AFRL-RI-RS-TR-2021-084

MULTIMODAL NEURAL DECODING: DATA-INTENSIVE APPROACHES TO UNDERSTANDING LONG-TERM, UNLABELED HUMAN BRAIN DATA

UNIVERSITY OF WASHINGTON

MAY 2021

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2021-084 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
STEVEN DRAGER
Work Unit Manager

/ S /
GREGORY HADYNSKI
Assistant Technical Advisor,
Computing & Communications Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
MAY 2021	FINAL TECHNICAL REPORT	NOV 2017 – NOV 2020

4. TITLE AND SUBTITLE	
MULTIMODAL NEURAL DECODING: DATA-INTENSIVE APPROACHES TO UNDERSTANDING LONG-TERM, UNLABELED HUMAN BRAIN DATA	**5a. CONTRACT NUMBER** FA8750-18-2-0259
	5b. GRANT NUMBER N/A
	5c. PROGRAM ELEMENT NUMBER 61101E
6. AUTHOR(S)	**5d. PROJECT NUMBER** DSD2
Bingni W. Brunton	**5e. TASK NUMBER** SU
	5f. WORK UNIT NUMBER WA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
University of Washington 4333 Brooklyn Ave NE, Box 359472 Seattle WA 98195-9472	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL/RITA 525 Brooks Road Rome NY 13441-4505	DARPA 675 North Randolph Street Arlington, VA 22203-2114	AFRL/RI
		11. SPONSOR/MONITOR'S REPORT NUMBER AFRL-RI-RS-TR-2021-084

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Fully automated decoding of human actions and intentions from neural signals is a major challenge in human-computer interactions. The current success of brain-computer interfaces (BCI) has hinged on labeled training data in laboratory conditions. To deploy BCIs in real-life, one must develop robust strategies to handle naturalistic disturbances. This research effort focused on decoding movements from large-scale human intracranial brain recordings, video, and audio, all continuously acquired over one week. Importantly, here the participants are simply behaving as they wish. This research demonstrated the ability to decode movements from neural recordings as well as that this decoder generalizes remarkably well on application to unseen participants.

15. SUBJECT TERMS

Brain-computer interfacing, transfer learning, machine learning, naturalistic human movements, intracranial brain recordings

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**STEVEN DRAGER**
U	U	U	UU	18	**19b. TELEPHONE NUMBER** (*Include area code*) **N/A**

TABLE OF CONTENTS

TABLE OF FIGURES

1.0 SUMMARY

Fully automated decoding of human actions and intentions through neural signals is a tantalizing challenge in human-computer interactions. The current success of brain-computer interfaces (BCI)—controlling robotic prostheses and computer software via brain signals—has hinged on availability of labeled training data collected in carefully controlled laboratory conditions. To deploy BCIs in practical, real-life applications, one must develop robust strategies and algorithms that can handle naturalistic disturbances and self-adapt to context.

This research focused on developing supervised, unsupervised, and semi-supervised approaches to decode neural states from long-term brain recordings acquired in a naturalistic setting. The multi-modal dataset comprises large-scale human intracranial brain recordings, video, audio, and depth recordings, all continuously and simultaneous acquired over at least one week. Importantly, unlike the majority of previous data sets used to train neural decoders, here the subjects being monitored are not instructed to perform specific tasks but are simply behaving as they wish, as shown in Figure 1.

To summarize the results from this 3-year seedling effort, this project has enabled, through direct funding as well as research opportunities leveraged through synergistic projects, the training of 2 postdoctoral researchers, 4 Ph.D. students, 1 master's student, and 5 undergraduate students. Research results have been described in 4 papers and 2 published conference proceedings; 1 paper has appeared as a publication in the Journal of Neural Engineering, and the other 3 papers are in various stages of peer-review (available as preprints). In addition, the research team has co-authored a review paper on the subject of data-driven modeling in human neuro-engineering.

2.0 INTRODUCTION

Brain-computer interfaces that interpret neural activity to control robotic or virtual devices have shown tremendous potential for assisting patients with neurological disabilities, including motor impairments, sensory deficits, and mood disorders. At the same time, brain-computer interfaces offer new insights about the function of neural circuits, including how sensorimotor information is represented in the brain.

