Inverted Encoding Models of Human Population Response Conflate Noise and Neural Tuning Width

Taosheng Liu,1 Dylan Cable,2 and Justin L. Gardner2

1Department of Psychology, Michigan State University, East Lansing, Michigan 48824, and 2Department of Psychology, Stanford University, Stanford, California 94305

Channel-encoding models offer the ability to bridge different scales of neuronal measurement by interpreting population responses, typically measured with BOLD imaging in humans, as linear sums of groups of neurons (channels) tuned for visual stimulus properties. Inverting these models to form predicted channel responses from population measurements in humans seemingly offers the potential to infer neuronal tuning properties. Here, we test the ability to make inferences about neural tuning width from inverted encoding models. We examined contrast invariance of orientation selectivity in human V1 (both sexes) and found that inverting the encoding model resulted in channel response functions that became broader with lower contrast, thus apparently violating contrast invariance. Simulations showed that this broadening could be explained by contrast-invariant single-unit tuning with the measured decrease in response amplitude at lower contrast. The decrease in response lowers the signal-to-noise ratio of population responses that results in poorer population representation of orientation. Simulations further showed that increasing signal to noise makes channel response functions less sensitive to underlying neural tuning width, and in the limit of zero noise will reconstruct the channel function assumed by the model regardless of the bandwidth of single units. We conclude that our data are consistent with contrast-invariant orientation tuning in human V1. More generally, our results demonstrate that population selectivity measures obtained by encoding models can deviate substantially from the behavior of single units because they conflate neural tuning width and noise and are therefore better used to estimate the uncertainty of decoded stimulus properties.

Key words: contrast; encoding model; orientation; population response

Significance Statement

It is widely recognized that perceptual experience arises from large populations of neurons, rather than a few single units. Yet, much theory and experiment have examined links between single units and perception. Encoding models offer a way to bridge this gap by explicitly interpreting population activity as the aggregate response of many single neurons with known tuning properties. Here we use this approach to examine contrast-invariant orientation tuning of human V1. We show with experiment and modeling that due to lower signal to noise, contrast-invariant orientation tuning of single units manifests in population response functions that broaden at lower contrast, rather than remain contrast-invariant. These results highlight the need for explicit quantitative modeling when making a reverse inference from population response profiles to single-unit responses.

Introduction

Bridging knowledge derived from measurements at different spatial and temporal scales is a significant challenge for understand-

The authors declare no competing financial interests.

Correspondence should be addressed to either of the following: Dr. Taosheng Liu, Department of Psychology, Michigan State University, 316 Physics Road, East Lansing, MI 48824, E-mail: tslui@msu.edu; or Dr. Justin L. Gardner, Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, E-mail: jlg@stanford.edu.

Received Aug. 25, 2017; revised Nov. 8, 2017; accepted Nov. 10, 2017.

Author contributions: T.L. and J.L.G. designed research; T.L. and J.L.G. performed research; T.L., D.C., and J.L.G. contributed unpublished reagents/analytic tools; T.L., D.C., and J.L.G. analyzed data; T.L. and J.L.G. wrote the paper.

This work was supported by RIKEN Brain Science Institute, Stanford’s BioX Undergraduate Summer Research Program, and VPUE Faculty Grant for Undergraduate Research to D.C., National Institutes of Health Grant R01 EY022727 to T.L., and Grants-in-Aid for Scientific Research 24300146, Japanese Ministry of Education, Culture, Sports, Science and Technology, and the Hellman Faculty Scholar Fund to J.L.G. We thank Kang Cheng (1962–2016) for his spirit of rigorous scientific inquiry and unfailing personal encouragement that has been, and continues to be, a guiding force of our research; Kenji Haruhana and members of the Support Unit for Functional Magnetic Resonance Imaging for assistance in data collection; Toshiko Ikari for administrative assistance; and John Serences for helpful comments on an earlier version of the manuscript.

The Journal of Neuroscience, Month XX, 2017• 37(XX): XXXX–XXXX • 1

DOI:10.1523/JNEUROSCI.2453-17.2017

Copyright © 2017 the authors 0270-6474/17/370001-11$15.00/0
Invasive animal experiments are difficult to verify and replicate in humans. It is therefore often unknown whether basic phenomena from the single-unit literature are applicable to humans, let alone how these phenomena will manifest at the larger scale of population activity that is typically interrogated by noninvasive measurement of the human brain.

