

Spectral Encoding of Seen and Attended Object Categories in the Human Brain

Zahraa Sabra,¹ Leonardo Bonilha,² and Thomas Naselaris¹

Departments of ¹Neuroscience, and ²Neurology, Medical University of South Carolina, Charleston, South Carolina 29425

Local field potentials (LFPs) encode visual information via variations in power at many frequencies. These variations are complex and depend on stimulus and cognitive state in ways that have yet to be fully characterized. Specifically, the frequencies (or combinations of frequencies) that most robustly encode specific types of visual information are not fully known. To address this knowledge gap, we used intracranial EEG to record LFPs at 858 widely distributed recording sites as human subjects (six males, five females) indicated whether briefly presented natural scenes depicted one of three attended object categories. Principal component analysis applied to power spectra of the LFPs near stimulus onset revealed a broadband component (1–100 Hz) and two narrowband components (1–8 and 8–30 Hz, respectively) that encoded information about both seen and attended categories. Interestingly, we found that seen and attended categories were not encoded with the same fidelity by these distinct spectral components. Model-based tuning and decoding analyses revealed that power variations along the broadband component were most sharply tuned and offered more accurate decoding for seen than for attended categories. Power along the narrowband delta–theta (1–8 Hz) component robustly decoded information about both seen and attended categories, while the alpha–beta (8–30 Hz) component was specialized for attention. We conclude that, when viewing natural scenes, information about the seen category is encoded via broadband and sub-gamma (<30 Hz) power variations, while the attended category is most robustly encoded in the sub-gamma range. More generally, these results suggest that power variation along different spectral components can encode qualitatively different kinds of visual information.

Key words: computational modeling; intracranial EEG; local field potentials; spectral patterns; vision; visual attention

Significance Statement

In this article, we characterize how changes in visual stimuli depicting specific objects (cars, faces, and buildings) and changes in attention to those objects affect the frequency content of local field potentials in the human brain. In contrast to many previous studies that have investigated encoding by variations in power at high (>30 Hz) frequencies, we find that the most important variation patterns are broadband (i.e., distributed across multiple frequencies) and narrowband, but in lower frequencies (<30 Hz). Interestingly, we find that seen and attended categories are not encoded with the same fidelity by these distinct spectral encoding patterns, suggesting that power at different frequencies can encode qualitatively different kinds of information.

Introduction

A long-standing challenge in systems neuroscience is to understand how variations in sensory and cognitive states are encoded in the power spectrum of local field potentials (LFPs). In visual neuroscience, encoding of visual stimuli is often studied by examining frequencies above 30 Hz exclusively—the so-called

“gamma” range. There is ample prior work to justify why the gamma band is the default frequency range of interest. For example, many studies have reported a robust and narrowly confined (40–60 Hz) increase in power in the gamma range in response to visual stimuli (Fries et al., 2001, 2008; Fries, 2009; Brunet et al., 2014). It has been argued that this gamma oscillation is required for the perception of both objects and natural scenes via the visual cortex (Tallon-Baudry et al., 1997; Hoogenboom et al., 2006; Fries et al., 2008; Brunet et al., 2014).

Recent studies have questioned the functional primacy of gamma oscillation responses in vision (Ray and Maunsell, 2010; Jia et al., 2013; Hermes et al., 2015). For example, Hermes et al. (2015) showed that while a robust narrowband gamma oscillation is induced by some types of visual stimuli, a narrowband gamma in response to pictures of faces or buildings was largely

Received April 20, 2019; revised Sept. 24, 2019; accepted Sept. 25, 2019.

Author contributions: Z.S., L.B., and T.N. designed research; Z.S. performed research; Z.S. and T.N. analyzed data; Z.S. wrote the paper.

This research was supported by National Eye Institute Grant R01-EY-023384. We thank the Medical University of South Carolina Comprehensive Epilepsy Center for support in the execution of this study.

The authors declare no competing financial interests.

Correspondence should be addressed to Thomas Naselaris at tnaselar@musc.edu.

<https://doi.org/10.1523/JNEUROSCI.0900-19.2019>

Copyright © 2020 the authors

absent for recording sites in the fusiform gyrus or parahippocampal gyrus—locales that are, respectively, highly responsive to these object categories (Kanwisher et al., 1997; Chao et al., 1999; Haxby et al., 2001). Rather, these authors noted stimulus-evoked power variations over a wide range of frequencies—termed the “broadband response”—in these high-level visual areas. Thus, the particular frequency bands that are most relevant for encoding information about object category is an unsettled question whose answer depends on brain region and the kind of stimuli presented.

Another unsettled question about the encoding of visual information in LFP spectra relates to visual attention. It is known that attention can modulate action potential firing rate (Moran and Desimone, 1985) and blood oxygenation level-dependent (BOLD; Çukur et al., 2013) responses to visual stimuli. However, there is some evidence from LFP studies that information about what is attended and what is seen may be encoded in different frequency bands (Bastos et al., 2015; Jensen et al., 2015; Michalareas et al., 2016). Indeed, there is accumulating evidence that top–down (e.g., visual attention) signals generally are encoded in the sub-gamma range (<30 Hz; Klimesch, 1999; Klimesch et al., 2011; Jensen et al., 2015; Michalareas et al., 2016; Helfrich et al., 2017). Thus, the frequency bands most relevant to the encoding of visual information may depend strongly on how attention varies during viewing.

In this work, we identify patterns of variation in spectral power that are most relevant to encoding natural scenes during sustained attention to a specific object category by analyzing LFP recordings using intracranial EEG in humans. We address key questions related to spectral encoding in this context, namely: which frequency bands are subject to the most power variation upon varying the category of attended/seen objects? How is information about attended and seen object categories encoded in patterns of variation in spectral power?

To address these questions, we prepared peristimulus power spectral density (PSD) functions for LFP signals recorded at sites on and beneath the surface of cortex. We decomposed the variation in PSDs across seen and attended stimulus categories using principal component analysis (PCA). The major principal components (PCs) revealed the combinations of frequency bands that explain a substantial fraction of the variation in spectral power across different seen/attended object categories and across recording sites. We then analyzed the variation in power along each PC at each site by building simple encoding models based upon seen/attended object category. We then used the encoding models to determine the spatial distribution of recording sites that encode seen/attended categories and the sharpness of tuning along each PC. Finally, we used the encoding models to perform model-based decoding to understand how the encoding of information about seen/attended stimulus category is distributed across the frequency spectrum.

Materials and Methods

Participants

Eleven people with epilepsy (6 males, 5 females) between 14 and 56 years old participated in the experiment. Before the experiment, each participant underwent a surgical procedure for the implantation of depth electrodes for the detection of seizures within cortical and subcortical structures. The location of the implanted electrodes was determined based solely on clinical considerations. All participants provided written consent approved by the Institutional Review Board at the Medical University of South Carolina. One participant (S2) had an electrode grid (a transparent numbered mesh of electrodes that is placed subdurally on the

Table 1. List of participants and electrodes location

| Participant ID | Electrodes location | Sex | Age (years) | Recording sites (n) |
|----------------|--|--------|-------------|---------------------|
| 1 | Bilateral temporal | Male | 44 | 28 |
| 2 | Left; grid: frontal and temporal; Depth: frontal | Female | 36 | 114 |
| 3 | Bilateral frontal, temporal, insular | Female | 33 | 115 |
| 4 | Bilateral temporal, occipital; right frontal | Male | 19 | 68 |
| 5 | Bilateral frontal, temporal, insular | Female | 45 | 132 |
| 6 | Right: temporal, occipital; left: frontal, temporal, insular | Male | 30 | 109 |
| 7 | Bilateral occipital; left temporal | Male | 32 | 45 |
| 8 | Left temporal | Female | 41 | 18 |
| 9 | Bilateral frontal, temporal, insular | Male | 14 | 96 |
| 10 | Left: frontal, temporal, insular | Male | 21 | 81 |
| 11 | Bilateral temporal | Female | 56 | 52 |

The table shows the distribution of 858 recording sites across participants and brain regions. Unless otherwise specified to be grids, electrodes are depth electrodes. Note that the nomenclature of electrodes is broadly construed: a “frontal” electrode goes from a deep brain region within frontal lobe up to frontal cortex. A “temporal” electrode goes from amygdala or hippocampus up to temporal cortex. “Insular” electrodes go from either anterior insula to frontal cortex or posterior insula to parietal cortex. Occipital electrodes go from medial occipital regions to occipital cortex and sometimes to parietal cortex.

brain surface) over the lateral surface of the left hemisphere in addition to depth electrodes.

Electrode features, placement, and location mapping

The location of subdural and depth electrodes (Ad-Tech) varied among participants (Table 1). The placement of electrodes was solely guided by the clinical evaluation of each participant. Each depth electrode has 10 recording sites separated along the electrode by 5 mm. Each electrode has a diameter of 2.29 mm. Participant S2 had an additional grid of 8 × 8 recording sites with 10 mm spacing between adjacent recording sites.

We used T1-weighted structural MR images taken for each participant postimplantation to determine the anatomical location of recording sites. First, electrodes were masked in the structural images using “cost function masking” (Brett et al., 2001) in MRICron (RRID:SCR_002403). Structural images for each participant were then normalized to MNI space using the Clinical Toolbox (Rorden et al., 2012) within SPM8 (RRID:SCR_007037; Clinical Toolbox for SPM (RRID:SCR_014096)). Mapping of individual recording sites on a template-normalized brain (ICBM152 brain) was visualized using BrainNet Viewer (RRID:SCR_009446). After excluding recording sites that were clinically determined to be a source of seizure (see details below), the 858 recording sites were distributed as follows: frontal (7 participants, 253 in left hemisphere, 68 in right hemisphere); temporal lobe, including medial temporal lobe (10 participants, 218 in left hemisphere, 148 in right hemisphere); insula (4 participants, 8 in left hemisphere, 20 in right hemisphere); occipital lobe (3 participants, 13 in left hemisphere, 33 in right hemisphere); parietal lobe (7 participants, 39 in left hemisphere, 28 in right hemisphere); and basal ganglia (4 participants, 23 in left hemisphere, 7 in right hemisphere).

Recording apparatus

LFP data were recorded using an XLTEK EEG system (Natus Medical). Sampling frequency varied across participants between 1 and 2 kHz. Recordings collected with a 2 kHz sampling rate were downsampled to 1 kHz for consistency of analysis across all 11 participants.

Experimental procedures

Participants sat upright and viewed a laptop screen during the course of the experiment. The screen was positioned at eye level at a distance from the participants of ~40 cm. Participants were encouraged to pause and take a break at any point during the experiment if they felt fatigued or distracted.

