Towards human SuperEEG

Lucy L. W. Owen[1], Andrew C. Heusser[1,2], and Jeremy R. Manning[1*]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA
[2]Akili Interactive, Boston, MA 02110, USA

Abstract

Human *SuperEEG*[1] entails measuring ongoing neural activity throughout the entire living brain with perfect precision and arbitrarily high spatiotemporal resolution. Although true SuperEEG is impossible using existing methods, here we present a model-based method for *inferring* full-brain neural activity at millimeter-scale spatial resolutions and millisecond-scale temporal resolutions using standard human intracranial recordings. Our approach assumes that different people's brains exhibit similar spatial correlations, and that activity and correlation patterns are spatially smooth. One can then ask, for an arbitrary individual's brain: given recordings from a limited set of locations in that individual's brain, along with the observed spatial correlations learned from other people's recordings, how much can be inferred about ongoing activity at *other* locations in that individual's brain?

Keywords: Electrocorticography (ECoG), intracranial electroencephalography (iEEG), local field potential (LFP), epilepsy, maximum likelihood estimation, Gaussian process regression

Introduction

Modern human brain recording techniques are fraught with compromise [33]. Commonly used approaches include functional magnetic resonance imaging (fMRI), scalp electroencephalography (EEG), and magnetoencephalography (MEG). For each of these techniques, neuroscientists and electrophysiologists must choose to optimize spatial resolution at the cost of temporal resolution (e.g., as in fMRI) or temporal resolution at the cost of spatial resolution (e.g., as in EEG and MEG). A less widely used approach (due to requiring work with neurosurgical patients) is to record from electrodes implanted directly onto the cortical surface (electrocorticography; ECoG)

[1]The term "SuperEEG" was coined by Robert J. Sawyer in his popular science fiction novel *The Terminal Experiment* [29]

or into deep brain structures (intracranial EEG; iEEG). However, these intracranial approaches also require compromise: the high spatiotemporal resolution of intracranial recordings comes at the cost of substantially reduced brain coverage, since safety considerations limit the number of electrodes one may implant in a given patient's brain. Further, the locations of implanted electrodes are determined by clinical, rather than research, needs.

An increasingly popular approach is to improve the effective spatial resolution of MEG or scalp EEG data by using a geometric approach called *beamforming* to solve the biomagnetic or bioelectrical inverse problem [28]. This approach entails using detailed brain conductance models (often informed by high spatial resolution anatomical MRI images) along with the known sensor placements (localized precisely in 3D space) to reconstruct brain signals originating from theoretical point sources deep in the brain (and far from the sensors). Traditional beamforming approaches must overcome two obstacles. First, the inverse problem beamforming seeks to solve has infinitely many solutions. Researchers have made traction towards constraining the solution space by assuming that signal-generating sources are localized on a regularly spaced grid spanning the brain and that individual sources are small relative to their distances to the sensors [1, 11, 34]. The second, and in some ways much more serious, obstacle is that the magnetic fields produced by external (noise) sources are substantially stronger than those produced by the neuronal changes being sought (i.e., at deep structures, as measured by sensors at the scalp). This means that obtaining adequate signal quality often requires averaging the measured responses over tens to hundreds of responses or trials (e.g., see review by [11]).

Another approach to obtaining high spatiotemporal resolution neural data has been to collect fMRI and EEG data simultaneously. Simultaneous fMRI-EEG has the potential to balance the high spatial resolution of fMRI with the high temporal resolution of scalp EEG, thereby, in theory, providing the best of both worlds. In practice, however, the signal quality of both recordings suffers substantially when the two techniques are applied simultaneously (e.g., see review by [13]). In addition, the experimental designs that are ideally suited to each technique individually are somewhat at odds. For example, fMRI experiments often lock stimulus presentation events to the regularly spaced image acquisition time (TR), which maximizes the number of post-stimulus samples. By contrast, EEG experiments typically employ jittered stimulus presentation times to maximize the experimentalist's ability to distinguish electrical brain activity from external noise sources such as from 60 Hz alternating current power sources.

The current "gold standard" for precisely localizing signals and sampling at high temporal resolution is to take (ECoG or iEEG) recordings from implanted electrodes (but from a limited set of locations in any given brain). This begs the following question: what can we infer about the activity exhibited by the rest of a person's brain, given what we learn from the limited intracranial recordings we have from their brain and additional recordings taken from *other* people's brains? Here we develop an approach, which we call *SuperEEG*, based on Gaussian process regression [27]. SuperEEG entails using data from multiple people to estimate activity patterns at arbitrary locations in each person's brain (i.e., independent of their electrode placements). We test our SuperEEG approach using two large datasets of intracranial recordings [7, 8, 12, 16–19, 21, 23, 30–32, 35, 40]. We show that the SuperEEG algorithm recovers signals well from electrodes that were held out of the training dataset. We also examine the factors that influence how accurately activity may be estimated (recovered), which may have implications for electrode design and placement in neurosurgical applications.

Approach

The SuperEEG approach to inferring high temporal resolution full-brain activity patterns is outlined and summarized in Figure 1. We describe (in this section) and evaluate (in *Results*) our approach using a two large previously collected dataset comprising multi-session intracranial recordings. Dataset 1 comprises multi-session recordings taken from 6876 electrodes implanted in the brains of 88 epilepsy patients [21, 23, 30–32]. Each recording session lasted from 0.2–3 h (total recording time: 0.3–14.2 h; Fig. S4A). During each recording session, the patients participated in a free recall list learning task, which lasted for up to approximately 1 h. In addition, the recordings included "buffer" time (the length varied by patient) before and after each experimental session, during which the patients went about their regular hospital activities (confined to their hospital room, and primarily in bed). These additional activities included interactions with medical staff and family, watching television, reading, and other similar activities. For the purposes of the Dataset 1 analyses presented here, we aggregated all data across each recording session, including recordings taken during the main experimental task as well as during non-experimental time. We used Dataset 1 to develop our main SuperEEG approach, and to examine the extent to which SuperEEG might be able to generate task-general predictions. Dataset 2 comprised multi-session recordings from 4436 electrodes implanted in the brains of 40 epilepsy patients [7, 8, 12, 16–19, 35, 40]. Each recording session lasted from 0.4–2.2 h (total recording time: 0.4–6.6 h; Fig. S4B). Whereas Dataset 1 included recordings taken as the patients participated in a variety of activities, Dataset 2 included recordings taken as each patient performed each of two specific experimental memory tasks: a random word list free recall task (Experiment 1) and a categorized word list free recall task (Experiment 2). We used Dataset 2 to further examine the ability of SuperEEG to generalize its predictions within versus across tasks. Figure S4 provides additional information about both datasets.