Recently, an encoding model approach has proven useful in the analysis of large-scale population activity measured by functional imaging (Naselaris et al., 2011; Serences and Saproo, 2012) and offers the promise of bridging knowledge from different species and scales of measurements.Encoding models are built on fundamental results in visual physiology, by encoding complex stimuli in lower dimensional representations, such as receptive field or channel models. The assumption is that, if these neural representations are operative in human cortex, then large-scale measurements of activity represent the aggregated responses of these basic neural operations. For example, a channel-encoding model (Brouwer and Heeger, 2009, 2013) has been used to examine continuous stimulus dimensions, such as color or orientation, where it is reasonable to expect that there are large groups of neurons, or channels with known selectivity, and that voxel responses can be modeled as linear combinations of such channels. These channel-encoding models have been used to examine responses for orientation, color, direction, and speed of motion and somatosensory response to better understand apparent motion (Chong et al., 2016), cross-orientation suppression (Brouwer and Heeger, 2011), normalization (Brouwer et al., 2015), speeded decision making (Ho et al., 2012), attention (Scolari et al., 2012; Garcia et al., 2013; Saproo and Serences, 2014; Ester et al., 2016), working memory (Ester et al., 2013, 2015), perceptual learning (Byers and Serences, 2014; Chen et al., 2015), biases in motion perception (Vintch and Gardner, 2014), and exercise (Bullock et al., 2017) using both functional imaging and EEG (Garcia et al., 2013; Bullock et al., 2017) measurements. Inverting these models to form predictions of channel response from cortical measurements produces tuned response profiles. The interpretations of these tuned response profiles are encouraging for the effort of bridging across measurements as they have shown results in concordance with expectations from electrophysiological measurement of phenomena, such as decision-making reliance on off-target populations (Purushothaman and Bradley, 2005; Scolari et al., 2012) and feature-similarity gain (Treue and Maunsell, 1996; Saproo and Serences, 2014) and response gain (McAdams and Maunsell, 1999; Garcia et al., 2013) modulation effects of attention.

Here we test the ability of the channel-encoding model approach to bridge single-unit and population-scale measurement by asking whether the well-known property of contrast-invariant orientation tuning is manifest in predicted channel responses from human primary visual cortex. We reasoned that examining whether orientation tuning bandwidth of human cortical population responses change with contrast would provide a good test case for the use of encoding models to bridge measurements because there is a clear prediction of invariance from single-unit measurements (Sclar and Freeman, 1982). However, contrary to the electrophysiology literature, we found that an encoding model produced channel response functions that increased in bandwidth as contrast was lowered. Computational modeling revealed that these effects can be explained by the reduced signal-to-noise ratio (SNR) of cortical responses at lower contrast. These results emphasize that bridging different levels of measurement through these analyses requires explicit quantitative statements of how properties of single units are expected to manifest in population activity.

Materials and Methods

Subjects. Six healthy volunteers (ages 33–42, two female) from the RIKEN Brain Science Institute community participated in the experiment; all had normal or corrected-to-normal vision and were experienced subjects in functional imaging experiments. The study protocol was approved by the RIKEN Functional MRI Safety and Ethics Committee, and all subjects gave written consent to experimental procedures in advance of participating in the experiment.

Stimuli. Stimuli were generated using MGL, a set of MATLAB (The MathWorks) routines for implementing psychophysical experiments (http://gru.stanford.edu/mgl). Stimuli were back projected onto a screen using an LCD projector (Silent Vision 6011; Avotec) at a resolution of 800 × 600 and a refresh rate of 60 Hz. Subjects viewed the screen via an angled mirror attached to the head coil. The projector was gamma corrected to achieve a linear luminance output.

Visual stimuli were sinusoidal gratings (spatial frequency: 0.7 cpd) in a circular aperture (10°), located to the left or the right of a central fixation point. All analyses were performed in the original (nontransformed) coordinate system.

Task and procedures. On each trial, two gratings were presented for 5.12 s, followed by a 3.84 s intertrial interval. During the gratings presentation, the phases of both gratings were updated every 0.13 s. The phase of each grating was randomly chosen from 1 of 16 uniformly distributed phases from 0 to 2π, and the starting time of the phase update of each grating was randomly determined such that the phase updates of the two gratings were asynchronous. The phase updates were implemented to reduce retinal adaptation and afterimages. The contrast and orientation of each grating were randomly chosen on each trial such that each combination of contrast (two levels) and orientation (eight levels) was presented three times in each run (48 trials in total). Although the intertrial interval was short, which could result in nonlinear summation of responses from the previous trial (Boynton et al., 1996), the trial randomization procedure served to minimize previous trial effects as, on average, they would come from a random trial type. In addition, a fixation period of 5.12 s preceded each run, making each run 435.2 s in the scanner. Subjects completed 9 runs in the scanner (432 trials in total), which yielded 27 trials per orientation/contrast combination.

While the gratings were presented in the periphery, subjects performed a luminance discrimination task at fixation. On each trial in this task, the fixation cross dimmed for 0.4 s twice, separated by a 0.8 s interval, and subjects had to indicate in which interval the cross appeared darker. The magnitude of dimming was held constant for one interval while the magnitude of dimming in the other interval was controlled by a one-up two-down staircase. Subjects pressed one of two keys (1 or 2) to indicate their response. The fixation task was performed continuously throughout a run and was asynchronous with the display of the grating stimuli. This task was used to control subjects’ attention and ensure a steady behavioral state and eye fixation. The independently randomized contrast and orientations of the two gratings on either side also served as an internal check of the fixation quality, as any systematic bias of eye position for one stimulus would not be systematic for the other.