Advances in brain-computer interfaces have been driven in part by improved neural decoding algorithms. Even so, these impressive demonstrations have relied on finely tuned models trained on experimentally derived labeled data acquired in well-controlled laboratory conditions. Thus, the remarkable feats of neural decoding to mobilize patients who have lost use of their limbs remain untested outside the laboratory.

One key challenge is how neural decoding may be approached "in the wild," where sources of behavioral and recording variability are significantly larger than what is found in the lab. Further, neural responses are known to differ between experimental and freely behaving conditions. However, it can be difficult to collect enough data to train decoders, especially given the non-stationary nature of the recorded signals, leading to decoders that generalize poorly to new data and require frequent re-calibrations.

This project, leveraged continuous clinical recording data from humans undergoing clinical epilepsy to develop supervised, unsupervised, and semi-supervised methods to understand and decode behavior from neural recordings. The research took advantage of recent innovations in computer vision, machine learning, and data-driven dynamic modeling to tackle the challenge of understanding this large and unstructured data. In addition to the research publications and presentations, all of the code developed during this project is available as GitHub repositories (links in each paper), and the curated datasets (minus personally identifying information) have been published as datasets to facilitate future developments of BCI's on naturalistic behaviors.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Dataset collection and curation

Over the course of this project opportunistic clinical recordings were collected from >60 participants during their clinical epilepsy monitoring (conducted at Harborview Medical Center in Seattle, WA). Of these, 12 participants (8 males, 4 females) were chosen for detailed analysis and modeling, due to the positions of their electrodes, their general activity level, and the cumulative duration of their monitoring. Participants were 29.4±7.9 years old at the time of recording (mean±standard deviation (SD)). The study was approved by the University of Washington Institutional Review Board for the protection of human participants. All participants provided written informed consent.

Participants for the study were selected who had Electrocorticography (ECoG) electrode coverage near primary motor cortex, with either one 8×8 or two 4×8 electrode grids placed subdurally on the cortical surface. Additional electrodes were implanted on the cortical surface for some participants, resulting in 87.0±12.9 total surface electrodes per participant (mean±SD). In addition, five participants had 23.2±12.1 intracortical depth electrodes (mean±SD). Electrodes were implanted primarily within one hemisphere for each participant (5 right hemi-sphere, 7 left hemisphere).

Participants underwent 24-hour clinical monitoring, involving semi-continuous ECoG and audio/video recordings over 7.4±2.2 days per participant (mean±SD). Some breaks occurred throughout monitoring (on average, 8.3±3.2 total breaks per participant, each lasting 1.9±2.4 hours [mean±SD]). For all participants, analysis was restricted to days 3–7 following the electrode implantation surgery, in order to exclude potentially anomalous neural and behavioral activity immediately following electrode implantation surgery. For several participants, some days were excluded due to corrupted or missing data files. During clinical monitoring, participants were observed during a variety of typical everyday activities, such as eating, sleeping, watching television, and socializing while confined to a hospital bed. ECoG and video were initially sampled at 1000 Hz and 30 frames per second, respectively.

The curated dataset has been made publicly available. The pre-processed ECoG, time-synced upper-limb joint tracking, and metadata on movement events extracted can be downloaded at https://figshare.com/s/ef4ea24d67d16233f73d .

3.2 Investigating naturalistic hand movements by behavior mining in long-term video and neural recordings

The project developed and validated a pipeline to extract temporally precise, interpretable movement events, by processing the video data through pose-estimation, pose time-series segmentation, event detection and finally, event metadata extraction [1]. To extract a participant's pose from raw video, a state-of-the-art markerless pose estimation tool was trained using manual annotations of 9 keypoints on each participant's body (nose, both wrists, elbows, shoulders, and ears). Next, the pose time-series was segmented into discrete, interpretable states, by applying a first-order auto-regressive hidden semi-Markov model (ARHSMM) with two latent states. Then string pattern matching was performed to identify movement initiation (0.5 seconds of rest followed by 0.5 seconds of movement) and rest events (3 continuous seconds of rest).