We first applied fourth order Butterworth notch filter to remove 60 Hz (± .5 Hz) line noise from every recording (from every electrode). Next, we downsampled the recordings (regardless of the original samplerate) to 250 Hz. (This downsampling step served to both normalize for differences in sampling rates across patients and to ease the computational burden of our subsequent analyses.) We then excluded any electrodes that showed putative epileptiform activity. Specifically, we excluded from further analysis any electrode that exhibited an average kurtosis

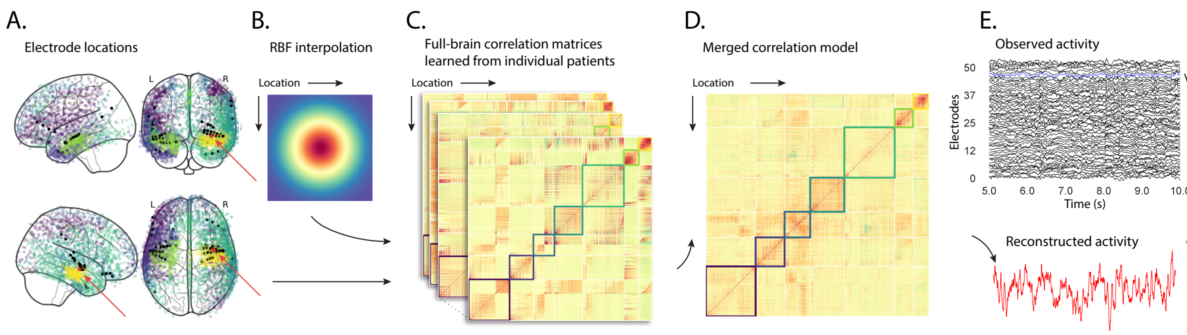

Figure 1: **Methods overview. A. Electrode locations.** Each dot reflects the location of a single electrode implanted in the brain of a Dataset 1 patient. A held-out recording location from one patient is indicated in red, and the patient's remaining electrodes are indicated in black. The electrodes from the remaining patients are colored by k-means cluster (computed using the full-brain correlation model shown in Panel D). **B. Radial basis function kernel.** Each electrode contributed by the patient (black) weights on the full set of locations under consideration (all dots in Panel A, defined as $\bar{R}$ in the text). The weights fall off with positional distance (in MNI space) according to an RBF. **C. Per-patient correlation matrices.** After computing the pairwise correlations between the recordings from each patient's electrodes, we use RBF-weighted averages to estimate correlations between all locations in $\bar{R}$. We obtain an estimated full-brain correlation matrix using each patient's data. **D. Merged correlation model.** We combine the per-patient correlation matrices (Panel C) to obtain a single full-brain correlation model that captures information contributed by every patient. Here we have sorted the rows and columns to reflect k-means clustering labels [using k=7; 41], whereby we grouped locations based on their correlations with the rest of the brain (i.e., rows of the matrix displayed in the panel). The boundaries denote the cluster groups. The rows and columns of Panel C have been sorted using the Panel D-derived cluster labels. **E. Reconstructing activity throughout the brain.** Given the observed recordings from the given patient (shown in black; held-out recording is shown in blue), along with a full-brain correlation model (Panel D), we use Equation 12 to reconstruct the most probable activity at the held-out location (red).

of 10 or greater across all of that patient's recording sessions. We also excluded any patients with fewer than 2 electrodes that passed this criteria, as the SuperEEG algorithm requires measuring correlations between 2 or more electrodes from each patient. For Dataset 1, this yielded clean recordings from 4168 electrodes implanted throughout the brains of 67 patients (Fig. 1A); for Dataset 2, this yielded clean recordings from 3159 electrodes from 24 patients. Each individual patient contributes electrodes from a limited set of brain locations, which we localized in a common space [MNI152; 10]; an example Dataset 1 patient's 54 electrodes that passed the predefined kurtosis test are highlighted in black and red.

The recording from a given electrode is maximally informative about the activity of the neural tissue immediately surrounding its recording surface. However, brain regions that are distant from the recording surface of the electrode also contribute to the recording, albeit (ceteris paribus) to a much lesser extent. One mechanism underlying these contributions is volume conduction. The precise rate of falloff due to volume conduction (i.e., how much a small volume of brain tissue at location x contributes to the recording from an electrode at location η) depends on the size of the recording surface, the electrode's impedance, and the conductance profile of the volume of brain between x and η. As an approximation of this intuition, we place a Gaussian radial basis function (RBF) at the location η of each electrode's recording surface (Fig. 1B). We use the values of the RBF at any brain location x as a rough estimate of how much structures around x contributed to the recording from location η:

$$\mathrm{rbf}(x|\eta, \lambda) = \exp\left\{-\frac{||x-\eta||^2}{\lambda}\right\}, \tag{1}$$

where the width variable λ is a parameter of the algorithm (which may in principle be set according to location-specific tissue conductance profiles) that governs the level of spatial smoothing. In choosing λ for the analyses presented here, we sought to maximize spatial resolution (which implies a small value of λ) while also maximizing the algorithm's ability to generalize to any location throughout the brain, including those without dense electrode coverage (which implies a large value of λ). Here we set $\lambda = 20$, guided in part by our prior work [22, 24], and in part by examining the brain coverage with non-zero weights achieved by placing RBFs at each electrode location in Dataset 1 and taking the sum (across all electrodes) at each voxel in a 4 mm^3 MNI brain. (We then held λ fixed for our analyses of Dataset 2.) We note that this value could in theory be further optimized, e.g., using cross validation or a formal model [e.g., 24].

A second mechanism whereby a given region x can contribute to the recording at η is through (direct or indirect) anatomical connections between structures near x and η. We use temporal correlations in the data to estimate these anatomical connections [2]. Let $\bar{R}$ be the set of locations at which we wish to estimate local field potentials, and let $R_s \subseteq \bar{R}$ be set of locations at which we observe local field potentials from patient s (excluding the electrodes that did not pass the kurtosis test described above). In the analyses below we define $\bar{R} = \cup_{s=1}^{S} R_s$. We can calculate the expected inter-electrode correlation matrix for patient s, where $C_{s,k}(i,j)$ is the correlation between the time series of voltages for electrodes i and j from subject s during session k, using:

$$\bar{C}_s = \mathrm{r}\left(\frac{1}{n}\left(\sum_{k=1}^{n} \mathrm{z}(C_{s,k})\right)\right), \text{ where} \tag{2}$$

$$\mathrm{z}(r) = \frac{\log(1+r) - \log(1-r)}{2} \text{ is the Fisher } z\text{−transformation and} \tag{3}$$

$$\mathrm{z}^{-1}(z) = \mathrm{r}(z) = \frac{\exp(2z) - 1}{\exp(2z) + 1} \text{ is its inverse.} \tag{4}$$

