Supporting Information

Okazawa et al. 10.1073/pnas.1415146112

SI Methods

Behavioral Task and Stimulus Presentation. Two female macaque monkeys (monkeys “SI” and “EV”; *Macaca fuscata*; weighing 5.3–6.2 kg) were seated in a primate chair with their heads fixed and facing the screen of a cathode ray tube monitor (frame rate: 100 Hz; Totoku Electric) situated at a distance of 57 cm from the monkeys. The monkeys were required to fixate on a small white spot (visual angle: <0.1°) at the center of the display (1.5–2.2° window). Eye position was monitored using an infrared eye camera system (ISCAN). Visual stimuli were presented on the monitor using a graphics board (VSG; Cambridge Research Systems) and calibrated with a colorimeter (CS200; KONICA MINOLTA). Image resolution was 800 × 600 pixels (20 pixels/°). A trial started with the presentation of the fixation spot, after which stimuli were presented five to six times within a trial. Each stimulus presentation lasted 200 ms with 200-ms blank intervals. Each stimulus was repeated at least five times and usually eight times in a session.

Electrophysiological Recording. Neuronal activity was recorded from the dorsal part of V4, in the prelunate gyrus. Under aseptic conditions and general anesthesia, a recording chamber was surgically attached to the skull at 0–5 mm posterior and 24–29 mm lateral (the stereotaxic coordinates). The electrical activity of well-isolated single neurons was recorded extracellularly using tungsten microelectrodes (200 μm in diameter, 1–2.5 MΩ at 1 kHz; FHC or Unique Medical). Receptive fields were manually plotted using small geometric stimuli, and their centers and sizes were determined. For cells that did not respond to the small stimuli, only the centers of the receptive fields were determined using a texture (6.4° × 6.4°). We recorded from 109 neurons that selectively responded to the presented 250–500 textures ($P < 0.0001$, Kruskal–Wallis test). Of these, 90 neurons (53 from SI and 37 from EV) were selected as samples for further analyses, based on their sparseness indices (<0.75) (1). This was defined as follows:

$$\text{sparseness index} = \left[1 - \left(\sum \frac{r}{n}\right)^2 \Big/ \sum\left(\frac{r^2}{n}\right)\right] \Big/ \left(1 - \frac{1}{n}\right), \quad \textbf{[S1]}$$

where r is the firing rate to the stimulus and n is the number of stimuli. We used this criterion because neurons with high sparseness indices (≥0.75) tended to respond only to a tiny fraction of the stimuli (on average, 2.0% of stimuli evoked more than the half-maximum response), which prohibits meaningful interpretation of quantitative analysis based on any model. Responses were computed using the mean firing rates during a 200-ms period beginning 50 ms after stimulus onset. We subtracted baseline activities, which were defined as the firing rates during the 300 ms before the onset of the first stimulus averaged across all trials in a session.

Generation of Control Images. In addition to the main textural stimuli, we also presented several control stimuli (Fig. 6*A*) for all recorded neurons. After at least five generations of adaptive sampling, we selected five textures from the presented stimuli sampled equally based on the evoked firing rates under the constraint that the images were included in the 4,400 noninterpolated images. From each of these five textures, we prepared five control stimuli, including the original stimulus: the Scramble image was generated by randomizing the phase of Fourier transform; the Rotation image was generated by rotating the original image by 90°; the Same image was synthesized using the same PS parameters as the original texture but from different random noise; and the Photo image was the photograph used to extract the synthesis parameters for the original image.

Analysis of the Efficiency of Adaptive Sampling. In *Results*, we determined whether the adaptive sampling successfully generated stimuli around the optimal textures for each neuron (Fig. 2*E*). To do this, we first defined the optimal textures as those evoking responses larger than 90% of the maximum for each neuron. The number of optimal textures was, on average, 2.4 ± 1.6. We then computed the Euclidean distance between each pair of stimuli and the nearest optimal texture in the seven-dimensional space for each stimulus. The average of the distances of all stimuli presented up to a given generation is plotted as the “Data” in Fig. 2*E*. As a control, we considered the case of simulated neurons that had the same level of selectivity for textures but whose preferred textures were randomly distributed in the sampling space (“Random tuning model” in Fig. 2*E*). To generate this control, we first randomly reassigned the response evoked by each stimulus in each neuron to all possible textures (n = 10,355) and produced the corresponding pseudoneural responses. We then simulated the same adaptive sampling experiment for these pseudoneurons and conducted exactly the same analysis with the simulated data. The results for these pseudoneurons are plotted as Random tuning model in Fig. 2*E*.