MRI methods. Imaging was performed with a Varian Unity Inova 4T whole-body MRI system (now Agilent Technologies) located at the RIKEN Brain Science Institute (Saitama, Japan). A volume RF coil (transmit) and a four-channel receive array (Nova Medical) were used to acquire both functional and anatomical images.

Each subject first participated in a separate scanning session to obtain their retinotopic maps (see below for more details), using standard procedures. During this session, a high-resolution 3D anatomical T1-weighted volume (MPRAGE; TR, 13 ms; TI, 500 ms; TE, 7 ms; flip angle, 11°; voxel size, 1 × 1 × 1 mm; matrix, 256 × 256 × 180) was obtained, which served as the reference volume to align all functional images. The reference volume was segmented to generate cortical surfaces using Free-surfacer (Dale et al., 1999). Subsequently, the anatomy volumes taken at the beginning of each session were registered to the reference volume so that the cortical regions in the functional scans were aligned with the retinotopy. All analyses were performed in the original (nontransformed) co-
ordinates before being mapped to the cortical surface and specific visual regions.

During the main experiment, functional images were collected using a T2*-weighted EPI sequence (TR, 1.28 s; TE, 25 ms; flip angle, 45°; sensitivity encoding with acceleration factor of 2). We collected 29 slices at an angle approximately perpendicular to the calcarine sulcus, with resolution of 3 × 3 × 3 mm (FOV, 19.2 × 19.2 cm; matrix size, 64 × 64). The first four volumes in each run were discarded to allow T1 magnetization to reach steady state. In addition, a T1-weighted (MPRAGE; TR, 11 ms; TI, 500 ms; TE, 6 ms; flip angle, 11°; voxel size, 3 × 3 × 3 mm; matrix, 64 × 64 × 64) anatomical image was acquired to be used for coregistration with the high-resolution reference volume collected in the retinotopic session.

Various measures were taken to reduce artifacts in functional images. During scanning, respiration was recorded with a pressure sensor, and heartbeat was recorded with a pulse oximeter. These signals were used to attenuate physiological signals in the imaging time series using retrospective estimation and correction in k space (Hu et al., 1995).

Retinotopic mapping procedure. In this separate scanning session, we mapped each subject’s occipital visual areas using well-established phase-encoding methods (Sereno et al., 1995; DeYoe et al., 1996; Engel et al., 1997), so only a brief description is provided here. We presented rotating wedges and expanding/contracting rings over multiple runs and averaged runs of the same type. Then a Fourier analysis was applied to the averaged time course to derive the polar angle map and eccentricity map from the wedge and ring data, respectively. Borders between visual areas were defined as phase reversals in the polar angle map of the visual field. The first map was visualized on computationally flattened representations of the cortical surface generated by FreeSurfer. For each subject, we could readily define many visual areas, including V1, V2, V3, and hV4. However, we will mainly focus on V1 in this study.

BOLD data analysis. Data were processed and analyzed using mrTools (http://gru.stanford.edu/mrTools) and other custom code in MATLAB. Preprocessing of functional data included head movement correction, linear detrend, and temporal high-pass filtering at 0.01 Hz. The functional images were then aligned to high-resolution anatomical images for each participant, using an automated robust image registration algorithm (Nestares and Heeger, 2000). Functional data were converted to percentage signal change by dividing the time course of each voxel by its mean signal over a run, and data from the 9 scanning runs were concatenated for subsequent analysis.

Voxel selection. We used an event-related (finite impulse response or deconvolution) analysis to select voxels in V1 that responded to the stimulus presentation. Each voxel’s time series was fitted with a GLM with regressors for 16 conditions (2 contrasts × 8 orientations) that modeled the BOLD response in a 25 s window after trial onset. The design matrix was then pseudo-inversed and multiplied by the time series to obtain an estimate of the response amplitude for each individual trial in each voxel. We call the set of response amplitudes across all voxels in a V1 hemisphere a response “instance.”

We fit the encoding model to the instances with a fivefold cross-validation scheme, in which four-fifths of the trials were randomly selected to be the training data and the remaining one-fifth of trials constituted the test data. This analysis was performed on instances for the high-contrast and low-contrast trials separately (216 trials per contrast). Our encoding model consisted of 8 evenly spaced channels (i.e., model basis functions) from 0° to 180°, with each channel a half-wave rectified sinusoid raised to the power of 7. These basis functions were chosen to approximate single-neuron’s orientation tuning function in V1. In the following exposition, we adopted the notation from Brouwer and Heeger (2009). The training instances can be expressed as a m × n matrix B1, where m is the number of voxels and n is number of trials in the training data. We then constructed hypothetical channel outputs given the stimulus orientation on each trial of the training dataset, which yielded a k × n matrix C1, where k is the number of channels (i.e., k = 8). Each column in the C1 matrix represented a set of ideal response to the stimulus orientation on that trial from the eight channels. A weight matrix W (m × k) relates the observed data B and the hypothetical responses as follows:

\[ B_k = W C_k \]  