Next, we use Equation 1 to construct a number of to-be-estimated locations by number of patient electrode locations weight matrix, W_s. Specifically, W_s approximates how informative the recordings at each location in R_s are in reconstructing activity at each location in $\bar{R}$, where the contributions fall off with an RBF according to the distances between the corresponding locations:

$$W_s(i, j) = \mathrm{rbf}(i|j, \lambda). \tag{5}$$

Given this weight matrix, W_s, and the observed inter-electrode correlation matrix for patient s, $\bar{C}_s$, we can estimate the correlation matrix for all locations in $\bar{R}$ ($\hat{C}_s$; Fig. 1C) using:

$$\hat{N}_s(x, y) = \sum_{i=1}^{|R_s|} \sum_{j=1}^{i-1} W(x, i) \cdot W(y, j) \cdot \mathrm{z}(\bar{C}_s(i, j)) \tag{6}$$

$$\hat{D}_s(x, y) = \sum_{i=1}^{|R_s|} \sum_{j=1}^{i-1} W(x, i) \cdot W(y, j). \tag{7}$$

$$\hat{C}_s = \mathrm{r}\left(\frac{\hat{N}_s}{\hat{D}_s}\right). \tag{8}$$

After estimating the numerator ($\hat{N}_s$) and denominator ($\hat{D}_s$) placeholders for each $\hat{C}_s$, we aggregate these estimates across the S patients to obtain a single expected full-brain correlation matrix ($\hat{K}$; Fig. 1D):

$$\hat{K} = \mathrm{r}\left(\frac{\sum_{s=1}^{S} \hat{N}_s}{\sum_{s=1}^{S} \hat{D}_s}\right). \tag{9}$$

Intuitively, the numerators capture the general structures of the patient-specific estimates of full-brain correlations, and the denominators account for which locations were near the implanted electrodes in each patient. To obtain $\hat{K}$, we compute a weighted average across the estimated patient-specific full-brain correlation matrices, where patients with observed electrodes near a particular set of locations in $\hat{K}$ contribute more to the estimate.

Having used the multi-patient data to estimate a full-brain correlation matrix at the set of locations in $\bar{R}$ that we wish to know about, we next use $\hat{K}$ to estimate activity patterns everywhere in $\bar{R}$, given observations at only a subset of locations in $\bar{R}$ (Fig. 1E).

Let α_s be the set of indices of patient s's electrode locations in $\bar{R}$ (i.e., the locations in R_s), and let β_s be the set of indices of all other locations in $\bar{R}$. In other words, β_s reflects the locations in $\bar{R}$ where we did not observe a recording for patient s (these are the recording locations we will want to fill in using SuperEEG). We can sub-divide $\hat{K}$ as follows:

$$\hat{K}_{\beta_s,\alpha_s} = \hat{K}(\beta_s, \alpha_s), \text{ and} \tag{10}$$

$$\hat{K}_{\alpha_s,\alpha_s} = \hat{K}(\alpha_s, \alpha_s). \tag{11}$$

Here $\hat{K}_{\beta_s,\alpha_s}$ represents the correlations between the "unknown" activity at the locations in β_s and the observed activity at the locations in α_s, and $\hat{K}_{\alpha_s,\alpha_s}$ represents the correlations between the observed recordings (at the locations in α_s).

Let Y_{s,k,α_s} be the number-of-timepoints (T) by $|\alpha_s|$ matrix of (observed) voltages from the electrodes in α_s during session k from patient s. Then we can estimate the voltage from patient s's k^{th} session at the locations in β_s using [27]:

$$\hat{Y}_{s,k,\beta_s} = ((\hat{K}_{\beta_s,\alpha_s} \cdot \hat{K}^{-1}_{\alpha_s,\alpha_s}) \cdot Y^{\mathrm{T}}_{s,k,\alpha_s})^{\mathrm{T}}. \tag{12}$$

This equation is the foundation of the SuperEEG algorithm. Whereas we observe recordings only at the locations indexed by α_s, Equation 12 allows us to estimate the recordings at all locations indexed by β_s, which we can define *a priori* to include any locations we wish, throughout the brain. This yields estimates of the time-varying voltages at *every* location in $\bar{R}$, provided that we define $\bar{R}$ in advance to include the union of all of the locations in R_s and all of the locations at which we wish to estimate recordings (i.e., a timeseries of voltages).

We designed our approach to be agnostic to electrode impedances, as electrodes that do not exist do not have impedances. Therefore our algorithm recovers voltages in standard deviation (z-scored) units rather than attempting to recover absolute voltages. (This property reflects the fact that $\hat{K}_{\beta_s,\alpha_s}$ and $\hat{K}_{\alpha_s,\alpha_s}$ are correlation matrices rather than covariance matrices.) Also, note that Equation 12 requires computing a T by T matrix, which can become computationally intractable when T is very large (e.g., for the patient highlighted in Fig. 2, $T = 12786750$). However, because Equation 12 is time invariant, we may compute Y_{s,k,β_s} in a piecewise manner by filling in Y_{s,k,β_s} one row at a time (using the corresponding samples from Y_{s,k,α_s}).

The SuperEEG algorithm described above and in Figure 1 allows us to estimate, up to a constant scaling factor, local field potentials (LFPs) for each patient at all arbitrarily chosen locations in the set $\bar{R}$, *even if we did not record that patient's brain at all of those locations*. We next turn to an evaluation of the accuracy of those estimates.

Results

We used a cross-validation approach to test the accuracy with which the SuperEEG algorithm reconstructs activity throughout the brain. For each patient in turn, we estimated full-brain correlation matrices (Eqn. 9) using data from all of the *other* patients. This step ensured that the data we were reconstructing could not also be used to estimate the between-location correlations that drove the reconstructions via Equation 12 (otherwise the analysis would be circular). For that held-out patient, for each of their electrodes in turn, we used Equation 12 to reconstruct activity at the held-out electrode location, using the correlation matrix learned from all other patients' data as $\hat{K}$, and using activity recorded from the other electrodes from the held-out patient as Y_{s,k,α_s}. We then asked: how closely did each of the SuperEEG-estimated recordings at those electrodes match the observed recordings from those electrodes (i.e., how closely did the estimated $\hat{Y}_{s,k,\beta_s}$ match the observed Y_{s,k,β_s}?).

