J Neurophysiol 112: 856–869, 2014.
First published May 21, 2014; doi:10.1152/jn.00485.2013.

Tuning and spontaneous spike time synchrony share a common structure in macaque inferior temporal cortex

Chia-pei Lin (林佳霈),[1,3]* Yueh-peng Chen (陳嶽鵬),[1]* and Chou P. Hung (洪洲伯)[1,2]
[1]*Institute of Neuroscience and Brain Research Center, National Yang-Ming University, Taipei, Taiwan;* [2]*Department of Neuroscience, Georgetown University, Washington, District of Columbia; and* [3]*RIKEN Brain Science Institute, Saitama, Japan*

Submitted 5 July 2013; accepted in final form 18 May 2014

Lin CP, Chen YP, Hung CP. Tuning and spontaneous spike time synchrony share a common structure in macaque inferior temporal cortex. *J Neurophysiol* 112: 856–869, 2014. First published May 21, 2014; doi:10.1152/jn.00485.2013.—Investigating the relationship between tuning and spike timing is necessary to understand how neuronal populations in anterior visual cortex process complex stimuli. Are tuning and spontaneous spike time synchrony linked by a common spatial structure (do some cells covary more strongly, even in the absence of visual stimulation?), and what is the object coding capability of this structure? Here, we recorded from spiking populations in macaque inferior temporal (IT) cortex under neurolept anesthesia. We report that, although most nearby IT neurons are weakly correlated, neurons with more similar tuning are also more synchronized during spontaneous activity. This link between tuning and synchrony was not simply due to cell separation distance. Instead, it expands on previous reports that neurons along an IT penetration are tuned to similar but slightly different features. This constraint on possible population firing rate patterns was consistent across stimulus sets, including animate vs. inanimate object categories. A classifier trained on this structure was able to generalize category "read-out" to untrained objects using only a few dimensions (a few patterns of site weightings per electrode array). We suggest that tuning and spike synchrony are linked by a common spatial structure that is highly efficient for object representation.

spike correlations; visual object recognition; population coding; object representation; multielectrode array

TO UNDERSTAND THE ALGORITHM of visual object recognition, it is necessary to investigate how the activities of single neurons and neuronal populations are linked in the ventral visual cortex (DiCarlo et al. 2012; Edelman 2002; Hung and DiCarlo 2012; Logothetis 2010; Ohki et al. 2006; Op de Beeck et al. 2008; Schyns 2010). This link has been explored as pairwise interactions between neurons in macaque inferior temporal (IT) cortex, in terms of their tuning correlations (similarity in their stimulus preferences) and spike time correlations. Investigating this link is important because the specific correlational structure is postulated to affect redundancy and representational capacity, depending on the relationship between tuning and spike time correlation and on the size of the neuronal population (Averbeck et al. 2006; Cohen and Kohn 2011; Ecker et al. 2011; Gawne and Richmond 1993). Spike correlation within a narrow time window (spike time synchrony) is also postulated to enable downstream neurons to integrate information via coincidence detection and specific timing delays (Schwarzlose et al. 2005) and may be important for learning, e.g., via spike timing-dependent plasticity (Dan and Poo 2006; Masquelier and Thorpe 2007; Yao et al. 2004). The link between tuning and spike time synchrony in IT is controversial, with some evidence that neurons that are similarly tuned are also synchronized (Tamura et al. 2004) and other evidence suggesting no link (Aggelopoulos et al. 2005; Gawne and Richmond 1993; Gochin et al. 1991).

A possible explanation for these differing results is that tuning and spike time synchrony may share a common functional structure in which most nearby neurons are weakly correlated in tuning and spike timing, but some neurons have more similar tuning and stronger synchrony, as our laboratory and others have found in retina and multiple cortical areas (Chu et al. 2014; Fukushima et al. 2012; Harris et al. 2003; Katsuki et al. 2013; Stevenson et al. 2012). Also, investigation during muscle relaxation (Tamura et al. 2004) has the advantage of avoiding eye movement-related synchrony that may obscure the underlying structure (Ito et al. 2011; Rajkai et al. 2008). The spatial characteristics of this structure are unknown. Is the link between tuning and synchrony mainly determined by cortical distance, possibly due to traveling waves (Benucci et al. 2007; Huang et al. 2010), or is it more related to findings that neurons along a penetration (within a cortical column) prefer similar but slightly different features (Fujita et al. 1992; Sato et al. 2009; Tanaka 2003; Tsunoda et al. 2001), including in primary visual cortex (Chu et al. 2014)?

Also, what is the object coding capability of this structure? Previous reports from widely distributed IT populations (Hung et al. 2005) do not indicate how local correlations affect population coding, and investigating this issue is necessary to develop local learning rules for computational models of object recognition (Masquelier and Thorpe 2007; Pinto et al. 2009; Serre et al. 2007; Ullman et al. 2002). Specifically, it has been postulated that local structures may be useful for unsupervised generalization learning (learning to recognize complex features while ignoring minor variations, described as "cortically local subspace untangling" or "manifold learning") and should be understood at an intermediate level of abstraction (Bengio 2009; DiCarlo et al. 2012), whereas differences in local populations may be useful for discrimination (Fujita et al. 1992; Tanaka 2003). This structure may be particularly useful for unsupervised learning if it persists in the absence of sensory stimulation, e.g., if it can be measured during spontaneous activity. However, the link between spontaneous spike time correlations, tuning structure, and generalization has not been

* C.-p. Lin and Y.-p. Chen contributed equally to this work.

Address for reprint requests and other correspondence: C. P. Hung, Dept. of Neuroscience, Georgetown Univ., NRB EP-04, 3970 Reservoir Rd. NW, Washington, DC 20007 (e-mail: ch486@georgetown.edu).

made. Here, we asked whether the structure is consistent across stimuli, and whether this constraint on possible firing rate patterns would enable a downstream neuron to efficiently (via a small set of synaptic weights) detect whether a new stimulus belongs to a learned category ("within-category generalization"). This is unknown, and previous reports that synchrony strength is dynamic and stimulus- or context-dependent hint instead at poor consistency that might limit generalization (Anderson et al. 2006; Hirabayashi and Miyashita 2005; Takeuchi et al. 2011; Tallon-Baudry et al. 2004; Uchida et al. 2006).

We investigated these issues by testing whether the spatial pattern of tuning correlations (the pattern of sites whose responses covary across stimuli) is quantitatively linked to the spatial pattern of spike synchrony measured during spontaneous activity. We used 64-site arrays to systematically and simultaneously sample along both vertical (columnar depth) and horizontal directions up to our array width of 1.4 mm (Fig. 1), and we developed a general linear model (GLM) to assess the contributions of tuning and cortical distance to synchrony strength. To assess the structure's stability and to characterize the structure at an intermediate level of abstraction (extending upon previous reports of IT feature columns), we applied k-means clustering to tuning responses and quantified the consistency of the clustering across different stimulus sets, including animate vs. inanimate objects. We then quantified the within-category generalization performance of this structure, vs. other possible structures (e.g., penetration averages), via a linear classifier. Overall, the results suggest that tuning and spontaneous spike synchrony are indeed quantitatively linked in some pairs, forming a spatial structure that is highly consistent across stimuli. This spatial structure is highly efficient: just a few dimensions of the structure, based on a few clusters or sites, were sufficient for category generalization.

MATERIALS AND METHODS

Surgery and recording. All experimental procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of National Yang-Ming University.

We recorded from the right lateral surface of anterior IT cortex (AP 16) of four adult Taiwanese macaque monkeys (*Macaca cyclopis*) (Chu et al. 2014) under light neurolept anesthesia (Brown et al. 2011) (0.9 $\mu g \cdot kg^{-1} \cdot h^{-1}$ iv Fentanyl, 70%:30% N_2O/O_2, 0.25 mg/kg im droperidol, and 0.3–0.5% isoflurane) and muscle relaxation (1.2 $mg \cdot kg^{-1} \cdot h^{-1}$ iv rocuronium bromide). Several reports have shown robust IT responses under neurolept anesthesia (Fujita et al. 1992; Sato et al. 2009, 2013; Tsunoda et al. 2001; Wang et al. 2000; Yamane et al. 2006). We inserted five electrode arrays in four monkeys, where each "array" is defined as a separate experiment with a single insertion of a 64-site array [A8×8–5mm-200–200-413, NeuroNexus Technologies, 8 penetrations and 8 channels ("sites") per penetration, 0.2 × 0.2 mm spacing, 1.4 × 1.4 mm total, 15 μm thick]. Recording directly from the lateral surface avoids distortions from electrode bending, cortical compression, uneven sampling, and fine blood vessel damage. We estimate that the angular deviation at the surface was less than 8° ($\tan^{-1}$ 1/7 = 8.13°) in the electrode plane, because we could see and hear spikes at the beginning of array insertion on the bottom row of all eight penetrations. We estimate that, at worst, our deviation is 0.2 mm away from the center of the column. Also, deviations in the orthogonal plane would have restricted tuning correlations to pairs at the same depth, which would appear in the 64 × 64 tuning correlation matrix (see Fig. 4*D*) as a diagonal lattice, whereas we observed correlations between the most superficial and deepest depths.

Spikes (400–5,000 Hz) and local field potentials (1–300 Hz) were filtered (48 dB/octave) and continuously digitized at 24.4 kHz (RZ2, Tucker-Davis Technologies). Spikes were isolated using Wave_clus 2.0 (Hung et al. 2005; Kreiman et al. 2006; Quiroga et al. 2004). We sorted the units by gradually increasing the temperature of the clustering algorithm until less than 5% of interspike intervals were <2.5 ms. We grouped units by site as "multiunit activity" (typically 1–2 single units per channel) for tuning analyses; single-unit activity (SUA) yielded similar results. Spike synchrony analysis was based on SUA of the most active unit at each site.

