



Mining naturalistic human behaviors in long-term video and neural recordings

Satpreet H. Singh^a, Steven M. Peterson^{b,e}, Rajesh P.N. Rao^{c,a,d,f}, Bingni W. Brunton^{b,e,f,*}

^a Department of Electrical and Computer Engineering, University of Washington, Seattle, USA

^b Department of Biology, University of Washington, Seattle, USA

^c Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, USA

^d Center for Neurotechnology, University of Washington, Seattle, USA

^e eScience Institute, University of Washington, Seattle, USA

^f University of Washington Institute for Neuroengineering, Seattle, USA

ARTICLE INFO

Keywords:

Naturalistic behavior
Computer vision
Neural decoding
Electrocorticography
Brain-computer interfaces

ABSTRACT

Background: Recent technological advances in brain recording and machine learning algorithms are enabling the study of neural activity underlying spontaneous human behaviors, beyond the confines of cued, repeated trials. However, analyzing such unstructured data lacking *a priori* experimental design remains a significant challenge, especially when the data is multi-modal and long-term.

New method: Here we describe an automated, behavior-first approach for analyzing simultaneously recorded long-term, naturalistic electrocorticography (ECoG) and behavior video data. We identify and characterize spontaneous human upper-limb movements by combining computer vision, discrete latent-variable modeling, and string pattern-matching on the video.

Results: Our pipeline discovers and annotates over 40,000 instances of naturalistic arm movements in long term (7–9 day) behavioral videos, across 12 subjects. Analysis of the simultaneously recorded brain data reveals neural signatures of movement that corroborate previous findings. Our pipeline produces large training datasets for brain-computer interfacing applications, and we show decoding results from a movement initiation detection task.

Comparison with existing methods: Spontaneous movements capture real-world neural and behavior variability that is missing from traditional cued tasks. Building beyond window-based movement detection metrics, our unsupervised discretization scheme produces a queryable pose representation, allowing localization of movements with finer temporal resolution.

Conclusions: Our work addresses the unique analytic challenges of studying naturalistic human behaviors and contributes methods that may generalize to other neural recording modalities beyond ECoG. We publish our curated dataset and believe that it will be a valuable resource for future studies of naturalistic movements.

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

* Corresponding author at: Department of Biology, University of Washington, Seattle, USA.
E-mail address: bbrunton@uw.edu (B.W. Brunton).

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. (2020). 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 surgery. Our video processing pipeline (Fig. 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

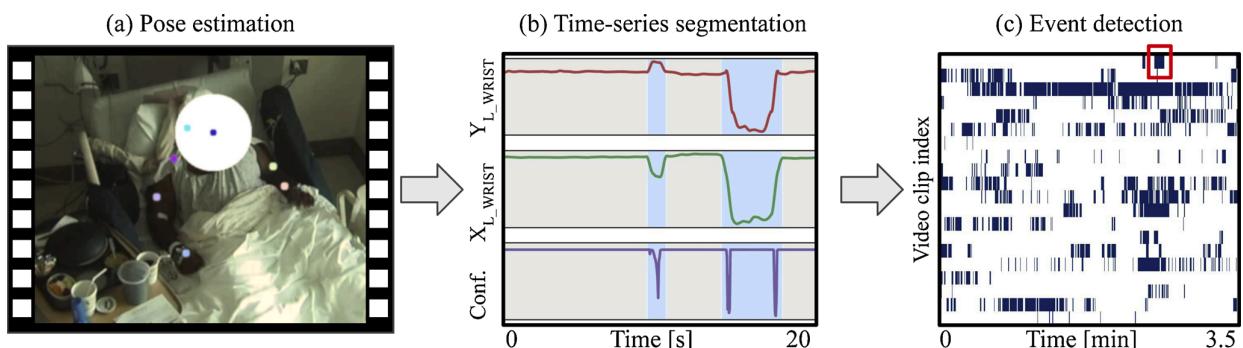


Fig. 1. Pipeline for behavioral video data processing. (a) Video frame showing estimated pose keypoints (colored dots) on human subject (see Supplementary Fig. A.13 for more examples). (b) Trajectories of estimated (x, y) pose coordinates and associated ‘confidence score’ for the left wrist. Autoregressive 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 (0.5 s) followed by 15 contiguous *move* states (0.5 s).

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, 2018a) 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 patient population comprised 8 males and 4 females, with an average age of 29.4 years (± 7.9 years s.d.) at the time of recording. 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 (Sony SNC-EP580 series, with day [RGB color] and night [near-infrared B/W] modes) 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. Patients continued to receive their clinician-determined doses of anti-epileptic and pain-relieving drugs throughout the recording period. Although there may be some cognitive impairment from medication, using electrocorticographic data from epilepsy patients for sensorimotor neuroscience and brain-computer interface tasks has become accepted practice in the field (Parvizi and Kastner, 2018b; Miller, 2019; Leuthardt et al., 2004, 2006).

Each patient had about 90 electrodes implanted under the skull and dura, directly on their brain surface, including either an 8×8 grid or two 8×4 grids of electrodes (see Supplementary Table A.1 for the exact number of electrodes per subject). These electrodes were placed to monitor a subset of cortical regions (right hemisphere for 5 and left hemisphere for 7 patients), predominantly in one brain hemisphere, as determined by individual clinical need for epilepsy resection surgery and independent of the analysis presented in this manuscript. Grid electrodes are constructed of 3-mm diameter platinum pads spaced 1-cm apart center-to-center and embedded in silastic (Ad-Tech Medical Instrument Corporation Epilepsy Long Term Monitoring subdural grid electrodes). The sampling rates for video and ECoG recordings (Natus Medical Incorporated XLTEK EMU 128 FS Headbox) were 30 frames/s (at 640 pixels \times 480 pixels resolution) 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 data preprocessing

We applied the following standard preprocessing steps to the ECoG data (Gabriel et al., 2019; Miller, 2019; Miller et al., 2007; Schalk et al., 2007): First, we removed DC drift by subtracting the median voltage (across entire recording) of each electrode. Next, in a manner similar to Delorme et al. (2001), we set to 0 V all electrode data for 2 s around high-amplitude artifacts, identified by times of abnormally high electrode-averaged absolute voltage (>50 interquartile range [IQR], across entire recording). This was implemented using custom Python code. Then, we band-pass filtered (1–200 Hz, FIR filter), notch filtered (60 Hz and harmonics up to 240 Hz, FIR notch filter) the data to minimize line noise, and downsampled the data to 500 Hz, all using the Python MNE package (version 0.22) (Gramfort et al., 2014) functions `filter.filter_data()` and `filter.notch_filter()` respectively.

