Emergence of a contrast-invariant representation of naturalistic texture in macaque visual cortex

Gerick M. Lee [1], Najib J. Majaj [1], C. L. Rodríguez Deliz [1], Lynne Kiorpes [1], J. Anthony Movshon [1,2]
[1]Center for Neural Science, New York University, New York, NY, USA
[2]Corresponding author: movshon@nyu.edu

ABSTRACT

Sensory stimuli vary across a variety of dimensions, like contrast, orientation, or texture. The brain must rely on population representations to disentangle changes in one dimension from changes in another. To understand how the visual system might extract separable stimulus representations, we recorded multiunit neuronal responses to texture images varying along two dimensions: contrast, a property represented as early as the retina, and naturalistic statistical structure, a property that modulates neuronal responses in V2 and V4, but not in V1. We measured how sites in these 3 cortical areas responded to variation in both dimensions. Contrast modulated responses in all areas. In V2 and V4, the presence of naturalistic structure both modulated responses and increased contrast sensitivity. Tuning for naturalistic structure was strongest in V4; tuning in both dimensions was most heterogeneous in V4. We measured how well populations in each area could support the linear readout of both dimensions. Populations in V2 and V4 could support the linear readout of naturalistic structure, but only in V4 did we find evidence for a robust representation that was contrast-invariant.

Significance Statement

Single neurons in visual cortex respond selectively to multiple stimulus dimensions, so signals from single neurons cannot distinguish changes in one dimension from changes in another. We measured responses from simultaneously recorded neural populations in three hierarchically linked visual areas – V1, V2, and V4 – using texture stimuli that varied in two dimensions, contrast and naturalistic image structure. We used linear decoding methods to extract information about each dimension. In all three areas, contrast could be decoded independently of image structure. Only in V4, however, could image structure be decoded independently of contrast. The reason is that selectivity for texture and contrast in V4 was much more diverse than in V1 or V2. This heterogeneity allows V4 to faithfully represent naturalistic image structure independent of contrast.

INTRODUCTION

The responses of sensory neurons are modulated by multiple stimulus dimensions, so a change in neuronal response does not unambiguously represent a change in any particular dimension. Algebraically, such a system is underdetermined because the representations of many stimulus dimensions are "entangled". To support the readout of multiple stimulus dimensions, there must be neurons that are diverse in their stimulus selectivity – a population of identically tuned neurons would fare no better than one.

Consider a case in point. The primates reading this paper can do so across a variety of contrast and lighting conditions, from backlit digital displays with strongly contrasting black text set against a bright white background, to poorly printed paper copies in a dim room. This "contrast-invariant" capacity requires the perceived identity of a word to remain stable across different lighting conditions. It also exemplifies the larger computational problem – the brain must represent a large number of sensory dimensions (e.g., contrast, stimulus identity), and must flexibly support the readout of any of these dimensions, based on the perceptual demands of the moment (DiCarlo and Cox, 2007; DiCarlo et al., 2012).

Here, we focused on two stimulus dimensions. One, visual contrast, is explicitly encoded by neurons in the retina (Enroth-Cugell and Robson, 1966; Shapley and Enroth-Cugell, 1984). The second concerns the structure of natural images. A characteristic feature of these images is the presence of structures that reflect spatial correlations, like edges, lines, and contours (Portilla and Simoncelli, 2000). Texture images that contain this "naturalistic" structure often drive neural activity in visual areas V2 and V4, more strongly than "noise" textures, matched in orientation and spatial frequency content (Freeman et al., 2013; Lee et al., 2024; Okazawa et al., 2017). The presence of naturalistic structure does not modulate neuronal activity in V1 (Freeman et al., 2013; Lee et al., 2024; Ziemba et al., 2024, 2019).

We recorded the responses of neural populations in areas V1, V2, and V4 of awake macaque monkeys during passive fixation (using multielectrode arrays). We presented texture images, with and without naturalistic structure, at a range of contrasts. We measured how contrast and texture were encoded at individual multiunit sites. Sites in all areas were modulated by contrast. Sites in V2 were often also modulated by naturalistic structure; sites in V4 more so. In V2 and V4, encoding of contrast and texture intersected: contrast sensitivities were higher in response to naturalistic textures than to noise textures. In V4, the distribution of tuning for texture and contrast was broad. We measured how well populations in all 3 areas could separately encode both contrast and naturalistic structure. As expected, the V1 population encoded only contrast, and could not distinguish naturalistic structure. The V2 population encoded both variables, but the representations of naturalistic structure and contrast in V2 were largely entangled – linear decoders for texture type were often confounded by changes in contrast. In V4, we found a contrast-invariant representation of naturalistic structure which could be extracted using a simple, biologically-plausible mechanism.