Synching stimulus and LFP time series

To align the time of presentation of each experimental stimulus with the recorded brain signals a photodiode was attached to the bottom right corner of the screen. Each stimulus frame displayed a small white box at

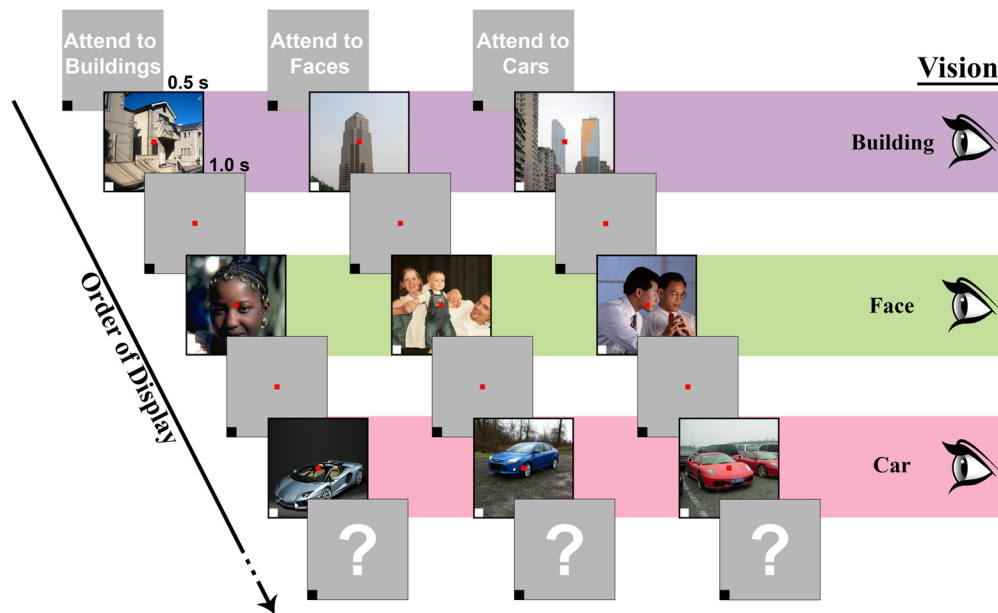


Figure 1. Overview of the experimental design. Participants viewed color photographs depicting either a car, building, or face. Photographs were presented for 0.5 s followed by a 1.0 s gray-screen ISI. For each participant, the experiment was partitioned into blocks of 30–50 photographs. At the beginning of each block, participants were cued to attend to one object category in the following order: buildings, faces, and cars. Photographs did not repeat across or within blocks. At random intervals, participants were presented with a question mark after the ISI and asked to verbally indicate whether the last displayed photograph belonged to the attended object category. This design allowed the study of nine combinations of seen and attended object categories (e.g., “see building, attend car”). In the illustration above, columns indicate series of photographs presented within one block. The attended object category is indicated at the top of each column. Photographs were selected to illustrate each of the nine possible combinations of seen (rows) and attended (columns) object categories.

the location of the photodiode at stimulus onset, and a small black box at stimulus offset. Leads from the photodiode were inserted into the same bedside preamplifier used to record brain signals. The photodiode time series (a series of on–off pulses) was then used to synchronize the brain signals with stimulus onsets and offsets off-line (Rorden and Hanayik, 2014).

Experimental design

Participants fixated on a dot at the center of the laptop screen throughout each experiment. Each experiment consisted of a stream of natural scenes that prominently displayed a human face, a building, or a car. Each image was displayed for 0.5 s, followed by an interstimulus interval (ISI) of 1000 ms. During ISIs, a blank screen (luminance set to average luminance of all images) was displayed; participants maintained fixation throughout each ISI.

Images were displayed in 9–15 blocks of 30–50 images each. All images were unique, and none of the images was repeatedly displayed. A total of 450 images were displayed. At the start of each block, participants were instructed to attend to the “face” images, the “car” images, or the “building” images. The attended category was fixed during each block; it alternated across blocks in the following order: building, face, car (Fig. 1). After a random number of selected ISIs during each block, a large question mark was displayed (an average of five question marks was displayed per block). Participants were instructed to then verbally indicate whether the last seen image was of the same category as the attended one. The experiment provided verbal feedback on their response to indicate correct and incorrect trials. All participants gave correct responses on at least 90% of the trials.

Given that all images belonged to one of three attended object categories, the experiment was composed of nine distinct task conditions, each corresponding to a distinct (seen category, attended category) pair. These task conditions are illustrated in Figure 1.

Signal screening and preprocessing

Before analyzing brain signals, we screened all recordings to identify and exclude recording sites that were reported by the patient’s clinical neurophysiologists to be associated with ictal or significant interictal epileptiform activity. We also excluded recording sites located above the surface of the brain, as identified by inspecting the postoperative structural images. A total of 286 of 1144 recording sites were excluded in this way.

After screening, all signals were referenced to a global mean. That is, for each participant independently a global average time series was computed and subtracted from the time series of each individual recording site.

Peristimulus power spectral density functions

For each recording site, the LFP, starting from stimulus onset and ending 0.5 s after (i.e., the total duration of each stimulus display), was extracted and used to estimate a PSD function (Miller et al., 2014, 2016). PSD frequencies ranged from 1 Hz (DC offset was excluded) to 100 Hz with an increment of 1 Hz; frequencies between 56 and 63 Hz were excluded to compensate for a 60 Hz line artifact. Each of these single-trial PSDs were normalized by calculating the logarithmic value of dividing single-trial PSDs over the mean power of all single-trial PSDs ($\log[\text{PSD}/\langle \text{PSD} \rangle]$), where $\langle \cdot \rangle$ indicates an average across all trials for the given recording site) as in the study by Miller et al. (2016). Single-trial PSDs for each recording site were then grouped according to task condition (i.e., according to the [seen category, attended category] pair associated with each trial) and averaged to construct a condition-specific average PSD. For each recording site, this process resulted in nine condition-specific average PSD functions. A baseline PSD was also calculated by averaging across PSDs prepared from the first 0.5 s of signal during each ISI. Thus, a total of 8580 condition-specific PSDs were calculated from 858 recording sites across 11 participants.

Principal component analysis

The 8580 PSDs were treated as independent 93-dimensional (1 dimension for each of the frequencies in the PSD) observation vectors in a PCA. This resulted in 93 orthogonal PCs. Each PC was itself a 93-dimensional PSD function, with each value indicating power at each of the studied 93 frequency points. In this work, we study the top three components, which accounted for ~40% of the variance.

PC-specific encoding models

We investigated how variations in spectral power along each PC encoded information about the variation in task conditions by constructing PC-specific encoding models. Encoding models were constructed independently for each recording site and participant. To construct models for participant k , we first applied PCA to the PSDs from the other 10

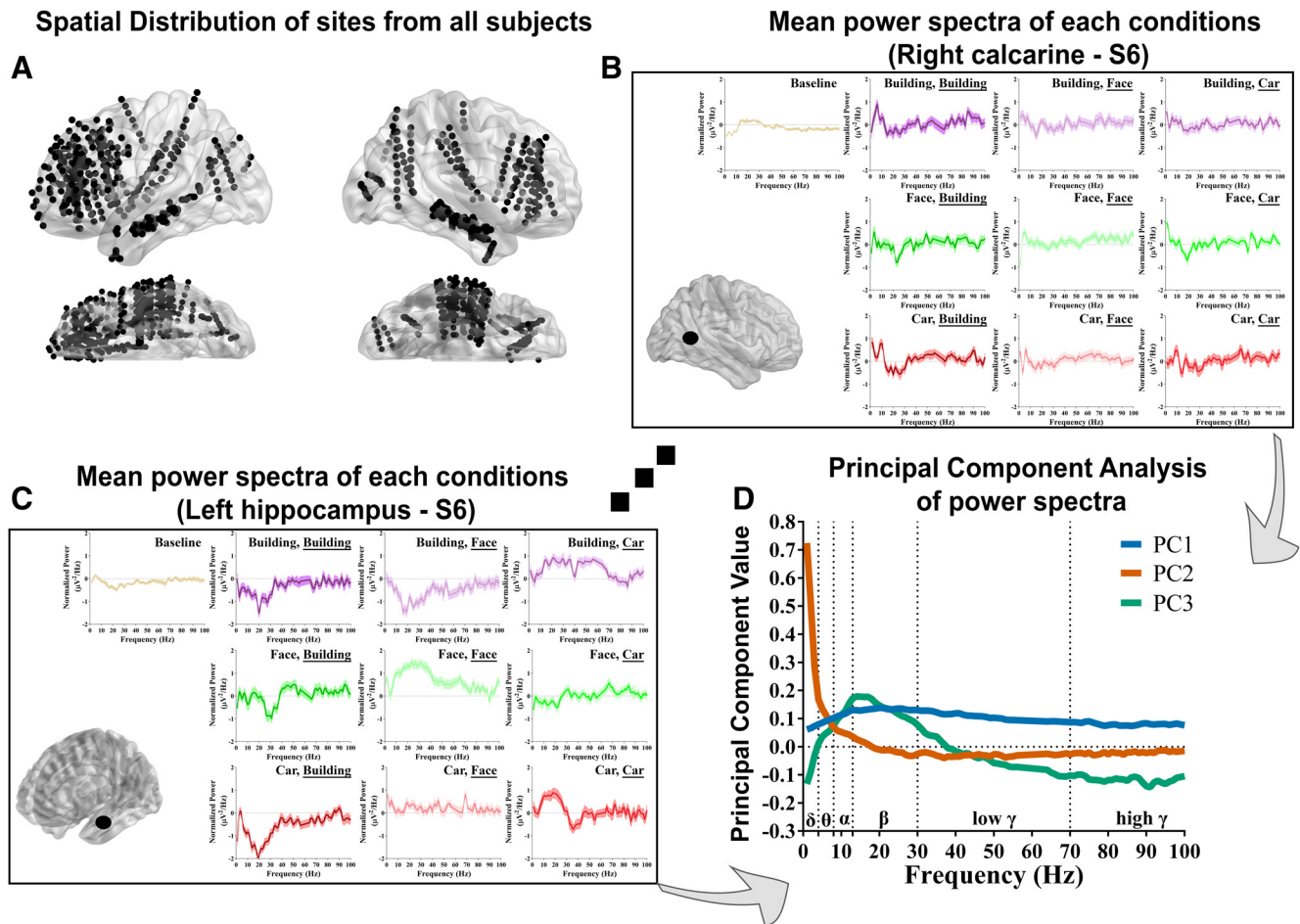


Figure 2. PCA of peristimulus PSD functions. **A**, Brain surface reconstructions illustrating the location of all recording sites from 11 participants. Each black dot indicates the location of one recording site, plotted in MNI space. L corresponds to the left hemisphere and R to the right hemisphere. Table 1 lists the distribution of electrode locations across participants. **B, C**, Peristimulus PSDs from a single recording site in the right calcarine cortex (**B**) and left hippocampus (**C**) from participant 6 (S6). Each subplot shows the average (shading indicates SEM) normalized, log-transformed PSD corresponding to each combination of seen and attended object categories (attended object category is underlined in each subplot) and for the ISI also referred to as baseline. These condition-specific averages were performed over all single-trial PSDs for the specified combination of (seen, attended) object category (or for the ISI). Single-trial PSDs are derived from the LFP beginning at the onset of stimulus for a duration of 0.5 s. A total of 8580 condition-specific PSDs were calculated from 858 recording sites across 11 participants. **D**, Results of PCA applied to all condition-specific PSDs. The top three PCs are shown (labeled PC1, PC2, and PC3). PC1 characterizes broadband variation in spectral power; PC2 and PC3 are selective for power variation in narrow, low-frequency bands. Figure 3 shows the results of PCA when applied to individual participants separately. Figure 4 shows the correspondence between PC2 and ERP.

participants. We then projected the single-trial PSDs for participant k onto the first, second, and third PCs resulting from this participant-specific PCA. Thus, for each recording site we have 450 projections (one per experimental stimulus) per PC describing the variation in spectral power along the PC. These projections are the “responses” that are predicted by each PC-specific encoding model. Formally, for a single recording site and participant, let \mathbf{q}_i be the PSD calculated for trial i and let \mathbf{p}_j be the j th principal component. Then $r_{ij} = \mathbf{q}_i \cdot \mathbf{p}_j$ is the single-trial response for PC_j that is predicted by each encoding model. The encoding model for a single recording site in participant k is specified as a simple linear combination of the following task conditions:

$$r_j = S\mathbf{w}^T + b + \varepsilon$$

where r_j is a $(N_{\text{train}} \times 1)$ vector of responses along PC_j , S is the $(N_{\text{train}} \times 9)$ design matrix of the experiment (each column indicates a task condition), \mathbf{w} is the (1×9) vector of condition weights (i.e., regression parameters), b is an intercept term, and ε is zero-mean Gaussian noise. N_{train} indicates the number of training samples.