Classification of Neurons Based on the Tolerance to Control Manipulations. To determine whether the fitting weights of neurons obtained using the minPS explained the responses to those control stimuli (Fig. 7), we classified neurons into those tolerant of the image manipulations performed under each control condition or those intolerant of the manipulations. To do this, we first determined whether each of the 29 parameters in the minPS was tolerant of the manipulations of the controls. We computed the minPS parameters for our 10,355 textures and those for the control stimuli (Scramble or Rotation) corresponding to each texture and calculated the correlation coefficients between the set of values for each parameter in the minPS. When the correlation was greater than 0.4, we regarded the parameter as tolerant of the manipulations. Typical tendencies of the tolerance for each group of parameters are summarized in Fig. 7*G*. It should be noted that, because we defined the tolerance based on the correlation, exact values could be changed by the manipulations, even for parameters assigned as tolerant. For each neuron, we summed up the absolute values of weights for tolerant and intolerant parameters separately. When the summed weight for tolerant parameters was larger than that for intolerant ones, we categorized the neuron as tolerant of that image manipulation.

Computation of the Neuronal Sensitivity to Textures. The number of neurons tuned to each group of statistical parameters (Fig. 4*E*) is related to, but is not directly comparable to, the contribution of each group to the human sensitivity to the textures estimated in Freeman et al. (2). In that study, perceptual sensitivity to a particular texture was psychophysically determined as the discrimination threshold between the textural image and a spectrally matched noise image. Those investigators measured perceptual sensitivities to 500 textures and linearly fitted them using the PS statistics to specify which group of PS statistics was incorporated into images showing high perceptual sensitivity. Here, we considered reproducing these results using our neuronal data. We first computed the predicted firing rates of recorded neurons for each of our 10,355 textures and their spectrally matched noise images using the weights obtained by the fitting to the minPS. From them, we were able to estimate the sensitivity of a neuron to a given texture (d') using Eq. **S2**:

$$d' = \sqrt{\frac{(FR_t - FR_n)^2}{\sigma^2}}, \qquad \textbf{[S2]}$$

where FR_t and FR_n indicate the predicted firing rates for the texture and its corresponding noise, and σ^2 indicates the variance of the neuron's responses computed by averaging the trial-by-trial variances of the firing rates of that neuron. We then used Eq. **S3** to sum up d' for all neurons showing significant fit to the minPS ($n = 83$):

$$d'_t = \sqrt{\sum_i (d'_{ti})^2}, \qquad \textbf{[S3]}$$

where the d'_t indicates the sensitivity of the population of neurons to a particular texture t, and d'_{ti} indicates the sensitivity of neuron i to texture t. Here, we assumed the independent summation of information from individual neurons. We supposed that the population sensitivity, d'_t, corresponds to the perceptual sensitivity to each texture. Next, the obtained population neural sensitivity was linearly regressed using the PS statistics, and the contribution of each group of parameters was computed using the so-called averaging-over-orderings procedure (3), which basically computes a difference in the R^2 of the fit before and after inclusion of a particular group of parameters as a measure of the amount of contribution of that group. Because this differential R^2 depends on the order in which the group is added, the method repeats all possible orders of additions and averages over them. This yields the average percent contributions of individual groups among the R^2 computed using all parameters (Fig. 8*A*). This is the same procedure that was done for the psychophysical sensitivity in Freeman et al. (2). We used a Pearson correlation coefficient to assess the similarity between our physiological data and the data obtained in Freeman et al. (2), and the statistical significance was tested using a permutation test. In that test, we shuffled the fitting weights of individual neurons and repeated the same analysis 2,000 times to obtain the correlation coefficients for the perceptual sensitivity.

1. Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287(5456):1273–1276.
2. Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA (2013) A functional and perceptual signature of the second visual area in primates. *Nat Neurosci* 16(7):974–981.
3. Grömping U (2007) Estimators of relative importance in linear regression based on variance decomposition. *Am Stat* 61(2):139–147.