Finally, any drifts common to all electrodes were removed by re-referencing the data at each time-step to the common median across electrodes. Electrodes that had abnormally high standard deviation (>5 IQR) or kurtosis (>10 IQR) compared to the across-electrodes median, were excluded from further analysis. All electrode positions were localized and converted to Montreal Neurological Institute and Hospital (MNI) coordinates using the Fieldtrip toolbox (version 20181206) (Oostenveld et al., 2011; Stolk et al., 2018) in MATLAB. Additional details about ECoG preprocessing and electrode localization are available in Peterson et al. (2020). To aid interpretability, we restricted our analysis in this paper to the 64 grid electrodes covering one hemisphere per subject. 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.

2.3. Video data preprocessing

Before processing the video data, we manually inspected and annotated it at a coarse (every 3 min or so) level of granularity to create an *omit-list* of video segments that were 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 Fig. 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 h per subject, depending on their clinical treatment regime, activity and sleep schedule, and length of hospital stay. When analyzing ECoG and video data together, the two data-streams were synchronized using metadata extracted using the equipment manufacturer's software (Natus Medical Incorporated, NeuroWorks Version 9.1.0.2658).

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 (Fig. 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 (Fig. 1a). Note that unlike some professional motion-tracking systems, there are no physical markers placed on the bodies of the subjects during any time.

During prediction, the pose-estimation tool produced the (x, y) coordinates and a *confidence* estimate between $[0, 1]$ for each keypoint per frame (Fig. 1b). To quantify the performance of keypoint tracking, we estimated the pixel-wise RMS error to be 1.54 mean ± 0.13 s.d. pixels on the training data and 5.97 mean ± 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 (Fig. 1b for left-wrist). This model converts each keypoint's continuous pose dynamics into discrete latent-state trajectories consisting of two distinct states, that we label as *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. As in Wiltschko et al. (2015), we fit the ARHSMM using the pyhsmm-autoregressive Python package (version/commit e6cfde5) (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 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 (0.5 s), followed by at least 15 consecutive *move* states (0.5 s). Similarly, *no-movement* events are state sequences of 90 *rest* states (3.0 s) 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. (See Supplementary Section A5.3 for additional details.)

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 (Fig. 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 1st, 2nd and 3rd-degree

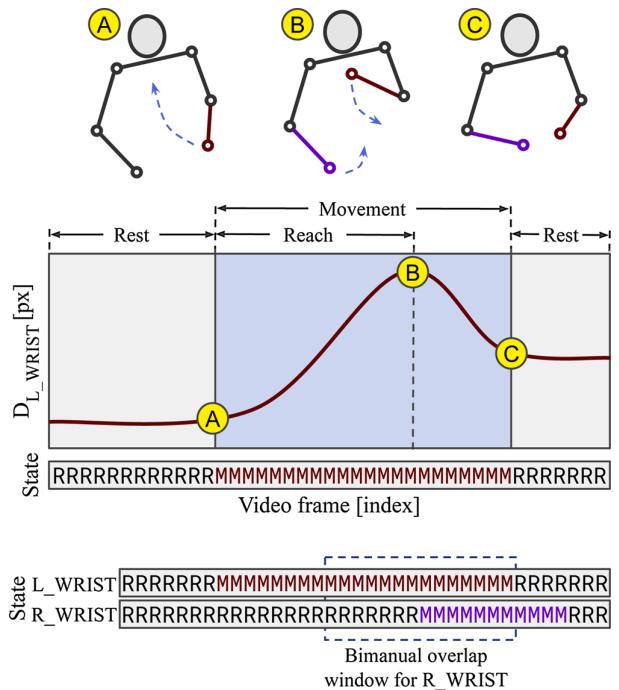


Fig. 2. Schematic of metadata extraction: [Top] Cartoon showing left-wrist movement at time of (A) movement initiation (B) maximum displacement from initiation (*reach*) and (C) movement end. [Middle] Time course of left-wrist radial distance [px: pixels], and discretized state (R: *rest*, M: *move*). Extracted movement metadata include duration, start and end coordinates, among others (Section 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.

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 (Fig. 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*).

Finally, we generated metadata that we used to remove false positives that arise from 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 body postures, 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. Validating curated events through application prototypes

We demonstrate the efficacy of our event generation pipeline as a data source for two prototype applications, neural correlates of movement initiation and decoding naturalistic movement initiation. These prototypes have been extensively developed upon in two companion papers (Peterson et al., 2020, 2021).

3.2.1. Neural correlates of movement initiation

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 positives ($17.8\% \text{ mean} \pm 9.9\% \text{ s.d. events}$) (see Supplementary Table A.4 for full list of events selected per subject).

3.2.2. Decoding naturalistic movement initiation

A grand challenge in neuroengineering is the development of BCIs that can be used to predict spontaneous activity and intentions outside the lab, in everyday settings (Shanechi, 2018, 2019; Smalley, 2019; 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 0.5 s before to 0.5 s 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 $200 \text{ ms} \times 5 \text{ Hz}$ 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. From these, we used events from the first two days as the training set and events from the last day as the test set (see Supplementary Table A.5 for exact list of days chosen for each subject). To eliminate the confound of movement initiation in the opposing wrist, we then excluded events with significant movement ($\geq 0.2 \text{ s}$) in the opposing wrist within the $\pm 0.5 \text{ s}$ 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 sets up a theoretical baseline performance of 50% accuracy for test set performance. Training and test supports were $633 \text{ mean} \pm 417 \text{ s.d.}$ and $331 \text{ mean} \pm 203 \text{ s.d.}$ examples, respectively (see Supplementary Table A.5 for exact per-subject values). 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. 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 ($5\% \text{ mean} \pm 10\% \text{ s.d.}$ for no-movement, $22\% \text{ mean} \pm 16\% \text{ s.d.}$ for contralateral, and $14\% \text{ mean} \pm 10\% \text{ s.d.}$ for ipsilateral events, aggregated over all subjects).

