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 with perfect precision and at arbitrarily high spatiotemporal resolution. Although true SuperEEG is impossible using existing methods, here we present a model-based method for *inferring* 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 (all else being equal) neural activity at nearby locations will tend to be similar. 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 in other people's recordings, what would recordings most likely have looked like 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 [2]. 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) or into deep brain structures (intracranial EEG; iEEG). However, these intracranial approaches also require compromise: the high temporal and spatial resolu-

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

tions 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 [3]. 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 [4–6]. 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 [6]).

Another approach to obtaining high spatial and temporal 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 [7]). In addition, the experimental designs that are ideally suited to each technique individually are somewhat at odds. For example, fMRI experiments typically 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 [8]. 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 SuperEEG approach using two large datasets of intracranial recordings [9–22]. 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 important 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 [9–13]. Each recording session lasted from XXX–XXX hours, and includes data recorded roughly from when the patients woke up each morning, to before they went to sleep at the end of each day. In addition to typical bed-ridden hospital patient activities (e.g., lying in bed, reading, watching television, using personal electronic devices, listening to music, visiting with family and friends, etc.), the patients also performed a variety of experimental cognitive tasks throughout their day (primarily list-learning memory

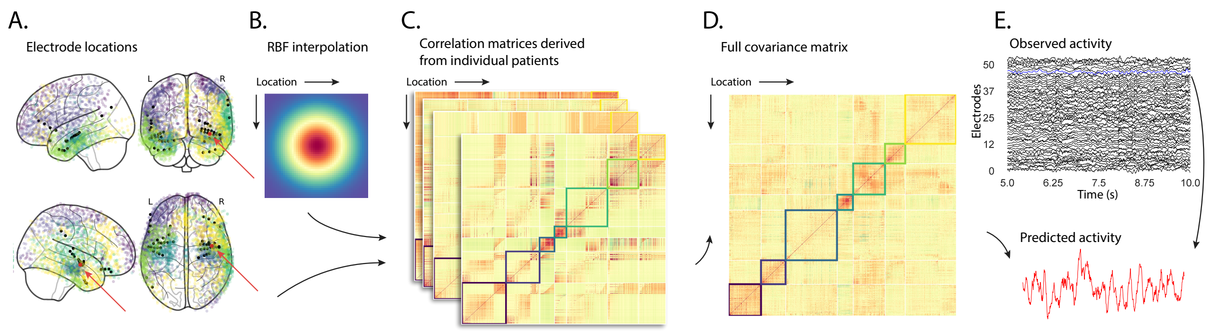

Figure 1: **Methods overview. A. Electrode locations.** Each dot reflects the location of a single electrode in dataset 1, colored according to 7 factor labels (see Panel D for details). One patient's electrode locations are highlighted in black and the to-be-reconstructed recording location is highlighted in red. **B. Radial basis function (RBF).** 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. Combined correlation matrix.** We estimate a single full-brain correlation matrix by averaging the patient-specific correlation matrices. We sort the resulting correlation matrix based on 7 factor labels obtained from k-means clustering [23]). **E. Reconstructing activity throughout the brain.** Given the observed activity from the patient's electrodes and the estimated correlation matrix (Panel D), we can compute a maximum likelihood estimate of the voltage trace at any location in $\bar{R}$. An example reconstruction (at the red dot in Panel A) is shown in red, and the actual recording at that location is highlighted above in blue.

tasks). For the purposes of the Dataset 1 analyses presented here, we aggregated all data across each recording session, ignoring the particular activities or tasks the patients were performing at any given moment. We used Dataset 1 to develop and debug 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 XXX electrodes implanted in the brains of XXX epilepsy patients [14–22]. Whereas Dataset 1 included recordings taken during a wide variety of behaviors, Dataset 2 included recordings taken as each patient performed each of two memory tasks: a random word list free recall task and a categorized word list free recall task. We used Dataset 2 to further examine the ability of SuperEEG to generalize its predictions within versus across tasks.

We first applied fourth order Butterworth notch filter to remove 60 Hz ($\pm$ .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 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 2975 electrodes from 24 patients. Each individual patient contributes electrodes from a limited set of brain locations, which we localized in a common space [MNI152; 24]; 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). Using our prior work as a guide [25, 26], we set $\lambda = 20$, although this could in theory be optimized, e.g., using cross validation or a formal model [e.g., 26].

