Correlated activity supports efficient cortical processing

Chou P. Hung[1,2]*, Ding Cui[1], Yueh-peng Chen[2], Chia-pei Lin[2] and Matthew R. Levine[1]

[1] Department of Neuroscience, Georgetown University, Washington, D.C., USA
[2] Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

Visual recognition is a computational challenge that is thought to occur via efficient coding. An important concept is sparseness, a measure of coding efficiency. The prevailing view is that sparseness supports efficiency by minimizing redundancy and correlations in spiking populations. Yet, we recently reported that "choristers", neurons that behave more similarly (have correlated stimulus preferences and spontaneous coincident spiking), carry more generalizable object information than uncorrelated neurons ("soloists") in macaque inferior temporal (IT) cortex. The rarity of choristers (as low as 6% of IT neurons) indicates that they were likely missed in previous studies. Here, we report that correlation strength is distinct from sparseness (choristers are not simply broadly tuned neurons), that choristers are located in non-granular output layers, and that correlated activity predicts human visual search efficiency. These counterintuitive results suggest that a redundant correlational structure supports efficient processing and behavior.

Keywords: object recognition, inferior temporal cortex, macaque, visual search, efficient coding

INTRODUCTION

Visual recognition engages neural mechanisms that are essential to our ability to learn and process complex information (Poggio and Bizzi, 2004). The key challenge of recognition is generalization, which requires that the representation is both object-specific and invariant to changes such as illumination and pose, even for novel objects. This is thought to occur via a hierarchy of cortical areas along the ventral visual pathway, ending in the inferior temporal (IT) cortex (Miyashita, 1993; Logothetis and Sheinberg, 1996; Tanaka, 1996; Tootell et al., 2003), but the underlying computations remain poorly understood (DiCarlo and Cox, 2007; DiCarlo et al., 2012). Current models and theories of recognition (Riesenhuber and Poggio, 1999; Masquelier and Thorpe, 2007; Mutch and Lowe, 2008; Bengio, 2009; Krizhevsky et al., 2012; Le et al., 2012; Zeiler and Fergus, 2014; Cadieu et al., 2014) are based on the idea that a hierarchy of simple and complex cells combine to increase specificity and invariance. To improve these models, it is necessary to understand the computations of local populations of neurons at an intermediate level of abstraction (DiCarlo et al., 2012).

A key concept is sparseness, a measure of coding efficiency. The current thinking is that sparseness increases efficiency by minimizing redundancy, correlation, and noise (Gawne and Richmond, 1993; Zohary et al., 1994; Vinje and Gallant, 2000; Olshausen and Field, 2004; Ecker et al., 2010; Renart et al., 2010; Xing et al., 2011; Hansen et al., 2012; King et al., 2013). Yet, reports in V1 slices and *in vivo* have shown the existence of neural ensembles that fire reliably in concert during spontaneous activity (Sadovsky and Maclean, 2014), and the same ensembles are active both without stimulation and in response to stimulation (Chu et al., 2014; Miller et al., 2014). We recently reported (Lin et al., 2014) that in macaque IT, correlated neurons "choristers" (Kenet et al., 2005; Carandini, 2014), neurons that have similar stimulus tuning and coincident spike timing, even during spontaneous activity, carry more generalizable object information than uncorrelated neurons ("soloists"). This surprising result hints that, counterintuitively, correlation *supports* efficient coding and that current thinking focused on sparsening, decorrelation, and denoising may be flawed.

The idea that the correlational structure, i.e., the spatial pattern of homogeneity vs. heterogeneity within a local population of neurons, may support efficient coding has been postulated in theory (Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wu et al., 2002; Dehaene and Changeux, 2005; Averbeck et al., 2006; Cohen and Kohn, 2011; Ecker et al., 2011; Eyherabide and Samengo, 2013; Shamir, 2014), but it has received little experimental support. Three novel aspects of our study allowed us to explore this hypothesis.

First, we used dense electrode arrays (64 sites across roughly two cortical columns, 0.2 mm resolution horizontally and in depth, **Figure 1A**) to characterize the correlational structure. High-density arrays allowed us to record neurons that have similar tuning, to measure redundancy as "Average Correlation Strength" (a site's average pairwise tuning similarity with all other sites in the array, where the tuning similarity between two sites is the Pearson correlation of their z-normalized stimulus responses, related to the concept of "population sparseness" Willmore et al., 2011). Because previous reports of efficient coding had insufficient sampling density to measure population sparseness, they instead measured "sparseness" as tuning sharpness, the selectivity of a neuron's response across stimuli, under the assumption that "sparseness" and "population sparseness" are interchangeable (that sparseness and correlation strength are inversely related) (Rolls and Tovee, 1995; Vinje and Gallant, 2000; Zoccolan et al., 2007; Willmore et al., 2011). When studies did examine functional correlation, it was in terms

FIGURE 1 | Experimental design and "pipe cleaner" model. (A) We inserted a dense multi-depth array (64 sites across ∼2 cortical columns) in macaque lateral IT (A16) and recorded spiking responses under light neurolept anesthesia. Stimuli were presented via rapid serial visual presentation, for 94 ms ON and 106 ms OFF (5 Hz), in pseudorandom order for 10 repetitions. Spike count from 100 to 200 ms post stimulus onset was averaged across repetitions. **(B)** A "pipe cleaner" model linking local correlational structure in neighboring columns to invariant representation. Most neurons are weakly correlated "soloists" (the bristles), tied to an underlying structure of correlated neurons ("choristers", the spine). Sampling a few points along the spine (a few choristers) is sufficient to reconstruct the overall structure. The model predicts that generalizable object information is carried by the choristers, and that the heterogeneity of the soloists may help to fine-tune the choristers to support generalization.

of individual pairs of neurons, without comparing the relationship between sparseness (tuning sharpness) and correlation (Gawne and Richmond, 1993; Sato et al., 2009; Takeuchi et al., 2011), or the comparison was limited to layers 2/3 (Tamura et al., 2014). Whether "correlation strength" and "sparseness" are related for a diverse sample of IT neurons remains untested (Willmore et al., 2011), and answering this question is important for understanding how local architecture relates to coding efficiency.

Second, the dense multi-depth arrays allowed us to examine layer specificity, which can tell us about input-output relationships. We and others (Sato et al., 2009; Lin et al., 2014; Tamura et al., 2014) previously reported that local IT populations have a correlational structure in which most neurons are weakly correlated and few neurons have strong tuning correlation and

significant spontaneous coincident spiking (∼6% of neuronal pairs in IT, vs. ∼50% of pairs in V1; Chu et al., 2014). Yet, these rare IT choristers are also highly efficient. Just 4–5 choristers per array (the top 6% as defined by k-means clustering, or 8% as defined by average pairwise tuning correlation) have the same object coding capability, for within-category generalization, as the entire array population (no more are needed given their object coding efficiency; Figure 7C of Lin et al., 2014). Based on this correlational structure and the much-better object coding capability of choristers vs. soloists, we previously proposed a "pipe cleaner" model (**Figure 1B**, a "fiber bundle" in mathematical terminology) in which the choristers (the spine) are the substrate of IT's output, encoding an invariant representation that supports generalization and recognition, and in which the soloists (the bristles) are IT's inputs, acting as heterogeneous tensors that fine-tune this high-dimensional representation (in the parlance of DiCarlo et al. (2012), to support "cortically local subspace untangling" and to "flatten object manifolds"). If so, choristers and soloists should be layer specific, with soloists tending to be in input layers and choristers tending to be in output layers. Such layer-specificity would be consistent with reports of decorrelated responses near layer 4 of V1 (Ecker et al., 2010; Hansen et al., 2012) and with reports that tolerance but not selectivity (sparseness) increases along the ventral visual pathway (Rust and Dicarlo, 2010; Willmore et al., 2011).

Third, we tested whether local correlated activity can predict visual search efficiency for complex naturalistic object stimuli. Previous reports have linked IT neuronal tuning to visual perception (Logothetis and Schall, 1989; Op de Beeck et al., 2001; Baker et al., 2002; Sigala and Logothetis, 2002; Mruczek and Sheinberg, 2007a; Sripati and Olson, 2010; Verhoef et al., 2012) and have linked perception to topography in V1 (Michel et al., 2013), but the interpretation was not linked to correlational structure. If local correlational structure, e.g., from short-range lateral inhibition in IT, predicts search efficiency, it would support that correlated activity and topography are linked to complex shape perception. It would also support recent reports that abnormal correlated activity and excitatory/inhibitory balance in object areas are linked to abnormal perception in autistics, linking these findings to spiking activity (Jiang et al., 2013; Robertson et al., 2013). Here, we asked whether local correlated activity predicts visual search for combinations of naturalistic objects. To avoid effects that might be driven by spatial attention or processes earlier in visual cortex, we used brief presentations at random locations followed by masking, and we equalized the stimuli for low level visual properties such as Fourier energy. Also, our stimuli were object combinations defined by local correlated activity in IT ("neurally defined features"; each "feature" is a set of objects), rather than abstract human-defined shapes as in previous reports, so that the predictions are specifically tied to contrastive coding of complex features by neighboring IT columns (e.g., from lateral inhibition).

Together with our previous report (Lin et al., 2014), these tests provide additional support for the hypothesis that correlated activity supports efficient processing and behavior. We

report that although the concepts of sparseness and decorrelation are often conflated, correlation strength and sparseness (when measured as tuning sharpness) should be considered as separate factors. We also provide additional support for choristers as the output neurons of IT, based on their cortical depth. Finally, we show that correlated activity in macaque IT predicts human visual search performance in a task with complex shapes.