The values of the encoding model weights \mathbf{w} and intercept b were determined for each recording site via simple linear regression that predicts the projection of each condition to be equal to the average projection of the normalized spectra of the trained trials for the studied condition. Regressions were performed by applying a 10-fold leave- k -out

procedure to a subset of 360 trials that was used only for training and evaluating encoding models (the remaining $N_{\text{val}} = 90$ trials were used for decoding analyses, as described below). This means that for each recording site, 10 separate regressions were performed. Each regression used a different set of $N_{\text{train}} = 288$ trials that were randomly sampled from the 360 trials used to construct encoding models. On each of these 10 folds, the regression model was evaluated on the remaining $N_{\text{test}} = 72$ trials. We used Pearson correlation between model predictions and the observed responses r_j . The “prediction accuracy” that we report for each encoding model is the average Pearson correlation across all 10 folds. Here an encoding model is considered “accurate” if the Pearson correlation exceeds a significance threshold ($p < 0.05$) determined by a random permutation test (5000 permutations of prediction and observed responses across trials). The encoding model weights that we discuss and analyze below are the average of the weights across all 10 folds.

Tuning width of encoding model

For analysis of the tuning width of the encoding model, weights for each recording site were normalized to span a (0.0, 1.0) range and then were ranked by magnitude. Tuning width is the rank of the lowest-ranking weight that has a value greater than the middle of the range (0.5). A “sharply tuned” encoding model would have a tuning width of 9, indicating only a single weight exceeds the middle of the (normalized) range

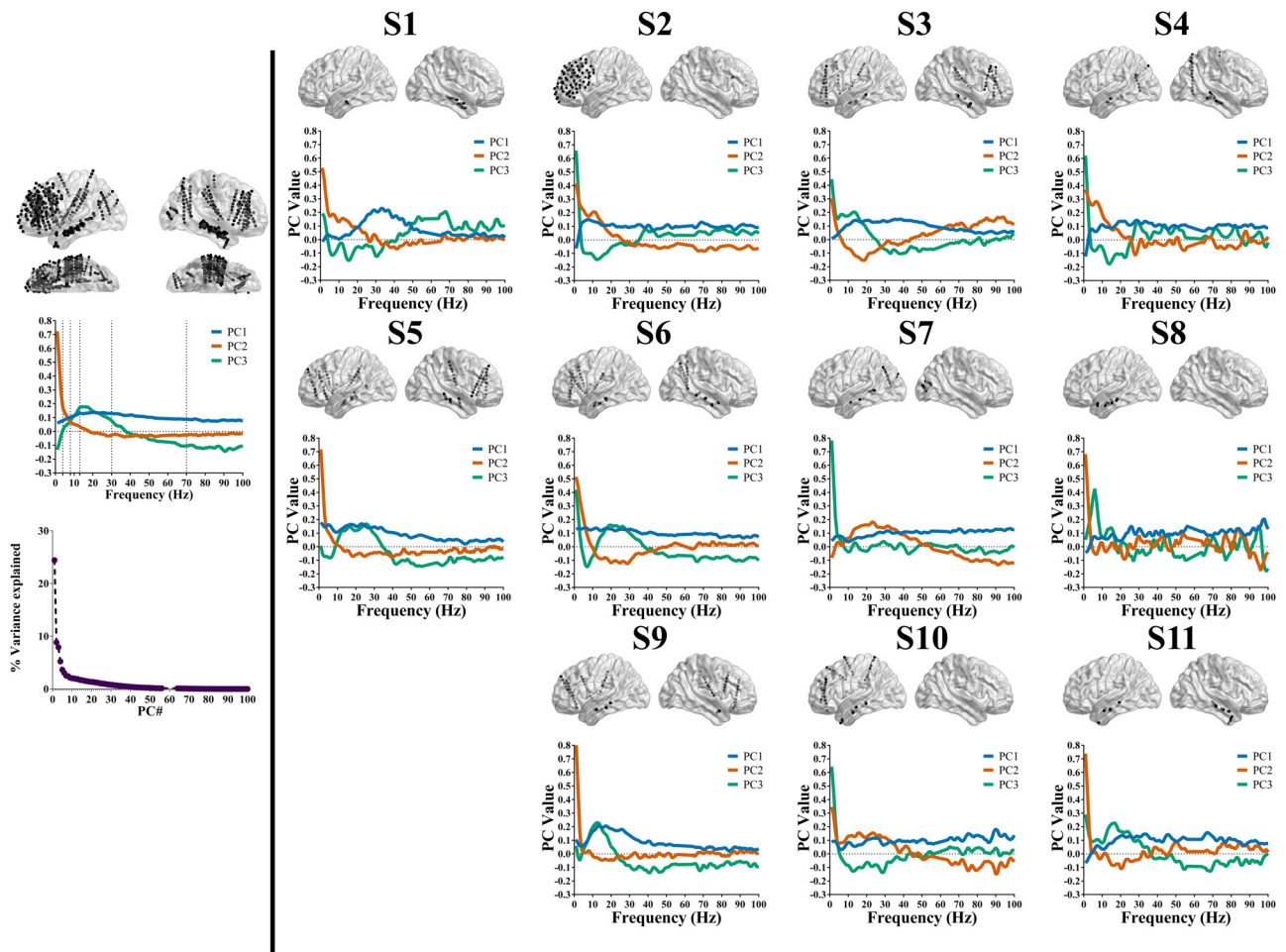


Figure 3. Results of PCA applied to individual participants. Left, Top, Recording sites for all participants mapped onto a reconstructed brain surface. The spectral patterns for the top three PCs are calculated based on condition-specific PSDs from all participants except the tested participant to avoid double dipping. Left, Middle, First three PCs for combined participant data. Left, bottom, Percentage of total variance explained by each PC shows that PC1 represents 24.4%, PC2 represents 8.8%, and PC3 represents 7.9%. Right, Each panel shows recording sites (top) and single-participant PCs (bottom). The number of recording sites used for PCA varies across participants between 28 and 132 recording sites.

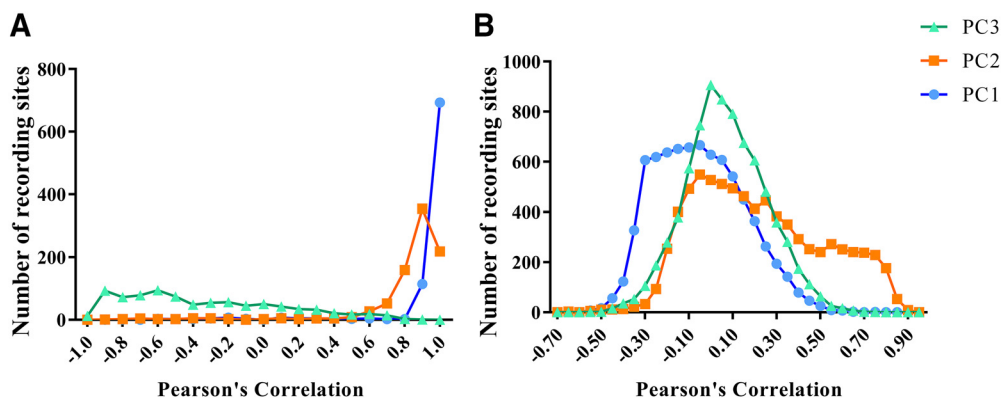


Figure 4. Relationship between principal components and the frequency content of event-related potentials. **A**, PC1 and PC2, but not PC3, encompass most of the frequency content of ERPs. Time series at each recording site were filtered by PC1 (blue), PC2 (orange), or PC3 (green). After filtering, ERPs were computed for signals at each recording site. The Pearson correlation between the “filtered” ERPs and standard, unfiltered ERPs was then calculated. The histogram of correlation coefficients for all 858 recording sites shows that the structure of ERPs was destroyed by PC3 filtering, but PC1 and PC2 filtering preserved most of the structure of ERPs. **B**, The PSD functions of the condition-specific ERPs are most closely correlated with PC2. PSD functions of the condition-specific ERPs at each recording site were computed and then correlated with each of the PCs. The histogram of correlation coefficients shows that PC1 and PC3 are uncorrelated with the PSD of ERPs. For many recording sites, however, PC2 is strongly correlated with the PSD of the ERPs.

of weight values. A “broadly tuned” encoding model would have a tuning width of 4 or 5, indicating an even distribution of weights about the middle of the range. Tuning width is plotted as weight value against weight rank. For the extreme case of broad and completely nonspecific

tuning, the resulting plot would appear as a flat line. For recording sites with highly specific encoding of a subset of task conditions, these plots will reveal curves with a sharp peak (corresponding to the largest model weight) and rapid falloff.