For each detected movement event, several metadata features were extracted from the continuous pose-dynamics associated with the movement. These include movement-associated metadata like the (x, y) coordinates of the keypoint at the start and end of the event, duration of the entire movement (up to next rest state), and rest duration before and after movement. Since people often move both hands at the same time (i.e. "bimanually"), each movement event was augmented with metadata about the opposing wrist's movement, if any.

To examine the neural correlates of naturalistic movement initiation, a time-frequency analysis was performed of the neural recordings by averaging event-locked spectrograms for each participant, using hundreds of movement initiation events chosen to match movement statistics (reach magnitude, onset velocity, and shape) of a previous controlled experimental study. Using the aforementioned metadata to guide the search, up to 200 events per day over 5 days for each of 12 participants was selected, and then further inspection of the video for each event occurred to remove any false positives.

The pipeline was also leveraged as a source of training data for a BCI decoder that detects wrist movement initiation events. Separate classifiers were trained, tailored to each participant, to discriminate between movement initiation events and no-movement events for each wrist using only features derived from the ECoG neural recordings of the 12 participants. The decoder used the Random Forest algorithm on time-frequency spectrograms of the neural data. The decoder was applied to 3 consecutive recording days for each participant, withholding the last day for testing.

3.3 Behavioral and neural variability of naturalistic arm movements

The raw ECoG data was processed using custom minimum norm estimates (MNE) Python scripts, performing high-amplitude artifact removal, band-pass filtering, notch filtering of 60 Hz line noise, and common median re-referencing, as shown in Figure 1 [2]. Electrode positions were localized by co-registering preoperative magnetic resonance imaging (MRI) and postoperative computerized tomography (CT) scans, which were then warped into Montreal Neurological Institute (MNI) space. Markerless pose estimation was performed on the raw video footage separately for each participant to determine wrist positions. Movement states were identified using a first-order autoregressive hidden semi-Markov model to each wrist trajectory. Next, timestamps accompanying clinical recordings were used to synchronize movement initiation events with ECoG recordings, generated 10-second ECoG segments centered around each event, computed spectral power using Morlet wavelets, and projected the spectral power at every electrode into common regions of interest defined by the automated anatomical labeling (AAL) atlas.

Multiple behavioral and environmental metadata features that quantified variations in movement parameters and environmental contexts were extracted. Using the 10 extracted behavioral features as independent variables, a separate linear regression model was fit to the spectral power at every electrode. For the dependent variable, spectral power was averaged over the first half second of movement onset for low-frequency (8–32 Hz) and high-frequency (76–100 Hz) bands. After training each regression model, model validation was performed by computing the R^2 on withheld data and also assessed against the contribution of each behavioral feature. To minimize bias in the selection of training and testing data, 200 random, independent train/test splits for every regression model was performed.

