Article

Current Biology

Selective representations of texture and motion in mouse higher visual areas

Highlights

- Texture stimuli are best represented in areas LM & LI, compared to areas AL & PM
- Motion stimuli are best represented in area AL compared to areas LM & PM
- Texture and motion components of naturalistic videos are similarly segregated
- Receptive field models show trade-off for representations of motion versus texture

Authors

Yiyi Yu, Jeffrey N. Stirman, Christopher R. Dorsett, Spencer L. Smith

Correspondence

sls@ucsb.edu

In brief

Mouse vision evolved for complex natural stimuli, not simple gratings. Using naturalistic stimuli, large-scale calcium imaging, and a data-driven approach, Yu et al. reveal how informative representations of motion and form are segregated to separate higher visual areas. They present modeling data for a potential mechanism of this segregation.

Yu et al., 2022, Current Biology *32*, 1–11
July 11, 2022 © 2022 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2022.04.091

CellPress

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

Current Biology

CellPress
OPEN ACCESS

Article

Selective representations of texture and motion in mouse higher visual areas

Yiyi Yu,[1] Jeffrey N. Stirman,[2] Christopher R. Dorsett,[2] and Spencer L. Smith[1,3,*]
[1]Department of Electrical & Computer Engineering, Center for BioEngineering, Neuroscience Research Institute, University of California, Santa Barbara, Santa Barbara, CA 93106, USA
[2]Neuroscience Research Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[3]Lead contact
*Correspondence: sls@ucsb.edu
https://doi.org/10.1016/j.cub.2022.04.091

SUMMARY

The mouse visual cortex contains interconnected higher visual areas, but their functional specializations are unclear. Here, we used a data-driven approach to examine the representations of complex visual stimuli by L2/3 neurons across mouse higher visual areas, measured using large-field-of-view two-photon calcium imaging. Using specialized stimuli, we found higher fidelity representations of texture in area LM, compared to area AL. Complementarily, we found higher fidelity representations of motion in area AL, compared to area LM. We also observed this segregation of information in response to naturalistic videos. Finally, we explored how receptive field models of visual cortical neurons could produce the segregated representations of texture and motion we observed. These selective representations could aid in behaviors such as visually guided navigation.

INTRODUCTION

Visual systems evolved to extract behaviorally relevant information from complex natural scenes. Visual stimuli contain information about texture, motion, objects, and other features of the environment around the animal. These components of visual stimuli have unequal relevance across behaviors. For example, optic flow (OF) and parallax motion information can help guide navigation behavior, but object recognition is often invariant to motion.

In mice, the axons from the neurons in the primary visual cortex (V1) extend out to an array of higher visual areas (HVAs), seven of which share a border with V1, and all of which have characteristic connectivity with other brain regions. Mouse visual cortical areas exhibit a level of hierarchical structure[1] and form two subnetworks anatomically[2,3] and developmentally.[4] HVAs receive functionally distinct afferents from the V1.[5,6] At least nine HVAs exhibit retinotopic topology,[7–10] and the neurons in HVAs have larger receptive fields (RFs) than those in the V1.[4,8,11] This organization and connectivity of mouse visual areas may have evolved to selectively propagate specific visual information to other brain regions.[5,6]

Gratings are classic visual stimuli for characterizing responses in visual cortical areas.[12–14] In mice, HVAs exhibit biases in their preferred spatial and temporal frequencies of gratings, but overall, their frequency passbands largely overlap,[5,15–17] and, thus, it can be challenging to determine how visual information is processed selectively in HVAs. Moreover, neural responses to more complex or naturalistic stimuli cannot be predicted from gratings. Even the neural responses to the superposition of two gratings, called plaid stimuli, are not well predicted by their component responses. Plaids can thus reveal cells that are selective to pattern motion, as opposed to the motion of the component gratings,[18] and plaids have been used to show that lateromedial (LM) and rostrolateral (RL) HVAs contain pattern-selective neurons.[19] Similarly, responses to gratings cannot predict invariant object recognition, which is exhibited by neurons in some rodent HVAs.[20–23] Thus, stimuli beyond simple gratings can help reveal functional specializations of HVAs.

Naturalistic visual stimuli contain complex multi-scale spatial features with statistical dependencies that are lacking in simple gratings. Mice can distinguish photographs of natural scenes,[24] which contain these features, but the features are often non-uniform and sparse. Parametric texture stimuli provide a more spatially uniform stimulus for determining how complex statistical features are represented by neuronal circuitry,[25] and the cortical area V2 in primates is specialized for processing texture stimuli.[26]

Motion is another component of visual stimuli, and it can be represented with some independence from the spatial components of naturalistic visual representations in the brain.[27] Texture and motion differentially contribute to neuronal activity in HVAs.[28] The representation of texture relies on the encoding of a combination of local spatial features,[26] whereas the representation of motion relies on computing integrated motion signals (e.g., opponent motion energy[29]). Thus, there is a potential computational rationale for expecting segregated representations of motion and texture.

In this study, we determined whether the L2/3 neurons in mouse HVAs exhibit biases in their representations of motion and texture. We used three classes of visual stimuli: drifting textures, random dot kinematograms (RDKs), and naturalistic

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

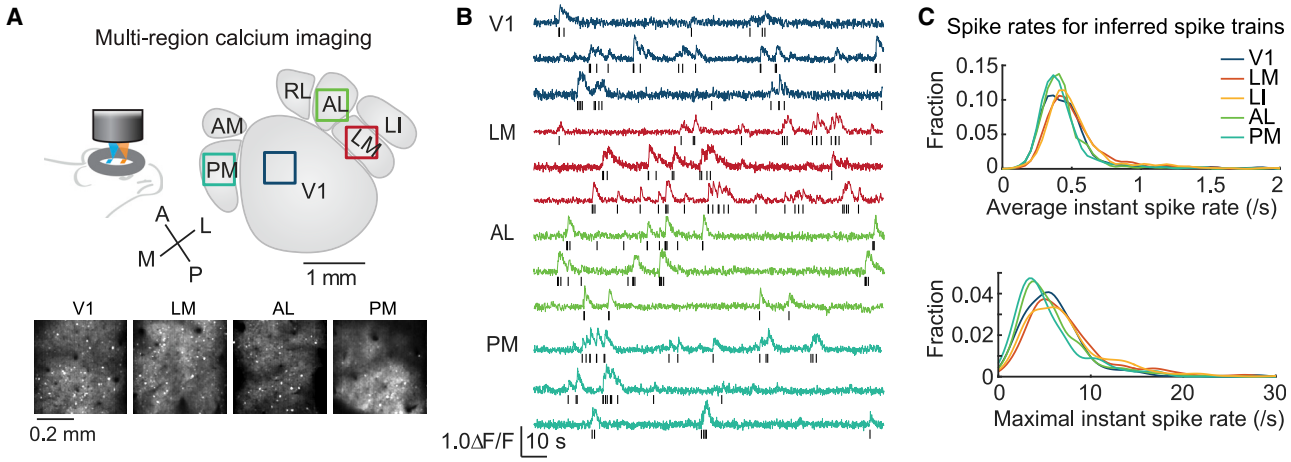

Figure 1. Multi-area calcium imaging of mouse visual cortex

(A) Neuronal activity was imaged in multiple HVAs simultaneously using large-field-of-view, temporally multiplexed, two-photon calcium imaging. In an example experiment, layer 2/3 excitatory neurons were imaged in the areas V1 and LM, AL, and PM simultaneously. The squares indicate the imaged regions, and the projections of raw image stacks are shown below.

(B) Image stacks were analyzed to extract calcium dynamics from cell bodies after neuropil subtraction. These traces were used to infer spike activity, as shown in raster plots below each trace.

(C) Statistics of the inferred spiking were similar to those of prior reports, indicating accurate inference. The mean and maximal instantaneous firing rates of the neurons in the V1 and HVAs are similar (mean, 0.5 ± 0.5 spike/s; max, 7 ± 11 spike/s; $p = 0.055$ and $p = 0.6$; one-way ANOVA).

videos. We examined how the texture/spatial and motion components of a naturalistic video are represented and found that high fidelity representations of these stimulus features are segregated to different HVAs. We then characterized the encoding properties of HVA neurons using a Gabor filter-based RF model. The results from these experiments reveal functional segregations of mouse HVAs for texture/spatial and motion components of visual stimuli.

RESULTS

Multi-area calcium imaging to distinguish the tuning properties of HVAs

To survey the tuning properties of multiple visual cortical areas, we performed population calcium imaging of the L2/3 neurons in the V1 and four HVAs (the LM, laterointermediate [LI], anterolateral [AL], and posteromedial [PM]) of awake mice using a multiplexing, large-field-of-view two-photon microscope with subcellular resolution developed in house[30] and transgenic mice expressing the genetically encoded calcium indicator GCaMP6s.[31,32] We located the V1 and HVAs of each mouse using retinotopic maps obtained by intrinsic signal optical imaging[4,16] (Figure S1A). The borders of HVAs were reliably delineated in most cases, with the exception being some experiments where the boundary between areas anteromedial (AM) and PM was not clearly defined. In those cases, to be conservative, we considered the population to be pooled between areas AM and PM but dominated by the latter and thus referred to as "PM" or "AM/PM." In cases where area AM neurons were positively identified, we did not observe a functional difference between putative AM and PM neurons. The large-field-of-view imaging system allows us to carry out flexible high-resolution recordings from up to four cortical visual areas simultaneously[11,30] (Figures 1A and 1B). In this study, most of the data were recorded in twin-region imaging mode, and two datasets were recorded in single-region imaging mode. Neuropil-corrected calcium signals were used to infer probable spike trains for each neuron (Figure S1B). The inferred spike train was accurate enough for computing spike-count-based statistical values.[11] During visual stimulation, the average and maximal firing rates inferred were similar across cortical areas and were typically around 0.5 spikes/s average and ranged up to 15–30 spikes/s maximal (Figure 1C), consistent with the previously reported values from electrophysiology.[33]

We characterized the neuronal responses to three types of visual stimuli: scrolling textures (hereafter referred to as "texture stimuli"), RDKs, and a naturalistic video mimicking home cage navigation. Hundreds of neurons were recorded for areas V1, LM, AL, and PM for all three stimuli, and the area LI was recorded for the texture stimuli (complete numbers are in Table S1). The neurons that exhibited reliable responses to a stimulus were included for further characterization (reliable neurons responded to >50% of the trials for RDKs or texture stimuli or exhibited trial-to-trial correlations >0.08 with 0.5 s bins to naturalistic stimuli). These criteria resulted in 20%–50% of all the recorded neurons being included for further analysis (Table S1).