Multiple detection of the same unit across different channels was tested via cross-correlation analysis with 1-ms histogram bins (Chu et al. 2014; Hung et al. 2007). Of 19,408 pairs tested, 25 (0.13%) were possible cases of multiple detection because they had narrow 1-ms peaks, defined by $B_0 > (B_1 + B_{-1})$ and $B_0 > 5\%$ of spike pairs, where B_0 is the height of the bin at 0, −1, or 1 ms (any of the 3) spike

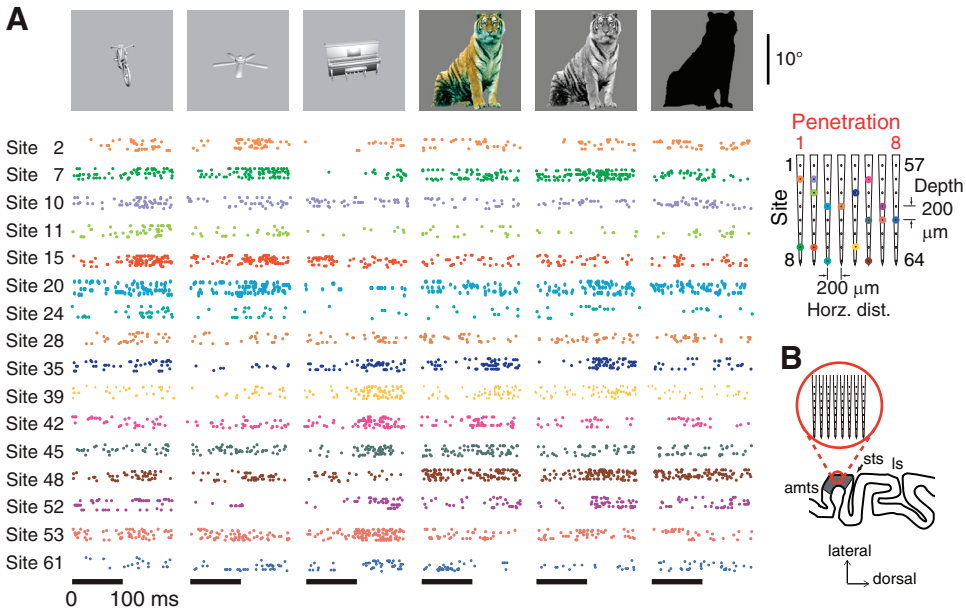

Fig. 1. Array recording of inferior temporal (IT) object responses. *A*, *top*: rasters show examples of reliable object-selective spiking responses at 16 sites of the 8 × 8 electrode array (*right*). Objects were serially presented (94 ms ON, 106 ms OFF with blank screen) in pseudorandom order for 10 repetitions. Two stimulus sets (*set 1*: 240 rendered gray-scale objects, *set 2*: 113 color, gray, and silhouette photographed objects) were presented as separate blocks. The 10° bar next to the *right-most* image shows the relative size of the stimuli. Black bars under rasters indicate ON period. *B*: we inserted arrays in the right anterior IT cortex (AIT, AP16), in a flat region of the lateral surface at least 3 mm from the superior temporal sulcus (sts). amts, Anterior middle temporal sulcus; ls, lateral sulcus. Note that array and cortex are not to scale.

time difference, and $[B_1, B_{-1}]$ are the heights of the neighboring bins. Of these 25 pairs, 16 were deemed not to be the same unit because they were separated by over 400 μm without intervening multiple-detection and because they showed 500-Hz peaks and long separations (up to 1.4 mm), indicating that they belong to a highly synchronized inhibitory network. The remaining 9 pairs had contamination of 13–98% (13–98 coincidences per 100 spike pairs) and were excluded by removing from each pair the unit with the smaller waveform (7 units total) and all of their pairs from analysis.

Stimuli. Stimuli consisted of 240 gray-scale rendered three-dimensional (3D) objects ("stimulus *set 1*"; *array 1* included two additional identical blank stimuli) and 113 color, gray-scale, and silhouette pictures of photographed objects (Sato et al. 2009) ("stimulus *set 2*"). Categories in *set 1* included animals ($n = 19$), faces ($n = 16$), plants ($n = 18$), foods ($n = 14$), tools ($n = 14$), vehicles ($n = 13$), appliances ($n = 13$), and furniture ($n = 21$), as well as other categories not included in classifier analysis. The monkeys had never seen these specific images, but they had prior exposure to a wide variety of natural and man-made objects. The full gallery of stimuli and neuronal responses will be shared at http://crcns.org.

To reduce brain pulsation, the stimuli and monkey were rotated 80° along the rostral/caudal axis. Stimuli were shown monocularly to the left eye, which was focused via contact lens upon a CRT monitor (ViewSonic P227fB) 57 cm away. Stimuli were positioned foveally via alignment of the optic disk (Garway-Heath et al. 2000). Object images, including background, subtended 20° (typical object sizes were 10° for *set 1* and 15° for *set 2*, see Fig. 1) and were presented in pseudorandom order, interleaved with a matching gray background at 5 Hz (94 ms ON/106 ms OFF, 85-Hz refresh, ambient illumination $<$ 0.4 Cd/m^2). Stimulus onset times were detected by photodiode [Matlab, Psychophysics Toolbox (Kleiner et al. 2007), and National Instruments Board]. To avoid response weakening due to the lack of microsaccades, we jittered the stimulus position at each video frame within a 0.2° window as is customary for anesthetized IT recordings (Sato et al. 2009). In a previous report in awake monkeys, a 0.2° shift in retinal position of 0.6° objects altered responses by $<$20% and did not alter tuning (DiCarlo and Maunsell 2003). Because our objects are larger, the variation is probably much less. Stimuli were presented for 10 repetitions (i.e., an entire set of 240 or 113 images was presented in random order, followed by presentation of the same set in a different random order, etc.). We presented the two stimulus sets via separate blocks lasting 8 and 4 min, respectively.

Tuning correlations. Tuning correlations were based on the *z*-normalized "evoked" (baseline-subtracted) trial-averaged multiunit activity (similar results were obtained with SUA) firing rates in the [100:200] ms poststimulus interval (i.e., one number per site per stimulus). Baseline was the [30:80] ms interval averaged across all stimuli (i.e., one number per site per block, not stimulus-specific). *Z*-normalization was across stimuli for each site. Results were similar for raw responses without baseline subtraction or *z*-normalization.

Even vs. odd trial correlation in Fig. 2*A* was calculated from all stimuli, including two sets. The analysis was corrected for the fewer number of trials (5 repetitions instead of 10) by assuming independent noise across trials. For Fig. 4*D*, we denoised the firing rate response by taking the first principal component of the temporally binned response (temporal principal component analysis), in which the bins are equivalent to peristimulus time histogram (PSTH) bins from 100:200-ms poststimulus onset, with 20-ms bin widths. The input to the principal component analysis was, therefore, a 5×240 array for 240 stimuli, i.e., we reduced the dimensionality of the stimulus representation from 5 to 1 by optimally weighting the contribution of the response given by different time bins (Richmond and Optican 1987). Results were similar for un-denoised responses.

Spontaneous spike time correlation (synchrony) analysis. We applied cross-correlation analysis to measure the differences in the spike times of two neurons (SUA) recorded from separate channels. This included spike jitter analysis to detect pairs with significant spike time correlations ("synchrony") and to estimate synchrony strength (see below) (Chu et al. 2014). Pairs were based on units with at least 2,000 spikes (results were similar at other thresholds). *Arrays 2–5* were recorded from 10 min of continuous spontaneous activity (blank screen). For *array 1*, we concatenated continuous periods (>30 s each) of spontaneous activity (blank screen at the start and end of each block). *Arrays 4* and *5* were excluded from further analysis because they had too few significant pairs (2 out of 244 tested at the standard 2,000-spikes threshold, 6 out of 942 tested at 1,000-spikes threshold), consistent with their many broken/inactive sites and with their narrower tuning (Fig. 2*B*). To avoid a dependence of the cross-correlo-

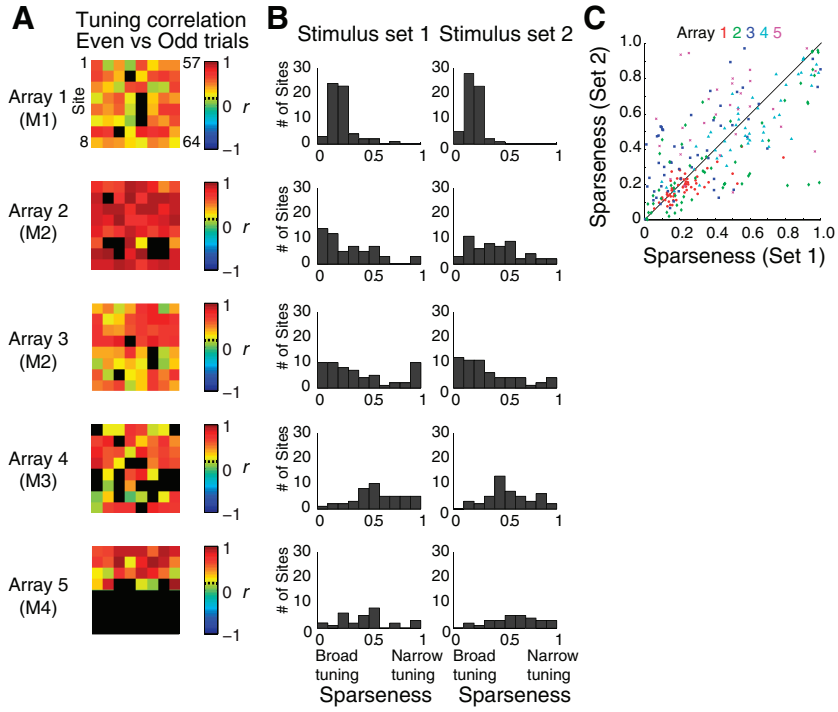

Fig. 2. Tuning reliability and tuning widths across 5 arrays. Response statistics across 5 array locations (5 recording sessions) in 4 monkeys are shown. *A*: tuning correlation between even and odd trials for all sites in each array, based on both stimulus sets. Tuning is based on evoked (baseline-subtracted) trial-averaged spiking activity measured in the [100:200] ms period, *z*-normalized for each site across stimuli. Baseline was the [30:80] ms period averaged across all stimuli (one number per site). Black indicates inactive channels. Nearly all active sites had significant tuning correlation across even vs. odd trials ($r > 0.16$, dotted line). *B*: distribution of sparseness (tuning width) across arrays and stimulus sets. Sites in *arrays 4* and *5* were more sparse (more narrowly tuned). According to Zoccolan et al. 2007, high sparseness is associated with poorer tolerance, which may explain their weaker consistency across stimulus sets. Sparseness was calculated according to Zoccolan et al. 2007 as $(1 - [\mathrm{Sum}(R_i/n)]^2/\mathrm{Sum}(R_i^2/n))/[1 - (1/n)]$, where R_i is the site response to the ith stimulus, and n is the number of stimuli in the set. *C*: comparison of sparseness across the two stimulus sets for all 250 sites, $r = 0.72$, $P < 10^{-39}$, color-coded by array.

gram height on recording duration, the histogram of interspike intervals between two neurons was divided by $\sqrt{(N_1 \times N_2)}$, where N_1 and N_2 are the number of spikes in the two spike trains (Hung et al. 2007; Smith and Kohn 2008).

Synchrony was significant ($P < 0.05$, "significant synchrony") for a pair of neurons if any bin in the [−5.5:5.5] ms interval of the cross-correlation histogram exceeded the significance threshold. The significance threshold was specific to each bin, defined by spike jittering. Other authors have previously applied this method to estimate the null distribution of spike time correlations that are slower than a predetermined time window (25 ms, results were similar at other sizes) (Anderson et al. 2006; Smith and Kohn 2008). The jitter predictor is better than shift or shuffle predictors, because it preserves the spike count within each 25-ms interval and spike timing interactions that are due to firing rate, while abolishing any remaining interactions faster than 25 ms.

Specifically, each spike train is divided into 25-ms intervals, and the spikes in each interval are resampled from other 25-ms intervals (other trials for the same stimulus) to preserve stimulus-specific firing rates, including rate changes faster than 25 ms. A jittered spike train would, therefore, be exactly the same as its original spike train, if both are binned at 25 ms, and PSTHs constructed from the jittered spike train would appear nearly identical to the originals, including firing rate changes faster than 25 ms. Note that this is different than simply shifting each spike by up to 25 ms, which would blur the binned spike train and the PSTHs. Because our synchrony analysis is based on spontaneous activity, the spikes are essentially drawn from a uniform distribution 25 ms wide, and so jittering simply preserves the spike count within each 25-ms interval.