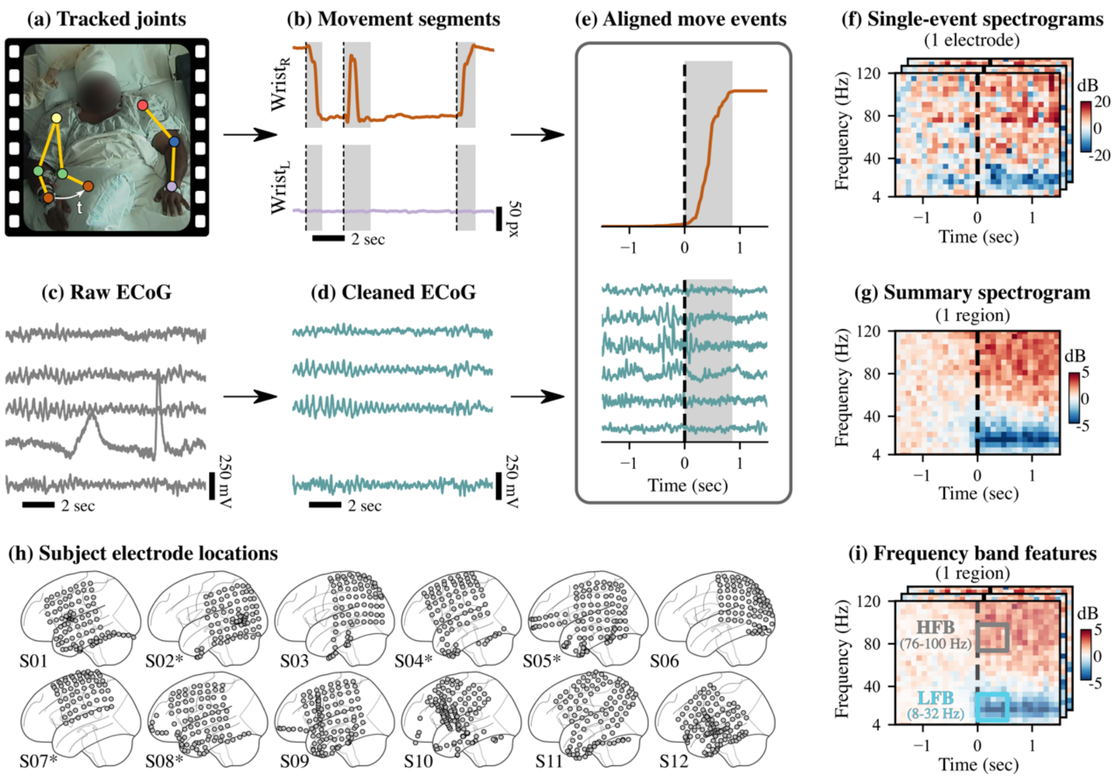

Figure 1: Schematic overview of data processing, analysis, and modeling framework. (a)–(b) Based on continuous video monitoring of each subject, trajectories of the left and right wrists (Wrist_L and Wrist_R in (b)) were estimated using neural networks and automatically segmented into move (gray) and rest (white) states as shown in (b). (c)–(d) Raw multi-electrode electrocorticography (ECoG) was filtered and re-referenced; bad electrodes (e.g., ones with artifacts) were removed from further analysis. (e) Movement onset events detected from video as shown in (b) were aligned with ECoG data using timestamps. (f) For each move event at each electrode, spectral power was computed and visualized as a log-scaled spectrogram. (g) Summarizing across events and electrodes, the spectral power was projected from electrodes onto 8 cortical regions based on anatomical registration and computed the median power across movement events. (h) The data included 12 subjects; their electrode placements are shown in MNI coordinates. Five of the subjects had electrodes implanted in their right hemispheres (denoted by asterisks). For consistency of later analyses, these electrode locations were mirrored as shown. (i) To partially explain the event-by-event neural variability in low-frequency (LFB: 8–32 Hz) and high-frequency (HFB: 76–100 Hz) spectral power, multiple linear regression models were fit at each electrode using behavioral features extracted from the videos.

3.4 Generalized neural decoders for transfer learning across participants and recordings modalities

The HTNet model builds upon EEGNet, a compact convolutional neural network for electroencephalogram (EEG) data. EEGNet has three convolution layers: (1) a one-dimensional convolution analogous to temporal band-pass filtering, (2) a depth-wise convolution to perform spatial filtering, and (3) a separable convolution to identify temporal patterns across the previous filters. HTNet, adds a Hilbert transform layer after the initial temporal convolution to compute relevant spectral power features using a data-driven filter-Hilbert analog [3]. Then a matrix multiplication layer was added to project electrode-level spectral power onto common brain regions of interest, using the pre-computed weight matrices.

HTNet decoding performance was compared against EEGNet, random forest, and minimum distance decoders, tested on the ECoG dataset and a publicly available EEG dataset. The decoding task for both datasets was to classify upper-limb "move" and "rest" events. For the ECoG dataset, the median DC drift and high-amplitude discontinuities were removed, band-pass filtered (1–200 Hz), notch filtered, re-referenced to the common median across electrodes, noisy electrodes removed, and generated as 2-second segments centered around each event. For the EEG dataset, the data was pre-processed by average referencing, 1 Hz high-pass filtering, resampling to 250 Hz, and generating 2-second segments centered around each event.

The decoder performance was assessed during three scenarios: (1) testing on an untrained recording day for the same ECoG participant (tailored decoder), (2) testing on an untrained ECoG participant (same modality), and (3) testing on participants from the EEG dataset after training only on the ECoG dataset (unseen modality), as shown in Figure. ***2***. Hyperparameter tuning was performed to identify optimal values for each decoder. In addition to testing generalizability, how much a generalized HTNet decoder improves when re-trained using data from the test participant, a process known as fine-tuning, was assessed. Lastly, the interpretability of HTNet's trained weights as well as the effects of electrode overlap and number of training participants on cross-participant performance was assessed.

