Mapping stimulus feature selectivity in macaque V1 by two-photon Ca$^{2+}$ imaging: Encoding-model analysis of fluorescence responses to natural movies

Koji Ikezoe$^{a,b,c}$, Mahya Amano$^c$, Shinji Nishimoto$^{a,b,**}$, Ichiro Fujita$^{a,b,*}$

$^a$ Graduate School of Frontier Biosciences, Osaka University, 1-4 Yamadaoka, Suita, Osaka, 565-0871, Japan
$^b$ Center for Information and Neural Networks, Osaka University and National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita, Osaka, 565-0871, Japan
$^c$ Faculty of Medicine, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan

** Corresponding author. Center for Information and Neural Networks, National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita, Osaka, 565-0871, Japan.

E-mail addresses: nishimoto@nict.go.jp (S. Nishimoto), fujita@fbs.osaka-u.ac.jp (I. Fujita).

ARTICLE INFO

Keywords:
- Two-photon microscopy
- Calcium imaging
- Striate cortex
- Orientation map
- Receptive field
- Motion-energy model

ABSTRACT

In vivo calcium (Ca$^{2+}$) imaging using two-photon microscopy allows activity to be monitored simultaneously from hundreds of individual neurons within a local population. While this allows us to gain important insights into how cortical neurons represent sensory information, factors such as photo-bleaching of the Ca$^{2+}$ indicator limit imaging duration (and thus the numbers of stimuli that can be tested), which in turn hampers the full characterization of neuronal response properties. Here, we demonstrate that using an encoding model combined with presentation of natural movies results in detailed characterization of receptive field (RF) properties despite the relatively short time for data collection. During presentation of natural movie clips to macaque monkeys, we recorded fluorescence signals from primary visual cortex (V1) neurons that had been loaded with a Ca$^{2+}$ indicator. For each recorded neuron, we constructed an encoding model that comprised an array of motion-energy filters that tiled over the RFs. We optimized the weight of each filter's output so that the linear sum of the outputs across the filters mimicked the neuron's Ca$^{2+}$-signal responses. These models were able to predict the neural responses to a different set of natural movies with a significant degree of accuracy. Moreover, the orientation tunings of neurons simulated by the model were highly correlated with those experimentally obtained when grating stimuli were presented to the monkeys. The model predictions were also consistent with what is known about spatial frequency tunings, the structure of excitatory subfields of RFs (i.e., classical RFs), and functional maps for these RF properties in V1. Further analysis revealed a new aspect of V1 functional architecture; the extent and distribution of suppressive RF subfields varied among nearby neurons, while those for excitatory subfields were shared. Thus, applying our encoding-model analysis to two-photon Ca$^{2+}$ imaging of neuronal responses to natural movies provides a reliable and efficient means of analyzing a wide range of RF properties in multiple neurons imaged in a local region.

INTRODUCTION

Sensory information is represented by the activity of large neuronal ensembles in the brain. Characterizing the response properties of as many neurons involved in these representations as possible is a key to understanding exactly how sensory information is represented. Two-photon imaging of neurons loaded with Ca$^{2+}$-fluorescent indicators allows spiking activity to be monitored simultaneously in hundreds of individual neurons within a local region (Stosiek et al., 2003; Grienberger and Konnerth, 2012). A great advantage of in vivo two-photon Ca$^{2+}$ imaging over multiple single-unit recordings with electrodes is its sub-micron spatial resolution (Denk et al., 1990) that allows precise localization of individually recorded neurons that can help determine functional micro-architecture (Ohki et al., 2005, 2006; Smith and Häusser, 2010). Recently developed versions of two-photon microscopes have an expanded field-of-view (~20 mm$^2$) and semi-simultaneous...
sampling at different depths (Göbel et al., 2007; Cotton et al., 2013; Ji et al., 2016; Sofroniew et al., 2016; Stryker et al., 2016; Szalay et al., 2016), which enables thousands of neurons to be recorded across multiple sensory areas. Two-photon Ca$^{2+}$- imaging is also applicable to non-human primates (Nauhaus et al., 2011a,b, 2016; Ikezoe et al., 2013; Sadakane et al., 2015; Li et al., 2017; Tang et al., 2018). The high spatial resolution of two-photon imaging enables long-term monitoring of neuronal activity in awake and behaving monkeys (Li et al., 2017; Tang et al., 2018).

Despite these exciting new developments, analyzing properties of individual neurons in a local population is still difficult because photo-bleaching of Ca$^{2+}$ indicators limits the duration during which neuronal responses to stimuli can be measured, thus precluding an examination of responses to different types of stimuli. Even if genetically encoded calcium indicators are used, this problem persists (Dana et al., 2016; Mizuno et al., 2018) when monitoring neuronal activity over an extended time period. In the visual system for example, even when a limited set of visual stimuli are used, they need to be optimized in position and size for the receptive fields (RFs) of individual neurons. This requires time, which increases with the number of neurons. Added to this, tuned features often differ among neurons (e.g., orientation vs. color), and sensory stimuli that can evoke responses in some neurons may be inappropriate for examining responses of other neurons recorded at the same time. Moreover, the time required for optimizing the stimuli prevents us from studying changes in response properties over time, such as those that occur through attentional shift, postnatal development, and learning. One solution is to examine the transfer function of individual neurons using system-identification techniques (Wu et al., 2006) as fast as possible.

Here, we demonstrate that extensive examination of RF properties from a large number of individual neurons can be achieved by applying an encoding-modeling approach to two-photon imaging data that are obtained over a relatively short period using movies of natural scenes. Encoding models quantitatively characterize neuronal stimulus-response functions using system-identification techniques (Wu et al., 2006). This approach has been used to model spiking neuronal activity (Pillow et al., 2008; Nishimoto and Gallant, 2011; Yamins et al., 2014) as well as hemodynamic responses recorded using functional magnetic resonance imaging (fMRI) techniques (Naselaris et al., 2011; Nishimoto et al., 2011; Huth et al., 2016; Güçli and van Gerven, 2017). This also revealed change of response properties through attentional shift (David et al., 2008; Çukur et al., 2013). Another merit of this technique is that characterizing neuronal stimulus-response functions using encoding models does not require explicit estimation of spike timing or instantaneous firing rates. Thus, in the fMRI studies referred to above, the models were created directly from hemodynamic responses. Similarly, we can create encoding models from fluorescence responses without estimating spike timing or firing rates.

We applied these techniques to the primary visual cortex (V1) of macaque monkeys. We first constructed models for individual V1 neurons using responses to natural movies. The models predicted the individual responses to another set of natural movies with significant accuracy. The models also predicted orientation tuning and orientation maps. Moreover, these predictions were consistent with those reported for spatial frequency tuning in previous studies (De Valois et al., 1982; Bredfeldt and Ringach, 2002; Nauhaus et al., 2012b). Finally, we found a novel feature of V1 functional architecture: excitatory RF subfields (i.e., classical RFs) of nearby neurons were similar in shape and position, whereas suppressive surround subfields were substantially different. Our encoding-model analysis, combined with natural movie stimuli and two-photon Ca$^{2+}$ imaging, thus provides a reliable and efficient means of analyzing a variety of RF properties for multiple neurons in a local region.

Materials and methods

All experimental procedures were approved by the animal experiment committee of Osaka University, and conform to the Guide for the Care and Use of Laboratory Animals issued by the National Institutes of Health, USA (DHEW Publication No. (NIH) 85–23, Revised 1996, Office of Science and Health Reports, DRR/NIH, Bethesda, MD 20205, USA).