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 and frequency components. Spectral features are indexed by electrode, time, and frequency. We define Feature Importance aggregated by Electrode $\text{FI}_E(e)$ for electrode $e \in \mathcal{E}$ as:

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

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 $\text{NFI}_E(e)$:

$$\text{NFI}_E(e) = \frac{\text{FI}_E(e)}{\max_{e \in \mathcal{E}} \text{FI}_E(e)}.$$

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

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

$$\text{NFI}_F(f) = \frac{\text{FI}_F(f)}{\max_{f \in \mathcal{F}} \text{FI}_F(f)}$$

4. Results

4.1. Characterizing naturalistic events

Our pipeline extracted 959 to 6745 individual wrist movement events per subject ($487 \text{ mean} \pm 215 \text{ per day}$) across 12 subjects. (Figs. 3 and A.11 for right and left wrists, respectively). We observed rich within-subject variability in the event metadata: Figs. 5 and A.12 visualize the distributions for several movement associated metadata for right and left wrists, respectively. A random sample of trajectories for one patient (S10) is shown in Fig. 4.

4.2. Validating curated events through application prototypes

4.2.1. Neural correlates of movement initiation

We observed (see Fig. 6) movement-associated power increases in a high-frequency band (76–100 Hz) and decreases in a low-frequency band (8–32 Hz) across four cortical areas (Precentral, Inferior Parietal, Postcentral and Supramarginal), with strongest responses in the Precentral area (Miller et al., 2007; Miller, 2019).

4.2.2. Decoding naturalistic movement initiation

Individual subject classifier performance varied widely between subjects, ranging from around chance levels (50%) to 80% on test accuracy (Fig. 8). Classifier performance tended to be associated with the extent of motor cortex coverage (Fig. 7 and A.10). Test set decoding accuracy (Fig. 8) was observed to be higher for the contralateral wrist in almost all subjects. Figs. 7 and 9 show the spatial and spectral (frequency) contributions of the classifier's features respectively. We found (Fig. 7; also see Supplementary Fig. A.10) that electrodes over sensorimotor cortex, when available, dominated feature importance (e.g. Subjects S07, S06, S03 and S11 in Fig. 8) in most but not all patients. As seen in Fig. 9, we found that a low-frequency band (<35 Hz) and a high-frequency band (around 100 Hz) dominates feature importance in most but not all patients. We extensively develop upon the limited prototype described here in our companion paper focused on decoding (Peterson et al., 2021).

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

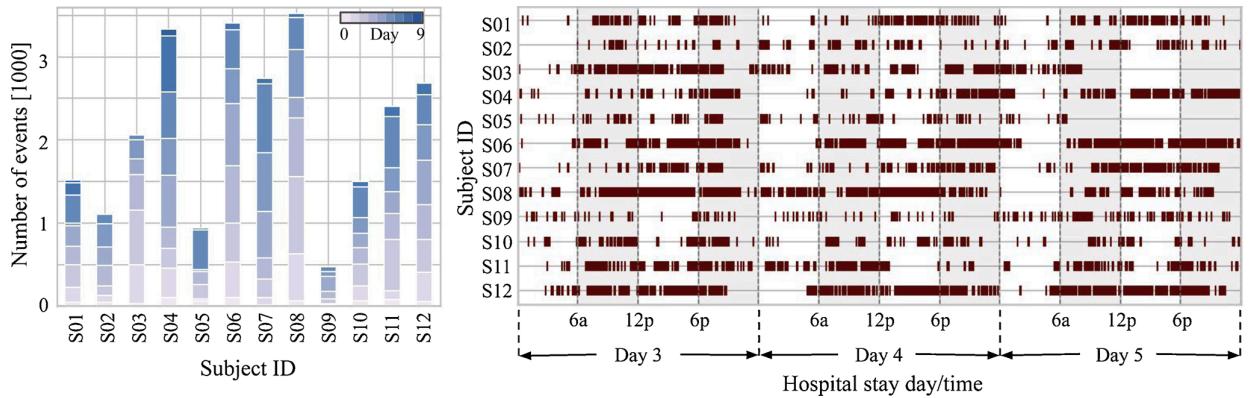


Fig. 3. [Left] Number of right-wrist movement initiation events discovered per day for each of 12 subjects, totaling 475 to 3526 events per subject across their entire duration of clinical observation (268 mean \pm 123 s.d. per day). [Right] Raster plot of right-wrist movement initiation event occurrences showing bursts of activity interspersed with periods of rest or omit-listed (Section 3.1.1) periods (see Fig. A.11 for equivalent plots for left-wrist).

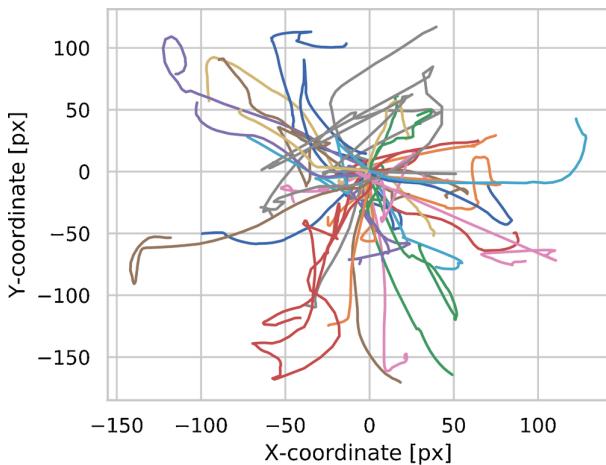


Fig. 4. A sample of 50 typical right-wrist trajectories (px; pixels; translated to start at origin) showing diversity of naturalistic reach movements for a single subject (S10). Different colors represent different individual trajectories. Note the large variability in the movements, compared to what is normally captured by controlled experiments (see Supplementary Fig. A.13 for instructions to access video-snippets of such movements).

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. We found large variability between subjects in the number of events discovered, which we attribute to inter-subject differences in times excluded in the omit-listing process, cycles of sleep and wakeful activity, and clinical treatment regimes (Figs. 3 and A.11).

We take a “mining” approach to analyzing these large datasets, which contain high variability coming from not only the diversity of patients and human behaviors, but also due to the opportunistic data-collection paradigm. Thus, large chunks of the recordings are discarded, so there is large inter-subject variability in the number of events discovered. We believe this highlights the open research challenges associated with analyzing such data. The within-subject variability observed in our event metadata (Figs. 4, 5, and A.12) is expected to closely reflect the natural statistics of human upper-limb movements while seated, since our subjects received no instructions for when and

how to move.