To illustrate our approach, we first examine an individual held-out raw LFP trace and its associated SuperEEG-derived reconstruction. Figure 2A displays the observed LFP from the red electrode in Figure 1A (blue), and its associated reconstruction (red), during a 5 s time window during one of the example patient's six recording sessions. The two traces match closely ($r = 0.86, p < 10^{-10}$). Figure 2B displays a two-dimensional histogram of the actual versus

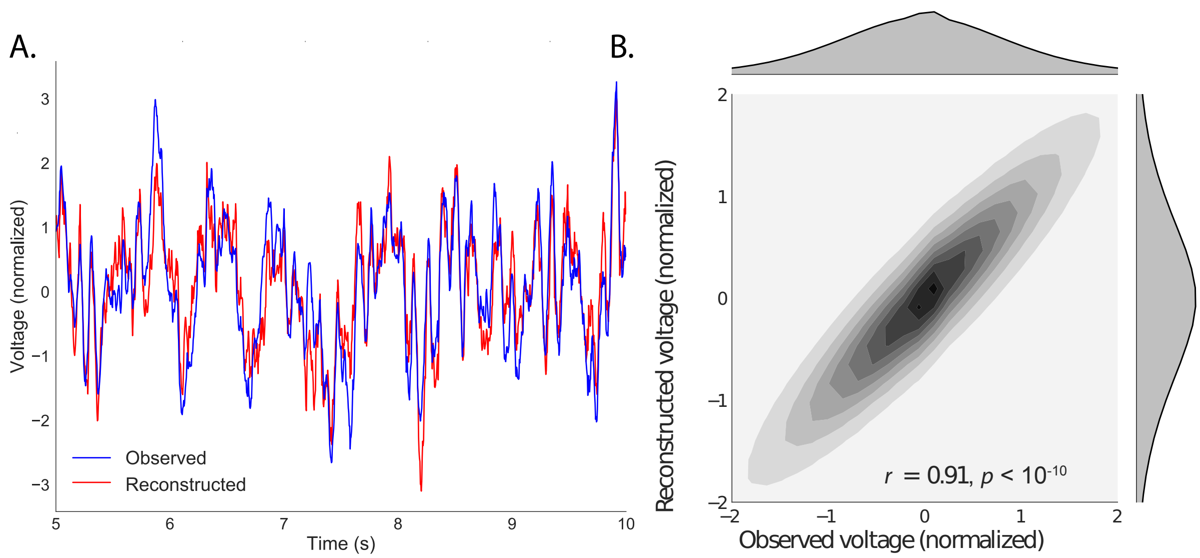

Figure 2: **Observed and reconstructed LFP from a single electrode. A. Example LFP.** A 5 s recording from the red electrode in Figure 1A is displayed in blue, and the reconstructed LFP during the same time window is shown in red. **B. Observed versus reconstructed LFP over 14.2 hours.** The two-dimensional histogram reflects the relation between distributions of observed versus reconstructed voltages from one patient, across the 14.2 hours of recorded data collected over 6 recording sessions. The correlation reported in the panel is between the observed and reconstructed voltages. Both panels: all voltages are represented in standard deviation units (computed within session).

reconstructed voltages for the entire 14.2 total hours of recordings from the example electrode (correlation: $r = 0.91, p < 10^{-10}$). This example confirms that the SuperEEG algorithm recovers the recordings from this single electrode well. Next, we used this general approach to quantify the algorithm's performance across the full dataset.

For each held-out electrode, from each held-out patient in turn, we computed the average correlation (across recording sessions) between the SuperEEG-reconstructed voltage traces and the observed voltage traces from that electrode. For this analysis we set $\bar{R}$ to be the union of all electrode locations across all patients. This yielded a single correlation coefficient for each electrode location in $\bar{R}$, reflecting how well the SuperEEG algorithm was able to recover the recording at that location by incorporating data across patients (black histogram in Fig. 3A, map in Fig. 3C). The observed distribution of correlations was centered well above zero (mean: 0.52; t-test comparing mean of distribution of z-transformed average patient correlation coefficients to 0: $t(66) = 25.08, p < 10^{-10}$), indicating that the SuperEEG algorithm recovers held-out activity patterns substantially better than random guessing.

As a stricter benchmark, we compared the quality of these across-participant reconstructions (i.e., computed using a correlation model learned from other patients' data) to reconstructions generated using a correlation model trained using the in-patient's data. In other words, for this within-patient benchmark analysis we estimated $\hat{C}_s$ (Eqn. 8) for each patient in turn, using recordings from all of that patient's electrodes except at the location we were reconstructing. These within-patient reconstructions serve as an estimate of how well data from all of the other electrodes from that single patient may be used to estimate held-out data from the same patient. This allows us to ask how much information about the activity at a given electrode might be inferred through (a) volume conductance or other sources of "leakage" from activity patterns measured from the patient's other electrodes and (b) across-electrode correlations learned from that single patient. As shown in Figure 3A (gray histogram), the distribution of within-patient correlations was centered well above zero (mean: 0.32; t-test comparing mean of distribution of z-transformed average patient correlation coefficients to 0: $t(66) = 15.16, p < 10^{-10}$). However, the across-patient correlations were substantially higher (t-test comparing average z-transformed within versus across patient electrode correlations: $t(66) = 9.62, p < 10^{-10}$). This is an especially conservative test, given that the across-patient SuperEEG reconstructions exclude (from the correlation matrix estimates) all data from the patient whose data is being reconstructed. We repeated each of these analyses on a second independent dataset and found similar results (Fig. 3B, D; within versus across reconstruction accuracy: $t(23) = 6.93, p < 10^{-5}$). We also replicated this result separately for each of the two experiments from Dataset 2 (Fig. S1). This overall finding, that reconstructions of held-out data using correlation models learned from *other* patient's data yield higher reconstruction accuracy than correlation models learned from the patient whose data is being reconstructed, has two important implications. First, it implies that distant electrodes provide additional predictive power to the data reconstructions beyond the information contained solely in nearby electrodes. (This follows from the fact that each patient's electrodes are implanted in a unique set of locations, so for any given electrode the closest electrodes in the full dataset are likely to come from the same patient.) Second, it implies that the spatial correlations learned using the SuperEEG algorithm are, to some extent, similar across people.

The recordings we analyzed from Dataset 1 comprised data collected as the patients performed a variety of (largely idiosyncratic) tasks throughout each day's recording session. That we observed reliable reconstruction accuracy across patients suggests that the spatial correlations derived from the SuperEEG algorithm are, to some extent, similar across tasks. We tested this finding more explicitly using Dataset 2. In Dataset 2, the recordings were limited to times when each patient was participating in each of two experiments (Experiment 1, a random-word list free recall task, and Experiment 2, a categorized list free recall task). We wondered whether a correlation model learned from data from one experiment might yield good predictions of data from the other experiment. Further, we wondered about the extent to which it might be beneficial or harmful to combine data across tasks.

