

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

**Satpreet H. Singh<sup>1</sup>, Steven M. Peterson<sup>2,6</sup>, Rajesh P. N. Rao<sup>3,4,5</sup>, and Bingni W. Brunton<sup>2,5,6,†</sup>**

<sup>1</sup> Department of Electrical and Computer Engineering, University of Washington, Seattle, USA

<sup>2</sup> Department of Biology, University of Washington, Seattle, USA

<sup>3</sup> Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, USA

<sup>4</sup> Center for Neurotechnology, Seattle, USA

<sup>5</sup> University of Washington Institute for Neuroengineering, Seattle, USA

<sup>6</sup> eScience Institute, University of Washington, Seattle, USA

†Author to whom correspondence should be addressed.

E-mail: bbrunton@uw.edu

## Abstract.

*Objective:* Recent technological advances in brain recording and artificial intelligence are propelling a new paradigm in neuroscience beyond the traditional controlled experiment. Rather than focusing on cued, repeated trials, *naturalistic neuroscience* studies neural processes underlying spontaneous behaviors performed in unconstrained settings. However, analyzing such unstructured data lacking *a priori* experimental design remains a significant challenge, especially when the data is multi-modal and long-term. Here we describe an automated approach for analyzing simultaneously recorded long-term, naturalistic electrocorticography (ECoG) and naturalistic behavior video data.

*Approach:* We take a behavior-first approach to analyzing the long-term recordings. Using a combination of computer vision, discrete latent-variable modeling, and string pattern-matching on the behavioral video data, we find and annotate spontaneous human upper-limb movement events. We then demonstrate applications of these naturalistic behavior events, along with their associated neural recordings, for neural encoding and decoding.

*Main results:* We show results from our approach applied to data collected for 12 human subjects over 7–9 days for each subject. Our pipeline discovers and annotates over 40,000 instances of naturalistic human upper-limb movement events in the behavioral videos. Analysis of the simultaneously recorded brain data reveals neural signatures of movement that corroborate prior findings from traditional controlled experiments. We also prototype a decoder for a movement initiation detection task to demonstrate the efficacy of our pipeline as a source of training data for brain-computer interfacing applications.

*Significance:* Our work addresses the unique data analysis challenges in studying naturalistic human behaviors, and contributes methods that may generalize to other neural recording modalities beyond ECoG. We publicly release our curated dataset, providing a resource to study naturalistic neural and behavioral variability at a scale not previously available.

**Keywords:** naturalistic behavior, computer vision, neural correlates, neural decoding, electrocorticography, brain-computer interfaces

## 1. Introduction

Neuroscience has long been interested in understanding brain activity associated with spontaneous behaviors in freely behaving subjects. Even so, hypotheses regarding brain function have typically been tested using carefully designed, well-controlled experimental tasks, where timing of cues, stimuli, and behavioral responses are known precisely. Fortunately, recent technological advances have enabled us to study increasingly naturalistic and longer brain recordings, giving rise to a new paradigm called “naturalistic neuroscience” (Nastase *et al.*, 2020; Huk *et al.*, 2018; Gabriel *et al.*, 2019; Markowitz *et al.*, 2018; Wang *et al.*, 2016) where neural computations associated with such spontaneous behaviors are studied. Understanding such unstructured, long-term, and multi-modal data poses a substantial analytic challenge, due in part to the lack of *a priori* experimental design and the difficulty of isolating interpretable behavioral events.

### 1.1. Related work

Our work is related to several areas of active research in neuroscience, neuroengineering, and neuroethology that integrate techniques from machine-learning, computer-vision, and statistical modeling. Many recent methodological innovations have addressed the automated analysis of non-human animal behavior (Batty *et al.*, 2019; Pereira *et al.*, 2019; Nassar *et al.*, 2019; Mathis *et al.*, 2018; Johnson *et al.*, 2016; Wiltschko *et al.*, 2015) (see also Mathis and Mathis (2020) for a recent survey and Anderson and Perona (2014) for a perspective on this emerging area). A typical non-human naturalistic neuroscience experiment (Johnson *et al.*, 2020; Markowitz *et al.*, 2018; Berman, 2018) first collects simultaneously recorded behavioral video and neural activity data from one or more freely behaving subjects in an uncontrolled but sufficiently confined environment. Next, the video recordings are processed through an extensive pipeline consisting of steps such as: segmenting the subject(s) from the background, transforming subject pose to common coordinates using affine transformations, estimating pose of body-parts across frames, and higher-level operations such as classifying pose or segmenting pose into actions. Combined with the simultaneously recorded neural data, such naturalistic behavior data are being used to shed light on previously intractable questions in behavioral neuroscience, often at unprecedented scale.

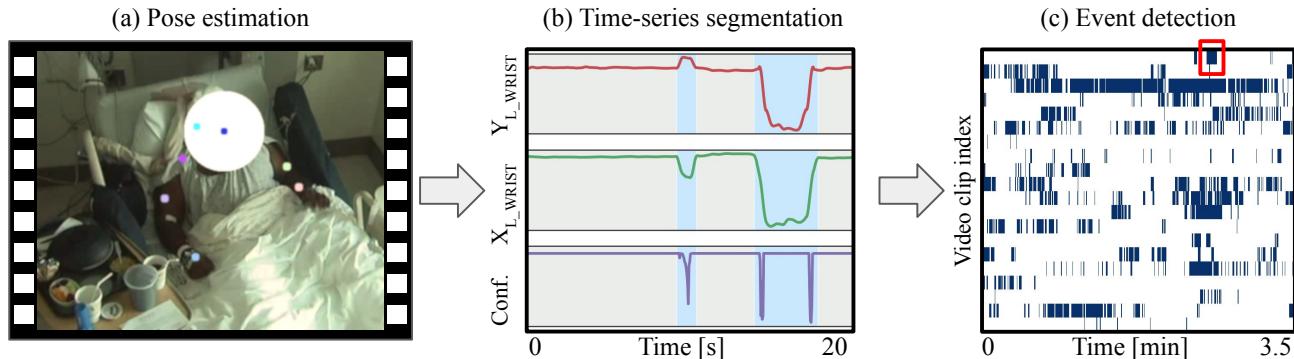
Human action-recognition methods from mainstream computer vision (Ramasamy Ramamurthy and Roy, 2018)

are relevant but not directly applicable to the needs of naturalistic human neuroscience. Traditionally, action-recognition research has concerned itself with discriminating activities at a coarse level, such as sitting vs. walking (Ghorbani *et al.*, 2020), and has often assumed the availability of a large corpus of labeled training data. In contrast, to study the kinds of behaviors that interest neuroscientists and neuroengineers, we seek to localize fine-grained movements to sub-second temporal resolution, and ideally use the fewest behavioral labels possible (Seethapathi *et al.*, 2019). Lastly, since it is not known which behaviors or behavioral characteristics will elicit neural responses worth studying further, a queryable representation that supports the flexibility to study several kinds of behaviors is desirable. Recent work by (Fu *et al.*, 2019) develops such representations for semi-automated exploration of scenes in general videos.

Our work is most closely related to recent work in human naturalistic neuroscience that combine computer vision with opportunistic clinical brain recordings, including Wang *et al.* (2016), Alasfour *et al.* (2019), and Chambers *et al.* (2019). In particular, we build on the work of Wang *et al.* (2018) using similar video data and estimate the pose of human upper-body keypoints using neural-networks. Gabriel *et al.* (2019) use optical flow and image partitioning to detect coarse limb movements from video taken in a clinical setting similar to ours and develop neural decoders for detecting these movements from brain data. Compared to Wang *et al.* (2018), who use a moving window heuristic on pose estimates to detect movements, we take a more principled approach to modeling the pose data. This allows us to localize movement events with finer temporal-resolution and characterize entire movement trajectories, which in turn enables novel applications described later in the paper. We also use newer, more efficient computer vision methods (Mathis and Mathis, 2020; Nath *et al.*, 2019) that allow us to process data at a scale that exceeds all of the aforementioned studies taken together in the number of subjects and duration of recordings analyzed. Finally, we focus on curating, characterizing, and making our dataset available to the research community to foster further research and development in this area.

### 1.2. Our approach

We present a scalable behavior-mining approach to analyze simultaneously recorded naturalistic brain and behavior data, obtained opportunistically from human subjects undergoing long-term clinical monitoring prior to epilepsy



**Figure 1.** Pipeline for behavioral video data processing. (a) Video frame showing estimated pose keypoints (colored dots) on human subject. (b) Autoregressive hidden Markov model (Section 3.1.2) robustly segmented pose trajectory into *rest* (shaded grey) and *move* (shaded light blue) states. (c) Raster plot of pose states (*move* in dark blue, *rest* in white) for several video-clips for pattern matching at scale. Red box depicts one *movement initiation* event matching a pattern of 15 contiguous *rest* states (0.5s) followed by 15 contiguous *move* states (0.5s).

surgery. Our video processing pipeline (Figure 1) first estimates the locations of keypoints (e.g. wrists and elbows) on the upper-body using a neural network trained on each subject (Mathis *et al.*, 2018). We then segment the trajectory of each keypoint in time using discrete latent-variable models, building a discrete representation of pose dynamics. Interestingly, having a discrete, sequential representation of upper-limb pose simplifies the problem of detecting behavioral *events* to pattern-matching on strings. Using regular-expressions corresponding to patterns of interest, we discover thousands of interpretable events per subject—an order of magnitude more observations than in a typical controlled human experiment. To study the rich naturalistic variability associated with these events, we also extract metadata including movement angle, magnitude, and duration.