Animal preparation

We recorded from three female macaque monkeys (Macaca fascicularis, body weight: 2.3–3.2 kg). Monkeys were prepared for repeated recordings through an initial aseptic surgery more than two weeks before the first session of recording (for details, see Ikezoe et al., 2013). On each day of a recording session, a small craniotomy and durotomy (2–3 mm in diameter) were performed over the region of V1 to be imaged. Anesthesia was initially induced with ketamine hydrochloride (Ketalar, Daichi-Sankyo, Tokyo; 11.5 mg/kg, i.m.) after an injection of atropine sulfate (Atropine Sulfate Injection, Mitsubishi Tanabe Pharma, Osaka; 0.025 mg/kg, i.m.), and was maintained with isoflurane (1%–3%) in a mixture of nitric oxide and oxygen (7:3). The exposed portion of the dura and cortex was covered with agar and a cover-glass. After surgery, we stopped administering isoflurane, and switched to fentanyl citrate (Fentanyl, Daiichi-Sankyo; 10 μg/kg/h, i.v.) for analgesia (Pupilsik and Kohn, 1997), and administered vecuronium bromide (Muculax, Merck Sharp and Dohme, Tokyo; 0.08 mg/kg, i.v.) to prevent eye movements. The monkeys were artificially ventilated with a mixture of nitrogen and oxygen. Body temperature was maintained at 37–38 °C and end-tidal CO$_2$ was kept at 4.0%–5.5%. Electrocardiogram, blood pressure, and arterial oxygen-saturation levels were continuously monitored and maintained within appropriate physiological ranges. We administered phenylephrine hydrochloride and tropicamide to the eyes to relax accommodation and dilate the pupils. Corneas were covered with hard contact lenses with pupils (3 mm in diameter) to prevent drying and to allow the images on the stimulus display to be focused on the retina. Contact lenses of appropriate curvature and power were chosen for individual eyes based on the measurement by using an auto kerato-refractometer (KR-7100, Topcon, Tokyo; Tamura et al., 2004).

At the end of each recording session, neostigmine methylsulfate (Vagostigmin, Shinongi, Osaka; 0.1 mg/kg, i.m.) was administered to aid the recovery of spontaneous respiration. Ketoprofen (Megeide, Nissin Pharmaceutical,., Yamagata, Japan, 0.8 mg/kg, i.m.) was administered for post-surgery analgesia. The monkeys were then returned to their home cage and given ketoprofen and antibiotics (Piperacillin sodium, Taisho Toyama Pharmaceutical, Tokyo) for one week post-experiment.

In vivo calcium imaging

We performed in vivo Ca$^{2+}$ imaging of a population of neurons in layers 2 and 3 of V1. We injected a membrane-permeable calcium indicator, Oregon Green 488 BAPTA-1 AM (OGB-1, Thermo Fisher Scientific, MA) or Cal-520 AM (Cal-520, AAT Bioquest, CA) at a depth of 250–300 μm from the cortical surface through a micropipette using air pressure via a Picospitzer III device (General Valve, Fairfield, NJ). OGB-1 (0.3 mM) or Cal-520 (1.0 mM) was dissolved in a solution containing 0.2% Pluronic F-127 (Thermo Fisher Scientific, 2.5% dimethyl sulfoxide (Sigma-Aldrich), 10 mM HEPES, 2.5 mM KCl, and 150 mM NaCl (pH 7.4). Sulforhodamine 101 (SR 101; 100 μM, Thermo Fisher Scientific) was injected together with the Ca$^{2+}$ indicator to identify astrocytes (Nimmerjahn et al., 2004). Fluorescence was imaged with a two-photon microscope (MOM, Sutter, CA) equipped with a mode-locked Ti:sapphire laser (MaiTai, Spectra Physics, CA). In most experiments (n = 17), the microscope was equipped with 16 × objective lenses (CFI75 LWD 16 × W, NA 0.8, Nikon, Tokyo), a resonant scanner for the X direction and a galvano-scanner for the Y direction. The microscope was controlled by MOM Computer System and Software 2.0 (Sutter). The frame rate was 30.9 frames/s. In the other two experiments, the microscope was equipped with a 40 × objective lens (LUMPLFLN 40 × W, NA 0.8, Olympus, Tokyo) and two galvano-scanners for the X and Y directions.
The microscope was controlled using ScanImage software (Polgárdi et al., 2003). The frame rate was 7.81 frames/s. The sizes of the imaged areas were 462 μm × 462 μm (2 regions, 1 monkey), 315 μm × 315 μm (8 regions, 1 monkey), 355 μm × 355 μm (8 regions, 1 monkey), and 630 μm × 630 μm (1 region).

Visual stimulation

We presented natural movie clips (10′ × 10′) on a gray background using a liquid crystal display (MDT231WG, Mitsubishi electric, Tokyo) or an organic electroluminescence display (PVM-1741, Sony, Tokyo). The center portion of the movies covered the minimum response fields of the neurons that had been determined earlier using multiple-unit recording or the fluorescence responses to presentations of 2′ × 2′ rectangular patches of natural movies at varying positions. The movie stimuli were composed of short clips (~10 s) showing natural scenes, moving animals, humans, letters, and sentences (Nishimoto and Gallant, 2011; Nishimoto et al., 2011; available at https://crcns.org/data-sets/vc/vim-2/about-vim-2). The movie frame-rate was 24 Hz. We showed six 10-min movies without any gaps between individual short clips. Three movies were used for training the encoding model (training sets) and the others were for testing the quality of the model (test sets, see below). For the training sets (10 min × 3 movies = 30 min), each short clip appeared only once, whereas in the test sets (10 min × 3 movies = 30 min), three different 1-min movies consisting of fixed sequences of multiple short clips were each pseudo-randomly presented 10 times. In each 10-min movie, the 1-min movies appeared three or four times. The training sets and the test sets were interleaved with intervals of 1–10 min. After showing the all movies, we also presented achromatic drifting sinusoidal gratings with orientations of 0°, 30°, 60°, 90°, 120°, and 150° for 17/19 imaged regions. In 15 experiments, the gratings were in a circular patch, 1′ or 2′ of visual angle in diameter, and presented over the minimum response fields of the neurons. In the other two experiments, the gratings had the same size as the natural movies (i.e., 10′ × 10′). The drifting directions of the grating were orthogonal to the orientation. Spatial frequencies of the gratings were 1, 2, and 4 cycles/degree.

Data analysis

We used a template to correct constant image-distortion caused by non-linear resonant scanning. We then used 2-dimensional cross-correlation to remove displacement of the images caused by motion of the brain (Guizar-Sicairos et al., 2008). To quantify the fluorescence of individual neurons, we manually selected the cell-body pixels for each neuron from the average across all acquired images. We then averaged the fluorescence across the cell-body pixels of each neuron. This was repeated for each image to obtain the time courses for fluorescence for each individual neuron. The fluorescence data were high-pass filtered (cut-off frequency: 0.02 Hz) to remove slow fluctuations of signal strength that can be caused by cortical movement in the depth direction or by photo-bleach. To examine responses to the stimuli, we normalized the changes (dF/F) in fluorescence signals by the DC component (FO) of the fluorescence (dF/F = F(t) - FO). The DC component was estimated by applying the median filter for the raw signals (filter width = 50 s).