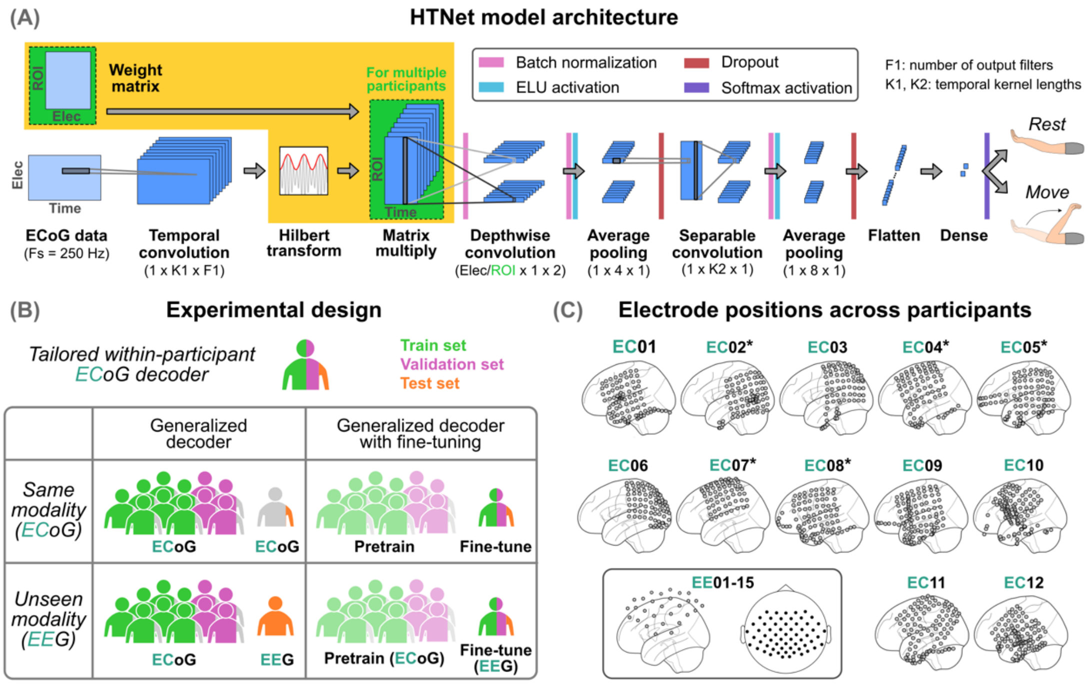

Figure. 2: Overview of HTNet architecture, experimental design, and electrode locations. (A) HTNet is a convolutional neural network architecture that extends EEGNet by handling cross-participant variations in electrode placement and frequency content. (B) Using electrocorticography data, both tailored within-participant and generalized multi-participant models were trained to decode arm movement vs. rest. Multi-participant decoders were tested separately on held-out data from unseen participants recorded with either the same modality as the train set (ECoG) or an unseen modality (EEG). These pre-trained decoders were fine-tuned using data from the test participant. (C) Electrode placement varies widely among the 12 ECoG participants. Asterisks denote five participants whose electrodes were mirrored from the right hemisphere.

3.5 Time-varying autoregression with low rank tensors

A windowed technique was developed to learn parsimonious time-varying autoregressive models from multivariate timeseries [4]. This unsupervised method uncovers interpretable spatiotemporal structure in data via non-smooth and non-convex optimization. In each time window, it was assumed the data followed a linear model parameterized by a system matrix, and this stack of potentially different system matrices was modeled as a low rank tensor. Because of its structure, the model is scalable to high-dimensional data and can easily incorporate priors such as smoothness over time. The components of the tensor were found by using alternating minimization and proved that any stationary point of this algorithm is a local minimum.