Next, we explore the use of these behavioral events for neuroscience and neuroengineering applications by analyzing the simultaneously recorded brain data. Event-averaged spectrograms associated with our naturalistic human upper-limb movement initiation events corroborate and strengthen previous findings from controlled experiments (Miller *et al.*, 2007) (see also (Peterson *et al.*, 2020)). Preliminary investigations also suggest that our workflow could produce data useful for training brain-computer interface (BCI) decoders; due to the use of larger sample sizes of training data representative of naturalistic variability, such decoders may perform more robustly in real-world deployments.

Our key contributions in this paper are as follows. First, we present a highly automated, novel workflow for analyzing simultaneously recorded naturalistic long-term human brain and behavioral video data. Second, we develop a domain-relevant, robust, temporally precise, and queryable representation of human upper-limb pose. Third, to showcase our workflow, we demonstrate example applications in neuroscience and neuroengineering, suggesting that our approach and results are of broad interest. Finally,

to support open science and facilitate further research in this area, we release our curated dataset consisting of annotated naturalistic events and associated neural recordings.

## 2. Dataset

### 2.1. Human subjects and data collection

Our dataset consists of human intracranial electrocorticography (ECoG) (Parvizi and Kastner, 2018) neural recordings and simultaneously recorded behavioral video recordings, obtained opportunistically from 12 patients with epilepsy for the duration of each patient’s long-term (7–9 days) continuous clinical observation. The University of Washington Institutional Review Board for the protection of human subjects approved our study and all patients provided their informed written consent. Patient behavior was continuously recorded by a wall-mounted camera (RGB/infrared,  $640 \times 480$  pixels) for real-time monitoring by an around-the-clock clinical team, except during intermittent equipment servicing or private times when the camera was switched off or turned away. Patients were observed performing their daily activities (including talking, eating, watching TV, using a computer or phone, sleeping, receiving clinical care etc.) from the hospital bed while being tethered to a brain-recording interface.

Each patient had about 90 electrodes implanted under the skull and dura, directly on their brain surface, including either an  $8 \times 8$  grid or two  $8 \times 4$  grids of electrodes. These electrodes were placed to monitor a subset of cortical regions, predominantly in one brain hemisphere, as determined by individual clinical need (right hemisphere for 5 and left hemisphere for 7 patients). Grid electrodes are spaced 1 cm apart center-to-center. The sampling rates for video and ECoG recordings were 30 fps and 1000 Hz respectively. Together, the ECoG and video (approximately 18 million frames for a week) totaled about 250 GB of data

per patient.

## 2.2. ECoG and video data preprocessing

We applied standard pre-processing steps to the ECoG data (Peterson *et al.*, 2020; Gabriel *et al.*, 2019; Miller, 2019; Miller *et al.*, 2007; Schalk *et al.*, 2007), including down-sampling to 500Hz, 60-Hz line noise removal, large-amplitude artifact removal, and median-centering using a common reference across all electrodes. All electrode positions were localized and converted to Montreal Neurological Institute and Hospital (MNI) coordinates using the Fieldtrip toolbox (Oostenveld *et al.*, 2011; Stolk *et al.*, 2018) in MATLAB. To aid interpretability, we restricted our analysis in this paper to the 64 grid electrodes covering one hemisphere per subject. We excluded electrodes with recording issues, such as persistent presence of artifacts. However, data for all available electrodes are provided in the publicly released dataset accompanying this paper. In most cases, we also did not analyze the neurally and behaviorally atypical data from the first two days of a patient’s hospital stay, since patients were usually heavily medicated during this time while recovering from electrode implantation surgery.

Before processing the video data, we manually inspected and annotated it at a coarse (every 3 minutes or so) level of granularity to create an *omit-list* that was excluded from further processing. The omit-list included long time-spans of sleep, times when a clinical or research team was actively working with the subject, private times, and times when applying computer-vision algorithms was impossible due to poor lighting conditions or severe occlusion of the subject’s body. Almost all camera movement occurs around times when the clinical team is actively working with the patient. Removing these times results in a mostly steady recording configuration as seen in Figure 1(a). We also labeled and excluded times when the clinical team had placed seizure restraints on the subjects’ hands, since these limited mobility and gave rise to unnatural movements. Completing these manual annotations took about 6–12 hours per subject, depending on their clinical treatment regime, activity and sleep schedule, and length of hospital stay. When analysing ECoG and video data together, the two data-streams were synchronized using metadata extracted using the equipment manufacturer’s (Natus Medical Incorporated) software.

## 3. Methods

### 3.1. Precision extraction of behavioral events at scale

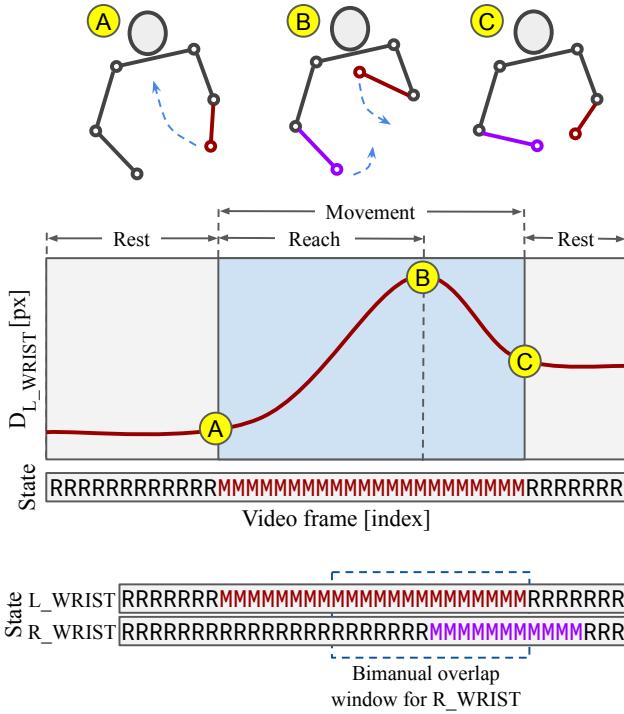
We developed and validated a pipeline to extract temporally precise, interpretable movement events, by processing the video data through pose-estimation, pose time-series segmentation, event detection and finally, event metadata extraction (Figure 1).

**3.1.1. Markerless pose estimation** To extract a subject’s pose from raw video, we trained a state-of-the-art markerless pose estimation tool (Mathis *et al.*, 2018) known for its speed and data efficiency (Nath *et al.*, 2019). For training data, we manually annotated around 1000 frames per subject chosen randomly from the entire duration of a subject’s video data, preferentially sampling active, daytime hours over times when the subject was asleep. For each frame, we annotated up to 9 keypoints on the subject’s body whenever visible (Bourdev and Malik, 2009). These keypoints were the nose, both wrists, elbows, shoulders, and ears (Figure 1a).

During prediction, the pose-estimation tool produced the  $(x, y)$  coordinates and a *confidence* estimate between  $[0, 1]$  for each keypoint per frame (Figure 1b). To quantify the performance of keypoint tracking, we estimated the pixel-wise RMS error to be  $1.54 \pm 0.13$  s.d. pixels on the training data and  $5.97 \pm 1.96$  s.d. pixels on the holdout data, both averaged across 12 subjects; holdout data is 5% of the manual annotations, excluding points below confidence threshold of 0.1. As an approximate scaling, 12 pixels in the video span about 4 cm in physical units, which is about the width of a human wrist. We estimated this scale by comparing standard human measurements (McDowell *et al.*, 2009) with median distance between shoulder keypoints (in pixels) at movement onset for a few subjects. On average, estimating pose for the entire duration of a subject’s video took 400 GPU-hours per subject using AWS p2.16xlarge NVIDIA K80 GPUs. We denoised the estimated pose trajectories by median filtering (window length 11 frames) and smoothing (window length 11, 2nd order Savitzky-Golay (Schafer, 2011) filter).

**3.1.2. Segmentation of pose trajectories** Next, we segmented the pose time-series into discrete, interpretable states while preserving the temporal precision of the keypoint tracking. We applied a first-order autoregressive hidden semi-Markov model (ARHSMM) (Murphy, 2012) with two latent states to each keypoint’s time-series (Figure 1b for left-wrist). This model converts each keypoint’s continuous pose dynamics into discretized dynamics consisting of *rest* and *move states*. Using a semi-Markov, rather than a Markov model, accounts for the bias that limbs tend to be at *rest* most of the time and mitigates unnecessary switching between latent states. Similar to (Wiltschko *et al.*, 2015), we fit the ARHSMM using the pyhsmm-autoregressive package in Python (Johnson and Willsky, 2013). The resulting states are at video frame-rate resolution and the segmentation is relatively robust to variation in lighting, camera angle, and level of activity in the video.