To test the task-specificity of the SuperEEG-derived correlation models, we repeated the above within- and across-patient cross validation procedures separately for Experiment 1 and Experiment 2 data from Dataset 2. We then compared the reconstruction accuracies for held-out electrodes, for models trained within versus across the two experiments, or combining across both experiments (Fig. S2). In every case we found that across-patient models trained using data from all other patients out-performed within-patient models trained on data only from the subject contributing the given electrode (ts$(24) > 6.50$, ps$< 10^{-5}$). All reconstruction accuracies

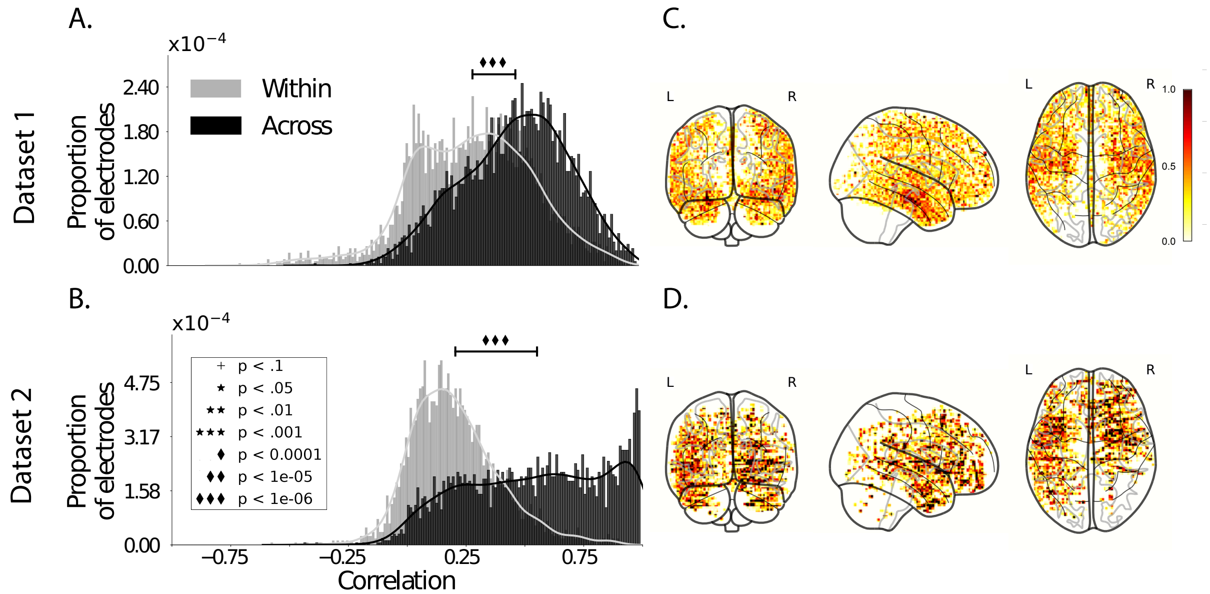

Figure 3: **Reconstruction quality across all electrodes in two ECoG datasets. A. Distributions of correlations between observed versus reconstructed activity by electrode, for Dataset 1.** The across-patient distribution (black) reflects reconstruction accuracy (correlation) using a correlation model learned from all but one patient's data, and then applied to that held-out patient's data. The within-patient distribution (gray) reflects performance using a correlation model learned from the same patient who contributed the to-be-reconstructed electrode. **B. Distributions of correlations for Dataset 2.** This panel is in the same format as Panel A, but reflects results obtained from Dataset 2. The histograms aggregate data across both Dataset 2 experiments; for results broken down by experiment see Figure S3. **C.–D. Reconstruction performance by location.** Each dot reflects the location of a single implanted electrode from Dataset 1 (Panel C) or Dataset 2 (Panel D). The dot colors denote the average across-session correlation, using the across-patient correlation model, between the observed and reconstructed activity at the given electrode location.

also reliably exceeded chance performance (ts$(24) > 8.00, p$s$< 10^{-8}$). Average reconstruction accuracy was highest for the across-patient models limited to data from the same experiment (mean accuracy: 0.68); next-highest for the across-patient models that combined data across both experiments (mean accuracy: 0.61); and lowest for models trained across task (mean accuracy: 0.47). This result also held for each of the Dataset 2 experiments individually (Fig. S3). Taken together, these results indicate that there are reliable commonalities in the spatial correlations of full-brain activity across tasks, but that there are also reliable differences in these spatial correlations across tasks. Whereas reconstruction accuracy benefits from incorporating data from other patients, reconstruction accuracy is highest when constrained to within-task data, or data that includes a variety of tasks (e.g., Dataset 1, or combining across the two Dataset 2 experiments).

Although both datasets we examined provide good full-brain coverage (when considering data from every patient; e.g. Fig. 3C, D), electrodes are not placed uniformly throughout the brain. For example, electrodes are more likely to be implanted in regions like the medial temporal lobe (MTL), and are rarely implanted in occipital cortex (Fig. 4A, B). Separately for each dataset, for each voxel in the 4 mm^3 voxel MNI152 brain, we computed the proportion of electrodes in the dataset that were contained within a 20 MNI unit radius sphere centered on that voxel. We defined the *density* at that location as this proportion. Across Datasets 1 and 2, the electrode placement densities were similar (correlation by voxel: $r = 0.56, p < 10^{-10}$). We wondered whether regions with good covererage might be associated with better reconstruction accuracy (e.g. Fig. 3C, D indicate that many electrodes in the MTL have relatively high reconstruction accuracy, and occipital electrodes tend to have relatively low reconstruction accuracy). To test whether this held more generally across the entire brain, for each dataset we computed the electrode placement density for each electrode from each patient (using the proportion of *other* patients' electrodes within 20 MNI units of the given electrode). We

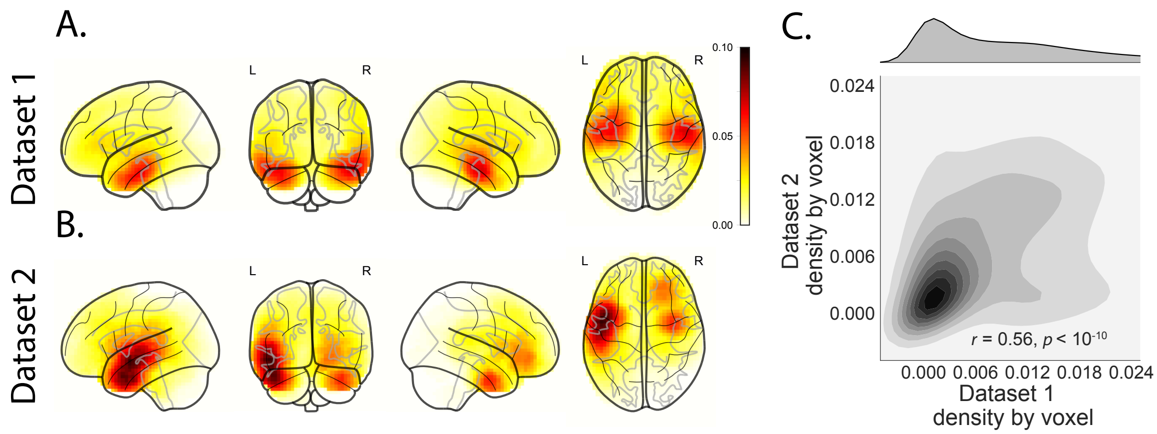

Figure 4: **Electrode sampling density by location. A. Electrode sampling density by voxel in Dataset 1.** Each voxel is colored by the proportion of total electrodes in the dataset that are located within a 20 MNI unit radius sphere centered on the given voxel. **B. Electrode sampling density by voxel in Dataset 2.** This panel displays the sampling density map for Dataset 2, in the same format as Panel A. **C. Correspondence in sampling density by voxel across Datasets 1 and 2.** The two-dimensional histogram displays the by-voxel densities in the two Datasets, and the one-dimensional histograms display the proportions of voxels in each dataset with the given density value. The correlation reported in the panel is across voxels in the 4 mm^3 MNI brain.

