Systems/Circuits

Orientation Decoding in Human Visual Cortex: New Insights from an Unbiased Perspective

Thomas A. Carlson

Perception in Action Research Centre and Department of Cognitive Science and Centre for Cognition and its Disorders, Macquarie University, Sydney, NSW 2109, Australia

The development of multivariate pattern analysis or brain "decoding" methods has substantially altered the field of fMRI research. Although these methods are highly sensitive to whether or not decodable information exists, the information they discover and make use of for decoding is often concealed within complex patterns of activation. This opacity of interpretation is embodied in influential studies showing that the orientation of visual gratings can be decoded from brain activity in human visual cortex with fMRI. Although these studies provided a compelling demonstration of the power of these methods, their findings were somewhat mysterious as the scanning resolution was insufficient to resolve orientation columns, i.e., orientation information should not have been accessible. Two theories have been put forth to account for this result, the hyperacuity account and the biased map account, both of which assume that small biases in fMRI voxels are the source of decodable information. In the present study, we use Hubel and Wiesel's (1972) classic ice-cube model of visual cortex to show that the orientation of gratings can be decoded from an unbiased representation. In our analysis, we identify patterns of activity elicited by the edges of the stimulus as the source of the decodable information. Furthermore, these activation patterns masquerade as a radial bias, a key element of the biased map account. This classic model thus sheds new light on the mystery behind orientation decoding by unveiling a new source of decodable information.

Key words: fMRI decoding; hyperacuity; multivariate pattern analysis; orientation columns; visual cortex

Introduction

Multivariate pattern analysis (MVPA) methods have made a large impact on fMRI research (Norman et al., 2006; Kriegeskorte and Kreiman, 2011). The inherent complexity of MVPA, however, can obscure the source of decodable information (Op de Beeck, 2010). This is exemplified in the debate about the (theoretical) spatial resolution of MVPA. In 2005, two impactful fMRI decoding studies showed that the orientation of gratings is decodable from primary visual cortex. From this work, it was postulated that MVPA confers hyperacuity or subvoxel resolution to fMRI (Boynton, 2005; Haynes and Rees, 2005; Kamitani and Tong, 2005). The foundation of this conclusion was knowledge of the organization of primary visual cortex (V1), famously described by Hubel and Wiesel (1972) using the ice-cube model (Hubel, 1988), in which hypercolumns represent orientation continuously within V1's retinotopic map. These early influential MVPA studies scanned at an insufficient resolution to measure hypercolumns directly, which led to the hyperacuity account.

How does decodable information about orientation arise at the resolution of fMRI? This question has been the subject of

recent vigorous debate. There are currently two theories. The first is the hyperacuity account, which suggests that imperfect sampling of orientation columns in fMRI voxels creates small biases, and that MVPA exploits these biases to recover orientation (Boynton, 2005; Haynes and Rees, 2005; Kamitani and Tong, 2005). The alternative is the biased map account, which argues that orientation decoding relies on coarse scale biases at the retinotopic map level, and in particular a bias for radial orientations (Sasaki et al., 2006; Mannion et al., 2009; Freeman et al., 2011). A number of approaches have been taken to study these possibilities. Spatial filtering of the fMRI data initially seemed to yield straightforward predictions (Op de Beeck, 2010). The results, however, have been equivocal (Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010; Swisher et al., 2010; Alink et al., 2013). Freeman et al. (2011) attempted to establish what information is "necessary" and "sufficient" for decoding; but again the findings were not so clear-cut (Alink et al., 2013). Finally there was a stimulus-based approach, spurred by the argument that radially balanced spirals should not be decodable if decoding relies solely on a radial bias, but this again was not definitive (Mannion et al., 2009; Clifford et al., 2011; Freeman et al., 2013).