Texture and RDK stimuli were best represented in separate HVAs

We tested the selectivity of neurons in the areas V1, LM, LI, AL, and PM to texture stimuli using a set of naturalistic textures that drifted in one of the four cardinal directions (Figure S2A). We generated four families of texture images based on parametric models of naturalistic texture patterns.[34] These stimuli allowed us to characterize the representation of both texture pattern information and drift direction information and thus test the tolerance of a texture-selective neuron to motion direction.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

CellPress
OPEN ACCESS

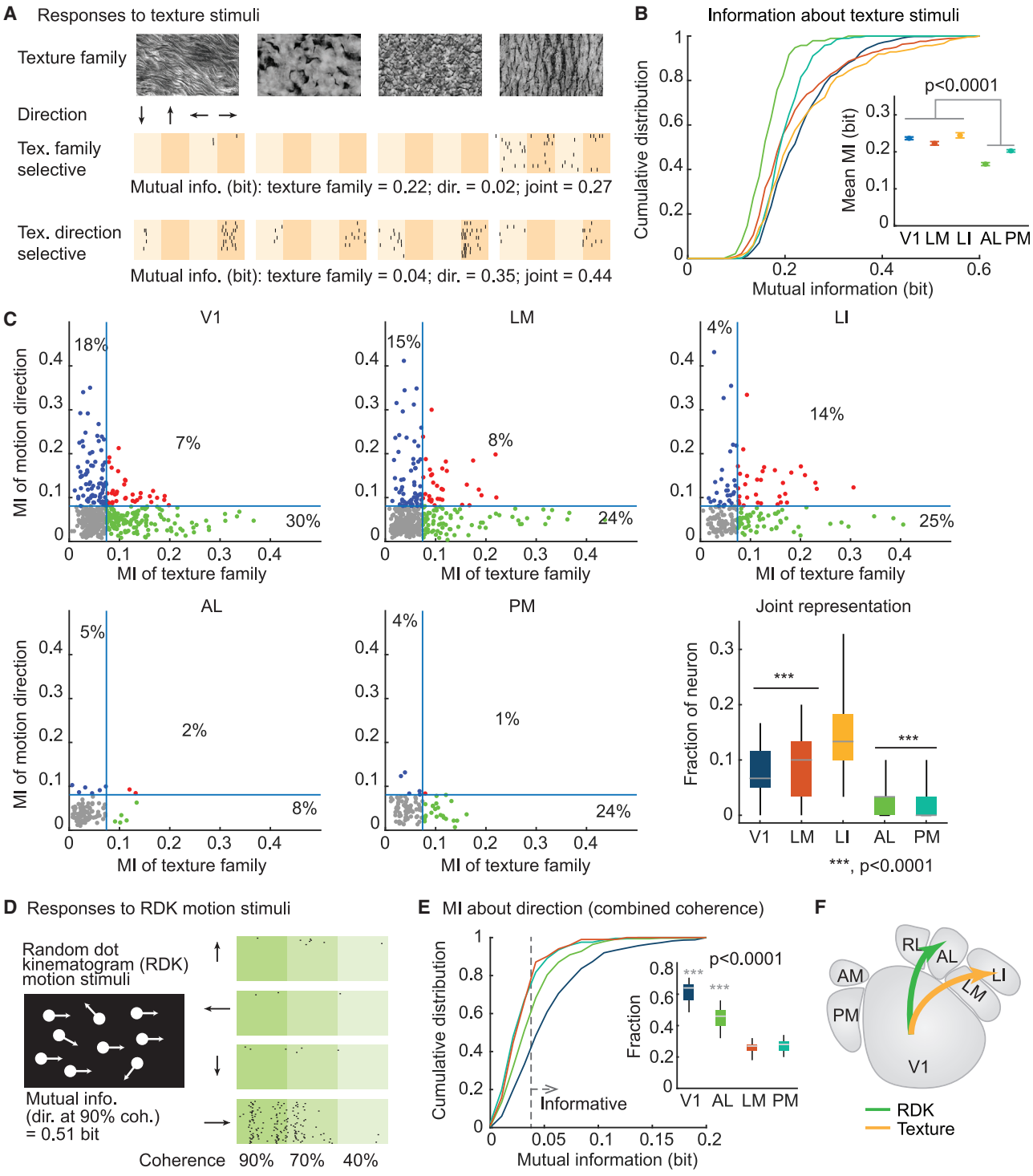

Figure 2. Segregated representations of textures and random dot kinematogram (RDK) motion in HVAs

(A) Mice were shown texture stimuli, each of which was from one of four texture families and drifted in one of four directions. The spike raster plots from two example neurons (10 trials shown for each) show that one neuron is selective for texture family and the other is more selective for texture direction. The amount of mutual information (MI, in bits) for the two stimulus parameters (texture family and motion direction) is written below each raster, along with the overall or joint (family and direction) MI.

(B) The areas V1, LM, and LI provide higher MI for texture stimuli than the area AL or PM ($p = 3.7 \times 10^{-17}$). The error bars in inset indicate SE.

(C) The relation between information about the moving direction and information about the texture family carried by the individual neuron per visual area. Each dot indicates one neuron. The blue line indicates the threshold of a significant amount of information, which was defined by shuffled data (mean + 3 * SD). Lower right: summary of the fraction of neurons exhibiting significant joint representation (red dots). Distribution generated by permutation.

(legend continued on next page)

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

We observed reliable responses to drifting textures in the areas V1, LM, LI, and PM. AL was less reliable in response to these stimuli (33%–41% of neurons in areas V1, LM, LI, and PM are reliable, whereas 20% of neurons in area AL are reliable; $p = 0.04$, one-way ANOVA; Figures S2A and S2B). The texture informative neurons exhibited various selectivity patterns (Figure 2A). We measured neuronal selectivity to texture family, motion direction, or joint selectivity using mutual information (MI) analysis. Higher bit values for a neuron-stimulus parameter pair mean that the activity from that neuron provides more information about that stimulus parameter (or a combination of stimulus parameters). Overall, the neurons in the areas V1 and LI were more informative about the texture stimuli, followed by neurons in area LM. By contrast, the neurons in the areas AL and PM were not informative about the texture stimuli (Figure 2B; $p = 3.7 \times 10^{-17}$; one-way ANOVA). To examine the tolerance of the texture encoding neurons to the translational direction, we computed the MI between neuronal responses and texture families (refer to the statistical pattern of a texture image). Area LI was the most informative about texture family out of all the tested visual areas, followed by the areas V1 and LM (Figure S2C; $p = 1.8 \times 10^{-11}$; one-way ANOVA). Meanwhile, areas V1, LM, and LI also carried more information about the motion direction of the texture stimuli, compared with the areas AL and PM (Figure S2D; $p = 1.2 \times 10^{-7}$, one-way ANOVA). Examining the information encoding of individual neurons, we found an increasing fraction of neurons that jointly encoded texture family and texture drift direction along the putative ventral pathway (V1: 7%, LM: 8%, LI: 14%, AL: 2%, and PM: 1%; $p = 9 \times 10^{-12}$; one-way ANOVA; Figure 2C), suggesting increasing joint coding along the putative ventral visual hierarchy (V1 → LM → LI).

To further examine joint coding along the putative ventral hierarchy, we fit an encoder model to the neurons that were significantly informative about the drift direction or the texture family (roughly half of the reliable neurons; Figure 2C). We decomposed the neuron model into the drifting direction component and the texture family component through singular value decomposition (SVD) decomposition, and we quantified the number of drifting directions and the number of the texture family one neuron was selective to in the SVD components (Figures S2E and S2F). Joint encoding neurons exhibited increasingly broader selectivity toward texture family and drifting directions along the putative ventral pathway (Figure S2G).

These results for texture encoding contrast with the results for standard drifting gratings. For gratings, we found motion direction information to be encoded broadly, differing <10% among HVAs (Figure S2H), whereas texture motion information did not propagate to visual areas outside the putative ventral pathway, differing >200% among HVAs (Figures S2C and S2D). The drift speeds were similar (32°/s for the textures and 40°/s for the gratings), so it is unclear which spatial structural differences between these stimuli drove the differences in encoding. Thus, we next examined the responses to a stimulus with less spatial structure and a greater focus on motion.

We examined the encoding of RDKs, which are salient white dots on a dark background with 40%–90% motion coherence (remaining dots move in random directions; Figure 2D). The RDK stimuli elicited reliable responses (responses to >50% trials) in neurons in areas V1, LM, AL, and PM. For this stimulus, reliable neurons were common in the V1 and rarer in area LM ($p = 0.001$, one-way ANOVA; Figures S3A and S3B). To characterize the direction selectivity of reliable neurons, we computed the MI between neuronal responses and the motion direction at each coherence level. Areas V1 and AL have larger fractions of informative neurons than areas LM and PM ($p = 2.3 \times 10^{-15}$, one-way ANOVA; Figure 2E). On average, areas V1 and AL were more informative than areas LM and PM about the motion direction of the RDKs at moderate-to-high coherence levels (≥70%; Figures S3C and S3D; $p = 3.3 \times 10^{-10}$, one-way ANOVA). At a low coherence level (40%), the differences among HVAs became insignificant (Figure S3D).

To further characterize the direction selectivity of neurons, we fit an encoder model to the neurons that were significantly informative about the RDK stimuli. We further decompose the neuron model into the direction component and the coherence component through SVD decomposition (Figures S3E and S3F; STAR Methods). We then identified the preferred direction and the preferred coherence level in these SVD components for individual neurons. Interestingly, the direction preference of the V1 was biased toward a downward motion or horizontal flows ($p = 8.6 \times 10^{-8}$, one-way ANOVA). However, in area AL, an RDK-representing HVA, such a bias was not obvious ($p = 0.09$, one-way ANOVA; Figure S3G). All the tested areas were modulated by coherence, and the highest coherence was preferred by the majority of neurons (Figure S3G).

In summary, texture-selective neurons were more abundant in areas V1, LM, and LI, whereas RDK direction selectivity neurons were more abundant in areas V1 and AL. Thus, information about drifting textures and RDK motion are relatively segregated to distinct HVAs (Figure 2F).

The features of naturalistic videos were best represented in separate HVAs

To determine whether this segregation of texture and motion information among HVAs could be detected within a more complex stimulus, we characterized the neuronal representation of a 64-s-long naturalistic video (Figure 3A). The naturalistic video stimulus contained time-varying visual features such as contrast,[35,36] luminance, edge density,[37] difference-of-Gaussian (DOG) entropy,[38] and OF speed and direction[39] (STAR Methods;

(D) Mice were shown random dot kinematogram (RDK) motion stimuli, which drifted in one of four directions with up to 90% coherence (fraction of dots moving in the same direction). A raster for an example neuron (30 trials) shows that it fires during rightward motion, with 0.51 bits of MI for motion direction at 90% coherence.