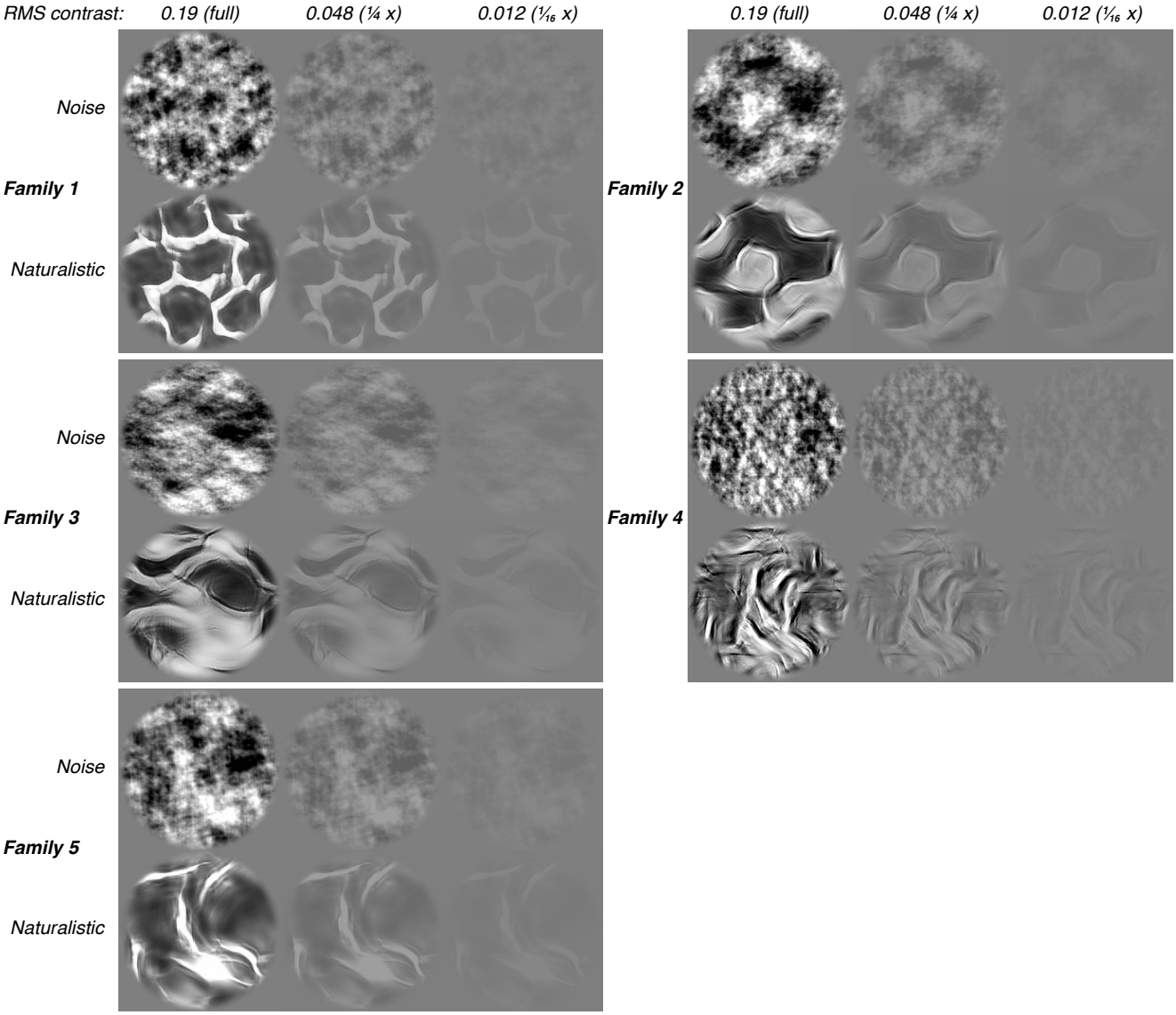

Figure 1: Synthetic texture stimuli. Texture images as used in our experiments. We synthesized "noise" (above) and "naturalistic" (below) textures to match the average spectral statistics of natural images. We used 5 natural images to generate the different texture "families" shown here. We displayed these images at 7 total contrasts, 3 of which are shown here. Note the presence of sharp, continuous elements (e.g., edges, lines, contours) in the naturalistic textures; these elements are largely absent from the noise textures. The difference between naturalistic and noise textures is perceptible at all contrasts, even near the threshold for seeing.

MATERIALS AND METHODS

Visual stimuli

As in previous experiments (Lee et al., 2024), we generated stimuli using the texture model of Portilla and Simoncelli (2000), using the methods detailed in Freeman et al. (2013). Briefly, we used photographs of repeating patterns, including both natural and artificial objects, to generate synthetic textures. We used each photograph to generate a texture "family". We first synthesized "noise" textures, which matched the average spectral content of the original image, as measured with steerable pyramid filters (Portilla and Simoncelli, 2000). We then synthesized "naturalistic" textures, which matched both the outputs of the oriented filters (as with the noise textures), as well as the correlations between filter types. For each texture family, we ran the synthesis process multiple times, starting with different white noise images. The resultant "samples" were all matched in their spectra and statistics, but differed in local spatial detail.

We used textures from 5 families (Figure 1). We used 5 different samples of naturalistic and noise textures from each family. We adjusted contrast in octave steps down from the root-mean-square contrast of the original image. Our image set totaled 400 images: 5 (families) $\times$ 5 (samples) $\times$ 7 (contrasts) $\times$ 2 (naturalistic or noise) = 350 texture images, plus 50 blank images used to estimate baseline firing rates.

We presented stimuli on a gamma-corrected CRT monitor with a mean luminance of 28 cd m^{-2}, a resolution of 1280 by 960 pixels, and a frame rate of 100 Hz. We seated animals in a custom primate chair 114 cm from the monitor, at which distance it subtended 20 by 15 deg.

Neurophysiology

Experimental procedure

We trained two juvenile *Macaca nemestrina* monkeys, both female, to fixate the central 3 deg of a computer monitor. The display background was gray, and the center was marked with a red square which was 0.1-0.2 deg across. Texture images appeared after 160 ms in a pseudorandom order. We presented blocks of 4 images, and showed each image for 200 ms, followed by a 200 ms blank interval.

All texture images subtended 6.4 deg and had the same mean luminance as the gray background. Receptive fields were centered within 1.5 deg of the center of gaze, so we centered stimuli at the center of the monitor. The visual stimuli covered the aggregate receptive fields of all recording sites.

We rewarded animals for holding fixation through a block with a juice reward. If an animal broke fixation during a stimulus, we interrupted presentation and presented the interrupted stimulus again later in the overall sequence. We recorded 10 repetitions of each stimulus.

We performed all animal procedures in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* (2011), and with the approval of the New York University Animal Welfare Committee.

Neural recording

We recorded multiunit neural responses at the ages of 32 (animal M1) and 30 (animal M2) weeks. In these same animals we have previously shown that the encoding of naturalistic structure is mature by this age (Lee et al., 2024)). The experiments reported in this paper were conducted as part of the same experimental series as those of Lee et al. (2024).

We implanted 96-electrode ("Utah") recording arrays into the left hemispheres under sterile conditions. Electrodes were arranged in a 10$\times$10 square (four positions were not used for recording), had shanks 1 mm

long and had an interelectrode spacing of 0.4 mm. Following implantation, we used anatomical landmarks (including gross anatomical landmarks such as sulci and vasculature, which we observed surgically and related to previously documented area boundaries Saleem and Logothetis (2012); Winters et al. (1969)), physiological response properties (e.g. response latencies), and receptive field characteristics to determine the cortical location of array sites. All sites on an array were typically within a single visual area, with one exception, where the array lay on the border between V1 and V2. We recorded a total of 58 responsive sites from V1 (from animal M1), 115 sites in V2 (32 from M1 and 83 from M2), and 108 sites in V4 (86 from M1 and 22 from M2).

We recorded bandpass filtered (250 Hz to 7.5 kHz) electrical activity, at a sampling rate of 30 kHz. To minimize the influence of common-mode signals, we subtracted the median sample-by-sample voltage across all sites, and then defined spike events (threshold crossings) as negative voltage deviations at least 3 times the root-mean-squared deviation of the baseline voltage.

Physiological data analysis

Single-site analysis

For analysis, we binned multiunit threshold crossings into 10 ms windows. To determine single-site response latencies, we separately measured the discriminability between blank stimuli, and either naturalistic texture or noise texture stimuli as a d' between response distributions: $\frac{\mu_{stim}-\mu_{blank}}{\sqrt{\frac{1}{2}(\sigma^2_{stim}+\sigma^2_{blank})}}$, where μ reflects the mean of a distribution, and σ its standard deviation. We defined the onset time for a given site as the first deviation above a d' of $\pm$ 0.4, as long as the following 2 windows were also suprathreshold (and of the same sign). The onset time for a given population was the median onset time across visually responsive sites. For most analyses, we then summed responses across the 200 ms window following stimulus onset.

To determine whether a site was visually responsive in a given recording session, we combined two metrics. We measured the stimulus-blank discriminability (as a d') for the first 200 ms of the visually-evoked response. We also measured the average response separately for even and odd repetitions of each stimulus. We measured the Pearson correlation between these vectors. We used sites with both a discriminability of 0.1 or greater, and a correlation of 0.4 or greater (the results reported here remained stable across a variety of different metrics and thresholds).