then correlated these density values with the across-patient reconstruction accuracies for each electrode. Contrary to our expectation, rather than positive correlations, we found weak (but reliable) negative correlations between reconstruction accuracy and density for both datasets (Dataset 1: $r = -0.07, p < 10^{-5}$; Dataset 2: $r = -0.18, p < 10^{-10}$). This indicates that the reconstruction accuracies we observed are not driven solely by sampling density, but rather may also reflect higher order properties of neural dynamics such as functional correlations between distant voxels [3].

In neurosurgical applications where one wishes to infer full-brain activity patterns, can our framework yield insights into where the electrodes should be placed? A basic assumption of our approach (and of most prior ECoG work) is that electrode recordings are most informative about the neural activity near the recording surface of the electrode. But if we consider that activity patterns throughout the brain are meaningfully correlated, are there particular implantation locations that, if present in a patient's brain, yield especially high reconstruction accuracies throughout the rest of the brain? For example, one might hypothesize that brain structures that are heavily interconnected with many other structures could be more informative about full-brain activity patterns than comparatively isolated structures.

To gain insights into whether particular electrode locations might be especially informative, we first computed the average reconstruction accuracy across all of each patient's electrodes (using the across-patients cross validation test; black histograms in Fig. 3A and B). We labeled each patient's electrodes in each dataset with the average reconstruction accuracy for that patient. In other words, we assigned every electrode from each given patient the same value, reflecting how well the activity patterns at those electrodes were reconstructed on average. Next, for each voxel in the 4 mm^3 MNI brain, we computed the average value across any electrode (from any patient) that came within 20 MNI units of that voxel's center. Effectively, we computed an *information score* for each voxel, reflecting the average reconstruction accuracy across any patients with electrodes near each voxel– where the averages were weighted to reflect patients who had more electrodes implanted near that location. This yielded a single map for each dataset, highlighting regions that are potentially promising implantation targets in terms of providing full-brain activity information via SuperEEG (Fig. 5A, B). Despite task and patient differences across the two datasets, we nonetheless found that the maps of the most promising implantation targets derived from both datasets were similar (voxelwise correlation between information scores across the two datasets: $r = 0.03, p < 10^{-10}$). While the correspondence between the two maps was not perfect, our finding that there were some commonalities between the two maps lends support to the notion that different brain areas are differently informative about full-brain activity patterns. Further, we found the intersection of the top 10% of the most informative voxels between the two datasets (Fig. S5) and found gray matter regions like the right postcentral gyrus, right supramarginal gyrus, and the right nucleus accumbens that

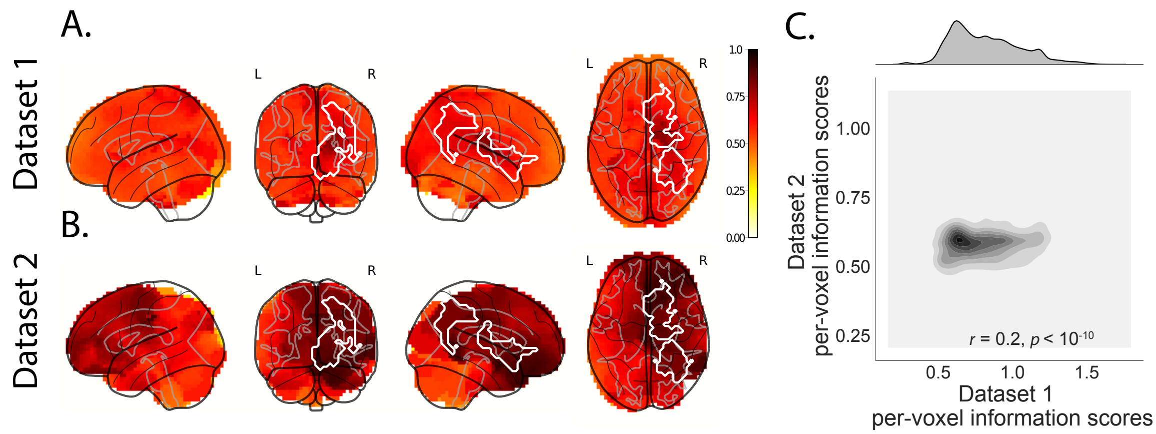

Figure 5: **Most informative electrode locations. A. Dataset 1 information score by voxel.** The voxel colors reflect the weighted average reconstruction accuracy across all electrodes from any patients with at least one electrode within 20 MNI units of the given voxel. **B. Dataset 2 information score by voxel.** This panel is in the same format as Panel A. In both panels the contours indicate the intersections between the top 10% most informative voxels in each map. **C. Correspondence in information scores by voxel across Datasets 1 and 2.** Same format as Figure 4C.

exhibited strong information scores in both datasets may be especially promising implantation targets.

Discussion

Are our brain's networks static or dynamic? And to what extent are the network properties of our brains stable across people and tasks? One body of work suggests that our brain's *functional* networks are dynamic [e.g., 24], person-specific [e.g., 9], and task-specific [e.g., 39]. In contrast, although the gross anatomical structure of our brains changes meaningfully over the course of years as our brains develop, on the timescales of typical neuroimaging experiments (i.e., hours to days) our anatomical networks are largely stable [e.g., 4]. Further, many aspects of brain anatomy, including white matter structure, are largely preserved across people [e.g., 15, 26, 37]. There are several possible means of reconciling this apparent inconsistency between dynamic person- and task-specific functional networks versus stable anatomical networks. For example, relatively small magnitude anatomical differences across people may be reflected in reliable functional connectivity differences. Along these lines, one recent study found that diffusion tensor imaging (DTI) structural data is similar across people, but may be used to predict person-specific resting state functional connectivity data [2]. Similarly, other work indicates that task-specific functional connectivity may be predicted by resting state functional connectivity data [5, 38]. Another (potentially complementary) possibility is that our functional networks are constrained by anatomy, but nevertheless exhibit (potentially rapid) task-dependent changes [e.g., 36].