(E) Cumulative distribution plot of information about the motion direction combining all coherence levels. The gray dashed line indicates the threshold for a significant amount of information, which was defined by shuffled data (mean + 3 * SD). Inset: a boxplot of the fraction of informative neurons generated by permutation.

(F) These results indicate a segregation of visual stimulus representations: texture stimuli to area LM and RDK motion stimuli to area AL.

(B, C, and E) The p values were generated by one-way ANOVA with Bonferroni correction.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

CellPress
OPEN ACCESS

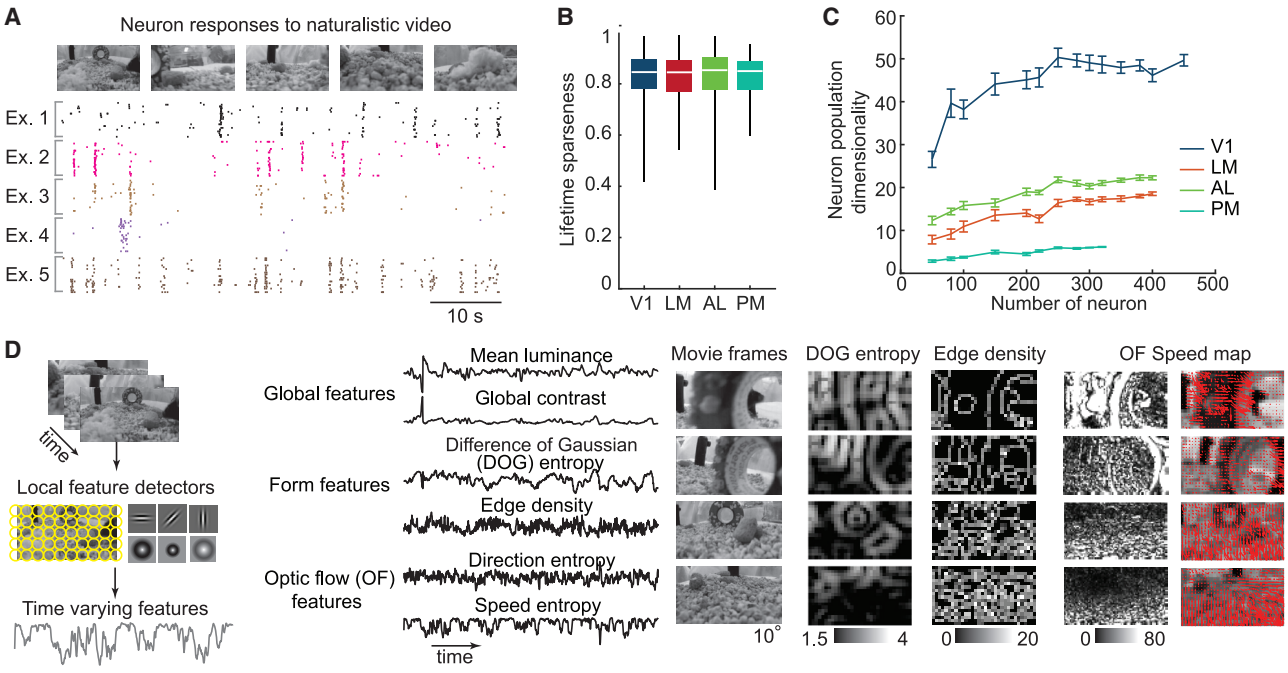

Figure 3. Parametric features of the naturalistic video stimulus

(A) Five example neurons show reliable, yet diverse, spike responses during a naturalistic video stimulus.

(B) The neurons in all four tested areas exhibited similarly high response sparseness to the naturalistic video ($p = 0.8$, one-way ANOVA).

(C) Dimensionality of neuron population responses versus the number of neurons per visual cortical area (error bars indicate SE of permutation).

(D) Form and motion components of the naturalistic video were extracted using a bank of linear filters with various sizes and locations (left). This provided time-varying signals correlated with global, form, and motion features, such as contrast, difference-of-Gaussian (DOG) entropy, and speed entropy (middle). To provide an intuitive feel for these features, example naturalistic video frames with the corresponding DOG entropy maps, edge density maps, and optical flow speed maps are shown.

See also Figures S4A–S4C.

Figure S4A–S4D). Both edge density and DOG entropy capture the spatial properties of a naturalistic video frame, whereas DOG entropy better supports texture family encoding (Figures S4E and S4F). Our results, thus far, suggested that the activity in area AL would be modulated by the motion information (OF speed or direction) in the naturalistic video and the activity in area LM would be modulated by the texture information (DOG entropy) in the same video.

About 45% of the V1 neurons and about 25% of the LM, AL, and PM neurons responded to the naturalistic video reliably (trial-to-trial correlation >0.08; Figure S5A). Neurons in the four imaged visual areas exhibited diverse and highly selective responses to the naturalistic video. Individual neurons responded to ~3% of the stimulus video frames, corresponding to a high lifetime sparseness (0.83 ± 0.09 [mean ± SD]; Figure 3B). The dimensionality of the population responses of the V1 was about three to ten times higher than those of the LM, AL, and PM areas (Figure 3C). The average response of a cortical-area neuron population converged with several hundreds of neurons (about 500 from the V1 and about 200 from HVAs; Figure S5B). Thus, the dataset has sufficiently covered the neuronal responses of areas V1, LM, AL, and PM to support further analysis. To take a data-driven approach to understand the diverse response patterns, we classified neurons using an unbiased clustering method (Gaussian mixture model [GMM]; STAR Methods). The neurons were partitioned into 25 tuning classes based on their responses to the naturalistic video, and 21 of these classes exhibited unique sparse response patterns, responding at specific time points of the naturalistic video (Figures S5C and S5D). All tuning classes were observed in the V1 and HVAs, but their relative abundance varied by area (Figures S5E and S5F). The V1 had a relatively more uniform distribution of tuning classes compared with HVAs (Figure S5E). The lower dimensionality and biased distribution of tuning classes of HVAs can indicate selective representations of visual features.

To examine how well the visual features (Figure S4D) of the naturalistic video were represented in HVAs, we computed the linear regression between the responses of individual neurons, or population-averaged activity, with the time-varying visual features of the naturalistic video (Figure 3D). We defined the modulation power of each feature as the variance of responses explained by the model (i.e., the r^2 of a linear fit) and the modulation coefficients as the slope. The visual features were computed at multiple spatial scales (image filtered by a Gaussian kernel with 1°–25° full width at half maximum [FWHM]), and qualitatively similar results were observed across a wide range of scales. Overall, the set of visual features explained 31% of the variance of the trial-averaged responses (with 0.5 s bins) of individual neurons across the visual areas. Representative results were obtained for edge density maps with a Gaussian kernel of 2.35° (FWHM) and DOG entropy maps with a Gaussian kernel of 11.75° (FWHM, inner kernel; the outer kernel is 2-fold larger in FWHM) (Figure 3D).

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

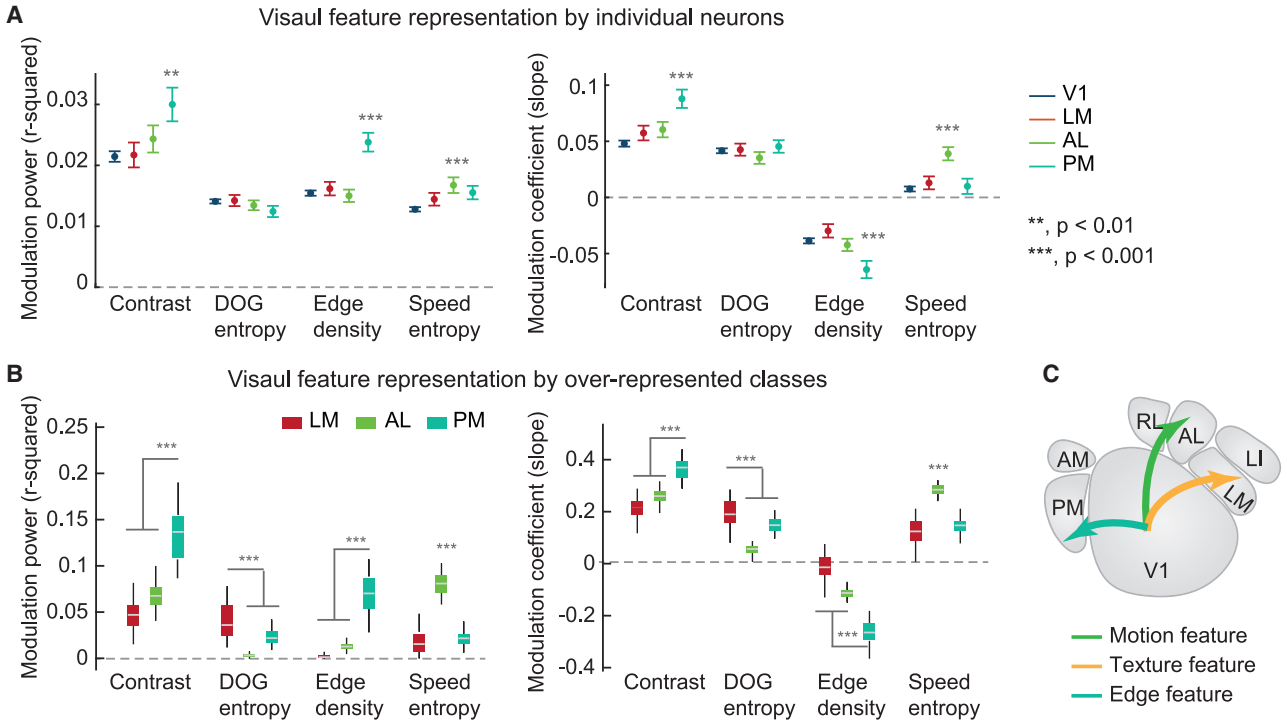

Figure 4. Segregated representations of spatial and motion features in naturalistic videos

(A) The average modulation power (left) and the average modulation coefficients (right) of individual neuronal activity by visual features of the naturalistic video. The modulation power is characterized by the r-squared values (variance explained) of the linear regression. Error bars indicate SE of permutation. p values are from one-way ANOVAs with the Bonferroni correction for multiple comparisons.

(B) The time-varying features were weighted to best match the average neuronal activity of enriched classes for a cortical area (N = 200 with permutation; classes with red stars in Figure S5F). The area LM was strongly modulated by DOG entropy. The area AL was strongly modulated by speed entropy. The area PM was modulated by contrast and edge density. The modulation coefficients were typically positive but were negative for edge density. Thus, the area PM is positively modulated by contrast but negatively modulated by edge density. (p values are from one-way ANOVAs with the Bonferroni correction for multiple comparisons.)