In the present study, we take a modeling approach. We constructed a model of visual cortex based on the classic ice-cube model (Hubel and Wiesel, 1972; Hubel, 1988), and then attempted the decode the orientation of gratings from the model's outputs. Importantly the "cubes" in our implementation were perfect, i.e., they had no biases whatsoever. We show the orien-

tation of gratings is decodable from an unbiased representation, a result that neither of the current theories put forth to explain orientation decoding can account for. We then identify the stimulus's edges as a source of decodable information for orientation decoding and show the model can account for a range of findings in the literature.

Materials and Methods

Model inputs/stimuli

To study the nature of decodable information in visual cortex, we examined five sets of stimuli. Each stimulus set was based on previous studies examining the basis of orientation decoding (Kamitani and Tong, 2005; Freeman et al., 2011, 2013). There were four sets of oriented gratings. The parameters of the gratings were taken from Freeman et al. (2011), which examined four grating patterns including a set modeled after the stimuli used by Kamitani and Tong (2005). The spiral stimuli were created using methods and parameters described by Freeman et al. (2013). To transform physical stimulus descriptions in visual angle to images for the model, we used a conversion of 25 pixels to 1° of visual angle. For descriptive purposes, we describe the results in terms of visual angle to maintain intuitive compatibility with the neuroimaging studies.

General stimulus parameters. Image intensity values were scaled to range from 0 to 255. The stimuli had an outer radius of 9.5° of visual angle (237.5 pixels). The grating stimuli were centered on a 25.5 × 25.5° (640 × 640 pixels) uniform gray background (intensity value = 127), which served as padding for the convolution of the images with model filters (see model below). The spiral stimuli were centered on a larger 32 × 32° background (800 × 800 pixels) background, which allowed sufficient padding the larger filters in the multiscale model (see model below).

Grating stimulus parameters. The gratings spatial frequency was 16 cycles/degree (cpd) of visual angle (0.025 cycles/pixel in the image). Each set of gratings had eight orientated exemplars (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5° rotations from vertical).

Square-wave gratings. The original claim that decoding methods confer hyperacuity used square-wave gratings (Kamitani and Tong, 2005). Freeman et al.'s (2011) rendering of their stimuli was a square-wave grating displayed in an annulus with an outer radius of 9.5°. The inner radius of the annulus was 0.5° (12.5 pixels).

Sine-wave gratings. Freeman et al. (2011) tested three types of sinusoidal gratings. The first had the same edge boundaries as the square-wave stimuli described above, outer radius (9.5°) and the inner radius (0.5°).

Sine-wave gratings with large annulus and blurred transition. The second set of sinusoidal gratings had a larger inner radius of 4.5° (112.5 pixels). For this set, Freeman et al. (2011) blurred the inner and outer edges of the annulus with a 1° raised cosine transition (25 pixels).

Sine-wave gratings with large annulus very blurred transition. Freeman et al. (2011) additionally tested a stimulus set with a wider edge transition. Here, the cosine transition was applied over half of the width of the annulus (2.5°), such that the stimulus was only full contrast at a single eccentricity. Following this, our rendered version was the same as above except with a transition period of 67.5 pixels over the inner and outer edges.

Log spiral gratings. The capacity to decode spiral gratings has been used to argue that multivariate patterns analysis methods must rely on more than global radial preference maps (Clifford et al., 2011; but see Freeman et al., 2013). Following the procedure of Freeman et al. (2013), we created a set of spiral gratings by plotting spiral lines and rasterizing the images in MATLAB (β factor = 0.7). After rendering the stimulus image, an image mask was used to set the inner radius 4.5° (112.5 pixels), the outer radius to 9.5°. Both edges were blurred using 1° (25 pixels) cosine transition.