Here we have taken a model-based approach to studying whether high spatiotemporal resolution activity patterns throughout the human brain may be explained by a static connectome model that is shared across people and tasks. Specifically, we trained a model to take in recordings from a subset of brain locations, and then predicted activity patterns during the same interval, but at *other* locations that were held out from the model. Our model, based on Gaussian process regression, was built on three general hypotheses about the nature of the correlational structure of neural activity (each of which we tested). First, we hypothesized that functional correlations are stable over time and across tasks. We found that, although aspects of the patients' functional correlations that were stable across tasks, we achieved better reconstruction accuracy when we trained the model on within-task data [we acknowledge that our general approach could potentially be extended to better model across-task changes, following 5, 38, and others]. Second, we hypothesized that some of the correlational structure of people's brain activity is similar across individuals. Consistent with this hypothesis, our model explained the data best when we trained the correlation model using data from *other* patients– even when compared to a correlation model trained on the same patient's data. Third, we resolved ambiguities in the data by hypothesizing that neural activity from nearby sources will tend to be similar, all else being equal. This hypothesis was supported through our finding that all of the models we trained that incorporated this spatial smoothness assumption predicted held-out data well above chance.

One potential limitation of our approach is that it does not provide a natural means of estimating the precise timing of single-neuron action potentials. Prior work has shown that gamma band and broadband activity in the LFP may be used to estimate the firing rates of neurons that underly the population contributing to the LFP [6, 14, 20, 25]. Because SuperEEG reconstructs LFPs throughout the brain, one could in principle use gamma or broadband power in the reconstructed signals to estimate the corresponding firing rates (though not the timings of individual action potentials).

Beyond providing a means of estimating ongoing activity throughout the brain using already implanted electrodes, our work also has implications for where to place the electrodes in the first place. Electrodes are typically implanted to maximize coverage of suspected epileptogenic tissue. However, our findings suggest that this approach could be further optimized. Specifically, one could leverage not only the non-invasive recordings taken during an initial monitoring period (as is currently done routinely), but also recordings collected from other patients. We could then ask: given what we learn from other patients' data (and potentially from the scalp EEG recordings of this new patient), where should we place a fixed number of electrodes to maximize our ability to map seizure foci? As shown in Figure 5, recordings from different locations are differently informative in terms of reconstructing the spatiotemporal activity patterns throughout the brain. This property might be leveraged in decisions about where to surgically implant electrodes in future patients.

Concluding remarks

Over the past several decades, neuroscientists have begun to leverage the strikingly profound mathematical structure underlying the brain's complexity to infer how our brains carry out computations to support our thoughts, actions, and physiological processes. Whereas traditional beamforming techniques rely on geometric source-localization of signals measured at the scalp, here we propose an alternative approach that leverages the rich correlational structure of two large datasets of human intracranial recordings. In doing so, we are one step closer to observing, and perhaps someday understanding, the full spatiotemporal structure of human neural activity.

Code availability

We have published an open-source toolbox implementing the SuperEEG algorithm. It may be downloaded here. Additionally, we have provided notebooks for all analyses and figures reported here.

Data availability

The dataset analyzed in this study was generously shared by Michael Kahana. A portion of Dataset 1 may be downloaded here. Dataset 2 may be downloaded here.

Acknowledgements

We are grateful for useful discussions with Luke J. Chang, Uri Hasson, Josh Jacobs, Michael Kahana, and Matthijs van der Meer. We are also grateful to Michael Kahana for generously sharing the ECoG data we analyzed in our paper, which was collected under NIMH grant MH55687 and DARPA RAM Cooperative Agreement N66001-14-2-4-032, both to MJK. Our work was also supported in part by NSF EPSCoR Award Number 1632738 and by a sub-award of DARPA RAM Cooperative Agreement N66001-14-2-4-032 to JRM. The content is solely the responsibility of the authors and does not necessarily represent the official views of our supporting organizations.

Author Contributions

J.R.M conceived and initiated the project. L.L.W.O. and A.C.H. performed the analyses. J.R.M. and L.L.W.O. wrote the manuscript.

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.R.M. (jeremy.r.manning@dartmouth.edu).

References

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

[2] Becker, C. O., Pequito, S., Pappas, G. J., Miller, M. B., adn D S Bassett, S. T. G., and Preciado, V. M. (2018). Spectral mapping of brain functional connectivity from diffusion imaging. *Scientific Reports*, 8(1411):https://www.doi.org/10.1038/s41598–017–18769–x.

[3] Betzel, R. F., Medaglia, J. D., Kahn, A. E., Soffer, J., Schonhaut, D. R., and Bassett, D. S. (2017). Inter-regional ecog correlations predicted by communication dynamics, geometry, and correlated gene expression. *arXiv*, page 1706.06088.

[4] Casey, B. J., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, 54(1–3):241–257.

[5] Cole, M. W., Ito, T., Bassett, D. S., and Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, 19(12):1718–1726.

[6] Crone, N. E., Korzeniewska, A., and Franaszczuk, P. (2011). Cortical gamma responses: searching high and low. *International Journal of Psychophysiology*, 79(1):9–15.

[7] Ezzyat, Y., Kragel, J. E., Burke, J. F., Levy, D. F., Lyalenko, A., Wanda, P., O'Sullivan, L., Hurley, K. B., Busygin, S., Pedisich, I., Sperling, M. R., Worrell, G. A., Kucewicz, M. T., Davis, K. A., Lucas, T. H., Inman, C. S., Lega, B. C., Jobst, B. C., Sheth, S. A., Zaghloul, K., Jutras, M. J., Stein, J. M., Das, S. R., Gorniak, R., Rizzuto, D. S., and Kahana, M. J. (2017). Direct brain stimulation modulates encoding states and memory performance in humans. *Current Biology*, 27:1–8.

[8] Ezzyat, Y., Wanda, P. A., Levy, D. F., Kadel, A., Aka, A., Pedisich, I., Sperling, M. R., Sharan, A. D., Lega, B. C., Burks, A., Gross, R. E., Inman, C. S., Jobst, B. C., Gorenstein, M. A., Davis, K. A., Worrell, G. A., Kucewicz, M. T., Stein, J. M., Gorniak, R., Das, S. R., Rizzuto, D. S., and Kahana, M. J. (2018). Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. *Nature Communications*, 9(365):https://www.doi.org/10.1038/s41467–017–02753–0.

[9] Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18:1664 – 1671.

[10] Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., and Collins, D. L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Medical Image Computing and Computer-Assistend Intervention*, 9:58–66.