(C) These results indicate that visual feature components are segregated among HVAs: texture features to the area LM, motion features to the area AL, and fine edge features to the area PM.

Regressing these visual features with individual neuronal responses revealed selective feature representations. The AL neurons were more sensitive to the modulation by the OF speed entropy, whereas the PM neurons were more sensitive to the modulation by contrast and edge density (Figure 4A). We further characterized the collective effects of the biased distributions of tuning classes (from the GMM clustering analysis) among HVAs. Some tuning classes were more abundant in specific HVAs (relative to their abundance in the V1), and we assessed their activity modulation by the features in the naturalistic video. Area LM was enriched with neurons that correlated with DOG entropy (Figure 4B). As expected, areas AL and PM were enriched with tuning classes that correlate with the OF speed entropy feature and the contrast and edge density features, respectively (Figure 4B).

Together, these results indicate that motion information and spatial information are differentially represented among HVAs due to the distribution of tuning classes among them. AL neurons provided superior representations of the motion features in the naturalistic video, and LM and PM neurons provided superior representations of the texture and edge features, respectively, in the same naturalistic video (Figure 4C).

Gabor models exhibited biased feature representations

To this point, the evidence from the naturalistic stimuli indicates an enriched representation of spatial and motion features in areas LM, PM, and AL. We then questioned whether the feature enrichment in HVAs could be explained by RF structure. To model the RF structures of neurons, we employed a rich set of Gabor models, which are classic models for visual cortical neurons.[40,41] The Gabor models were simulated using a base model of a linear-nonlinear (LNL) cascade with Gabor-filter-based linear kernels (Figure 5A). The set of models included 2D Gabor models with single or multiple subunits combined linearly or quadratically (Figure 5B), as multiple Gabor kernels are required for predicting the V1 neuron responses in mice[41] and generating tolerances to rotation, translation, and scale.[26,42] The set of models also included 3D Gabor kernels with or without untune normalization units (Figure 5B), as normalization is critical for capturing the diverse response profiles of the V1 neurons to naturalistic stimuli.[43] For the 2D Gabor-filter-based models, we examined both linear and energy models. These are similar to the models of complex cells in which input from multiple simple cells with similar orientation preferences but varying phases are integrated.[12] Other combinations were used as well (cross

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

CellPress
OPEN ACCESS

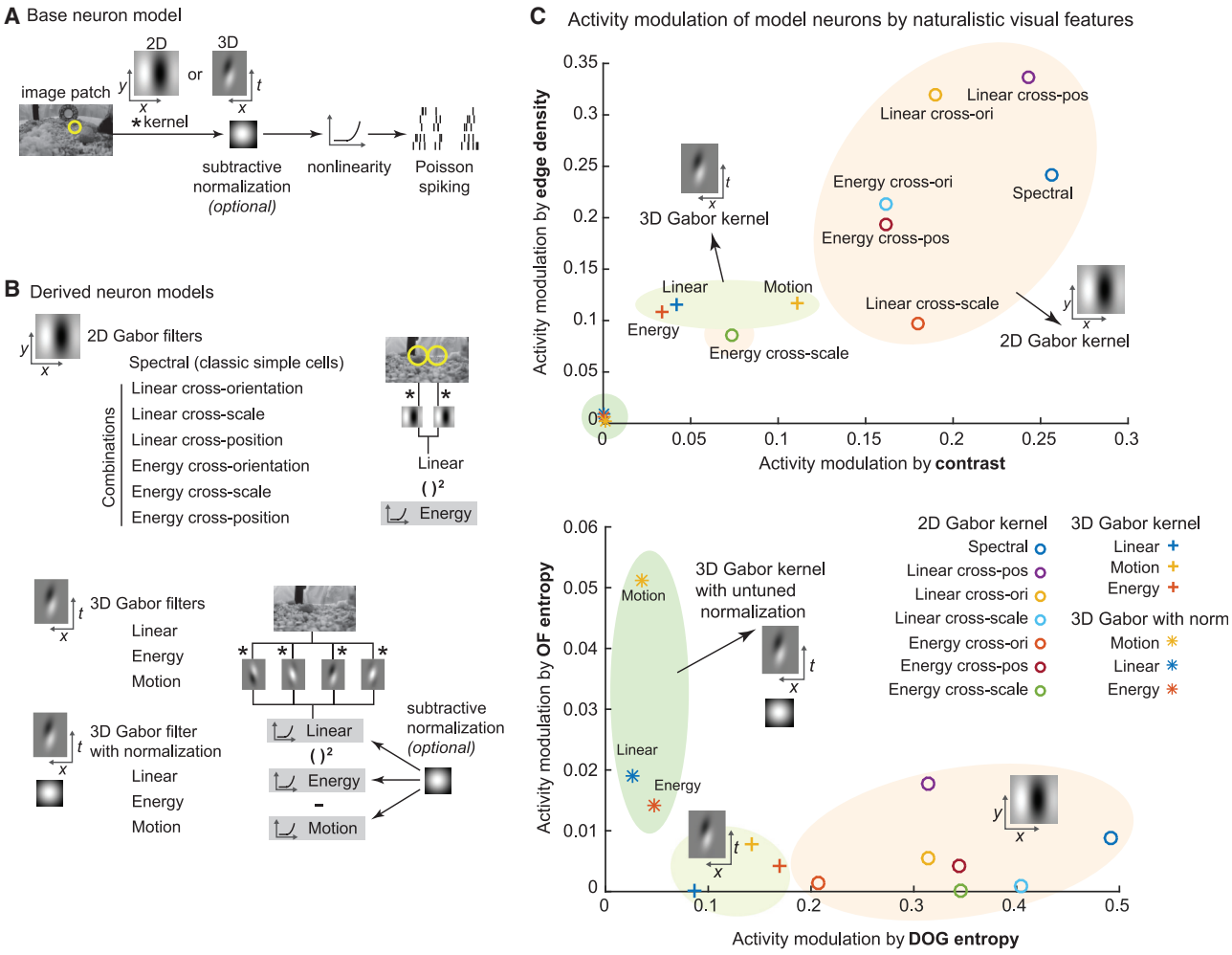

Figure 5. Spatial and motion feature encoding by variants of Gabor-filter-based models

(A) The general architecture is a linear-nonlinear-Poisson (LNP) cascade neuron model. Neurons were simulated by various 2D and 3D Gabor-like linear kernels, with or without an untuned subtractive normalization.

(B) From the base LNP model, variations were derived and organized into three classes: 2D Gabor-based, 3D Gabor-based without normalization, and 3D Gabor-based with normalization. Both linear and energy responses (akin to simple cells and complex cells) were computed from combinations of 2D Gabor filters. Linear, energy, and motion responses (akin to simple cells, complex cells, and speed cells) were computed from the 3D Gabor filters.

(C) These three classes of models varied in how much their activity was modulated by the global, form, and motion features in naturalistic videos. The neuron models are plotted by their modulation in feature spaces. The location of a neuron model was defined by the modulation power (same as Figure 4).

orientation, cross scale, etc.; Figure 5B). For the 3D Gabor-filter-based models, we also examined motion models (opponent motion energy[29]; Figure 5B). The simulations were performed with multiple spatial scales and, for the 3D Gabor-filter-based models, temporal scales and sampled uniformly in space.

We determined the feature extraction properties of the models by simulating responses to the texture and RDK stimuli and characterizing them (Figure S6). Neuron models varied in the encoding power of different types of stimuli or visual features. We noted that the 2D Gabor models exhibited specific tuning to the texture family while remaining tolerant to motion directions, especially the cross-orientation and linear cross-position models (Figure S6A), which are the best models for texture family encoding. On the other hand, the 3D Gabor models with normalization performed the best in encoding the motion direction in RDK stimuli (Figure S6B).

In response to the naturalistic videos, the 2D Gabor models, especially linear cross-position and linear cross-orientation models, exhibited better sensitivity to the contrast, edge density, and the DOG entropy, whereas the 3D Gabor models with untuned normalization exhibited better sensitivity to the OF entropy (Figure 5C). These simulation results confirmed the apparent trade-off in representation fidelity for spatial features and motion features, with the 2D Gabor kernels performing better on the former and the 3D Gabor kernels with or without normalization on the latter.

Gabor models reproduced specific feature representations of mouse visual cortex

Next, we determined the best Gabor model for the neuron responses recorded *in vivo* to the naturalistic video stimuli. We fit individual neuronal responses with the Gabor-based models.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

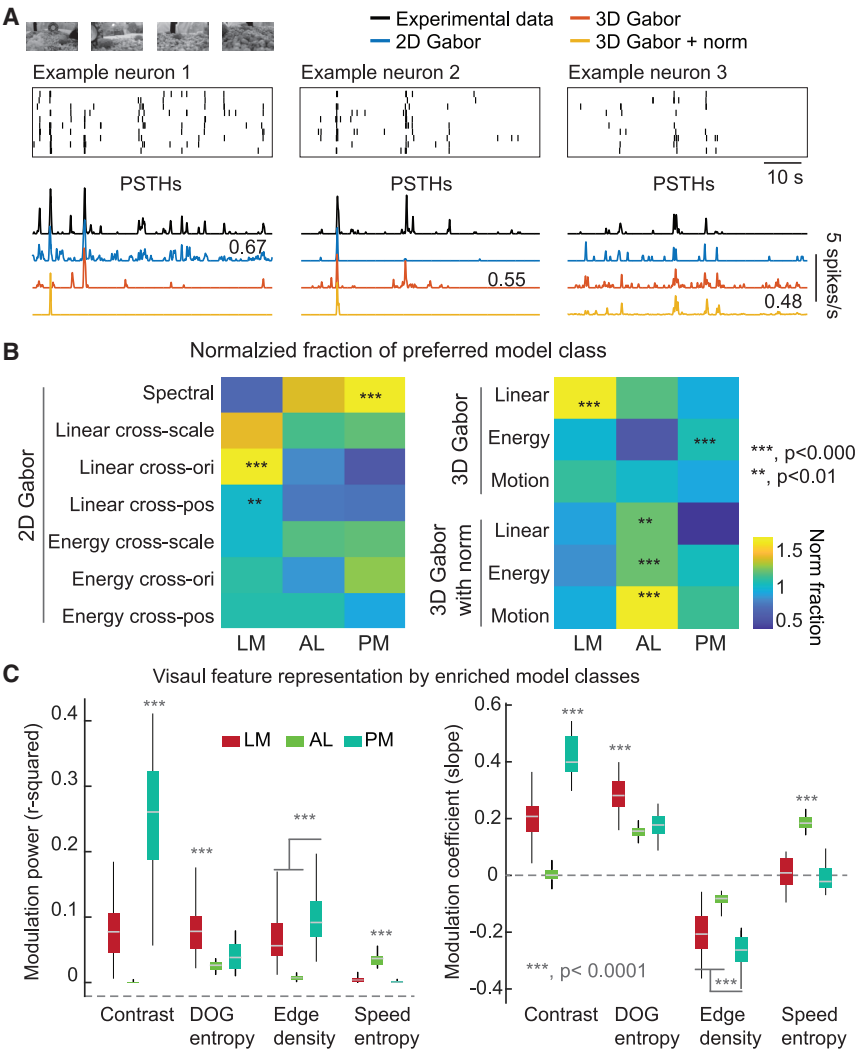

Figure 6. HVA-specific enrichment of Gabor model types

(A) Data from example neurons are shown in raster plots (top) and PSTHs (bottom) along with the best fits (as PSTHs) from each of the three model classes: 2D Gabor, 3D Gabor, and 3D Gabor with normalization. The numbers near the model PSTHs indicate the variance explained by the best models for that example neuron. One best model was selected for each neuron for the analysis in the following panels.