A second mechanism whereby a given region x can contribute to the recording at η is through anatomical connections between structures near x and η. We use spatial correlations in the data to estimate these anatomical connections [27]. Let $\bar{R}$ be the set of locations at which we wish to estimate local field potentials, and let R_s 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 patients to obtain a single expected full-brain correlation matrix ($\hat{K}$; Fig. 1D):

$$\hat{K} = \mathrm{r}\left(\frac{\sum \hat{N}_s}{\sum \hat{D}_s}\right). \tag{9}$$

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 $\bar{R}$, 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 length(α_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 [8]:

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

This equation is the foundation of the SuperEEG algorithm. Whereas we observe recordings only at the locations in α_s, Equation 12 allows us to estimate the recordings at all locations in 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 α_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 directly 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, $T = 20458799$). 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 trained on all other patients' data at $\hat{K}$, 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 and 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 (red), and its associated reconstruction (blue), during a 5 s time window during one of the example patient's six recording sessions. The two traces match closely ($r = XXX, p = XXX$). Figure 2B displays a two-dimensional histogram of the actual versus reconstructed voltages for the entire 14.2 total hours of recordings from the example electrode (correlation: $r = XXX, p = XXX$). This example confirms that the SuperEEG algorithm recovers the recordings from this single electrode well.

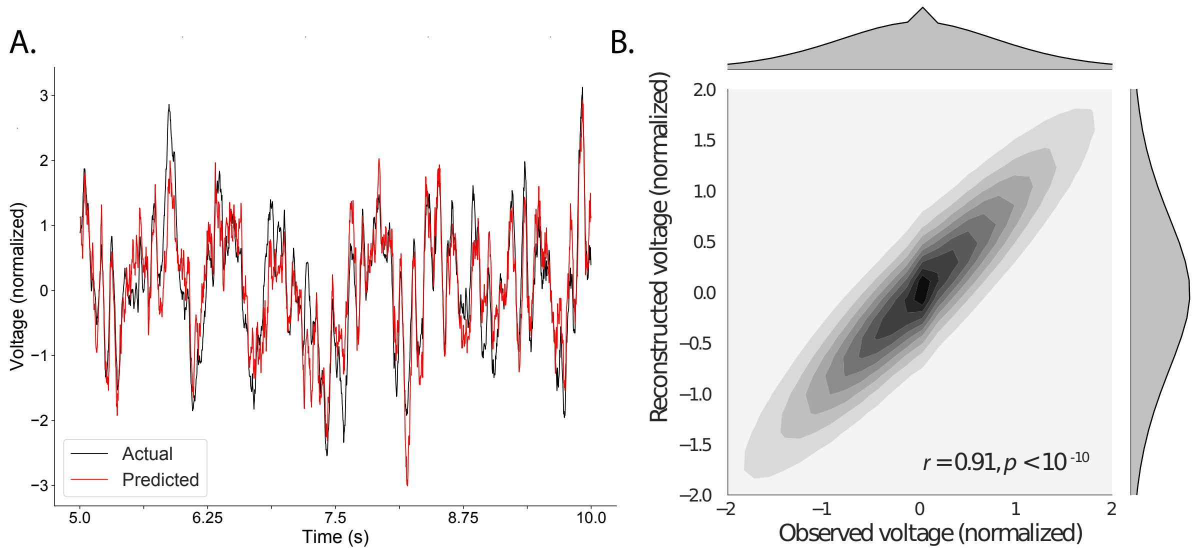

Figure 2: **Observed and reconstructed LFP from a single electrode. A. Example LFP.** A 5 s recording from the red electrode in Figure 1A is displayed in red, and the reconstructed LFP during the same time window is shown in blue. **B. Observed versus reconstructed LFP over 14.2 hours.** The two-dimensional 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).

Next, we quantify the algorithm's performance across the full dataset.