Each row of W represents the relative contribution of the eight channels to that voxel’s response. The least-square estimate of W was obtained with the following equation (T indicates the transpose of the matrix):

\[ \hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1} \]

The test instances can be expressed as a m × p matrix B2, where p is the number of trials in the test data. The estimated channel response to each test stimulus (Ĉk) can then be estimated using the weights W as follows:

\[ \hat{C}_k = (W^T W)^{-1} W^T B_2 \]

The test instances can be expressed as a m × p matrix B2, where p is the number of trials in the test data. The estimated channel response to each test stimulus (Ĉk) can then be estimated using the weights W as follows:
participant, we obtained channel response functions for both the contralateral and ipsilateral stimulus, separately for the high-contrast and low-contrast conditions.

In addition to the channel response function, we also calculated a goodness-of-fit measure of the encoding model. In each cross-validation fold, after we obtained estimates of the channel weights \( W \), we also constructed another set of hypothetical channel outputs given the actual stimulus orientation in the test data, \( C_2 \) (this is a matrix similar to \( C_0 \), but for the test trials, and differs from \( C_0 \), which is the estimated channel responses based on the voxel responses). The predicted voxel response can be obtained via the following:

\[
B_{\text{pred}} = W C_2
\]  

(4)

To the extent that the encoding model provides a good fit to the data, the predicted response \( B_{\text{pred}} \) should be similar to the observed response, \( B_r \). We can thus calculate the amount of variance explained by the model as follows:

\[
r^2 = 1 - \frac{\sum (B_{\text{pred}} - B_r)^2}{\sum (B_r - \bar{B})^2}
\]  

(5)

Here, \( \bar{B} \) is the mean voxel responses across voxels and trials, and the summation was performed across voxels and trials (i.e., all values in the \( m \times p \) matrix). We calculated \( r^2 \) from each fold of the cross validation and averaged them across the folds to obtain a single measure of the goodness of fit, for each contrast. This \( r^2 \) is different from the \( r^2 \) that represents the goodness of fit of the event-related model (see Voxel selection). In the remainder of this report, we will focus on this \( r^2 \) value that indexes the goodness of fit of the channel-encoding model.

Quantifying the channel response function. We fitted a circular bell-shaped function (von Mises) to the channel response function as follows:

\[
y = y_0 + Ae^{\kappa \cos(x - \mu)}
\]  

(6)

Where \( \kappa \) is the concentration parameter that controls the width of the function, \( \mu \) is the mean, \( x \) is the orientation, \( A \) is the amplitude parameter, and \( y_0 \) is the baseline. Thus, there were four free parameters of the fit: baseline \( (y_0) \), amplitude \( (A) \), mean \( (\mu) \), and concentration \( (\kappa) \). Because orientation is on \([0, \pi]\), whereas the von Mises function spans the interval \([0, 2\pi]\), orientation values were multiplied by 2 during the fit, after which the fitting results were scaled back to \([0, \pi]\). Fitting was performed using a nonlinear least square method, as implemented in MATLAB. To ease the interpretation of the results, we report the half-width at half-maximum instead of the concentration parameter \( \kappa \) because the latter is inversely related to the variance.

Linking neuronal tuning to channel response: a computational model. We implemented a computational model that links neuronal tuning to channel response functions, using the assumptions underlying the channel-encoding model. Given that the channel response function is a highly derived statistic, this modeling effort was used to clarify how various assumptions of neural tuning and signal to noise would manifest in channel response functions. Specifically, we used this model to fit our observed data and simulate other scenarios to test the validity of the encoding model. The schematic of the model is outlined in Figure 3. The model contains 100 V1 voxels, comparable with the actual number of voxels used in our data analysis (see above). Each voxel is assumed to contain neurons tuned to all orientations, whose tuning functions are described by von Mises functions (see above). The preferred orientations \((\mu)\) are evenly distributed across all possible orientations in 1° increments (left column, “Neural tuning functions”), forming 180 classes of neurons. The width of the neural tuning function is specified by the concentration parameter, \( \kappa \), which is the same for all neurons but can be manipulated across different simulations. The area under the orientation tuning function for neurons was normalized to 1 so that the average firing rate for each neuron would not vary with tuning width. For each voxel, a weight vector, \( w \), was generated by randomly sampling 180 numbers from \([0, 1]\). This weight vector specifies how much each class of neurons contributes to the voxel’s response. For an input stimulus with an orientation, \( \theta \), the response of each neuron is calculated according to its tuning function (see Eq. 6). For each voxel, the responses of individual neurons are multiplied by the weight vector, \( w \), and then summed to arrive at a predicted response (middle column, “Random weights”). To calculate the final voxel response, Gaussian noise, \( N(0, \sigma) \), is then added to this response to simulate physiological and thermal noise in BOLD measurements. Thus, each voxel’s response is determined by neuronal tuning width \( (\kappa) \), weight vector \( (w) \), and noise \( (\sigma) \). \( w \) is randomly generated for each voxel in each simulation, whereas \( \kappa \) and \( \sigma \) are parameters that we examined systematically in several simulations (see Results).