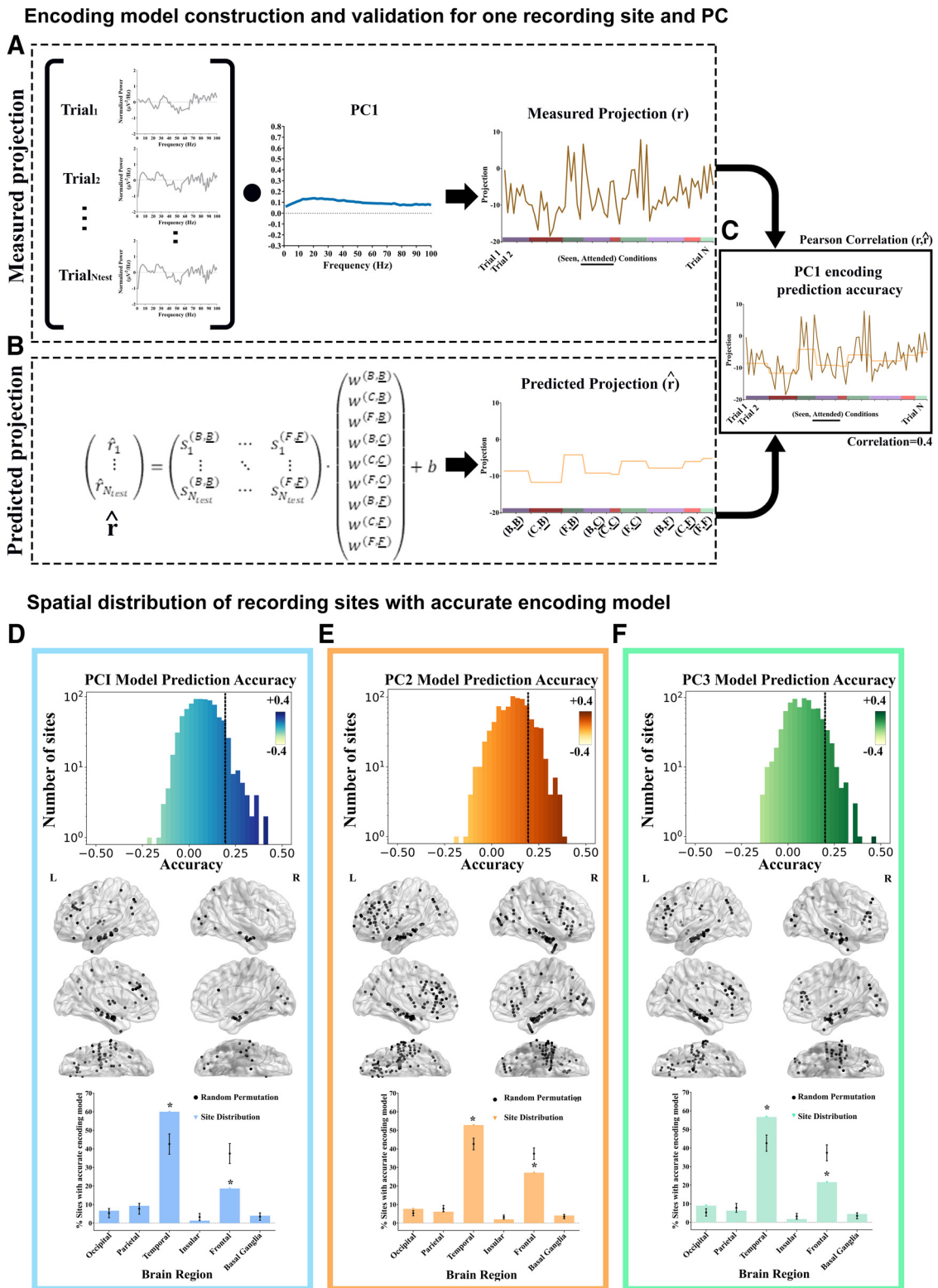


Figure 5. Spatial distribution of recording sites with an accurate encoding model for each PC. **A–C**, Encoding model construction and validation for one recording site and PC. The encoding model predicts power along each PC as a function of the nine conditions (i.e., each pairing of seen and attended object category) in the experiment. This panel shows the construction of an encoding model for the recording site illustrated in Figure 1. **A**, Measured PSDs (left) for a single recording site in the left hippocampus of participant 6 (S6) for several test trials. Measured projection values (r , right) are calculated by projecting each PSD onto the principal component (PC1 in this example, middle). The seen, attended category of each trial is color coded on the x-axis. **B**, Predicted projections (\hat{r} , beige curve at right) are calculated using the encoding model (the encoding model equation shown in the first panel was trained on an independent set of 80% of total trials). Here, s is the ($N_{test} \times 9$) design matrix of the test set (each column indicates a task condition), w is (9×1) vector of condition weights (i.e., regression parameters), and capital letters in the superscript indicate the (seen, attended) object categories, where B = building, F = face, and C = car. **C**, Comparison of the measured and predicted PC projections. Prediction accuracy is the Pearson correlation (0.4 in this example) between the measured and predicted projections. **D–F**, The same method described in **A–C** is applied to all 858 recording sites across PC1–PC3. Top panels, Histogram (log-scale) of encoding model prediction accuracy across all recording sites for each PC. Recording sites to the right of the accuracy threshold (dashed line, Pearson (Figure legend continues.)

Model-based decoding analysis

A decoding analysis was performed on the $N_{\text{val}} = 90$ trials that were not used to estimate or evaluate the encoding models. Decoding analyses were performed using data from participants S6–S11 only, as these were the participants for which the largest common number of trials (consisting of 432 trials) was acquired.

Here, the decoding objective is to predict the seen or attended object category associated with a set of responses sampled from a population of recording sites. Decoding is performed independently for each PC. As above, a response, r_j , associated with the j th PC p_j is defined as the projection of a PSD, q , onto the PC, $r_j = q \cdot p_j$. Below, we will suppress the subscript j , as it is understood that all responses refer to a specific PC.

Let $s1$, $s2$, and $s3$ indicate the seen object categories of “face,” “building,” and “car,” respectively. To decode a seen object category, say $s2$, we first form a response vector r_{s2} associated with category $s2$. Thus, we randomly select a trial associated with the category $s2$ under each attention condition, then we concatenate the responses of all recording sites within this population on this same selected trial. For example, to form a response vector associated with the seen category “building,” we randomly sample a trial associated with “see building, attend face,” “see building, attend building,” and “see building, attend car,” and we use those same sampled trials across all studied recording sites, as follows:

$$r_{s2} = (r_{s2:a1}^1, r_{s2:a2}^1, r_{s2:a3}^1, \dots, r_{s2:a1}^M, r_{s2:a2}^M, r_{s2:a3}^M),$$

where the superscript indexes a recording site; M is the total number of recording sites in the population; the subscripts $a1$, $a2$, and $a3$ index the three attended categories “face,” “building,” and “car”; and the notation $s2:a1$ indicates a trial associated with “see building, attend face,” and so on. The length of a response vector for a given seen object category is thus $3M$. Note that response vectors for the attended category can be formed in a similar way by fixing the attended category and allowing the object category to vary.

To perform model-based decoding of the seen object category from a response vector, say r_{s2} , we use the encoding model for each recording site in the response vector to generate a predicted response vector, r'_{s1} , r'_{s2} , r'_{s3} , for each of the three possible object categories. We then calculate a Pearson correlation between the response vector r_{s2} and each of the predicted response vectors r'_{s1} , r'_{s2} , r'_{s3} . Decoding is successful if the correlation between the response vector and its matching predicted response vector is higher than the correlation between the response vectors and the other two, nonmatching, predicted response vectors. In this example, decoding would be successful if the correlation between r_{s2} and r'_{s2} were larger than the correlation between r_{s2} and r'_{s1} , and also larger than the correlation between r_{s2} and r'_{s3} . We refer to successful decoding of a response vector as a “hit.”

To calculate the decoding performance for seen object categories, we formed 300 response vectors for each seen object category, for a total of 900 response vectors. Decoding performance is the number of hits divided by the total number of response vectors. Chance performance is 33%, since each response vector has a one-third chance of being successfully decoded by randomly selecting a seen object category. The decoding performance for the seen object category was calculated independently for each PC.

To perform model-based decoding of attended object category response vectors r_{a1} , r_{a2} , r_{a3} were correlated with predicted response vectors r'_{a1} , r'_{a2} , r'_{a3} as described above. Performance was calculated by

←

(Figure legend continued.) correlation > 0.19 , $p < 0.05$; permutation test) are designated as accurate. Middle panels, Recording sites (black dots) with accurate encoding models displayed on the surface of a template (MNI space) cortical surface reconstruction across the 11 participants for PC1 (D; $N = 75$ recording sites), PC2 (E; $N = 195$ recording sites), and PC3 (F; $N = 111$ recording sites). L, Left hemisphere; R, right hemisphere. Bottom panels, Percentage (bars) of recording sites with accurate encoding models in each of the labeled brain regions for each PC, as well as mean \pm SEM percentages under a null distribution (permutation test, $N = 10000$; stars indicate significance at $p < 10^{-3}$). Figure 6B shows their distribution across the 11 studied participants. And Figure 6C shows the spatial and spectral distribution of recording sites with an accurate encoding model.

counting “hits” across 900 response vectors, as described above. Decoding performance for attended object category was calculated independently for each PC.

Decoding performance was calculated for populations of varying sizes (i.e., populations with varying numbers of recording sites) ranging from 1 to 401 recording sites. To select recording sites for a population of a given size, all recording sites were rank ordered according to their prediction accuracy. Thus, the population of size one contained the single recording site with the highest prediction accuracy. The population of size two contained the recording sites with the highest and second-highest prediction accuracy, and so on. We emphasize that none of the trials used to measure prediction accuracy were used to measure decoding performance.

Analysis of event-related potentials

To determine the relationship between the PCs derived from single-trial PSDs and the PSD of event-related potentials (ERPs), we computed the time-domain average of signals locked to stimulus onset.

First, we compared ERPs derived from the time-domain signal to ERPs obtained after first filtering the time-domain signal with each PC (Miller et al., 2009). To perform this comparison, we applied a Morelet wavelet transform to signals at each recording site. We then multiplied the amplitude at each frequency by the corresponding PC power at that frequency. We then applied the inverse wavelet transform on the data to obtain a filtered time-domain signal. We then correlated for each recording site the ERPs derived from the PC-filtered time-domain signal with the unfiltered ERP.

Second, we compared the PSD of the unfiltered condition-specific ERPs to the PCs themselves. To do this, we computed the PSD of condition-specific ERPs. We then computed the Pearson correlation between these PSDs and each of the PCs.

Statistical analysis

Statistical analyses were performed using MATLAB 2018b (MathWorks) and GraphPad Prism 8 (GraphPad Software). Permutation tests were performed for comparisons to random distributions. PCA analysis was performed as described above. A Kruskal–Wallis test was used to compare the ranks of the tuning curves. A Wilcoxon signed-rank test was used to compare accuracy of decoding models to chance. Unless otherwise specified, statistical significance is reported at $\alpha < 0.05$.

Code accessibility

Source codes for all analyses and tests performed in this study are available upon request from the senior author.

Results

Eleven participants performed a category-based visual attention task in which they viewed photographs (presented for 0.5 s and followed by a 1 s interstimulus interval) of buildings, cars, and faces while attending to only one of the three categories (Fig. 1). At randomly selected times, participants were required to perform a one-back task in which they indicated whether the previous photograph depicted an object from the attended category. Each experimental condition thus corresponded to one of nine distinct pairs of seen and attended object categories (e.g., “see building, attend face”).

Broadband and narrowband, low-frequency components explain almost half the variance in spectral power across seen and attended object categories

To determine the combinations of frequency bands that explain the most variance in spectral power across seen and attended categories, we first prepared single-trial, peristimulus PSD functions for each recording site in the dataset (858 recording sites across 11 participants). Single-trial PSDs were log transformed and normalized ($\log[\text{PSD}/(\text{PSD})]$), where $\langle \cdot \rangle$ indicates an average across all trials for the given recording site; Miller et al., 2016).