Perfect cube model

We constructed a model of visual processing to study sources of decodable information beyond local and global biased representations. The model was based on the ice-cube model described by Hubel and Wiesel (1972; Hubel, 1988). Ocular dominance columns were not included in the model as eye of origin information was not relevant. In the model,

orientation is encoded using Gabor filters (Fig. 1B), which simulate the response of simple cells (Hubel and Wiesel, 1959). The model encoded eight orientations (0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5° rotation from vertical), the same as the stimulus orientations. The filters were 2D sine-wave gratings with a spatial frequency of 16 cpd (0.025 cycles/pixel in the image) multiplied times a 2D Gaussian kernel (σ = 0.5°/ 20 pixels) with a width of 3° (120 pixels). Note the spatial frequency of the filters matches that of the sine-wave grating stimuli, and thus are optimally suited to encode these patterns. The spatial frequency of the filters also matched the first harmonic (the fundamental frequency) of the square-wave gratings, and thus was also well suited to encode these stimuli. We also implemented a multiscale perfect cube model (PCM; Fig. 1) model (described below) that encoded the higher order harmonic frequencies of the square-wave stimulus. The model outputs for these frequencies were not substantially different from that of the fundamental (data not shown).

Each of the grating stimuli was convolved with the eight oriented Gabor filters to generate an activation map representing the spatial response of the filter. To remove the effect of the stimulus phase, we followed the approach used by Freeman et al. (2011). In their fMRI study, the authors presented subjects 16 equally spaced phase exemplars at a high rate (250 ms/exemplar). By presenting the stimuli at a high rate, the slow hemodynamic response was used to temporally average the response to the individual phase exemplars, generating a phase-independent pattern of activation in the cortex. In the model, we generated a phase-independent response by pooling the model's response to 64 equally spaced phase-stimulus exemplars (Fig. 1A). In this procedure, each phase exemplar was convolved with each filter (Fig. 1B) producing 64 responses for each filter. These responses were squared individually (to make the response output non-negative) and then summed to produce a single spatial map describing the activity from each filter (independent of stimulus phase; Fig. 1C). Finally, the filter response maps were summed into a single topographic map, which is the pattern response of the PCM model. Each pixel's intensity in the pattern response is the summed activity of all the orientation channels, and can be thought of as a perfect cube or as an unbiased fMRI voxel. The collective pattern represents the activation of simple cells in V1's topographic representation of the stimulus.

The PCM model has two key properties. (1) Orientation is implicit. By summing the filters' responses, activity at a given location indicates only that there is orientation "energy" at that location. The individual contribution of any orientation channel is ambiguous. Decodable information therefore can only arise at the pattern or map level. (2) Local orientation is perfectly balanced. All eight oriented filters contribute equally to the response at each location (pixel/voxel). This is critical as both the current theories of orientation decoding assume biases (either random or organized in a global map) are the source of decodable information for orientation decoding.

Multiscale PCM. For spiral stimuli, the spatial frequency of the pattern changes as a function of eccentricity. Although a single-level model is sufficient for the grating stimuli, in which there is a dominant frequency, the spiral stimuli require a model that encodes multiple spatial frequencies. We thus constructed a multilevel model in which orientation is encoded at multiple scales, akin to the S1 layer of the HMAX (Riesenhuber and Poggio, 1999; Serre et al., 2007). The spatial frequency channels were consisted of six different sized filters. The filters had a fixed frequency of three cycles and varied in size (30, 60, 90, 120, 150, and 180 pixels). This scaling resulted in filters that encode the stimuli at 2.5, 1.25, 0.825, 0.625, 0.50, and 0.425 cycles/°, respectively. The responses of the spatial frequency channels were analyzed both individually, and as a summed response map (see Results). All other aspects of the model were the same as the single level PCM model.