METHODS

NEUROPHYSIOLOGY AND STIMULUS PRESENTATION

All experimental procedures in monkeys (*Macaca cyclopis*) 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. The procedures for the experiments were described in Lin et al. (2014) and are briefly summarized here. We inserted dense microelectrode arrays that had 64 sites (8 shanks and 8 contacts per shank, at 0.2 mm spacing spanning 1.4×1.4 mm horizontally and in depth, NeuroNexus A8×8-5mm-200-200-413) spanning all cortical depths and ~2–4 neighboring cortical columns (Figure 5 of Lin et al. (2014)). Recordings were made from 5 arrays, where each array was a separate insertion in a separate recording session, across 4 monkeys.

Initial surgery for headpost, EEG, and recording chamber implant was under isoflurane anesthesia, followed by repeated recording sessions under light neurolept anesthesia (Fujita et al., 1992; Wang et al., 2000; Tsunoda et al., 2001; Yamane et al., 2006; Sato et al., 2009, 2013; Brown et al., 2011) (0.9 μg/kg/hr i.v. Fentanyl, 70%/30% N_2O/O_2, 0.25 mg/kg i.m. droperidol, and 0.3–0.5% isoflurane) and muscle relaxation (1.2 mg/kg/hr i.v. rocuronium bromide). The fentanyl concentration is $100\times$ lower than in a recent report that contrasted awake vs. anesthetized signals (Ecker et al., 2014), and $10\times$ lower than in reports that did not find an effect on neuronal dynamics (Loughnan et al., 1987; Constantinople and Bruno, 2011). Our signals also lacked artifacts such as prolonged responses and up/down fluctuations reported with other anesthetics (Contreras et al., 1997; Haider et al., 2013). Compared to awake recordings, light anesthesia and muscle relaxation have the advantage of excluding potential effects from task-related top-down signals (Sigala and Logothetis, 2002; Maier et al., 2007; Ruff and Cohen, 2014) or eye movements (Rajkai et al., 2008; Ito et al., 2011), and a recent report suggests that activity during running resembles activity under anesthesia and is dissimilar to "visually detached" activity during quiet wakefulness (Froudarakis et al., 2014).

Single units were analyzed for coincident spiking and to remove cases of multiple detection of the same neuron across different contacts. All "site" responses were based on multi-unit activity (MUA) pooled from isolated single units at the same contact. We report only "site" responses because lower spike counts and the possibility of oversorting with single unit activity (SUA) can artificially weaken correlation measurements (Cohen and Kohn, 2011), and because the conclusions were the same as for MUA. The stimuli were 240 grayscale rendered objects or 113 colored photographed objects presented via rapid serial visual presentation (94 ms ON/ 106 ms OFF).

ANALYSES

Analyses were based on spike count from 100 to 200 ms after stimulus onset. Each site's tuning function was calculated as its trial-averaged response, z-normalized across stimuli. The same matrix of trial-averaged and z-normalized tuning responses across the array (i.e., a 240×64 matrix for 240 stimuli and 64 sites) was used as the input for correlation analysis, k-means, principal component analysis (PCA), and classifier analysis as described in Lin et al. (2014).

We classified each site as a "chorister" or "soloist" based on the site's average pairwise tuning correlation with other sites from the same array (the same calculation as in Lin et al. (2014) Figure 7C, brown line, but here "choristers" are random sites in the top 30%ile instead of rank-ordered sites in the top 8%ile). This top 30%ile corresponds to 16 sites per array for arrays 1–3 and 8 sites per array for arrays 4 and 5 that had more inactive sites. "Soloists" are the remaining sites (**Figure 2A** black dots). Choristers and soloists lie along a continuum of average pairwise tuning correlation strengths (**Figure 2A**). For object classification (**Figure 2B**) and noise covariation analyses (**Figure 2C**), we compared choristers (top 30%ile) against soloists in the 45–65%ile. For cortical depth (**Figure 3B**), the soloists are the bottom 30%ile. Layer-specificity is not seen for soloists in the 45–65%ile.

We used a linear support vector machine classifier to estimate the ability of a hypothetical downstream neuron (e.g., in prefrontal cortex) to read out the category of an untrained object (within-category generalization) based on the population activity in IT. The classifier output is based on the weighted sum of spiking activity from a set of IT neurons followed by a decision threshold. Because there were 8 possible categories, the classifier learned a one-vs.-all decision hyperplane for each of 8 categories and output the category that had the highest certainty.

Sparseness was calculated according to Vinje and Gallant (2000) and Zoccolan et al. (2007) as:

$$S = \left(1 - \frac{\left(\sum \frac{R_i}{n}\right)^2}{\sum \frac{R_i^2}{n}}\right) \bigg/ \left(1 - \frac{1}{n}\right),$$

where R_i is the site response to the i-th stimulus and n is the number of stimuli in the set.

We could estimate the cortical depth because we were able to visually see individual sites disappear into the brain during insertion and, because of the small footprint of the array shanks (15 μm thick, 33 μm wide), we could also track individual units as they transitioned from the deepest to the most superficial sites during array insertion. Anatomical confirmation of depth was impossible due to damage from later recording sessions. However, we estimate that the deviation of the array from vertical was less than 8 deg (less than 0.2 mm horizontal offset at the deepest site), based on anatomical confirmation of our V1 recordings using the same arrays (Supplemental Figure 1 in Chu et al., 2014).

For PCA, each PC consists of relative site activities (e.g., 1×64 matrix of coefficients for 64 sites, normalized to unit length) and stimulus-related scores (e.g., 1×240 matrix of weights for 240 stimuli) for that PC. The z-normalized response of a site to a stimulus can be back-calculated by summing, across all PCs, the product of the site's coefficient for each PC and the stimulus's score for that PC.

HUMAN TESTING
Observers
Procedures were approved by the Institutional Review Board of Georgetown University and informed consent was obtained from all observers. Six observers (3 male, 3 female, including the second author) participated in the experiments. All observers had normal or corrected-to-normal vision. Apart from the second author, observers were naïve as to the purpose of the experiment and were paid for participation.

Apparatus
Stimuli were controlled by computer using Matlab and Psychtoolbox 3 (Kleiner et al., 2007) and displayed on a 17″ cathode ray tube (CRT) (Sony Trinitron Multiscan 17sfII) with spatial resolution 1024×768 pixels and refresh rate of 60 Hz. Eye-screen distance was 57 cm, so that each pixel subtended approximately 0.03°. Ambient illumination was <4 Cd/m².

Stimuli
Object stimuli belonging to neurally defined features (grayscale rendered objects, "Set 1") were resized to 64×64 pixels (1.9° × 1.9°) and convolved with a 3×3 pixel Difference-of-Gaussians filter to match the background gray. Because the IT correlational structure is slightly more stable across stimulus sets for z-normalized responses than for raw responses, we constructed stimuli using the neurally defined features from z-normalized responses. Object stimuli were then equated for low-level image properties using the SHINE toolbox (Willenbockel et al., 2010). Groups of object stimuli were then randomly tiled to create the background (5 different objects), target (3 different objects), and distractors (3 different objects). Tiling position combinations were restricted to avoid lines of the same object. Target and distractors luminances

FIGURE 2 | Local correlational structure, sparseness, and generalization performance. (A) Sparseness (tuning width) and average correlation strength were only weakly related across 250 sites in anesthetized IT ($r = -0.09$, $p = 0.04$). Sparseness was calculated according to Zoccolan et al. (2007) and Vinje and Gallant (2000). Average correlation strength was the average site-to-site tuning correlation between each site and all other sites in the same array. Sparseness and average correlation strength were each highly consistent across two stimulus sets ($r = 0.72$ and 0.70, $p < 10^{-37}$ and $p < 10^{-39}$ resp.). Choristers (brown) are the 30%ile of
(Continued)

FIGURE 3 | Cortical depth vs. correlation strength and sparseness. (A,B) We sorted sites by their average strength of tuning correlation with other sites in the same array, then grouped the top ∼30% of sites per array as "choristers" and the bottom ∼30% as "soloists". Choristers are rarer in layer 4 (1.0–1.2 mm depth), whereas soloists are more common at 0.2 and 1.2 mm depth. The number of sites selected per group was higher for arrays 1–3 (16 choristers and 16 soloists per array) than for arrays 4 and 5 (8 choristers/soloists per array), because arrays 4 and 5 had fewer active channels. Average correlation strengths of choristers and soloists were 0.15 ± 0.04 and 0.02 ± 0.03, resp. **(C,D)** Same analysis based on sparseness (tuning sharpness). Average sparseness of broadly tuned and sharply tuned (sparse) sites were 0.12 ± 0.08 and 0.69 ± 0.25, resp.

were darkened by 5%, to make them more visible against the background. Mask stimuli were specific for each trial, created by scrambling the background (without target and distractors) at 0.24° resolution. Fixation point was a black square of 0.45° × 0.45°.

General procedure

Each block consisted of 144 trials comprising 48 "opposite", 48 "related", and 48 "unrelated" conditions, all with the same stimulus onset asynchrony (SOA) and adaptation duration. To minimize effects spanning across trials, each trial was preceded by an inter-trial interval (trial was initiated by key press), a blank fixation screen (1.5 s), and an adapting background of up to 8 s. In addition, objects were balanced across targets, distractors, and background and across conditions.