For each held-out electrode, from each held-out patient in turn, we 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: XXX; t-test comparing mean of distribution of z-transformed per-electrode correlation coefficients to 0: $t(XXX) = XXX, p = XXX$), 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 derived 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 a single patient may be used to estimate held-out data. 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: XXX; t-test comparing mean of distribution of z-transformed per-electrode correlation coefficients to 0: $t(XXX) = XXX, p = XXX$). However, the across-patient correlations were substantially higher (t-test comparing average z-transformed within versus across patient electrode correlations: $t(XXX) = XXX, p = XXX$). 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 datset and found similar results (Fig. 3B, D; within versus across reconstruction accuracy: $t(23) = 6.93, p < 10^{-5}$). Our finding, that when reconstructing held-out data from a given patient correlation models derived from *other* patient's data yield higher reconstruction accuracy than correlation models derived from that patient, has two important implications. First, it implies that distant electrodes provide additional predictive power to the data reconstructions beynd 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 derived from 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

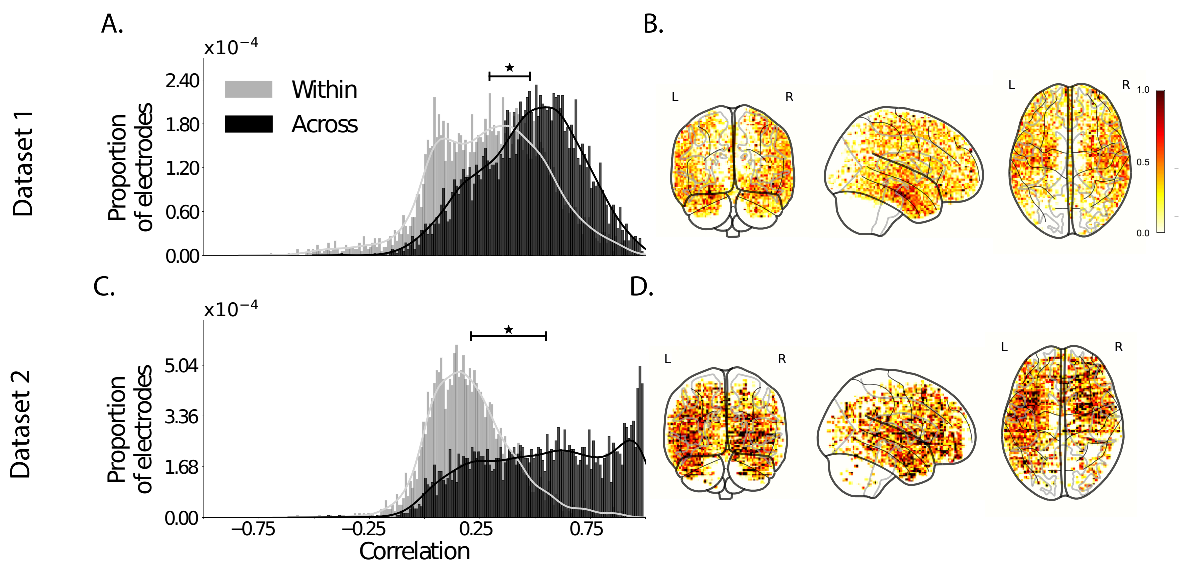

Figure 3: **Reconstruction quality across all electrodes in two ECoG datasets. A. Distributions of correlations between observed versus reconstructed activity by electrode, for Dataset 1.** The across-patient distribution (black) reflects reconstruction performance using a correlation model trained on all but one patient, and then applied to the held-out patient's data. The within-patient distribution (gray) reflects performance using a correlation model trained on 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. **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.

variety of (essentially uncontrolled) 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 trained using data only from one experiment might yield good predictions for data from the other experiment. Further, we wondered about the extent to which it might be benefitial or harmful to combine data across tasks.

JRM STOPPED HERE

We were interested in the task specific contributions to the reconstruction accuracy. Each patient in the the second dataset participated in two free recall experiments. We ran similar analyses for both experiments and found that activity was best reconstructed when limiting the training data to within task, as opposed to across task or incorporating data from both tasks (Fig. S1(mean reconstruction accuracy incorporating data within task: 0.55, across task: 0.37, all tasks: .50)). Although reconstruction accuracy in the across task analysis was still better than the volume conductance model alone (paired t-test between z-transformed mean correlation coefficients by patient: $t(47) = 5.65, p < 10^{-5}$), these results suggests that having a common tasks for patients may yield better reconstruction accuracy.