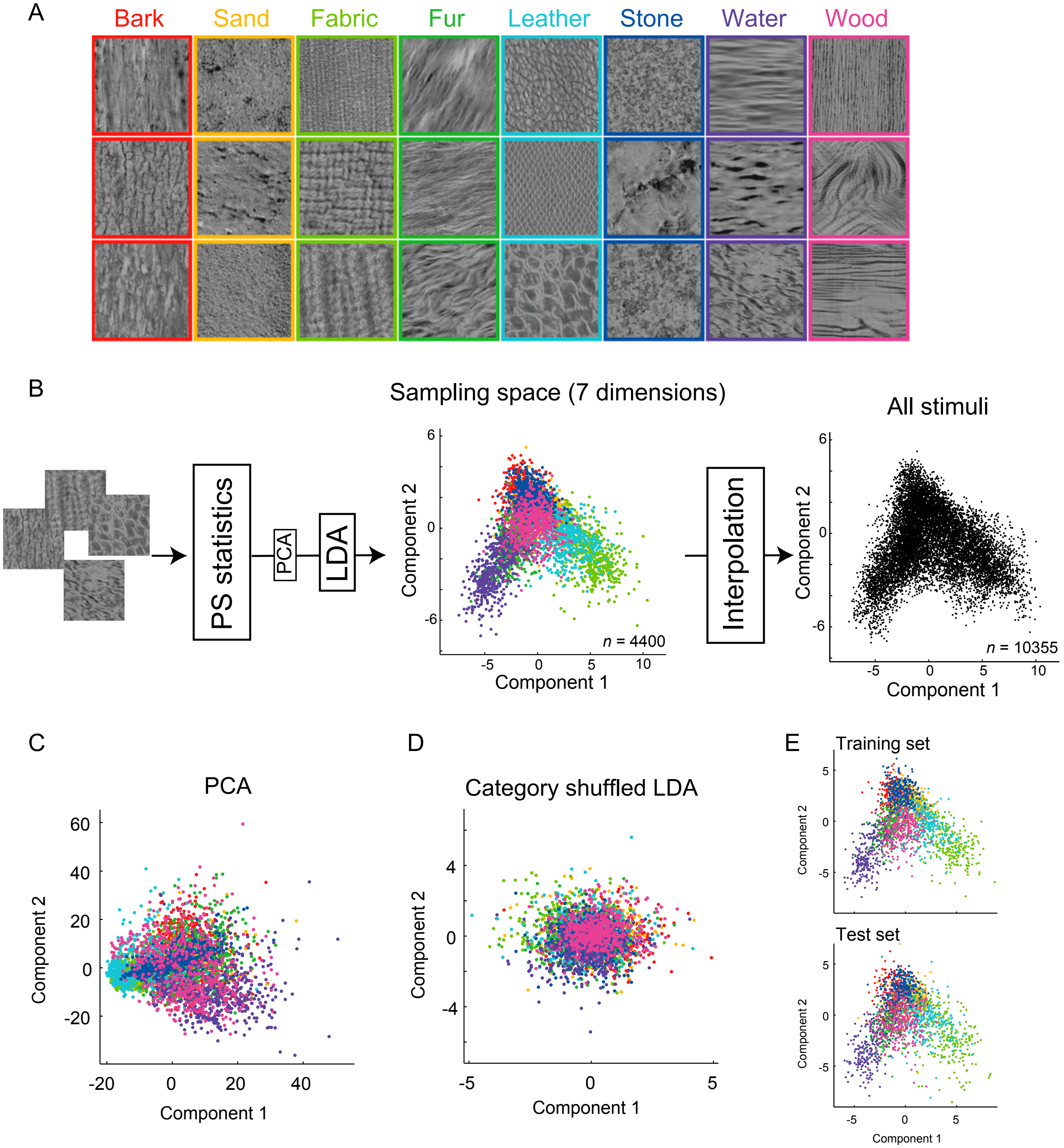


Fig. S1. Procedure used to generate the sampling space and its validation. (*A*) Examples of synthesized texture stimuli. Each texture originated from photographs of one of the eight material categories (bark, sand, fabric, fur, leather, stone, water, and wood). Frame colors indicate the corresponding material categories. (*B*) Method for generating the sampling space and stimuli for the adaptive sampling procedure. We first computed the synthesis parameters (PS statistics) of all 4,400 images (740 parameters). We then normalized individual parameters across the 4,400 images, denoised them using principal-component analysis (PCA), and finally projected them into a seven-dimensional space using Fisher's linear discriminant analysis (LDA). The middle panel depicts the distribution of textures in the first two dimensions of the resulting seven-dimensional sampling space (the same as in Fig. 1*A*). Textures in the same category tended to aggregate within this space. Because the density of the sampling was not uniform, we interpolated stimuli between neighboring textures through linear interpolation of the PS statistics. The rightmost panel shows the distribution of samples, including the interpolated ones. (*C*) When the dimension reduction was performed using PCA instead of LDA, no clear segregations of categories were observed. (*D*) When the tags of categories assigned to textures were shuffled, categories were not segregated, even with LDA. This indicates that the categorical segregation in *B* originated from the intrinsic structure hidden in the synthesis parameters. (*E*) Cross-validation of the categorical segregation in LDA. LDA coefficients were calculated using one-half of the stimuli (training set), whereas the other half (test set) was visualized using the obtained coefficients. The distributions of textures and categories were consistent between the training and test sets.