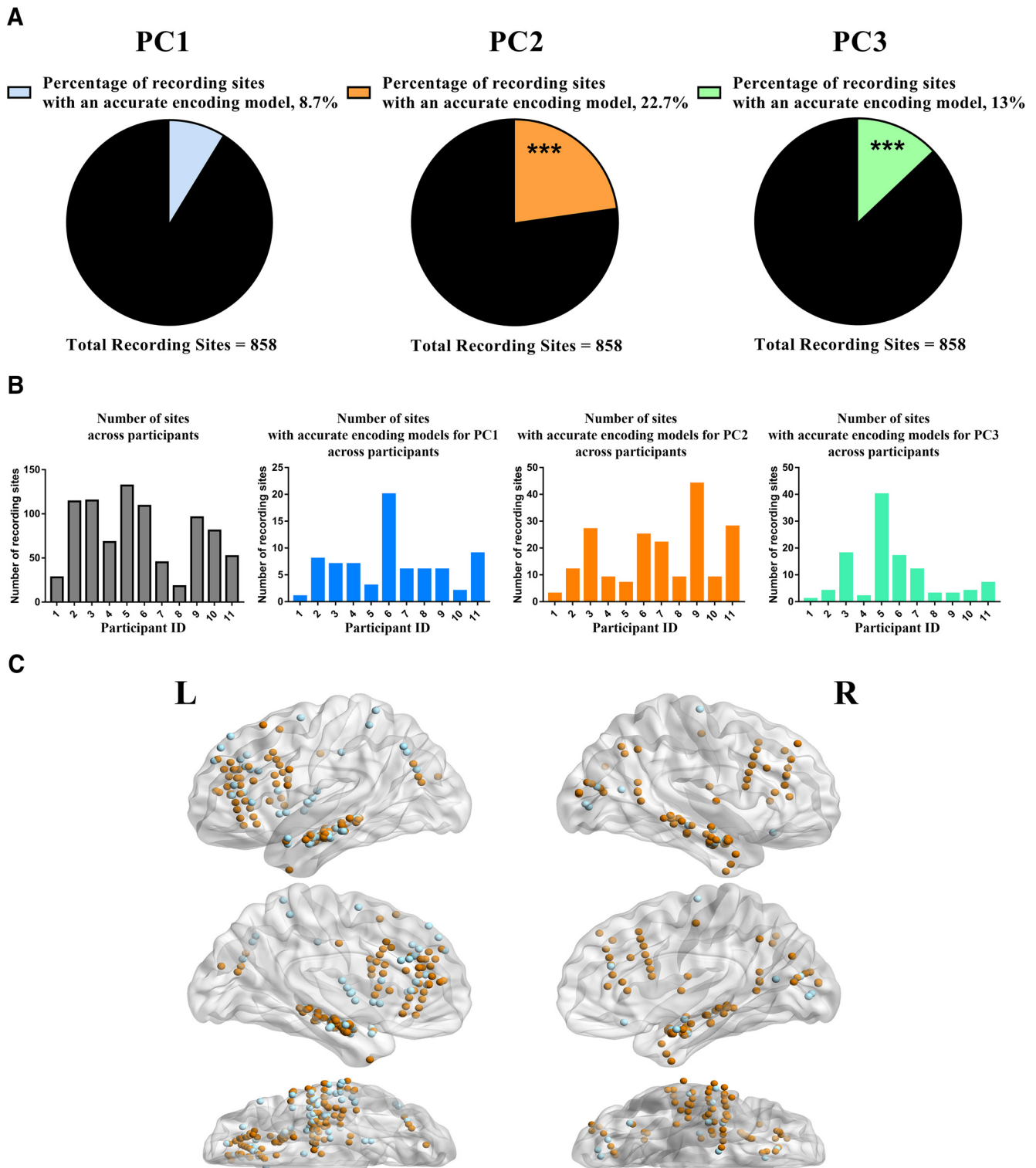


Figure 6. Spatial and spectral distribution of recording sites with an accurate encoding model. **A**, Pie charts show the proportion of recording sites with an accurate encoding model for the studied PC. There was a significantly higher proportion of recording sites with an accurate encoding model for PC2 and PC3 compared with PC1 ($N = 848$ total recording sites; $***p = 10^{-5}$ comparing each PC to PC1 permutation test with 1×10^5 permutations). **B**, From left to right: histogram showing the number of recording sites across the studied 11 participants, then the number of recording sites with accurate encoding models for PC1, PC2, and PC3. **C**, Locations of recording sites with an accurate encoding model for PC1 (broadband pattern, light blue) and PC2/3 (narrowband pattern, orange). Data shown in this panel are the same as the ones presented in Figure 5D–F, middle panels, but plotted here on one brain.

Normalized PSDs were then grouped according to the combination of seen and attended object category associated with each trial, and then averaged within group. This resulted in nine normalized mean PSDs (one for each pair of seen object category and at-

tended object category). An additional PSD corresponding to an average across all the interstimulus intervals was also prepared.

PCA was performed on pooled data from all participants (Fig. 2). We found that variance was rather broadly distributed across

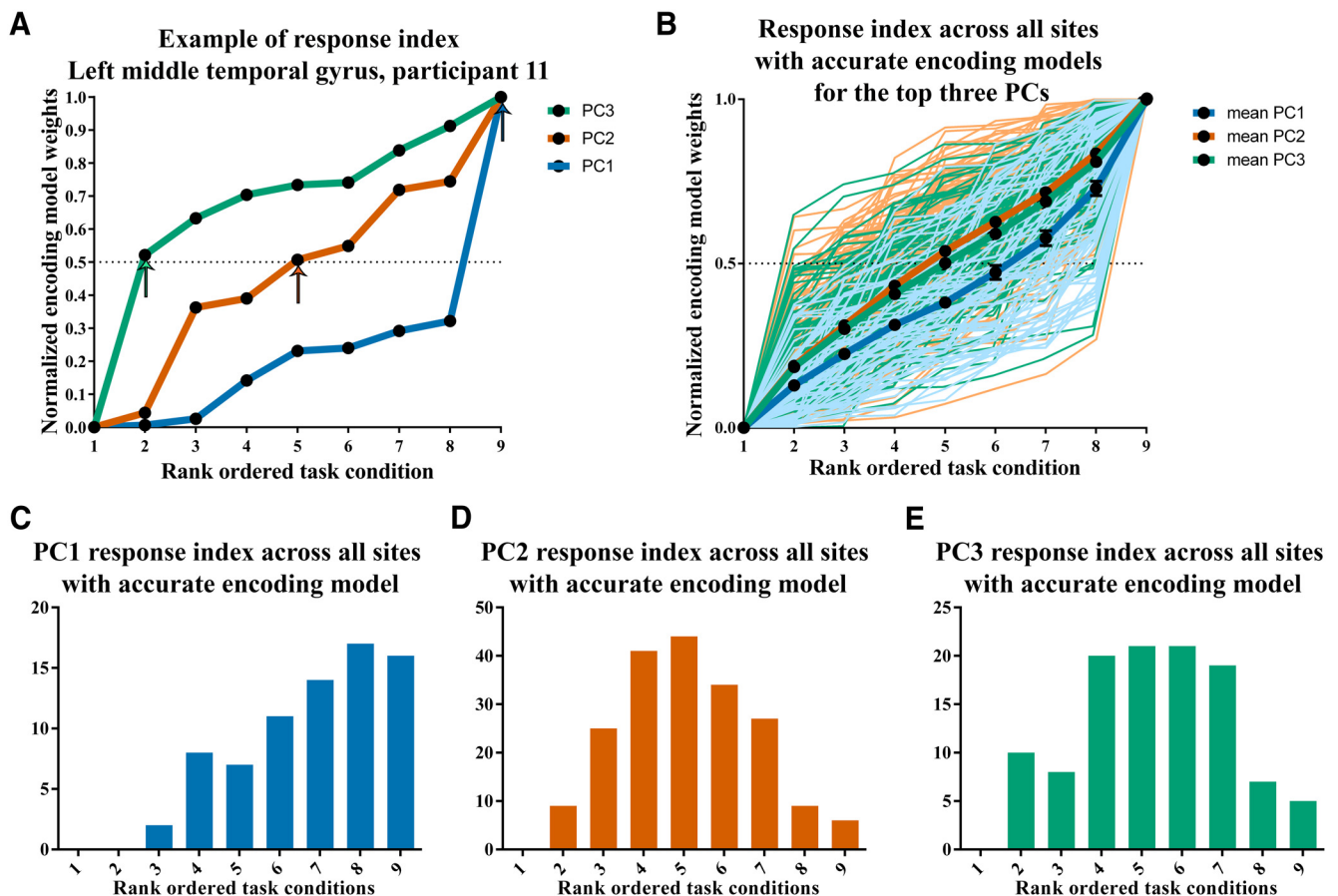


Figure 7. Selectivity index and sharpness of tuning to seen and attended object categories. **A**, Comparison of sharpness of tuning across the top three PCs for one recording site that had accurate encoding model for all three PCs. Parameters (normalized) of the encoding model for PC1–3 are plotted in blue, orange, and green, respectively. For each PC, the parameters of the encoding model were rank ordered along the x-axis. The rank of the smallest-valued parameter (arrows) above the middle of the normalized range (dashed line at 0.5) was then identified. High (8–9) or low (1–2) values of this response index indicate relatively sharp tuning. Midrange values (3–7) indicate relatively broad tuning. For the recording site shown here (participant S11, middle temporal gyrus), the index for PC1 is 9 (blue arrow), indicating sharp tuning of the response along the broadband component. Indices for PC2 and PC3 are 3 (orange arrow) and 6 (green arrow), respectively, indicated relatively broad tuning in the response along the low-frequency components. The detailed steps for calculating the selectivity index is presented in Figure 8. **B**, Rank-ordered tuning plots for every single recording site (thin lines; color corresponds to PC) and cross-site mean (thick lines, error bars are SEM) with an accurate encoding model for PC1, PC2, and PC3. **C–E**, Histogram of the number of recording sites with accurate encoding models for each of PC1 (**C**), PC2 (**D**), and PC3 (**E**) for each value of the response index. Response indices for PC1 are skewed toward the extremum relative to PC2–3, indicating sharper tuning of the broadband PC1 response than the low-frequency PC2 and PC3 responses. Locations of recording sites with an accurate PC1 and PC2–3 encoding model and high response index within the temporal and occipital lobes are shown in Figure 9.

PCs, with the top three PCs accounting for slightly less than half (41.7%) of the total variance in spectral power (Fig. 3). We focused our analysis efforts on these three PCs (Figs. 2D, 3). PC1 (24% of total variance) had a broadband profile indicated by elevated (nonzero) power across all frequencies in the studied range (1–100 Hz). PC2 (8.8%) and PC3 (7.9%) had narrowband, low-frequency profiles, as indicated by nonzero power over a narrow frequency range. Specifically, PC2 concentrated most nonzero power in a delta–theta frequency range (1–8 Hz), which for many recording sites was highly correlated with the PSD of the event-related potential (Fig. 4B). PC3 concentrated nonzero power primarily in an alpha–beta frequency range (8–30 Hz).

PCA applied to single-participant data revealed some variation in the ranking of the top three PCs (Fig. 3), but the basic motif of one broadband and two narrowband low-frequency components was conserved across participants. Thus, as seen and attended categories varied in our experiment, the largest variation in spectral power was along a broadband component (corresponding approximately to a variation in the sum of power across all frequencies), while the second and third largest varia-

tions in spectral power were restricted to low (sub-gamma) frequencies. Higher PCs explained vanishingly small variance and were not stable across participants; consequently, they were excluded from further analysis.

Broadband and narrowband, low-frequency components encode information about seen and attended object categories

To determine how information about seen and attended categories was encoded in each principal component, we independently estimated an encoding model for each of the top three PCs, and for each recording site. The encoding model predicted, for each trial, the spectral power along each PC as a function of the seen/attended object category associated with the trial (Fig. 5A–C). Thus, the encoding models used here were quite simple: for a given PC and recording site, a design matrix (independent variable) indicating the seen and attended category associated with each trial was regressed onto the projection of the PSD onto the PC (see Materials and Methods above). We estimated encoding models independently for each PC and recording site using linear regression.