(B) For each Gabor model type and HVA, we computed how often that model was preferred for that HVA. Then, for each HVA, we normalized the model type fractions by the frequency for model types in the V1.

(C) The modulation power (left) and the modulation coefficients (right) of the average responses of enriched model classes for HVAs (N = 50 with permutation; classes with stars in B).

(B and C) The p values from one-way ANOVAs to test the statistical difference among the three HVA areas.

One best model among the whole set, the "preferred" model, was selected for individual neurons. The preferred models captured the greatest amount of peristimulus time histograms (PSTHs) variance of the *in vivo* neuronal responses (Figure 6A). In general, the preferred models explained 40% ± 17% (mean ± SD) of the variance of the PSTH (no difference between visual areas; p = 0.06, one-way ANOVA).

The V1 and HVAs varied in model preference. All classes were present in the V1, but HVAs varied in the classes that were more or less frequently observed (Figure 6B). We found that area LM was enriched with texture feature encoding models, such as the 2D models with linear cross-orientation subunits and linear cross-position subunits (Figure 6B). Area AL was enriched with motion encoding models, such as the 3D motion models with normalization (Figure 6B). By contrast, area PM had an abundance of single-unit 2D models (spectral filters), which were sensitive to the contrast visual features (Figure 6B).

Using the preferred models that were enriched in each HVA, we reproduced the linear regression results between the features of the naturalistic video (Figure 3D) and the HVA-specific model responses (Figure 6C). In general, the model selection reflected the segregated representation of spatial and motion features by areas LM, AL, and PM. Overall, the Gabor model analysis revealed that areas LM, AL, and PM were enriched with different RF structures. Area LM had more 2D linear cross-orientation and linear cross-position RFs, as well as 3D simple cells. Area AL had more 3D simple, complex, and motion cells and received surround normalization input. Area PM had more 2D simple cells and 3D complex cells. These insights into the RF structure of the neurons in HVAs provide clues about potential neural circuitry underlying the varying representations of features in naturalistic stimuli.

DISCUSSION

The results revealed the unique encoding properties of the L2/3 neurons of the V1 and multiple HVAs in representing textures, motion, and naturalistic videos. Our results show that area LM neurons provide high fidelity representations of spatial features such as DOG entropy and textures but are poor at representing motion. By contrast, area AL neurons provide high fidelity representations of motion features but are poor at representing spatial features. Area PM neurons provide high fidelity representations of some spatial features such as edge density and contrast but not DOG entropy or textures like area LM neurons do. Relatedly, the areas LM, LI, and AL are all poor at representing edge density. These findings show that visual features are represented in segregated neuronal populations, implying trade-offs in encoding. To investigate potential trade-offs, we examined the RF structures of the neurons in the HVAs that contributed to specialized feature representations. Indeed, we found that

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

different RF models were required for reproducing the *in vivo* results in separate HVAs. These findings provide new insights into the neural circuitry that can generate distributed representations of visual stimuli in HVAs.

The rodent visual system evolved in response to the ecological niche mice found themselves in. We do not expect such a process to result in a neural circuitry that performs neat, absolute segregations of information about visual scenes. Instead, we expect a neural circuitry that supports adaptive behavior for the mouse's ecological niche, such as a visual mechanism for predator avoidance.[44,45] The principles of that circuitry are likely quite different from those of any systematic, mathematically compact approach for parsing a visual scene in terms of the known RF properties of visual cortical neurons. Thus, here, we used a data-driven approach to gain a conservative foothold into the complex visual scene processing in mice. We explored how segregated representations might emerge using a modeling approach based on the known RF properties of visual cortical neurons or at least popular models thereof. This analysis showed that a linear combination of edges and motion energy model with normalization provided accurate accounts for distinguishing texture and motion stimuli.

The enrichment of specific representations of motion or texture in the L2/3 of the AL and LM areas, respectively, could arise from specific connectivity from other brain regions (e.g., the V1) that preserve the selectivity that arose there[5] or from converging input that results in enhanced selectivity (i.e., more invariant selectivity) for a visual feature.[12,41,46] We generally cannot distinguish those two possibilities with this dataset. Dense feedback and feedforward connectivity between HVAs and the V1 make it difficult to pinpoint where selectivity arises. However, area LI neurons exhibited more complex selectivity than neurons in area V1, so it appears as though preserved selectivity from V1 projecting to area LI would be insufficient to produce such selectivity. Even then, we cannot rule out thresholding effects, which could play a role in increasing the apparent selectivity.

The dual-stream framework describes two subnetworks of visual circuitry, a ventral one for object recognition and a dorsal one for motion and action.[27,47] In primates, the ventral stream—from the V1, through the V2, V3, and V4, to the inferior temporal lobe— develops selective activity for specific recognized objects, including faces. The dorsal stream—from the V1 through the V2, middle temporal area (MT), medial superior temporal area (MST), and the parietal lobe—processes the spatial and motion information in visual scenes. Anatomically, cats and ferrets have similar visual hierarchies as primates. Cat areas 17, 18, and 19 are analogous to the V1, V2, and V3 of primates, and area 21a and posterior medial lateral suprasylvian sulcus (PMLS) are analogous to the primates' V4 and MT, respectively.[48–53] Robust representations of oriented gratings or edges are commonly found in the V1, but how do representations change along the dorsal and ventral streams? The neurons in the areas V2 and V4 can encode a combination of local features, such as multiple edges to detect curves and shapes.[37,42] The neurons in MT and MST areas can encode opponent motion energy.[29,54–56] However, conflicts and complexities have been identified for these apparently stream-segregated feature representations in primates.[57,58] Cats and ferrets may have a cortical circuitry that is functionally analogous to the V4 and MT in primates;[48,53,59] nevertheless, our knowledge of intermediate-level feature representations in these animals is limited. Anatomical connectivity with downstream brain regions supports functional distinctions in mouse HVAs between putative ventral and dorsal streams. For example, the putative ventral areas are strongly connected to temporal and parahippocampal cortices, whereas the putative dorsal areas are preferentially connected to the parietal, motor, and limbic areas.[3] However, there are also major differences. The cortical visual system of mice is distinct from those of primates, cats, and ferrets, which have a single V2 adjacent to the V1.[60] Instead, mice have more than seven HVAs that share a border with the V1.[8,60] These anatomical differences complicate the search for the dual-stream homologies in mice and motivate functional studies to elucidate the information processing in mouse HVAs.[27] The functional similarities between the mouse LM and LI areas and macaque V2 and V4 and between the mouse AL area and macaque MT are perhaps superficial but could also indicate that the dual-stream framework for visual pathways in primates could have an analog in mice.[3,61] Anatomical and RF mapping studies suggest that the mouse LM and AL areas likely serve as the ventral and dorsal gateways in the mouse visual hierarchy.[4,8,62]

Our experimental results reveal the segregations of visual encoding, or representations, among HVAs in mice. The modeling results demonstrate how classic models of visual cortex neuron RFs can entail reciprocal trade-offs for representing spatial features and motion features of visual stimuli, thus providing a potential functional substrate for this segregation. These insights show how mouse HVAs likely play distinct roles in visual behaviors and may share similarities with the dual processing streams in primates.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subjects
- METHOD DETAILS
  - Surgery
  - Intrinsic signal optical imaging (ISOI)
  - *In vivo* two-photon imaging
  - Visual stimuli
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Calcium imaging processing
  - Reproducibility and lifetime sparseness
  - Information analysis
  - Informative neurons
  - Tuning pattern of informative neurons
  - Population response dimensionality
  - Gaussian mixture model
  - Visual features of the naturalistic video

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

- Modulation power of naturalistic features
- Gabor-based receptive field models
- Fit neuron responses with Gabor-based models
- Statistical analysis and data availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.04.091.

ACKNOWLEDGMENTS

The funding was provided by grants from the NIH (R01EY024294, RF1NS121919, and R01NS091335), the NSF (1707287 and 1450824), the Simons Foundation (SCGB325407), and the McKnight Foundation to S.L.S.; a Helen Lyng White Fellowship to Y.Y.; a career award from Burroughs Welcome to J.N.S.; and a training grant support to C.R.D. (T32NS007431).