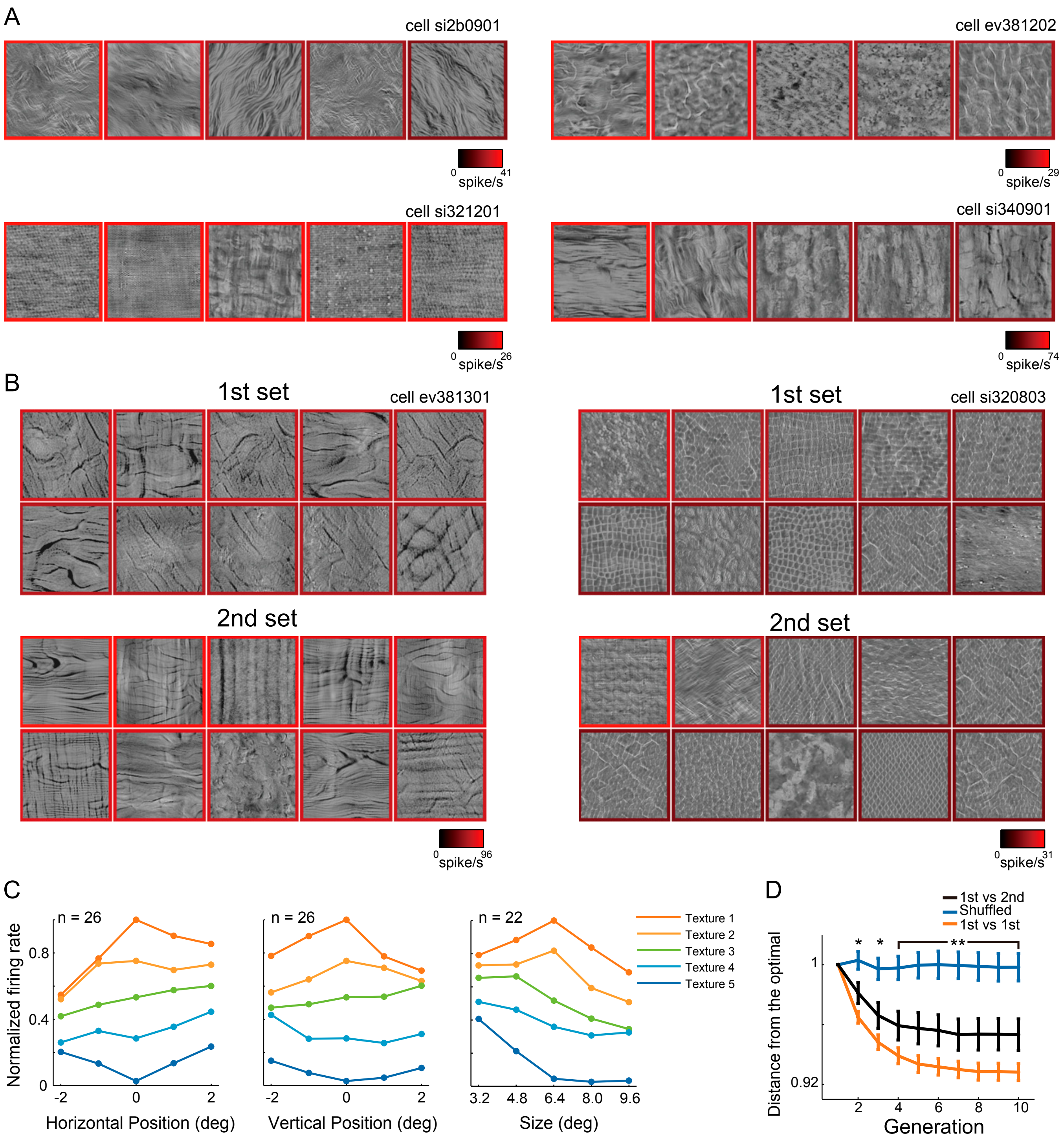
