4286	0.6022	**0.6055**
	Little	0.2760	0.2985	0.3377	0.3702	**0.3806**
	Average	0.2591	0.4750	0.4123	**0.4947**	0.4494
2 (cc)	Thumb	0.4402	0.6587	0.3799	**0.6979**	0.6674
	Index	0.3851	**0.5545**	0.4347	0.5010	0.4589
	Middle	0.3136	0.4903	0.3451	**0.4961**	0.4666
	Ring	0.3910	0.5684	0.5034	**0.6084**	0.6022
	Little	0.4269	0.5586	0.5580	**0.6161**	0.5326
	Average	0.3914	0.5661	0.4442	**0.5839**	0.5455
3 (ht)	Thumb	0.2353	0.2196	0.1568	**0.2643**	0.1937
	Index	0.3075	**0.4294**	0.3031	0.3446	0.3396
	Middle	0.1037	**0.1665**	0.0104	0.0865	0.0624
	Ring	0.3032	**0.4562**	0.3163	0.4185	0.3930
	Little	0.2885	**0.3908**	0.2272	0.2426	0.3274
	Average	0.2477	**0.3325**	0.2027	0.2713	0.2632
4 (jc)	Thumb	0.3194	**0.6130**	0.3967	0.4613	0.5306
	Index	0.2764	**0.5582**	0.4056	0.4523	0.5205
	Middle	0.1986	0.1825	0.1999	0.2177	**0.2453**
	Ring	0.3076	0.4542	0.2937	0.4579	**0.4589**
	Little	0.2415	**0.5136**	0.2791	0.3678	0.3676
	Average	0.2687	**0.4643**	0.3150	0.3914	0.4246
5 (jp)	Thumb	**0.6819**	0.0369	0.4857	0.5367	0.5691
	Index	0.4527	0.2387	0.3705	**0.4889**	0.4881
	Middle	**0.1699**	0.1030	0.1374	0.1524	0.1474
	Ring	0.4339	0.2781	0.3265	0.4248	**0.4474**
	Little	0.3191	-0.0385	0.3053	**0.3773**	0.3678
	Average	**0.4115**	0.1236	0.3251	0.3960	0.4040
6 (mv)	Thumb	0.1742	0.2340	0.1857	**0.4709**	0.4065
	Index	0.5354	0.2163	0.4816	**0.6040**	0.5315
	Middle	0.2672	0.1241	0.0728	0.3314	**0.3455**
	Ring	0.3691	0.2085	0.3213	0.5206	**0.5317**
	Little	0.4181	0.1416	0.3897	**0.6437**	0.6077
	Average	0.3528	0.1849	0.2902	**0.5141**	0.4846
7 (wc)	Thumb	0.3533	0.3160	0.2449	0.4026	**0.4221**
	Index	0.4112	0.3253	0.2187	0.4755	**0.5435**
	Middle	0.1782	0.2768	0.2572	0.3516	**0.3670**
	Ring	0.2515	0.2656	0.0772	0.1706	**0.3249**
	Little	0.2209	0.0756	0.1521	0.2007	**0.2593**
	Average	0.2830	0.2519	0.1900	0.3202	**0.3834**
8 (wm)	Thumb	0.3680	0.0799	0.0139	0.1359	**0.3757**
	Index	0.1271	0.1092	0.0876	0.0747	**0.1320**
	Middle	0.1727	0.2481	0.1385	**0.2558**	0.1938
	Ring	**0.4014**	0.0244	0.1421	0.2367	0.2989
	Little	**0.1154**	0.0154	0.0007	-0.0136	**0.1154**
	Average	**0.2369**	0.0954	0.0766	0.1379	0.2232
9 (zt)	Thumb	**0.4268**	0.3982	0.4110	0.3778	0.2878
	Index	0.3573	0.1392	0.2982	0.4374	**0.4534**
	Middle	0.3410	0.1506	0.3080	0.4618	**0.4695**
	Ring	0.4261	0.3759	0.3312	0.4704	**0.4981**
	Little	0.3079	0.2417	0.2629	0.3935	**0.4026**
	Average	0.3718	0.2611	0.3223	**0.4282**	0.4223
Grand Average		$0.3137_{\pm 0.1171}$	$0.3061_{\pm 0.1865}$	$0.2865_{\pm 0.1400}$	$0.3931_{\pm 0.1619}$	$\mathbf{0.4000_{\pm 0.1398}}$
p-Value		***	*	***	$n.s.$	-

Asterisks indicate the significance level (with FDR correction) between the ST-EEGFormer-s and counterparts.

the adapted EEG-FM—capture more flexion dynamics than traditional machine learning models like RF.