RESULTS

The neurophysiological data here is based on reanalysis of a previously reported dataset collected in monkeys under light neurolept anesthesia (Lin et al., 2014). Briefly, analyses are based on trial-averaged z-normalized responses (250 multi-unit "sites" and 6462 site pairs from 359 neurons) to stimuli that were presented via rapid serial visual presentation (**Figure 1A**). We begin by addressing a few concerns about our previous report:

that the 6% cutoff of choristers is arbitrary and that in fact "choristers" and "soloists" are not two types of neurons, and that perhaps the better object coding performance of choristers is due to multiple detection of the same neuron across contacts, or because soloists are less visually driven. In fact, the distribution of average correlation strengths is continuous, and the separation into "choristers" and "soloists" is merely for convenience of comparison, not to say that there are two distinct cell types. Correlation strength and within-category generalization performance both decline smoothly, so missing a few top "choristers" during sampling should not affect the resulting structure very much. To increase the population size for testing the effect of average correlation strength, we relaxed the definition of choristers as random sites in the top 30%ile of average pairwise correlation strength per array (**Figure 2A**, brown), and of soloists as random sites in the median 30%ile (45–65%ile; black is lower 70%ile). This 30%ile threshold for choristers corresponds to minimum average correlation strengths of 0.12, 0.15, 0.16, 0.09, and 0.06 for the 5 arrays. These thresholds are similar for sites separated by at least 0.6 mm horizontal distance (0.09, 0.11, 0.14, 0.07, and 0.05). Although correlated neurons do tend to be more visually driven than uncorrelated neurons (c.f. Figure 5C of Tamura et al. (2014)), we still observed higher performance for choristers when choristers and soloists were matched for visual drive (**Figure 2B**, ∼12 Sp/s baseline-subtracted

response to each site's preferred stimulus; based on 5 arrays and 2 sites per array, at least 0.6 mm horizontal distance between sites).

SPARSENESS AND CORRELATION STRENGTH ARE MOSTLY UNRELATED

In previous reports, sparseness (measured as tuning sharpness) was thought to support efficient coding by reducing correlated activity (Young and Yamane, 1992; Rolls and Tovee, 1995; Baddeley, 1996; Olshausen and Field, 1996; Bell and Sejnowski, 1997; Vinje and Gallant, 2000; Zoccolan et al., 2007). This would predict that soloists should have better object coding capability (whereas our results suggest that choristers have better object coding, at least for within-category generalization) and that soloists should be sharply tuned. Conversely, a trivial explanation of the better object coding capability of choristers is that perhaps choristers are broadly tuned and therefore have better tolerance to stimulus variations.

We report that neither prediction is correct. Sparseness (measured as the modified sparseness index of Vinje and Gallant (2000), Lin et al. (2014) and average correlation strength are mostly uncorrelated across sites. Within each of 5 arrays (5 separate array insertions across 4 monkeys), the relationship between sparseness and average correlation strength was non-significant, and it was weak and barely significant when pooled across all arrays (Pearson $r = -0.09$, $p = 0.04$, $N = 250$ sites, **Figure 2A**). This weakness was not due to noise in either measurement, because sparseness and average correlation strength were each highly consistent across two stimulus sets ($r = 0.72$ and 0.70, $p < 10^{-37}$ and $p < 10^{-39}$, resp.). This dissociation between sparseness and correlation is consistent with a previous conjecture that these measures are unrelated (Willmore et al., 2011) and with a recent report that found a weak (albeit positive, $r = 0.07$, $p < 0.001$, rather than negative) dissociation in layer 2/3 (Tamura et al., 2014). Our data show that the dissociation also holds for a wider sample of IT neurons across supragranular, granular, and infragranular depths.

CORRELATED NEURONS ARE MOSTLY IN OUTPUT LAYERS

A key issue in linking neural activity to models is the cortical layer of different functional elements. An ongoing debate is whether neurons are correlated or uncorrelated, and whether these are in input or output layers. In V1, a recent study suggested that noise correlations are much lower than previously thought (Ecker et al., 2010), but alternatively it has been reported that noise correlation is layer-dependent and is lower, with better coding efficiency, in the granular layer (Hansen et al., 2012).

We suggest that neither view is entirely correct in IT. Here, we report that correlated neurons (choristers, with more efficient coding) are almost exclusively found in supragranular and infragranular layers. In IT, signal (tuning) correlation and noise correlation are related, and choristers tend to have stronger noise correlation (choristers (brown): Rsc = 0.13; soloists (45–65%ile, red): Rsc = 0.04; $p < 10^{-22}$, unpaired t-test; **Figure 2C**), including pairs separated by at least 0.6 mm horizontal distance and with similar visual drive (mean baseline-subtracted response to preferred stimulus of each cell is between 10 and 30 spikes/sec) (choristers: Rsc = 0.16; soloists: Rsc = 0.06; $p < 10^{-13}$; blue).

Of the 64 choristers, most were in supragranular and infragranular layers and only five were between 1.0–1.2 mm depth, near layer 4 (**Figure 3A**). Conversely, the most uncorrelated soloists (the ~30% of sites with the lowest correlation strength per array) were more prevalent at 0.2 and 1.2 mm depth (layers 1 and 4), although roughly half were in supragranular and infragranular layers (**Figure 3B**). The result was similar for single-unit activity. The proportion of choristers vs. soloists was significantly lower in the granular layer (1.0–1.2 mm) compared to supragranular and infragranular layers ($p = 0.0007$ and $p = 0.0003$, two-sided Fisher's test), and the difference between supragranular and infragranular layers was non-significant. This layer-specificity is consistent with a recent report in V1 that also measured correlated variability (Hansen et al., 2012). In contrast to correlation strength, sparseness, a measure of coding efficiency that is commonly based on tuning sharpness (Young and Yamane, 1992; Rolls and Tovee, 1995; Vinje and Gallant, 2000; Zoccolan et al., 2007), was not layer specific (**Figures 3C,D**, n.s. for all comparisons).

THE LOW DIMENSIONAL CORRELATIONAL STRUCTURE IS ALSO IN OUTPUT LAYERS

A recent perspective article (DiCarlo et al., 2012) highlighted the need to understand the processing of local populations of neurons at an intermediate level of abstraction. Covariation analysis (e.g., k-means clustering and PCA) is a useful form of abstraction because it directly ties the correlational structure to the idea of a low-dimensional manifold representation of object features (DiCarlo et al., 2012) and to our pipe cleaner model (Lin et al., 2014). We often encounter novel objects and novel environments (Vaziri et al., 2014) that must be categorized, and it is thought that the visual system learns useful shape statistics of the animal's environment (Srihasam et al., 2014). A key concept of the model is that the invariant representation, which supports generalization across rotation-in-depth, changes in illumination, and variations within an object category (studied here), has a spatial organization that is concentrated in a low-dimensional correlational structure. Such a low-dimensional correlational structure could be very useful for decoding by downstream neurons and for generalization learning, by providing a smoothly differentiable structure that is stable across categories, by reducing the number of inputs that must be pooled (instead of listening to all neurons, a downstream neuron could conceivably identify the most useful neurons in a population based solely on coincident timing, even during spontaneous activity), and by supporting robustness for noisy spiking populations.

We previously used k-means clustering (Lin et al., 2014) to identify clusters of sites that behaved more similarly across stimuli. Note that this is different from the typical approach, where the same data is clustered as groups of stimuli according to their response similarity (Kiani et al., 2007). Also, to focus on the local correlational structure, we focused our analysis specifically within each array, rather than pooling across the entire population (all arrays). Here, we extend our approach to PCA, to tie the low-dimensional structure to output layers and to behavior. Because many of the conclusions drawn from PCA regarding

FIGURE 4 | Cortical depth vs. low-dimensional correlational structure. (A) Percent of response variance explained by PCs 1–5, based on z-normalized responses. Chance and 5–95%ile distributions are indicated by open circles and red bars, based on shuffling of response IDs across trials. **(B,C)** Explained variance for Arrays 2 and 3, from two separate array insertions (separate recording sessions) in monkey 2. **(D)** Cortical depth vs. percent explained variance of PCs 1 and 2 across 5 arrays. **(E)** Comparison of distributions in **(D)** among different depths. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, •p = n.s.

spatial structure are similar to those from k-means clustering and from pairwise site correlations, we will discuss only the highlights below.

Compared to k-means clustering, PCA has the advantage of explaining more of the variance using fewer dimensions, because each PC is exactly aligned to maximally explain the remaining variance, whereas k-means clustering will include sites that are uncorrelated (soloists). Thus, whereas k-means clustering tends to highlight columnar organization (changes in tuning across cortex), PCA can reveal the layer-specificity of the low-dimensional correlational structure. For these dense arrays, the spatial patterns of site covariation were nearly identical between the lower PCs and k-means clustering and corresponded roughly to the differential activation of neighboring columns (at higher PCs, it is less likely that the orthogonal PCs are relatable to biological processes). Specifically, PC1 (the dimension of maximum variance within an array) corresponds roughly to activation vs. suppression of most sites in the array, which may contribute to invariant representation by encoding how strongly a feature or feature contrast is present within an object category. PC2 (the dimension that best explains the remaining variance) corresponds roughly to the differential activation of two neighboring cortical columns (i.e., the sign and relative strength of a feature contrast) and appears virtually identical to k-means clustering at $k = 2$ (Figure 5 of Lin et al., 2014).