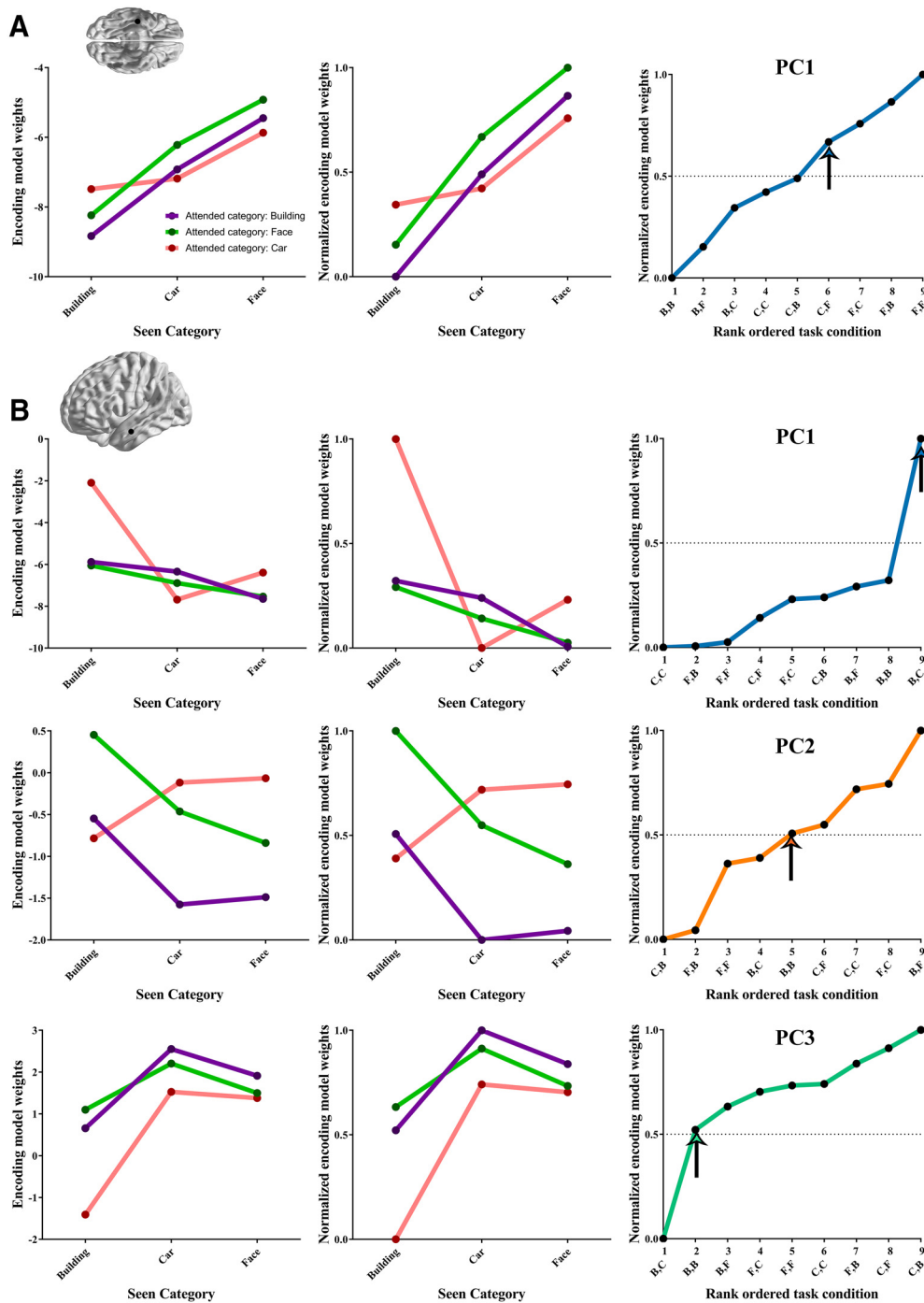


Figure 8. Examples of encoding model weights. **A**, Example of encoding model weights for a recording site in the ventral temporal lobe (location is shown as a black dot on the normalized brain template). Left, The encoding model weights for PC1 (y-axis) for each of the shown categories (x-axis) across the three attended categories (attended to building in purple, to face in green, and to car in pink). In this example, seeing a face had a higher projection than seeing a building or a car independently from the attended category. Middle, To compute a selectivity index, the encoding model weights of the left panel are then normalized to the range 0–1 (y-axis). Right, The nine (attended, seen) conditions are then rank ordered to sort the normalized encoding weights of the middle panel from the lowest to the highest value. Now each rank ordered task has a new abstract value from 1 to 9, representing its rank order position on the x-axis. The selectivity index is the x-axis value at which the normalized encoding weight reaches above the middle of the range (0.5). In this example, the selectivity index is 6 (pointed to with an arrow), and the conditions are rank ordered based on the seen category. **B**, The same steps as in **A** are shown for a recording site within the middle temporal gyrus (location is shown as a black dot on the brain template) for PC1 (first row, in blue), PC2 (second row, in orange), and PC3 (third row, in green). The summary of the selectivity index of this recording site across the top three PCs is shown in Figure 7.

We evaluated the prediction accuracy of the encoding model for each recording site and PC by calculating the Pearson correlation between the predictions of the model and the observed spectral power along the PC across a held-out set of testing trials (Fig. 5C). For many recording sites, this model achieves fairly high prediction accuracy (Fig. 5D–F).

Encoding models were considered accurate if their prediction accuracy exceeded a common threshold (Pearson correlation > 0.19, $p < 0.05$; permutation test). Recording sites with accurate encoding models for all three PCs were identified in the temporal, prefrontal, occipital, insular, and parietal cortex (Fig. 5D–F). There were no recording sites within the precentral gyrus that

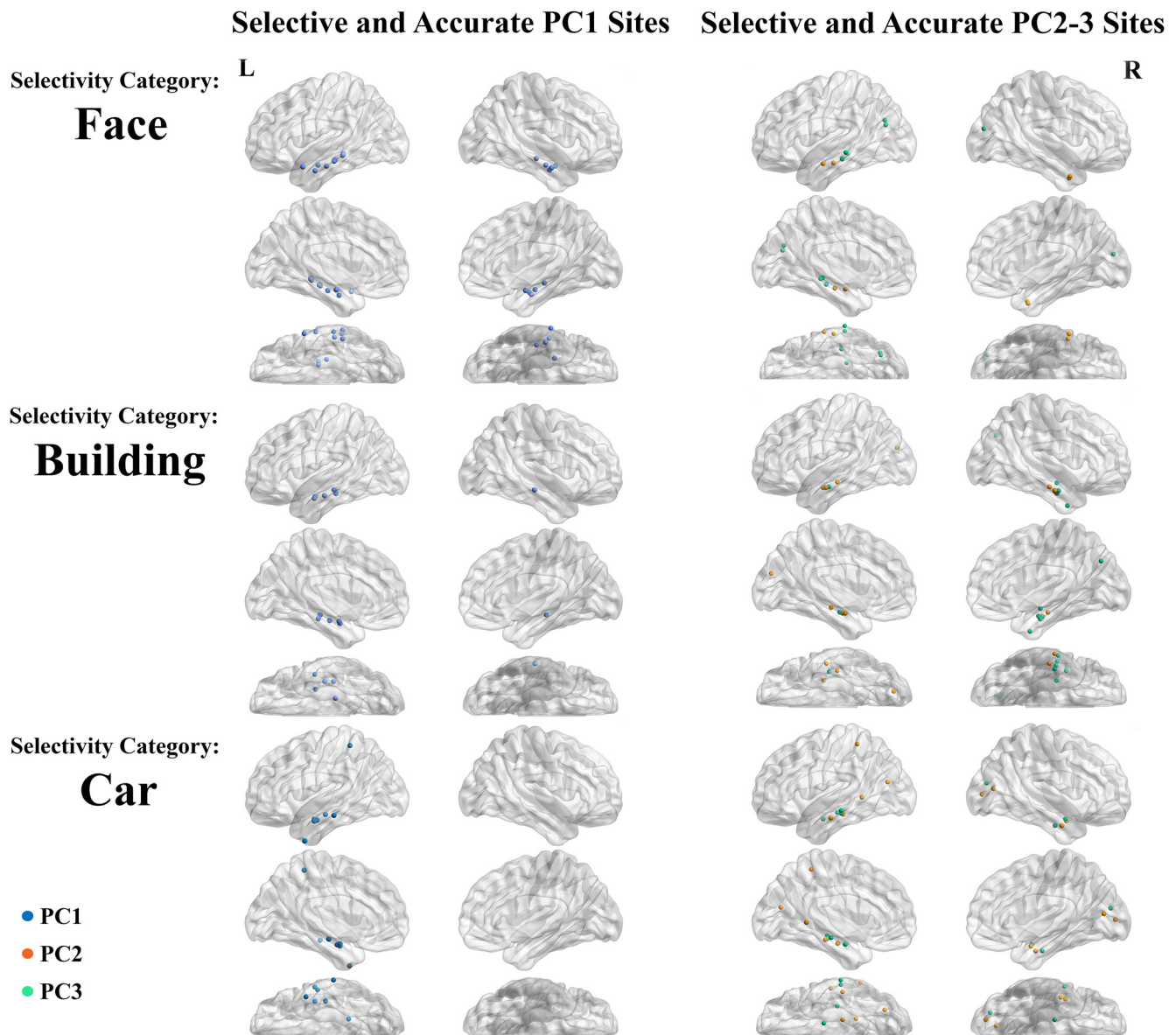


Figure 9. Locations of recording sites with an accurate PC1–3 encoding model and high response index. In the left column of the brain maps, shown in light blue are dots representing each recording site that had accurate encoding on PC1 and was sharply tuned for the specified object category with a response index ≥ 7 within the temporal and occipital lobes. In the right column of the brain maps, shown in orange and in green are dots representing each recording site that had accurate encoding on PC2 and PC3, respectively, and was sharply tuned for the specified object category with a response index ≥ 7 within the temporal and occipital lobes.

showed accurate encoding for any of the three PCs (0 of 8 studied recording sites). We observed significantly fewer recording sites with an accurate encoding model for PC1 compared with either PC2 or PC3 (75, 195, and 111 sites with an accurate encoding model for PC1, PC2, and PC3, respectively; $p = 0.00000$, permutation test with 100,000 permutations, $n = 848$; Fig. 6A). The recording sites with accurate encoding models for PC2 and PC3 appeared to be spatially distributed more broadly than recording sites with accurate encoding models for PC1 (Fig. 6C). Recording sites with accurate encoding models for PC1–3 were present in all 11 participants (Fig. 6B).

Although recording sites with accurate encoding models were distributed widely across the brain, we found that in temporal lobe the number of recording sites with accurate encoding models was significantly higher than would be expected by random sampling for all PCs (45, 103, and 63 sites for PC1, PC2, and PC3, respectively; $p = 0.0009$, 0.0006, and 0.0001, permutation test).

In the prefrontal lobe, fewer recording sites than would be expected by such random sampling had accurate encoding models for PC1–3 (14, 53, and 24 sites for PC1, PC2, and PC3, respectively; $p = 0.0002$, 0.0005, and 0.0000, permutation test). For all other regions, the number of recording sites that exceeded the prediction accuracy threshold was not statistically different from chance (Fig. 5D–F, bottom).

PC1 encoding models are more sharply tuned than PC2 or PC3 encoding models

Are power variations along each PC equally dependent on all seen/attended categories, or only on select ones? To address this question, we analyzed the distribution of weight values of the encoding model for each recording site and PC (Fig. 7). A recording site where power along one of the PCs increased in response to only one of the nine experimental conditions would be highly selective or sharply tuned. In this case, we would expect only one