**3.1.3. Behavioral event mining** Discretizing the pose trajectories facilitates the description of scientifically interesting behaviors performed spontaneously by the



**Figure 2.** Schematic of metadata extraction: [Top] Cartoon showing left-wrist movement at time of (A) movement initiation (B) maximum displacement from initiation (*reach*) and (C) movement end. [Middle] Time course of left-wrist radial distance [px: pixels], and discretized state (R: *rest*, M: *move*). Extracted movement metadata include duration, start and end coordinates, among others (Sec. 3.1.4). [Bottom] Discretized state sequence for both left and right-wrists, now showing movement initiation in the right-wrist while the left-wrist is still in motion. The dashed line shows the window over which *bimanual overlap* metadata is calculated, corresponding to the number of frames for which the opposing (left) wrist is in motion over that duration.

subject, even though they vary greatly in duration. Specifically, the task of finding different types of behavioral **events** thus reduces to string pattern matching on the discretized dynamics. For the behaviors we explore in the rest of this paper, we looked for *movement initiation* events by matching a pattern of 15 consecutive *rest* states (0.5s), followed by at least 15 consecutive *move* states (0.5s). Similarly, *no-movement* events are state sequences of 90 *rest* states (3.0s) across both wrists and the nose. To create our database of wrist movements, we use regular expressions to quickly find thousands of non-overlapping instances of such patterns in the discretized pose dynamics for each subject.

Parameters for smoothing, hyperparameters for the ARHSMM segmentation model, and the choice of regular expressions for event detection, were picked empirically by assessing performance on pose time-series derived from a small representative set of subject videos. We confirmed that the temporal accuracy of event boundaries matched our expectations by manually inspecting a few dozen random events of each movement type for each subject.

**3.1.4. Event metadata extraction** For each detected movement event, we extracted several metadata features from the continuous pose-dynamics associated with the movement. These include movement-associated metadata (Figure 2) 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.

Observed naturalistic hand movements often consisted of a hand reaching out, touching, or grabbing an object, then bringing the hand back to the body. Therefore, we defined the *reach* of a wrist movement to be its maximum radial displacement during the course of the event, as calculated from its location at the start of the event. We extracted the *magnitude*, *angle*, and *duration* for each reach.

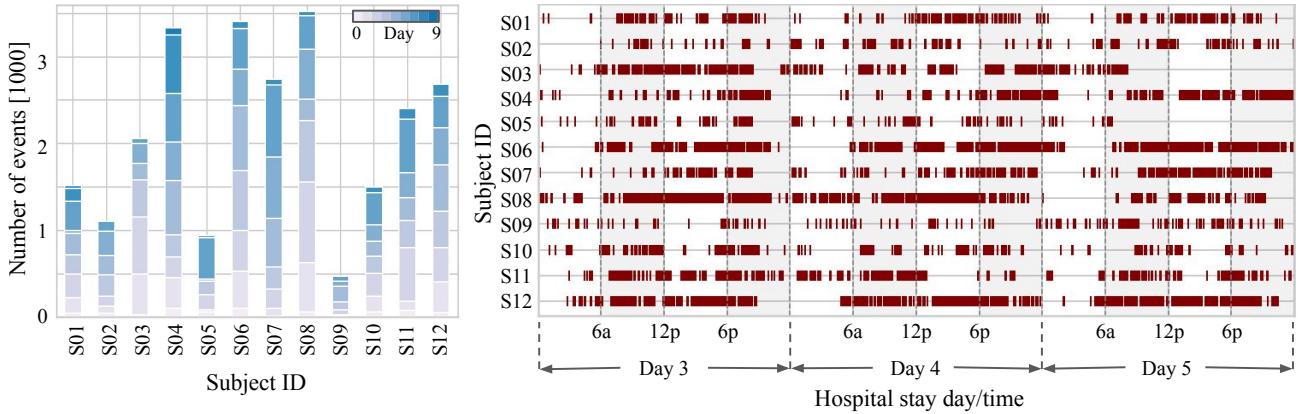
To measure the *shape* of a movement, we fit 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>-degree polynomials to a keypoint's displacement trajectory. Differences between the quality of the fit (as measured by  $R^2$ ) to each polynomial type provide a rough measure of the “curviness” of the movement trajectory. We also estimated a movement's onset and offset *speeds*, by calculating the keypoint's displacement change within short time windows around the start and end of the movement.

Since people often move both hands at the same time (i.e. “bimanually”), we augmented each movement event with metadata about the opposing wrist's movement, if any (Figure 2). By juxtaposing the discrete state sequence of both wrists, we calculated when the opposing hand starts to move (*lead/lag time difference*) and how long this movement overlaps with that of the primary hand (*overlap duration*).

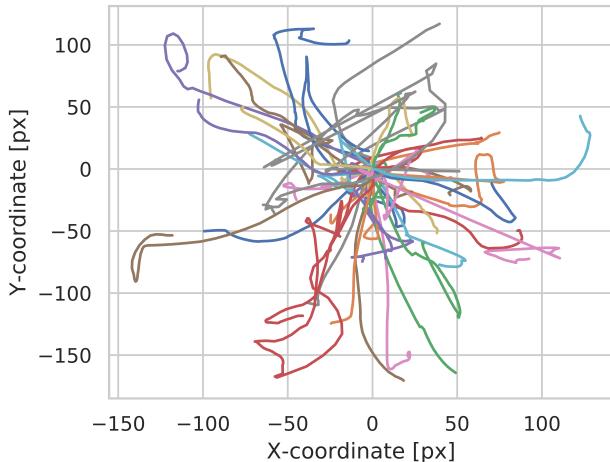
False positives in event discovery were still present in the data at this stage due to pose estimation failures and unusual pose states. To compensate for failures in 2D pose estimation, we calculated movement-weighted *confidence scores* for each event and removed those below a manually determined threshold. To eliminate outlier pose states, we calculated mean *distance* and mean *angle* between shoulder keypoints, then removed events from the top and bottom 5 percentiles of these quantities.

### 3.2. Neural correlates of movement

A core scientific question in systems neuroscience is how behaviors are encoded by the coordinated activation of brain regions. To examine the neural correlates of naturalistic movement initiation, we performed a time-frequency (TF) analysis of the neural recordings (Cohen, 2014) by averaging event-locked spectrograms for each subject, using hundreds of movement initiation events chosen to match movement statistics (reach magnitude, onset velocity, and shape) of a previous controlled experimental study (Miller *et al.*, 2007). Using the aforementioned metadata to guide our search, we selected up to 200 events per day over 5 days for each of 12 subjects, and then further inspected the video for each event to remove any false



**Figure 3.** [Left] Number of right-wrist movement initiation events discovered per day for each of 12 subjects, totaling 475 to 3526 events per subject across their entire duration of clinical observation ( $268 \pm 123$  s.d. per day). [Right] Raster plot of right-wrist movement initiation event occurrences showing bursts of activity interspersed with periods of rest or omit-listed (Section 3.1.1) periods. See Figure A2 for equivalent plots for left-wrist.



**Figure 4.** A sample of 50 typical right-wrist trajectories (px; pixels; translated to start at origin) showing diversity of naturalistic reach movements for a single subject (S10). Different colors represent different individual trajectories. Note the large variability in the movements, compared to what is normally captured by controlled experiments.

positives (17.8% mean  $\pm$  9.9% s.d. events).

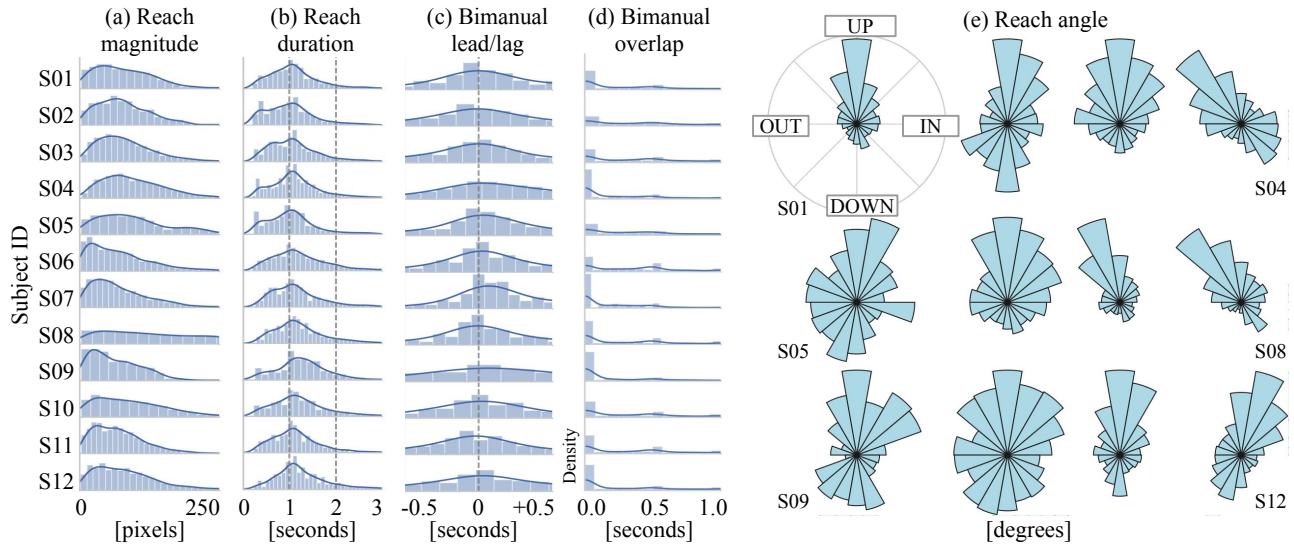
### 3.3. Decoding naturalistic movement initiation

A grand challenge in neuroengineering is development of BCIs that can be used to predict spontaneous activity and intentions outside the lab, in everyday settings (Shanechi, 2019; Smalley, 2019; Shanechi, 2018; Warren *et al.*, 2016; Shenoy and Chestek, 2012). Here we performed a preliminary study leveraging our pipeline as a source of training data for a BCI decoder that detects wrist movement initiation events. Specifically, we trained separate classifiers, tailored to each subject, to discriminate between movement initiation events and no-movement

epochs for each wrist using only features derived from the ECoG neural recordings.