We simulated experiments with the same basic setup as our empirical study: 8 possible orientations, with each orientation shown 27 times (trials). For each trial, we obtained a vector of voxel responses (instances), calculated as above. Then all the trial instances were subjected to the same analysis as the real data described above (i.e., cross-validation in which four-fifths of data were used to obtain the channel weights and one-fifth of the data was used to obtain channel response functions). We used the same exact code to analyze the synthetic and real data.

Computation of the posterior distributions. We computed the probability of different stimulus values given the test data (i.e., the posterior distribution) using a technique from van Bergen et al. (2015). The method begins with finding the weights of the channel-encoding model as above. After removing the signal due to the encoding model, a noise model is fit to the residual response. The noise model assumes that each individual voxel’s variability is Gaussian with one component that is independent among voxels and another that is shared across all voxels.

Each of the channels is modeled to have independent, identically distributed Gaussian noise. This leads to a covariance matrix for the noise as follows:

\[
\Omega = \rho \sigma^2 I + (1 - \rho) \Theta \Sigma^2 + \sigma^2 W W^T
\]  

(7)

Where \( \Omega \) is the noise model’s covariance matrix, \( I \) is the identity matrix, \( \Theta \) denotes element-wise multiplication, \( \Sigma \) is a vector containing each voxel’s independent standard deviation, \( \rho \) is a scalar between 0 and 1, which controls the amount of shared variability among voxels, \( \sigma \) is the SD of each channel, and \( W \) is the estimated weight matrix from Equation 2. This noise model with parameters, \( \rho, \sigma, \) and \( \sigma \) is fit via maximum likelihood estimation to the residual and can be used to compute the probability of generating any particular response given a stimulus value. Inversion of this equation using Bayes’ rule and a flat prior allows one to compute the probability of any stimulus value given a response, the posterior distribution (for a derivation and detailed explanation, see van Bergen et al., 2015). All analysis followed the same fivefold cross-validation scheme used for the encoding model by which four-fifths of the data were used to fit the model weights and noise model parameters and the one-fifth left-out data were used to compute the posterior distribution using Bayes’ rule. Results are shown averaged across all five left-out folds.

Results

We measured BOLD responses from retinotopically defined V1 to oriented sinusoidal gratings (Fig. 1) and used the resulting data to train and test a channel-encoding model. We first report average (univariate) activity across voxels. Although responses averaged across subjects and voxels in V1 did not show any apparent selectivity for orientation (Fig. 2A), they were, as expected (Boyalton et al., 2012), greater for higher-contrast compared with low-contrast contralateral stimuli (compare red vs yellow). Consistent with the laterality of V1, responses did not vary with contrast of the ipsilateral stimulus (black). The presence of a strong contrast response for the contralateral stimuli, but a complete absence of such a response for ipsilateral stimuli, also suggests that subjects maintained stable central fixation.

Despite this lack of orientation selectivity, channel response functions obtained from the encoding model and averaged across subjects displayed peaked responses at the true orientation of the contralateral stimulus for both high-contrast (Fig. 2B, red) and low-contrast (Fig. 2C, yellow) stimuli. We computed \( r^2 \), a ma-
functions might imply broader neuronal tuning at low contrast in human visual cortex, we instead considered whether it might be flat for low-contrast stimuli. Below, we focus on results from V1, which was 14.2 degrees and channel response function was essentially lowered. Finally, in hV4, tuning width for high-contrast stimuli was 21.3 to 75.5 degrees and amplitude went from 0.25 to 0.19 as contrast was lowered. In V3, tuning width went from 17.8 to 88.6 degrees and amplitude went from 0.23 to 0.18 as contrast was decreased (achieved by varying the SD of the added noise; $p < 0.042$, $t_{(11)} = 2.60$, one-tailed paired $t$ test). Ten of 12 hemispheres had lower amplitude responses result in lower SNR of the measurements at low contrast could also result in flatter (i.e., broader) channel response functions.

To test whether reduced SNR at low contrast, rather than changes in neural tuning, could account for the increased channel response bandwidth at low contrast, we built simulations (Fig. 3; see Materials and Methods). Briefly, each simulated voxel received randomly weighted responses from simulated orientation-tuned neurons. Different voxels had different weightings of the neuronal responses, thus resulting in weak, but different, orientation selectivity across voxels. We added random Gaussian noise to these voxel responses and trained and tested the encoding model using the same procedures as we did for the actual BOLD data. We varied the SD of the added noise ($\sigma$) to produce channel response functions that best fit the empirical data (in the least-squares sense) from the high-contrast trials (Fig. 4A). We then noted that, in the empirical data, there was a 42.2% decrease in neural response from high to low contrast across voxels (Fig. 2A). We therefore decreased neural response by this amount for all the simulated neurons and found that the resulting channel response function to be a reasonable fit for the low-contrast data (Fig. 4B). Importantly, this good correspondence between model predictions and data was achieved without fitting any parameter because the only thing we changed in the simulation was to decrease the magnitude of response across all neurons according to the value found from the empirical data. This suggests that reductions in response magnitude, and therefore SNR, are sufficient to produce changes in channel response bandwidth commensurate with what we observed.