Results

The orientation of gratings is decodable from an unbiased visual representation

Although fMRI voxels undoubtedly will have small biases in the number of neurons representing different orientations, there is no evidence that these biases are sufficiently large to be measured
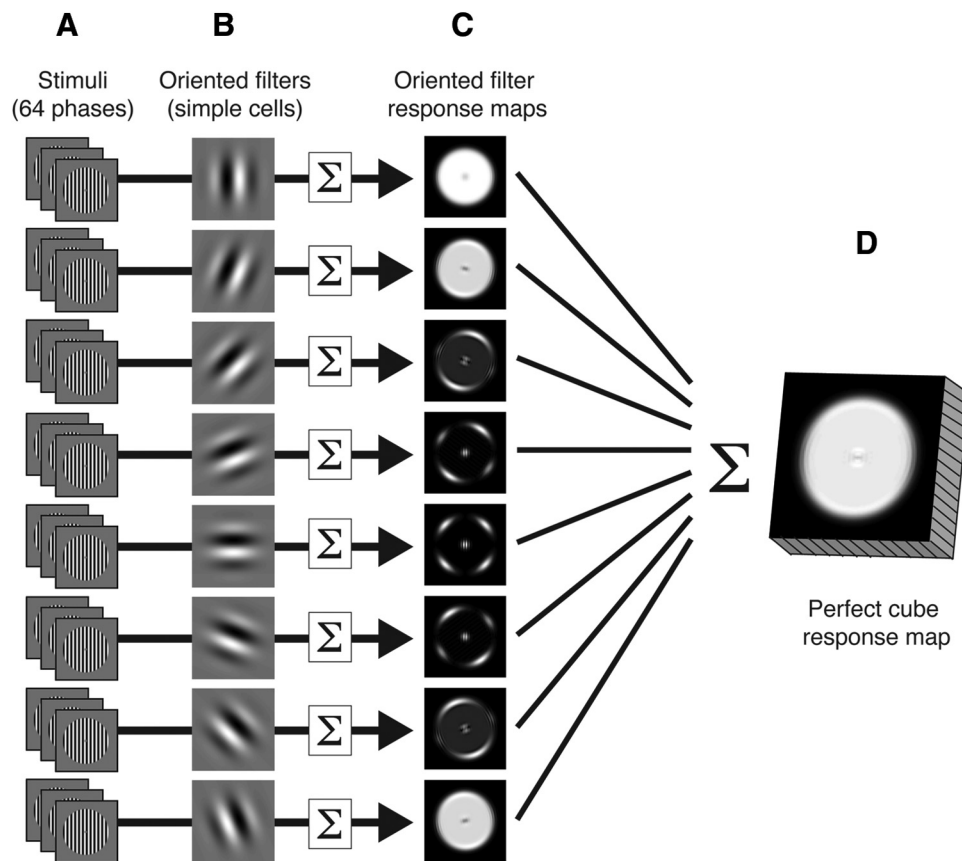
A
Stimuli
(64 phases)

B
Oriented filters
(simple cells)

C
Oriented filter
response maps

D

Perfect cube
response map

Figure 1. The PCM. **A**, Sixty-four phase-shifted stimulus exemplars. **B**, Oriented filters acting as simple cells encoding orientation are convolved with the stimulus exemplars. **C**, The response of each filter to each exemplar is squared, and then the exemplar responses are summed to make an activity map for each orientation filter. **D**, The filters' response maps are summed into a single pattern response map, in which each pixel is a "cube" that has a perfectly balanced contribution from each orientation channel, representing an unbiased fMRI voxel.

from the noisy fMRI signal. These biases are a key assumption of both the hyperacuity and global biased map accounts of orientation decoding. To date, no one to our knowledge has tested whether this assumption is necessary to account for orientation decoding. To study this, we constructed a model of early visual processing that lacked any such biases, which we refer to as the PCM (Fig. 1).