AUTHOR CONTRIBUTIONS

All the experiments and analyses were performed by Y.Y. The imaging system was built by J.N.S. Animal handling was assisted by C.R.D. Study design and supervision by S.L.S. Y.Y. and S.L.S. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 13, 2021
Revised: March 22, 2022
Accepted: April 28, 2022
Published: May 23, 2022

REFERENCES

1. Siegle, J.H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T.K., Choi, H., Luviano, J.A., et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. Nature *592*, 86–92.
2. Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. J. Neurosci. *35*, 1905–1918.
3. Wang, Q., Sporns, O., and Burkhalter, A. (2012). Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. J. Neurosci. *32*, 4386–4399.
4. Smith, I.T., Townsend, L.B., Huh, R., Zhu, H., and Smith, S.L. (2017). Stream-dependent development of higher visual cortical areas. Nat. Neurosci. *20*, 200–208.
5. Glickfeld, L.L., Andermann, M.L., Bonin, V., and Reid, R.C. (2013). Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat. Neurosci. *16*, 219–226.
6. Kim, M., Znamenskiy, P., Iacaruso, M.F., and Mrsic-Flogel, T.D. (2018). Segregated subnetworks of intracortical projection neurons in primary visual cortex highlights. Neuron, 1–9.
7. Garrett, M.E., Nauhaus, I., Marshel, J.H., and Callaway, E.M. (2014). Topography and areal organization of mouse visual cortex. J. Neurosci. *34*, 12587–12600.
8. Wang, Q., and Burkhalter, A. (2007). Area map of mouse visual cortex. J. Comp. Neurol. *502*, 339–357.
9. Kalatsky, V.A., and Stryker, M.P. (2003). New paradigm for optical imaging: temporally encoded maps of intrinsic signal. Neuron *38*, 529–545.
10. Schuett, S., Bonhoeffer, T., and Hübener, M. (2002). Mapping retinotopic structure in mouse visual cortex with optical imaging. J. Neurosci. *22*, 6549–6559.
11. Yu, Y., Stirman, J.N., Dorsett, C.R., and Smith, S.L. (2018). Mesoscale correlation structure with single cell resolution during visual coding. Preprint at bioRxiv. https://doi.org/10.1101/469114.
12. Hubel, D.H., and Wiesel, T. (1962). AND FUNCTIONAL ARCHITECTURE IN THE CAT S VISUAL CORTEX From the Neurophysiolojy Laboratory, Department of Pharmacology central nervous system is the great diversity of its cell types and inter- receptive fields of a more complex type (Part I) and to. J. Physiol. *160*, 106–154.
13. Jin, M., and Glickfeld, L.L. (2020). Mouse higher visual areas provide both distributed and specialized contributions to visually guided behaviors. Curr. Biol. *30*, 4682–4692.e7.
14. Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S.E., and Pachitariu, M. (2021). High-precision coding in visual cortex. Cell *184*, 2767–2778.e15.
15. Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L., and Reid, R.C. (2011). Functional specialization of mouse higher visual cortical areas. Neuron *72*, 1025–1039.
16. Marshel, J.H., Garrett, M.E., Nauhaus, I., and Callaway, E.M. (2011). Functional specialization of seven mouse visual cortical areas. Neuron *72*, 1040–1054.
17. Goltstein, P.M., Reinert, S., Bonhoeffer, T., and Hübener, M. (2021). Mouse visual cortex areas represent perceptual and semantic features of learned visual categories. Nat. Neurosci. *24*, 1441–1451.
18. Tinsley, C.J., Webb, B.S., Barraclough, N.E., Vincent, C.J., Parker, A., and Derrington, A.M. (2003). The nature of V1 neural responses to 2D moving patterns depends on receptive-field structure in the marmoset monkey. J. Neurophysiol. *90*, 930–937.
19. Juavinett, A.L.L., and Callaway, E.M.M. (2015). Pattern and component motion responses in mouse visual cortical areas. Curr. Biol. *25*, 1759–1764.
20. Zoccolan, D. (2015). Invariant visual object recognition and shape processing in rats. Behav. Brain Res. *285*, 10–33.
21. Tafazoli, S., Safaai, H., De Franceschi, G., Rosselli, F.B., Vanzella, W., Riggi, M., Buffolo, F., Panzeri, S., and Zoccolan, D. (2017). Emergence of transformation-tolerant representations of visual objects in rat lateral extrastriate cortex. eLife *6*, 1–39.
22. Chang, L., and Tsao, D.Y. (2017). The code for facial identity in the primate brain. Cell *169*, 1013–1028.e14.
23. Froudarakis, E., Cohen, U., Diamantaki, M., Walker, E.Y., Reimer, J., Berens, P., Sompolinsky, H., and Tolias, A.S. (2020). Object manifold geometry across the mouse cortical visual hierarchy. Preprint at bioRxiv. https://doi.org/10.1101/2020.08.20.258798.
24. Yu, Y., Hira, R., Stirman, J.N., Yu, W., Smith, I.T., and Smith, S.L. (2018). Mice use robust and common strategies to discriminate natural scenes. Sci. Rep. *8*, 1379.
25. Victor, J.D., Conte, M.M., and Chubb, C.F. (2017). Textures as probes of visual processing. Annu. Rev. Vis. Sci. *3*, 275–296.
26. 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*, 974–981.
27. Cadieu, C.F., and Olshausen, B.A. (2012). Learning intermediate-level representations of form and motion from natural movies. Neural Comput. *24*, 827–866.
28. Zeki, S., and Shipp, S. (1988). The functional logic of cortical connections. Nature *335*, 311–317.
29. Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.
30. Stirman, J.N., Smith, I.T., Kudenov, M.W., and S.L.S.. (2016). Wide field-of-view, multi-region two-photon imaging of neuronal activity. Nat. Biotechnol. *34*, 857–862.
31. Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C., et al. (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron *85*, 942–958.
32. Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

33. Niell, C.M., and Stryker, M.P. (2008). Highly selective receptive fields in mouse visual cortex. J. Neurosci. *28*, 7520–7536.
34. Portilla, J., and Simoncelli, E.P. (2000). Parametric texture model based on joint statistics of complex wavelet coefficients. Int. J. Comput. Vis. *40*, 49–70.
35. Prusky, G.T., and Douglas, R.M. (2004). Characterization of mouse cortical spatial vision. Vision Res. *44*, 3411–3418.
36. Goldbach, H.C., Akitake, B., Leedy, C.E., and Histed, M.H. (2021). Performance in even a simple perceptual task depends on mouse secondary visual areas. eLife *10*, 1–39.
37. Ziemba, C.M., Freeman, J., Movshon, J.A., and Simoncelli, E.P. (2016). Selectivity and tolerance for visual texture in macaque V2. Proc. Natl. Acad. Sci. USA *113*, E3140–E3149.
38. Duan, H., and Wang, X. (2015). Visual attention model based on statistical properties of neuron responses. Sci. Rep. *5*, 8873.
39. Sit, K.K., and Goard, M.J. (2020). Distributed and retinotopically asymmetric processing of coherent motion in mouse visual cortex. Nat. Commun. *11*, 3565.
40. Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. J. Neurosci. *17*, 8621–8644.
41. Yoshida, T., and Ohki, K. (2020). Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. Nat. Commun. *11*, 872.
42. Okazawa, G., Tajima, S., and Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. Proc. Natl. Acad. Sci. USA *112*, E351–E360.
43. Cowley, B.R., Smith, M.A., Kohn, A., and Yu, B.M. (2016). Stimulus-driven population activity patterns in macaque primary visual cortex. PLoS Comput. Biol. *12*, e1005185.
44. Yilmaz, M., and Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. Curr. Biol. *23*, 2011–2015.
45. Shang, C., Chen, Z., Liu, A., Li, Y., Zhang, J., Qu, B., Yan, F., Zhang, Y., Liu, W., Liu, Z., et al. (2018). Divergent midbrain circuits orchestrate escape and freezing responses to looming stimuli in mice. Nat. Commun. *9*, 1232.
46. de Vries, S.E.J., Lecoq, J.A., Buice, M.A., Groblewski, P.A., Ocker, G.K., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., et al. (2020). A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nat. Neurosci. *23*, 138–151.
47. Mishkin, M., and Ungerleider, L.G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. *6*, 57–77.
48. Dreher, B., Wang, C., Turlejski, K.J., Djavadian, R.L., and Burke, W. (1996). Areas PMLS and 21a of cat visual cortex: two functionally distinct areas. Cereb. Cortex *6*, 585–599.
49. Manger, P.R., Nakamura, H., Valentiniene, S., and Innocenti, G.M. (2004). Visual areas in the lateral temporal cortex of the ferret (Mustela putorius). Cereb. Cortex *14*, 676–689.
50. Connolly, J.D., Hashemi-Nezhad, M., and Lyon, D.C. (2012). Parallel feedback pathways in visual cortex of cats revealed through a modified rabies virus. J. Comp. Neurol. *520*, 988–1004.
51. Tusa, R.J., and Palmer, L.A. (1980). Retinotopic organization of areas 20 and 21 in the cat. J. Comp. Neurol. *193*, 147–164.
52. Pan, H., Zhang, S., Pan, D., Ye, Z., Yu, H., Ding, J., Wang, Q., Sun, Q., and Hua, T. (2021). Characterization of feedback neurons in the high-level visual cortical areas that project directly to the primary visual cortex in the cat. Front. Neuroanat. *14*, 1–14.
53. Lempel, A.A., and Nielsen, K.J. (2019). Ferrets as a model for higher-level visual motion processing. Curr. Biol. *29*, 179–191.e5.
54. Khawaja, F.A., Liu, L.D., and Pack, C.C. (2013). Responses of MST neurons to plaid stimuli. J. Neurophysiol. *110*, 63–74.
55. Adelson, E.H., and Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. Nature *300*, 523–525.
56. Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.
57. Ferrera, V.P., and Maunsell, J.H.R. (2005). Motion processing in macaque V4. Nat. Neurosci. *8*, 1125. 1125.
58. Konen, C.S., and Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. Nat. Neurosci. *11*, 224–231.
59. Dreher, B., Michalski, A., Ho, R.H.T., Lee, C.W.F., and Burke, W. (1993). Processing of form and motion in area 21a of cat visual cortex. Vis. Neurosci. *10*, 93–115.
60. Rosa, M.G.P., and Krubitzer, L.A. (1999). The evolution of visual cortex: where is V2? Trends Neurosci *22*, 242–248.
61. Saleem, A.B. (2020). Two stream hypothesis of visual processing for navigation in mouse. Curr. Opin. Neurobiol. *64*, 70–78.
62. Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. J. Neurosci. *31*, 1905–1918.
63. Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L.F., Carandini, M., and Harris, K.D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at bioRxiv. https://doi.org/10.1101/061507.
64. Pnevmatikakis, E.A., Merel, J., Pakman, A., and Paninski, L. (2013). Bayesian spike inference from calcium imaging data. Adv. Neural Inf. Process. Syst. *26*, 1250–1258.
65. Smith, S.L., and Häusser, M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. Nat. Neurosci. *13*, 1144–1149.
66. Stirman, J.N., Townsend, L.B., and Smith, S.L. (2016). A touchscreen based global motion perception task for mice. Vis. Res. *127*, 74–83.
67. Harris, K.D., Quiroga, R.Q., Freeman, J., and Smith, S.L. (2016). Improving data quality in neuronal population recordings. Nat. Neurosci. *19*, 1165–1174.
68. Yu, Y., Burton, S.D., Tripathy, S.J., and Urban, N.N. (2015). Postnatal development attunes olfactory bulb mitral cells to highfrequency signaling. J. Neurophysiol. *114*, 2830–2842.
69. Froudarakis, E., Berens, P., Ecker, A.S., Cotton, R.J., Sinz, F.H., Yatsenko, D., Saggau, P., Bethge, M., and Tolias, A.S. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. Nat. Neurosci. *17*, 851–857.
70. Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. Nature *529*, 345–350.
71. Canny, J. (1986). A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. *8*, 679–698.
72. Hermundstad, A.M., Briguglio, J.J., Conte, M.M., Victor, J.D., Balasubramanian, V., and Tkačik, G. (2014). Variance predicts salience in central sensory processing. eLife *3*, 1–28.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: Emx1-cre; Ai94D (TITL-GCamp6s); ROSA-tTA	The Jackson Laboratory	Strain: jax #005628, Jax #024104, Jax # 011008
Software and algorithms		
MATLAB R2015b, 2019b	Mathworks	https://www.mathworks.com/products/matlab.html
suite2p calcium imaging processing toolbox	Pachitariu et al.[63]	https://github.com/cortex-lab/Suite2P
MCMC spike deconvolution toolbox	Pnevmatikakis et al.[64]	https://github.com/flatironinstitute/CaImAn-MATLAB/tree/master/deconvolution
Gabor modeling and naturalistic video processing	This paper	https://github.com/yuyiyi/HVA_tuning.git