We measured receptive field locations and sizes by recording neural responses to a small spot, which was tiled across the visual field. We used d' values, summed for 200 ms following response onset, between spot stimuli and blanks to estimate receptive field centers and sizes. All receptive fields were centered roughly 1.5 deg from the center of gaze (below and right of the horizontal and vertical meridian, respectively).

As a single site measure of naturalistic texture discriminability, we measured the d' between naturalistic and noise texture evoked response distributions.

To model the relationship between stimulus contrast and single-site response magnitude, we measured the average response to all stimuli at a given contrast, which we fit using the supersaturating model of Peirce (2007):

$$R(c) = R_{maxcon}\frac{c^n}{c^{sn}_{50m} + c^{sn}} + B,$$

where R is the response to a given contrast, R_{maxcon} is the fit asymptotic value (not necessarily the

maximum overall response) of the site. Both n and s are fit exponents, and B is the fit baseline response. This model is an adaptation of the earlier Naka-Rushton model (Naka and Rushton, 1966). Peirce's principal innovation is the supersaturation parameter s, which enables the model to fit non-monotonic contrast-response data. The presence of the s parameter means that the c_{50m} and R_{maxcon} parameters do not refer to the half-maximum contrast, nor the maximum response rate, as in the Naka-Rushton model. Here, we refer to them as c_{50m} and R_{maxcon} to indicate that the former parameter is only relevant for the model (the m in c_{50m}), while the latter refers to the response at maximum contrast, not necessarily the overall maximum response. For our reported half-maximum contrast values, we parametrically extracted the actual contrast value, referred to as c_{50}, which represents the contrast eliciting a half-maximum response.

Population analysis

To measure population discriminability, we used a training set M_{train}, containing responses to naturalistic and noise texture images from a given texture family, to compute linear classifiers. A linear classifier is a hyperplane $\vec{b}$ that separates responses by class in M_{train}. To compute these classifiers, we used a linear discriminant, given by the weighted sum of responses from all sites that best separates two response distributions defined by a stimulus feature.

Each column in the M_{train} matrix contains data from a single recording site, each row corresponds to a response to an individual stimulus presentation, each entry contains the response of a given site on that trial. To measure discriminability we projected held out data M_{test}, onto the orthogonal axis as $M_{test}\vec{b}$. We measured these distances in d' units. We recorded responses to 5 texture samples for each texture family and contrast. For cross-validation, we used 4 of these samples in M_{train}, and the fifth in M_{test}. We repeated this five-fold cross-validation using all 5 samples.

We measured performance across a number of population sizes, octave-spaced from 2 to 64. Because different recording arrays yielded different numbers of sites, we diagonalized the covariance matrix to remove the impact of correlations between sites. For the results reported here, we used a population of size 32, the largest size that allowed us to compare results across areas. Our results were robust across the other population sizes. To estimate variability, for each population size n_{pop}, we used 40 random subsamples of size n_{pop}, drawn from the total sample of visually-responsive sites in that area. Population subsamples were nested: for example, for subsamples of size 4, the first 2 sites were the same 2 sites as those used in our subsampled population of size 2.

We trained two classifiers for each texture family: one to discriminate naturalistic textures from noise, using responses to all contrasts, and another to discriminate higher contrast textures from lower contrasts, using responses to both naturalistic textures and noise. This yielded two linear axes in the population response space: a naturalistic-noise axis, and a contrast axis. We used Gram-Schmidt orthogonalization to compute a third axis, a naturalistic-noise axis which is independent of contrast. We measured discriminability between naturalistic textures and noise using both the original naturalistic-noise axis, and the contrast-independent axis (Figure 4 B and C).

Statistical testing

Where necessary, we used bootstrap resampling to estimate confidence intervals. We performed permutation ANOVAs, after Anderson (2001), to measure area-level effects. When significant, we then performed pairwise permutation tests to compare effects either between or within areas.

RESULTS

Neuronal responses in V2 and V4 are often modulated by the presence of naturalistic structure within images, while responses in V1 are not. To understand how this modulation is affected by changes in contrast, we used data from 58 multiunit sites from V1, 115 sites from V2, and 108 from V4. We presented naturalistic textures, and their spectrally matched noise counterparts, at different contrasts (Figure 1).

Characterizing contrast sensitivity

In all 3 areas, we observed strong effects of contrast on both the magnitude and onset latency of responses (Figure 2A). Responses in V1 were not modulated by the presence of naturalistic structure. The average response in V2 and V4 to naturalistic textures was stronger than the response to spectrally-matched noise textures. In V2, the response difference between naturalistic and noise textures varied with contrast. In V4, the response difference persisted across all contrasts. In V4, low contrast stimuli evoked an onset response which was weaker and slower than to high contrast stimuli. Yet even at low contrasts, naturalistic textures evoked a larger response than noise textures.

We measured the relationship between contrast and response at the level of individual sites (Figure 2B). In V1, responses increased monotonically with contrast. Responses to naturalistic and noise textures were comparable at all contrasts. Responses in V2 also increased monotonically with contrast. At higher contrasts, both example sites responded more to naturalistic textures than to noise textures. In V4, we saw two changes relative to V1 and V2 – first, we noticed a divergence between sites' contrast sensitivity measured using naturalistic textures, versus their noise texture evoked contrast sensitivity. As a result, while V4 response magnitudes varied with contrast, a selectivity for naturalistic texture was visible across a wide range of contrasts. Second, we noticed an increased prevalence of non-monotonic contrast-response relationships, typified by the bottom right example in Figure 2B.

To quantify the contrast response function across all sites in our sample, we fit responses from all sites with the model of Peirce (2007), which uses a modified Naka-Rushton function to relate response and contrast. The model includes a supersaturation parameter, which allows it to capture non-monotonic contrast-response relationships. We used the model fits to measure the contrast at which responses reached half of their maximum (c_{50}), which we take as a proxy for contrast threshold, the inverse of which is contrast sensitivity. Lower c_{50} values indicate higher contrast sensitivities. In the examples shown in Figure 2B, contrast sensitivity was higher in V2 and V4 than in V1.

Figure 2C shows the contrast response functions for all sites, in response to naturalistic (above) and noise textures (below). For naturalistic textures, contrast response functions taken in V4 (orange traces) were typically to the left of contrast response functions taken in V1 and V2 (green and blue traces, respectively), suggesting that V4 sites had higher contrast sensitivity for naturalistic textures. This difference was less evident for noise textures. Supersaturating contrast response functions were also more prevalent in V4 than in V1 or V2, especially for responses to naturalistic textures.