We also wondered whether reconstruction quality (measured as the correlation between the observed and reconstructed data) varied with the electrode locations (Fig. 3B & D). In general, reconstruction quality remained high throughout the brain. Although reconstruction accuracy appeared high in the medial temporal lobe, which is a common epileptic focus (and therefore a common target for electrode implantation), we observed a weak but statistically reliable negative correlation between reconstruction quality and electrode density (defined as the proportion of electrodes within 20 MNI units for each location; dataset 1: $r = -0.07, p < 10^{-5}$, dataset 2: $r = -0.16, p < 10^{-10}$). This provides some evidence that our reconstruction accuracy results cannot be driven only by volume conductance. Qualitatively, it

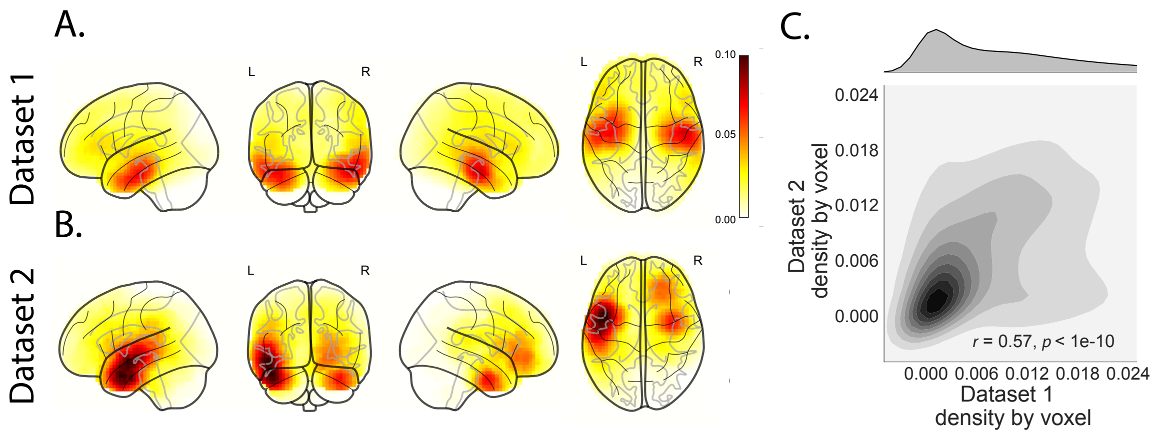

Figure 4: **Sampling density and reconstruction quality. A. & B.** The glass brain maps show sampling density by voxel location for dataset 1 and dataset 2. **C.** Correlation of sampling density by voxel location for dataset 1 vs. dataset 2.

appeared that the distribution of electrodes was similar across the datasets, suggesting potential commonalities of target locations across patients and similarities in surgical decisions. Indeed, we found a relatively strong correlation between the electrode densities within the two datasets (defined as the proportion of electrodes within 20 MNI units for each 34686 voxels (Fig. 4A, B); $r = 0.57, p < 10^{-10}$).

In addition to exploring how reconstruction quality varies with location, we also wondered whether there might be effects of electrode placements on reconstruction quality. For example, are there particular implantation locations that yield especially high reconstruction accuracies at other locations throughout the brain? To gain insights into this questions, we computed the average reconstruction correlation for each patient, then computed the average patient reconstruction correlation for any patients who had electrodes within a 20 MNI unit diameter sphere centered on each voxel location. The resulting maps highlight the locations of implanted electrodes from patients whose reconstructions were especially accurate (Fig. 5A and B). We found that the most informative locations were consistent across datasets which lends support to the notion that different electrode location are more informative about activity across patients (Fig. 5C); $r = 0.22, p < 10^{-10}$). The locations in dark red might therefore be good candidate implantation targets for neurosurgeons and neurologists who wish to use SuperEEG to reconstruct