RESOURCE AVAILABILITY

Lead contact

Requests for additional information or resources related to the study should be directed to Spencer L. Smith (sls@ucsb.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Deconvolved spike data, and key functions for Gabor modeling and naturalistic video processing are available in: https://github.com/yuyiyi/HVA_tuning.git.
- Code for calcium imaging processing is available in: https://github.com/yuyiyi/CaSoma_Proc.git.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

All animal procedures and experiments were approved by the Institutional Animal Care and Use Committee of the University of North Carolina at Chapel Hill or the University of California Santa Barbara and performed in accordance with the regulation of the US Department of Health and Human Services. Total of 34 GCaMP6s-expressing transgenic adult mice of both sexes were used in this study. Mice were 110 – 300 days old for data collection. GCaMP6s-expressing mice were induced by triple crossing of the following mouse lines: TITL-GCaMP6s (Allen Institute Ai94), Emx1-Cre (Jackson Labs #005628), and ROSA:LNL:tTA (Jackson Labs #011008).[31] Mice were housed under a 12 h light / 12 h dark cycle, and experiments were performed during the dark cycle of mice.

METHOD DETAILS

Surgery

For cranial window implantation, mice were anesthetized with isoflurane (1.5 – 1.8% in oxygen) and acepromazine (1.5 – 1.8 mg/kg body weight). Carprofen (5 mg/kg body weight) was administered prior to surgery. Body temperature was maintained using physically activated heat packs or homeothermic heat pads during surgery. Eyes were kept moist with ophthalmic ointment during surgery. The scalp overlaying the right visual cortex was removed, and a custom steel headplate with a 5 mm diameter opening was mounted to the skull with cyanoacrylate glue (Oasis Medical) and dental acrylic (Lang Dental). A 4 mm diameter craniotomy was performed over the visual cortex and covered with a #1 thickness coverslip, which was secured with cyanoacrylate glue.

Intrinsic signal optical imaging (ISOI)

Prior to two-photon imaging, the locations of primary and HVAs were mapped using ISOI, as previously reported.[4,30,65] Pial vasculature images and intrinsic signal images were collected using a CCD camera (Teledyne DALSA 1M30) and a tandem lens

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

macroscope. A 4.7 × 4.7 mm^2 cortical area was imaged at 9.2 μm/pixel spatial resolution and at 30 Hz frame rate. The pial vasculature was illuminated and captured through green filters (550 ± 50 nm and 560 ± 5 nm, Edmund Optics). The ISO images were collected after focusing 600 μm down into the brain from the pial surface. The intrinsic signals were illuminated and captured through red filters (700 ± 38 nm, Chroma and 700 ± 5 nm, Edmund Optics). Custom ISOI instrumentation was adapted from Kalatsky and Stryker.[9] Custom acquisition software for ISOI imaging collection was adapted from David Ferster.[30] During ISOI, mice were 20 cm from a flat monitor (60 × 34 cm^2), which covered the visual field (110° x 75°) of the left eye. Mice were lightly anesthetized with isoflurane (0.5%) and acepromazine (1.5 – 3 mg/kg). The body temperature was maintained at 37 °C using a custom electric heat pad.[30] Intrinsic signal responses to vertical and horizontal drifting bars were used to generate retinotopic maps for azimuth and elevation. The retinotopic maps were then used to locate V1 and HVAs (Figure S1A). Borders between these areas were drawn using features of the elevation and azimuth retinotopic maps, such as reversals, manually.[4,16] The vasculature map provided landmarks to identify visual areas in two-photon imaging.

In vivo two-photon imaging

Two-photon imaging was performed using a custom Trepan2p microscope controlled by custom LabView software.[30] Two regions were imaged simultaneously using temporal multiplexing.[30] Two-photon excitation light from an ultrafast Ti:Sapph laser tuned to 910 nm (MaiTai DeepSee; Newport Spectra-Physics) laser was split into two beams through polarization optics, and one path was delayed 6.25 ns relative to the other. The two beams were steered independently from each other using custom voice coil steering mirrors and tunable lenses. This way, the X, Y, Z plane of the two paths can be independently positioned anywhere in the full field (4.4 mm diameter). The two beams were raster scanned synchronously about their independently positioned centers by a 4 kHz resonant scanner and a linear scanner (Cambridge Technologies). Photons were detected (H7422P-40, Hamamatsu) and demultiplexed using fast electronics. For four-region scanning, the steering of the two beams was alternated every other frame.

In the current study, two-photon imaging of 500 x 500 $μm^2$ was collected at 13.3 Hz for two-region imaging, or 6.67 Hz for large-field single path imaging (Table S1). We typically imaged neurons in V1 and one or more HVAs simultaneously. Up to 500 neurons (V1: 129 ± 92; HVAs: 94 ± 72; mean ± SD) were recorded per imaging region (500 x 500 $μm^2$). Imaging was performed with typically <80 mW of 910 nm excitation lights out of the front of the objective (0.45 NA), including both multiplexed beams together. Mice were head-fixed about 11 cm from a flat monitor, with their left eye facing the monitor, during imaging. The stimulus display monitor covered 70° x 45° the left visual field. Two-photon images were recorded from awake mice. During two-photon imaging, we monitored the pupil position and diameter using a custom-controlled CMOS camera (GigE, Omron) at 20 – 25 fps. No additional illumination was used for pupil imaging.

Visual stimuli

Visual stimuli were displayed on a 60 Hz LCD monitor (9.2 x 15 cm^2). All stimuli were displayed in full contrast.

The texture stimuli (Figure S2A) were generated by panning a window over a large synthesized naturalistic texture image at one of the cardinal directions at the speed of 32 °/s. We generated the large texture image by matching the statistics of naturally occurring texture patterns.[34] The texture pattern families were: animal fur, mouse chow, rocks, and tree trunk. Each texture stimulus ran for 4 s and was interleaved by a 4 s gray screen.

The random dot kinematogram (**RDK**) stimuli contained a percentage (i.e., coherence) of white dots that move in the same direction (i.e., global motion direction) on a black background (Figure S3A). We presented the animal with RDK at three coherence levels (40%, 70%, and 90%) and four cardinal directions. The dot diameter was 3.8° and the dot speed was 48 °/s. White dots covered about 12.5% of the screen. The lifetime of individual dots were about 10 frames (1/6 s). These parameters were selected based on mouse behavior in a psychometric RDK task.[66] Each RDK stimulus ran for 3 – 7 s (responses in the first 3 s were used for analysis) and interleaved with 3 s gray screen. The same RDK pattern was looped over trials.

Two naturalistic videos (Figure 3A) were taken by navigating a mouse home cage, with or without a mouse in the cage. Each video had a duration of 32 s and were presented with interleaved 8 s long periods with a gray screen. For the convenience of analysis, we concatenated the responses to the two videos (total 64 s).

QUANTIFICATION AND STATISTICAL ANALYSIS

Calcium imaging processing

Calcium imaging processing was carried out using custom MATLAB codes. Two-photon calcium imaging was motion corrected using Suite2p subpixel registration module.[63] Neuron ROIs and cellular calcium traces were extracted from imaging stacks using custom code adapted from Suit2p modules[63] (https://github.com/cortex-lab/Suite2P). Neuropil contamination was corrected by subtracting the common time series (1st principal component) of a spherical surrounding mask of each neuron from the cellular calcium traces.[15,67]

Neuropil contamination corrected calcium traces were then deconvolved using a Markov chain Monte Carlo (**MCMC**) methods[64] (https://github.com/flatironinstitute/CaImAn-MATLAB/tree/master/deconvolution; Figure S1B). For each calcium trace, we repeated the MCMC simulation for 400 times, and measured the reproducibility of MCMC spike train inference for each cell. We computed the Pearson correlation of the entire inferred spike train (tens of mins duration) binned at 13.3 fps. For all subsequent analysis, only cells with stable spike train inference results were included (correlations between MCMC simulations > 0.2). The number of neurons

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

passed this criterion was defined as the total number of neurons (Table S1). The MCMC quality control was defined empirically. For all following analysis, we randomly pick one simulation trial for each neuron. Randomly picking a different trial did not affect results quantitatively.

Neurons in V1 and HVAs exhibited similar instantaneous firing rates, defined as inverse of inter-spike-interval (Figure 1C). Maximal instantaneous firing rates were computed as the inverse of the minimal inter-spike-interval, while average instantaneous firing rates were computed as the inverse of the average inter-spike-interval for individual neurons.

Reproducibility and lifetime sparseness

The reproducibility of responses to naturalistic videos was defined as the trial-to-trial Pearson correlation between inferred spike trains of each neuron binned in 500 ms bins. The reproducibility of responses to texture stimuli and RDK were computed as the fraction of trials that a neuron fired to its preferred stimulus within a time window (4 s for texture stimuli and 3 s for RDK). These definitions were commonly used in previous studies.[46,68] Only reliably responsive neurons were included in the latter analysis (Pearson correlation > 0.08 to naturalistic video; fired on > 50% trials to the texture and RDK stimuli). The qualitative results were not acutely sensitive to the selection criteria.

The lifetime sparseness defines how often one neuron response to a particular stimulus was computed as (Equation 1)[69]:

$$S = \frac{1 - \frac{1}{N} * \frac{\left(\sum_i r_i\right)^2}{\sum_i r_i^2}}{1 - \frac{1}{N}} \quad \text{(Equation 1)}$$

For lifetime sparseness, r_i is trial-averaged response to i^{th} stimulus and N is the length of the stimuli. The sparseness to naturalistic videos was computed using 500 ms bins. The qualitative results of reliability and sparseness were not acutely sensitive to the bin size.