We defined the significance threshold for each time bin as the maximum of the cross-correlogram values obtained across 220 spike jitter permutations in that specific time bin. We used 220 permutations because there are 11 bins in the [−5.5:5.5] ms interval, and so the likelihood that any bin exceeds the significance threshold is 1/220 $\text{bin}^{-1} \times 11$ bins (Bonferroni correction) = 11/220, i.e., $P = 0.05$. For a typical array with ~2,000 pairs, a $P = 0.05$ threshold would yield 100 spuriously "significant" correlations. To avoid such spurious correlations, we split the recording into two periods and required that the peak exceed the $P = 0.05$ threshold in both periods, effectively resulting in $P = 0.0025$, i.e., ~5 spuriously "significant" correlations per array. Because we observed extremely few suppressive interactions, we report only positive interactions.

We defined "synchrony strength" as the height of the cross-correlation histogram in the [−5.5:5.5] ms window after subtracting the 50th percentile of the jitter predictor. Synchrony strength, therefore, measures the likelihood of coincident spiking that is faster than the 25-ms jitter window.

Comparison between tuning and spike time correlation (synchrony). Comparison between tuning correlation and synchrony strength was based on Fisher-transformed tuning correlation and log-transformed synchrony strength (see Fig. 4), followed by Pearson correlation of these transformed values. GLM for synchrony strength was as follows:

$$\text{Log}(Y_i) = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2, \quad i = 1, 2, \ldots, N \qquad (1)$$

where Y_i is the synchrony strength for cell pair i, x_{i1} is cell pair i's Fisher-corrected tuning correlation, x_{i2} is the pair's cortical separation distance in mm (diagonal, vertical, or horizontal distance), and β are the coefficients for each factor, for N total pairs. β_0 is the offset for each array. N was 800 because we subsampled 100 pairs per distance (e.g., 0, 0.2, . . . 1.4 mm for horizontal and vertical distance). We applied cell-wise cross-validation (i.e., no reuse of the same cells for training and testing).

Tuning covariation structure (k-means clustering). We applied k-means clustering to array tuning responses to identify groups of sites whose responses covary across stimuli, to characterize the population at an intermediate level of abstraction (at the scale of cortical columns) and to examine the consistency of the clustering patterns across stimulus sets and across semantic (animate vs. inanimate) categories. Such consistency may be useful for generalization by allowing the same set of synaptic weights to be used to detect a feature across different stimuli. The input for the k-means clustering analysis was a matrix of 64 sites × 240 stimulus responses for *set 1* (fewer for inactive channels and *set 2*), i.e., the same *z*-normalized "evoked" (baseline-subtracted) trial-averaged responses that were used to calculate tuning correlations. The resulting output was in a space defined by image identities (i.e., 240 stimuli). Note that this is the reverse of what is commonly done in population coding studies, where population vectors corresponding to different images are analyzed within the space defined by the number of recorded neurons/sites. The clustering consistency across stimulus sets or across animate/inanimate categories was measured for each array via the Jaccard index ("*J*"), defined as the proportion of sites whose cluster label from one stimulus set was unchanged ("matched") when the analysis was rerun on the other stimulus set. *J* could thus range from 0 (no matching sites) to 1 (matching clusters at all sites). To estimate the likelihood that *J* can be explained by the sampling of the array (e.g., more broken channels in *array 5*), we shuffled the cluster labels within each penetration ("vertical shuffle") and across the entire array ("full shuffle"). Results were similar when we shuffled within each depth ("horizontal shuffle", not shown). However, we note that, if there were more than one cluster per penetration or row, the shuffling would destroy any pattern.

We estimated for each array the distribution of chance outcomes based on 1,000 shuffles. Because the k-means clustering algorithm does not assign each cluster the same label on each run, even when the clustering is identical, we needed to relabel the clusters following each run. For example, following each run with $k = 4$, we tried all $4! = 24$ combinations of relabeling of the 4 clusters, e.g., 4-1-3-2 instead of 1-3-2-4, to find the labeling that produced the highest *J*. This search among all possible combinations of relabeling produces a slight upward bias in *J*. The bias is specific for each array (because each array has different relative cluster sizes), chosen number of clusters *k* based on Bayesian information criterion (BIC) model selection, and shuffling method (see Fig. 5*C*). Because the time needed to test all possible combinations increases exponentially with *k*, we were only able to test shuffles up to $k = 8$. Based on results for $k = 2–8$, we expect that the chance distribution assumes a U-shaped dependence on *k*, with chance *J* reaching 1 when *k* equals the number of sites in the array.

Comparing tuning covariations vs. significant synchrony. We compared tuning-based k-means clustering with spatial patterns of significant spontaneous synchrony (binary 0 or 1) (see Figs. 3, 4 and 6). We defined "Clustering match probability" as the proportion (0 to 1) of neuronal pairs of a certain type, either pairs with significant synchrony or pairs that were asynchronous, that are pairs of neurons belonging to the same ("matched") cluster [e.g., prob (site A's cluster ID = site B's cluster ID) for all pairs AB that have significant synchrony; see Fig. 6]. To assess the spatial specificity of the link, we examined how clustering match probability depended on *k*. Additionally, for null hypothesis, we spatially scrambled cluster labels, which either preserved (vertical scramble) or ignored (full array scramble) topography. Null distributions were based on 200 scrambles. Shuffles and pairs were limited to sites with at least 2,000 spikes.

Classifier analysis. All stimuli for the classifier analysis were chosen from eight categories in *set 1* (gray-scale rendered 3D objects, i.e., no color or texture): animals, faces, plants, foods, tools, vehicles, appliances, and furniture as described in *Stimuli* in MATERIALS AND METHODS. Objects in other categories were not included. We measured category readout via a linear support vector machine classifier using Matlab's bioinformatics toolbox. Overall, classifier performance was based on the average performance across 20 independent runs. For each run, we built eight binary classifiers, one classifier for each

category, to detect whether the object belonged to that category vs. all other categories. Also for each run, we chose a random subset of 8 "training" objects and 5 unseen "test" objects per category (64 training objects and 40 test objects per run) from a pool of 13–21 objects per category. Different sets of eight training and five test objects were used for each run. The output category was chosen as the binary classifier (out of 8) whose output was farthest from its decision boundary (i.e., the classifier with the highest certainty).

To compare the effects of different methods of site clustering, we averaged the tuning responses (one number per site per stimulus, the same as for calculating tuning correlations) across sites within each cluster and used these average responses as input to the classifier. The same clusters were used for training and testing.

For example, for classifier analysis based on k-means clustering, the input to k-means clustering was a matrix of 64 sites × 200 stimuli, excluding the 40 test stimuli, and the input for classifier training was a matrix of k clusters × 64 stimuli. The same clusters were then used for the test phase of the classifier analysis. Thus, during testing, the classifier only had access to clustering patterns defined without test stimuli, to avoid over-fitting and circular logic (Kriegeskorte et al. 2009).

In Fig. 7*A*, we varied the dimensionality of the input by varying k, the number of clusters per array. For example, for $k = 3$ k-means clusters, the input for classifier training was a 15 × 64 matrix (5 arrays × 3 k-means clusters/array × 64 stimuli), and the input for classifier testing was a 15 × 40 matrix. For "Random Clusters," each array was uniformly divided into k clusters of sites. For within-penetration averages ("Penetration Averages") and within-depth averages ("Depth Averages"), only the mean responses of a random subset of penetrations or depths were used. $k = 1$ is the same as averaging the responses along a single randomly chosen penetration in each array, i.e., the input for classifier training had size 5 × 64 (5 arrays × 1 penetration per array × 64 training stimuli). For "Depth Averages," because *array 5*'s active sites were limited to the upper four depths, the input to the classifier for $k = 1$–8 had sizes 5, 10, 15, 20, 24, 28, 32, and 36 × the number of stimuli. Short dashed line in Fig. 7*A* was based on using, as input for classifier training, the matrix of 250 sites (all sites in all arrays) × 64 training stimuli, and as input for classifier testing, the matrix of 250 sites × 40 test stimuli (i.e., no clustering).

For Fig. 7*B*, the dimensionality of the input to the classifier was 5 arrays × 2 clusters per array ($k = 2$) × the number of stimuli. So, the training matrix had 10 × 64 dimensions for 64 stimuli, and the test matrix had 10 × 40 dimensions for 40 stimuli.

RESULTS

We recorded 359 neurons in 250 sites in response to 240 gray-scale rendered 3D objects ("*set 1*") and 113 photographed objects with color, gray-scale, and silhouette variants ("*set 2*"), totaling 10 blocks (5 array × 2 stimulus sets, i.e., 5 separate sessions across 4 monkeys). Responses were stimulus selective ("tuned"), strong, and reliable, consistent with previous reports of IT responses under neurolept anesthesia (Sato et al. 2009; Tsunoda et al. 2001). For the most responsive five sites of each array, mean baseline-subtracted response to the preferred stimulus was 45–124 spikes/s and signal-to-noise ratio was 22–75 means/SD. Tuning was consistent between even and odd trials ($r = 0.80$–0.96 for the top site in each array, mean $r = 0.57$ for all sites and two sets, Fig. 2*A*). Nearly all active sites had significant tuning correlation across even vs. odd trials ($r > 0.16$, dotted line). Significant responses appeared as early as 90 ms, and most responses had onset latencies of 100–120 ms, consistent with our laboratory's previous recordings in awake monkeys (Hung et al. 2005). Tuning width ["sparseness" (Zoccolan et al. 2007)] varied across sites and across arrays, but was highly consistent across stimulus sets ($r = 0.72$, $P < 10^{-39}$, Fig. 2, *B* and *C*).

Link between tuning and spontaneous synchrony strength for neuronal pairs. Figure 3 shows an example of spike time correlations (synchrony) measured from a pair of single units during spontaneous activity (blank screen). We focused on fast correlations (<25 ms) by using spike jittering (Smith and Kohn 2008) and calculated the binary significance and strength of synchrony within a ±5-ms coincidence window. Chance was $P = 0.0025$, or ~5 spuriously "significant" pairs out of ~2,000 tested per array. Of 6,340 pairs tested in *arrays 1–3*, 391 pairs (6%) had significant synchrony. *Arrays 4* and *5* had too few (≤6) significant pairs and so were excluded from further analysis.