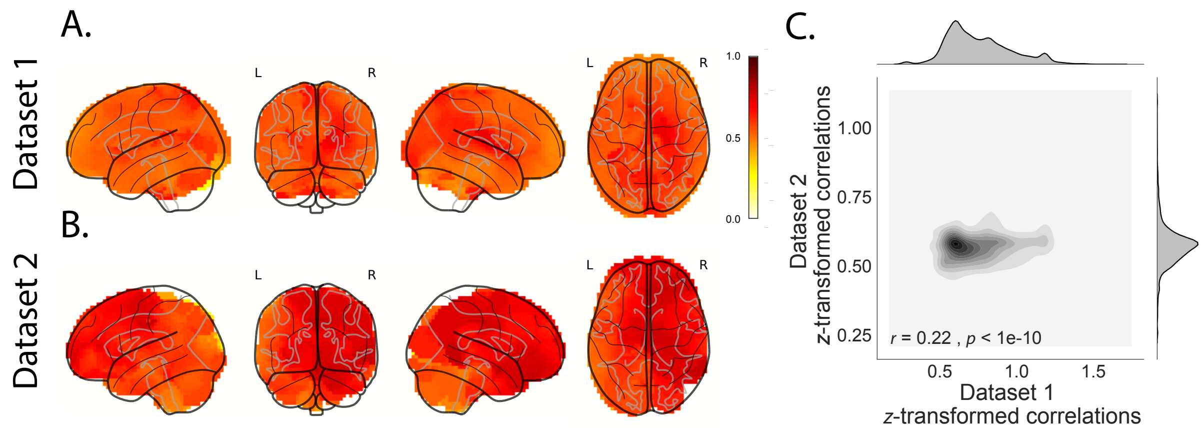

Figure 5: **Most informative electrode locations. A. & B.** The glass brain maps displays the average reconstruction correlations (by patient, across all electrodes) for patients with electrodes within a 20 MNI unit diameter sphere centered on each location for dataset 1 and dataset 2. **C.** Correlation between z-transformed correlations by voxel for dataset 1 vs. dataset 2.

full-brain electrophysiological signals. The above findings, that one can infer brain activity throughout a person's brain using recordings from a limited number of locations from that person's brain in conjunction with recordings from other people's brains, have deep implications for the structure of brain data. The first implication is that the correlational structure of different people's brain data is largely preserved across individuals. Despite recent evidence that different people have stable but reliably different resting state connectome [28], our results suggest that the correlational structure of different people's brain data is preserved enough across individuals to provide meaningful information.

Discussion

SuperEEG infers full-brain activity patterns by leveraging correlations in those patterns of brain activity within and across people. Although the approach may, in principle, be used to infer brain activity *anywhere* in the brain, the inferences perform slightly better for regions with dense electrode sampling across patients. (Taken to the logical extreme, we could not hope to accurately recover activity patterns from brain areas where no recordings existed from any patient.) As more data are included in the inference procedure, this suggests that reconstruction accuracy should improve.

A fundamental assumption of the SuperEEG algorithm is that the data covariance matrix is stable over time and across people. This is a useful simplification. However, a growing body of evidence from the fMRI community suggests that the data covariance matrix changes in meaningful ways over time (for example, the data covariance matrix changes from moment-to-moment during story listening, serving as a unique "fingerprint" for each moment of the story; further, these task-driven timepoint-specific covariance fingerprints appear to be largely preserved across people [29, 30]). These findings indicate that the full-brain covariance matrix is not stable over time. Other recent work has shown that people's resting state connectivity matrices may be used to uniquely identify individuals and predict fluid intelligence scores [28]. This indicates that the full-brain covariance matrix is not stable across people. If the fundamental stability assumptions that SuperEEG relies on are violated, how can the SuperEEG algorithm still accurately recover LFP data? It is important to recognize that the fact that variability (over time or across people) is predictive (e.g., of cognitive states during story listening or fluid intelligence scores) does not necessarily mean that this variability is large in magnitude. Rather, we have long known that brain structure is tightly preserved across individuals (and over time, at least on the timescale of typical clinical and experimental recording sessions), and any functional changes must occur within the framework of the underlying structural anatomy. Nevertheless, one could imagine future improvements to the SuperEEG approach that leverage resting state fMRI or structural data [e.g., diffusion tensor imaging (DTI)] to estimate Bayesian priors over the correlation matrices inferred, in the current framing, using only ECoG data. Further, relaxing the assumption that the covariance matrix is stable (over time and/or across people), and/or incorporating more detailed brain conductance models (e.g., informed by structural MRI scans) may improve the predictive performance of the approach.

One potential limitation of the SuperEEG approach is that the above assumption of covariance stability across people may be violated even more if different patients are performing different cognitive tasks. To understand of the extent to which the current findings generalize across cognitive tasks, we replicated our initial findings using a dataset in which patients participated in two tasks, and limited the training data to either within task, across task, or using both tasks. Since we found the most accurate reconstructions using task-specific data, this would suggest building up new databases for estimating each task-specific covariance matrix. Or, using a more sophisticated approach, one could create a hierarchical model whereby each task-specific covariance matrix was modeled as a perturbation of a "global" task-unspecific covariance matrix (which could in turn be informed by fMRI or DTI data).