We next compared the adapted EEG-FM to the state-of-the-art (SOTA) decoder, HiLoFuseNet (Sun et al., 2025), which leverages both low-frequency signals and high-gamma activity (70–200 Hz). As shown in Figure A.2, although the adapted EEG-FM yields lower overall performance metrics compared to the SOTA decoder, it exhibits a remarkable ability to track the complex temporal dynamics and flexion trends of all five fingers. Specifically, the EEG-FM closely mirrors the ground-truth trajectories during movement onset and peak transitions, even without access to the information-rich high-gamma band—a component widely considered critical for precise intracranial decoding. The remaining performance gap, particularly in terms of peak magnitude estimation and baseline noise suppression, likely stems from the absence of these high-frequency features in our current model input. Nevertheless, these results underscore the efficacy of cross-modality transfer, suggesting that large-scale pre-training allows the model to extract significant motor information from low-frequency representations alone.

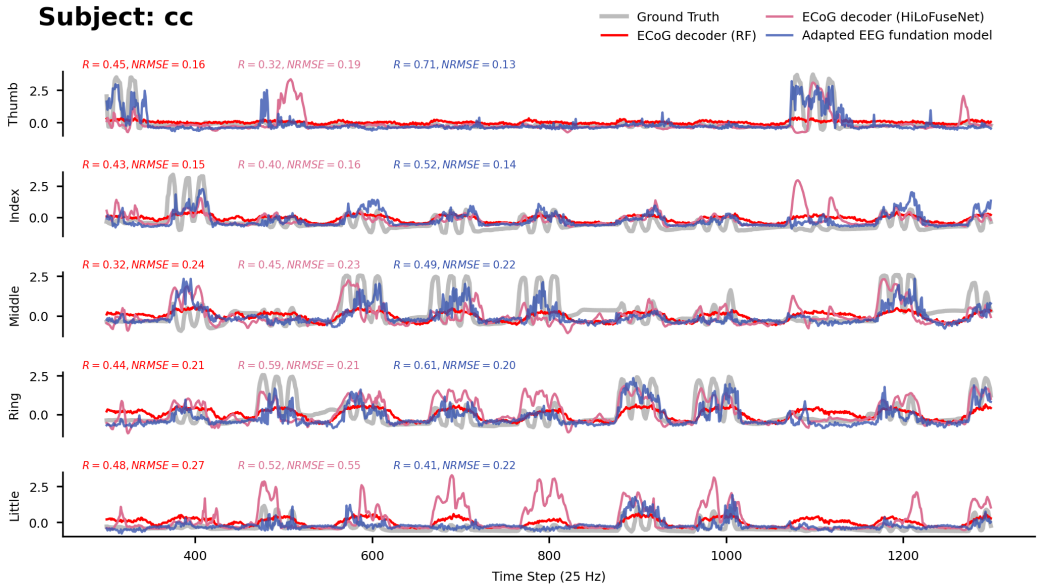

Figure A.1: Comparison of movement trajectory predictions for Subject cc. The adapted EEG-FM (ST-EEGFormer-s) is benchmarked against conventional ECoG decoders RF (time-frequency features) and HiLo-FuseNet.

A.8 Visualization of ECoG-to-EEG spatial projections

Figure A.3 visualizes the learned spatial weights of the CLP module, which define the mapping from scalp EEG channels to intracranial ECoG electrodes. Subjects **cc** and **mv** were selected for visualization as they exhibited the highest decoding performance across the cohort, as detailed in Table A.4. Each topomap represents a weighting vector from the affinity matrix $\mathbf{W}$, illustrating how information from the EEG cap converges to represent a specific ECoG electrode. Unexpectedly, the learned spatial weights exhibit a distributed global projection pattern rather than a strictly localized anatomical alignment. For instance, projections for the M1 electrode (**Subjects cc**) incorporate contributions from across the scalp. However, more investigation is needed to interprete the findings.