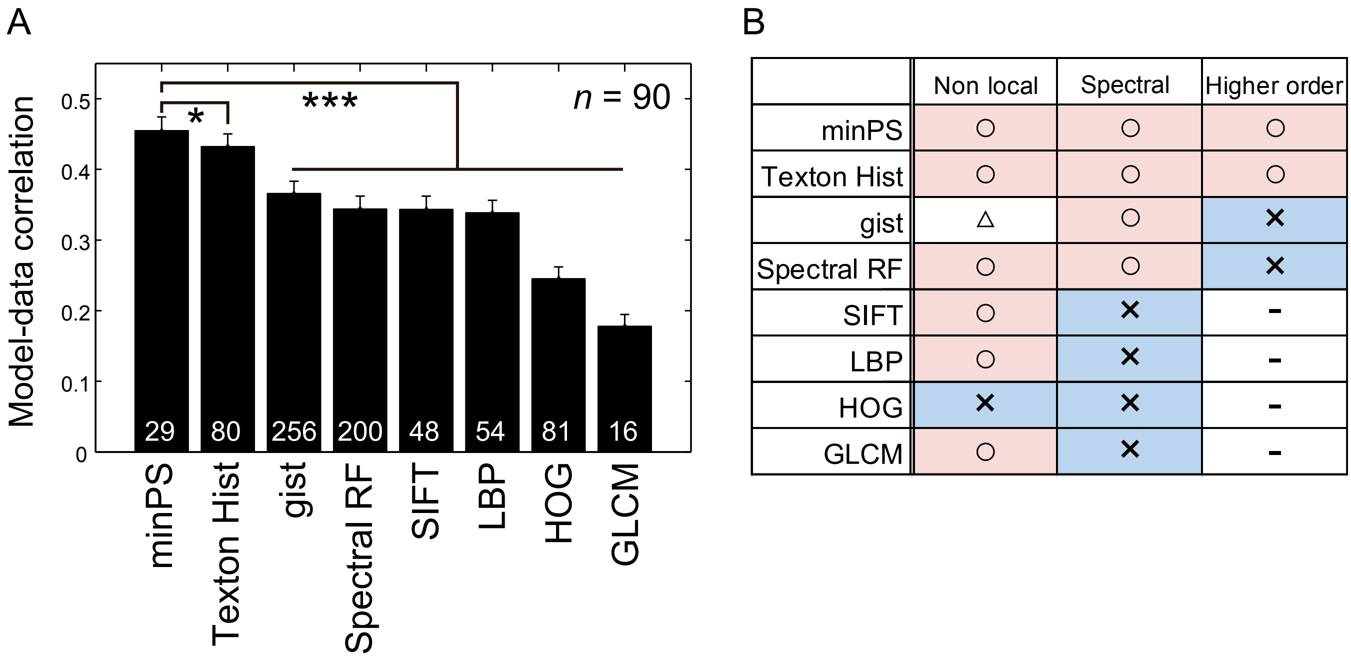