Although tuning width is not expected to change with contrast, what if we had tested a property for which we expect a neural tuning width change? Do channel response functions track changes in neural tuning width? We simulated neural tuning functions from 5 to 40 degrees half-width at half-height as well as a “stick function,” which responds only to a single orientation maximally and does not respond to any other orientation and computed channel response functions under different amounts of noise (Fig. 5A; cyan to magenta curves represent different neural tuning widths). We found that the resulting channel response functions did indeed track the neural tuning widths, but as the goodness-of-fit $r^2$ increased (achieved by varying the SD of the added noise; Fig. 5B, abscissa), the difference in channel response widths was diminished (Fig. 5B, larger splay of curves on the left vs right side). Thus, channel response functions can reflect underlying neural tuning widths; but perhaps counterintuitively, the better the goodness of fit of the encoding model, the less difference the neural tuning width makes.

To understand why better goodness of fit implies worse model discriminability of underlying neural tuning functions, it is important to note that the absolute width of channel response functions is, in the limit of no noise, determined by the basis functions used in the encoding model and not by the neural tuning widths.
themselves. We simulated two variations of the encoding model in which we varied the model basis function widths. We increased the basis function width by decreasing the exponent on the sinusoidal basis to 3 (Fig. 5A) and decreased the basis function width by using stick filters (Fig. 5C). For all the simulations, as the goodness of fit increases, the recovered channel response functions approach the width of the model basis function (dashed line) rather than the neuronal response width (for both the channel response functions and the model basis functions, we use the fitted half-width at half-height as our measure of tuning width; therefore, the stick functions do not have infinitely narrow tuning). Given that the encoding model is essentially a linear regression model, this is not an unexpected outcome. Linear weights are being determined to best map the voxel responses unto ideal channel responses. As long as the voxel responses are determined by stimulus orientation, then the regression model will be able to recreate exactly any model basis functions that can be formed as linear combinations of the represented orientations.

The above analysis suggests that, although absolute neural tuning width may not be readily determined from the channel response function width, changes in tuning width might be meaningful if SNR does not change between conditions. That is, reading vertically for one level of goodness of fit in Figure 5A–C, the channel response function width changes systematically as a function of neural tuning width. Might we be able to determine that neural tuning width has changed if we observe data in which we have matched goodness of fit of the encoding model? Although this does not occur for changes in contrast, this could be the case for potential modulation of tuning width by cognitive factors, such as attention and learning.

However, changes in neural tuning width also result in changes in the model’s goodness of fit, thus complicating the possibility of interpreting changes in channel response width. We simulated different neural tuning widths from 5 to 45 degrees half-width at half-height (Fig. 6, abscissa) for different amounts of additive noise (SD ranging on a logarithmic scale from 0.001 to 1, blue to yellow curves). As can be appreciated by the downward slopes of the curves in Figure 6, as the neural tuning widths get wider, there is less information about orientation available to fit the model. In the extreme, a flat neural tuning function would result in no orientation-specific response and the encoding model would fail to fit the data completely. Each curve in the simulation is what one might expect to measure if neural tuning width is the only variable that changes in the experiment and noise is due mostly to external factors that do not change with conditions. That is, if one expects only an increase in neural tuning width, the resulting channel response functions would be expected to have both wider tuning and lower SNR. However, changes in neural tuning width instead result in changes in the model’s goodness of fit, thus complicating the possibility of interpreting changes in channel response width. We simulated different neural tuning widths from 5 to 45 degrees half-width at half-height (Fig. 6, abscissa) for different amounts of additive noise (SD ranging on a logarithmic scale from 0.001 to 1, blue to yellow curves). As can be appreciated by the downward slopes of the curves in Figure 6, as the neural tuning widths get wider, there is less information about orientation available to fit the model. In the extreme, a flat neural tuning function would result in no orientation-specific response and the encoding model would fail to fit the data completely. Each curve in the simulation is what one might expect to measure if neural tuning width is the only variable that changes in the experiment and noise is due mostly to external factors that do not change with conditions. That is, if one expects only an increase in neural tuning width, the resulting channel response functions would be expected to have both wider tuning and lower SNR.