We first attempted to decode orientation from PCM's outputs. Our study tested a range of grating stimuli used by studies investigating the nature of decodable information in visual cortex (Kamitani and Tong, 2005; Freeman et al., 2011). There were four sets of gratings, each with eight orientations. To determine whether orientation is decodable from the PCM, we computed the correlation distance $(1 − \rho)$ between PCM's pattern response for all possible pairwise comparisons between orientations. If the correlation distance between the PCM's outputs is greater than zero, a correlation classifier (and virtually all other classifiers as well) would be capable of "decoding" the patterns. Figure 2A–D shows the decoding results formatted as dissimilarity matrices (DSM). Each entry in the DSM is the correlation distance between a pair of stimuli. The diagonal is the correlation distance between the model outputs for the same stimulus, which by definition is 0. For all the sets of stimuli, we found that orientation is decodable. It is notable that there is a relationship between orientation disparity and decodability. The diagonal stripe pattern in the DSMs shows that larger orientation differences are more decodable, concordant with the findings of Kamitani and Tong (2005). It is also noteworthy that not all the stimulus sets were equally decodable. The correlation distance between the heavily

blurred grating stimuli is more than a factor of 2 lower than the other stimulus sets (note the scale change from Figs. 2A–C and D). This suggests that the edge blurring manipulation used by Freeman et al. (2011) is degrading the decodability of the patterns.

Using the PCM and stimuli used previously to study the nature of decodable information in visual cortex, our analysis thus far has shown that decodable orientation information need not arise from biases in fMRI voxels. These findings suggest that current theories may have overlooked alternative sources of information that might account for orientation decoding.