We compared half maximum response (c_{50}) values, measured separately for naturalistic and noise textures, for all areas (Figure 2D). Taking these c_{50} values as measures of contrast threshold, the figure reveals two main trends in the data. First, contrast thresholds fell (i.e., sensitivity increases) from V1 to V2, and again from V2 to V4. Second, the difference in contrast threshold between responses to naturalistic and noise textures was largest in V4, smaller in V2, and absent in V1. For V2 and V4, thresholds were lower

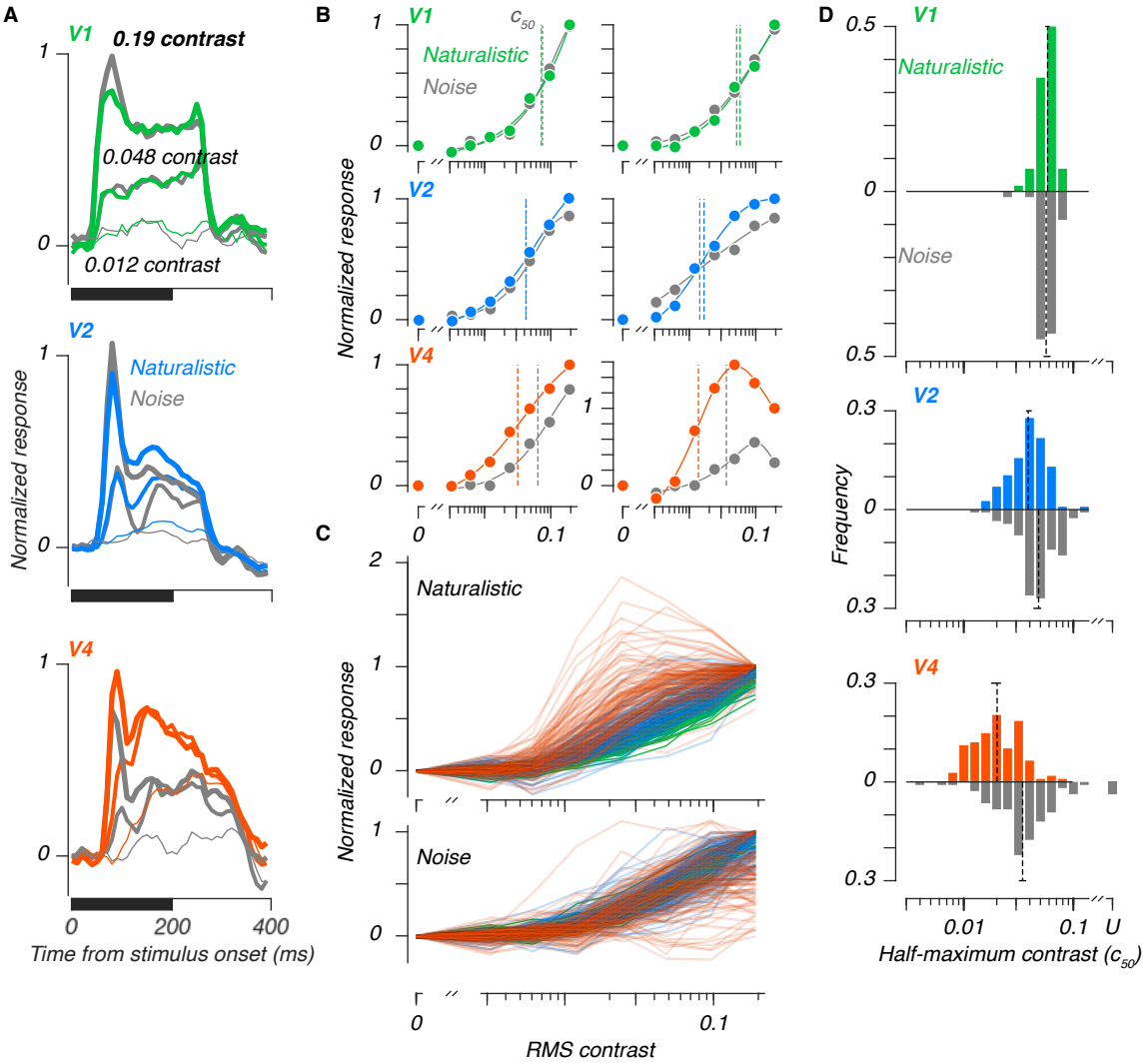

Figure 2: Multiunit neural responses to texture images at different contrasts. A: Post-stimulus time histograms, each normalized to the maximum response, then averaged across all responsive sites, to naturalistic and noise textures at 3 different contrasts (the same contrasts shown in Figure 1). B: Responses for six example sites, for the first 200 ms of the visually evoked response. Responses are normalized to the naturalistic texture response at maximum contrast. Points show average responses to naturalistic (full color) and noise textures (gray), solid lines show model fits, using the model of Peirce (2007). Half-maximum contrast values for naturalistic and noise textures (c_{50}) are marked with vertical dashed lines (colored and gray, respectively). C: Contrast responses to texture stimuli for all sites. The upper panel shows responses to naturalistic textures, The lower panel shows responses to noise textures. Both sets of curves are normalized to the largest response at maximum contrast for each site, whether it was evoked by naturalistic or noise stimuli. D: Distributions of half maximum contrast values (c_{50}) for each area and stimulus type. Vertical dashed lines and numerical values show medians. The horizontal line at the extreme end of each dashed line shows the 95% confidence interval around the median. For some sites in V4, a half maximum contrast could not be fit for the data to noise textures, because the response was too weak or variable. These "undefined" sites (U) are offset to the right, and were not used to compute medians. We used the c_{50} values as a proxy for contrast threshold.

for naturalistic textures than for noise.

Barring the naturalistic-noise difference in V1, these effects were statistically significant: we tested interareal contrast thresholds via permutation ANOVA, followed by pairwise permutation tests (permutation ANOVA, Anderson (2001), naturalistic: $F = 58, p < 0.001$; noise: $F = 11, p < 0.001$, pairwise permutation tests: V1-V2 noise texture difference: $p < 0.003$; all other comparisons $p < 0.001$). We tested the differences between naturalistic and noise texture evoked contrast thresholds via pairwise permutation tests (V1: $p > 0.05$, V2: $p < 0.001$, V4: $p < 0.001$).

We also measured the monotonicity of each contrast response function, using the fits to the model of Peirce (2007). As shown above (Figure 2B and C), all V1 responses to naturalistic and noise textures increased monotonically with contrast. In V2, 4 of 115 sites (3%) had non-monotonic contrast responses. In V4, 26 of 108 sites (24%) of sites had a non-monotonic contrast response to noise textures, while 38 of 108 (35%) had a non-monotonic contrast response to naturalistic textures.

Single site encoding of naturalistic structure and contrast

Having characterized the relationship between stimulus contrast and neural response, we wanted to measure the selectivity of each site for naturalistic image structure, first between full contrast naturalistic and noise texture pairs, then for all other contrast pairs.