Fig. S2. Examples of the texture preferences of neurons. (*A*) Five most preferred textures for four different neurons. The preferred textures of each neuron shared common appearance, whereas different neurons preferred textures with different appearances. (*B*) To evaluate the dependency of the obtained selectivity on the first generation selected for the adaptive sampling, we conducted the adaptive procedure twice using different first generations with a subset of neurons ($n = 13$). The panels show the obtained best 10 textures with different first sets for two example neurons. The appearances of the preferred textures in the two sets were quite similar. The frame colors indicate firing rates evoked by the textures. (*C*) Tests for position and size invariances performed in a subset of neurons ($n = 26$ for position and $n = 22$ for size tests). For these control tests, five textures were selected for each cell, and they were equally sampled based on the evoked firing rates. For the position invariance test, we examined the neural responses to stimuli at five different retinal positions (0°, ±1°, ±2°) shifted horizontally (*Left*) or vertically (*Middle*) in the visual field for each of the five textures. For the size invariance test (*Right*), we examined five images of different size generated by rescaling the original images for each of the five textures. In all panels, colored lines indicate the responses to textures ranked according to the evoked firing rates in the main experiment. The firing rates are normalized by the maximum response of each cell. (*D*) To examine whether two independent samplings (lineages) starting from different first generations converged to the same peak, we determined the optimal stimuli of a given cell for one lineage and computed the average distance between the points corresponding to these optimal stimuli and points corresponding to all sampled stimuli in the

Legend continued on following page

other lineage. The panel shows the average distance between all of the textures up to a given generation in one lineage and the optimal stimuli identified in the other lineage (“first vs. second,” black line; $n = 13$). As a control, we also examined the case in which the relationships between two lineages were shuffled across cells (“Shuffled,” blue line; $n = 13$). As expected, the stimuli did not converge to the peak in the shuffled condition. The actual data (first vs. second) showed significant departures from the shuffled condition in generations 2–10, indicating that the two lineages measured from a single cell converged to the similar points in the space. For reference, the distance between the stimuli and the optimal stimuli extracted from the same lineage is shown (“first vs. first,” orange line; $n = 90$). The plot is the same with that in Fig. 2*E* (Data; orange line). *$P < 0.05$, **$P < 0.01$.



	Non local	Spectral	Higher order
minPS	○	○	○
Texton Hist	○	○	○
gist	△	○	×
Spectral RF	○	○	×
SIFT	○	×	-
LBP	○	×	-
HOG	×	×	-
GLCM	○	×	-

Fig. S3. Comparison of fitting performance achieved with minPS and other models. (*A*) The fitting performances by various models describing image features. These include models describing textures (Texton Hist, LBP, GLCM), those describing objects or scenes (gist, SIFT, HOG), and those previously used to describe neuronal activities of V4 (Spectral RF). All of these models explicitly compute the statistical properties of images. White numbers in the bars indicate the number of parameters for each model. The error bars indicate SEM across the neurons. *$P < 0.05$, ***$P < 0.001$, Wilcoxon test. Although a recent study showed that biologically plausible hierarchical network models achieve good fitting performance to the neural responses in V4 (1), we did not include them because the models do not explicitly represent the statistical parameters to compute, and the image features that evoked the neuronal responses are not shown. (*B*) Lists of features (columns) of the models (rows) arranged in the order of their fitting performance. “Non local” means that features computed in a given model do not depend on the particular spatial positions in an image. The triangle in Non local indicates that features are partially localized. “Spectral” means that a model computes features related to spatial frequency components such as Gabor-like filters. “Higher order” means that a model also incorporates features related to the correlations among different spatial frequency components. Models with good fitting performances had all three of these characteristics. Implementation of the models: We implemented the models essentially as in the original papers. We implemented or made use of available codes of many existing models and computed their statistical parameters for all presented textures. Because each model has several variants and some hyperparameters, we will briefly describe how we implemented the models. “Texton histogram” (2) is the model that describes the co-occurrence of subband features as a texton and counts up the numbers of occurrences of individual textons in an image. The number of textons is up to the users, and we set it at 80. “Gist” (3) is a model that concatenates the spatial frequency components of subregions of an image. We split an image into four by four subregions. “Spectral RF” (4) is introduced to explain the V4 responses to spatial frequency components. We simply computed the amplitudes of Fourier transforms in a range of spatial frequencies from the DC component to the 10 cycles per image components. Because we intended to compute the fitting performance, we did not take into account the intrinsic correlations between pairs of spectral channels in the stimuli, as was done in the paper (4). “SIFT” (5) is the model that captures the local gradient pattern. To do so, the model first finds several salient key points in an image and computes the local orientations of gradients around each key point. To describe an image, the model counts up the numbers of occurrences of several dominant local patterns. The number of dominant local patterns is up to the users, and we set it as 48. Similarly, “LBP” (6) counts up the numbers of occurrences of several dominant local patterns, but the way of representing local patterns is different. “HOG” (7, 8) computes the local gradients in subregions of an image. We split an image into three by three subregions. “GLCM” (9) first computes the autocorrelations of pixels and extracts several features describing the autocorrelations. We specifically extracted features called “Contrast,” “Energy,” “Homogeneity,” and “Entropy.”