5.1. Validating curated events through application prototypes

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 initiation, and decoding naturalistic movement initiation using ECoG data. Key to the success of our applications 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).

Our prototype spectrograms (Fig. 6) showed a pattern of movement-associated power increases in a high-frequency band (76–100 Hz) and decreases in a low-frequency band (8–32 Hz) in the same areas as reported in previous controlled laboratory experiments (Miller et al., 2007; Miller, 2019). In our concurrently released preprint (Peterson et al., 2020), we further investigate the consequences of the relatively higher variance of naturalistic movement statistics, model the contributions of the various movement metadata to the observed neural responses, and contrast prior controlled experimental findings with those from our dataset to highlight differences due to the different time-scales of data-collection (a few hours for controlled experiments vs. over several days for ours). In particular, analyses comparing naturalistic and controlled experimental data in Peterson et al. (2020) were enabled by our ability to match movement statistics between naturalistic and controlled experiments by selecting events by reach magnitude, onset velocity, and shape metadata.

Similarly, our decoding results show (Fig. 9) that a low-frequency band (<35 Hz) and a high-frequency band (around 100 Hz) dominates feature importance for most subjects (with some exceptions). This pattern is similar to what has been observed with controlled brain-computer interface studies such as Miller et al. (2007), Volkova et al. (2019), and Yuan and He (2014). Our decoding performance, as measured by test set accuracy, is similar to those in previous studies that also use such opportunistically collected ECoG data (Gabriel et al., 2019; Wang et al., 2018). Unlike these previous studies, the ability to select events without opposing wrist activity allowed us to disambiguate confounds when comparing movement decoders for opposing wrists. In most subjects, contralateral limb movement decoding is more accurate than it is for ipsilateral movements, which is expected due to hemispheric lateralization (Tam et al., 2019). When motor cortex electrode coverage is limited or unavailable (Fig. 7), decoding is still possible (Fig. 8), which is likely due to the decoders exploiting correlations between brain-regions (Tam et al., 2019; Miller et al., 2007; Schalk et al., 2007). When motor cortex coverage is available, we see decoder feature

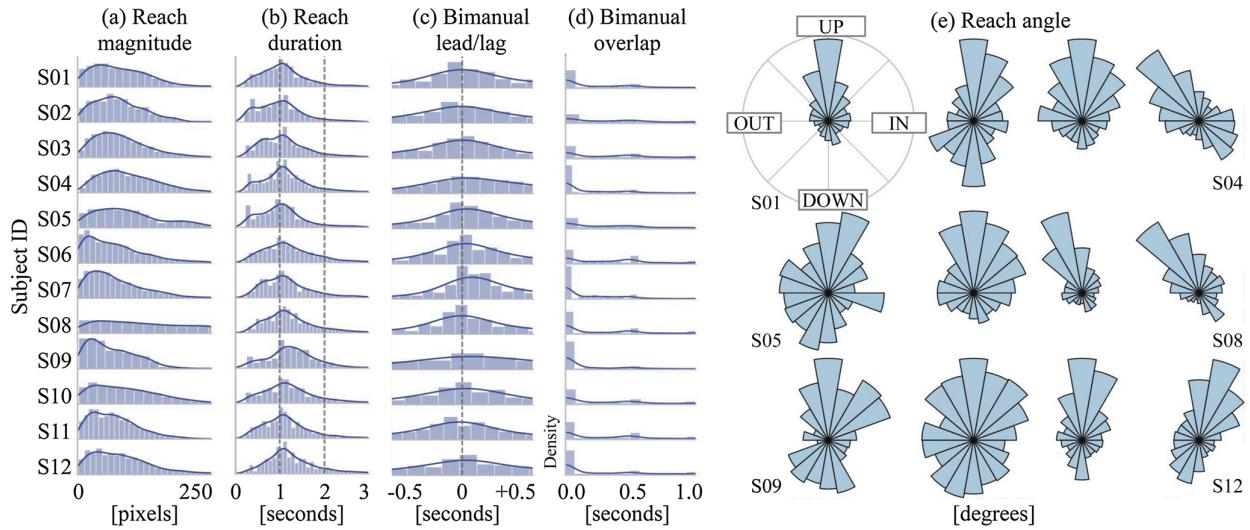


Fig. 5. Histograms of right-wrist movement initiation event metadata per subject for their entire duration of clinical observation: (a) reach magnitudes show a dominance of small movements (92 pixels/30.6 cm mean \pm 22 pixels/7.3 cm s.d. of subject-wise medians), (b) Reach durations tended to be quite short (1.2 s mean \pm 0.1 s s.d. of subject-wise medians), (c) when both hands moved together (“bimanually”), they tended to start at about the same time, with the left hand start very slightly earlier (leading by 0.2 s mean \pm 0.1 s s.d. of subject-wise medians; negative values in plot imply opposing wrist leading primary), (d) duration of time left hand was moving during a \pm 0.5 s window around time of right hand movement initiation (0.3 s mean \pm 0.06 s s.d. of subject-wise medians). (e) Polar histograms show that many subjects primarily made upward-downward reaches. However, reaches in almost every other direction were also observed. (95° mean \pm 19° s.d. of subject-wise histogram mode bin, with angles measured counterclockwise) (for subject-level metadata statistics, see Supplementary Table A.2). See Fig. A.12 for equivalent plots for left-wrist.

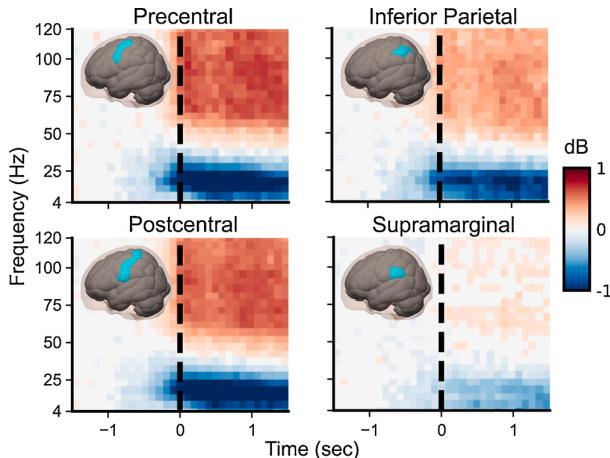


Fig. 6. Neural correlates of movement initiation: event-locked spectrograms, averaged by brain region (cyan color in insets) across 12 subjects, showed movement-associated high-frequency power increase and low-frequency power decrease. These patterns corroborate and strengthen previous findings from 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.

importance concentrate itself on this area (Miller et al., 2007; Gabriel et al., 2019; Wang et al., 2018). The large dataset produced by this approach has also critically enabled the development of a generalized neural-network based decoder adept at transferring across subjects and across recording modalities (Peterson et al., 2021).