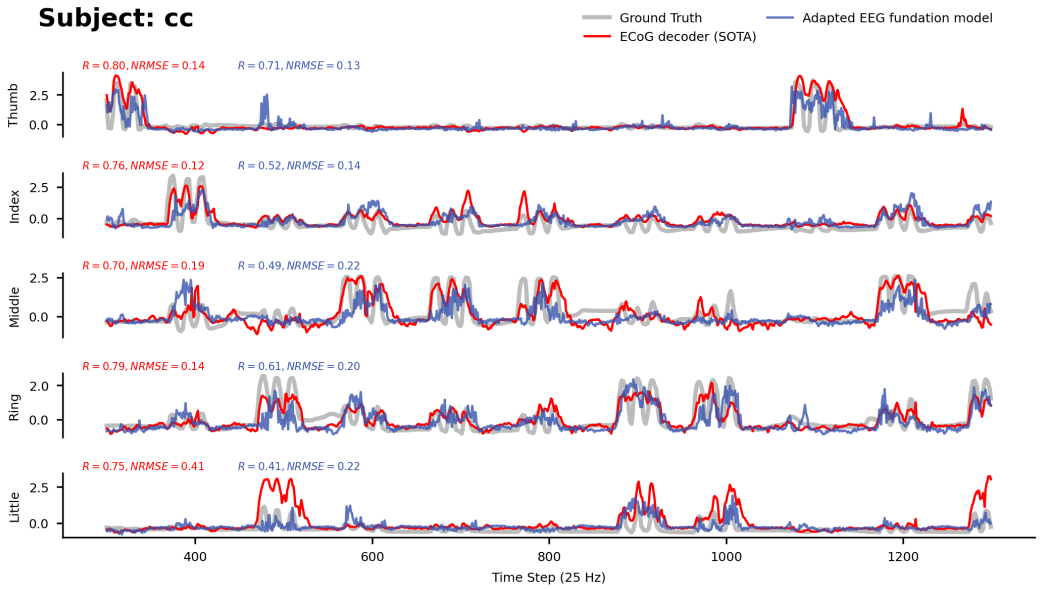

Figure A.2: Comparison of movement trajectory predictions for Subject cc. The adapted EEG-FM (ST-EEGFormer-s) is benchmarked against the SOTA ECoG decoder.

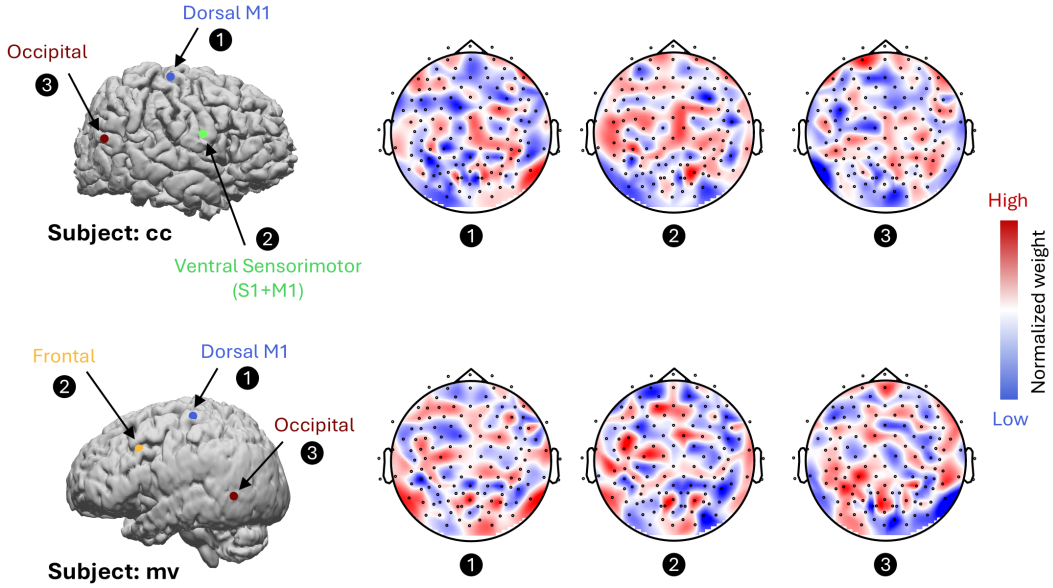

Figure A.3: Spatial projection weights for representative ECoG electrodes in **Subjects cc** and **mv**. The brain plots (left) indicate the anatomical sites of the ECoG electrodes, while the topomaps (right) show the corresponding EEG weight distributions from the learned matrix $\mathbf{W}$.