By examining how well the lowest PCs explain the variance of individual sites, we can determine the layer-specificity of the low-dimensional correlational structure. Because of the high proportion of soloists and the rarity of choristers (**Figure 2A** and Gawne and Richmond, 1993; Sato et al., 2009; Lin et al., 2014), the first few PCs explained only a small fraction of the total variance within each array (**Figures 4A–C**, example arrays 1–3; responses were averaged across repetitions and z-normalized), even though

the tuning of individual sites was highly consistent across even vs. odd trials ($r \sim 0.6$–0.9, Figure 2A of Lin et al. (2014)). The first two PCs explained only about 25% of the response variance across stimuli. A previous study reported 15% explained variance for two PCs, based on recordings from random penetrations across IT (Baldassi et al., 2013), and another study reported ∼70% for two PCs (∼60% for PC1) based on recordings from electrode bundles targeted to the centers of IT optical imaging domains (Figures 10, 16 of Sato et al. (2009)). A possible explanation for the large difference across studies is that in addition to layer-specific heterogeneity of choristers vs. soloists, there is also topographical heterogeneity that cannot be attributed to bias from image guided electrode targeting.

For individual sites, the percent of variance explained (%EV) by the first two PCs varied widely and was lower for sites in layers 1 and 4 (0.2 and 1.2 mm depths) (**Figures 4D,E**). It was more widely distributed, with up to 77%EV, for sites in supragranular and infragranular layers. The specificity of the correlational structure to output layers hints that it is shaped by local networks, rather than by feedforward or thalamic input to layer 4 or feedback from higher areas to layer 1. The lower explained variance at 0.2 mm and 1.2 mm depths (many sites are <5%EV), despite good even-vs.-odd trial tuning consistency at all depths (Figure 2A of Lin et al. (2014)), suggests that the inputs to layers 1 and 4 are nearly orthogonal to the correlated activity, consistent with our "pipe cleaner" model. In comparison, scrambling the stimulus IDs of the scores of the first two PCs, without altering the PCA coefficients, resulted in −23%EV on average (i.e., the difference between the scrambled prediction and the actual response has a total variance that is on average 123% that of each site's actual total variance, **Figure 4D**). The fact that the EVs of scrambled predictors are negative, and not zero, for sites in all layers further supports that these sites are

visually driven and selective. Only layer 4 (1.2 mm depth) had average response below baseline (−0.5 spikes/s), consistent with suppression by local or feedforward inhibition. Overall, this result extends upon the layer-specificity of choristers and soloists as measured by average correlation strength by showing that the correlational structure is concentrated in a few dimensions (a low-dimensional manifold), mainly in a subpopulation of neurons in output layers.

NEURALLY DEFINED FEATURES BASED ON CORRELATED ACTIVITY

To link the correlational structure to behavior, we constructed "neurally defined features" based on the tuning of neighboring IT columns. Previous reports used a variety of methods (feature reduction, k-means clustering, PCA, or simply averaging the tuning along a penetration) to characterize IT tuning (Young and Yamane, 1992; Tsunoda et al., 2001; Baldassi et al., 2013; Sato et al., 2013). However, because their analyses focused on the tuning of single neurons or random IT populations, behavior has not been tied to the concept of a cortically local low-dimensional manifold and lateral inhibition.

Here, to focus on the differential coding by local populations, we defined "neurally defined features" as sets of stimuli determined by PCA of each array (the same PCs as in **Figure 4**, e.g., computed from a 240 stimuli × 64 site matrix of z-normalized tuning responses). The "neurally defined features" are sets of stimuli with extreme PCA scores, treated collectively without altering or blending the images. For example, feature "Array 1 PC1+" is the set of 10 stimuli with the most positive PC1 scores for Array 1, allowing that some of the same stimuli may also belong to PC2+ or PC2− or to features of another array (however, for behavioral testing we did not allow "reuse" of stimuli across background, target, and distractor within a trial). Although one "Array 1 PC1+" stimulus may have a higher PC1 score than another "Array 1 PC1+" stimulus, we treat them equally because it is the collective effect of the set of "Array 1 PC1+" stimuli that dilutes away stimulus-specific effects, to increase the feature's specificity to that neuronal population (e.g., vs. populations in early visual cortex or elsewhere in IT). In each array, PC2+ and PC2− correspond to differential activation of neighboring cortical columns, and PC1+ and PC1− correspond to co-activation and co-suppression of most neurons in the array (differential activation at a larger spatial scale).

Figure 5A shows examples of stimulus responses along PCs 1 and 2 for Array 1. The red and blue matrices show examples of baseline-subtracted firing rates across the 8 × 8 array to specific stimuli. Across different levels of overall activation and suppression (different PC1 scores), stimuli that differentially activated the left column more than the right column (PC2− stimuli) tended to be objects with protrusions, whereas stimuli that differentially activated the right column more than the left column (PC2+ stimuli) tended to be objects with internal features. Although these semantic descriptions are qualitative and are not part of the feature definition, the positive and negative PC extremes appeared to prefer contrastive features ("rumpled" vs. "smooth", "upward" vs. "downward vertex with gradient") that were consistent across two stimulus sets

(grayscale rendered objects and color/grayscale/silhouette photographed objects). Therefore, we assigned the positive and negative extremes to separate features, resulting in 12 neurally defined features derived from 6 PC feature dimensions (PCs 1 and 2 from 3 arrays, i.e., 3 separate recording sessions across 2 monkeys, **Figure 5B**).

Why use PCA, instead of k-means clustering or penetration averaging? Because neighboring columns have correlated tuning and because k-means clustering does not distinguish soloists from choristers, features from one k-means cluster are less visually distinguishable from those of neighboring clusters. Unlike PCA, k-means clustering or penetration averaging would have ordered stimuli according to how strongly they activated each column, causing stimuli that appear very different to group together (e.g., the bike (#46) and the fence (#11) for the right column, or the shears (#128) and the couch (#87) for the left column). PCA features appear more different, particularly PC2+ vs. PC2−. We note that this advantage of PCA may be specific to local populations sampled by densely spaced electrode arrays. The differential coding along PC2 is consistent with previous reports of a "shape-contrast" effect in perception (Suzuki and Cavanagh, 1998) and in IT responses (Leopold et al., 2006), although it is distinct from the idea of norm-based encoding (Valentine, 1999) because it is primarily driven by shape rather than by semantic category or low-level properties such as color and texture (Baldassi et al., 2013; Lin et al., 2014).

These PC feature dimensions were uncorrelated across arrays (Pearson correlations of PC scores were non-significant), even the features measured from different sessions 3 mm apart in the same monkey (M2), indicating that the features are not simply due to familiarity (Mruczek and Sheinberg, 2005, 2007b; Hein et al., 2007; Anderson et al., 2008) or coarse topography (Op de Beeck et al., 2007; Sato et al., 2013). Also, the monkeys had never seen these stimuli previously.