For individual neurons, we calculated the signal-to-noise ratio of fluorescence signals, which is the ratio of the maximum fluorescence-signal magnitude for the period of entire stimulus presentation to the noise level. The noise level was estimated as the standard deviation of the fluorescence signals during a 30-s period before movie presentation (5 s × 6 trials). We evaluated consistency of responses across 10 presentations of the test-set movies by calculating the explainable variance (Sahani and Linden, 2003) as follows:

\[ \bar{\gamma}(t) = \frac{1}{N} \sum_{n=1}^{N} y_n(t), \]

where \( y_n(t) \) is the fluorescence signal (dF/F) for a time bin in the n-th trial, \( N \) is the number of trials, and \( \bar{\gamma}(\cdot) \) denotes the variance. Explainable variance indicates a proportion of variance that can be explained by stimulus-related component \( \gamma(t) \) among overall variance of fluorescence signals \( y(t) \). Explainable variance is 1 when fluorescence responses to the repeated presentations of an identical movie are completely the same across the repeats. When fluorescence responses are inconsistent across trials, the value is close to zero.

Encoding-model analysis

To estimate stimulus-feature selectivity for each neuron, we fit a motion-energy encoding model to the fluorescence signals that were evoked by the natural movies (Fig. 1A). The details of the model design have been described elsewhere (Nishimoto and Gallant, 2011; Nishimoto et al., 2011). First, color movies were converted into the Commission International de l’Eclairage (CIE) L*A*B* color space and only the luminance (L*) information was used for the following filtering stages. The model consists of two stages of filter arrays (Fig. 1B).

The first stage was an array of nonlinear, spatiotemporal motion-energy filters (Fig. 1B left; Adelson and Bergen, 1985; Watson and Ahumada, 1985). Each filter detects motion signals within the movies for varying directions of motion, spatial frequencies, temporal frequencies, and spatial locations. The center positions of the filters were positioned on a two-dimensional Cartesian grid that tiled a localized analysis window. The dimensions of the filter array varied across recording sites and typically spanned an analysis window that covered roughly twice as much areas as a classical RF and had 12 movement directions (6 orientations), a spatial frequency of 0–24 cycles/analysis window (stepped in octaves, except for 0-cycle filters), and a temporal frequency of 0–6 Hz (stepped in octaves, except for 0-cycle filters). We chose the filter ranges so as to cover the known ranges of temporal and spatial frequency selectivity in V1 (De Valois et al., 1982; Foster et al., 1985). The numbers of directions, frequencies, sizes, and positions were determined to mimic typical sampling intervals in physiological experiments (e.g., sampling direction selectivity at a 30-degree step and frequency selectivity at one-octave step). Outputs of filters for opponent directional motions were pooled without weighting to reduce the total numbers of second stage filters. The filtered signals were then compressed with a log-transform and down-sampled to 6 Hz. The log-transform partially explained by stimulus-related component \( \gamma(t) \) among overall variance of fluorescence signals \( y(t) \). Explainable variance is 1 when fluorescence responses to the repeated presentations of an identical movie are completely the same across the repeats. When fluorescence responses are inconsistent across trials, the value is close to zero.

\[ \text{Explainable Variance} = 1 - \frac{\text{Var}(y(t) - \bar{\gamma}(t))}{\text{Var}(y(t))} \]

The second stage was an array of linear temporal filters that followed each of the first stage filters (Fig. 1B right). These temporal filters were designed to capture temporally delayed contributions of motion-energy signals to the fluorescence signals of each neuron (Smetters et al., 1999). The temporal filters spanned the delays from 0 to 1 s at a step of 1/6 s. The model output was the linear sum of the filtered signals. Because responses of V1 neurons to natural images are sparser than those to bars and gratings (Vinje and Gallant, 2000), the linear relationship between fluorescence and firing rate might be maintained (Ikezoe et al., 2013).

We modeled fluorescence signals for individual neurons by optimizing the second stage filters using L2-regularized (ridge) linear regression (Huth et al., 2012). To find the optimal regularization parameters, we divided the entire 30-min training-data samples into 50 chunks (i.e., each chunk was 36 s of consecutive data), and randomly picked 90% of the chunks for regression. For each recording site, we
chose the regularization parameter that gave the highest average response-prediction accuracy for the remaining 10% of the data (the optimal regularization parameter). The final model was then obtained by regressions that included all training data and the optimal regularization parameter. Neuron-wise modeling accuracy was then quantified using Pearson’s correlation coefficient \( r \) between the measured and predicted fluorescence signals for the 30-min test sets that were reserved for this purpose. We also calculated goodness-of-fit \( R^2 \), which evaluates modeling accuracy in terms of the magnitude and shape of the signals. In our data, the goodness-of-fit \( R^2 = 0.14 \pm 0.15 \), mean ± s.d.) was close to the square of Pearson’s correlation coefficient \( r = 0.40 \pm 0.15 \), mean ± s.d.), suggesting that our model fit not only the shape, but also the magnitude and the offset of the time course of the fluorescence responses. For this reason, we used Pearson’s correlation coefficient to quantify the modeling accuracy. Note that the degree of the prediction performance cannot be directly compared with the degree of the explainable variance.

**In silico examination of neuronal response properties**

Once we fit the encoding models using the responses to the natural movies, we estimated visual feature selectivity of each neuron using a simulated system-identification procedure (in-silico simulations; Nishimoto et al., 2011). We simulated the responses of these model neurons to a set of moving grating stimuli on the virtual screen (Fig. 1C). The size, shape, location, spatial frequency, and temporal frequency of the stimuli were identical to those used in the in vivo experiments. A blank (uniformly gray) stimulus was added for estimating the response baseline. For each neuron, we estimated model responses to the stimulus set to construct predicted orientation and spatial frequency-tuning functions. Models whose orientation and spatial frequency tunings were fit well with von Mises function and Gaussian function, respectively (goodness-of-fit, \( R^2 \geq 0.7 \)), were used for further tunings analysis.

Similarly, we estimated excitatory RF center and suppressive RF surrounds for each neuron via in-silico simulations (Fig. 8a; Nishimoto et al., 2011). In this case, we simulated the responses of model neurons to dynamic white noise stimuli presented at various positions across the virtual screen. The size and position of the virtual screen were identical to those of the analysis window used for motion-energy modeling analysis. The noise stimuli tiled the screen in a 7 × 7 grid. The responses were aggregated into a two-dimensional selectivity map for visualization. The responses are shown relative to the predicted response for the uniform gray stimulus.

**Analysis of orientation tunings**

Neurons whose fluorescence magnitude varied according to the grating orientation \( p < .05 \), Kruskal-Wallis test) were defined as orientation selective. The mean fluorescence change to repeated presentations of a grating was used as the response. Neuronal or predicted preferred orientation was determined as follows:

\[
\text{Preferred orientation} = \frac{1}{2} \arctan \left( \sum \frac{|r_i| \exp (2\theta_i)}{\sum r_i^2} \right)
\]

where \( k \) is a stimulus number, \( r_k \) is the magnitude of the neuronal or predicted responses to orientation \( \theta_k \), and \( i \) is an imaginary unit. The strength of orientation selectivity was estimated with the circular variance (CV) of the predicted responses. The CV was calculated as follows:

\[
\text{Circular variance (CV)} = 1 - \frac{\sum |r_i| \exp (2\theta_i)}{\sum r_i^2}
\]

We quantified the degree of heterogeneity for the predicted orientation preferences around a given pixel by calculating a local heterogeneity index (LHI) as follows (Ikezoe et al., 2013):

\[
\text{Local heterogeneity index} = 1 - \frac{\sum D_n \exp (2\theta_n)}{\sum D_n}
\]

where \( n \) is the number that identifies a neuron, \( d_n \) is the distance between a given pixel and the neuron, \( n \), \( \theta_n \) is the predicted preferred orientation of the neuron \( n \), and \( \sigma \) equals 50 μm.