For pairs with significant spontaneous synchrony, we found a quantitative link between tuning correlation and spontaneous synchrony strength [Fig. 4, *A–C*, red dots, *r_sig* (Pearson's r) = 0.62, 0.47, and 0.33, $P = 0.0015$, $P < 0.001$, and $P < 0.001$ for *arrays 1–3*, respectively], and these pairs had stronger tuning correlation and stronger synchrony (black marginals)

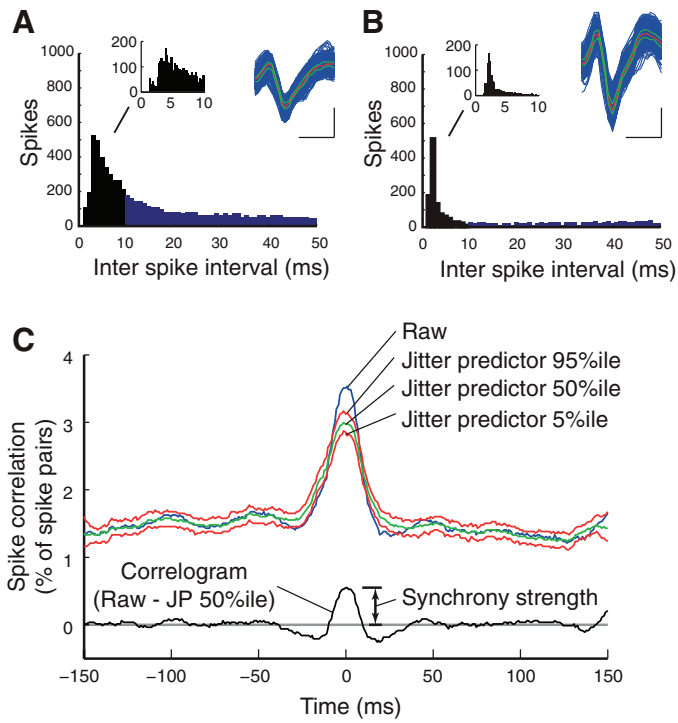

Fig. 3. Coincident spiking between a pair of IT neurons. *A* and *B*: interspike interval histograms for two units recorded from deep and superficial layers, respectively (1.6 and 0.6 mm from cortical surface; 1-ms bins). Both units were well-isolated, with <<5% of interspike intervals in the 2.5-ms refractory period (*left insets*, 0.2 ms bins). *Right insets* show example waveforms (scale bars 1 ms × 0.1 mV; red = median; green = 25 and 75th percentiles). Trains total 10,407 and 4,139 spikes, respectively. *C*: cross-correlogram histogram ("Raw", blue) based on cells *A* and *B*, showing spike time intervals between the two spike trains (abscissa) and each interval's percentage of all spike pairs (ordinate). Trains were a continuous (10 min) block of spontaneous activity (no stimuli). We defined coincident firing (synchrony) as a significant peak near 0 ms. Binary significance was based on whether any bin in the [−5.5:5.5] ms interval exceeded the significance threshold (upper red line, rank 1 of 220 jitter randomizations per bin × 11 bins = 1/20 randomizations, $P = 0.05$; 25-ms jitter window; 7-ms boxcar filter). Significant pairs passed this test twice (for two periods of spontaneous spike trains) for a total $P = 0.0025$. Synchrony strength was measured as the height of the jitter-corrected correlogram (Raw minus the Jitter predictor 50th percentile).

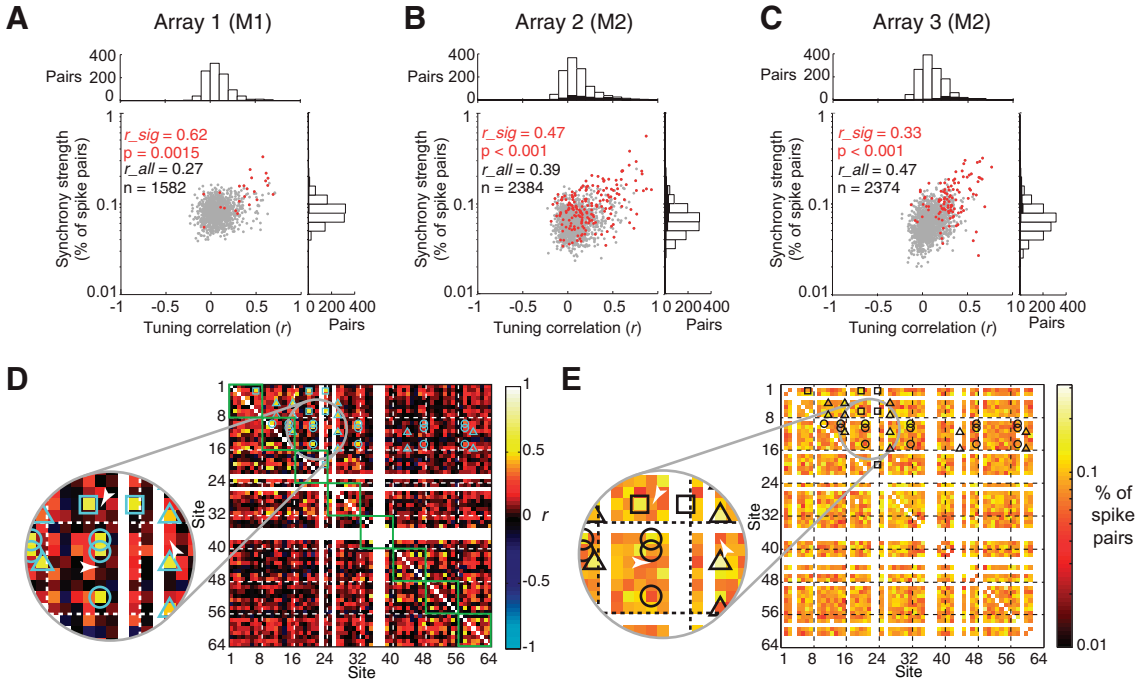

Fig. 4. Quantitative link between site-to-site tuning correlations and spontaneous synchrony strength. *A–C*: tuning correlation vs. synchrony strength among all pairs of sites in *arrays 1–3*. Pairs with significant synchrony (red dots, black marginals) are rare (6% of all pairs) and show quantitative link (*r_sig*, Pearson *r*) between tuning similarity and synchrony strength. The link for all pairs (*r_all*, red and gray dots) is weaker for *arrays 1* and *2* and is associated with weak tuning correlation and weak synchrony (white marginals). *D*: site-to-site tuning correlations for *array 1*, *set 1*. Stronger correlations (orange and yellow, $r > 0.3$) were rare, even for sites along the same penetration (green boxes), and appeared to belong loosely to groups of sites (marked by circles, triangles, and squares), as determined by eye based on stronger correlations within each group and weak correlations at intersections between sites from different groups (darker red or black, see arrowheads). White crosses indicate broken and inactive sites. *E*: site-to-site synchrony strengths, marked as in *D*. Intersections between groups (arrowheads) also had weaker synchrony. White crosses are broken sites and sites with less than 2,000 spikes.

compared with all pairs (white marginals). This link between tuning similarity and synchrony strength is consistent with an earlier report (Tamura et al. 2004). Additionally, because our measure of synchrony was based on spontaneous activity, it cannot be attributed to co-fluctuations in visually evoked responses. The link also cannot be due to eye movements (Ito et al. 2011; Rajkai et al. 2008), because it was measured under muscle relaxation. The rarity of significant synchrony, and the weaker link for *arrays 1* and *2* when the analysis was based on all pairs (*array 1*: $r\_all = 0.27$, $n = 1{,}582$ pairs, red and gray dots; *array 2*: $r\_all = 0.39$, $n = 2{,}384$ pairs), are consistent with previous reports that did not find a link between tuning and spike time correlation (Aggelopoulos et al. 2005; Gawne and Richmond 1993; Gochin et al. 1991). Thus, when examined as pairs of neurons, the link between tuning and spontaneous synchrony is sparse, appearing in a small subset of pairs.

What is the spatial structure of this link? We used a GLM to test the hypothesis that synchrony strength is mainly explained by cortical distance, and to test whether tuning correlation can explain synchrony strength after factoring for the contribution of cortical distance. The GLM factors were tuning correlation (Pearson correlation of the stimulus preferences of the two cells) and cortical separation distance. We subsampled the pairs so that the number of pairs was the same at each distance, and we applied cell-wise cross-validation (training and testing the GLM on separate populations of cells, i.e., no reuse of the same cells). For single-factor GLM, tuning correlation was a significant predictor ($P < 0.05$ to 0.005) and explained 7–18% of the variance across the three arrays (Table 1). Diagonal (site-to-site) cortical distance was also significant in the single-factor model, but explained a smaller portion of the variance (0.8–2.3%). Diagonal distance was significant in the combined factors model (combining tuning correlation and diagonal distance) for *array 3* only, but its β-coefficient remained small. Horizontal and vertical distances were not significant, and none of the distance factors was significant when we limited the GLM to pairs with significant synchrony (Table 2).

This link between tuning and spike synchrony strength, and the weakness of distance as a predictor, was also evident from examining the patterns of site-to-site tuning correlations and synchrony. Most pairs within the same penetration (pairs in which both neurons belonged to the same penetration, i.e., green boxes along diagonal in Fig. 4*D*) had weak tuning correlation [$r < 0.3$, red and black, for 569/713 (80%) pairs across the 5 arrays], consistent with a previous report (Sato et al. 2009). Only 58/713 (8%) pairs in the same penetration had tuning correlation stronger than $r = 0.5$ (yellow). These pairs with similar tuning (marked by circles, squares, and triangles) also tended to have stronger synchrony (yellow in Fig. 4*E*), and this relationship was mostly absent from other pairs in the same penetration.

The GLM result suggests that, at short distances up to our array width of 1.4 mm, spontaneous synchrony is associated more with tuning similarity than with cortical distance. In other words, spontaneous synchrony at the scale of neighboring IT feature columns is more related to the "minor" variations in tuning along a penetration (Fujita et al. 1992) than to topographic mapping of across short cortical distances. These results are consistent with our results in primary visual cortex (Chu et al. 2014), where cortical distance became a nonsignif-

Table 1. *GLM β-coefficients and explained variance, all pairs*

	β-Coefficients			Explained Variance of Single-Factor Mean, % (Probability)		
	Array 1	*Array 2*	*Array 3*	*Array 1*	*Array 2*	*Array 3*
β1 (tuning correlation)						
Single factor	0.22	0.36	0.52	7.1 (<0.005)	15 (<0.005)	18 (<0.005)
Combined factors	**0.21**	**0.36**	**0.50**	**(<0.05)**	**(<0.005)**	**(<0.005)**
β2 (diagonal distance)						
Single factor	−0.03	−0.04	−0.06	0.8 (<0.05)	1.1 (<0.01)	2.3 (<0.005)
Combined factors	**−0.02**	**−0.02**	**−0.06**	**NS**	**NS**	**(<0.005)**
β3 (horizontal distance)						
Single factor	−0.05	−0.03	−0.07	NS	NS	NS
Combined factors	**N/A**	**N/A**	**N/A**			
β4 (vertical distance)						
Single factor	−0.001	−0.02	−0.02	NS	NS	NS
Combined factors	**N/A**	**N/A**	**N/A**			

General linear model used tuning similarity and cortical distance to predict synchrony strength measured during spontaneous activity, with cross-validation across sites and even sampling across distance. The β-coefficient of each factor is shown for the single-factor model (regular font) and for the combined-factors model (boldface, β1 + β2, β1 + β3, and β1 + β4). NS, nonsignificant; N/A, not applicable.

icant predictor of spontaneous synchrony strength when combined with multiple tuning factors. Together, they suggest a closer association between tuning and spike timing than was previously reported.