5.2. Limitations and future work

Our work has a number of limitations that can be improved upon in future work. First, our strategy of discretizing individual keypoint time-

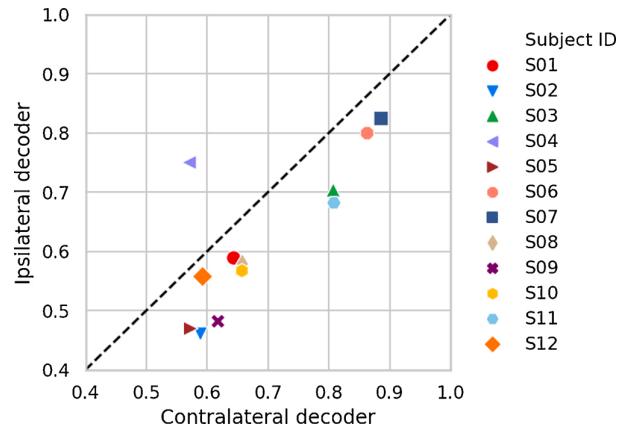


Fig. 7. Contralateral and ipsilateral wrist movement initiation decoder feature importance scores aggregated by electrode (NFI_i), showing spatial contributions of different brain regions. Scores are normalized by dividing by highest electrode score for each decoder. Electrode coverage over motor cortex is associated with higher decoder accuracy; for instance, subjects having good motor cortex coverage (S07, S06, S03 and S11) have the highest decoding performance (Fig. 8). See Fig. A.10 for plot with all 12 subjects.

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

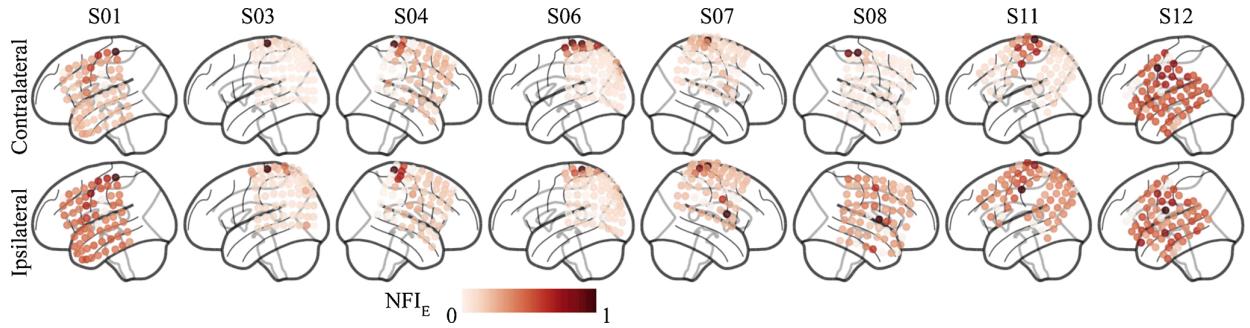


Fig. 8. Test set decoding accuracy for initiation of movement of contralateral (side opposite electrode implant) and ipsilateral (same side) wrists: as expected, decoding of contralateral movements is slightly more accurate than ipsilateral in almost all cases.

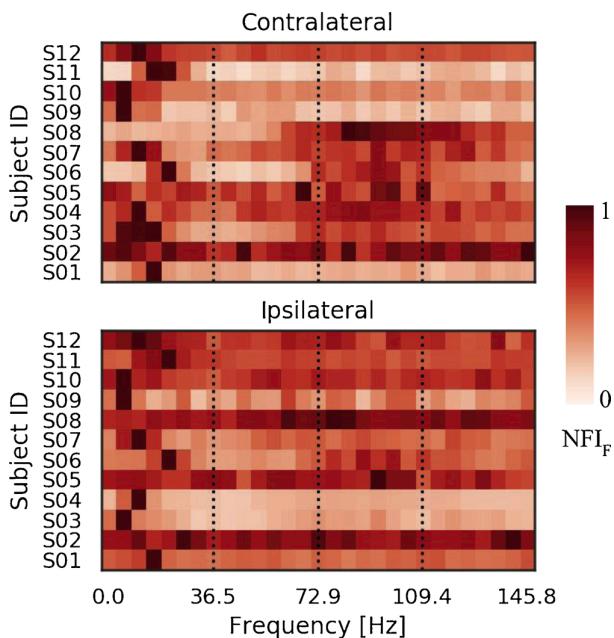


Fig. 9. Decoder feature importance scores aggregated by frequency and normalized by score of highest frequency bin per subject (NFI_F). Heatmaps show that the most relevant spectral features tend to come from a low-frequency band (<35 Hz) and a high-frequency band (around 100 Hz), similar to prior findings (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 associated with lower decoding accuracy.

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. For example, we are exploring the neural correlates and decoding of coarse reach directions (inwards, outwards, upwards and downwards) instead of exact reach angles in ongoing work. 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 trade-off 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.

5.3. 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 10–100s 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 believe that our dataset could serve several other lines of scientific inquiry. As a follow-up to the movement-initiation task described in this paper, we are currently exploring the use of neural network based decoders for (coarse-grained) event metadata such as reach angle, reach goal (what activity/object the reach was for) and large changes in body posture. The abundant availability of neural data allows us to learn representations that are interpretable and task-specific (Pailla et al., 2019; Shiraishi et al., 2020). Our curated 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,b; Ly et al., 2018; Cole and Voytek, 2019; Zhao and Park, 2016), dynamical modeling of the electrocorticographic spectrum (Chaudhuri et al., 2018; Beck et al., 2018; Donoghue et al., 2020; 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, 2017; Abbaspourazad et al., 2018), 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 and data to reproduce several key plots in this manuscript is publicly available at: <https://github.com/BruntonUWBio/mining2021>. Our complete curated dataset consisting of events and their metadata and associated neural data (≈ 40 GB) can also be downloaded following the instructions provided at the aforementioned URL. Video examples of (left, right, & bimanual) wrist movements for one subject (S01) can also be found there.