Stimulus edges produce nonuniform spatial responses at the pattern response level
We found decodability was lower for the patterns with heavily blurred edges, indicating a candidate source of decodable information is the edges. We examined this first by looking at the response of the individual filters, before summing their responses. Figure 3A shows the individual filter responses to a vertical exemplar from the four stimulus sets. The vertical filter that is aligned with the orientation of the stimulus produces a uniform spatial response shaped like the stimulus. The filters that are not in alignment with the stimulus orientation, however, show nonuniform spatial responses. The 45° filter, for example, shows more activity in the upper right and lower left quadrants. It is also apparent that the distortions are localized to the edges of the stimulus. Note that the stimulus sets with the small inner radius (Fig. 3A, top two rows) have distortions in a small region in the center, whereas the stimulus sets with a larger inner radius (Fig.
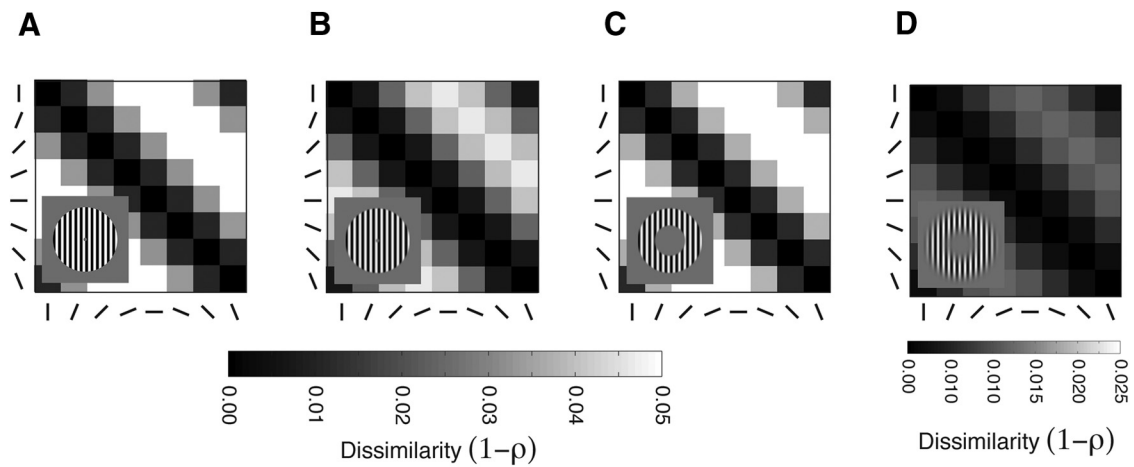
Figure 2. Decoding results. ***A–D***, Correlation classifier DSMs for the four sets of grating stimuli. Rows and columns are stimulus orientation. Individual entries in the DSM are correlation distances (i.e., the decodability) between pairs of oriented gratings. Inset, Example stimuli. ***A***, Square-wave grating with a small inner radius edge. ***B***, Sine-wave grating with a small inner radius edge. ***C***, Sine-wave grating with a large inner radius edge. ***D***, Sine-wave grating with a small inner radius edge and heavy blur function. Note the scale change in dissimilarity for ***A–C*** DSMs to the DSM for ***D***.
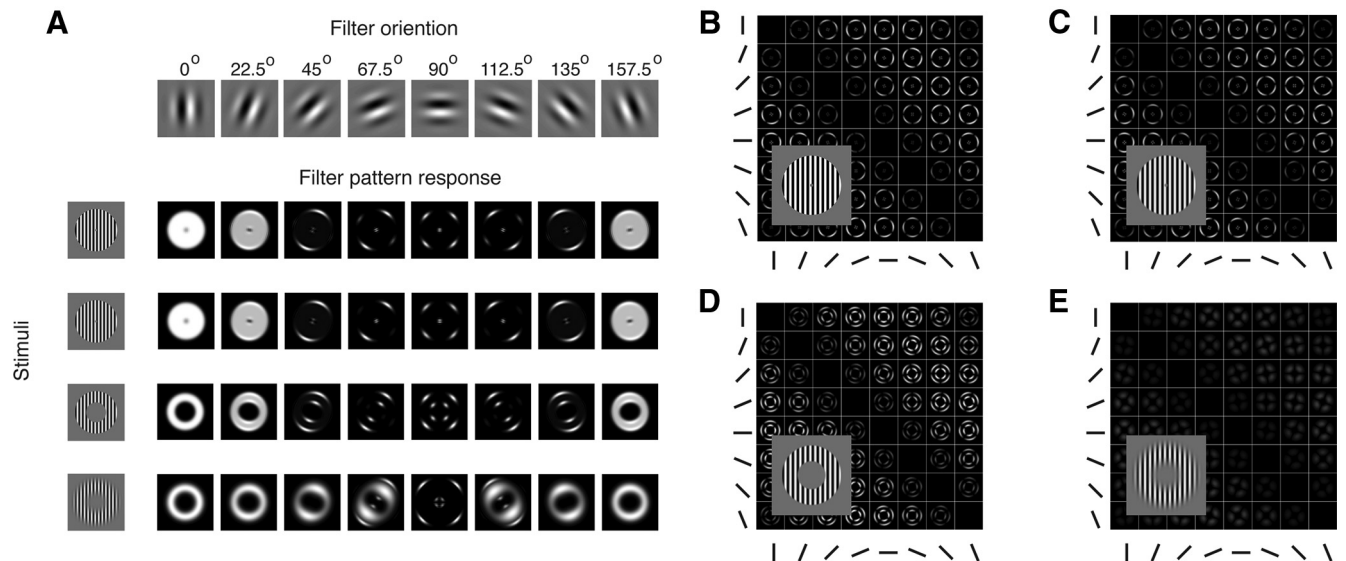

Figure 3. Edge related activity. ***A***, Orientation filter-response patterns for a vertical stimulus. The oriented filters are shown at the top of each column. Vertically oriented exemplars from each stimulus set are shown at the beginning of the rows. Pattern responses from the filters to the vertical exemplar from each sets shown as images. Pattern responses are scaled individually from 0 to 1. ***B–E***, DSM of the differential response between oriented stimuli. Rows and columns are exemplar orientations. Each entry in the DSM is the absolute difference in the model response to a pair of exemplars. The four DSMs were normalized collectively to range from 0 to 1 for the purposes of comparing responses across stimulus sets. Example stimuli are shown in the insets. ***B***, Square-wave grating with a small inner radius edge. ***C***, Sine-wave grating with a small inner radius edge. ***D***, Sine-wave grating with a large inner radius edge. ***E***, Sine-wave grating with a small inner radius edge and heavy blur function.