Consistency of the structure across stimulus sets. Are the spatial patterns of tuning consistent across stimulus sets? This question of structural consistency and tolerance to image variations is important from the perspective of understanding how IT conveys generalizable information to downstream areas, i.e., as available to simple weighted summation decoding schemes, to recognize new objects in a learned category or a learned feature in a new category. The weak tuning correlations of individual pairs within a penetration (Sato et al. 2009) may pose a problem for generalization if tuning correlations were highly stimulus dependent. Ideally, if two neurons along a penetration were correlated for one set of stimuli, they should also be correlated for a different set of stimuli (if both sets contain the same feature), even if the stimuli belonged to a different semantic category. Indeed, the pairwise tuning correlations were very consistent across animate vs. inanimate sets, after factoring for the contribution of diagonal cortical distance (partial correlation $r = 0.32$ to 0.74 across 5 arrays, $P < 10^{-10}$ in each array). They were also consistent across *sets 1* and *2*, indicating that the consistency was not simply due to color or surface texture, which were absent from *set 1* (*arrays 1–4*: partial correlation $r = 0.37$ to 0.63, $P < 10^{-34}$ in each array; *array 5*: $r = 0.11$, $P = 0.034$). This indicates that the spatial pattern is highly consistent across stimuli, including semantic categories, at the level of pairs of neurons.

We next turned to a different type of analysis based on how larger populations ($n > 2$ units) covary across stimuli. It has been proposed that, to understand the algorithm of object recognition, more development is needed at intermediate levels of abstraction, at the level of populations of ~40k neurons that roughly comprise an IT feature column (DiCarlo et al. 2012). A useful analogy is that studying aerodynamics, rather than feathers, is needed to understand flight. Pairwise interactions are at a too-low level of abstraction, making the interpretation complicated because of neuronal heterogeneity. Analyzing at an intermediate level of abstraction would have the benefit of reducing the dimensionality (e.g., from n^2 pairwise interactions to n sites), making the problem more tractable. The question is whether the format of the intermediate abstraction (i.e., the method of dimensionality reduction) is appropriate, i.e., does it preserve the essential structure that supports generalized object recognition? Below, we examine whether neuronal covariation patterns, as defined by k-means clustering, are also consistent across semantic categories, and whether neuronal covariations preserve or lose object content.

An example illustrating the spatial covariation structure can be seen in the patterns of site-to-site tuning correlations (Fig. 4*D*) and spontaneous synchrony strengths (Fig. 4*E*) for *array 1*. Pairs with significant synchrony appeared to cluster according to specific groups of sites (example clusters are marked by

Table 2. *GLM β-coefficients and explained variance, significant pairs only*

	β-Coefficients			Explained Variance of Single-Factor Mean, % (Probability)		
	Array 1	*Array 2*	*Array 3*	*Array 1*	*Array 2*	*Array 3*
β1 (tuning correlation)						
Single factor	0.42	0.31	0.28	38.1 (<0.005)	23 (<0.005)	7.6 (<0.05)
β2 (diagonal distance)						
Single factor	0.19	−0.006	−0.20	NS	NS	NS
β3 (horizontal distance)						
Single factor	−0.12	−0.08	−0.27	NS	NS	NS
β4 (vertical distance)						
Single factor	0.21	0.29	0.02	NS	NS	NS

This table is the same as Table 1, except general linear model analysis is limited to pairs with significant synchrony. None of the distance factors was significant in the single-factor model, so we did not test combined factors.

squares, triangles, and circles). Notably, both tuning correlations and synchrony strength appeared to be weaker at intersections across groups (arrowheads), even if both sites were in the same penetration and each site was correlated and synchronized with other sites in the same penetration. This hints that the pairwise correlations were tied to larger clustering patterns and were not simply tied to the average correlation strength of each neuron. The other arrays had similar spatial patterns (not shown).

To examine the clustering consistency across stimuli, we applied k-means clustering to population tuning responses (e.g., 64 sites × 240 stimulus responses per site). We asked whether clustering patterns were consistent across different stimulus sets and different semantic categories (animate vs. inanimate), and we assessed via site shuffling whether this consistency was higher than expected from topographical organization. We quantified the consistency of the clustering as the Jaccard coefficient J, which measures, on a scale from 0 to 1, the fraction of sites whose label did not change across different stimulus sets or animate/inanimate categories. The clustering patterns were highly consistent across stimulus sets [Fig. 5*A*; k chosen via BIC model selection; weaker clustering for *arrays 4* and *5* may be due to fewer channels and narrower tuning (Zoccolan et al. 2007)]. This indicated that the consistency was not simply due to color or texture, because both were absent in *set 1*.

The clustering patterns were also consistent across animate vs. inanimate objects (Fig. 5*B*), suggesting that the patterns were not due to a difference in semantic category coding across clusters. We considered that such clustering consistency could occur if one cluster was tuned to animate and the other to inanimate objects, but the cluster tuning preferences indicated that they were not. This shape-based, rather than semantic category-based, tuning is consistent with a recent paper (Baldassi et al. 2013) and is in explicit disagreement with Kiani et al. (2007). The clustering pattern was also not simply due to differences in tuning width (sparseness), which produced different clustering patterns that were not significantly matched.

Clustering consistency J was high for a wide range of k clusters (Fig. 5*C*), and model selection (BIC) indicated that the optimal cluster sizes (triangles) were within the range of k tested. We estimated the 95th percentile of J's null distribution and J's significance by shuffling sites within each penetration ("vertical shuffling", p_v) and across the full array ("full shuffling", p_f) for *set 2* and for inanimate objects. J was significant ($P < 0.05$, and in most cases $P < 0.001$) for almost all k, showing that the structure was consistent across a range of spatial scales. Such high consistency could also occur if two neighboring columns occupied different sites along the same penetration. However, we saw such overlaps at virtually all penetrations, and the consistency was sustained for a wide range of k beyond $k = 2$, up to at least $k = 8$ (Fig. 5, *D* and *E*), whereas previous reports examined clustering that is on the order of $k = 2$ neighboring columns (Sato et al. 2009; Tanaka 2003; Tsunoda et al. 2001) or coarser (Sato et al. 2013).

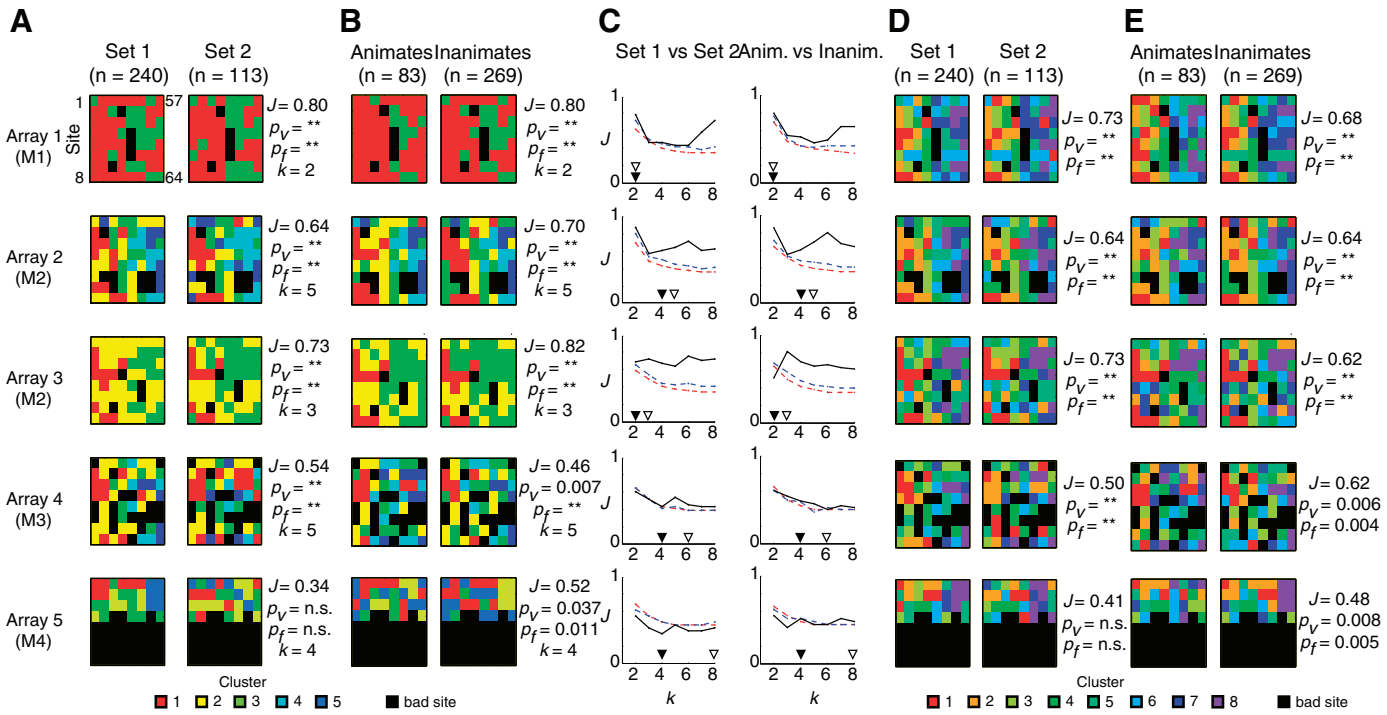

Fig. 5. Consistency of site clustering patterns across stimuli. *A*: consistency of clustering patterns (see color legend at bottom) across stimulus sets, shown for $k = 2$ to 5 clusters based on Bayesian information criterion (BIC) of each array. Input to k-means clustering was the same as for computing tuning correlations, e.g., trial-averaged z-normalized tuning matrix of 64 sites × 240 stimuli. Colors indicate sites belonging to different clusters. Black indicates broken or inactive sites. Jaccard coefficient (J) indicates proportion of sites whose cluster label were unchanged ("matched") when reanalyzed with the other stimulus set. J ranges from 0 (no sites with matched clusters) to 1 (all sites matched). Match significance was determined by vertical within-penetration shuffling (p_v) and by full array shuffling (p_f) of cluster labels, based on 1,000 shuffles (**$P < 0.001$). *B*: clustering consistency between animate (*left*) vs. inanimate (*right*) categories. Animates comprise faces and animals from both stimulus sets, and inanimates comprise all other stimuli. *C*: dependence of J on k. Red and blue dashed lines indicate 0.05 significance threshold for p_f and p_v, respectively. Filled and open triangles indicate optimal number of clusters for *sets 1* and *2*, respectively, as estimated by BIC. *D* and *E*: clustering consistency at $k = 8$ for *set 1* vs. *set 2*, and for animates vs. inanimates.

These results indicate that the local functional structure is insensitive to minor feature differences (color, surface texture) or to semantic category. The consistency of this structure across multiple scales ($k = 2–8$) is consistent with the idea of a "smooth" representation that is useful for learning (e.g., in computer models via gradient descent). Because all of these stimuli were novel to the monkeys, it suggests that this smoothness extends to novel stimuli as well. Overall, the clustering consistency hints that synaptic weights based on such clustering patterns may be useful for generalization, and the scaling of this consistency beyond $k = 2$ suggests that the structure is smooth across multiple spatial scales. We speculate that synaptic weights might, therefore, focus at different scales to serve different levels of discrimination and generalization.