Information analysis

Mutual information (**MI**) evaluates the information the neuronal response (r) has about certain aspects of the stimulus, and it is computed in units of bits. It was computed using the following equation.

$$MI(r,s) = \sum_r \sum_s p_{r,s}(r,s) * log_2 \frac{p_{r,s}(r,s)}{p_r(r) * p_s(s)} \quad \text{(Equation 2)}$$

We computed the MI between neuron responses and the visual stimulus (s has 16 categories for texture stimuli, $p_s(s) = 1/16$; s has 12 categories for RDK, $p_s(s) = 1/12$). We also computed the MI between neuron responses and the texture family (s has 4 categories for texture stimuli, $p_s(s) = 1/4$), and the MI between neuron responses and the moving directions (s has 4 categories for both texture stimuli and RDK, $p_s(s) = 1/4$). The probability of neuron responses was computed from spike count distributions within a stimulus window (4 s for texture stimulus and 3 s for RDK). Reliable RDK and texture reliable neurons (response to >50% of trials) were included for the MI analysis.

Informative neurons

We defined threshold for informative neurons from shuffled data (> mean + 3 SD of MI of shuffled data). That is the spike count ($p_{r,s}(r,s)$) of a neuron shuffled on each trial independently.

Tuning pattern of informative neurons

To estimate the tuning pattern of informative neurons, i.e., which texture pattern or motion direction one neuron responded to, or how many texture patterns one neuron responded to, we decomposed the neuronal responses into motion direction components, and texture family or RDK coherence components using singular value decomposition (**SVD**). To be more robust, instead of using trial-averaged response, we first estimated the neuronal responses by linearly regressing with a unit encoding space (Figures S2E and S3E). Lasso regularization was applied to minimize overfitting. The regularization hyper-parameters were selected by minimizing the cross-validation error in predicting single trial neuronal responses. The linear regression model performance was measured by the Pearson correlation between the trial-averaged neuron response and the model. All informative neurons were well-fit by the model (Figures S2F and S3F).

We then characterized the SVD components of all informative neurons. The SVD vectors reflect the selectivity of a neuron. Absolute values close to one indicate a preferred stimulus while values close to zeros indicate a null stimulus. Some texture informative neurons were activated by multiple texture patterns or motion directions, which were identified as the components that have a >0.4 absolute SVD vector value (qualitative results hold with similar thresholds). We quantified the number of texture patterns and the number of motion directions one texture informative neuron was responsive to (Figure S2G). Also, we identified the preferred motion direction and coherence level for RDK informative neurons (Figure S3G).

Population response dimensionality

We measured the dimensionality of neuron populations with certain size through principal components analysis of trial-averaged response of neurons (spike trains were binned in 500 ms). We generated principal components of each population by the trial-averaged

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

responses that were computed using randomly selected half of the trials (training data), and then used the first x number of principal components to recover the trial-averaged responses of the remaining half of the trials (testing data). The dimensionality was defined as the number of principal components that reproduced the testing data that minimize the mean squared error.

Gaussian mixture model

To characterize the tuning properties in an unbiased manner, neurons were clustered using a Gaussian mixture model[70] (**GMM**) based on the trial-averaged responses to the naturalistic video. Only reliably responsive neurons were included for GMM analysis (trial-to-trial Pearson correlation of the inferred spike trains > 0.08, after spike trains were binned in 500 ms bins). Neuronal responses of the whole population, pooled over all cortical areas, were first denoised and reduced in dimension by minimizing the prediction error of the trial-averaged response using principal component (PC) analysis. 45 PCs were kept for population responses to the naturalistic videos. We also tested a wide range of PCs (20 – 70) to see how this parameter affected clustering, and we found that the tuning group clustering was not acutely affected by the number of PCs used. Neurons collected from different visual areas and different animals were pooled together in training the GMM (3839 neurons). GMMs were trained using the MATLAB build function *fitgmdist* with a range of numbers of clusters. A model of 25 classes was selected based on the Bayesian information criterion (BIC). We also examined models with different numbers of classes (20, 30, 45, or 75), and found that the main results held regardless of the number of GMM classes. Neurons with similar response patterns were clustered into the same class. Figure S5C shows the response pattern of GMM classes to the naturalistic video. The size of the naturalistic video classes is shown in Figure S5D. To examine the reproducibility of the GMM classification, we performed GMM clustering on 10 random subsets of neurons (90% of all neurons). We found the center of the Gaussian profile of each class was consistent (Pearson correlation of class centers, 0.74 +/- 0.12). About 65% of all neurons were correctly (based on the full data set) classified, while 72% of neurons in classes that are over-represented in HVAs were correctly classified. Among misclassifications, about 78% were due to confusion between the three untuned classes with tuned classes. Thus, most of the classes to come out of the GMM analysis appear to be reproducible, and are not sensitive to specific subsets of the data.

Visual features of the naturalistic video

We characterized various visual features of the naturalistic video (Figure S5).

Average luminance

The average pixel value of each frame.

Global contrast

The ratio between the standard deviation of pixel values in a frame, and the average luminance of that same frame.

Edge density

The local edges were detected by a Canny edge detector.[71] The algorithm finds edges by the local intensity gradient and guarantees to keep the maximum edge in a neighborhood while suppressing non-maximum edges. We applied the Canny edge detector after Gaussian blurring of the original image at multiple scales (1°-10°). A binary edge map was generated as the result of edge detection (Figure S4A). The edge density was computed as the sum of positive pixels in the binary edge map of each frame.

Difference of Gaussian (DOG) entropy

We characterized local luminance features following the difference of Gaussian filtering at multiple scales, and then computed the entropy of these features within a local neighborhood (Figure S4B).

Optical flow entropy

We estimated the direction and speed of salient features (e.g., moving objects) using the Horn-Schunck method at multiple spatial scales. Then we computed the entropy of the OF direction and speed at each frame. Since the OF estimation relies on the saliency of visual features, the moving texture and RDK stimuli resulted in distinct OF entropies, with the latter being larger (Figure S5C).

Visual features were computed either by average over space or by computing a spatial variance value (i.e., entropy). These measurements were inspired by the efficient coding theory,[72] which suggested that the neuron population coding is related to the abundance or the variance of visual features in the natural environment.

Discrimination of texture images by DOG features

We computed the pairwise distance between texture images (Figure S4E) within the same class or from different classes (Figure S4F). The Euclidean distance was computed using DOG entropy (11.75° spatial filter size) or edge density (2.35° spatial filter size) feature maps. We then trained a support vector machine (**SVM**) classifier to discriminate texture images within and across classes, based on this pairwise distance (using the MATLAB built-in function *classify*). We reported the cross-validation classification error rate (Figure S4F).

Modulation power of naturalistic features

A linear regression model was fit to the neuronal activity of individual neurons or average population response with individual features (Figure 4). These features are described above in the section ''visual features of the naturalistic video''. We then evaluated a feature's contribution in modulating the average population responses by the variance explained (r-squared) of each model (Figure 3D). Features were computed over multiple spatial scales. The spatial scales that best modulated (highest r-squared) the neuronal response was used for this analysis.

Please cite this article in press as: Yu et al., Selective representations of texture and motion in mouse higher visual areas, Current Biology (2022), https://doi.org/10.1016/j.cub.2022.04.091

Neuronal activity of individual neurons or population average response were binning into 50 ms bins (50 – 500 ms bins were tested and generated qualitatively similar results). To evaluate the functional contribution of over-represented classes of HVAs, we pooled neurons from over-represented tuning classes of one HVA (200 neurons per pool with permutation; 50 – 200 pool size was tested and results hold), and we averaged activity over the pool, and then determined which features modulated activity the over-represented classes in an HVA (Figure 4B).

Gabor-based receptive field models

The neuron models used the structure of a linear-nonlinear (**LNL**) cascade. The spiking of model neurons was simulated following a nonhomogeneous Poisson process with a time varying Poisson rate. The rate was calculated by convolving visual stimuli with a linear kernel or a combination of linear kernels, followed by an exponential nonlinearity (Figure 5A). Linear kernels were modeled by 2D (XY spatial) or 3D Gabor (XYT spatiotemporal) filters defined over a wide range of spatiotemporal frequencies and orientations. We simulated neurons with simple cells, complex cells and motion cells models[29] (Figure 5B). The three differed in the linear components of the LNL cascade: simple cells (called *linear model*, or *spectral model* for the 2D Gabor kernels) used the linear response of a Gabor filter; complex cells (called *energy model*) used the sum of the squared responses from a quadrature pair of Gabor filters (90° phase shifted Gabor filter pairs); speed cells (called *motion model*) used the arithmetic difference between the energy responses from an opponent pair of complex cells. We also modeled neurons based on the cross product of the linear or energy responses from two 2D Gabor filters. In particular, we simulated the following three combination models: 1. 2D Gabor filters matched in spatial scale and location but tuned to different orientations (*cross-orientation model*); 2. 2D Gabor filters tuned to the same orientation and location with different spatial scales (*cross-scale model*); 3. 2D Gabor filters with matched tuning properties but offset in visual space (*cross-position model*) (Figure 5B). In addition, we included a subtractive normalization before taking the nonlinearity in some models. A total of 13 Gabor model types were used (Figure 5B).

To examine feature encoding by these neuron model types, we performed 10 – 20 repeats of simulation for each neuron model to each stimulus. Either the simulated spike trains or peristimulus time histograms (**PSTH**) were used for characterizing the feature encoding. We analyzed the model responses in the same way as we had done for the mouse experimental data. We computed the mutual information between simulated neuron responses and texture stimuli or RDK stimuli, and characterized the selectivity of simulated neurons to texture families or RDK directions (Figure S6). Next, we examined the modulation of simulated population responses by visual features of the naturalistic video. Neuron models were located in the feature space by how much of the population response variance was explained by individual features (Figure 5C).

Fit neuron responses with Gabor-based models

We fit individual neuronal responses with models following a linear regression equation (Equation 3). The linear coefficients were optimized by minimizing the cross-validation error. We also tested a sigmoidal nonlinear fitting (Equation 4). Sigmoidal parameters were optimized through gradient descent. As sigmoidal nonlinearity did not significantly improve the modeling performance, we reported the results from the linear fitting.

$$neuron\ response = a * x; \quad \text{(Equation 3)}$$

$$neuron\ response = \frac{a}{1 + \exp(-b * x + c)}; \quad \text{(Equation 4)}$$

Where x is the simulated response and a – c are parameters for fitting. Neuron models were grouped into three categories: 2D Gabor models, 3D Gabor models, and 3D Gabor models with normalization. One model of each category, which minimizes the cross-validation error, was kept for each neuron. Then, we selected the one among the three that maximize the variance explained of the PSTH for each neuron (Figure 6A). HVAs varied in their preference of different model types (Figure 6B). We then examined the modulation power of naturalistic visual features of enriched neuron models.

Statistical analysis and data availability

P-values were generated by one-way ANOVA with Bonferroni multiple comparison if not otherwise stated. Detailed statistical values were provided in the figure legends. Boxplot whiskers indicate the full data range.