We used the separation between naturalistic and noise texture response distributions (Figure 3A) to measure texture selectivity. Measured at full contrast, this comported with prior observations – on average, sites in V1 barely preferred noise textures over naturalistic ones (mean d' [95% CI]: -0.09 [-0.13, -0.04]); sites in V2 and V4 preferred naturalistic textures to noise (V2 mean d' [95% CI]: 0.16 [0.11, 0.22]; V4: 0.59 [0.49, 0.69]). We compared each of the 3 distributions, and found a significant effect of area on selectivity (permutation ANOVA: $F = 23, p < 0.001$). Followup comparisons confirmed that V2 was significantly more selective than V1 (mean difference: 0.25; permutation test, $p < 0.001$). V4 was significantly more selective than V2 (mean difference: 0.43, permutation test, $p < 0.001$), and V1 (mean difference: 0.67, permutation test, $p < 0.001$). These relationships largely agree with prior observations, with the exception that the V4-V2 difference reported here is larger than in our previous study (Lee et al., 2024). For that study, we recorded neuronal responses to 35 texture families, only at maximum contrast. In this study, we used only 5, chosen from the larger set on the basis of their ability to elicit a strong differential response between naturalistic and noise. When we analyzed the subset containing these 5 families in the previous dataset, we observed a comparable V4-V2 difference.

Having confirmed that V2 and V4 were both sensitive to the presence of naturalistic structure within texture images, we related each site's selectivity for naturalistic texture to its corresponding contrast selectivity – the difference in response between full contrast stimuli (including both naturalistic and noise textures), and the lowest (non-zero) contrast (Figure 3B). In V1, we observed a broad distribution of selectivity for contrast, while texture selectivity was negligible. Sites in V2 and V4 varied in their tuning for both contrast and texture. There was a large degree of variation along both axes in V4 – contrast and texture selectivity were heterogeneously distributed.

We computed the mean naturalistic-noise texture selectivity for all possible contrast pairs. Each heatmap in Figure 3C shows the texture selectivity for each of the 49 possible contrast pairs (7 naturalistic contrasts by 7 noise contrasts). The top right corner of each heatmap shows the same full contrast comparison as Figure 3A. Values along the identity diagonal – plotted from bottom left to top right –

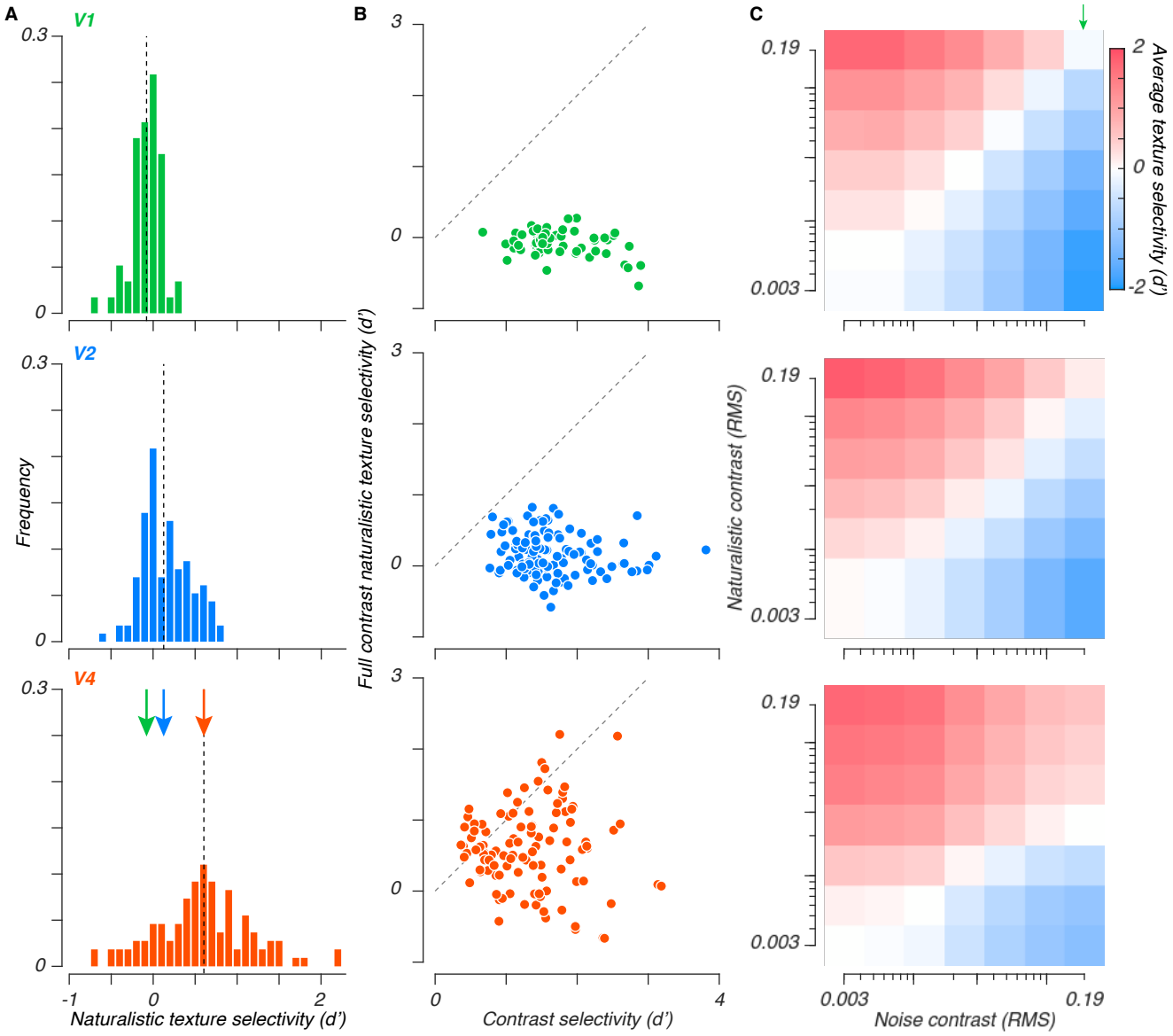

Figure 3: Single site texture selectivity. A: Distributions of single site texture selectivity, measured between the responses to full contrast naturalistic and noise texture images. Dashed lines and arrows show means. For reference, the selectivities for the example sites in Figure 2B were: 0.03 and 0.02 (V1), 0.32 and 0.29 (V2), 0.65 and 0.64 (V4), respectively. B: Contrast selectivity, measured as the d' separating all responses to full contrast, from all responses to the lowest (non-zero) contrast, versus full contrast texture selectivity, as measured in panel A. Note that the identity lines (dashed gray) are vertically offset from the center of the plot. C: Mean single site texture selectivities, measured between responses to naturalistic and noise textures for all possible contrast pairs. The full-contrast selectivities shown in panel A are at the top right corner, indicated with a green arrow. In V1 and V2, selectivity is dominated by contrast – most heatmap cells above the identity diagonal are red, while most below are blue. In V4, the interaction between contrast and image structure is shown by the larger red region, showing contrast pairs for which naturalistic textures evoke stronger responses.

show the 7 matched contrast comparisons. Off-diagonal values show selectivities which reflect an area's combined modulation by contrast and texture.

Responses in V1 were not selective for naturalistic structure – values were close to zero along the identity diagonal. Responses in V1 were selective for contrast – values for textures differing in contrast (off-diagonal) were large and reflected the ratio of contrasts (i.e., the difference in log contrasts) – selectivities were largest for pairs with the largest contrast ratio.