Authors' contribution

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.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank John So for extensive help with manual annotation of the video data, and Nathan Davis for help fixing a bug with a video annotation script. 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 and a CJ and Elizabeth Hwang Professorship to RPNR, the Alfred P. Sloan Foundation (BWB), and the Washington Research Foundation (BWB).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jneumeth.2021.109199>.

References

- Abbaspourazad, H., Wong, Y., Pesaran, B., Shafechi, M.M., 2018. Identifying multiscale hidden states to decode behavior. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 3778–3781.
- Alasfour, A., Gabriel, P., Jiang, X., Shamie, I., Melloni, L., Thesen, T., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., et al., 2019. Coarse behavioral context decoding. *J. Neural Eng.* 16, 016021.
- Anderson, D., Perona, P., 2014. Toward a science of computational ethology. *Neuron* 84, 18–31. <https://doi.org/10.1016/j.neuron.2014.09.005>.
- Aznan, N.K.N., Atapour-Abarghouei, A., Bonner, S., Connolly, J.D., Al Moubayed, N., Breckon, T.P., 2019. Simulating brain signals: creating synthetic EEG data via neural-based generative models for improved SSVEP classification. 2019 International Joint Conference on Neural Networks (IJCNN) 1–8.
- Batty, E., Whiteway, M., Saxena, S., Biderman, D., Abe, T., Musall, S., Gillis, W., Markowitz, J., Churchland, A., Cunningham, J., et al., 2019. BehaveNet: nonlinear embedding and Bayesian neural decoding of behavioral videos. *Advances in Neural Information Processing Systems*, pp. 15680–15691.
- Beck, A.M., Stephen, E.P., Purdon, P.L., 2018. State space oscillator models for neural data analysis. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 4740–4743.
- Berman, G., 2018. Measuring behavior across scales. *BMC Biol.* 16 <https://doi.org/10.1186/s12915-018-0494-7>.
- Berman, G.J., Bialek, W., Shaevitz, J.W., 2016. Predictability and hierarchy in drosophila behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11943–11948.
- Blankertz, B., Muller, K.R., Curio, G., Vaughan, T.M., Schalk, G., Wolpaw, J.R., Schlogl, A., Neuper, C., Pfurtscheller, G., Hinterberger, T., et al., 2004. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans. Biomed. Eng.* 51, 1044–1051.
- Blankertz, B., Muller, K.R., Krusinski, D.J., Schalk, G., Wolpaw, J.R., Schlogl, A., Pfurtscheller, G., Millan, J.R., Schroder, M., Birbaumer, N., 2006. The BCI competition iii: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159.
- Bourdev, L., Malik, J., 2009. Poslets: body part detectors trained using 3D human pose annotations. *International Conference on Computer Vision*.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brunton, B.W., Johnson, L.A., Ojemann, J.G., Kutz, J.N., 2016. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods* 258, 1–15.
- Cao, Z., Simon, T., Wei, S.E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 7291–7299.
- Chambers, C., Seethapathi, N., Saluja, R., Loeb, H., Pierce, S.R., Bogen, D.K., Prosser, L., Johnson, M.J., Kording, K.P., 2020. Computer vision to automatically assess infant neuromotor risk. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2431–2442.
- Chaudhuri, R., He, B.J., Wang, X.J., 2018. Random recurrent networks near criticality capture the broadband power distribution of human ECoG dynamics. *Cereb. Cortex* 28, 3610–3622.
- Cohen, M., 2014. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.
- Cole, S., Voytek, B., 2019. Cycle-by-cycle analysis of neural oscillations. *J. Neurophysiol.* 122, 849–861.
- Datta, S.R., 2019. Q&A: understanding the composition of behavior. *BMC Biol.* 17, 44.