4.0 RESULTS AND DISCUSSION

4.1 Investigating naturalistic hand movements by behavior mining in long-term video and neural recordings

The automated, data-driven annotation pipeline was applied to the behavioral data collected for 12 human participants over 7–9 days for each participant [1]. The pipeline discovered and annotated over 40,000 instances of naturalistic human upper-limb movement events in the behavioral videos. Analysis of the simultaneously recorded brain data revealed neural signatures of movement in the high-frequency (76-100 Hz) and low-frequency (8-32 Hz) bands that corroborate prior findings from traditional controlled experiments. A decoder was prototyped for a movement initiation detection task to demonstrate the efficacy of the pipeline as a source of training data for brain-computer interfacing applications. Individual classifier performance varied widely among participants, ranging from around chance levels to 80% on test accuracy. It was found that the most important electrodes for good decoding performance were located in the sensorimotor cortex. Additionally, low-frequencies (<35 Hz) and high-frequency (~100 Hz) bands were the most important frequency features for decoding. This research addressed the unique data analysis challenges in studying naturalistic human behaviors, and contributes methods that may generalize to other neural recording modalities beyond ECoG.

4.2 Behavioral and neural variability of naturalistic arm movements

This project characterized the variability of both naturalistic upper-limb reaching movements and the corresponding changes in cortical spectral power [2]. Based on findings from controlled experiments, it was hypothesized that naturalistic reaches would be associated with transient decreases in low-frequency power and increases in high-frequency power, localized to frontoparietal sensorimotor cortices. The results support this hypothesis on average, see Figure. *3*; however, it is shown that there is considerable variability in spectral power both within and across participants. Multiple-variable linear regression modelling partially explains this single-event neural variability using reach angle and day of recording features, but much of the neural variability remains unexplained by behavioral and environmental features. In general, it is found that results from controlled upper-limb reaching tasks do generalize to naturalistic movements on average, but naturalistic movements involve considerable event-by-event neural variability that cannot be fully explained by simple behavioral and environmental measures.

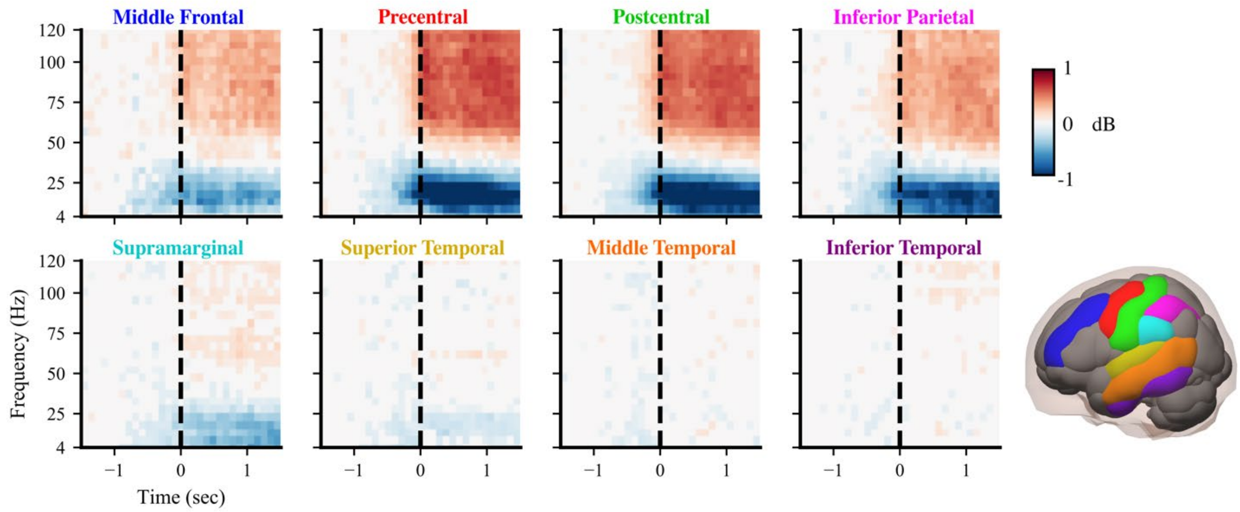

Figure. 3: Group-level cortical spectral power changes are consistently localized to sensorimotor regions. Spectrograms show movement event-triggered spectral power patterns for 8 cortical regions (highlighted in lower right) summarized across all 12 subjects. Spectral power was projected based on anatomical registration from electrodes onto 8 regions of interest: middle frontal (blue), precentral (red), postcentral (green), inferior parietal (magenta), supramarginal (cyan), superior temporal (yellow), middle temporal (orange), and inferior temporal (purple). The baseline power of 1.5–1 seconds was subtracted before movement initiation. Non-significant differences from baseline power were set to 0 ($p > 0.05$).