Our decoder uses the Random Forest (RF) algorithm (Breiman, 2001; Murphy, 2012), which is typically considered one of the best off-the-shelf classification algorithms for small/medium sized datasets (Hastie *et al.*, 2009). We used ECoG data 0.5s before to 0.5s after each event to compute TF spectrograms at each of the 64 grid electrodes and used the flattened vector of TF bins as features for the classifier (TF bins were 200ms  $\times$  5Hz resolution, truncated at 150 Hz; approximately 9000 features total).

Given that the brain's response can drift over the course of days (Farshchian *et al.*, 2019; Klosterman *et al.*, 2016), a reduced subset of events from 3 consecutive days (typically days 3 through 5 of clinical monitoring) were used. We used events from the last day as the test set. To eliminate the confound of movement initiation in the opposing wrist, we further filtered events to exclude those with significant movement ( $\geq 0.2$  seconds) in the opposing wrist within the  $\pm 0.5$ s window used for ECoG data. Positive (movement initiation) and negative (no-movement) examples were balanced by down-sampling negative examples. This balancing eliminated bias in the training set and set up a baseline performance of 50% accuracy for test set performance. Training and test supports were  $633 \pm 417$  s.d. and  $331 \pm 203$  s.d. examples, respectively. We tuned the RF using a 20-trial randomized search over two hyperparameters: number of trees (range: [50, 250]) and maximum tree-depth (range: [3, 15]). For each set of hyperparameters, 5-fold cross-validation holdout accuracy was used to measure performance. Final performance reported is from training using best hyperparameters and corresponds to classifier accuracy on events from the withheld test day.



**Figure 5.** Histograms of right-wrist movement initiation event metadata per subject for their entire duration of clinical observation: (a) Reach magnitude shows a dominance of small movements, (b) Reach durations tended to be concentrated around  $\approx 1$ s, (c) When both hands moved together (“bimanually”), they tended to start at about the same time. (d) Duration of time left hand was moving during a  $\pm 0.5$ s window around time of right hand movement initiation. (e) Polar histograms show that many subjects primarily made upward-downward reaches. However, reaches in almost every other direction were also observed. See Figure A3 for equivalent plots for left-wrist.

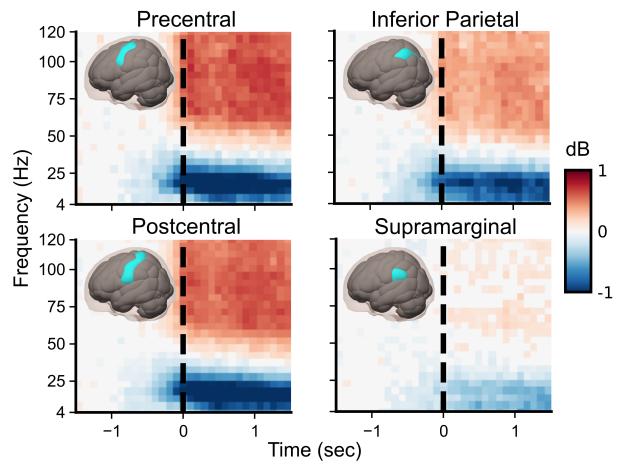
## 4. Results

### 4.1. Characterizing naturalistic events

Our pipeline extracted 959 to 6745 individual wrist movement events per subject ( $487 \pm 215$  per day) across 12 subjects (Figures 3 and A2). We found large variability between subjects in the number of events discovered, which we attribute to inter-subject differences in cycles of sleep and wakeful activity and clinical treatment regimes (Figure 3 and A2). We also observed rich within-subject variability in the event metadata (Figures 4, 5, and A3), which further differentiates our dataset from those collected in controlled experiments. Since our subjects received no instructions for when and how to move, we expect the observed movement statistics to closely reflect the natural statistics of human upper-limb movements while seated.

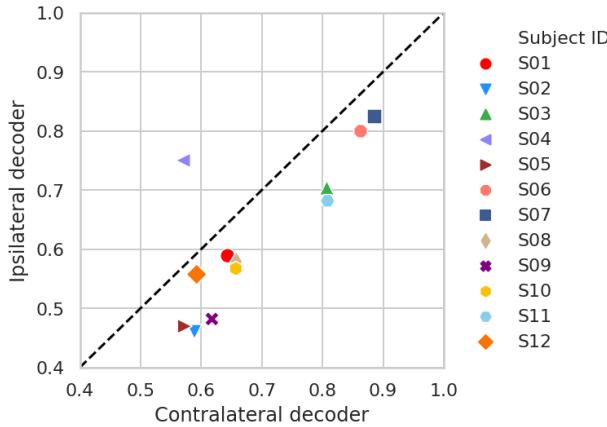
### 4.2. Neural correlates of movement initiation

We observed movement-associated power increases in a high-frequency band (76–100 Hz) and decreases in a low-frequency band (8–32 Hz) across many cortical areas (Figure 6). This pattern is similar to what has been observed with controlled experimental trials in Miller *et al.* (2007), Volkova *et al.* (2019), and Yuan and He (2014). Furthermore, we strengthen prior findings by showing that these movement-associated patterns hold across 5 consecutive days, well beyond the timescale of a typical controlled experiment (less than hours). To the best of our knowledge, this is the first reported instance of a TF analysis of spontaneous naturalistic



**Figure 6.** Neural correlates of movement initiation: Event-locked spectrograms, averaged by brain region (cyan color in insets) across 12 subjects, showed movement-associated high-frequency power increase and low-frequency power decrease. These patterns corroborate and strengthen previous findings from controlled experiments (Miller *et al.*, 2007). See our companion preprint Peterson *et al.* (2020) for a deeper exploration of the behavioral and neural variability of these movements.

movements using events discovered by an automated workflow. We elaborate on these results in our concurrently released preprint (Peterson *et al.*, 2020), where we further investigated the consequences of the relatively higher variance of naturalistic movement statistics and modeled the contributions of the various movement metadata to the observed neural responses.



**Figure 7.** Test set decoding accuracy for initiation of movement of contralateral (side opposite electrode implant) and ipsilateral (same side) wrists: As expected, decoding of contralateral movements is slightly more accurate than ipsilateral in almost all cases.

#### 4.3. Decoding naturalistic movement initiation

Individual subject classifier performance varied widely between subjects, ranging from around chance levels to 80% on test accuracy (Figure 7), comparable to previously reported work (Gabriel *et al.*, 2019; Wang *et al.*, 2018). Classifier performance tended to be correlated with extent of motor cortex coverage (Figure 8 and A1). Due to hemispheric lateralization of brain function, decoding contralateral limb movements is expected to be more accurate than decoding ipsilateral movements (Tam *et al.*, 2019). Consequently, test set decoding accuracy (Figure 7) was higher for the contralateral wrist in almost all subjects. Since false positives (FPs) in the event data establish a ceiling on classifier accuracy, we estimated their prevalence by manually inspecting 100 randomly sampled events per event-type from each subject ( $5\% \pm 10\%$  s.d. for no-movement,  $22\% \pm 16\%$  s.d. for contralateral, and  $14\% \pm 10\%$  s.d. for ipsilateral events). With more stringent rejection of FPs, we expect improved decoding performance and potentially more pronounced differences between contralateral and ipsilateral decoding accuracy.

To interpret the importance of spectral features in the decoder, we visualized Random Forest feature importance scores (Breiman, 2001; Hastie *et al.*, 2009). We aggregated these scores in two ways to gain insight into their spatial (Figure 8 and A1) and frequency (Figure 9) components. Spectral features are indexed by electrode, time, and frequency. We define Feature Importance aggregated by Electrode  $FI_E(e)$  for electrode  $e \in \mathcal{E}$  as:

$$FI_E(e) = \sum_{f \in \mathcal{F}, t \in \mathcal{T}} FI(t, f, e),$$

where  $\mathcal{E}$ ,  $\mathcal{F}$  and  $\mathcal{T}$  are the sets of electrodes, frequency-bins, and time-bins over which the spectral features are

calculated, respectively. For the purpose of visualization, we normalized these values to get Normalized Feature Importance aggregated by Electrode  $NFI_E(e)$ :

$$NFI_E(e) = \frac{FI_E(e)}{\max_{e \in \mathcal{E}} FI_E(e)}.$$

As seen in Figures 8 and A1, we found that electrodes over sensorimotor cortex, when available, dominated feature importance (e.g. Subjects S07, S06, S03 and S11 in Figure 7). When motor cortex coverage is limited or unavailable, decoding is still possible because inter-region correlations are known to exist in the brain (Tam *et al.*, 2019; Miller *et al.*, 2007; Schalk *et al.*, 2007) and were likely exploited by the classifier, but with limited decoding capacity.