NEURALLY DEFINED FEATURES PREDICT VISUAL SEARCH EFFICIENCY

To link correlated activity to behavior, we designed a human visual search task in which the target, distractors, and background were disjoint sets of objects from monkey neurally defined features. Previous reports based on simple features such as orientation, color, and size hint that visual performance is associated with horizontal processes and lateral inhibition in early visual cortex (Butler et al., 2008; Yoon et al., 2010; Michel et al., 2013). Here, we asked whether lateral inhibition among complex feature representations in IT might also predict visual performance. Because of the short horizontal range of lateral inhibition in macaque IT, which may translate to longer-range inhibition in humans if similar statistical feature mechanisms are useful for representation, we tested whether the target would be more salient from the background if they were contrastive ("opposite" sign) features from the same array (differentially activating neighboring columns, e.g., Array 2 PC2+ target vs. PC2− background), than if they were "related" features (different PCs of the same array, activating the same column at different scales, e.g., Array 2 PC2+ target vs. PC1+ background) or "unrelated" features (PCs from different arrays, activating distant
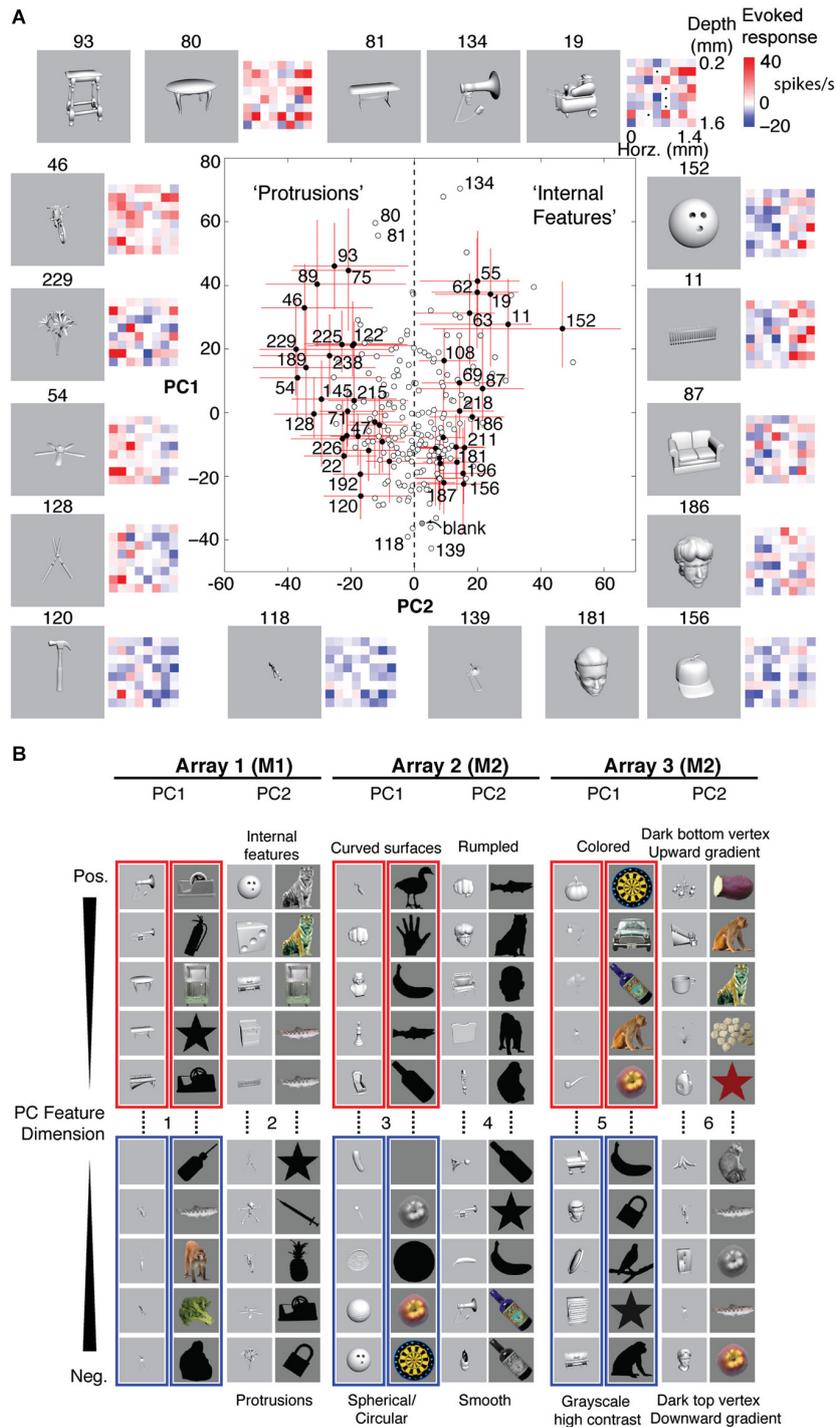
FIGURE 5 | Neurally defined features. (A) Neurally defined features based on Array 1's PC1 and PC2 scores. Each red and blue 8 × 8 matrix shows baseline-subtracted response to one stimulus across the 64 sites, spanning all depths and neighboring IT columns. PC1+ stimuli activated most sites. PC2+ and PC2− stimuli differentially activated sites on the right and left sides of the array. Numbers indicate stimulus IDs. Black dots in stimulus 19's matrix indicate inactive sites. Red lines indicate 5–95%ile, and filled circles indicate stimuli with significant PC2. "Protrusions" and "Internal Features" are labels to help see the pattern of PC2− and PC2+ stimuli, but the labels are not part of the feature definition. **(B)** Features from 3 array insertions (3 recording sessions) in two monkeys and two stimulus sets. Only the stimuli with the most extreme scores are shown, out of 240 object stimuli for set 1 (grayscale rendered 3D objects) and 113 stimuli for set 2 (color, grayscale, and silhouette photographs). The slight difference between panels **(A)** and **(B)** is because the scores in A are calculated from unnormalized responses, to scale with the baseline-subtracted firing rates in the matrices, whereas the scores in **(B)** are from z-normalized responses, for better consistency of spatial covariation patterns across stimulus sets.
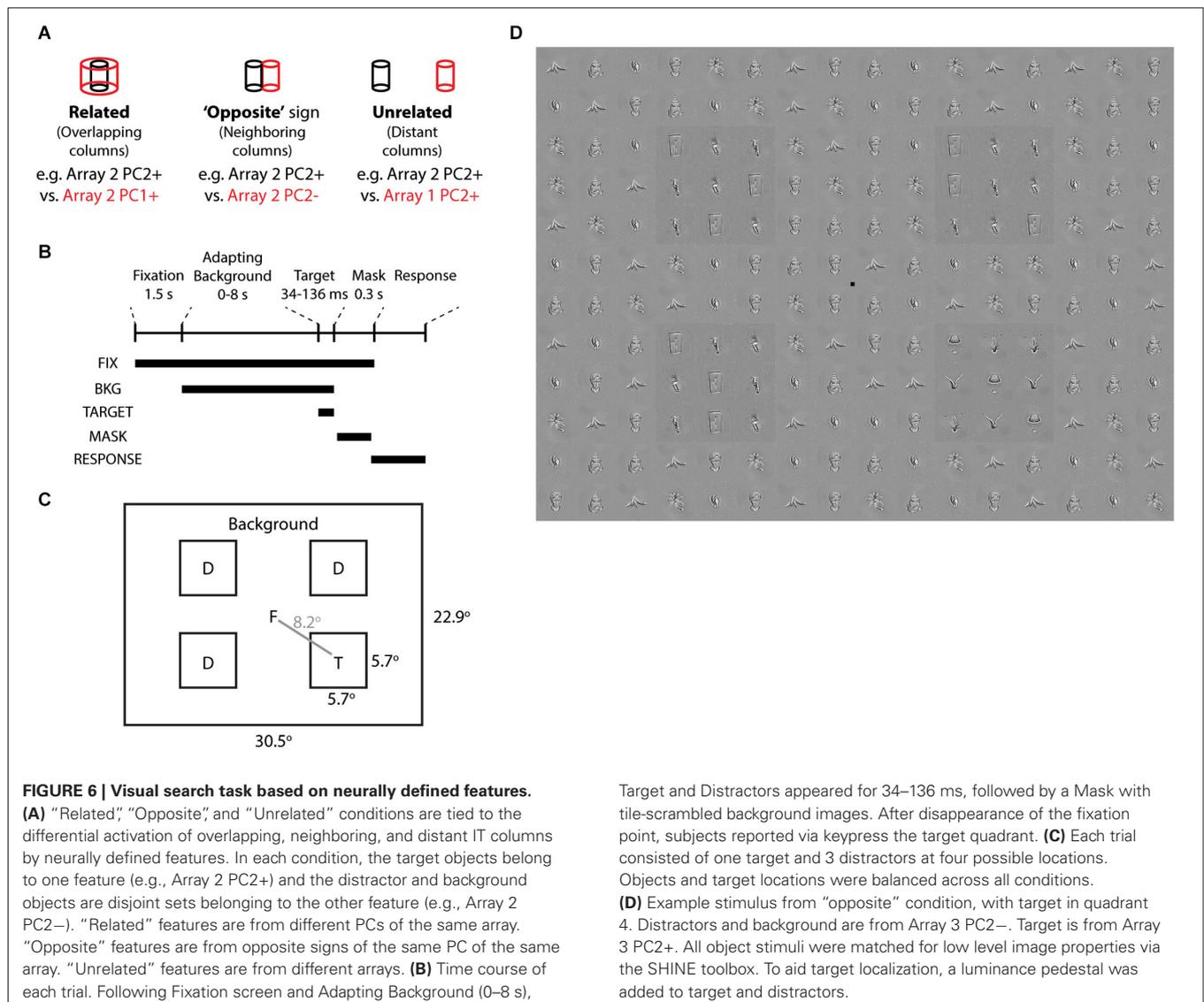
FIGURE 6 | Visual search task based on neurally defined features.
(A) "Related", "Opposite", and "Unrelated" conditions are tied to the differential activation of overlapping, neighboring, and distant IT columns by neurally defined features. In each condition, the target objects belong to one feature (e.g., Array 2 PC2+) and the distractor and background objects are disjoint sets belonging to the other feature (e.g., Array 2 PC2−). "Related" features are from different PCs of the same array. "Opposite" features are from opposite signs of the same PC of the same array. "Unrelated" features are from different arrays. **(B)** Time course of each trial. Following Fixation screen and Adapting Background (0–8 s),

Target and Distractors appeared for 34–136 ms, followed by a Mask with tile-scrambled background images. After disappearance of the fixation point, subjects reported via keypress the target quadrant. **(C)** Each trial consisted of one target and 3 distractors at four possible locations. Objects and target locations were balanced across all conditions. **(D)** Example stimulus from "opposite" condition, with target in quadrant 4. Distractors and background are from Array 3 PC2−. Target is from Array 3 PC2+. All object stimuli were matched for low level image properties via the SHINE toolbox. To aid target localization, a luminance pedestal was added to target and distractors.

columns >3 mm apart, e.g., Array 2 PC2+ target vs. Array 1 PC2+ background) (**Figure 6A**). The distractors and background were disjoint sets of objects sharing the same neurally defined feature.

To induce a temporary visuoperceptual distortion as in Leopold et al. (2001), we adapted the subject to the background for up to 8 s, followed by a brief (34–136 ms) presentation of the target and distractors and then a mask (scrambled background) (**Figures 6B,C**). Subjects indicated via key press the quadrant in which the target appeared. Subjects were instructed to search for the quadrant whose pattern appeared different from the other three quadrants. We measured visual search efficiency as the reporting accuracy of the target quadrant (chance = 25% correct). Such visual search displays are commonly used to study early perceptual processes and have only recently been applied to neurally related complex shapes (Sripati and Olson, 2010). A strength of the task is that the brief stimulus appearance and the target location randomization preclude artifacts from

differences in spatial attention or eye position. To focus the task on complex shapes rather than early visual processes, we used the SHINE toolbox (Willenbockel et al., 2010) to equalize the objects in terms of low-level cues including luminance, contrast, and orientation-specific Fourier power (including spatial frequency) (**Figure 6D**).