[11] Hillebrand, A., Singh, K. D., Holliday, I. E., Furlong, P. L., and Barnes, G. R. (2005). A new approach to neuroimaging with magnetoencephalography. *Human Brain Mapping*, 25:199–211.

[12] Horak, P. C., Meisenhelter, S., Song, Y., Testorf, M. E., Kahana, M. J., Viles, W. D., Bujarski, K. A., Connolly, A. C., Robbins, A. A., Sperling, M. R., Sharan, A. D., Worrell, G. A., Miller, L. R., Gross, R. E., Davis, K. A., Roberts, D. W., Lega, B., Sheth, S. A., Zaghloul, K. A., Stein, J. M., Das, S. R., Rizzuto, D. S., and Jobst, B. C. (2017). Interictal epileptiform discharges impair word recall in multiple brain areas. *Epilepsia*, 58(3):373–380.

[13] Huster, R. J., Debener, S., Eichele, T., and Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: an introductory review. *The Journal of Neuroscience*, 32(18):6053–6060.

[14] Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V., and Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences, USA*, 107(14):6487–6482.

[15] Jahanshad, N., Kochunov, P. V., Sprooten, E., Mandl, R. C., Nichols, T. E., Almasy, L., Blangero, J., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Martin, N. G., McMahon, K. L., Medland, S. E., Mitchell, B. D., Olvera, R. L., Peterson, C. P., Starr, J. M., Sussmann, J. E., Toga, A. W., Wardlaw, J. M., Wright, M. J., Pol, H. E. H., Pastin, M. E., McIntosh, A. M., Deary, I. J., Thompson, P. M., and Glahn, D. C. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the enigma–dti working group. *NeuroImage*, 81(1):455–469.

[16] Kragel, J. E., Ezzyat, Y., Sperling, M. R., Gorniak, R., Worrell, G. A., Berry, B. M., Inman, C., Lin, J.-J., Davis, K. A., Das, S. R., Stein, J. M., Jobst, B. C., Zaghloul, K. A., Sheth, S. A., Rizzuto, D. S., and Kahana, M. J. (2017). Similar patterns of neural activity predict memory formation during encoding and retrieval. *NeuroImage*, 155:70–71.

[17] Kucewicz, M. T., Berry, B. M., Kremen, V., Brinkmann, B. H., Sperling, M. R., Jobst, B. C., Gross, R. E., Lega, B., Sheth, S. A., Stein, J. M., Das, S. R., Gorniak, R., Stead, S. M., Rizzuto,

D. S., Kahana, M. J., and Worrell, G. A. (2017). Dissecting gamma frequency actiivty during human memory processing. *Brain*, 140:1337–1350.

[18] Kucewicz, M. T., Berry, B. M., Miller, L. R., Khadjevand, F., Ezzyat, Y., Stein, J. M., Kremen, V., Brinkmann, B. H., Wanda, P., Sperling, M. R., Gorniak, R., Davis, K. A., Jobst, B. C., Gross, R. E., Lega, B., Gompel, J. V., Stead, S. M., Rizzuto, D. S., Kahana, M. J., and Worrell, G. A. (2018). Evidence for verbal memory enhancement with electrical brain stimulation in the lateral temporal cortex. *Brain*, 141(4):971–978.

[19] Lin, J.-J., Rugg, M. D., Das, S., Stein, J., Rizzuto, D. S., Kahana, M. J., and Lega, B. C. (2017). Theta band power increases in the posterior hippocampus predict successful episodic memory encoding in humans. *Hippocampus*, 27:1040–1053.

[20] Manning, J. R., Jacobs, J., Fried, I., and Kahana, M. J. (2009). Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *The Journal of Neuroscience*, 29(43):13613–13620.

[21] Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences, USA*, 108(31):12893–12897.

[22] Manning, J. R., Ranganath, R., Norman, K. A., and Blei, D. M. (2014). Topographic factor analysis: a Bayesian model for inferring brain networks from neural data. *PLoS One*, 9(5):e94914.

[23] Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., and Kahana, M. J. (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *The Journal of Neuroscience*, 32(26):8871–8878.

[24] Manning, J. R., Zhu, X., Willke, T. L., Ranganath, R., Stachenfeld, K., Hasson, U., Blei, D. M., and Norman, K. A. (2018). A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*.

[25] Miller, K. J., Shenoy, P., den Nijs, M., Sorensen, L. B., Rao, R. P. N., and Ojemann, J. G. (2008). Beyond the gamma band: the role of high-frequency features in movement classification. *IEEE Transactions on Biomedical Engineering*, 55(5):1634 – 1637.

[26] Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., Toga, A. W., Pike, G. B., Neto, P. R., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P., and Mazziotta, J. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. *NeuroImage*, 40(2):570–582.

[27] Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press.

[28] Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.*, 32(1):11–22.

[29] Sawyer, R. J. (1995). *The Terminal Experiment*. HarperPrism.

[30] Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J., and Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, 23(34):10809–10814.

[31] Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., Litt, B., Brandt, A., and Kahana, M. J. (2007a). Gamma oscillations distinguish true from false memories. *Psychological Science*, 18(11):927–932.

[32] Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., Tully, M. S., and Kahana, M. J. (2007b). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, 17(5):1190–1196.

[33] Sejnowski, T. J., Churchland, P. S., and Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, 17:1440–1441.

[34] Snyder, A. Z. (1991). Dipole source localization in the study of EP generators: a critique. *Electroencephalography and Clinical Neurophysiology*, 80(4):321–325.

[35] Solomon, E. A., Gross, R., Lega, B., Sperling, M. R., Worrell, G., Sheth, S. A., Zaghloul, K. A., Jobst, B. C., Stein, J. M., Das, S., Gorniak, R., Inman, C., Seger, S., Kragel, J. E., Rizzuto, D. S., and Kahana, M. J. (2018). Mtl functional connectivity predicts stimulation-induced theta power. *Nature Communications*, In press.

[36] Sporns, O. and Betzel, R. F. (2016). Modular brain networks. *Annual Review of Psychology*, 67:613–640.

[37] Talairach, J. and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. Verlag, Stuttgart.

[38] Tavor, I., Jones, O. P., Mars, R. B., Smith, S. M., Behrens, T. E., and Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220.

[39] Turk-Browne, N. B. (2013). Functional interactions as big data in the human brain. *Science*, 342:580–584.

[40] Weidemann, C. T., Kragel, J. E., Lega, B. C., Worrell, G. A., Sperling, M. R., Sharan, A. D., Jobst, B. C., Khadjevand, F., Davis, K. A., Wanda, P. A., Kadel, A., Rizzuto, D. S., and Kahana, M. J. (2018). Neural activity reveals interactions between episodic and semantic memory systems during retrieval. *Journal of Experimental Psychology: General*, In press.

[41] Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimieni, J. R., Fischl, B., Liu, H., and Buckner,

R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophsiology*, 106(3):1125–1165.