3*A*, bottom two rows) have a ring shaped distortion pattern near the inner edge of the stimulus. It is important to note this analysis is done at the individual filter level, which is not accessible to the classifier. If these distortions vary systematically with orientation exist after summing the filter responses, then a classifier could use these patterns to "decode" orientation from an unbiased representation.

We next turned our examination to the summed output of the model. One simple way to determine the source of the decodable information is to look at the difference in the model's response to pairs of stimuli. Figure 3*B–E* show the absolute difference in pattern response for all possible pairwise combinations between the stimuli. The difference between stimuli is clearly localized to the edges, as expected from the analysis of the individual filters.

The square-wave stimulus (Fig. 3*B*) has a small differential response in the center and a large ring-shaped differential response, coinciding with the inner and outer edges. The stimulus set with the same edge profile (Fig. 3*C*) has nearly the identical pattern. Figure 3*D* is the stimulus set that has an inner edge with a larger radius. In comparing Figure 3, *C* and *D*, the differential activity shifts in accordance with the relocation of the edge. Finally, Figure 3*E* shows the differential response for the stimulus set with an extremely blurred edge. This notably was Freeman et al.'s (2011) best effort to address the possibility of an edge artifact. Although the edge related differential activity is moderated substantially, there remain differences.

This analysis shows that (1) orientation filters not aligned with the stimulus produce distorted spatial responses, (2) that these

distortions are a source of decodable information as they produce differential responses at the map level, and (3) that these distortions are produced by the edges of the stimulus.

Edges-related activity masquerades as a radial bias
Freeman et al. (2011) presented a parsimonious account of orientation decoding based on the radial bias. We have now shown that the edges of stimuli generate decodable activation patterns in a model of visual processing without any such bias. We next examined whether there is a relationship between the spatial distribution of activity elicited by edges and the radial bias as observed in fMRI studies (Sasaki et al., 2006; Mannion et al., 2010; Freeman et al., 2011). Briefly the radial bias in fMRI is increased activation in retinotopic maps locations that are in correspondence with the orientation of the stimulus; e.g., a vertically oriented grating stimulus would evoke more activity in the upper and lower visual field, and a horizontal grating would evoke more activity in the left and right visual field. The accepted interpretation of this result is that there are a greater proportion of neurons representing orientations pointing toward the fovea in visual cortex (Sasaki et al., 2006; Mannion et al., 2010; Freeman et al., 2011).

Is there a relationship between the edge distortions in PCM's pattern response and the radial bias? Figure 4A shows the model output for the eight orientated exemplars from the large inner annulus sine-wave stimulus set. The top row is the exemplars, the middle row is PCM's pattern response to the exemplar, and the bottom row is the 2D Fourier transform (2DFT) of PCM's pattern response. Note the pattern responses are slightly elongated along the axis corresponding to the stimulus orientation. The vertical stimulus, for example, is stretched vertically in PCM's pattern response. This nonuniform spatial response is more easily seen in the 2DFT, which clearly shows more orientation energy aligned with the stimulus orientation. This response pattern accords with the radial bias prediction, which in the case of the vertically stimulus predicts greater activity in the upper and lower visual field. For all the orientations, PCM's pattern response is elongated along an axis in accordance with the stimulus orientation, matching the radial bias prediction.

The 2DFT of PCM's pattern response (Fig. 4A, bottom row) is organized with the low spatial frequencies in the center and increasingly higher spatial frequencies going outward. Note the energy in the 2DFT is neither strongly localized to the center (low frequencies) nor a ring pattern corresponding to high spatial frequencies. The response is broadband. This is in agreement with studies showing that information for orientation decoding is not strongly localized to either low or high spatial frequency patterns of activation in the cortex, which are the predictions of the global map and hyperacuity accounts respectively (Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010; Op de Beeck, 2010; Swisher et al., 2010; Freeman et al., 2011; Alink et al., 2013), albeit this is somewhat oversimplified.