Unlike in V1, responses in V2 were selective for naturalistic structure, evidenced by the positive selectivities along the identity diagonal. But as in V1, responses in V2 were selective for contrast. For all off-diagonal contrast pairs, values reflected the contrast ratio, not texture. All comparisons below the identity diagonal were negative, indicating that tuning for contrast was stronger than tuning for naturalistic structure (i.e., sites responded more vigorously to higher contrast noise textures, than to lower contrast naturalistic textures).

Responses in V4 were selective for naturalistic structure. Unlike in V2, naturalistic structure outweighed contrast for a number of unequal contrast pairs – demonstrated by the presence of a number of positive selectivities below the identity diagonal.

Even in V4, selectivities were negative if the contrast ratio between lower contrast naturalistic and higher contrast noise was great enough (e.g., the bottom right corner of Figure 3C). Taken at face value, this result would make a paradoxical prediction about texture perception, best demonstrated by Figure 1 – specifically, the rightmost (lowest contrast) naturalistic textures, and leftmost (highest contrast) noise textures. For these comparisons, a typical human observer can readily identify which is the more naturalistic. If behavioral texture discrimination resulted from the simple average of single site activity, as seen in Figure 3C, observers would not only be unable to correctly distinguish low contrast naturalistic textures from full contrast noise (i.e., chance performance), but also they would consistently perform *below* chance, and mislabel high contrast noise textures as naturalistic. A glance at Figure 1 demonstrates that this is not consistent with human experience. We therefore wanted to consider how ensembles of neurons might be able to disambiguate contrast from naturalistic structure.

The population code for naturalistic structure generalizes across contrast in V4

We trained two linear classifiers, one to discriminate contrasts (using both naturalistic and noise textures), and another to discriminate naturalistic from noise textures (using responses to all pairs of contrasts). We trained and tested each classifier on multiple randomly sampled populations of 32 sites. At the population level, contrast discrimination was comparable between V1, V2, and V4 – all off-diagonal regions of the heatmaps are colored to indicate that contrasts are correctly discriminated by the populations regardless of naturalistic structure (Figure 4A).

We used the second classifier to measure performance across all contrast pairs for naturalistic versus noise textures (Figure 4B) – a correct classifier for naturalistic texture should yield positive classification performance for all contrast pairs, i.e. all heatmap cells should be red. V1 could not reliable discriminate naturalistic textures from noise, and instead often mislabeled low contrast stimuli as naturalistic textures. This shows that the population did not reliably represent naturalistic statistics, leaving the classifier to latch on to the nuisance dimension of contrast. Populations in both V2 and V4 supported texture discrimination for many contrast pairs. Yet as was the case for the single-site average (Figure 3B), the representations of texture and contrast were entangled, leading to below-chance performance for contrast

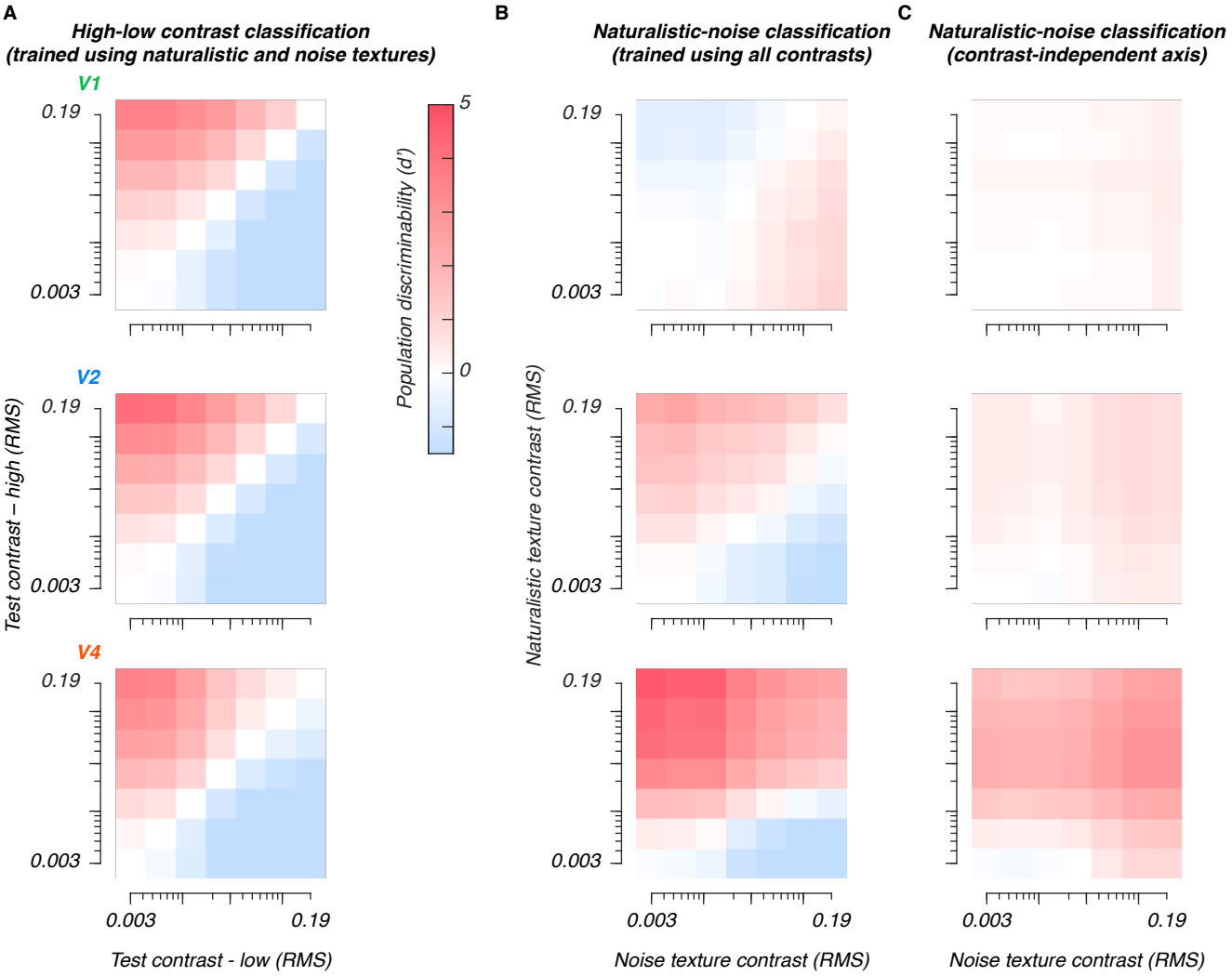

Figure 4: Population contrast and texture discriminabilities. A: Mean contrast discriminability, based on populations of 32 sites. We trained a classifier to discriminate high from low contrast textures, using neural responses to both naturalistic and noise textures. We tested performance separately for each contrast pair (using held-out data). Perfect performance for this decoder would yield a heatmap that is red above the identity diagonal and blue below it – very much as shown. B: Mean naturalistic-noise texture discriminability, based on populations of 32 sites. We trained a classifier to discriminate naturalistic textures from noise textures, using neural responses to all contrasts. We tested performance separately for each contrast pair. Perfect performance for this decoder would yield a heatmap that is everywhere red. Plainly performance is incorrect (blue) in many cases. C: Mean naturalistic-noise texture discriminability, based on populations of 32 sites. We used the Gram-Schmidt process to obtain a third classifier: a naturalistic-noise classifier, orthogonal to the contrast classifier in panel A. We used this contrast-independent classifier to test performance separately for each contrast pair. Perfect performance for this decoder would again yield a heatmap that is everywhere red. Performance is generally excellent for V4, and often correct though much weaker in V2 and V1.