1. Yamins DLK, et al. (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 111(23):8619–8624.
2. Varma M, Zisserman A (2005) A statistical approach to texture classification from single images. *Int J Comput Vis* 62(1-2):61–81.
3. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42(3):145–175.
4. David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96(6):3492–3505.
5. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110.
6. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24 (7):971–987.
7. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *Comput Vis Pattern Recog* 1:886–893.
8. Ludwig O, Delgado D, Goncalves V, Nunes U (2009) Trainable classifier-fusion schemes: An application to pedestrian detection. *12th International IEEE Conference on Intelligent Transportation Systems* (IEEE, St. Louis), pp 1–6.
9. Haralick RM, Shanmuga K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern Syst* SMC3(6):610–621.

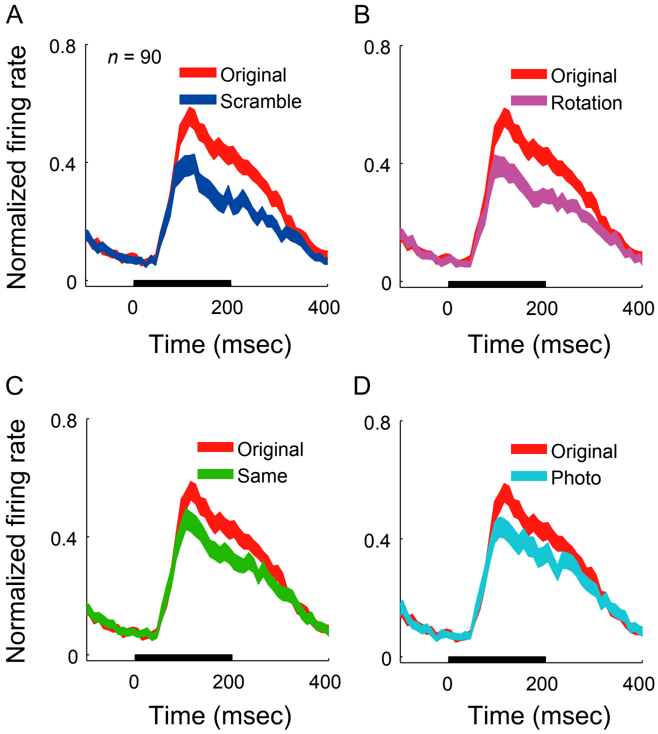


Fig. S4. Time courses of the responses to the control images. (*A*–*D*) Each panel corresponds to the indicated control condition. For an explanation of each control condition, see *SI Methods*. For each condition, five stimuli were presented and the time courses show the responses averaged across all stimuli. The time courses were computed by counting spikes in a sliding, nonoverlapping, 10-ms window. The amplitudes were normalized to each neuron's maximum response across the five stimuli in the original condition and across the time points. The line thicknesses indicate SEM across the neurons. Horizontal bars below the PSTH indicate the timing of the stimulus presentation.