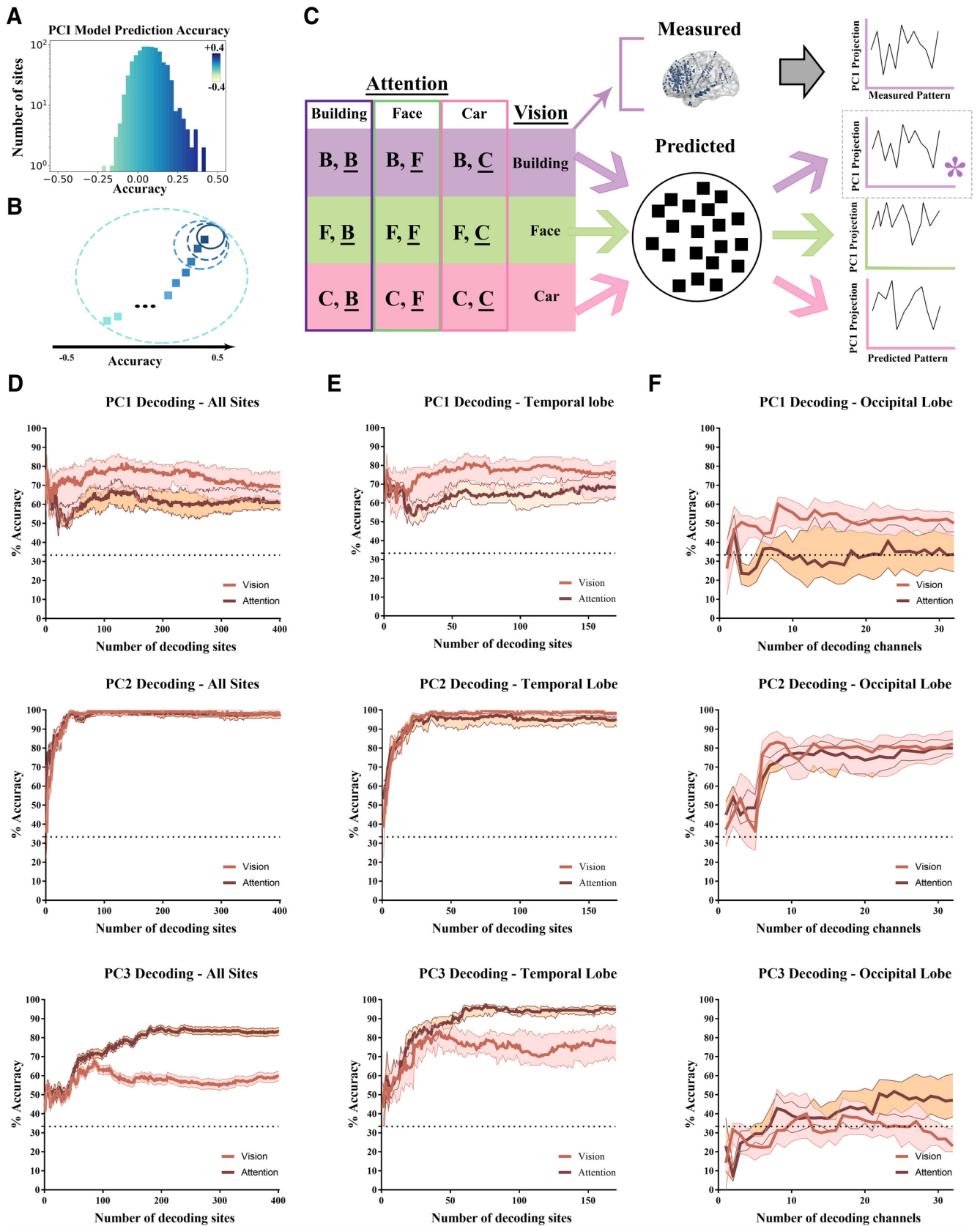


Figure 10. Model-based decoding of seen and attended object categories. Model-based decoders were used to identify the seen or attended object category encoded in spectral power along PC1, PC2, and PC3 from populations of recording sites. **A, B**, How populations were selected. **A**, Recording sites were selected on the basis of their encoding model prediction accuracy. The plot shows a histogram of encoding model accuracies for PC1 (similar histograms were constructed for PC2 and PC3). Color indicates prediction accuracy. **B**, Populations of varying size were constructed by selecting recording sites in order of encoding model prediction accuracy. The population of size one (squares indicate recording sites; circles delimit populations) contains the single recording site whose encoding model has the highest prediction accuracy. The population of size two contains the recording sites whose encoding models have the highest and (Figure legend continues.)

of the encoding model weights to have a value well above or below the middle of the range of values of the weights for that model. For recording sites where activity was nonselective or broadly tuned, we would expect to see all weight values distributed close to the middle of the range. Thus, a simple index for sharpness of tuning is the rank of the encoding model weight that exceeded the middle of the range of weight values (Fig. 8). By this measure, we find that encoding models for PC1 are, as a population, significantly more sharply tuned than encoding models for PC2 and PC3 ($n = 75, 195,$ and 111 for PC1, PC2, and PC3, respectively; rank-order comparison via Kruskal–Wallis test with Dunn’s correction for multiple comparisons, $p < 0.0001$; Fig. 7B). This was true when tested for all recording sites (Fig. 7C) and when tested only for recording sites in the temporal lobe (data not shown). In the temporal lobe, recording sites where activity was driven most strongly by faces tended to cluster within the fusiform gyrus; recording sites where activity was driven most strongly by buildings tended to cluster near parahippocampal gyrus (Fig. 9, left panels, but see right panels for locations of recording sites with accurate PC2–3 encoding models). The encoding of faces and buildings by power along PC1 is thus consistent with well known findings of fMRI studies (Kanwisher et al., 1997; Haxby et al., 2001; Joseph, 2001).

Information about seen and attended object categories are unequally distributed across the first three principal components

Given previous results (Klimesch, 1999; Klimesch et al., 2011; Miller et al., 2014, 2016; Jensen et al., 2015; Michalareas et al., 2016; Helfrich et al., 2017), we suspected that variations in spectral power along the top three PCs might encode different amounts of information about seen versus attended object categories.

←

(Figure legend continued.) second-highest prediction accuracies, and so on. The largest population contains all recording sites. **C**, The procedure for decoding the seen object category from a population of recording sites (a similar procedure was applied to decode the attended object category). The first three columns on the left of the matrix indicate the nine experimental conditions. Capital letters indicate the seen, attended object categories, where B is building, F is face, and C is car. The attended object category is fixed along each column; the seen object category is fixed along each row. The rightmost column indicates the seen object category to be identified from measured spectral power (along PC1 in this example) in the population of recording sites. In this example, the “true” seen object category is building. Thus, measured spectral power was obtained from three randomly selected trials (top row of the matrix) in which the seen object was building and the attended object was car, face, and building, respectively. Measured spectral power for each recording site and trial were concatenated to form the “measured pattern,” which is illustrated by the plot at top right. In this plot, recording site and condition are arranged along the x -axis, and measured spectral power is indicated on the y -axis. To decode the seen object category from this pattern of measured spectral power, the encoding model for each recording site in the population was used to generate “predicted patterns” (bottom three plots at right) corresponding to each of the three seen object categories. That is, for each seen object category predicted spectral power for each recording site is concatenated across the three attended object categories. Each of the three predicted patterns was then correlated with the measured pattern. If the predicted pattern corresponding to the true seen object (in this example, building, indicated with an asterisk) was most highly correlated with the measured pattern, decoding was considered successful and marked as a hit. **D–F**, Decoding performance (curves indicate median, shading is 95% confidence interval) for spectral power measured along PC1 (top row), PC2 (middle row), and PC3 (bottom row). Recording sites were sampled from all available sites (left column), temporal lobe only (middle column), or occipital lobe only (right column). Each plot shows the median percentage of hits (y -axis) across 900 attempts (300 attempts per each object category) for populations of varying size (x -axis). Decoding of the seen object category (red curves) and the attended object category (brown curves) was performed independently. The quantification of accuracy is detailed in Figure 11.

Inspection of encoding model weights revealed a complicated pattern of interactions between the encoding of seen and attended object categories. Although some recording site channels showed, for some PCs, a classical amplification of signal when the preferred category was seen (Fig. 8A), most recording sites showed a more complicated form of interaction between attention and vision across categories and across PCs that was not easily apprehended by visual inspection (Fig. 8B). Thus, we used a model-based decoding analysis (Fig. 10A–C) to compare the amount of information about the seen object category encoded by each PC to the amount of information about attended object category encoded by each PC. This analysis used the encoding models to decode from measured brain activity the seen or attended object category associated with single trials (Kay et al., 2008). The measure of brain activity in this case was the projection of the peristimulus PSD of each recording site onto a single PC (i.e., the dependent variable of the encoding model for each recording site). Thus, the analysis allowed us to independently decode seen and attended object categories from variations in power along each PC. By comparing the accuracy of the decoded seen object category to the accuracy of the decoded attended object category, we were able to quantify the relative amount of information that each PC encodes about vision and attention.

We observed a consistent pattern of differential decoding accuracy for seen and attended object categories regardless of whether brain activity was sampled from all lobes (Fig. 10D, 401 recording sites), the temporal lobe only (Fig. 10E, 170 recording sites), or the occipital lobe only (Fig. 10F, 32 recording sites). For PC1, the seen object category was decoded more accurately than the attended category; in particular, only the seen category was decoded more accurately than chance from recording sites in the occipital lobe (Wilcoxon signed rank test comparison to chance, $p < 0.01$). For PC2 decoding, accuracy for seen and attended categories was at parity and significantly better than chance in all brain areas (Wilcoxon test, $p < 0.0001$). For PC3, the attended object category was decoded more accurately than the seen category; in particular, only the attended category was decoded more accurately than chance from recording sites in the occipital lobe (Wilcoxon test, $p < 0.0001$). Thus, variations in broadband spectral power encoded more information about the seen than the attended object category; variations in delta–theta narrowband power robustly encoded both seen and attended categories; variations in narrowband α – β power encoded more information about the attended than the seen object category. A more granular analysis of decoding accuracy for individual object categories (Fig. 11) was consistent with this pattern.

Discussion

Summary of results

We analyzed how variations in the spectral power of local field potentials encode variations in seen and attended object categories. Using PCA, we found that ~40% of the variance in spectral power is explained by variation in broadband (PC1) power, and by variation in low (delta–theta for PC2, alpha–beta for PC3) frequencies specifically. Power variation along the broadband and low-frequency PCs encoded information about variations in seen and attended object categories at recording sites in occipital, temporal, insular, parietal, and prefrontal cortices. Recording sites with accurate encoding models were especially prevalent in the temporal lobe, which is well known to maintain representations of object category (Kanwisher et al., 1997; Haxby et al., 2001; Joseph, 2001).

Although broadband PC1 explained the single largest fraction of variance, it was less tightly coupled than PC2 and PC3 to vari-