- Delorme, A., Makeig, S., Sejnowski, T., 2001. Automatic artifact rejection for eeg data using high-order statistics and independent component analysis. Proceedings of the Third International ICA Conference 9–12.
- Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., et al., 2020. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* 23, 1655–1665.
- Farshchian, A., Gallego, J.A., Cohen, J.P., Bengio, Y., Miller, L.E., Solla, S.A., 2019. Adversarial domain adaptation for stable brain-machine interfaces. 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019.
- Fu, D., Crichton, W., Hong, J., Yao, X., Zhang, H., Truong, A., Narayan, A., Agrawala, M., Ré, C., Fatahalian, K., 2019. Rekall: Specifying Video Events using Compositions of Spatiotemporal Labels. arXiv:1910.02993 [cs].
- Gabriel, P., Chen, K., Alasfour, A., Pailla, T., Doyle, W., Devinsky, O., Friedman, D., Dugan, P., Melloni, L., Thesen, T., Gonda, D., Sattar, S., Wang, S., Gilja, V., 2019. Neural correlates of unstructured motor behaviors. *J. Neural Eng.* 16, 066026. <https://doi.org/10.1088/1741-2552/ab355c>.
- Ghorbani, S., Mahdaviani, K., Thaler, A., Kording, K., Cook, D.J., Blohm, G., Troje, N.F., 2020. Movi: A Large Multipurpose Motion and Video Dataset. arXiv:2003.01888.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hännäläinen, M.S., 2014. Mne software for processing meg and eeg data. *Neuroimage* 86, 446–460.
- Gupta, S., Gomez-Marin, A., 2019. A context-free grammar for *Caenorhabditis elegans* behavior. *BioRxiv* 708891.
- Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M., 2018. Co-teaching: robust training of deep neural networks with extremely noisy labels. Advances in Neural Information Processing Systems, pp. 8527–8537.
- Hansen, L., Siebert, M., Diesel, J., Heinrich, M.P., 2019. Fusing information from multiple 2D depth cameras for 3D human pose estimation in the operating room. *Int. J. Comput. Assist. Radiol. Surg.* 14, 1871–1879.
- Hartmann, K.G., Schirrmeyer, R.T., Ball, T., 2018. EEG-GAN: Generative Adversarial Networks for Electroencephalographic (EEG) Brain Signals. arXiv:1806.01875.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics.
- Huk, A., Bonnen, K., He, B., 2018. Beyond Trial-based paradigms: continuous behavior, ongoing neural activity, and natural stimuli. *J. Neurosci.* <https://doi.org/10.1523/JNEUROSCI.1920-17.2018>.
- Johnson, M., Duvenaud, D., Wiltschko, A., Adams, R., Datta, S., 2016. Composing graphical models with neural networks for structured representations and fast inference. Advances in Neural Information Processing Systems 29. Curran Associates, Inc, pp. 2946–2954.
- Johnson, M.J., Willsky, A.S., 2013. Bayesian nonparametric hidden semi-Markov models. *J. Mach. Learn. Res.* 14, 673–701.
- Johnson, R.E., Linderman, S., Panier, T., Wee, C.L., Song, E., Herrera, K.J., Miller, A., Engert, F., 2020. Probabilistic models of larval zebrafish behavior reveal structure on many scales. *Curr. Biol.* 30, 70–82.
- Karashchuk, P., Rupp, K.L., Dickinson, E.S., Sanders, E., Azim, E., Brunton, B.W., Tuthill, J.C., 2020. Anipose: a toolkit for robust markerless 3D pose estimation. *BioRxiv*.
- Klosterman, S.L., Estep, J.R., Monnin, J.W., Christensen, J.C., 2016. Day-to-day variability in hybrid, passive brain-computer interfaces: comparing two studies assessing cognitive workload. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 1584–1590.
- Kobayashi, T., Aoki, Y., Shimizu, S., Kusano, K., Okumura, S., 2019. Fine-grained action recognition in assembly work scenes by drawing attention to the hands. 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) 440–446.
- Krause, A., McMahan, H.B., Guestrin, C., Gupta, A., 2008. Robust submodular observation selection. *J. Mach. Learn. Res.* 9, 2761–2801.
- Leuthardt, E.C., Miller, K.J., Schalk, G., Rao, R.P., Ojemann, J.G., 2006. Electrocorticography-based brain computer interface – the seattle experience. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 194–198.
- Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G., Moran, D.W., 2004. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* 1, 63.
- Li, Y., Yang, J., Song, Y., Cao, L., Luo, J., Li, L.J., 2017. Learning from noisy labels with distillation. Proceedings of the IEEE International Conference on Computer Vision 1910–1918.
- Luxem, K., Fuhrmann, F., Kuersch, J., Remy, S., Bauer, P., 2020. Identifying behavioral structure from deep variational embeddings of animal motion. *BioRxiv*.
- Ly, K., Wu, J., Levinson, L., Shuman, B., Steele, K., Ojemann, J., Rao, R., 2018. Electrocorticographic dynamics predict sustained grasping and upper-limb kinetic output. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) 110–114.
- Markowitz, J., Gillis, W., Beron, C., Neufeld, S., Robertson, K., Bhagat, N., Peterson, R., Peterson, E., Hyun, M., Linderman, S., Sabatini, B., Datta, S., 2018. The striatum organizes 3D behavior via moment-to-moment action selection. *Cell* 174. <https://doi.org/10.1016/j.cell.2018.04.019>, 44–58.e17.
- Mathis, A., Mamidanna, P., Cury, K., Abe, T., Murthy, V., Mathis, M., Bethge, M., 2018. DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts With Deep Learning. Nature Publishing Group. Technical Report.
- Mathis, M., Mathis, A., 2020. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr. Opin. Neurobiol.* 60, 1–11.
- McDowell, M.A., Fryar, C.D., Ogden, C.L., 2009. Anthropometric Reference Data for Children and Adults: United States, 1988–1994. Vital and Health Statistics. Series 11, Data from the National Health Survey, pp. 1–68.
- Miller, K., Leuthardt, E., Schalk, G., Rao, R., Anderson, N., Moran, D., Miller, J., Ojemann, J., 2007. Spectral changes in cortical surface potentials during motor movement. *J. Neurosci.* 27, 2424–2432. <https://doi.org/10.1523/JNEUROSCI.3886-06.2007>.
- Miller, K.J., 2019. A library of human electrocorticographic data and analyses. *Nat. Hum. Behav.* 3, 1225–1235.
- Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. The MIT Press.
- Nassar, J., Linderman, S., Bugallo, M., Park, I.M., 2019. Tree-structured recurrent switching linear dynamical systems for multi-scale modeling. International Conference on Learning Representations.
- Nastase, S.A., Goldstein, A., Hasson, U., 2020. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* 222, 117254.
- Natarajan, N., Dhillon, I.S., Ravikumar, P.K., Tewari, A., 2013. Learning with noisy labels. Advances in Neural Information Processing Systems, pp. 1196–1204.
- Nath, T., Mathis, A., Chen, A., Patel, A., Bethge, M., Mathis, M., 2019. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* 14, 2125–2126.
- Omigodou, A., Doyle, W.K., Devinsky, O., Friedman, D., Thesen, T., Gilja, V., 2016. Hidden-Markov factor analysis as a spatiotemporal model for electrocorticography. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 1632–1635.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience 2011, 156869. <https://doi.org/10.1155/2011/156869>.
- Pailla, T., Miller, K.J., Gilja, V., 2019. Autoencoders for learning template spectrograms in electrocorticographic signals. *J. Neural Eng.* 16, 016025.
- Pandarinath, C., Ames, K.C., Russo, A.A., Farshchian, A., Miller, L.E., Dyer, E.L., Kao, J.C., 2018a. Latent factors and dynamics in motor cortex and their application to brain-machine interfaces. *J. Neurosci.* 38, 9390–9401.
- Pandarinath, C., O’Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., et al., 2018b. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 1.
- Parvizi, J., Kastner, S., 2018a. Human intracranial EEG: promises and limitations. *Nat. Neurosci.* 21, 474.
- Parvizi, J., Kastner, S., 2018b. Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* 1.
- Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.H., Murthy, M., Shaevitz, J.W., 2019. Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117.
- Peterson, S.M., Singh, S.H., Wang, N.X., Rao, R.P., Brunton, B.W., 2020. Behavioral and neural variability of naturalistic arm movements. *BioRxiv*.
- Peterson, S.M., Steine-Hanson, Z., Davis, N., Rao, R.P., Brunton, B.W., 2021. Generalized neural decoders for transfer learning across participants and recording modalities. *J. Neural Eng.* 18, 026014.
- Piergiovanni, A., Ryoo, M.S., 2018. Fine-grained activity recognition in baseball videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 1740–1748.
- Ramasamy Ramamurthy, S., Roy, N., 2018. Recent trends in machine learning for human activity recognition – a survey. Wiley Interdisc. Rev.: Data Mining Knowl. Discov. 8, e1254.
- Rohrbach, M., Amin, S., Andriluka, M., Schiele, B., 2012. A database for fine grained activity detection of cooking activities. 2012 IEEE Conference on Computer Vision and Pattern Recognition 1194–1201.
- Rolnick, D., Veit, A., Belongie, S., Shavit, N., 2017. Deep Learning Is Robust to Massive Label Noise. arXiv:1705.10694.
- Sajda, P., Gerson, A., Muller, K.R., Blankertz, B., Parra, L., 2003. A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 184–185.
- Sarafianos, N., Boteanu, B., Ionescu, B., Kakadiaris, I.A., 2016. 3D human pose estimation: a review of the literature and analysis of covariates. *Comput. Vision Image Understand.* 152, 1–20.
- Schafer, R.W., 2011. What is a Savitzky-Golay filter? *IEEE Signal Process. Mag.* 28, 111–117.
- Schalk, G., Kubanek, J., Miller, K., Anderson, N., Leuthardt, E., Ojemann, J., Limbrick, D., Moran, D., Gerhardt, L., Wolpaw, J., 2007. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *J. Neural Eng.* 4, 264.
- Seethapathi, N., Wang, S., Saluja, R., Blohm, G., Kording, K.P., 2019. Movement Science Needs Different Pose Tracking Algorithms. arXiv:1907.10226.
- Shanechi, M.M., 2018. Brain-machine interfaces. Dynamic Neuroscience. Springer, pp. 197–218.
- Shanechi, M.M., 2019. Brain-machine interfaces from motor to mood. *Nat. Neurosci.* 22, 1554–1564.
- Shao, D., Zhao, Y., Dai, B., Lin, D., 2020. Finegym: a hierarchical video dataset for fine-grained action understanding. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2616–2625.
- Shenoy, K., Chestek, C., 2012. Neural prosthetics. *Scholarpedia* 7, 11854.
- Shenoy, P., Krauledat, M., Blankertz, B., Rao, R.P., Müller, K.R., 2006. Towards adaptive classification for BCI. *J. Neural Eng.* 3, R13.
- Shiraishi, Y., Kawahara, Y., Yamashita, O., Fukuma, R., Yamamoto, S., Saitoh, Y., Kishima, H., Yanagisawa, T., 2020. Neural decoding of electrocorticographic signals using dynamic mode decomposition. *J. Neural Eng.*
- Smalley, E., 2019. The business of brain-computer interfaces. *Nat. Biotechnol.* 37, 978.