To understand the contributions of various frequency-bins to decoding, we define analogous formulas for feature importances aggregated by frequency-bin:

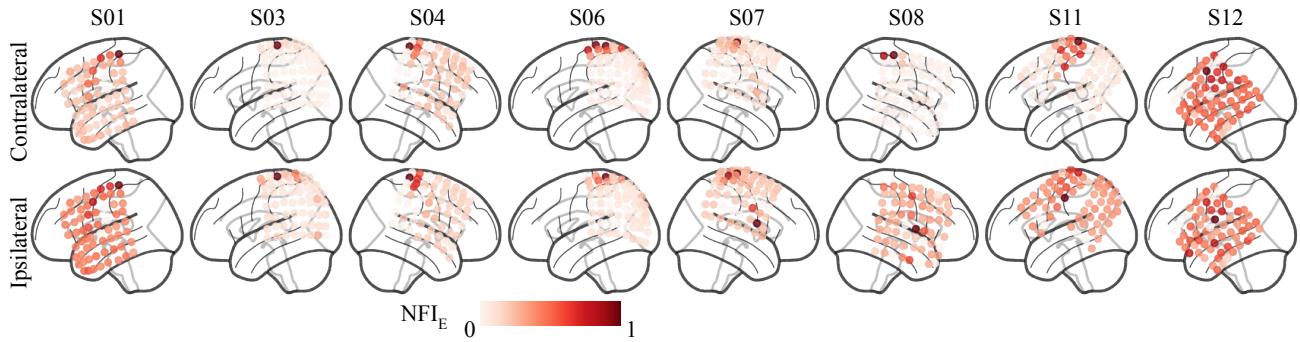
$$FI_F(f) = \sum_{e \in \mathcal{E}, t \in \mathcal{T}} FI(t, f, e)$$

$$NFI_F(f) = \frac{FI_F(f)}{\max_{f \in \mathcal{F}} FI_F(f)}$$

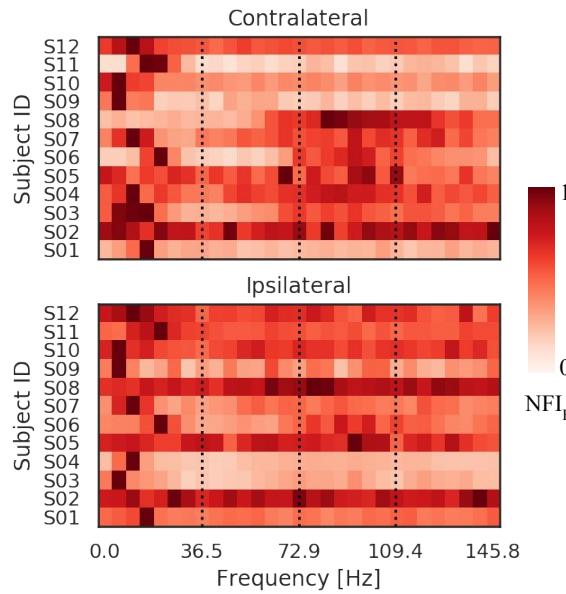
As seen in Figure 9, we found that a low-frequency band ( $< 35\text{Hz}$ ) and a high-frequency band (around 100 Hz) dominates feature importance, similar to prior findings (Miller *et al.*, 2007). If ipsilateral wrist movement was being decoded, or when motor cortex electrode coverage was lacking (e.g. contralateral S02 and S05 in Figures 8 and 1), spectral feature contributions tended to be more broadly distributed across the frequency spectrum and correlated with lower decoding accuracy.

## 5. Discussion

In summary, we have developed a highly automated and scalable approach for analyzing long-term datasets of simultaneously collected human brain and naturalistic behavior data. Our workflow robustly uncovered and annotated thousands of human upper-limb movement events in behavior videos. To detect movement events, we first discretized pose time-series for each wrist into two latent states, indicating movement or rest, and then used regular expressions to look for user-specified patterns in the latent state sequences. This semi-supervised strategy allowed us to rapidly explore movements and their associated brain responses. Importantly, our curated naturalistic dataset supported direct comparison with existing literature from controlled experiments. To demonstrate the applicability of our workflow, we analyzed the brain data associated with the annotated events from two perspectives: characterizing neural correlates of movement, and decoding naturalistic movement initiation using ECoG data. Key to the success of our applications



**Figure 8.** Contralateral and ipsilateral wrist movement initiation decoder feature importance scores aggregated by electrode ( $NFI_E$ ), showing spatial contributions of different brain regions. Scores are normalized by dividing by highest electrode score for each decoder. Electrode coverage over motor cortex is highly correlated with decoder accuracy; for instance, subjects having good motor cortex coverage (S07, S06, S03 and S11) have the highest decoding performance (Figure 7). See Figure A.1 for plot with all 12 subjects.



**Figure 9.** Decoder feature importance scores aggregated by frequency and normalized by score of highest frequency bin per subject ( $NFI_F$ ). Heatmaps show that the most relevant spectral features tend to come from a low-frequency band (< 35 Hz) and a high-frequency band (around 100 Hz), similar to prior findings (Miller *et al.*, 2007). When motor cortex electrode coverage is lacking (e.g. contralateral S02 and S05) or if ipsilateral wrist movement is being decoded, spectral feature contributions tend to be more broadly distributed across the frequency spectrum, and correlated with lower decoding accuracy.

is the availability of a large number of repeated instances of movement initiation events, all available with high temporal precision, which is an essential requirement for generating event-averaged spectrograms (Cohen, 2014). The ability to select movements by magnitude, onset velocity, and complexity (using shape metadata) allowed us to match movement statistics between naturalistic and controlled experimental data, enabling a fair comparison. Furthermore, the ability to select events without opposing wrist activity allowed us to disambiguate confounds when

comparing movement decoders for opposing wrists.

### 5.1. Limitations and future work

Our work has a number of limitations that can be improved with further development. First, our strategy of discretizing individual keypoint time-series to two latent states and then pattern-matching on latent state sequences may be challenging with more complex behaviors involving coordinated movement of more keypoints. When we increased the number of latent states in the pose segmentation process, we also noticed that behavioral states were harder to interpret and associated ECoG responses were not easily separable. The automated analysis of behavior for simple model organisms such as worms (Gupta and Gomez-Marin, 2019), zebrafish (Johnson *et al.*, 2020), flies (Berman *et al.*, 2016) and mice (Luxem *et al.*, 2020; Markowitz *et al.*, 2018; Datta, 2019), has advanced to the extent of being able to automatically extract hierarchies of coordinated behavioral sequences (or *grammars*) from naturalistic videos. Except for some very limited work (Summers-Stay *et al.*, 2012; Yang *et al.*, 2014), such progress has been elusive in human computer vision, possibly due to the sheer complexity and variability of human movements in various contexts. Though not tailored to our temporal precision requirements, future research in fine-grained human action recognition in sports (Shao *et al.*, 2020; Piergiovanni and Ryoo, 2018), domestic (Rohrbach *et al.*, 2012) and industrial (Kobayashi *et al.*, 2019) contexts could eventually provide methods that enable the collection of massive datasets of finely annotated human behavior.

All of our data was acquired opportunistically and videos were recorded from a single clinical monitoring camera. Thus, a primary drawback of the event metadata generated by our pipeline is that they are derived from pose-estimation on single-camera RGB images, implying that all pose coordinates are 2D projections and that the fidelity of pose-derived metadata is limited. However,

one can still extract utility from the event metadata by coarse-graining or binning them. The kinematic dataset could be made significantly richer by the use of additional hardware, such as an RGB-D (RGB with depth) camera or a stereoscopic camera system, that would enable pose-estimation and object-tracking in 3D (Karashchuk *et al.*, 2020; Hansen *et al.*, 2019; Sarafianos *et al.*, 2016).

We controlled false positives in the event discovery process using a combination of pose-estimation confidence and a tedious manual omit-listing process. We found the confidence estimate provided by our pose-estimation tool to perform well under conditions of good visibility, but it was sensitive to variations arising from naturalistic lighting and occlusions. One potential source of improvement could come from using pose-estimation algorithms that employ body models, such as OpenPose (Cao *et al.*, 2017). In our assessment, DeepLabCut (Mathis and Mathis, 2020; Nath *et al.*, 2019) offered a better speed (cost) vs. accuracy tradeoff at the scale we deployed for pose-estimation. Future work is poised to take advantage of rapid innovations in computer vision, as more tools become available and accessible. While manual creation and review of an omit-list cannot be completely avoided for compliance with human research protocols, we believe that a stereoscopic or depth based camera system could also help detect occlusions better and lead to a reduction in false positives.

Finally, two limitations arise from the opportunistic data-collection paradigm itself. First, we have limited our study to a subject’s wrists because they are relatively unconstrained and can perform spontaneous naturalistic movements compared to the rest of the subject’s body. Our subjects’ heads are tethered to a brain recording device that partially restricts the movement of the rest of their upper body. The study of more naturalistic, especially more active, behaviors would require wireless recording. Second, ECoG data such as ours has been obtained opportunistically from a neuro-atypical patient population undergoing long-term monitoring preceding invasive epilepsy resection surgery. We note with caution that conclusions from analyzing such data might not generalize well to the broader, neuro-typical population.