pairs that contained low contrast naturalistic textures and high contrast noise (blue zones at the bottom right corner of both heatmaps). A suitable measure of classifier peformance in this case is the mean d' for all values below the identity diagonal, i.e. for all cases in which the naturalistic texture was of lower contrast than the noise texture. These means were weakly positive for V1 (0.48), and negative (i.e. incorrect) for V2 (-0.54) and V4 (-0.14).

If neurons in a population differ from one another in their tuning for each dimension, information can be read out by orthogonalizing the representation of the nuisance dimension (in this case, contrast). If all sites are identically tuned for a pair of stimulus dimensions, the system is underdetermined, and the dimensions will remain entangled. Our measurements in V4 suggested that tuning for contrast and texture was heterogeneous (Figure 3B). We therefore expected that a correctly chosen linear classifier could leverage the diversity in texture and contrast tuning seen in the V4 population to support contrast-invariant texture discrimination.

The two classifiers we trained – the contrast discriminant and the texture discriminant – were both vectors in the same population response space. As a result, we were able to use standard linear methods (Gram-Schmidt process) to project the contrast discriminant axis (Figure 4A) out of the texture discriminant axis (Figure 4B). This resulted in a third discriminant – a new texture discriminant axis *orthogonal* to the contrast discriminant. We used this contrast-independent axis to again measure how well each area could discriminate naturalistic textures from noise (Figure 4C). Performance in V1 was close to zero for all contrast pairs, suggesting that the original learned representation of naturalistic structure was tuned to stimulus contrast, not texture statistics. In V2, performance using the orthogonalized classifier was also relatively low for all contrast pairs, evidence that the representation of texture was entangled with that of contrast. In V4, performance was relatively stable across all contrasts, consistent with a contrast-invariant texture representation. All three areas do show correct performance for all contrast pairs (all cells are to some degree red). Using the same figure of merit as above, we now observed robust performance in V4 even for cases in which contrast and naturalistic statistics were in conflict (mean d' of heatmap cells below the identity diagonal = 1.5). Performance was weakly positive in V2 (mean d' = 0.45) and barely discernible in V1 (mean d' = 0.21).

Comparison of Figure 4B and C shows that the orthogonalized classifier performed worse for contrast pairs featuring high contrast naturalistic, and low contrast noise textures, but its performance was superior for matched contrasts, and for contrast pairs where the noise texture was higher in contrast than the naturalistic one. Critically, this shift in performance was possible using exclusively linear operations. Our results thus demonstrate that a biologically-plausible, contrast-invariant representation of naturalistic texture begins to emerge in V2 and is robustly clear in V4.

Citation	Areas recorded	c_{50} (RMS contrast)
Sclar et al. (1990)	V1	0.12
Albrecht and Hamilton (1982)*	V1	0.068
Zheng et al. (2007)	V1	0.066
This study – Noise	**V1**	**0.056**
This study – Naturalistic	**V1**	**0.058**
Zheng et al. (2007)	V2	0.071
This study – Noise	**V2**	**0.048**
This study – Naturalistic	**V2**	**0.039**
Williford and Maunsell (2006)	V4	0.053
Sani et al. (2013)†	V4	0.018
This study – Noise	**V4**	**0.034**
This study – Naturalistic	**V4**	**0.02**

Table 1: Comparison with previously reported average contrast sensitivity values. c_{50} values, as reported in previous studies. We converted Michelson contrast values to RMS contrast for comparison with our own results. For papers which reported both means and medians, we reported medians. The value reported by Albrecht and Hamilton (1982), marked with an asterisk, included neurons recorded from V1 of both macaques and cats. The study of Sani et al. (2013), marked with an obelisk, used a rectangular bar. Other studies used gratings or gabor patterns.

DISCUSSION

To understand how neural populations might build separate representations of stimulus dimensions, we measured how the responses of neurons and neural populations are modulated by changes in contrast and naturalistic structure. Contrast modulates activity in V1, V2, and V4 (Albrecht and Hamilton, 1982; Sani et al., 2013; Sclar et al., 1990; Williford and Maunsell, 2006; Zheng et al., 2007). Naturalistic structure modulates activity in V2 and V4 (Arcizet et al., 2008; Freeman et al., 2013; Lee et al., 2024; Okazawa et al., 2017; Ziemba et al., 2024, 2019), but not in V1. By varying both dimensions together, we were able to compare contrast sensitivities across areas, and to establish how the encoding of naturalistic structure in V2 and V4 is impacted by changes in contrast. We found that contrast sensitivities were higher in V2 than in V1, and higher still in V4. Contrast sensitivities in V2 and V4 were texture-specific: sensitivities were higher for naturalistic textures than for noise. At the population level, we found that while both V2 and V4 represent naturalistic structure, heterogeneous neural tuning for contrast and texture in V4 provided a basis for a robust contrast-invariant readout of naturalistic structure.

Contrast sensitivity for large texture images

Previous studies of contrast sensitivity in the ventral stream typically used stimuli, often gratings, which were adjusted to the preferred size, orientation, movement, and spatial frequency of the single neuron under study (Albrecht and Hamilton, 1982; Sani et al., 2013; Sclar et al., 1990; Williford and Maunsell, 2006; Zheng et al., 2007). Instead of optimized gratings, we used large texture images, each containing a wide variety of orientations and spatial frequencies. We also recorded multiunit activity – we did not sort out single units. As shown in Table 1, we found higher contrast sensitivities than prior work which used grating or Gabor pattern stimuli of a given spatial frequency. The combination of wide-band stimuli, plus our use of multiunit activity, may explain this difference: by averaging responses across multiple units, we may have increased the resulting signal-to-noise ratio. By using stimuli containing multiple spatial frequencies and orientations, any individual neuron within a multiunit recording would be more likely to respond. In particular, we found that the contrast sensitivity of sites in V4 was similar to that seen by Sani et al. (2013), who used sharp-edged bar stimuli (which contained energy at many spatial

frequencies). In general, while the relationship in sensitivity across areas was consistent, the relationship between stimulus type and sensitivity is not fully explained.