Link between clustering pattern and spontaneous synchrony. The GLM analysis showed a quantitative link between tuning and spontaneous spike synchrony for significant pairs, and that this link was not due to cortical distance. Also, tuning clustering patterns were highly consistent across stimulus sets. These combined results suggest that synchrony may also form spatial groups of sites within a penetration. Examples of such groups are hinted by the site-to-site correlations in Fig. 4, *D* and *E*, for *array 1*.

Figure 6*A* illustrates examples of this link between synchronized groups of sites (black and white lines) and tuning-based clustering patterns for *arrays 1* and *2* (same as Fig. 5*D*, $k = 8$). Lines of the same color, i.e., groups of synchronized neurons, link sites that belonged to the same k-means cluster (*left*, black lines link red sites, white lines link orange sites, for *array 1*; *right*, same for *array 2*). Consistent with the GLM result that synchrony strength was unrelated to short cortical distance, the different synchronized groups were interdigitated across distinct sets of sites within the same penetrations.

To quantify the link between clustering patterns and synchrony, we defined clustering match probability as the probability that neurons in a synchronized pair belong to the same k-means cluster (Fig. 6*B*). Clustering match probability was surprisingly high, occurring at >50% on average at $k = 8$ (Fig. 6*C*, chance 12.5%, dashed line). Controls based on scrambling of the cluster labels across the full array, or within each vertical penetration (in depth) or within each row support that distance cannot explain this link ($P < 0.001$ for group and all individual comparisons). The Match probability for nonconnected pairs was roughly at chance. This link between tuning and spike synchrony was found across a wide range of k (Fig. 6*D*). Overall, these results show that tuning and spontaneous spike timing patterns are very tightly coupled in IT.

Object coding capability of this structure. The tuning and spike synchrony structure, and its consistency across stimuli, hints that the structure of the local population code may be useful for generalization. To test this, we used a linear classifier to measure the ability of downstream areas to read out the category membership of "unseen" objects within learned cat-

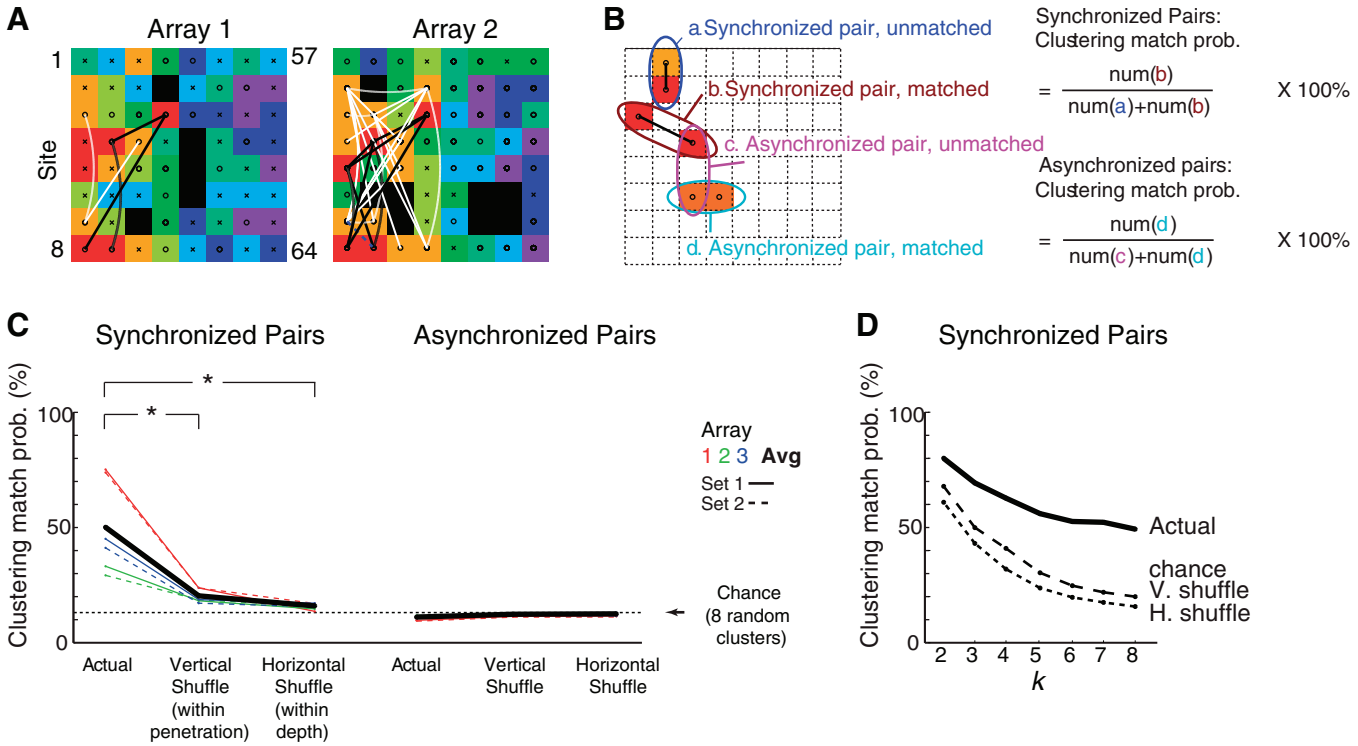

Fig. 6. Spatial specificity of match between spike synchrony and tuning. *A*: example overlays of spontaneous spike synchrony and tuning-based clusters. Lines indicate examples of pairs with significant spike synchrony ("functionally connected") during spontaneous activity. White and black lines show synchrony within each of two k-means clusters in the same cortical column. Blue dashed line between *sites 7* and *16*, *array 2*, shows example of synchrony across different clusters ("unmatched"). *B*: method of comparing spontaneous synchrony vs. clustering. We grouped pairs into four types "*a*" to "*d*" depending on synchrony significance (synchronized vs. asynchronized) and whether both sites in a pair had the same cluster ID (matched vs. unmatched). We then defined clustering match probability as the percentage of pairs that had matched clusters. This was calculated separately for synchronized pairs only [$b/(a + b)$] and for asynchronized pairs only [$d/(c + d)$]. *C*: clustering match probability for actual data, vertical, and horizontal within-depth shuffling (i.e., controlling for horizontal and vertical cortical distance, respectively), with $k = 8$ clusters. Colors indicate *arrays 1–3*, solid/dashed lines are stimulus *sets 1–2*, and black indicates group mean. *$P < 0.001$ for group and all individual comparisons. Synchrony analysis was limited to units with at least 2,000 spikes. *D*: dependence of clustering match probability on number of clusters k. Dashed and dotted lines indicate expected clustering match probability for vertical and horizontal (within depth) shuffling of cluster labels.

egories ("within-category generalization") via a simple weighted summation decoding scheme. The specific pattern of weights determines which sites are averaged for input to the classifier. For example, $k = 4$ clusters per array means that, for each array, we averaged the responses within each of four k-means clusters, four random penetrations (all responses averaged within each penetration), four random depths (all responses averaged within each depth), or divided the array into four random groups of sites. Specifically, are population responses that are more similar to the k-means clustering patterns [aligned to tuning covariations, i.e., "choristers" (Carandini 2014; Kenet et al. 2005)] better for generalization, compared with alternative patterns? Also, how does generalization performance vary with the number of clusters?

Classifier performance was best for site clusters based on k-means clustering. Surprisingly, just two clusters ($k = 2$ per array, 40% correct, chance was 12.5%, for 8 categories) were sufficient to boost performance to nearly the maximum readout performance (43%), which was based on all sites without site averaging (Fig. 7*A*). Performance was above chance for all eight categories (Fig. 7*B*). Using random clusters, penetrations, or depths decreased performance by nearly one-half for $k = 2$ [decreased from 40% correct down to 27–31% correct vs. chance level 12.5% correct, i.e., $(27 - 40)/(40 - 12.5) = -47\%$]. This suggests that the spatial structure of tuning and spike synchrony is helpful for generalization, because two k-means clusters per array ($k = 2$) were sufficient to produce high performance. In contrast, one would expect performance to increase with k simply because, with low k, each random cluster pools over many sites, while at larger k, each cluster pools over fewer sites (and therefore results in lower loss of tuning information).

Next, to fairly compare among different clustering methods with the same k, we limited all clusters to at most eight sites each. The performance of k-means clustering (32% at $k = 2$ clusters, 8 random sites within each cluster) was still higher than random clustering (29%). It is remarkable that performance is so good for k-means, even at $k = 2$, because all sites in a cluster are weighted equally, even sites at the edge of the cluster.

In fact, performance was even better when the classifier input was limited to sites that had the most similar tuning to the cluster average (i.e., "choristers," defined without test objects). Four "choristers" (4 sites), each chosen from one of four k-means clusters ($k = 4$) per array, were sufficient for performance to reach plateau (42.4% correct, Fig. 7*C*, black). Performance was also good for "choristers" defined as sites with the highest average pairwise tuning correlation (36.1% for 4 sites per array, reaching plateau with 42.8% correct at 5 sites per array, brown). Performance was much poorer for "soloists," sites that were less correlated with other neurons (near the 50th percentile of sites rank-ordered by average pairwise tuning correlation, violet).

Overall, the classifier results suggest that correlated and covarying tuning responses (responses that lie along a local population's low-dimensional manifold, "choristers") support generalization performance, at least for categories that are highly distinct from each other. This has been postulated by several reports (Averbeck et al. 2006; DiCarlo et al. 2012; Ecker et al. 2011; Fujita et al. 1992; Tanaka 2003), and several studies have investigated population coding by IT neurons at coarse spatial scales (Hung et al. 2005; Rolls and Tovee 1995; Sato et al. 2013). However, to our knowledge, our study is the first to explicitly test this for correlated responses within and across neighboring columns, from which most nearby neurons