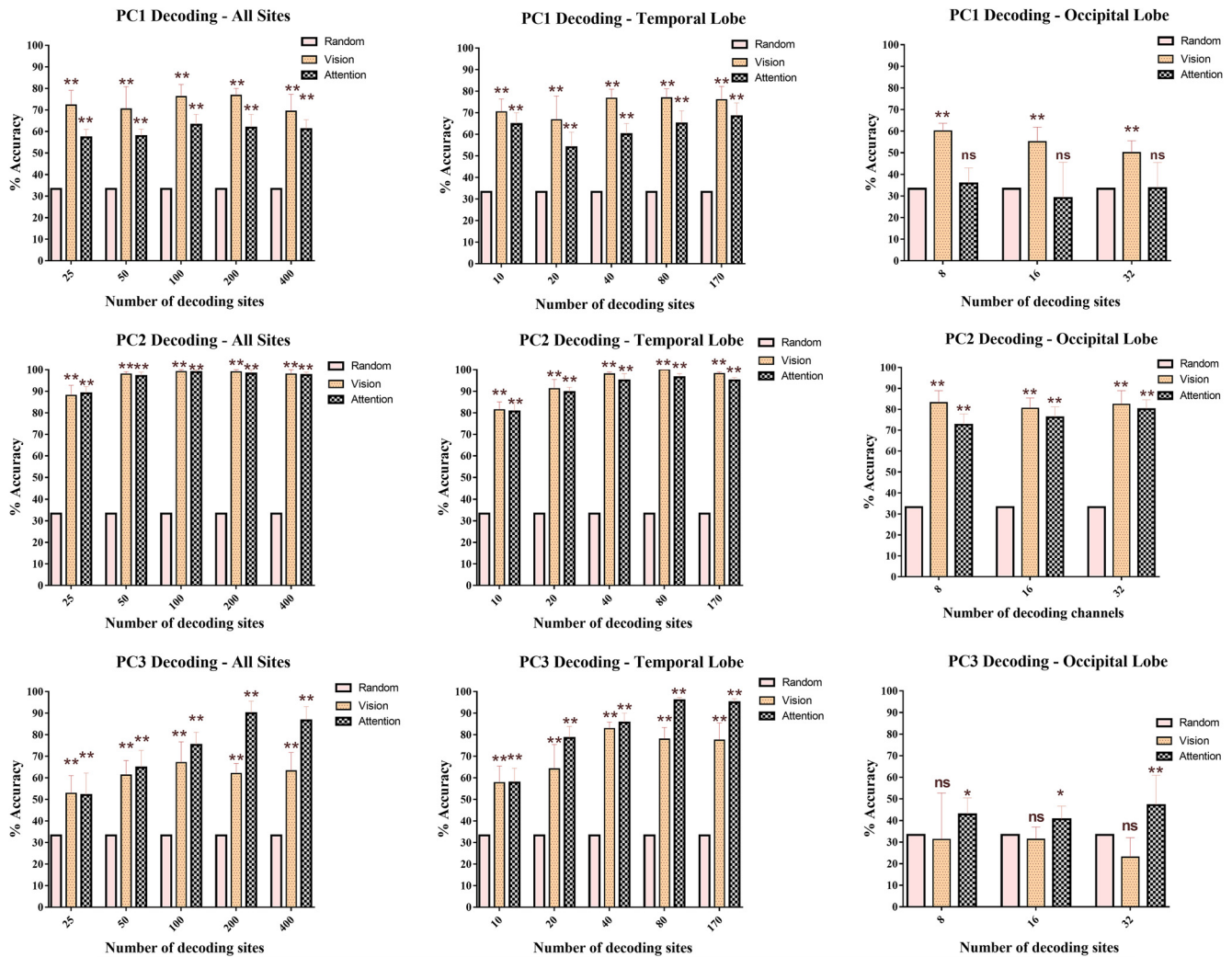


Figure 11. Category-specific decoding accuracy for each PC. Median decoding accuracy (error bars represent the 95% confidence interval) for populations of varying size. A Wilcoxon signed rank test was used to compare each sample to accuracy expected by chance (33.33% accuracy) for decoding specific categories (face, building, or car). * $p < 0.01$, ** $p < 0.0001$. ns, not statistically significant.

ations in seen and attended object categories, as evidenced by the fact that fewer recording sites had an accurate PC1 encoding model than they did for the low-frequency PC2 or PC3 encoding models. This means that, relative to PC2 and PC3, a larger fraction of the variance explained by PC1 was not related to variations in seen or attended object categories. Thus, in our experiment, broadband power may have encoded low-level visual features (Winawer et al., 2013) and may have reflected variations in signal properties across different participants and different brain areas, as well as the inevitable trial-to-trial variability of signals recorded at one site (i.e., the SEM of the within-group average PSDs that were entered into the PCA).

For recording sites with an accurate PC1 encoding model, the variation in power along the broadband PC1 was more sharply tuned than for the low-frequency PC2 and PC3. This finding is consistent with prior work showing that during finger movement, the broadband pattern was more sparse than the narrowband pattern (Miller et al., 2009). The selective tuning of power along the broadband PC1 is reminiscent of the highly selective BOLD responses to object category observed in temporal lobe (Bar et al., 2001; Haxby et al., 2001). Indeed, previous work suggests that broadband LFP and the BOLD response are tightly coupled (Hermes et al., 2017).

Although information about seen and attended object categories was encoded by power variations along each PC, the relative amount of information about vision and attention varied across PCs. Broadband PC1 encoded more information about the seen category than the attended category. The delta–theta band PC2 robustly encoded information about both the seen and attended categories. The alpha–beta band PC3 encoded more information about the attended category than the seen category.

Information about seen and attended object category is encoded by variation in power at many frequencies

Our findings are consistent with previous work that reported the encoding of object category in broadband and low-frequency power variations. (Bosman et al., 2012; Miller et al., 2014, 2016; Jensen et al., 2015; Michalareas et al., 2016). In this work, we studied the encoding of both the seen and attended object categories across a spatially broad distribution of recording sites. We found that the encoding of attended object categories is, like the encoding of seen object categories, encoded by power variation across a range of frequencies.

We did not observe a principal component that resembled a narrowband gamma response (Fries et al., 2008; Fries, 2009). The absence of an obvious narrowband gamma-component does not

mean that gamma oscillations are not an important component of the neural processing of seen and attended object categories. Our findings simply indicate that a narrowband gamma response did not account for enough of the variance in spectral power during our experiment to merit its own PC. Also notable is an absence of a spectral component resembling a broad (high-pass) gamma response (Jacobs and Kahana, 2009; Vidal et al., 2010). One possible explanation for this finding is that broad, high-pass gamma responses reflect a broadband effect that is masked by a suppressed low-frequency (sub-gamma) narrowband response (Miller et al., 2014). Therefore, by decoupling the broadband and low-frequency narrowband patterns, the PCA performed here unmasked the low-frequency components observed in PC2–3 and revealed the underlying broadband effect.

Our results may be interpreted as endorsing an expansive, inclusive approach to analyzing the relationship between variations in spectral power and task conditions, particularly for tasks that involve interacting top–down and bottom–up signals. Focusing exclusively on a single frequency band, such as gamma, may obscure components of the spectrum that encode a significant amount of information.

The broadband response

Our results provide a detailed characterization of the broadband response to seen and attend object categories. We now briefly summarize its most salient characteristics.

Broadband responses are promiscuous. Projecting a PSD onto the broadband components effects a nearly equally weighted sum of powers at all frequencies. Therefore, power variations at any frequency will register as nonzero variance along the broadband PC. This makes broadband power variations sensitive to both signal and noise, broadly construed. This is most likely why the broadband PC explains the most variance in PSDs but at the same time offers the fewest recording sites with an accurate encoding model.

Interestingly, where the broadband response is coupled to variation in seen and attended categories (i.e., at recording sites with an accurate encoding model), it is more selective than low-frequency components (PC2–3). Of the three PCs, PC1 was the most sharply tuned. In the temporal lobe, the recording sites where broadband response was selective for faces tend to locate more lateral than recording sites where the broadband response was most selective for buildings, which in turn were more medially located. This is generally consistent with relative locations of the place- and face-selective regions revealed in fMRI studies (Kanwisher et al., 1997; Haxby et al., 2001; Joseph, 2001).

Across the population of recording sites studied here, broadband responses were quite redundant. This was evidenced by plots showing the change in decoding performance as a function of population size. Relative to the low-frequency PCs, decoding performance plateaued rapidly when decoding was performed with broadband responses. This observation held regardless of whether recording sites were sampled from only occipital lobe, only temporal lobe, or all available sites.

Finally, in early visual areas, broadband responses during our experiment were “bottom–up.” This was evidenced by the fact that recording sites in the occipital lobe permitted above-chance decoding performance for the seen object category, but not the attended object category. In contrast, low-frequency PC2 responses in occipital lobe recording sites permitted above-chance decoding of both seen and attended object category, while PC3 responses in occipital lobe recording sites permitted the above-chance decoding of the attended object only.

In summary, we find that broadband PC1 responses were promiscuous (i.e., sensitive to multiple sources of variance), sharply tuned, generally consistent with findings from fMRI studies, encoded redundant information across the recording sites studied here, and, in occipital lobe, encoded only bottom–up sensory information.

Decoding attention and vision from spectral power variation

Many previous studies have reported that attention modulates the encoding of seen stimuli in firing rate (Moran and Desimone, 1985) and BOLD responses (Çukur et al., 2013), and can increase the strength of a response to a preferred stimulus category (Downing et al., 2001; Baldauf and Desimone, 2014). In our hands, analysis of the interactions among attended category, preferred category, and spectral component did not admit any clearly interpretable pattern of dependence. However, one clear and straightforward pattern was revealed by the model-based decoding analyses presented here: we found that information about attention and the seen object category was differentially distributed across the frequency spectrum, as summarized by the model-based decoding analyses above. This finding suggests that attention and vision have different spectral signatures. In particular, variation in power within the alpha–beta band of frequencies (PC3) was highly selective for the attended object category. This finding is consistent with emerging evidence for asymmetric encoding of top–down and bottom–up signals in the variation of spectral power (Bastos et al., 2015; Jensen et al., 2015; Michalareas et al., 2016).

References

- Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science* 344:424–427.
- Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529–535.
- Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85:390–401.
- Bosman CA, Schoffelen JM, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P (2012) Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* 75:875–888.
- Brett M, Leff AP, Rorden C, Ashburner J (2001) Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage* 14:486–500.
- Brunet N, Vinck M, Bosman CA, Singer W, Fries P (2014) Gamma or no gamma, that is the question. *Trends Cogn Sci* 18:507–509.
- Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913–919.
- Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci* 16:763–770.
- Downing P, Liu J, Kanwisher N (2001) Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia* 39:1329–1342.
- Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci* 32:209–224.
- Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560–1563.
- Fries P, Scheeringa R, Oostenveld R (2008) Finding gamma. *Neuron* 58:303–305.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Helfrich RF, Huang M, Wilson G, Knight RT (2017) Prefrontal cortex mod-

- ulates posterior alpha oscillations during top-down guided visual perception. *Proc Natl Acad Sci U S A* 114:9457–9462.
- Hermes D, Miller KJ, Wandell BA, Winawer J (2015) Stimulus dependence of gamma oscillations in human visual cortex. *Cereb Cortex* 25:2951–2959.
- Hermes D, Nguyen M, Winawer J (2017) Neuronal synchrony and the relation between the blood-oxygen-level dependent response and the local field potential. *PLoS Biol* 15:e2001461.
- Hoogenboom N, Schoffelen JM, Oostenveld R, Parkes LM, Fries P (2006) Localizing human visual gamma-band activity in frequency, time and space. *Neuroimage* 29:764–773.
- Jacobs J, Kahana MJ (2009) Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *J Neurosci* 29:10203–10214.
- Jensen O, Bonnefond M, Marshall TR, Tiesinga P (2015) Oscillatory mechanisms of feedforward and feedback visual processing. *Trends Neurosci* 38:192–194.
- Jia X, Xing D, Kohn A (2013) No consistent relationship between gamma power and peak frequency in macaque primary visual cortex. *J Neurosci* 33:17–25.
- Joseph JE (2001) Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. *Cogn Affect Behav Neurosci* 1:119–136.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355.
- Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* 29:169–195.
- Klimesch W, Fellinger R, Freunberger R (2011) Alpha oscillations and early stages of visual encoding. *Front Psychol* 2:118.
- Michalareas G, Vezoli J, van Pelt S, Schoffelen JM, Kennedy H, Fries P (2016) Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89:384–397.
- Miller KJ, Zanos S, Fetz EE, den Nijs M, Ojemann JG (2009) Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans. *J Neurosci* 29:3132–3137.
- Miller KJ, Honey CJ, Hermes D, Rao RP, den Nijs M, Ojemann JG (2014) Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. *Neuroimage* 85:711–720.
- Miller KJ, Schalk G, Hermes D, Ojemann JG, Rao RP (2016) Spontaneous decoding of the timing and content of human object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. *PLoS Comput Biol* 12:e1004660.
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784.
- Ray S, Maunsell JH (2010) Differences in gamma frequencies across visual cortex restrict their possible use in computation. *Neuron* 67:885–896.
- Rorden C, Hanayik T (2014) StimSync: open-source hardware for behavioral and MRI experiments. *J Neurosci Methods* 227:90–99.
- Rorden C, Bonilha L, Fridriksson J, Bender B, Karnath HO (2012) Age-specific CT and MRI templates for spatial normalization. *Neuroimage* 61:957–965.
- Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1997) Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *J Neurosci* 17:722–734.
- Vidal JR, Ossandón T, Jerbi K, Dalal SS, Minotti L, Ryvlin P, Kahane P, Lachaux JP (2010) Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and ERP response selectivity. *Front Hum Neurosci* 4:195.
- Winawer J, Kay KN, Foster BL, Rauschecker AM, Parvizi J, Wandell BA (2013) Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. *Curr Biol* 23:1145–1153.