**Results**

**Fluorescence responses of neurons to natural movie presentation**

We recorded fluorescence signals from 2379 V1 neurons in 19 imaged regions of three monkeys. Fluorescence changes differed across neurons during presentation of natural movies in a given imaged area (Fig. 2A and B). Some neurons showed a frequent increase in fluorescence (neuron 1), while others showed a sparse increase (neuron 2). These changes in fluorescence were robust and consistent across 10 repeated presentations of the movies, indicating that they were neuronal responses to the visual stimuli (Fig. 2B). We measured the consistency of fluorescence change across the repeated presentations of the same movie stimuli by calculating explainable variance (see Materials and Methods). Explainable variance depends on the variability in neuronal response (i.e., neuronal noise) and the recording noise (e.g., electric noise in the imaging system or fluctuation in fluorescence caused by movement of the cortex). The explainable variance of the 2379 neurons ranged from 0.06 to 0.87, and the median was 0.19 (\( n = 2379 \); Fig. 2D), suggesting that at this resolution, approximately 19% of the variance over time can be...
explained by the stimulus sequence. We examined how the explainable variance changed with the number of trials (2–10 trials) using a bootstrap analysis. The explainable variance decreased gradually as a function of the total number of trials with the slope becoming shallower. The standard deviation of the explainable variance for 10 trials was 0.04 (mean ± s.d.), suggesting that our estimates were accurate.

The explainable variance was correlated with the signal-to-noise ratio of the fluorescence signals (see Materials and Methods; \( r_s = 0.54, p < .0001 \), Spearman’s rank correlation coefficient; Fig. 2E), as expected. Neurons with lower explainable variance were often located in the peripheral regions of the imaged regions (Fig. 2C) where fluorescence was weak and susceptible to recording noise.

**Evaluation of the encoding-model analysis: prediction performance for novel movies**

For each neuron imaged, we constructed an encoding model comprising a large set of motion-energy and temporal filters (see Materials and Methods, Fig. 1A and B). The temporal filters were fit using neuronal responses to a training set of 30-min natural movies (training movies). In the constructed encoding model for each neuron, 22.8 ± 1.1% (mean ± s.d.) of the filters contributed to 50% of the sum of absolute weight and 49.4 ± 1.1% contributed to 80%. The weights peaked at 0 s and decreased to 60% at 1/6 s and almost 0% at 5/6 s on average. To test the performance of the model, we virtually presented (i.e., fed to the model as inputs) a 3-min movie (test movie) that had not been used for model construction. The output of the model (Fig. 3A, orange) generally followed the averaged response of the corresponding neuron across 10 repeated presentation of the movie (see Materials and Methods, Fig. 3A, blue). However, the model outputs did not follow the large fluorescence changes because the model includes a compressive non-linearity filter for stabilization that prevents excessively large outputs (“log” in Fig. 1B). We evaluated the model’s prediction accuracy by calculating Pearson’s correlation coefficients between the model output and the responses of the corresponding neurons. The correlation coefficient for the example shown in Fig. 3A was 0.70 (\( p < .0001 \)). In the imaged region shown in Fig. 3B, the mean correlation coefficient across neurons was 0.62 ± 0.09 (mean ± s.d., \( n = 126 \); Fig. 3C). For the entire population of neurons (19 imaging sites), the correlation coefficient was 0.40 ± 0.15 (Fig. 3D) and significantly larger than zero (\( p < 10^{-5} \), Wilcoxon signed-rank test), indicating that the encoding models predicted the responses of neurons with a significant degree of accuracy.

We next investigated how the variability in prediction performance arose among the neurons. First, neurons were not clustered, or were only weakly clustered, across the cortical surface according to their prediction performance. In the imaged region shown in Fig. 3B, the difference between the prediction performances between pairs of neurons did not correlate with their physical distance across the cortex (Fig. 3E; \( r_s = 0.01, p = .30, 7875 \) pairs). The prediction for the entire population from the 19 sites was very weak (Fig. 3F; \( r_s = 0.09, 10^{-5}, 171,755 \) pairs). Second, the prediction performance positively correlated with the explainable variance (Fig. 3G; \( r_s = 0.43, p < .0001 \)); the response of a neuron with consistent fluorescence signals across trials was more accurately predicted by its encoding model. In contrast, prediction performance as a whole hardly correlated at all with the signal-to-noise ratio of the fluorescence for individual neurons (Fig. 3H; \( r_s = 0.07, p = .0007 \); Note that the \( p \)-value was very low because of the large number of data points). The responses of neurons with high signal-to-noise ratio were not always predicted more accurately by the model. Indeed, the responses of many neurons located in dark peripheral regions (i.e., weaker optical signals) in Fig. 3B were well predicted by the corresponding models. These results indicate that prediction performance was higher for neurons that responded to the stimuli consistently across trials than for those with variable responses, irrespective of the signal-to-noise ratio of their fluorescence. For further analysis, we excluded neurons with less than 0.2 of prediction performances (\( n = 248, 10.4\% \); open bins in Fig. 3D). Changing this threshold from 0.2 to 0.1 or 0.3 did not affect the overall conclusions reported in this paper.
Relationship between prediction performance and the signal-to-noise ratio. Relationship between prediction performance and explainable variance. (H) imaged region shown in B; F: a density plot for the population data). (G)

physical distance between the two neurons (E: a scatter plot of data from the experimentally determined orientation tunings with those estimated by the models). We then compared the orientation tunings of the imaged neurons. We did not evaluate the spatial frequency tuning curves of the neurons because we tested only three different frequencies (1, 2, and 4 cycles/degree). We therefore compared the model predictions to the predicted tuning curves (Spearman's rank correlation coefficient, $r_s < 0.25$, $p < .001$; $r_s = 0.30$, $p < .001$, respectively, Fig. 4G and H). Thus, some models that provided highly accurate predictions for natural movies did not do so for gratings orientation, suggesting that selectivity for aspects other than orientation, such as luminance, had a significant impact on the neuronal activity. Further, we calculated the confidence intervals for the preferred orientations determined in vivo using a resampling technique. The magnitudes of confidence intervals were positively correlated with the differences between the preferred orientations estimated in vivo and in silico ($r = 0.31$, Pearson's correlation coefficient, $p = .002$). This correlation indicated that the discrepancy was partially ascribable to the limited precision with which the preferred orientations of neurons could be estimated. The difference between the model-based and experimentally determined preferred orientation was correlated with the strength of orientation tunings of the neurons ($r_s = 0.45$, $p < 10^{-5}$, Spearman's rank correlation coefficient, Fig. 4I). Furthermore, the tuning correlation was negatively correlated with the tuning strength ($r_s = -0.48$, $p < 10^{-5}$, Spearman's rank correlation coefficient, Fig. 4J). These results indicated that weak orientation tunings of the neurons lowered the precision of prediction on the properties of orientation tuning.

Using gratings, we have previously shown that V1 neurons in a local region with heterogeneous orientation preferences are broadly tuned to orientation (Ikezoe et al., 2013; see also Ohki et al., 2006; Nauhaus et al., 2008). The functional arrangement of neurons revealed by the models was consistent with this previous result. In the images seen in Fig. 5A and B, regions can be seen in which the preferred tunings differed by more than 45° for some neurons, and the tuning curves (predicted vs. experimentally determined) were negatively correlated (Fig. 4B–F). Prediction of the response to the natural movies was not necessarily poor in these neurons (Fig. 3B). For the entire population (17 imaging regions), only a weak correlation was observed between prediction performance for natural movies and that for preferred orientations and tuning curves (Spearman's rank correlation coefficient, $r_s = 0.01$, $p = .30$).