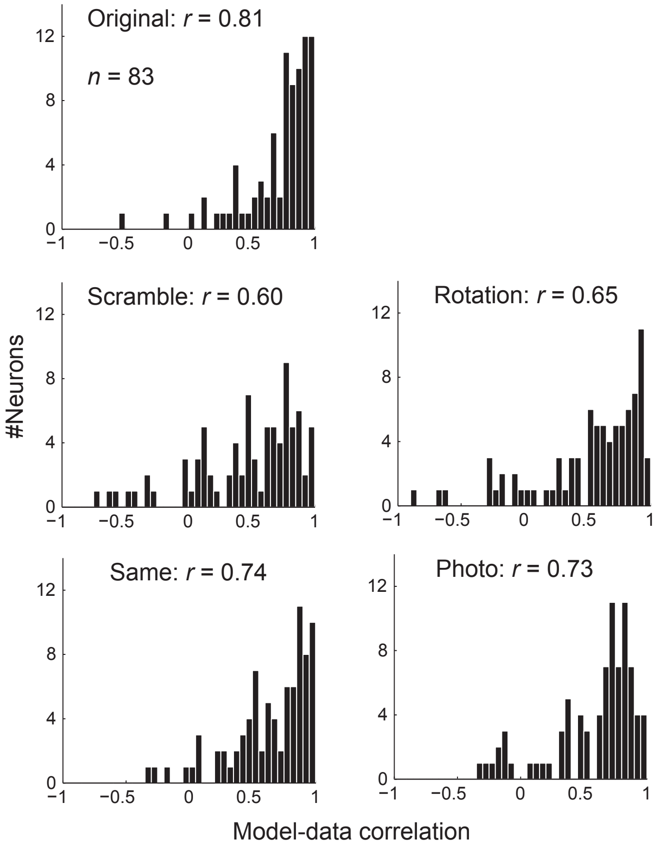


Fig. S5. Distributions of prediction performances for the responses to control stimuli. From the fitting weights obtained in the main adaptive sampling experiment (Fig. 4*B*), we estimated the predicted firing rate for each control stimulus based on the PS statistics of the image. For each neuron, the performances were evaluated as the Pearson correlation coefficients between the observed and predicted firing rates elicited by five texture images in each control condition. Numbers shown at the top of each panel indicate the median correlation coefficients across all neurons.

Table S1. Number of parameters in PS and minPS statistics

Group	Equation to derive no. parameters	No. parameters	In minPS
Spectral	$N*K + 2$	18	4
Marginal	$3 + (N + 1)*2$[†]	13	1
Linear cross position	$(N + 1) * (M^2 + 1)/2$	125	4
Linear cross scale	$2 * K * K * (N - 1)$	96	4
Energy cross position	$N * K * (M^2 + 1)/2$	400	6
Energy cross orientation	$N * ((K * (K - 1))/2 + K)$	40	6
Energy cross scale	$K * K * (N - 1)$	48	4
Total		740	29

N, number of filter scales ($N = 4$); *K*, number of filter orientations ($K = 4$); *M*, number of spatially neighboring pixels used to compute "Position" statistics ($M = 7$); No. parameters, number of parameters in each group in the PS statistics; In minPS, number of parameters in each group in the minPS statistics.

[†] "Marginal" includes skewness and kurtosis, but does not include mean, SD, minimum, and maximum because those parameters were equated across textural images. It also includes the variance of high-pass residual.

Table S2. Summary of minPS statistics

			Scale		Ori		
No.	Group	Description	F	C	V	H	Figs.
1	Spectral	Fine, vertical	○		○		5*A*, 7*A*
2		Fine, horizontal	○			○	
3		Coarse, vertical		○	○		
4		Coarse, horizontal		○		○	
5	Marginal	Skewness					5*B*, 7*B*
6	Linear cross position	PC1					
7		PC2					5*C*, 7*C*
8		PC3					
9		PC4					
10	Linear cross scale	Fine, vertical	○		○		
11		Fine, horizontal	○			○	
12		Coarse, vertical		○	○		
13		Coarse, horizontal		○		○	
14	Energy cross position	Vertical, PC1			○		
15		Horizontal, PC1				○	5*D*, 7*D*
16		Vertical, PC2			○		
17		Horizontal, PC2				○	
18		Vertical, PC3			○		
19		Horizontal, PC3				○	
20	Energy cross orientation	Fine, vertical vs. oblique	○		○		
21		Fine, vertical vs. horizontal	○		○	○	5*E*, 7*E*
22		Fine, horizontal vs. oblique	○			○	
23		Coarse, vertical vs. oblique		○	○		
24		Coarse, vertical vs. horizontal		○	○	○	
25		Coarse, horizontal vs. oblique		○		○	
26	Energy cross scale	Fine, vertical	○		○		5*F*, 7*F*
27		Fine, horizontal	○			○	
28		Coarse, vertical		○	○		
29		Coarse, horizontal		○		○	

C, coarse; F, fine; H, horizontal; Ori, orientation; PC, principal component; V, vertical. “Figs.” indicates figure numbers showing a representative neuron weighted on the parameter.