4.3 Generalized neural decoders for transfer learning across participants and recordings modalities

The results show that this approach to movement decoding, HTNet, is generalizable and tunable, capable of learning common patterns from the training data that transfer to unseen participants and recording modalities [3]. HTNet consistently outperformed state-of-the-art decoders when tested on unseen participants (Figure. 4), even when a different recording modality was used. By fine-tuning these generalized HTNet decoders, the research achieved performance approaching the best tailored decoders with as few as 50 ECoG or 20 EEG events. The project was also able to interpret HTNet's trained weights and demonstrate its ability to extract physiologically-relevant features. By generalizing to new participants and recording modalities, robustly handling variations in electrode placement and allowing participant-specific fine-tuning with minimal data, HTNet is applicable across a broader range of neural decoding applications compared to current state-of-the-art decoders.

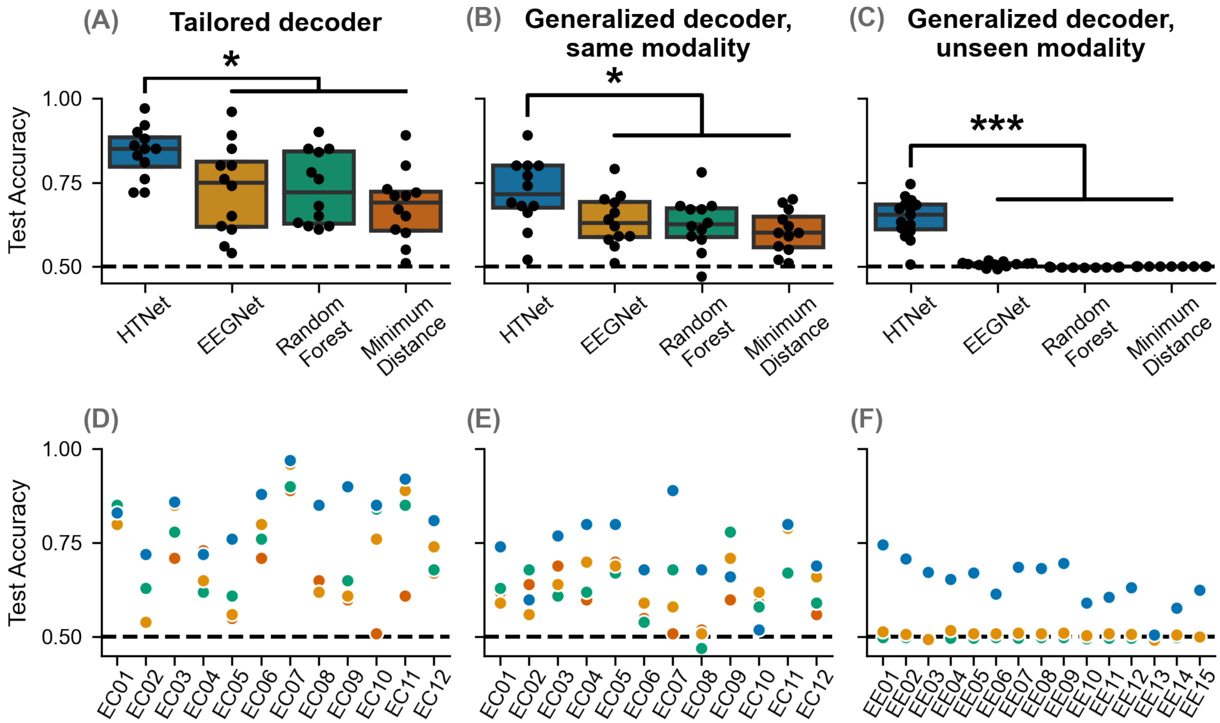

Figure. 4: HTNet generalizes better than EEGNet and other decoders. HTNet achieves significantly higher test accuracy than EEGNet, random forest, and minimum distance decoders across all three scenarios: (A) tailored ($p<0.05$), (B) same modality ($p<0.05$), and (C) unseen modality ($p\leq0.001$). Note that the trained models for same and unseen modality conditions are identical; only the test set differs. (D–F) Bottom row displays decoder performance grouped by test participant for each fold.