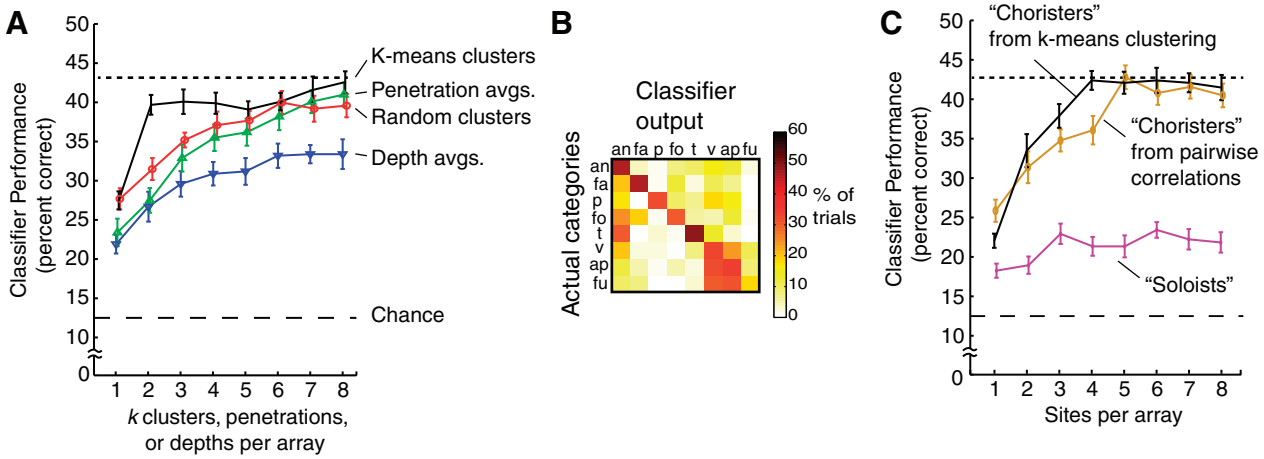

Fig. 7. Category generalization performance vs. clustering method. *A*: classifier performance vs. number of clusters for different site clustering methods. Tuning responses were grouped as k-means clusters (black), random clusters (red), within-penetration averages (green), and within-depth averages (blue). Input to the classifier was based on the average response within each cluster (same clusters for training and testing). K-means clusters were based on 200 objects in *set 1*, excluding the test objects in that run. Performance was based on 20 runs. In each run, 8 binary classifiers were trained on each of 8 categories (64 training objects, 8 objects per category), then tested on categorization of novel objects from the same categories (5 objects per category). A random group of 64 training and 40 test objects were chosen for each run, from a pool of 13–21 objects per category, excluding objects in other categories. k indicates number of clusters or random penetrations or depths per array. For example, $k = 2$ means that the input to classifier training was a 10×64 matrix, for 5 arrays $\times$ 2 clusters/array $\times$ 64 training objects. Dotted line indicates performance of classifier trained, tested on all 250 sites without clustering. Dashed line is chance (12.5%). *B*: confusion matrix showing pattern of mistakes made by the classifier for $k = 2$ k-means clusters per array. Each row indicates the actual category shown to the monkey, and each column indicates the classifier predictions. Categories include animals (an), faces (fa), plants (p), foods (fo), tools (t), vehicles (v), appliances (ap), and furniture (fu). *C*: performance of "choristers" vs. "soloists." "Choristers" are sites whose tuning is most similar to that of each k-means cluster (black, same clusters as in *A*), or sites with the highest average pairwise tuning correlation (brown), for nontest objects. "Soloists" are sites that have medium correlation with other sites (at the 50th percentile of sites rank-ordered by average pairwise tuning correlation for nontest objects, violet).

receive their input, and to link this to spatial patterns of spontaneous spike time synchrony.

DISCUSSION

Overall, these results improve our understanding of the spatial structure within and across neighboring IT cortical columns, in terms of spike time synchrony and object category coding. First, the results showed a quantitative link between tuning and spontaneous spike time synchrony, after factoring out the contribution of cortical distance. Although the link was limited to a small subset (6%) of neuronal pairs that had significant spike synchrony, significant synchrony was also linked to spatial patterns of tuning covariation across many neurons, based on k-means clustering. Second, these clustering patterns were highly consistent across stimulus sets, including animate vs. inanimate categories, suggesting that this spatial structure constrains possible firing rate patterns. Third, population responses aligned to this spatial structure enabled a linear classifier to generalize to new objects within learned categories ("within-category generalization") using only a few neurons per array. This link between tuning, spike time synchrony, and object coding by local neuronal populations should be useful for theories and models of unsupervised generalization learning for object recognition.

Why was cortical distance nonsignificant? The GLM results showed that the link between tuning and spike time synchrony was not simply due to cortical distance. Although tuning correlation and synchrony strength do decline with distance, consistent with previous reports, tuning correlation remained a significant contributor to synchrony strength, after factoring out the effect of cortical distance up to the 1.4-mm width of our array. For two of three arrays, cortical distance was not significant when combined with tuning correlation in the GLM. Our results show that, at short distances, synchrony is closely linked to the minor variations in tuning along a penetration, expanding on previous reports of such tuning variations (Fujita et al. 1992; Sato et al. 2009).

Several reasons may explain why distance was not a significant factor. First, we measured synchrony at short distances. To our knowledge, no study has examined IT spike synchrony at long distances, although oscillatory synchrony of ~17 Hz has been reported for electrodes separated by 3–6 mm and was not significant at longer distances (Tallon-Baudry et al. 2004). Second, our use of spike jitter correction and muscle relaxation avoided slower correlations that may be associated with traveling waves and eye movement (Huang et al. 2010; Ito et al. 2011; Rajkai et al. 2008). Third, recording synchrony during spontaneous activity avoided stimulus-related differences in synchrony, which may be a separate signal on top of the inherent structure in our report (Hirabayashi and Miyashita 2005). When we applied these methods in V1, we also found that when a combination of V1 tuning properties was used to predict spike synchrony, cortical distance was no longer a significant factor (Chu et al. 2014).

Why does local redundancy help object coding? Previous reports suggested that redundancy, in the form of correlated tuning within an IT cortical column, may aid in generalizing across minor variations of a feature, whereas differences in tuning may be useful for discriminating between similar objects (Fujita et al. 1992; Sato et al. 2009; Tanaka 2003). Our classifier results confirm the usefulness of local redundancy for generalization, but they do not test the usefulness for discrimination. It has been postulated that categorization and identification may be computationally equivalent (Hung et al. 2005) and so may not require separate mechanisms (however, see Sugase et al. 1999).

The categories we tested for classification were diverse, and we speculate that discriminating categories that are more similar (e.g., generalizing across different images of the same individual, vs. other individuals) may require denser sampling along the local population's low-dimensional manifold (sampling more clusters or sites per array). Indeed, learning the appropriate representational densities for different categorization tasks, and how to generalize to more difficult tasks with few training examples, is one of the challenges of unsupervised generalization learning (Kriegeskorte et al. 2008; Poggio and Bizzi 2004). Whether representational density is linked to the difficulty of generalization remains to be explored. Our results suggest that, at least for categories that are highly different, downstream neurons may not need to sample densely. As few as four sites per 1.4-mm-wide array were sufficient to reach the same generalization performance as listening to all neurons in the array, for sites that were most aligned with the clusters' average tuning [neurons along the low-dimensional "manifold" of the local feature subspace (DiCarlo et al. 2012), i.e., "choristers" (Carandini 2014; Kenet et al. 2005)]. A useful analogy here is that of a pipe cleaner: although the bristle directions are heterogeneous like individual neurons, the spine is what gives the pipe cleaner its overall shape, and sampling a few points along the spine is sufficient to reconstruct its overall contour. This suggests that the representation may be highly efficient for generalization, in that downstream neurons may be able to find the right neurons to listen to, based only on tuning and spike timing covariations.

Implications for semantic vs. shape-based coding. Whether the IT code is primarily based on semantic category or shape has emerged as a contentious issue. Although many reports have shown that IT encodes complex shapes (Desimone et al. 1984; Fujita et al. 1992; Meyers et al. 2008) and 3D structure (Hung et al. 2012; Janssen et al. 2003; Yamane et al. 2008), recent reports in monkeys and humans have highlighted the categorical and semantic structure in the code (Bell et al. 2009, 2011; Kiani et al. 2007; Kriegeskorte et al. 2008; Sato et al. 2013). This issue is difficult to resolve, because objects that belong to the same category also tend to have similar shapes. Previous reports approached this issue by examining how stimulus responses were clustered, e.g., whether animate and inanimate object responses were strongly segregated. We took a different approach by asking which site responses were clustered, i.e., whether the grouping of sites by response covariation across stimuli depended on semantic category. Our intuition is that, if the coding is based on semantic category, then the covariation patterns should be different for animate vs. inanimate objects, because (extending upon Hebb's hypothesis) only cells that collectively perform the same function should wire together and show synchronized spiking activity, even in the absence of visual stimulation (Hung and DiCarlo 2012). However, we observed that the site clustering patterns were the same for both animate and inanimate stimuli, across a wide range of clustering sizes k. We also observed the same site clustering patterns for both gray-scale nontextured stimuli

(*set 1*) and for colored photographs (*set 2*). This implies that the underlying structure in our recordings was not based on semantic category, color, or texture, but rather shape. This result is consistent with a recent report (Baldassi et al. 2013) that the IT code is primarily shape based, rather than semantic category based.

Implications for microstimulation, neuroimaging, and modeling. These results suggest that the link between tuning and spontaneous correlations is not limited to large-scale networks (Fukushima et al. 2012; Kenet et al. 2003), but also includes interactions within local networks (neighboring columns), which comprise the bulk of synaptic connections. Our results show that this inherent structure is linked to the object-coding capability of local patches of cortex. The spatial precision of this structure is consistent with a recent study showing that microstimulation activates highly specific networks within a cortical volume, and that advancing the electrode by just 30 μm activates non-overlapping networks (Histed et al. 2009).

In human neuroimaging, recent studies have linked the temporal correlation of neighboring voxels (regional homogeneity "ReHo" and "Hcorr") to several neurological diseases (Deshpande et al. 2009; Zang et al. 2004). Specifically, local voxel correlations in the fusiform face area are quantitatively linked to the severity of face discrimination behavior in autistic subjects (Jiang et al. 2013). The link between local tuning and spontaneous synchrony in this report and our laboratory's recent report in V1 (Chu et al. 2014) strengthens the neuronal basis of these measures and may be useful for investigating the mechanisms (e.g., an imbalance in inhibition) that are thought to underlie disease.

A possible limitation of these results is that they were recorded under light anesthesia. In awake monkey IT, synchrony is dynamic and layer specific, depending on the behavioral state of the animal (Takeuchi et al. 2011). Our data also hinted at layer specificity, consistent with a feed-forward interaction from *layer 4* to *layers 2/3*, but we did not verify this anatomically. Although awake recordings avoid the issue of interpreting responses under anesthesia, previous reports found that, in awake monkeys, LFP and synchrony are time-locked to saccades and microsaccades (Hermens and Walker 2010) and are postulated to spread outward from foveal representations (Ito et al. 2011; Rajkai et al. 2008). Recording under light anesthesia and muscle relaxation, as in this study, allows the contribution of eye movement to be considered separately.

These results should be useful for theories and models of unsupervised generalization learning. In particular, models based on spike timing-dependent plasticity (Masquelier and Thorpe 2007) rely on close temporal association between units that have similar tuning. A challenge in building these models is that temporal associations may be highly dependent on the specific pattern of visual experience (Li and DiCarlo 2008). Our results show that such tuning and timing associations for generalization may be maintained in the presence of visual stimulation from unrelated object categories and even in the absence of visual stimulation. Moreover, the consistency of the tuning covariation patterns across different spatial scales, and its ability to support category generalization, support previous reports that a smooth representation is useful for learning [i.e., a low-dimensional manifold for complex feature representation (DiCarlo and Cox 2007; DiCarlo et al. 2012; Poggio and Bizzi 2004; Tanaka 2003; Wang et al. 1996)]. Our results hint that this smooth structure may persist as synchronous patterns in the absence of visual stimulation. We also note that evoked responses had spatial patterns of spike synchrony that were virtually identical to those during spontaneous activity. In this report, we focused only on patterns during spontaneous activity to exclude the possibility of contamination by stimulus-related artifact. It will be interesting to test whether such spontaneous patterns persist during sleep, supporting "offline" learning of complex shapes (Baeck et al. 2012, 2014).