We began by comparing, in one subject, how performance depended on stimulus condition, target duration (stimulus onset asynchrony "SOA" between target/distractors and mask) and adaptation duration. At 34 ms SOA, performance was consistently higher across different durations of adaptation when the target feature was "opposite" in sign to the distractors and background (i.e., when target and background were contrastive features that differentially drive neighboring IT columns) (**Figure 7A**, blue) than when target and background were "related" features (red) ($p = 0.038$, Cochran-Mantel-Haenszel test). Although performance was slightly higher at 2 s adaptation, the odds ratios were not heterogeneous across different levels of adaptation

FIGURE 7 | Visual search performance for neurally defined features.
(A) Performance of one subject for target whose neurally defined feature is "opposite" (blue) or "related" (red) to that of the distractors and background, across different adaptations and different stimulus onset asynchrony (SOA) between stimulus and mask. Accuracy is higher for "opposite" at 34 ms SOA, and the difference between "opposite" and "related" is more consistent at longer adaptation. Error bars show 95% CI, based on 48 trials per condition. **(B)** Performance across 6 subjects at 2 s adaptation and 34 ms SOA for "opposite" (blue), "related" (red), and "unrelated" (green) conditions. Black line indicates average performance across subjects. The difference between "opposite" and "related" is significant at p = 0.03, based on Wilcoxon signed ranks test. **(C)** Average performance across 6 subjects at 34, 68, and 136 ms SOA, 2 s adaptation.

(p = n.s., Breslow-Day test). An opposite effect was seen at 68 ms SOA at 0 and 0.5 s adaptation, but the effect reversed with longer adaptation. At the longest adaptation (8 s), performance was higher for "opposite" than for "related" features at all SOAs. Based on this pattern, we surmised that the effect was most reliably consistent with the prediction at the shortest SOA (34 ms) and with longer adaptation. To avoid tiring the subjects with the 8 s long adaptation or a possible flooring effect at shorter adaptation (e.g., 0.5 s), we tested all subjects at 2 s adaptation.

Across 6 subjects, visual search efficiency was consistently higher for "opposite" features than for "related" features at 34 ms SOA and 2 s adaptation (**Figure 7B**, p = 0.03, Wilcoxon signed rank test). As with the first subject, the higher performance for "opposite" features was only observed at the shortest SOA of 34 ms across the 6 subjects (**Figure 7C**). The persistence of the effect across different durations of adaptation at the short 34 ms SOA hints that the effect is likely driven by feed-forward processing and short-range lateral interactions in IT, because 34 ms is likely too brief for feedback (Bansal et al., 2014; Scholl et al., 2014) or long-range lateral interactions (Singer and Kreiman, 2014; Tang et al., 2014). We suggest that the mechanism is associated with short-range lateral inhibition (e.g., between neighboring columns) in IT, similar to reports of lateral interactions in early visual cortex (Das and Gilbert, 1999; Michel et al., 2013), rather than a distance-dependent effect in IT, because the search efficiency of "related" (0 mm cortical separation) and "unrelated" (>3 mm separation) features was not significantly different. Also, "opposite" features had higher search performance than "unrelated" features in 4 of 6 subjects, but this difference did not reach significance.

DISCUSSION

Our results suggest that correlated activity contributes to efficient coding and human visual search efficiency. The main findings are that correlation strength and sparseness are only weakly related and should be considered as separate factors, that correlated activity is primarily located in output layers, and that correlated activity in monkey IT predicts human visual search efficiency. Together, these results suggest that correlated activity may be the substrate of IT's output and that, contrary to previous reports, correlated activity contributes to coding efficiency.

"POPULATION SPARSENESS" VS. "SPARSENESS" IN EFFICIENT CODING

These results suggest that a fundamental shift is needed in our approach to understanding efficient coding. Previous reports of efficient coding assumed that population sparseness and tuning sharpness (conventionally termed "sparseness"; Rolls and Tovee, 1995; Vinje and Gallant, 2000; Zoccolan et al., 2007; Willmore et al., 2011) are interchangeable. Instead, our results suggest that correlation strength (inversely related to population sparseness) is better than tuning sharpness as a measure of population redundancy, and that these two measures are mostly unrelated. Surprisingly, they show that the representation is more correlated in output layers than in input layers, which is opposite to the expectation that increasing sparseness supports efficient coding. This layer-specific increase in correlation is unrelated to tuning sharpness. Together with our previous report showing the better object coding capability of choristers vs. soloists, these results highlight the role of correlation in efficient coding.

WHY ARE CORRELATION STRENGTH AND SPARSENESS UNRELATED?
The main reason for this apparent discrepancy is that previous studies did not measure population sparseness. Their wider electrode spacing meant that neuronal tuning was too dissimilar to

compute population sparseness. Dense sampling, on the order of 64 neurons per mm^3, is necessary to measure population sparseness, because neuronal tuning is heterogeneous even within a cortical column and because choristers are rare (Sato et al., 2009; Lin et al., 2014). It is unclear what mechanism might enable correlation of sharply tuned neurons in output layers and decorrelation of broadly tuned neurons in input layers. The prevalence of soloists in input layers 1 and 4 suggests that the feedforward and feedback inputs to IT are already decorrelated, or that they are actively decorrelated by inhibition. Conversely, our finding that correlated activity is mostly in output layers is consistent with the layer specificity of local circuits and horizontal fibers. However, the consistency of our V1 and IT results in terms of tuning and spike timing correlational structure suggest that they are probably driven more by local circuitry than by long range fibers, which have different patterns in V1 vs. IT (Tanigawa et al., 2005).

IMPLICATIONS FOR VISUAL SEARCH EFFICIENCY

Overall, these results support that a human homolog of IT, previously shown by many studies (Grill-Spector et al., 2001; Tootell et al., 2003; Orban et al., 2004; Kriegeskorte et al., 2008), guides search based on complex features. In relation to classical theories of visual search based on feature integration theory (FIT; Treisman and Gelade, 1980), these results differ in two key aspects. First, whereas FIT posits that fast visual search relies on early visual areas, our results support an accumulating body of evidence that later visual areas also contribute to fast visual search (Hochstein and Ahissar, 2002). Second, FIT posits that pre-attentive, parallel search is more efficient for low-level features than for feature conjunctions. Our results show that preattentive, parallel search is also more efficient for specific types of complex features, contextually dependent on the complex features present in the background, and that this contextual dependency is specifically linked to cortical neighborhood relationships and correlated activity in IT. This supports a model by Duncan and Humphreys that all search is parallel and depends on representational similarity and competition for resources across multiple levels of the visual system (Duncan and Humphreys, 1989, 1992).

Consistent with a previous report that linked macaque IT responses to human visual search efficiency (Sripati and Olson, 2010), our results suggest that this context dependency is due to a stimulus-specific competition for resources that can be explained by local contrastive mechanisms such as lateral inhibition (Wang et al., 2000; Leopold et al., 2001). Our results strengthen the case that this mechanism is tied to competition for local resources in IT, vs. in earlier areas, because it depends on IT cortical proximity. Also, by linking search efficiency to correlational structure, our results support an assumption in the previous report (Sripati and Olson, 2010), that a population of heterogeneous neurons (e.g., within an IT column) can be modeled by the discriminative capacity of their correlated activity, as the activity of a few neurons (choristers). One difference from the previous report (Sripati and Olson, 2010) is that their behavioral and neural responses were predicted by the coarse footprint difference of the objects, i.e., the

spatial overlap of the blurred images, whereas in our data the coarse footprint difference does not predict better performance in the "opposite" condition (unpaired t-test of distributions of coarse footprint index in correct vs. incorrect trials was non-significant). This difference, together with our use of brief presentations and masking, further supports that low-level features are insufficient to account for our results. Also, because the correlational structure was tied to shape rather than semantic category (Lin et al., 2014), and because there was no difference in category overlap across stimulus conditions, our results support that the contrastive mechanism was feature-based, not semantically-based.

IMPLICATIONS FOR COMPUTATIONAL MODELS OF RECOGNITION

An ongoing debate in computational modeling of recognition and generalization learning is how to design the architecture, e.g., whether it is necessary to simulate populations of binary spiking neurons (Masquelier and Thorpe, 2007; Chan et al., 2011; Merolla et al., 2014), or whether convolutional networks are sufficient or even superior. Although convolutional networks outperform spiking networks on datasets like ImageNet, and their performance approximates ideal observers on object categorization and exceeds that of randomly sampled IT neurons (surprising because IT is the last stage of the ventral pathway) (Krizhevsky et al., 2012; Cadieu et al., 2013, 2014; Zeiler and Fergus, 2014), their performance remains far worse than that of humans on real-world vision. In a recent model that approximates IT and ideal observers (Yamins et al., 2014), the approximation to IT is as low as 20%EV for single sites (mean 48.5%EV), and IT split-half data still outperforms the best model on predicting representational dissimilarity for image generalization, object generalization, and category generalization. Our results suggest that part of this gap is due to the much poorer coding capability of soloists vs. choristers and due to the rarity of choristers. We suggest that the comparison (for both convolutional and spiking networks) should be against IT choristers, rather than a random pool of IT neurons. Also, our layer-specific correlation results show that correlation strength increases from input to output layers within a cortical column. Increasing the cell and layer specificity of modeling could in principle favor spiking network models with individual cores that simulate computations within cortical columns, as in Merolla et al. (2014).