A second potential limitation of the SuperEEG 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 [31]. 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), but also recordings collected from other patients. We could then ask: given everything we know about the other patients and from the scalp 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 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 a large dataset 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 released an open-source SuperEEG Python toolbox. All of the code used in this manuscript is on GitHub, and the code may be shared using a GitHub account accessible to the reviewers upon request.

Data availability

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

Acknowledgements

We are grateful for useful discussions with Luke J. Chang and Matthijs van der Meer. We are also grateful to Michael J. Kahana for generously sharing the ECoG dataset we analyzed in our paper, which was collected under NIMH grant MH55687 to MJK. Our work was also supported in part by NSF EPSCoR Award Number 1632738. 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 and Notes

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

[2] T. J. Sejnowski, P. S. Churchland, J. A. Movshon, *Nature Neuroscience* **17**, 1440 (2014).

[3] J. Sarvas, *Phys. Med. Biol.* **32**, 11 (1987).

[4] A. Z. Snyder, *Electroencephalography and Clinical Neurophysiology* **80**, 321 (1991).

[5] S. Baillet, J. C. Mosher, R. M. Leahy, *IEEE Signal Processing Magazine* **18**, 14 (2001).

[6] A. Hillebrand, K. D. Singh, I. E. Holliday, P. L. Furlong, G. R. Barnes, *Human Brain Mapping* **25**, 199 (2005).

[7] R. J. Huster, S. Debener, T. Eichele, C. S. Herrmann, *The Journal of Neuroscience* **32**, 6053 (2012).

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

[9] P. B. Sederberg, M. J. Kahana, M. W. Howard, E. J. Donner, J. R. Madsen, *Journal of Neuroscience* **23**, 10809 (2003).

[10] P. B. Sederberg, *et al.*, *Cerebral Cortex* **17**, 1190 (2007).

[11] P. B. Sederberg, *et al.*, *Psychological Science* **18**, 927 (2007).

[12] J. R. Manning, S. M. Polyn, G. Baltuch, B. Litt, M. J. Kahana, *Proceedings of the National Academy of Sciences, USA* **108**, 12893 (2011).

[13] J. R. Manning, M. R. Sperling, A. Sharan, E. A. Rosenberg, M. J. Kahana, *The Journal of Neuroscience* **32**, 8871 (2012).

[14] Y. Ezzyat, *et al.*, *Current Biology* **27**, 1 (2017).

[15] P. C. Horak, *et al.*, *Epilepsia* **58**, 373 (2017).

[16] J. E. Kragel, *et al.*, *NeuroImage* **155**, 70 (2017).

[17] M. T. Kucewicz, *et al.*, *Brain* **140**, 1337 (2017).

[18] J.-J. Lin, *et al.*, *Hippocampus* **27**, 1040 (2017).

[19] E. A. Solomon, *et al.*, *Nature Communications* **In press** (2018).

[20] C. T. Weidemann, *et al.*, *Journal of Experimental Psychology: General* **In press** (2018).

[21] Y. Ezzyat, *et al.*, *Nature Communications* **9**, https://www.doi.org/10.1038/s41467 (2018).

[22] M. T. Kucewicz, *et al.*, *Brain* **141**, 971 (2018).

[23] B. T. T. Yeo, *et al.*, *Journal of Neurophsiology* **106**, 1125 (2011).

[24] G. Grabner, *et al.*, *Medical Image Computing and Computer-Assistend Intervention* **9**, 58 (2006).

[25] J. R. Manning, R. Ranganath, K. A. Norman, D. M. Blei, *PLoS One* **9**, e94914 (2014).

[26] J. R. Manning, *et al.*, *NeuroImage* (2018).

[27] C. O. Becker, *et al.*, *Scientific Reports* **8**, https://www.doi.org/10.1038/s41598 (2018).

[28] E. S. Finn, *et al.*, *Nature Neuroscience* **18**, 1664 (2015).

[29] E. Simony, C. J. Honey, J. Chen, U. Hasson, *Nature Communications* **7**, 1 (2016).

[30] J. R. Manning, *et al.*, *bioRxiv* p. 106690 (2017).

[31] J. R. Manning, J. Jacobs, I. Fried, M. J. Kahana, *The Journal of Neuroscience* **29**, 13613 (2009).