## 5.2. Opportunities using curated dataset

Accompanying our manuscript, we have publicly released our curated dataset comprising neural data and event metadata for over 40,000 instances of naturalistic human upper-limb movement events, and an equal number of rest events, across 12 subjects over about a week of clinical monitoring each. We expect our dataset to be broadly applicable to BCI research like previously released datasets such as the BCI competitions I–IV datasets (Sajda *et al.*, 2003; Blankertz *et al.*, 2004, 2006; Tangermann *et al.*, 2012), or other ECoG data libraries (Miller, 2019) that were generated through controlled experimentation. While the aforementioned datasets consist of 10s–100s of repeated

instances of a behavior per subject, our dataset provides 1000s of instances per subject. It captures rich naturalistic variability across multiple axes relating to the neural activity (subject, seizure foci, day of observation, electrode placement, and recording fidelity) and the behavior (subject activity profile, medication and treatment regime; wrist movement times, movement handedness and sequencing; and other event metadata).

We are working on refinements and extensions of the preliminary investigations presented here, and believe that our dataset could serve several other lines of scientific inquiry. A thorough analysis of neural encoding of wrist movement initiation is available in our simultaneously released preprint (Peterson *et al.*, 2020). As a follow-up to the limited prototype described here, we are currently exploring the use of neural networks to build more expressive decoders (Roy *et al.*, 2019) for events and their associated event metadata. The abundant availability of neural data also allows us to explore representation learning to obtain interpretable task-specific neural features (Pailla *et al.*, 2019; Shirashi *et al.*, 2020), and transfer learning to adapt decoders trained for one subject to another (Wu *et al.*, 2020; Elango *et al.*, 2017; Shenoy *et al.*, 2007).

Our dataset is amenable to several types of modeling goals and approaches including unsupervised latent factor modeling to extract single-trial neural dynamics (Pandarinath *et al.*, 2018a; Ly *et al.*, 2018; Cole and Voytek, 2019; Pandarinath *et al.*, 2018b; Zhao and Park, 2016), dynamical modeling of the electrocorticographic spectrum (Chaudhuri *et al.*, 2018; Beck *et al.*, 2018; Haller *et al.*, 2018; Brunton *et al.*, 2016), probabilistic modeling to better understand neural data variability across trials, subjects and brain-regions (Omigodun *et al.*, 2016; Yang *et al.*, 2019; Abbaspourazad *et al.*, 2018; Yang *et al.*, 2017), generative modeling to generate synthetic brain data (Hartmann *et al.*, 2018; Aznan *et al.*, 2019) and modeling the non-stationarity of the brain signal over long recording time spans (Farshchian *et al.*, 2019; Klosterman *et al.*, 2016; Shenoy *et al.*, 2006). We hope that our dataset will enable further research on models of neural function that incorporate naturalistic variability.

The presence of false positives in the data also motivates exploring algorithms for machine learning with noisy labels (Rolnick *et al.*, 2017; Natarajan *et al.*, 2013; Han *et al.*, 2018). This paradigm has been well studied for other applications of machine learning such as computer vision (Li *et al.*, 2017), where amassing large datasets with noisy labels is relatively inexpensive, but quality labeling is expensive to obtain. The large behavioral variability associated with our neural data could also be used to investigate optimal training set selection, i.e. what types of and how much training data could be ideal for training a decoder (Wei *et al.*, 2015, 2014; Krause *et al.*, 2008). Such characterizations could be used to inform the engineering of BCIs, making them significantly more robust to the variations present in real-world deployments.

## Code and dataset release

Code to reproduce several key plots in this manuscript is publicly available at: <https://github.com/BruntonUWBio/singh2020>. Our complete curated dataset consisting of events and their metadata and associated neural data can also be downloaded following the instructions provided at the aforementioned URL.

## Acknowledgements

We thank John So for extensive help with manual annotation of the video data. This work benefited from and was enabled by the groundwork laid by Nancy X. R. Wang towards study approval, initial clinical data procurement, preprocessing, and manual annotation, and establishing the plausibility of movement initiation prediction using a subset of this clinical data. We thank the neurosurgeons Dr. Jeffrey G. Ojemann and Dr. Andrew Ko, and the staff and consenting patients at the University of Washington Harborview Medical Center in Seattle, for supporting this research. We thank Pierre Karashchuk, Kameron D. Harris, James Wu, Nile Wilson, Ariel Rokem, Renshu Gu, David J. Caldwell, and Preston Jiang for helpful discussions and suggestions. This work was funded by NSF award (1630178) and DOD/DARPA award (FA8750-18-2-0259) to BWB and RPNR, NSF award EEC-1028725 to RPNR, the Alfred P. Sloan Foundation (BWB), and the Washington Research Foundation (BWB).

## Author Contributions

SHS, RPNR, and BWB conceived of the study/analysis. SHS and SMP performed the data analysis. SHS, SMP, RPNR, and BWB interpreted the results. SHS and BWB wrote the manuscript. SHS, SMP, RPNR and BWB edited the manuscript. RPNR and BWB acquired funding for the project.

## References

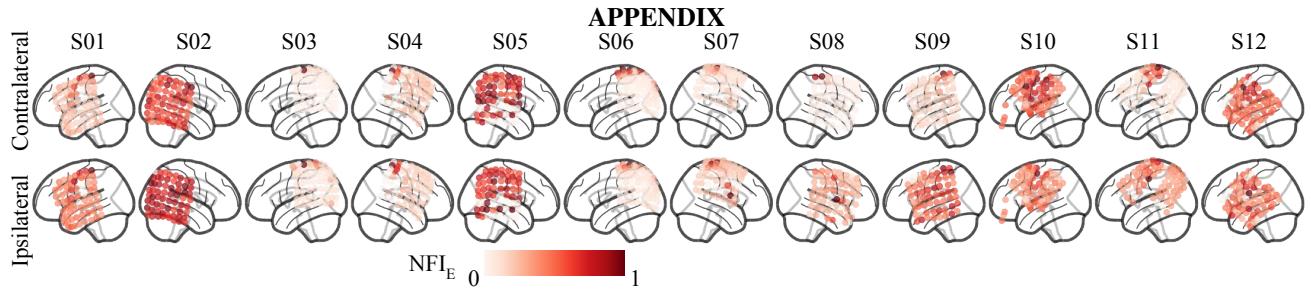
- H. Abbaspourazad, Y. Wong, B. Pesaran, and M. M. Shafechi. Identifying multiscale hidden states to decode behavior. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3778–3781. IEEE, 2018.
- A. Alasfour, P. Gabriel, X. Jiang, I. Shamie, L. Melloni, T. Thesen, P. Dugan, D. Friedman, W. Doyle, O. Devinsky, et al. Coarse behavioral context decoding. *Journal of Neural Engineering*, 16(1):016021, 2019.
- D. Anderson and P. Perona. Toward a Science of Computational Ethology. *Neuron*, 84(1):18–31, October 2014.
- N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. Al Moubayed, and T. P. Breckon. Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved SSVEP classification. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.
- E. Batty, M. Whiteway, S. Saxena, D. Biderman, T. Abe, S. Musall, W. Gillis, J. Markowitz, A. Churchland, J. Cunningham, et al. BehaveNet: nonlinear embedding and bayesian neural decoding of behavioral videos. In *Advances in Neural Information Processing Systems*, pages 15680–15691, 2019.
- A. M. Beck, E. P. Stephen, and P. L. Purdon. State space oscillator models for neural data analysis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4740–4743. IEEE, 2018.
- G. J. Berman, W. Bialek, and J. W. Shaevitz. Predictability and hierarchy in drosophila behavior. *Proceedings of the National Academy of Sciences*, 113(42):11943–11948, 2016.
- G. Berman. Measuring behavior across scales. *BMC Biology*, 16(1), December 2018.
- B. Blankertz, K.-R. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Biomedical Engineering*, 51(6):1044–1051, 2004.
- B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer. The BCI competition iii: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006.
- L. Bourdev and J. Malik. Poselets: Body part detectors trained using 3D human pose annotations. In *International Conference on Computer Vision*, sep 2009.
- L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.
- B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, 2016.
- Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.
- C. Chambers, N. Seethapathi, R. Saluja, H. Loeb, S. Pierce, D. Bogen, L. Prosser, M. J. Johnson, and K. P. Kording. Computer vision to automatically assess infant neuromotor risk. *BioRxiv*, page 756262, 2019.