ACKNOWLEDGMENTS

We thank C. Baker, J. DiCarlo, M. Riesenhuber, and M. Tanifuji for helpful discussions and C. Tan, T. Sato, T. Serre, G. Kreiman, T. Poggio, and M. Tanifuji for the stimuli.

GRANTS

This study was supported by the Taiwan Ministry of Education Five-Year Aim for the Top University Plan, NSC-96-2811-B-010-501 and NSC-97-2321-B-010-007, and by Georgetown University Medical Center GD4235-619.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: C.-p.L., Y.-p.C., and C.P.H. performed experiments; C.-p.L., Y.-p.C., and C.P.H. analyzed data; C.-p.L., Y.-p.C., and C.P.H. interpreted results of experiments; C.-p.L., Y.-p.C., and C.P.H. prepared figures; C.-p.L. and C.P.H. drafted manuscript; C.-p.L., Y.-p.C., and C.P.H. edited and revised manuscript; C.-p.L., Y.-p.C., and C.P.H. approved final version of manuscript; C.P.H. conception and design of research.

REFERENCES

Aggelopoulos NC, Franco L, Rolls ET. Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. *J Neurophysiol* 93: 1342–1357, 2005.

Anderson B, Harrison M, Sheinberg DL. A multielectrode study of the inferotemporal cortex in the monkey: effects of grouping on spike rates and synchrony. *Neuroreport* 17: 407–411, 2006.

Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366, 2006.

Baeck A, Rentmeesters N, Holtackers S, Op de Beeck HP. The effect of sleep in perceptual learning with complex objects. *Vision Res* 99: 180–185, 2014.

Baeck A, Windey I, Op de Beeck HP. The transfer of object learning across exemplars and their orientation is related to perceptual similarity. *Vision Res* 68: 40–47, 2012.

Baldassi C, Alemi-Neissi A, Pagan M, DiCarlo JJ, Zecchina R, Zoccolan D. Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Comput Biol* 9: e1003167, 2013.

Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG. Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *J Neurophysiol* 101: 688–700, 2009.

Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RB, Ungerleider LG. Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J Neurosci* 31: 12229–12240, 2011.

Bengio Y. Learning deep architectures for AI. *Found Trends Mach Learn* 2: 1–127, 2009.

Benucci A, Frazor RA, Carandini M. Standing waves and traveling waves distinguish two circuits in visual cortex. *Neuron* 55: 103–117, 2007.

Brown EN, Purdon PL, Van Dort CJ. General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu Rev Neurosci* 34: 601–628, 2011.

Carandini M. Soloists and choristers in a cortical population. In: *Computational and Systems Neuroscience Workshop: Scalable Models for High-Dimensional Neural Data*. Snowbird, UT: COSYNE, 2014.

Chu CCJ, Chien PF, Hung CP. Tuning dissimilarity explains short distance decline of spontaneous spike correlation in macaque V1. *Vision Res* 96: 113–132, 2014.

Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.

Dan Y, Poo MM. Spike timing-dependent plasticity: from synapse to perception. *Physiol Rev* 86: 1033–1048, 2006.

Deshpande G, LaConte S, Peltier S, Hu X. Integrated local correlation: a new measure of local coherence in fMRI data. *Hum Brain Mapp* 30: 13–23, 2009.

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.

DiCarlo JJ, Cox DD. Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341, 2007.

DiCarlo JJ, Maunsell JH. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89: 3264–3278, 2003.

DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 73: 415–434, 2012.

Ecker AS, Berens P, Tolias AS, Bethge M. The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31: 14272–14283, 2011.

Edelman S. Constraining the neural representation of the visual world. *Trends Cogn Sci* 6: 125–131, 2002.

Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343–346, 1992.

Fukushima M, Saunders RC, Leopold DA, Mishkin M, Averbeck BB. Spontaneous high-gamma band activity reflects functional organization of auditory cortex in the awake macaque. *Neuron* 74: 899–910, 2012.

Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 107: 1809–1815, 2000.

Gawne TJ, Richmond BJ. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13: 2758–2771, 1993.

Gochin PM, Miller EK, Gross CG, Gerstein GL. Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp Brain Res* 84: 505–516, 1991.

Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsaki G. Organization of cell assemblies in the hippocampus. *Nature* 424: 552–556, 2003.

Hermens F, Walker R. What determines the direction of microsaccades? *J Eye Mov Res* 3: 1–20, 2010.

Hirabayashi T, Miyashita Y. Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. *J Neurosci* 25: 10299–10307, 2005.

Histed MH, Bonin V, Reid RC. Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron* 63: 508–522, 2009.

Huang X, Xu W, Liang J, Takagaki K, Gao X, Wu JY. Spiral wave dynamics in neocortex. *Neuron* 68: 978–990, 2010.

Hung CC, Carlson ET, Connor CE. Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74: 1099–1113, 2012.

Hung CP, DiCarlo JJ. Ultrafast decoding from cells in the macaque monkey. In: *Visual Population Codes: Toward a Common Multivariate Framework for Cell Recording And Functional Imaging*, edited by Kriegeskorte N and Kreiman G. Cambridge, MA: MIT, 2012, p. 275–305.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866, 2005.

Hung CP, Ramsden BM, Roe AW. A functional circuitry for edge-induced brightness perception. *Nat Neurosci* 10: 1185–1190, 2007.

Ito J, Maldonado P, Singer W, Grun S. Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cereb Cortex* 21: 2482–2497, 2011.

Janssen P, Vogels R, Liu Y, Orban GA. At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron* 37: 693–701, 2003.

Jiang X, Bollich A, Cox P, Hyder E, James J, Gowani SA, Hadjikhani N, Blanz V, Manoach DS, Barton JJS, Gaillard WD, Riesenhuber M. A quantitative link between face discrimination deficits and neuronal selectivity for faces in autism. *Neuroimage Clin* 2: 320–331, 2013.

Katsuki F, Qi XL, Meyer T, Kostelic PM, Salinas E, Constantinidis C. Differences in intrinsic functional organization between dorsolateral prefrontal and posterior parietal cortex. *Cereb Cortex*. In press, 2013.

Kenet T, Arieli A, Tsodyks M, Grinvald A. Are single cortical neurons soloists or are they obedient members of a huge orchestra? In: *Problems in Systems Neuroscience*, edited by van Hemmen JL and Sejnowski TJ. New York: Oxford University Press, 2005.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A. Spontaneously emerging cortical representations of visual attributes. *Nature* 425: 954–956, 2003.

Kiani R, Esteky H, Mirpour K, Tanaka K. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97: 4296–4309, 2007.

Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? (Abstract). *Perception* 36, *Suppl*: 14, 2007.

Kreiman G, Hung CP, Kraskov A, Quiroga RQ, Poggio T, DiCarlo JJ. Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron* 49: 433–445, 2006.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126–1141, 2008.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12: 535–540, 2009.

Li N, DiCarlo JJ. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321: 1502–1507, 2008.

Logothetis NK. Perception and the brain. In: *Towards a Theory of Thinking*, edited by Glatzeder BM. Heidelberg, Germany: Springer-Verlag, 2010, p. 161–175.

Masquelier T, Thorpe SJ. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput Biol* 3: e31, 2007.

Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol* 100: 1407–1419, 2008.

Ohki K, Chung S, Kara P, Hubener M, Bonhoeffer T, Reid RC. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* 442: 925–928, 2006.

Op de Beeck HP, Dicarlo JJ, Goense JB, Grill-Spector K, Papanastassiou A, Tanifuji M, Tsao DY. Fine-scale spatial organization of face and object selectivity in the temporal lobe: do functional magnetic resonance imaging, optical imaging, and electrophysiology agree? *J Neurosci* 28: 11796–11801, 2008.

Pinto N, Doukhan D, DiCarlo JJ, Cox DD. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Comput Biol* 5: e1000579, 2009.

Poggio T, Bizzi E. Generalization in vision and motor control. *Nature* 431: 768–774, 2004.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* 16: 1661–1687, 2004.

Rajkai C, Lakatos P, Chen CM, Pincze Z, Karmos G, Schroeder CE. Transient cortical excitation at the onset of visual fixation. *Cereb Cortex* 18: 200–209, 2008.

Richmond BJ, Optican LM. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J Neurophysiol* 57: 147–161, 1987.

Rolls ET, Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73: 713–726, 1995.

Sato T, Uchida G, Lescroart MD, Kitazono J, Okada M, Tanifuji M. Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *J Neurosci* 33: 16642–16656, 2013.

Sato T, Uchida G, Tanifuji M. Cortical columnar organization is reconsidered in inferior temporal cortex. *Cereb Cortex* 19: 1870–1888, 2009.

Schwarzlose RF, Baker CI, Kanwisher N. Separate face and body selectivity on the fusiform gyrus. *J Neurosci* 25: 11055–11059, 2005.

Schyns PG. Grand challenges in perception science: modeling the future. *Front Psychol* 1: 10, 2010.

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* 29: 411–426, 2007.

Smith MA, Kohn A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28: 12591–12603, 2008.

Stevenson IH, London BM, Oby ER, Sachs NA, Reimer J, Englitz B, David SV, Shamma SA, Blanche TJ, Mizuseki K, Zandvakili A, Hat-

sopoulos NG, Miller LE, Kording KP. Functional connectivity and tuning curves in populations of simultaneously recorded neurons. *PLoS Comput Biol* 8: e1002775, 2012.

Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400: 869–873, 1999.

Takeuchi D, Hirabayashi T, Tamura K, Miyashita Y. Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* 331: 1443–1447, 2011.

Tallon-Baudry C, Mandon S, Freiwald WA, Kreiter AK. Oscillatory synchrony in the monkey temporal lobe correlates with performance in a visual short-term memory task. *Cereb Cortex* 14: 713–720, 2004.

Tamura H, Kaneko H, Kawasaki K, Fujita I. Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons. *J Neurophysiol* 91: 2782–2796, 2004.

Tanaka K. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13: 90–99, 2003.

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4: 832–838, 2001.

Uchida G, Fukuda M, Tanifuji M. Correlated transition between two activity states of neurons. *Phys Rev E Stat Nonlin Soft Matter Phys* 73: 031910, 2006.

Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. *Nat Neurosci* 5: 682–687, 2002.

Wang G, Tanaka K, Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272: 1665–1668, 1996.

Wang Y, Fujita I, Murayama Y. Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat Neurosci* 3: 807–813, 2000.

Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11: 1352–1360, 2008.

Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M. Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J Neurophysiol* 96: 3147–3156, 2006.

Yao H, Shen Y, Dan Y. Intracortical mechanism of stimulus-timing-dependent plasticity in visual cortical orientation tuning. *Proc Natl Acad Sci U S A* 101: 5081–5086, 2004.

Zang Y, Jiang T, Lu Y, He Y, Tian L. Regional homogeneity approach to fMRI data analysis. *Neuroimage* 22: 394–400, 2004.

Zoccolan D, Kouh M, Poggio T, DiCarlo JJ. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27: 12292–12307, 2007.