Another aspect that may favor spiking network models is the relationship between correlational structure and learning. Whereas learning in convolutional networks occurs via genetic algorithms that guide connection patterns based on overall performance, learning in spiking networks is more local and can in principle be tied to our "pipe cleaner" model and spike timing dependent plasticity (STDP). This difference in approach manifests in convolutional networks as a gradual increase in performance along the hierarchy (Serre et al., 2007), whereas in our data the near-chance performance of soloists hints that performance increase may be staggered along the hierarchy, alternating between low performing input layers and high performing output layers for each cortical "area". This alternation between low-performing soloists and high-performing choristers may be critical to learning and maintaining an

invariant representation (cortically local subspace untangling; DiCarlo et al., 2012). Another aspect that could be modeled is that choristers are rarer in IT (Lin et al., 2014) than in V1 (Chu et al., 2014). We speculate that the increasing rarity of choristers is because of increasing complexity (increasingly high-dimensional feature spaces) along the visual hierarchy, which requires ever larger and more heterogeneous populations of soloists within each column to develop and maintain an invariant low-dimensional (manifold) representation. These emphases on local learning could also benefit from architectures based on multicore networks.

HOW MIGHT THE CORRELATIONAL STRUCTURE FUNCTION ALGORITHMICALLY?

How might a homogeneous/heterogeneous spiking network support generalization learning? Our conceptual "pipe cleaner" model (Lin et al., 2014) predicts that the feedforward and feedback inputs to IT may act as tensors, enabling the fine adjustments that may be necessary to build and maintain an invariant representation. The near-orthogonality of the inputs vs. the manifold (**Figure 4D**) indicates that they are optimally tuned to alter the manifold (i.e., co-alignment with the manifold would be inefficient and could conceivably result in uneven coverage). Such adjustments could occur via STDP, because soloists (mostly in the input layers) that are better tuned to the feedforward (environmental) and feedback (behavioral context) input statistics will spike more quickly, shaping the tuning of the choristers that support the invariant representation. This prediction is consistent with a recent report that found layer-specific temporal sequencing in perirhinal cortex (Takeuchi et al., 2011). Because the invariant representation in IT is based on shape rather than semantic category, invariance training on any category would also improve invariance to other categories that share the same feature, supporting generalization from few examples. We speculate that the combination of heterogeneity (population sparseness) in the input layers and redundancy/smoothness (overlap in tuning) in the output layers may be important for populations of spiking neurons, to achieve sufficient bit resolution from binary spiking neurons operating in high dimensional feature space. This problem of poor bit resolution has been criticized as a fundamental weakness of spiking network models vs. convolutional network models of recognition, and homogeneous/heterogeneous networks may be a key part of the brain's solution.

ON TECHNICAL APPROACHES TO STUDY EFFICIENT CODING

Recent technical advances have improved cell-specificity, sampling density, and anatomical co-registration. Our results suggest that, to better understand how the local correlational structure contributes to efficient coding, simultaneous sampling across multiple depths down to at least 1.0–1.2 mm is essential, to map both the inputs and the outputs within a column. Sampling density on the order of 64 neurons/mm^3 is also critical, to measure correlational structure (not just tuning sharpness) and to detect choristers that are rare in IT. Finally, high temporal resolution is necessary, to link the correlational structure to mechanisms such as STDP and to learning behavior. Currently,

dense electrode arrays (e.g., the NeuroNexus Matrix Array) are the only technology that meets these design requirements in terms of deep sampling and sampling individual spikes *in-vivo* in behaving mammals. Unlike other technologies that are still in development, this technology is available today, and its potential for transforming neuroscience remains largely untapped.

REFERENCES

Abbott, L. F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101. doi: 10.1162/089976699300016827

Anderson, B., Mruczek, R. E., Kawasaki, K., and Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cereb. Cortex* 18, 2540–2552. doi: 10.1093/cercor/bhn015

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Baddeley, R. (1996). Searching for filters with 'interesting' output distributions: an uninteresting direction to explore? *Network* 7, 409–421. doi: 10.1088/0954-898x/7/2/021

Baker, C. I., Behrmann, M., and Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat. Neurosci.* 5, 1210–1216. doi: 10.1038/nn960

Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., and Zoccolan, D. (2013). 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. doi: 10.1371/journal.pcbi.1003167

Bansal, A. K., Madhavan, R., Agam, Y., Golby, A., Madsen, J. R., and Kreiman, G. (2014). Neural dynamics underlying target detection in the human brain. *J. Neurosci.* 34, 3042–3055. doi: 10.1523/jneurosci.3781-13.2014

Bell, A. J., and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Res.* 37, 3327–3338. doi: 10.1016/s0042-6989(97)00121-1

Bengio, Y. (2009). Learning deep architectures for AI. *Found. Trends Mach. Learn.* 2, 1–127. doi: 10.1561/2200000006

Brown, E. N., Purdon, P. L., and Van Dort, C. J. (2011). General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu. Rev. Neurosci.* 34, 601–628. doi: 10.1146/annurev-neuro-060909-153200

Butler, P. D., Silverstein, S. M., and Dakin, S. C. (2008). Visual perception and its impairment in schizophrenia. *Biol. Psychiatry* 64, 40–47. doi: 10.1016/j.biopsych.2008.03.023

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *arXiv*:1406.3284.

Cadieu, C. F., Hong, H., Yamins, D., Pinto, N., Majaj, N. J., and Dicarlo, J. J. (2013). "The neural representation benchmark and its evaluation on brain and machine," in *International Conference on Learning Representations (ICLR)*, (Scottsdale, AZ). *arXiv*:1301.3530.

Carandini, M. (2014). "Soloists and choristers in a cortical population," in Computational and Systems Neuroscience Workshop: Scalable models for high-dimensional neural data (Snowbird, UT).

Chan, V. H., Hunzinger, J. F., and Behabadi, B. F. (2011). *Method and Apparatus for Neural Temporal Coding, Learning and Recognition.* USA patent application 13/211,091.

Chu, C. C. J., Chien, P. F., and Hung, C. P. (2014). Tuning dissimilarity explains short distance decline of spontaneous spike correlation in macaque V1. *Vision Res.* 96, 113–132. doi: 10.1016/j.visres.2014.01.008

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Constantinople, C. M., and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron* 69, 1061–1068. doi: 10.1016/j.neuron.2011.02.040

Contreras, D., Destexhe, A., Sejnowski, T. J., and Steriade, M. (1997). Spatiotemporal patterns of spindle oscillations in cortex and thalamus. *J. Neurosci.* 17, 1179–1196.

Das, A., and Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature* 399, 655–661.

Dehaene, S., and Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol.* 3:e141. doi: 10.1371/journal.pbio.0030141

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341. doi: 10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Duncan, J., and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–458. doi: 10.1037//0033-295x.96.3.433

Duncan, J., and Humphreys, G. (1992). Beyond the search surface: visual search and attentional engagement. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 578–588. doi: 10.1037//0096-1523.18.2.578

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., et al. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82, 235–248. doi: 10.1016/j.neuron.2014.02.006

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327, 584–587. doi: 10.1126/science.1179867

Ecker, A. S., Berens, P., Tolias, A. S., and Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* 31, 14272–14283. doi: 10.1523/jneurosci.2539-11.2011

Eyherabide, H. G., and Samengo, I. (2013). When and why noise correlations are important in neural decoding. *J. Neurosci.* 33, 17921–17936. doi: 10.1523/jneurosci.0357-13.2013

Froudarakis, E., Berens, P., Ecker, A. S., Cotton, R. J., Sinz, F. H., Yatsenko, D., et al. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* 17, 851–857. doi: 10.1038/nn.3707

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360, 343–346. doi: 10.1038/360343a0

Gawne, T. J., and Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* 13, 2758–2771.

Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1409–1422. doi: 10.1016/s0042-6989(01)00073-6

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature* 493, 97–100. doi: 10.1038/nature11665

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron* 76, 590–602. doi: 10.1016/j.neuron.2012.08.029

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/jneurosci.1740-07.2007

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804. doi: 10.1016/S0896-6273(02)01091-7

Ito, J., Maldonado, P., Singer, W., and Grun, S. (2011). Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cereb. Cortex* 21, 2482–2497. doi: 10.1093/cercor/bhr020

Jiang, X., Bollich, A., Cox, P., Hyder, E., James, J., Gowani, S. A., et al. (2013). A quantitative link between face discrimination deficits and neuronal selectivity for faces in autism. *Neuroimage Clin.* 2, 320–331. doi: 10.1016/j.nicl.2013.02.002

Kenet, T., Arieli, A., Tsodyks, M., and Grinvald, A. (2005). "Are single cortical neurons soloists or are they obedient members of a huge orchestra?," in *Problems in Systems Neuroscience*, eds J. L. Van Hemmen and T. J. Sejnowski (New York: Oxford University Press), 160–181.

Kiani, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J. Neurophysiol.* 97, 4296–4309. doi: 10.1152/jn.00024.2007

King, P. D., Zylberberg, J., and Deweese, M. R. (2013). Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. *J. Neurosci.* 33, 5475–5485. doi: 10.1523/jneurosci.4188-12.2013

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP Abstract Supplement.*

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* 25, 1106–1114.

Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., et al. (2012). "Building high-level features using large scale unsupervised learning," in Proceedings of the 29th International Conference on Machine Learning, (Edinburgh, Scotland), 11.

Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442, 572–575. doi: 10.1038/nature04951

Leopold, D. A., O'toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* 4, 89–94. doi: 10.1038/82947

Lin, C.-P., Chen, Y.-P., and Hung, C. P. (2014). Tuning and spontaneous spike time synchrony share a common structure in macaque inferior temporal cortex. *J. Neurophysiol.* 112, 856–869. doi: 10.1152/jn.00485.2013

Logothetis, N. K., and Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science* 245, 761–763. doi: 10.1126/science.2772635

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.

Loughnan, B. L., Sebel, P. S., Thomas, D., Rutherfoord, C. F., and Rogers, H. (1987). Evoked potentials following diazepam or fentanyl. *Anaesthesia* 42, 195–198. doi: 10.1111/j.1365-2044.1987.tb02999.x

Maier, A., Logothetis, N. K., and Leopold, D. A. (2007). Context-dependent perceptual modulation of single neurons in primate visual cortex. *Proc. Natl. Acad. Sci. U S A* 104, 5620–5625. doi: 10.1073/pnas.0608489104

Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 3:e31. doi: 10.1371/journal.pcbi.0030031.eor

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). Artificial brains. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 668–673. doi: 10.1126/science.1254642

Michel, M. M., Chen, Y., Geisler, W. S., and Seidemann, E. (2013). An illusion predicted by V1 population activity implicates cortical topography in shape perception. *Nat. Neurosci.* 16, 1477–1483. doi: 10.1038/nn.3517

Miller, J. E., Ayzenshtat, I., Carrillo-Reid, L., and Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. *Proc. Natl. Acad. Sci. U S A* 111, E4053–E4061. doi: 10.1073/pnas.1406077111

Miyashita, Y. (1993). Inferior temporal cortex: where visual perception meets memory. *Annu. Rev. Neurosci.* 16, 245–263. doi: 10.1146/annurev.neuro.16.1.245

Mruczek, R. E., and Sheinberg, D. L. (2005). Distractor familiarity leads to more efficient visual search for complex stimuli. *Percept. Psychophys.* 67, 1016–1031. doi: 10.3758/bf03193628

Mruczek, R. E., and Sheinberg, D. L. (2007a). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *J. Neurosci.* 27, 2825–2836. doi: 10.1523/jneurosci.4102-06.2007

Mruczek, R. E., and Sheinberg, D. L. (2007b). Context familiarity enhances target processing by inferior temporal cortex neurons. *J. Neurosci.* 27, 8533–8545. doi: 10.1523/jneurosci.2106-07.2007

Mutch, J., and Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comput. Vis.* 80, 45–57. doi: 10.1007/s11263-007-0118-0

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Op de Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G., and Dicarlo, J. J. (2007). A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cereb. Cortex* 18, 1676–1694. doi: 10.1093/cercor/bhm196

Op de Beeck, H., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* 4, 1244–1252. doi: 10.1038/nn767

Orban, G. A., Van Essen, D., and Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.* 8, 315–324. doi: 10.1016/j.tics.2004.05.009

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. *Nature* 431, 768–774. doi: 10.1038/nature03014

Rajkai, C., Lakatos, P., Chen, C. M., Pincze, Z., Karmos, G., and Schroeder, C. E. (2008). Transient cortical excitation at the onset of visual fixation. *Cereb. Cortex* 18, 200–209. doi: 10.1093/cercor/bhm046

Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., and Baker, C. I. (2013). Slower rate of binocular rivalry in autism. *J. Neurosci.* 33, 16983–16991. doi: 10.1523/jneurosci.0448-13.2013

Rolls, E. T., and Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* 73, 713–726.

Ruff, D. A., and Cohen, M. R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. *Nat. Neurosci.* 17, 1591–1597. doi: 10.1038/nn.3835

Rust, N. C., and Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30, 12978–12995. doi: 10.1523/jneurosci.0179-10.2010

Sadovsky, A. J., and Maclean, J. N. (2014). Mouse visual neocortex supports multiple stereotyped patterns of microcircuit activity. *J. Neurosci.* 34, 7769–7777. doi: 10.1523/jneurosci.0169-14.2014

Sato, T., Uchida, G., Lescroart, M. D., Kitazono, J., Okada, M., and Tanifuji, M. (2013). Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *J. Neurosci.* 33, 16642–16656. doi: 10.1523/jneurosci.5557-12.2013

Sato, T., Uchida, G., and Tanifuji, M. (2009). Cortical columnar organization is reconsidered in inferior temporal cortex. *Cereb. Cortex* 19, 1870–1888. doi: 10.1093/cercor/bhn218

Scholl, C. A., Jiang, X., Martin, J. G., and Riesenhuber, M. (2014). Time course of shape and category selectivity revealed by EEG rapid adaptation. *J. Cogn. Neurosci.* 26, 408–421. doi: 10.1162/jocn_a_00477

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/tpami.2007.56

Shamir, M. (2014). Emerging principles of population coding: in search for the neural code. *Curr. Opin. Neurobiol.* 25, 140–148. doi: 10.1016/j.conb.2014.01.002

Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320. doi: 10.1038/415318a

Singer, J. M., and Kreiman, G. (2014). Short temporal asynchrony disrupts visual object recognition. *J. Vis.* 14:7. doi: 10.1167/14.5.7

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* 64:051904. doi: 10.1103/physreve.64.051904

Srihasam, K., Vincent, J. L., and Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783. doi: 10.1038/nn.3855

Sripati, A. P., and Olson, C. R. (2010). Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *J. Neurosci.* 30, 1258–1269. doi: 10.1523/jneurosci.1908-09.2010

Suzuki, S., and Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1315–1341. doi: 10.1037//0096-1523.24.5.1315

Takeuchi, D., Hirabayashi, T., Tamura, K., and Miyashita, Y. (2011). Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* 331, 1443–1447. doi: 10.1126/science.1199967

Tamura, H., Mori, Y., and Kaneko, H. (2014). Organization of local horizontal functional interactions between neurons in the inferior temporal cortex of macaque monkeys. *J. Neurophysiol.* 111, 2589–2602. doi: 10.1152/jn.00336.2013

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139. doi: 10.1146/annurev.neuro.19.1.109

Tang, H., Buia, C., Madhavan, R., Crone, N. E., Madsen, J. R., Anderson, W. S., et al. (2014). Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron* 83, 736–748. doi: 10.1016/j.neuron.2014.06.017

Tanigawa, H., Wang, Q., and Fujita, I. (2005). Organization of horizontal axons in the inferior temporal cortex and primary visual cortex of the macaque monkey. *Cereb. Cortex* 15, 1887–1899. doi: 10.1093/cercor/bhi067

Tootell, R. B. H., Tsao, D., and Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. *J. Neurosci.* 23, 3981–3989.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.

Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* 4, 832–838. doi: 10.1038/90547

Valentine, T. (1999). *Face-space Models of Face Recognition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Vaziri, S., Carlson, E. T., Wang, Z., and Connor, C. E. (2014). A channel for 3D environmental shape in anterior inferotemporal cortex. *Neuron* 84, 55–62. doi: 10.1016/j.neuron.2014.08.043

Verhoef, B. E., Vogels, R., and Janssen, P. (2012). Inferotemporal cortex subserves three-dimensional structure categorization. *Neuron* 73, 171–182. doi: 10.1016/j.neuron.2011.10.031

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276. doi: 10.1126/science.287.5456.1273

Wang, Y., Fujita, I., and Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat. Neurosci.* 3, 807–813. doi: 10.1038/77712

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/BRM.42.3.671

Willmore, B. D., Mazer, J. A., and Gallant, J. L. (2011). Sparse coding in striate and extrastriate visual cortex. *J. Neurophysiol.* 105, 2907–2919. doi: 10.1152/jn.00594.2010

Wu, S., Amari, S., and Nakahara, H. (2002). Population coding and decoding in a neural field: a computational study. *Neural Comput.* 14, 999–1026. doi: 10.1162/089976602753633367

Xing, D., Ringach, D. L., Hawken, M. J., and Shapley, R. M. (2011). Untuned suppression makes a major contribution to the enhancement of orientation selectivity in macaque v1. *J. Neurosci.* 31, 15972–15982. doi: 10.1523/jneurosci.2245-11.2011

Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A. N., and Tanifuji, M. (2006). Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J. Neurophysiol.* 96, 3147–3156. doi: 10.1152/jn.01224.2005

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and Dicarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U S A* 111, 8619–8624. doi: 10.1073/pnas.1403112111

Yoon, J. H., Maddock, R. J., Rokem, A., Silver, M. A., Minzenberg, M. J., Ragland, J. D., et al. (2010). GABA concentration is reduced in visual cortex in schizophrenia and correlates with orientation-specific surround suppression. *J. Neurosci.* 30, 3777–3781. doi: 10.1523/JNEUROSCI.6158-09.2010

Young, M. P., and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science* 256, 1327–1331. doi: 10.1126/science.1598577

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *ECCV 2014, Part I, LNCS 8689*, eds D. Fleet et al. (Switzerland: Springer International Publishing), 818–833.

Zoccolan, D., Kouh, M., Poggio, T., and Dicarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J. Neurosci.* 27, 12292–12307. doi: 10.1523/jneurosci.1897-07.2007

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/371358c0