- R. Chaudhuri, B. J. He, and X.-J. Wang. Random recurrent networks near criticality capture the broadband power distribution of human ECoG dynamics. *Cerebral Cortex*, 28(10):3610–3622, 2018.
- M. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press, January 2014.
- S. Cole and B. Voytek. Cycle-by-cycle analysis of neural oscillations. *Journal of Neurophysiology*, 122(2):849–861, 2019.
- S. R. Datta. Q&A: Understanding the composition of behavior. *BMC Biology*, 17(1):44, 2019.
- V. Elango, A. N. Patel, K. J. Miller, and V. Gilja. Sequence transfer learning for neural decoding. *BioRxiv*, page 210732, 2017.
- A. Farshchian, J. A. Gallego, J. P. Cohen, Y. Bengio, L. E. Miller, and S. A. Solla. Adversarial domain adaptation for stable brain-machine interfaces. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.
- D. Fu, W. Crichton, J. Hong, X. Yao, H. Zhang, A. Truong, A. Narayan, M. Agrawala, C. Ré, and K. Fatahalian. Rekall: Specifying Video Events using Compositions of Spatiotemporal Labels. *arXiv:1910.02993 [cs]*, October 2019.
- P. Gabriel, K. Chen, A. Alasfour, T. Pailla, W. Doyle, O. Devinsky, D. Friedman, P. Dugan, L. Melloni, T. Thesen, D. Gonda, S. Sattar, S. Wang, and V. Gilja. Neural Correlates of Unstructured Motor Behaviors. *Journal of Neural Engineering*, 16(6):066026, October 2019.
- S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje. Movi: A large multipurpose motion and video dataset. *arXiv preprint arXiv:2003.01888*, 2020.
- S. Gupta and A. Gomez-Marin. A context-free grammar for *caenorhabditis elegans* behavior. *BioRxiv*, page 708891, 2019.
- M. Haller, T. Donoghue, E. Peterson, P. Varma, P. Sebastian, R. Gao, T. Noto, R. T. Knight, A. Shestyuk, and B. Voytek. Parameterizing neural power spectra. *BioRxiv*, page 299859, 2018.
- B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in Neural Information Processing Systems*, pages 8527–8537, 2018.
- L. Hansen, M. Siebert, J. Diesel, and M. P. Heinrich. Fusing information from multiple 2D depth cameras for 3D human pose estimation in the operating room. *International Journal of Computer Assisted Radiology and Surgery*, 14(11):1871–1879, 2019.
- K. G. Hartmann, R. T. Schirrmeyer, and T. Ball. EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals. *arXiv preprint arXiv:1806.01875*, 2018.
- T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics, 2009.
- A. Huk, K. Bonnen, and B. He. Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli. *The Journal of Neuroscience*, pages 1920–17, July 2018.
- M. J. Johnson and A. S. Willsky. Bayesian Nonparametric Hidden Semi-Markov Models. *Journal of Machine Learning Research*, 14:673–701, February 2013.
- M. Johnson, D. Duvenaud, A. Wiltschko, R. Adams, and S. Datta. Composing graphical models with neural networks for structured representations and fast inference. In *Advances in Neural Information Processing Systems 29*, pages 2946–2954. Curran Associates, Inc., 2016.
- R. E. Johnson, S. Linderman, T. Panier, C. L. Wee, E. Song, K. J. Herrera, A. Miller, and F. Engert. Probabilistic models of larval zebrafish behavior reveal structure on many scales. *Current Biology*, 30(1):70–82, 2020.
- P. Karashchuk, K. L. Rupp, E. S. Dickinson, E. Sanders, E. Azim, B. W. Brunton, and J. C. Tuthill. Anipose: a toolkit for robust markerless 3D pose estimation. *BioRxiv*, 2020.
- S. L. Klosterman, J. R. Estep, J. W. Monnin, and J. C. Christensen. Day-to-day variability in hybrid, passive brain-computer interfaces: Comparing two studies assessing cognitive workload. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1584–1590. IEEE, 2016.
- T. Kobayashi, Y. Aoki, S. Shimizu, K. Kusano, and S. Okumura. Fine-grained action recognition in assembly work scenes by drawing attention to the hands. In *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 440–446. IEEE, 2019.
- A. Krause, H. B. McMahan, C. Guestrin, and A. Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9(Dec):2761–2801, 2008.
- Y. Li, J. Yang, Y. Song, L. Cao, J. Luo, and L.-J. Li. Learning from noisy labels with distillation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1910–1918, 2017.
- K. Luxem, F. Fuhrmann, J. Kuersch, S. Remy, and P. Bauer. Identifying behavioral structure from deep variational embeddings of animal motion. *BioRxiv*, 2020.

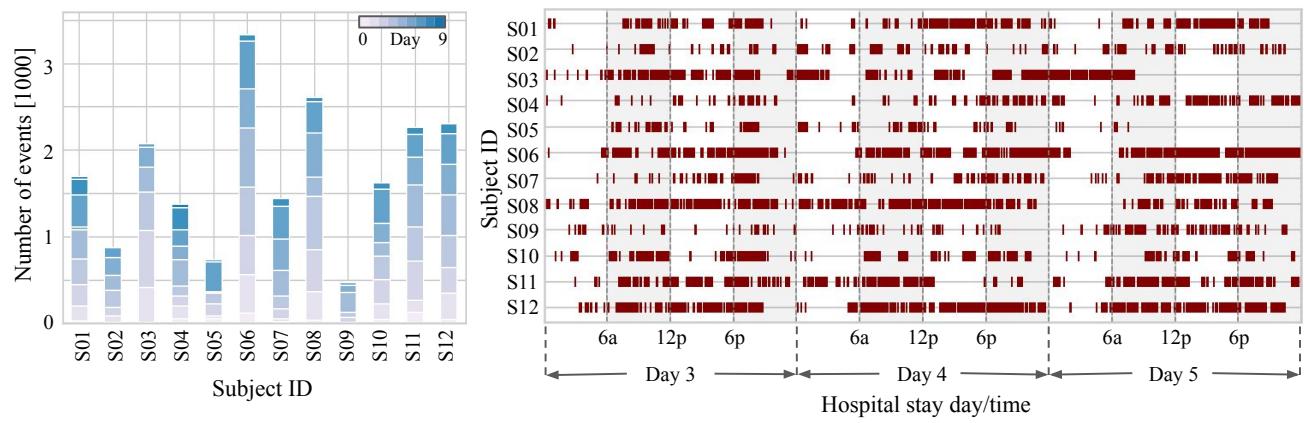
- K. Ly, J. Wu, L. Levinson, B. Shuman, K. Steele, J. Ojemann, and R. Rao. Electrocorticographic dynamics predict sustained grasping and upper-limb kinetic output. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 110–114. IEEE, 2018.
- J. Markowitz, W. Gillis, C. Beron, S. Neufeld, K. Robertson, N. Bhagat, R. Peterson, E. Peterson, M. Hyun, S. Linderman, B. Sabatini, and S. Datta. The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell*, 174(1):44–58.e17, June 2018.
- M. Mathis and A. Mathis. Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology*, 60:1–11, 2020.
- A. Mathis, P. Mamidanna, K. Cury, T. Abe, V. Murthy, M. Mathis, and M. Bethge. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Technical report, Nature Publishing Group, 2018.
- M. A. McDowell, C. D. Fryar, and C. L. Ogden. Anthropometric reference data for children and adults: United states, 1988–1994. *Vital and health statistics. Series 11, Data from the national health survey*, (249):1–68, 2009.
- K. Miller, E. Leuthardt, G. Schalk, R. Rao, N. Anderson, D. Moran, J. Miller, and J. Ojemann. Spectral Changes in Cortical Surface Potentials during Motor Movement. *Journal of Neuroscience*, 27(9):2424–2432, February 2007.
- K. J. Miller. A library of human electrocorticographic data and analyses. *Nature Human Behaviour*, 3(11):1225–1235, 2019.
- K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- J. Nassar, S. Linderman, M. Bugallo, and I. M. Park. Tree-structured recurrent switching linear dynamical systems for multi-scale modeling. In *International Conference on Learning Representations*, 2019.
- S. A. Nastase, A. Goldstein, and U. Hasson. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *PsyArXiv*, pages 1–12, 01 2020.
- N. Natarajan, I. S. Dhillon, P. K. Ravikumar, and A. Tewari. Learning with noisy labels. In *Advances in Neural Information Processing Systems*, pages 1196–1204, 2013.
- T. Nath, A. Mathis, A. Chen, A. Patel, M. Bethge, and M. Mathis. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14:2152–2176, 2019.
- A. Omigbodun, W. K. Doyle, O. Devinsky, D. Friedman, T. Thesen, and V. Gilja. Hidden-markov factor analysis as a spatiotemporal model for electrocorticography. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1632–1635. IEEE, 2016.
- R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen. Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:156869–156869, 2011.
- T. Pailla, K. J. Miller, and V. Gilja. Autoencoders for learning template spectrograms in electrocorticographic signals. *Journal of Neural Engineering*, 16(1):016025, 2019.
- C. Pandarinath, K. C. Ames, A. A. Russo, A. Farshchian, L. E. Miller, E. L. Dyer, and J. C. Kao. Latent factors and dynamics in motor cortex and their application to brain-machine interfaces. *Journal of Neuroscience*, 38(44):9390–9401, 2018.
- C. Pandarinath, D. J. O’Shea, J. Collins, R. Jozefowicz, S. D. Stavisky, J. C. Kao, E. M. Trautmann, M. T. Kaufman, S. I. Ryu, L. R. Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, page 1, 2018.
- J. Parvizi and S. Kastner. Human intracranial EEG: promises and limitations. *Nature Neuroscience*, 21(4):474, 2018.
- T. D. Pereira, D. E. Aldarondo, L. Willmore, M. Kislin, S. S.-H. Wang, M. Murthy, and J. W. Shaevitz. Fast animal pose estimation using deep neural networks. *Nature Methods*, 16(1):117, 2019.
- S. M. Peterson, S. H. Singh, N. X. Wang, R. P. Rao, and B. W. Brunton. Behavioral and neural variability of naturalistic arm movements. *BioRxiv*, 2020.
- A. Piergiovanni and M. S. Ryoo. Fine-grained activity recognition in baseball videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1740–1748, 2018.
- S. Ramasamy Ramamurthy and N. Roy. Recent trends in machine learning for human activity recognition - A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1254, 2018.
- M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele. A database for fine grained activity detection of cooking activities. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1194–1201. IEEE, 2012.
- D. Rolnick, A. Veit, S. Belongie, and N. Shavit. Deep learning is robust to massive label noise. *arXiv preprint arXiv:1705.10694*, 2017.

- Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, 2019.
- P. Sajda, A. Gerson, K.-R. Muller, B. Blankertz, and L. Parra. A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):184–185, 2003.
- N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris. 3D human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, 152:1–20, 2016.
- R. W. Schafer. What is a Savitzky-Golay filter? *IEEE Signal Processing Magazine*, 28(4):111–117, 2011.
- G. Schalk, J. Kubanek, K. Miller, N. Anderson, E. Leuthardt, J. Ojemann, D. Limbrick, D. Moran, L. Gerhardt, and J. Wolpaw. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *Journal of Neural Engineering*, 4(3):264, 2007.
- N. Seethapathi, S. Wang, R. Saluja, G. Blohm, and K. P. Kording. Movement science needs different pose tracking algorithms. *arXiv preprint arXiv:1907.10226*, 2019.
- M. M. Shanechi. Brain-machine interfaces. In *Dynamic Neuroscience*, pages 197–218. Springer, 2018.
- M. M. Shanechi. Brain-machine interfaces from motor to mood. *Nature Neuroscience*, 22(10):1554–1564, 2019.
- D. Shao, Y. Zhao, B. Dai, and D. Lin. Finegym: A hierarchical video dataset for fine-grained action understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2616–2625, 2020.
- K. Shenoy and C. Chestek. Neural prosthetics. *Scholarpedia*, 7(3):11854, 2012.
- P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller. Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3(1):R13, 2006.
- P. Shenoy, K. J. Miller, J. G. Ojemann, and R. P. Rao. Generalized features for electrocorticographic BCIs. *IEEE Transactions on Biomedical Engineering*, 55(1):273–280, 2007.
- Y. Shiraishi, Y. Kawahara, O. Yamashita, R. Fukuma, S. Yamamoto, Y. Saitoh, H. Kishima, and T. Yanagisawa. Neural decoding of electrocorticographic signals using dynamic mode decomposition. *Journal of Neural Engineering*, 2020.
- E. Smalley. The business of brain-computer interfaces. *Nature Biotechnology*, 37(9):978, 2019.
- A. Stolk, S. Griffin, R. van der Meij, C. Dewar, I. Saez, J. J. Lin, G. Piantoni, J.-M. Schoffelen, R. T. Knight, and R. Oostenveld. Integrated analysis of anatomical and electrophysiological human intracranial data. *Nature Protocols*, 13(7):1699–1723, 2018.
- D. Summers-Stay, C. L. Teo, Y. Yang, C. Fermüller, and Y. Aloimonos. Using a minimal action grammar for activity understanding in the real world. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4104–4111, 2012.
- W. Tam, T. Wu, Q. Zhao, E. Keefer, and Z. Yang. Human motor decoding from neural signals: a review. *BMC Biomedical Engineering*, 1(1):22, 2019.
- M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, et al. Review of the BCI competition iv. *Frontiers in Neuroscience*, 6:55, 2012.
- K. Volkova, M. A. Lebedev, A. Kaplan, and A. Ossadtchi. Decoding movement from electrocorticographic activity: A review. *Frontiers in Neuroinformatics*, 13, 2019.
- N. Wang, J. Olson, J. Ojemann, R. Rao, and B. Brunton. Unsupervised Decoding of Long-Term, Naturalistic Human Neural Recordings with Automated Video and Audio Annotations. *Frontiers in Human Neuroscience*, 10, April 2016.
- N. Wang, A. Farhadi, R. Rao, and B. Brunton. AJILE movement prediction: Multimodal deep learning for natural human neural recordings and video. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- D. J. Warren, S. Kellis, J. G. Nieven, S. M. Wendelken, H. Dantas, T. S. Davis, D. T. Hutchinson, R. A. Normann, G. A. Clark, and V. J. Mathews. Recording and decoding for neural prostheses. *Proceedings of the IEEE*, 104(2):374–391, 2016.
- K. Wei, Y. Liu, K. Kirchhoff, and J. Bilmes. Unsupervised submodular subset selection for speech data. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4107–4111. IEEE, 2014.
- K. Wei, R. Iyer, and J. Bilmes. Submodularity in data subset selection and active learning. In *International Conference on Machine Learning*, pages 1954–1963, 2015.
- A. Wiltschko, M. Johnson, G. Iurilli, R. Peterson, J. Katon, S. Pashkovski, V. Abraira, R. Adams, and S. Datta. Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135, December 2015.
- D. Wu, Y. Xu, and B. Lu. Transfer learning for EEG-based brain-computer interfaces: A review of progresses since 2016. *arXiv preprint arXiv:2004.06286*, 2020.

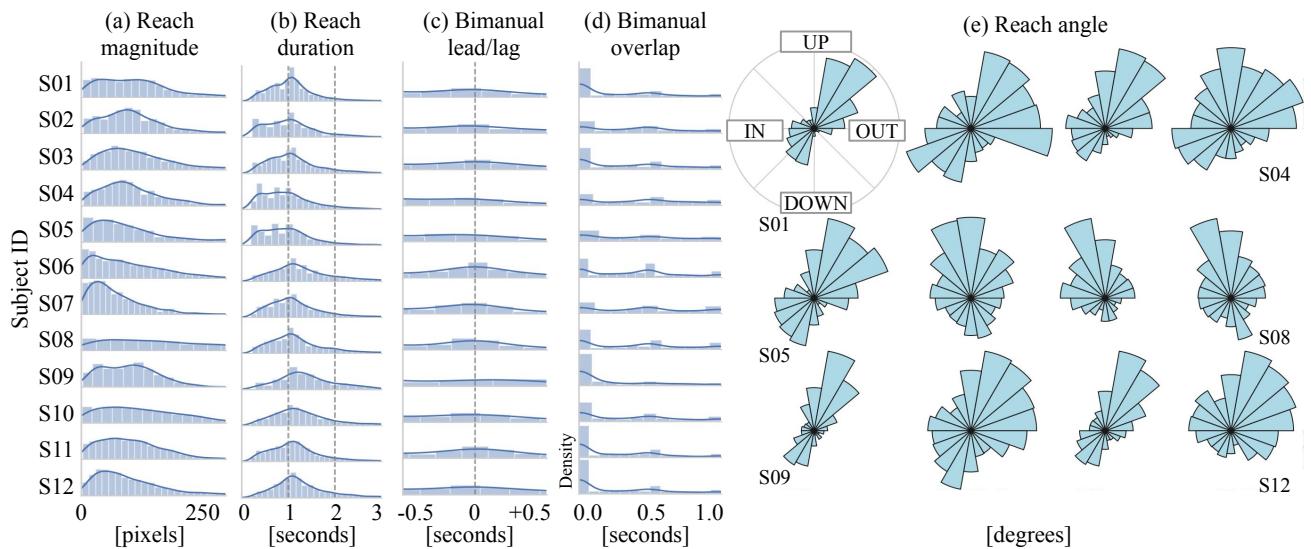
- Y. Yang, A. Guha, C. Fermüller, Y. Aloimonos, and A. V. Williams. A cognitive system for understanding human manipulation actions. In *Advances in Cognitive Systems*, pages 67–86, 2014.
- Y. Yang, E. F. Chang, and M. M. Shanechi. Dynamic tracking of non-stationarity in human ECoG activity. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1660–1663. IEEE, 2017.
- Y. Yang, O. G. Sani, E. F. Chang, and M. M. Shanechi. Dynamic network modeling and dimensionality reduction for human ECoG activity. *Journal of Neural Engineering*, 16(5):056014, 2019.
- H. Yuan and B. He. Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61:1425–1435, 2014.
- Y. Zhao and I. M. Park. Interpretable nonlinear dynamic modeling of neural trajectories. In *Advances in Neural Information Processing Systems*, pages 3333–3341, 2016.



**Figure 1.** Contralateral and ipsilateral wrist movement initiation decoder normalized feature importance scores aggregated by electrode (NFI<sub>E</sub>), showing spatial contributions of different brain regions for all 12 subjects. We see the same trend of motor cortex coverage being correlated with decoding accuracy, as was noted in Figure 8. Subjects having good motor cortex coverage (S07, S06, S03 and S11) have the highest decoding performance (Figure 7). Additionally, we see that electrodes with high normalized feature importance tend to be more spatially localized in the case where good motor cortex coverage is available.



**Figure 2.** Number of left-wrist movement initiation events discovered per day for each of 12 subjects, totaling 484 to 3338 events per subject across their entire duration of clinical observation ( $219 \pm 104$  s.d. per day). [Right] Raster plot of left-wrist movement initiation occurrences. See Figure 3 for equivalent plots for the right-wrist.



**Figure 3.** Histograms of left-wrist movement initiation event metadata per subject for their entire duration of clinical observation. We see all the same trends for the left-wrist as were noted for the right-wrist in Figure 5.