We further validated the models by examining how well they predicted neuronal orientation selectivity (Fig. 4). For 17/19 recorded regions, we presented drifting gratings to the animals and examined orientation tunings of the imaged neurons. We then compared the experimentally determined orientation tunings with those estimated by virtually presenting the same stimulus to the encoding models of the corresponding neurons. To facilitate an accurate comparison of orientation selectivity between in vivo and in silico experiments, we re-fit the encoding models to the fluorescence responses to the natural movie for the visual field where gratings were presented in vivo. The models accurately predicted preferred orientations for most of the orientation selective neurons. In the imaged region shown in Fig. 4A, the preferred orientations of the neurons (blue) were approximately 90° (vertical) in the center of the imaged region and gradually changed to 45° (right-tilted) toward the right portion. This spatial arrangement was reproduced by the model (orange) remarkably well. The preferred orientation of the neurons and their models were highly correlated (Fig. 4B; Circular correlation coefficient, $r = 0.68$, $p < 10^{-5}$). However, for some neurons in the periphery of the imaged region, we found large differences between predicted and experimentally determined preferred orientations. The median difference between them was 7.1° for this region (Fig. 4C). Calculation of Pearson's correlation coefficients between the orientation tuning curve of the neuron and that of the prediction revealed that they matched well with each other (Fig. 4D; the median $r = 0.82$). For the entire population (17 imaging regions), 62% (360) of the models for 578 orientation-selective neurons (Kruskal-Wallis test, $p < .05$) predicted the preferred orientations with less than 25° of difference (Fig. 4E; median of the differences = 14.8°), and 51% (296/578) of the models predicted tuning curves with a high accuracy ($r > 0.65$; Fig. 4F).

Evaluation of encoding models: prediction performance for grating stimuli

Fig. 3. Evaluation of model performance by examining responses to a test set of natural movies. (A) Responses of a neuron (blue) and the predicted responses (orange). (B) A cortical map of prediction performance for the encoding models (i.e., correlation coefficients between responses and predictions). (C, D) Frequency distributions of the model performance (C; the imaged region shown in B; D: population data). Data exhibiting < 20% performance accuracy (open bins) were excluded from further analyses. (E, F) Relationship between a pair's difference in correlation coefficients and the physical distance between the two neurons (E: a scatter plot of data from the imaged region shown in B; F: a density plot for the population data). (G) Relationship between prediction performance and explainable variance. (H) Relationship between prediction performance and the signal-to-noise ratio.
findings reported in the literature. The preferred spatial frequency, defined as the peak frequency of the fitted Gaussian curve, ranged from 0.3 to 4.5 cycles/degree. The median of preferred spatial frequency across the neurons was 2.3 cycles/degree, which is comparable to that reported in previous single-unit recording studies (Fig. 6A; De Valois et al., 1982; Bredfeldt and Ringach, 2002). Additionally, the predictions could generate functional maps for the preferred spatial frequency, as has been experimentally demonstrated in previous reports (Silverman et al., 1989; Nauhaus et al., 2012b). Neurons with similar preferred spatial frequency were located closely together, and the predicted spatial frequency of neurons changed gradually with the tangential position of the neurons in the cortex. In the imaged region of Fig. 6B, neurons preferring higher spatial frequencies are found in the lower left portion, whereas the neurons preferring lower spatial frequencies are found in the upper right portion. The degree to which preferred spatial frequencies differed between two neurons was correlated with their physical distance ($r_p = 0.17$, Spearman’s rank correlation coefficient, $p < 10^{-5}$). In the individual 19 imaged regions, the correlation coefficients were not very high, and ranged from $-0.04$ to $0.24$ ($0.089 \pm 0.06$). This indicates that the neurons were arranged according to their spatial frequency preference, but with substantial scatter, as has been reported earlier (Nauhaus et al., 2012b).

Our model also predicted a strong negative correlation between the preferred spatial frequency and the circular variance of orientation tuning; neurons with weaker orientation tuning (higher circular variance) preferred lower spatial frequencies (Fig. 6C for the imaged region shown in 6B; Fig. 6D for all 2049 neurons across the 19 imaged regions).

In layers 2 and 3 of monkey V1, neurons with similar preferred orientations or spatial frequencies are clustered orthogonally to the cortical surface in a columnar fashion (Hubel and Wiesel, 1968; Nauhaus et al., 2012b). Our models also reproduced orientation columns and spatial frequency columns. In one experiment, we collected data from the same tangential position in V1, but at three different depths (136, 156, and 181 μm; Fig. 7). We created orientation maps and spatial frequency maps for the three depths based on the neuronal tuning to these parameters predicted by the models. When the three maps were superimposed, we found that irrespective of different cortical depths, neurons located at nearby tangential positions shared similar preferred orientations and spatial frequencies (Fig. 7A and B).

To quantify the consistency of the orientation maps across depths, we calculated LHIIs (see Materials and Methods; Ikezoe et al., 2013) for individual neurons in the overlaid map and in the individual maps. LHI values were small and comparable between the two types of maps (median: 0.13 vs. 0.11, Wilcoxon signed-rank test, $p = .03$). This indicates that the orientation preferences were consistent along the depth direction. For the spatial frequency map, we calculated standard deviations of the preferred spatial frequencies for the neurons within a 50-μm radius of

![Fig. 4. Model prediction of orientation tuning and orientation maps. (A) Orientation maps for the neurons (blue) and model predictions (orange) of an imaged region. The bar shows preferred orientation of a neuron located at that position. (B) Relationship between the preferred orientations of the neurons and the model predictions. (C, E) Frequency distributions of the difference between preferred orientations of the neurons and those of the predictions (C: data from the imaged region shown in A; E: population data). (D, F) Frequency distributions of the correlation coefficients between orientation tuning curves for the neurons and those for the predictions (C: data from the imaged region shown in A; E: population data). (G) Relationship between prediction performance for validation movie stimuli and the difference in preferred orientation between the neurons and the predictions. (H) Relationship between prediction performance for validation movie stimuli and the correlation between tunings for the neurons and for the predictions. (I) Relationship between the strength of orientation selectivity (circular variance) of neurons and the difference in preferred orientation between the neurons and the predictions. (J) Relationship between the strength of orientation selectivity (circular variance) of neurons and the correlation between tunings for the neurons and for the predictions.](https://example.com/fig4)
each neuron in the overlaid and individual maps. Standard deviations were small and comparable between the two maps (median: 0.85 vs. 0.82 cycles/degree, Wilcoxon signed-rank test, \( p = 0.15 \)). These analyses may possibly be confounded by the difference in the number of neurons among different planes; the plane with the largest number of neurons has the strongest impact on the superimposed map. To reduce this confounding factor, we next assessed map consistency by calculating pixel-based maps, which were obtained by blurring the cell-based maps with a Gaussian envelope (sigma = 50 \( \mu \text{m} \)). Then, for tangential positions of neurons we calculated difference between preferred orientations (and preferred spatial frequency) for all pairs of three imaging sites at different depths. The absolute value of differences were 0.7, 5.7, 20.2 cycles/degree (10, 50 and 90 percentiles, orientation) and 0.04, 0.23, 0.51 cycles/degree (10, 50 and 90 percentiles, spatial frequency). The consistency of the orientation and spatial frequency maps across the three depths, which were imaged and analyzed independently, provides further support that our modeling approach accurately predicted the neuronal tuning properties.
Excitatory and suppressive RF subfields within RFs predicted by encoding models

We estimated the spatial distribution of excitatory and suppressive subfields within the RF of each neuron by virtually presenting dynamic white noise patterns to its encoding model (see Materials and Methods). Note that the excitatory and suppressive subfields mentioned here are not the ON and OFF subfields of simple-cells. This is because our encoding models incorporate non-linear filters. These subfields correspond to the classical RF and surrounding suppressive field, respectively (Hubel and Wiesel, 1968; Walker et al., 1999; Cavanaugh et al., 2002a, b; Tanaka and Ohzawa, 2009; Hallum and Movshon, 2014). In the imaged region of 315 μm × 315 μm shown in Fig. 8A, most neurons had an oval excitatory subfield (red) surrounded by suppressive subfields (blue). The absolute peak magnitude of the suppressive subfields was 31% smaller than that of the excitatory subfields (median: p < 10^-5, Wilcoxon signed-rank test).