The importance and basis of contrast invariance

Responses at single sites vary in response to multiple dimensions of stimulus variation. It is therefore impossible to decode the cause of a response change without considering a neural population. For example, while responses in V4 are typically enhanced by naturalistic structure, this enhancement could be overcome for most sites by an opposing change in contrast (Figure 3C). Satisfyingly, after projecting out the axis corresponding to stimulus contrast, V4 populations could robustly discriminate low contrast naturalistic textures from high contrast noise, using a linear decoder (Figure 4C). This is possible because in V4, individual site responses varied widely in both contrast and texture selectivity (Figure 3B), making it possible to locate a subspace independent of contrast. Supersaturating contrast-response relationships, especially prevalent in V4, may be of particular benefit in establishing this contrast-invariant subspace.

In addition to texture, neurons in V4 are often tuned for the boundary curvature of a given stimulus (El-Shamayleh and Pasupathy, 2016; Kim et al., 2019; Nandy et al., 2013; Pasupathy and Connor, 1999). This curvature tuning is robust to changes in position (Pasupathy and Connor, 1999), scale (El-Shamayleh and Pasupathy, 2016), and color (Bushnell et al., 2011). The tuning of V4 neurons for combinations of shape and texture (Kim et al., 2019), or combinations of shape and blur (Oleskiw et al., 2018) is largely separable, suggesting that decoders constructed according to the principles we have employed here could be used to represent these and other stimulus dimensions of interest and relevance to behavior.

Invariance is of course not limited to V4: populations in inferotemporal cortex (IT) encode the identity of visual objects (Hung et al., 2005; Rust and DiCarlo, 2010), in a manner robust to changes in position, scale, and context. Nor is invariance limited to vision – measurements from somatosensory cortex have documented separation between touch-related variables (Lieber and Bensmaia, 2020). In motor cortex, the population representation of movement preparation lies in a dimension orthogonal to the representation of the movement itself (Kaufman et al., 2014). Finally, population activity in prefrontal cortex can flexibly encode task-relevant stimulus parameters, and disregard nuisance dimensions (Mante et al., 2013). The results reported here demonstrate how separable representations may manifest along the visual hierarchy, in a manner that may support the ability of downstream areas to flexibly read out information.

Acknowledgements

We are grateful to members of the Visual Neuroscience Laboratory for advice and discussion. R.T. Raghavan helped with model fitting, Justin Lieber provided useful insight about population analysis. This work was supported by grants from the National Institutes of Health – including R01 EY022428 (JAM), R01 EY024914 (LK, JAM), R01 EY031446 (NAM), and F31 EY031249 (GML).

Author contributions

L.K., J.A.M., and N.J.M. designed the experiment. G.M.L. and C.L.R.D. conducted the experiments. G.M.L. analyzed the data and did the modeling. G.M.L. and J.A.M. wrote the paper.

Declaration of interests

The authors declare no competing interests.

REFERENCES

Albrecht, D. G. and Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.*, 48(1):217–237.

Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.*, 58(3):626–639.

Arcizet, F., Jouffrais, C., and Girard, P. (2008). Natural textures classification in area V4 of the macaque monkey. *Exp. Brain Res.*, 189(1):109–120.

Bushnell, B. N., Harding, P. J., Kosai, Y., Bair, W., and Pasupathy, A. (2011). Equiluminance cells in visual cortical area V4. *J. Neurosci.*, 31(35):12398–12412.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.*, 11(8):333–341.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3):415–434.

El-Shamayleh, Y. and Pasupathy, A. (2016). Contour curvature as an invariant code for objects in visual area V4. *J. Neurosci.*, 36(20):5532–5543.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol.*, 187:517–552.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.*, 16(7):974–81.

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866.

Kaufman, M. T., Churchland, M. M., Ryu, S. I., and Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nat. Neurosci.*, 17(3):440–448.

Kim, T., Bair, W., and Pasupathy, A. (2019). Neural coding for shape and texture in macaque area V4. *J. Neurosci.*, 39(24):4760–4774.

Lee, G. M., Rodríguez Deliz, C. L., Bushnell, B. N., Majaj, N. J., Movshon, J. A., and Kiorpes, L. (2024). Developmentally stable representations of naturalistic image structure in macaque visual cortex. *Cell Rep.*, 43(8):114534.

Lieber, J. D. and Bensmaia, S. J. (2020). Emergence of an invariant representation of texture in primate somatosensory cortex. *Cereb. Cortex*, 30(5):3228–3239.

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84.

Naka, K. I. and Rushton, W. A. H. (1966). S-potentials from luminosity units in the retina of fish (*Cyprinidae*). *J. Physiol.*, 185(3):587–599.

Nandy, A. S., Sharpee, T. O., Reynolds, J. H., and Mitchell, J. F. (2013). The Fine Structure of Shape Tuning in Area V4. *Neuron*, 78(6):1102–1115.

National Research Council (2011). *Guide for the Care and Use of Laboratory Animals*. National Academies Press, Washington, D.C, 8th edition.

Okazawa, G., Tajima, S., and Komatsu, H. (2017). Gradual development of visual texture-selective properties between macaque areas V2 and V4. *Cereb. Cortex*, 27(10):4867–4880.

Oleskiw, T. D., Nowack, A., and Pasupathy, A. (2018). Joint coding of shape and blur in area V4. *Nat. Commun.*, 9(1).

Pasupathy, A. and Connor, C. E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.*, 82(5):2490–2502.

Peirce, J. W. (2007). The potential importance of saturating and supersaturating contrast response functions in visual cortex. *J. Vis.*, 7(6):13.

Portilla, J. and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.*, 40(1):49–71.

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(39):12978–95.

Saleem, K. S. and Logothetis, N. K. (2012). *A combined MRI and histology atlas of the rhesus monkey brain in stereotaxic coordinates*. Academic Press.

Sani, I., Santandrea, E., Golzar, A., Morrone, M. C., and Chelazzi, L. (2013). Selective tuning for contrast in macaque area V4. *J. Neurosci.*, 33(47):18583–18596.

Sclar, G., Maunsell, J. H. R., and Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vis. Res.*, 30(1):1–10.

Shapley, R. and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Prog. Retin. Res.*, 3:263–346.

Williford, T. and Maunsell, J. H. R. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol.*, 96(1):40–54.

Winters, W. D., Kado, R., and Adey, W. R. (1969). *A stereotaxic brain atlas for Macaca nemestrina*. University of California Press.

Zheng, J., Zhang, B., Bi, H., Maruko, I., Watanabe, I., Nakatsuka, C., Smith III, E. L., and Chino, Y. M. (2007). Development of temporal response properties and contrast sensitivity of V1 and V2 neurons in macaque monkeys. *J. Neurophysiol.*, 97(6):3905–3916.

Ziemba, C. M., Goris, R. L. T., Stine, G. M., Perez, R. K., Simoncelli, E. P., and Movshon, J. A. (2024). Neuronal and behavioral responses to naturalistic texture images in macaque monkeys. *J. Neurosci.*, 44(42).

Ziemba, C. M., Perez, R. K., Pai, J., Kelly, J. G., Hallum, L. E., Shooner, C., and Movshon, J. A. (2019). Laminar differences in responses to naturalistic texture in macaque V1 and V2. *J. Neurosci.*, 39(49):9748–9756.