4.4 Time-varying autoregression with low rank tensors

This project illustrated the model's utility and superior scalability over extant methods when applied to several synthetic and real-world examples including: two types of time-varying linear systems, worm behavior, sea surface temperature, and monkey brain datasets. With synthetic data generated by a switching or smoothly varying linear dynamical system, it was shown that time-varying autoregressive models with low rank tensors (TVART) can recover the true dynamics and is competitive with other state-of-the-art techniques [4]. In real-world data examples, it was found that the recovered modes are interpretable and can correspond to important dynamical regimes.

5.0 CONCLUSIONS

This project has supported the collection and publication of a naturalistic human ECoG dataset. The research has focused on automatic extraction of interpretable features from long-term video monitoring, evaluation of multimodal features with deep learning, and decoding intent to move. Importantly, both supervised and unsupervised approaches to examining dynamic features of the ECoG data were described. Based on what has been learned, the latest paper [3] published highlights HTNet, a new neural network architecture that outperforms state-of-the-art decoders when generalizing to entirely unseen, new participants. The publications and the datasets published as part of this research project are significant advances on the path to BCIs that are deployable in practical, real-life applications.

6.0 REFERENCES

[1] S. H. Singh, S. M. Peterson, R. P. N. Rao and B. W. Brunton, "Investigating naturalistic hand movements by behavior mining in long-term video and neural recordings," *arXiv,* 2020.

[2] S. M. Peterson, S. H. Singh, N. X. R. Wang, R. P. N. Rao and B. W. Brunton, "Behavioral and neural variability of naturalistic arm movements," *bioRxiv,* 2020.

[3] S. M. Peterson, Z. Steine-Hanson, N. Davis, R. P. N. Rao and B. W. Brunton, "Generalized neural decoders for transfer learning across participants and recording modalities," *Journal of Neural Engineering,* 2021.

[4] K. D. Harris, A. Aravkin, R. Rao and B. W. Brunton, "Time-varying autoregression with low rank tensors," *arXiv,* 2019.

APPENDIX – List of Publications Supported by This Award (chronological order)

Brunton, B. W. & Beyeler, M.
Data-driven models for human neuroscience and neuroengineering. *Current Opin Neurobiol (2019), 58, 21-29.*

Singh, S. H., Peterson, S. M., Rao, R. P. N. & **Brunton, B. W.**
Enabling naturalistic neuroscience through behavior mining: Analysis of long-term human brain and video recordings
2019 Conference on Cognitive Computational Neuroscience (2019).

Singh, S. H., Peterson, S. M., Rao, R. P. N. & **Brunton, B. W.**
Towards naturalistic human neuroscience and neuroengineering: behavior mining in long-term video and neural recordings. *arXiv:2001.08349.*

Peterson, S. M.†, Singh, S. H.†, Wang, N. X. R., Rao, R. P. N. & **Brunton, B. W.** Behavioral and neural variability of naturalistic arm movements.
bioRxiv doi: https://doi.org/10.1101/2020.04.17.047357.

Azadian, E., Velchuru, G., Wang, N. X. R., Peterson, S. M., Staneva, V. & **Brunton, B. W.**
Decoding happiness from neural and video recordings.
NeurIPS 2020, workshop on Learning Meaningful Representations of Life.

Peterson, S. M., Steine-Hanson, Z., Davis, N., Rao, R. P. N. & **Brunton, B. W.**
Generalized neural decoders for transfer learning across participants and recording modalities.
to appear in Journal of Neural Engineering (2021), bioRxiv doi: https://doi.org/10.1101/2020.10.30.362558.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

AAL	automated anatomical labeling
ARHSMM	auto-regressive hidden semi-Markov model
BCI	brain-computer interface
CT	computerized tomography
ECoG	electrocorticography
EEG	electroencephalogram
MNE	minimum norm estimates
MNI	Montreal Neurological Institute
MRI	magnetic resonance imaging
R^2	not acronym, coefficient of determination
SD	standard deviation
TVART	time-varying autoregressive models with low rank tensors