The excitatory subfields were similar in shape and position for most neurons within this local region. The median of the two-dimensional spatial correlation coefficients between the excitatory subfields of any two neurons was 0.90 in the imaged region (1378 pairs, Fig. 8B, red). For all pairs of neurons from the 19 imaged regions, the median correlation coefficient was 0.71 (136,600 pairs, Fig. 8C, red). We next examined the centroid positions of the excitatory subfields for the individual neurons. Each RF was first divided into 7 × 7 compartments. The centroid was determined for the compartments in which the responses were higher than the median magnitude of excitatory responses. For each imaged site, the locations of the excitatory subfield centroids were similar across neurons: the median distance between them was 0.08–0.30°. We also examined whether the position of the centroid progressively changed with the tangential position of the neuron over the cortex. We mapped x or y positions of the centroids across the imaged region (Fig. 9), and fit the x- or y-position map with a 2-dimensional plane. However, most imaged regions lacked linear progression of the centroid; the variance explained by the fitted plane was less than 30% in all but the largest imaged region (630 μm × 630 μm), in which the explained variance was 31.1%. The cortical magnification factor (Talbot and Marshall, 1941; Daniel and Whitteridge, 1961; Nauhaus et al., 2016) was 8.1 mm²/degree² for this region (Fig. 9).

In stark contrast, suppressive subfields differed substantially between

---

Fig. 8. Excitatory and suppressive RF subfields. (A) Two-dimensional map of the subfields obtained using the fit model. Relative positions of the squares maintain the positions of the corresponding neurons in the imaged region. (B, C) Frequency distributions of the correlation coefficients between the spatial distribution of excitatory subfields (red) and suppressive subfields (blue) for a pair of neurons (B: data from the imaged region shown in A; C: population data from 19 imaging sites). (D, E) Frequency distributions for the distance of excitatory subfield (red) and suppressive subfield (blue) peaks between a pair of neurons (D: data from the imaged region shown in A; E: population data). Arrow heads: medians of the distributions.

Excitatory and suppressive subfields within RFs predicted by encoding models

---

Fig. 9. Retinotopic maps for excitatory subfields of neurons in an example imaged region. Dot color indicates the positions of excitatory subfields in the visual field. (A) Horizontal position. (B) Vertical position.
neurons in a local region. For example, neuron 1 in Fig. 8A had suppressive subfields in the left portion of its RF, whereas those for neuron 2 were in the upper portion. The median correlation coefficient between the suppressive subfields was 0.21 in the imaged region (Fig. 8B, blue) and −0.07 for all pairs of neurons (Fig. 8C, blue), which was substantially lower than that for the excitatory subfields (p < 10−5, Wilcoxon signed-rank test). We measured the variability of the estimated subfields using bootstrap resampling (a total of 10,800 training samples were re-sampled with replacement, and RFs were re-estimated; the entire procedure was repeated 31 times; Efron and Tibshirani, 1993). The median correlation coefficient between suppressive subfields within each neuron obtained from the bootstrap was 0.75, which was larger than that between suppressive subfields of different neurons (−0.07; Mann-Whitney U test, p < 10−5). Thus, the differences we observed between suppressive subfields of different neurons is genuine, and not a coincidence deriving from any unreliability in model estimation.

We next examined peak positions of excitatory and suppressive subfields. In the example region, peak positions of excitatory subfields across neurons were close to each other (median distance between peaks: 0°, i.e., the same grid; Fig. 8D, red), whereas suppressive subfields showed substantial variability in their peak positions (median distance between peaks: 2.7°; Fig. 8D, blue). For the entire population of neurons (19 imaging sites), the median peak distance between excitatory subfields was 0.7° (Fig. 8E, red), which was significantly smaller than that between suppressive subfields (median = 2.9°; Fig. 8E, blue; p < 10−5, Wilcoxon signed-rank test). The positions of excitatory subfields (i.e., classical RFs) were thus shared by nearby neurons, while those for suppressive subfields were not.

To examine whether the suppressive components of the fit models contributed to the representation of the visual inputs, we nullified the outputs of the motion-energy filters with negative weights. Note that the suppressive components of the models are not identical to the suppressive subfields in predicted RFs. The suppressive components are necessary for the suppressive subfields, but they can also contribute to the shaping of the excitory subfields. The models without negative-weight filters predicted the responses to the natural movies less accurately than the original models (mean correlation coefficients: 0.32 vs. 0.40, p < 10−5, Wilcoxon signed-rank test). This suggests that the suppressive components of the neuronal responses play a role in encoding the natural visual inputs.

Discussion

We applied an encoding-model analysis to two-photon imaging data of Ca2+ signals from V1 neurons in response to movies of natural scenes. The encoding models predicted individual neuronal responses to new stimuli, including natural movies and drifting gratings, with reasonable accuracy. The orientation maps deduced from the model responses were highly correlated with those determined experimentally, and predictions of spatial frequency tuning and a loose map for spatial frequency were consistent with those previously reported. By extending our analysis to the subfield organization of RFs, we revealed that nearby neurons had similar excitatory subfields, but different suppressive subfields. We thus demonstrate that combining the encoding-model analysis with natural movie stimuli and two-photon Ca2+ imaging is an efficient way to conduct extensive analyses of neuronal response properties and functional architecture.

Our motion-energy encoding model predicted neuronal responses to new stimuli with an average accuracy of r = 0.40 (Fig. 3D). This range of accuracy is similar to that reported for models of neuronal spiking activity in macaque V1 neurons (Willmore et al., 2010; r = 0.40) and in human V1 activity measured using fMRI (Nishimoto et al., 2011; r = 0.4–0.5). Given that V1 is one of the most quantitatively studied cortical areas and that the motion-energy model is a standard model for V1, our results could form a baseline accuracy level for future studies using a similar framework.

Despite our model’s success, the prediction accuracy for neuronal responses to natural movies could be improved. The prediction accuracy was limited primarily by the variation in responses to repeated presentations of the same stimulus (Fig. 3G). Fluorescence signals from neurons contain stimulus-unrelated components, such as changes in the internal state of the animal. For a minority of neurons that exhibit low prediction accuracy despite having high response consistency, longer data sampling for model construction and incorporating different types of response properties into our model might help improve the predictions. Our model already successfully simulates many properties of V1 neurons by tiling motion-energy filters in the visual field. For example, as shown in Fig. 8, the current model reproduced a form of surround suppression that can be described by a linear sum of non-linear stimulus energy in and around the classical RFs (Walker et al., 1999; Tanaka and Ohzawa, 2009). However, our current model is insensitive to higher-order selectivity that requires non-linear summation of stimulus energy (Anzai et al., 2007; Tang et al., 2018). Recently, Tang et al. (2018) demonstrated that V1 superficial layer neurons show higher-order, complex selectivity (e.g., strong curvature selectivity) that might not simply be explained by a linear sum of non-linear stimulus energy. Their results suggest that, for future studies, we could incorporate the higher-order selectivity into an encoding model to identify more precise roles of V1 neurons during natural vision. We also reasoned that adding direction and color selectivity to the model filters might improve predictions; however, a preliminary analysis indicated no clear change in prediction accuracy. Thus, more complex models with higher-order components (e.g., textures) might be needed to predict responses more accurately than our model.