These empirical and simulation results suggest caution in interpreting changes in channel response functions because they could be due to either or both changes in neural tuning width and changes in signal strength. What better ways are there for interpreting these functions? A recent report (van Bergen et al., 2015) proposes to transform these channel response functions into posterior distributions that show the probability of different stimulus values given the measured response. We applied the same analysis to our channel response functions, by estimating the
distribution of noise in the voxels and channels to determine the probability of measuring responses given any oriented stimulus and then applying Bayes’ rule with a flat prior to obtain the probability of various stimuli given the responses we measured in a left-out validation dataset (see Materials and Methods; Fig. 7). This resulted in posterior distributions that were peaked around the actual orientation for the contralateral stimulus (red and yellow curves) but flat for the ipsilateral stimulus (black curves), thus replicating the results on channel responses. This transformation of the results into posterior distributions allows for a more straightforward interpretation of the encoding model approach, encouraging interpretation in terms of the certainty by which a given neural response tells us what the stimulus was, rather than what it implies about underlying neural tuning functions.

**Discussion**

Using an encoding model approach, we built channel response functions for orientation and found that, unlike contrast-invariant single units, they were broader at lower contrast. Simulations showed that this effect could be fully explained by the measured decrease in overall neural response between high and low contrast, which results in lower SNR. As signal to noise is increased, channel response functions become narrower, until, in the limit of no noise, they approximate the shape of the model basis function (and not the underlying neural tuning function). While changes in underlying neural selectivity in our model could be reflected in channel response functions, our results demonstrate that changes in channel response functions do not necessarily reflect changes in underlying neural selectivity.
Orientation selectivity of single-unit responses has been shown to be invariant to image contrast (Sclar and Freeman, 1982; Skottun et al., 1987; Carandini et al., 1997), suggestive of a general neural computational mechanism (Carandini and Heeger, 2011) by which visual perception can remain relatively unaffected by differences in visibility of stimuli. Despite this central theoretic importance, obtaining noninvasive measurement of selectivity bandwidth from human cortex has been technically difficult because orientation selectivity is organized into cortical columns (Hubel and Wiesel, 1962, 1968; Blasdel and Salama, 1986; Bonhoeffer and Grinvald, 1991), much smaller than the typical spatial resolution of BOLD (Ogawa et al., 1990, 1992) measurements. While direct measurements of such columnar structures in humans have been achieved (Cheng et al., 2001; Sun et al., 2007; Yacoub et al., 2007, 2008), multivariate analysis using pattern classification approach to decode orientation and motion direction (Haynes and Rees, 2005; Kamitani and Tong, 2005, 2006) from distributed activity patterns has become a more common approach (Norman et al., 2006). However, this classification approach generally produces a categorical outcome, for example, which of two orientations was more likely to have resulted in the measured response pattern and thus is not typically used for probing selectivity bandwidth of neural representations. The encoding model approach allows one to reconstruct a response profile for a stimulus that has a tuning bandwidth that can be inspected across different contrasts.

Although we found an increase in the bandwidth of channel response functions for lower-contrast stimuli from human V1, this increase could be fully accounted for by the measured reduction of response amplitude due to contrast, thus reconciling our data with contrast-invariant orientation tuning. We recognize that contrast invariance at the population level as measured with BOLD is not guaranteed, even if single-unit spiking responses display contrast invariance. Systematic relationships between contrast sensitivity and selectivity for orientation could result in population responses changing selectivity with contrast. For example, if the least orientation-selective neurons saturate their responses at lower contrast than the most selective neurons, then population response would become more selective as contrast increases because population response would be dominated by the most selective neurons. However, no such systematic relationship has been observed and population spiking responses appear contrast-invariant in cats (Busse et al., 2009), consistent with our results. Furthermore, BOLD measurements may be better correlated with local field potentials than spiking activity (Logothetis et al., 2001), which could also result in deviations of BOLD population measures of contrast invariance and spiking activity of neurons. If BOLD measures are sensitive to subthreshold, synaptic activity that can contribute to local field potentials, broadening of channel response functions that we observed could be reflective of subthreshold activity, if such activity is not contrast-invariant. However, intracellular measurements of membrane potentials show that selectivity does not broaden at lower contrast. Indeed, selectivity is slightly increased at low contrast (Finn et al., 2007), consistent with our interpretation that channel response function broadening at low contrast is due to reduction in signal to noise.

Given our results, bridging effects of attention on single units with effects uncovered using encoding models of functional imaging measurements (Sprague et al., 2015) may be similarly complicated as bridging contrast invariance effects. Single-unit studies have suggested that neurons change gain, not selectivity bandwidth (McAdams and Maunsell, 1999; David et al., 2008) with spatial attention, a key finding that has shaped our understanding of neural mechanisms of attention (Carrasco, 2011; Ling et al., 2015). In human population measurements, improved orientation encoding has been found when orientation (but not contrast) is task relevant (Jehee et al., 2011; Ling et al., 2015). Although it would be of interest to know whether these population effects of attention reflect differences in neural tuning bandwidth, selective attention, similar to image contrast, also modulates response amplitudes (Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Kastner et al., 1999; Kastner and Ungerleider, 2000; Reynolds and Chelazzi, 2004) and thus is expected to improve SNR for population measures. Similarly to contrast effects, attention should be expected to bias channel response functions toward a narrower tuning, even if neural tuning bandwidth does not change.