- Stolk, A., Griffin, S., van der Meij, R., Dewar, C., Saez, I., Lin, J.J., Piantoni, G., Schoffelen, J.M., Knight, R.T., Oostenveld, R., 2018. Integrated analysis of anatomical and electrophysiological human intracranial data. *Nat. Protoc.* 13, 1699–1723. <https://doi.org/10.1038/s41596-018-0009-6>.
- Summers-Stay, D., Teo, C.L., Yang, Y., Fermüller, C., Aloimonos, Y., 2012. Using a minimal action grammar for activity understanding in the real world. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* 4104–4111.
- Tan, W., Wu, T., Zhao, Q., Keefer, E., Yang, Z., 2019. Human motor decoding from neural signals: a review. *BMC Biomed. Eng.* 1, 22.
- Tangermann, M., Müller, K.R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K.J., Mueller-Putz, G., et al., 2012. Review of the BCI competition iv. *Front. Neurosci.* 6, 55.
- Volkova, K., Lebedev, M.A., Kaplan, A., Ossadtchi, A., 2019. Decoding movement from electrocorticographic activity: review. *Front. Neuroinformatics* 13.
- Wang, N., Farhadi, A., Rao, R., Brunton, B., 2018. AJILE movement prediction: multimodal deep learning for natural human neural recordings and video. *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Wang, N., Olson, J., Ojemann, J., Rao, R., Brunton, B., 2016. Unsupervised decoding of long-term, naturalistic human neural recordings with automated video and audio annotations. *Front. Human Neurosci.* 10 <https://doi.org/10.3389/fnhum.2016.00165>.
- Warren, D.J., Kellis, S., Nieven, J.G., Wendelken, S.M., Dantas, H., Davis, T.S., Hutchinson, D.T., Normann, R.A., Clark, G.A., Mathews, V.J., 2016. Recording and decoding for neural prostheses. *Proc. IEEE* 104, 374–391.
- Wei, K., Iyer, R., Bilmes, J., 2015. Submodularity in data subset selection and active learning. *International Conference on Machine Learning* 1954–1963.
- Wei, K., Liu, Y., Kirchhoff, K., Bilmes, J., 2014. Unsupervised submodular subset selection for speech data. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4107–4111.
- Wiltschko, A., Johnson, M., Iurilli, G., Peterson, R., Katon, J., Pashkovski, S., Abraira, V., Adams, R., Datta, S., 2015. Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135. <https://doi.org/10.1016/j.neuron.2015.11.031>.
- Yang, Y., Chang, E.F., Shafechi, M.M., 2017. Dynamic tracking of non-stationarity in human ECoG activity. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 1660–1663.
- Yang, Y., Guha, A., Fermüller, C., Aloimonos, Y., Williams, A.V., 2014. A cognitive system for understanding human manipulation actions. *Advances in Cognitive Systems*, pp. 67–86.
- Yang, Y., Sani, O.G., Chang, E.F., Shafechi, M.M., 2019. Dynamic network modeling and dimensionality reduction for human ECoG activity. *J. Neural Eng.* 16, 056014.
- Yuan, H., He, B., 2014. Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Trans. Biomed. Eng.* 61, 1425–1435.
- Zhao, Y., Park, I.M., 2016. Interpretable nonlinear dynamic modeling of neural trajectories. *Advances in Neural Information Processing Systems*, pp. 3333–3341.