Although prediction accuracies for natural movies and gratings were correlated with each other, the correlation was not very strong (Fig. 4G and H). For some neurons, the model predicted responses to natural movies accurately, but did not work well for gratings. This discrepancy has several possible explanations. The most likely is that orientation selectivity is not the only factor driving the responses of these neurons to natural movies. In addition to orientation, natural movies contain numerous stimulus attributes including luminance, spatial frequency, and texture. If neuronal responses depend more strongly on one or several of these other parameters, our model might fail to capture orientation selectivity despite being able to predict the responses to natural movies fairly well. Another theory is that response properties of V1 neurons change depending on whether they are exposed to gratings or natural movies (David et al., 2004). The models obtained using natural stimuli might predict responses better for natural stimuli than artificial (geometrical) stimuli (Talebi and Baker, 2012). Inherent qualities of natural stimuli might evoke nonlinear responses that are different from those evoked by gratings. A third explanation for this discrepancy is that, as our analysis suggests, the precision of preferred-orienation estimation is limited in vivo. This can occur when the orientation tunings in labeled neurons are weak (Fig. 4I and J) or fluorescence responses are unstable.

By using the encoding models, we were able to identify functional architecture in the structure of RF subfields (Fig. 8). Because the subfields were measured using dynamic random dot stimuli with white, gray, or black brightness, they were the nonlinear subfields that complex cells often have. The excitatory subfields of individual neurons were round in shape and surrounded by suppressive subfields. This structure corresponds to classical RFs surrounded by suppressive fields. The suppressive fields were asymmetric, which has been suggested to be useful for responding to higher-order visual features such as oriented contours and textures (Walker et al., 1999; Tanaka and Ohzawa, 2009 for area 17 of cats; Hallum and Movshon, 2014 for V1 of monkeys). The positions of classical RFs were shared by nearby neurons, but those for the suppressive subfields were not, suggesting that V1 neurons are not systematically arranged for higher-order visual features. This result has no support by direct characterization of subfields at this moment, and should be tested using visual stimuli optimized to examine them (i.e., small patches of random dots and gratings).

Two-photon Ca2+ imaging allows simultaneous recording of visual
responses from multiple neurons. Because visual cortical neurons often have RFs with different positions and sizes, examining their response properties is difficult when one must first optimize the stimulus parameters for each neuron. To avoid this, previous studies used periodic stimuli that cover the full visual field (Ohki et al., 2005; Ikezoe et al., 2013). However, this often weakens neuronal responses because of surround suppression (Knierim and van Essen, 1992; Cavanaugh et al., 2002a). Indeed, our results showed suppressive subfields surrounding excitatory subfields (Fig. 8A). Spike-triggered averages of responses to white noise stimuli have been successfully applied to two-photon imaging data in efforts to analyze RF structure in mouse V1 (Smith and Häusser, 2010; Bonin et al., 2011). However, this technique reveals the linear kernel of responses and is useful only for neurons with simple-cell like response properties. In layers 2 and 3 of monkey V1, half of neurons show complex-cell like responses (Ringach et al., 2002) that are non-linear and thus cannot be analyzed with this technique. In contrast, encoding-model analysis consists of motion-energy filters that enable analysis of non-linear responses (Nishimoto and Gallant, 2011), as we determined the response properties of almost all visually responsive V1 neurons.

Encoding-model analysis can also be applied to neurons in visual areas beyond V1. Because neurons in downstream visual regions such as the inferior temporal cortex have selectivity for complex visual features, simple artificial stimuli such as gratings and random dots do not span the relevant feature space efficiently (Fujita, 2002; Felsen and Dan, 2005; Kourtzi and Connor, 2011; DiCarlo et al., 2012). Because natural movies contain varying types of visual features, including complex shapes, textures, colors, motion, objects, and faces, they have a better chance of evoking responses from these higher-order neurons. Once we obtain responses to the movies, post hoc analysis using the encoding-model approach enables us to determine stimulus selectivity for different visual features. With this approach, we can use the same set of stimuli to examine responses in different visual areas, facilitating systematic investigations into how sensory representation is transformed along the visual processing stream. In human fMRI studies, this encoding model approach was combined with presentation of natural movies to successfully reveal higher-order functional properties in lower to higher visual areas (Huth et al., 2012; Çukur et al., 2016). The number of video frames used in these fMRI studies was about double what we used in the present study. Although imaging fluorescence for 2 h using a chemical fluorescence dye might be difficult, genetically encoded calcium indicators (GECIs) allow a longer period of imaging. Combining our approach with GECIs will thus be helpful for characterizing response properties of higher-order visual neurons in more detail.

Finally, we note that our approach can help interpret the results from fMRI studies. fMRI allows the highest spatial resolution for measuring non-invasive brain activity in humans. However, even using high-field fMRI at 1 x 1 x 1 mm voxel resolution (Yacoub et al., 2008), one voxel volume contains at least 80,000-320,000 neurons (Christensen et al., 2007; Collins et al., 2010; Insel et al., 2013). Two-photon Ca²⁺ imaging reveals much finer temporal and spatial representations that are not resolvable via fMRI. Additionally, the relationship between individual neuronal activity and the blood-oxygen level dependent (BOLD) fMRI signal remains controversial (Logothetis, 2008; Boynton, 2011), whereas no such controversy surrounds calcium imaging (Stosiek et al., 2003; Grienberger and Konnerth, 2012; Ikezoe et al., 2013). Understanding the functional organization within a small volume of cortex might allow more informed interpretation of fMRI signals.

Conclusions

We combined two-photon Ca²⁺ imaging, an encoding model, and natural movie presentation for characterization of RF properties of individual neurons and the functional architecture in monkey V1. We achieved a prediction accuracy of the encoding model for Ca²⁺ signals comparable to that previously reported for single unit responses and fMRI signals. The results provide a baseline accuracy for future work with a similar approach. Furthermore, our analysis predicts a new aspect of the functional architecture of V1, i.e., clustering of neurons based on excitatory and inhibitory subfields of their RFs. This combination of the techniques is also applicable to fine-scale functional mapping of higher visual regions. Further improvement of the prediction accuracy of the encoding model will provide a promising approach for understanding visual information representation in the brain.

Data availability

Visual stimuli, analysis code for model construction, and a part of fluorescence data used in this report can be found online at several websites (visual stimuli: https://crcns.org/data-sets/vc/vim-2/about-vim-2; analysis code: https://github.com/gallantlab/motion_energy_matlab; fluorescence data: https://visiome.neuroinf.jp/modules/xoonips/detail.php?item_id=7261&m_lang=en).