A similar disconnect between single-unit and population measures impacts even simpler measures of cortical response that do not require multivariate approaches. Contrast sensitivity can be directly imaged because single units monotonically increase response with contrast (Albrecht and Hamilton, 1982; Sclar et al., 1990; Busse et al., 2009), resulting in a population response that

---

**Figure 6.** Model simulations of how goodness of fit of the encoding model ($r^2$) varies with neural tuning width and noise level in the synthetic data. Different colors represent different amounts of Gaussian noise added to the simulated neural response.

**Figure 7.** Posterior distributions from the Bayesian analysis. These functions represent the probability that a given stimulus (measured by the offset from the true orientation, x-axis) caused the observed BOLD response. Red represents posterior distributions for high-contrast contralateral stimuli. Yellow represents posterior distributions for low-contrast contralateral stimuli. Black represents posterior distributions for ipsilateral stimuli. Shaded region represents the SE over subjects and hemispheres.
also monotonically increases (Tootell et al., 1998; Boynton et al., 1999; Logothetis et al., 2001; Avidan and Behrmann, 2002; Olman et al., 2004; Gardner et al., 2005). Spatial attention has generally been shown to shift contrast response vertically upward when measured with functional imaging (Buracas and Boynton, 2007; Li et al., 2008; Murray, 2008; Pestilli et al., 2011; Hara and Gardner, 2014), which appears to be different from the variety of effects from contrast-gain to response-gain reported for single units (Reynolds et al., 2000; Martinez-Trujillo and Treue, 2002; Williford and Maunsell, 2006; Lee and Maunsell, 2010; Pooremaeli et al., 2010; Sani et al., 2017). Consideration of normalization and the size of the attention field relative to stimulus-driven responses can give rise to effects that can account for single-unit responses and EEG measures (Reynolds and Heeger, 2009; Ittipipat et al., 2014). But predictions of this normalization model of attention may differ for single units and population measures as different neurons in a population can be exposed to different balance of attention field and stimulus drive, giving rise to additive shifts when considered as a population (Hara et al., 2014). Relatedly, response gain changes may also manifest as additive shifts when directly examining voxel feature selectivity (Saparo and Serences, 2010).

While neural tuning width can be reflected in channel response functions, neural tuning width and signal-to-noise changes are intertwined, making it hard to disentangle their effects. For example, one might examine conditions in which signal to noise is matched and then hope to attribute changes in channel response function bandwidth solely to changes in neural tuning bandwidth. However, our simulations show signal-to-noise measures, such as the variance accounted for by the encoding model ($r^2$), covary with neural tuning width. As neural tuning width broadens, there is less modulation of voxel response with orientation; thus, the encoding model shows a decrease in $r^2$. Therefore, even pure changes in neural tuning width would result in conditions with lower $r^2$, making it hard to attribute changes in channel response functions solely to changes in neural tuning width.

The results of our simulation are agnostic to the source of selectivity for orientation in voxels. One possible source of orientation information are the irregularities of columnar organization, which could give rise to small, idiosyncratic biases in voxels (Boynton, 2005; Swisher et al., 2010). However, large-scale biases for cardinal (Furmanski and Engel, 2000; Sun et al., 2013) and radial (Sasaki et al., 2006) orientations have been reported, and these biases have been shown to be an important source of information to drive classification (Freeman et al., 2011, 2013; Beckett et al., 2012; Wang et al., 2014; Larsson et al., 2017; but see Alink et al., 2013; Pratte et al., 2016). Large-scale biases may result from vascular (Gardner, 2010; Kriegeskorte et al., 2010; Shmuel et al., 2010) or stimulus aperture (Carlson, 2014) related effects. Our simulations do not require, or exclude, any topographic arrangement of biases. Regardless of the source of orientation bias, channel response function widths would be expected to broaden as signal to noise decreases.

More generally, our results suggest a “reverse-inference” problem (Aguiirre, 2003; Poldrack, 2006) when interpreting outputs from inverted encoding models. Forward encoding from hypothetical neural responses to population activity is a powerful tool, but reversing this process to infer about neural responses is problematic when there is not a one-to-one mapping between single-unit and population measures. Consequently, this reverse-inference problem is not restricted to channel-encoding models but will occur for other encoding model approaches, such as population receptive fields (Dumoulin and Wandell, 2008) or Gabor wavelet pyramids (Kay et al., 2008), if one were to invert these models to infer properties of the underlying neural responses. For contrast and orientation, both increases in response amplitude and neural selectivity can result in narrower bandwidth of the channel response functions, so reverse inference requires taking both into account. Regardless of which neural change has occurred, read-out of these responses, be they in the brain or from external measurement, will have less certainty about what stimulus has caused those responses. Techniques that represent the output of encoding models as posterior distributions (van Bergen et al., 2015) offer a straightforward interpretation of the uncertainty in determining stimulus properties from cortical responses.

References


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline


Medline