Funding sources

This work was supported by the Japan Society for the Promotion of Science/Japanese Ministry of Education, Culture, Sports, Science and Technology [KAKENHI grant numbers JP15H01437, JP16H01673, and JP17H01381 to I.F., JP15H05311 to S.N., and JP25830012 and JP15K18357 to K.I.]; the National Institute of Information and Communications Technology; and the Ministry of Internal Affairs and Communications.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We thank Ryoosuke Takeuchi, Takanori Fukazawa, Yusuke Saito, and Kohei Iino for technical help. We thank Adam Phillips, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

Adelson, E.H., Bergen, J.R., 1985. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A 2, 284–299.
Anzai, A., Peng, X., van Essen, D.C., 2007. Neurons in monkey visual area V2 encode combinations of orientations. Nat. Neurosci. 10, 1313–1321.
Bonin, V., Histed, M.H., Yurgenson, S., Reid, R.C., 2011. Local diversity and fine-scale organization of receptive fields in mouse visual cortex. J. Neurosci. 31, 18506–18521.
Boynton, G.M., 2011. Spikes, BOLD, attention, and awareness: a comparison of electrophysiological and fMRI signals in V1. J. Vis. 11, 12.
Bredfeldt, C.E., Ringach, D.L., 2002. Dynamics of spatial frequency tuning in macaque V1. J. Neurosci. 22, 1976–1984.
Cavanaugh, J.R., Bair, W., Movshon, J.A., 2002a. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J. Neurophysiol. 88, 2530–2546.
Cavanaugh, J.R., Bair, W., Movshon, J.A., 2002b. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J. Neurophysiol. 88, 2547–2556.
Christensen, J.R., Larsen, K.R., Lisabry, S.H., Scalia, J., Arango, V., Dworzak, A.J., Pakkenberg, B., 2007. Neocortical and hippocampal neuron and glial cell numbers in the rhesus monkey. Anat. Rec. (Hoboken) 290, 330–340.
Collins, C.E., Airey, D.C., Young, N.A., Leitch, D.B., Kaas, J.H., 2010. Neuron densities vary across and within cortical areas in primates. Proc. Natl. Acad. Sci. U. S. A. 107, 15927–15932.
Cotton, R.J., Foudarakis, E., Storer, P., Saggau, P., Tolias, A.S., 2013. Three-dimensional mapping of microcircuit correlation structure. Front. Neural Circuits 7, 151.
Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2016. Functional subdomains within scene-selective cortex: parahippocampal place area, retrosplenial complex, and occipital place area. J. Neurosci. 36, 10257–10273.
Çukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770.
Dana, H., Mohar, B., Sun, Y., Narayan, S., Gordus, A., Hasseman, J.P., Tsegaye, G., Holt, G.T., Hu, A., Walpita, D., Patel, R., Macklin, J.J., Bargmann, C.I., Ahrens, M.B., Schreiter, E.R., Jayaraman, V., Looger, L.L., Svoboda, K., Kim, D.S., 2016. Sensitive red protein calcium indicators for imaging neural activity. eLife 5, e12727.
Nishimoto, S., Galant, J.L., 2011. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J. Neurosci. 31, 4551–4564.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Galant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21, 1641–1646.

Ohki, K., Chang, S., Ch'ng, Y.H., Kira, P., Reid, R.C., 2005. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature 433, 597–603.

Ohki, K., Chang, S., Kira, P., Hübener, M., Bonhoeffer, T., Reid, R.C., 2006. Highly ordered arrangement of single neurons in orientation pinwheels. Nature 442, 925–928.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., Simoncelli, E.P., 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999.

Pologruto, T.A., Sahabini, B.L., Svoboda, K., 2003. ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online 2, 13.

Poplack, S.J., Kohn, D.F., 1997. Anesthesia and analgesia in nonhuman primates. In: Kohn, D.F., Wixson, S.K., White, W.J., Benson, G.J. (Eds.), Anesthesia and Analgesia in Laboratory Animals. Academic Press, San Diego, pp. 233–255.

Ringach, D.L., Shapley, R.M., Hawken, M.J., 2002. Orientation selectivity in macaque V1: diversity and laminar dependence. J. Neurosci. 22, 5639–5651.

Sadakane, O., Masamizu, Y., Watakabe, A., Terada, S., Ohtsuka, M., Takaji, K., Mizukami, H., Ozawa, K., Kawasuki, H., Matsuzaki, M., Yamamoto, T., 2015. Long-term two-photon calcium imaging of neuronal populations with subcellular resolution in adult non-human primates. Cell Rep. 13, 1989–1999.

Sahani, M., Linden, J.F., 2003. How linear are auditory cortical responses?. In: Becker, S., Thrun, S., Obermayer, K. (Eds.), Advances in Neural Information Processing Systems, vol. 15. MIT Press, Cambridge, pp. 109–116.

Silverman, M.S., Grosof, D.H., De Volois, R.L., Elfar, S.D., 1989. Spatial-frequency organization in primate striate cortex. Proc. Natl. Acad. Sci. U.S.A. 86, 711–715.

Smeters, D., Majewska, A., Yuste, R., 1999. Detecting action potentials in neuronal fields and functional architecture of monkey striate cortex. J. Physiol. 525, 215–243.

Smith, S.L., Häusser, M., 2010. Parallel processing of visual space by neighboring neurons in mouse visual cortex. Nat. Neurosci. 13, 1144–1149.

Sofroniew, N.J., Flickinger, D., King, J., Svoboda, K., 2016. A large field of view two-photon microscope with subcellular resolution for in vivo imaging. eLife 5, e14742.

Stirman, J.N., Smith, I.T., Kudrnov, M.W., Smith, S.L., 2016. Wide-field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. Nat. Biotechnol. 34, 857–862.

Storey, C., Garachou, O., Hollobuff, K., Konnerath, A., 2003. In vivo two-photon calcium imaging of neuronal networks. Proc. Natl. Acad. Sci. U.S.A. 100, 7319–7324.

Szalay, G., Judik, L., Katona, G., Ocsai, K., Juhasz, G., Veress, M., Szadai, Z., Feher, A., Tompa, T., Chiervini, B., Maak, P., Rozsa, B., 2016. Fast 3D imaging of spine, dendritic, and neuronal assemblies in behaving animals. Nature 539, 732–738.

Talbot, S.A., Marshall, W.H., 1941. Physiological studies on neural mechanisms of visual localization and discrimination. Am. J. Ophthalmol. 24, 1255–1264.

Talebi, V., Baker, C.L., 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. J. Neurosci. 32, 1560–1576.

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

Tompa, T., Chiovini, B., Maak, P., Rozsa, B., 2016. Fast 3D imaging of spine, dendritic, and neuronal assemblies in behaving animals. Nature 539, 732–738.

Talebi, V., Baker, C.L., 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. J. Neurosci. 32, 1560–1576.

Talebi, V., Baker, C.L., 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. J. Neurosci. 32, 1560–1576.

Vinje, W.E., Gallant, J.L., 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287, 1273–1276.

Walker, G.A., Ohzawa, I., Freeman, R.D., 1999. Asymmetric suppression outside the classical receptive field of the visual cortex. J. Neurosci. 19, 10536–10553.

Watson, A.B., Ahumada, A.J., 1985. Model of human visual-motion sensing. J. Opt. Soc. Am. A 2, 322–341.

Willmore, B.D., Frueger, R.J., Gallant, J.L., 2010. Neural representation of natural images in visual area V2. J. Neurosci. 30, 2102–2114.

Wu, M.C., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. 29, 477–505.

Yacoub, E., Harel, N., Ugurbil, K., 2008. High-field fMRI unveils orientation columns in humans. Proc. Natl. Acad. Sci. U.S.A. 105, 10607–10612.

Yamins, D.L., Hong, H., Cadiou, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U.S.A. 111, 8619–8624.