We next examined the relationship between PCM's pattern response and its relationship to the radial bias more directly. Figures 4B–E show the mean activation of the model as a function of polar angle for a region defined by a double wedge (see icons below the plots) for different orientations. For all four sets of grating patterns, there is a clear relationship between polar angle and stimulus orientation. Vertically oriented stimuli evoke greater activity in the top and bottom of the map; stimuli tilted 45° evoke greater activity in the upper right and lower left regions of the map; horizontal stimuli evoke greater activity in the left and regions of the map, etc; exactly matching the radial bias prediction.

Next, to hone in on whether it is edge-related activity specifically that is mimicking the radial bias, we plotted the data at three radiuses corresponding to the inner edge, the outer edge, and an intermediate eccentricity in the middle of the annulus for the sine-wave grating stimuli with a large inner radius (Fig. 5A). At the inner edge (Fig. 5A, left plot) and outer edge (Fig. 5A, right plot) activity waxes and wanes in agreement with the radial bias prediction. In the middle of the annulus (Fig. 5A, middle plot) activation is at ceiling, indistinguishable across different polar angles. This shows that it is the edges of the stimulus that are the source of the activation patterns mimicking the radial bias.

Last, we considered whether the observed edge activity is sufficiently large to be measured with fMRI. There are many factors that would contribute to the magnitude of the effect at the level of BOLD, e.g., the contrast response function, surround suppression, and the translation of neuronal activity to BOLD response. One way to examine whether the observed edge related activity is sufficiently large to be measured with fMRI is to compare its magnitude to activity driven purely by the stimulus, which clearly can be measured using fMRI. Figure 5A shows PCM's response to both the edges and the stimulus normalized to range from 0 to 1. To estimate the stimulus activity, we can take the response from the intermediate eccentricity (Fig. 5A, middle plot). Expectedly the PCM responds strongly to the stimulus (value of 1). In comparison, edge activation ranges from 0.25 to 0 to the inner edge (Fig. 5A, left plot) and from 0.15 to 0 to the outer edge (Fig. 5A, right plot). The edge effect is thus ~20% the magnitude of the stimulus for this stimulus set. Figures 5B–D show the same data for the square wave, the small inner radius sine wave, and the heavily smoothed stimulus. For these stimuli, the edge effect ranges from 20 to 10%. Of the four stimulus sets, the edge effect is weakest for the heavily smoothed stimulus; however, note the radial bias effect bleeds into the center of the stimulus (presumably due to smoothing). This suggests that decodable edge related activity for the heavily smoothed stimulus would be more distributed. Importantly, whereas the edge effect is clearly lower than the stimulus effect, 10–20% of the magnitude of a full contrast stimulus (typical in fMRI orientation decoding studies) would likely be sufficient to be measured in the fMRI signal.

To summarize, we have shown that (1) oriented stimuli elicit activation patterns that mimic the radial bias in PCM, (2) that these patterns of activation are broadband, (3) the source of the activation patterns are the stimulus's edges, and (4) the magnitude of edge activity is likely sufficient to be detectable with fMRI.

Orientation decoding and edge activity
We have shown decodable activity for orientation emerges from the edges of the stimulus in the PCM, and that this activity masquerades as a radial bias. We next examined how filters with varying orientation disparity (OD) relative to the stimulus (i.e., stimulus orientation–filter orientation) contribute to the edge effect mimicking the radial bias (Fig. 6). Figure 6A shows the responses of five filters with varying ODS (0°, 22.5°, 45°, 67.5°, and 90°). By visual inspection, the filter aligned with the stimulus (i.e., 0° OD) generates a uniform spatial response in the shape of the stimulus. In contrast, the filters with ODS greater than zero all showed spatially nonuniform responses. The filter with the smallest OD (22.5°) evoked an ellipse-shaped pattern response tilted slightly away from the orientation of the stimulus in the direction of the filter's orientation. The filter notably also showed a stronger response to the edges (note the bright areas corresponding to the stimulus edges along the principle axis of the ellipse). The filters with ODS >22.5° did not respond to the inner
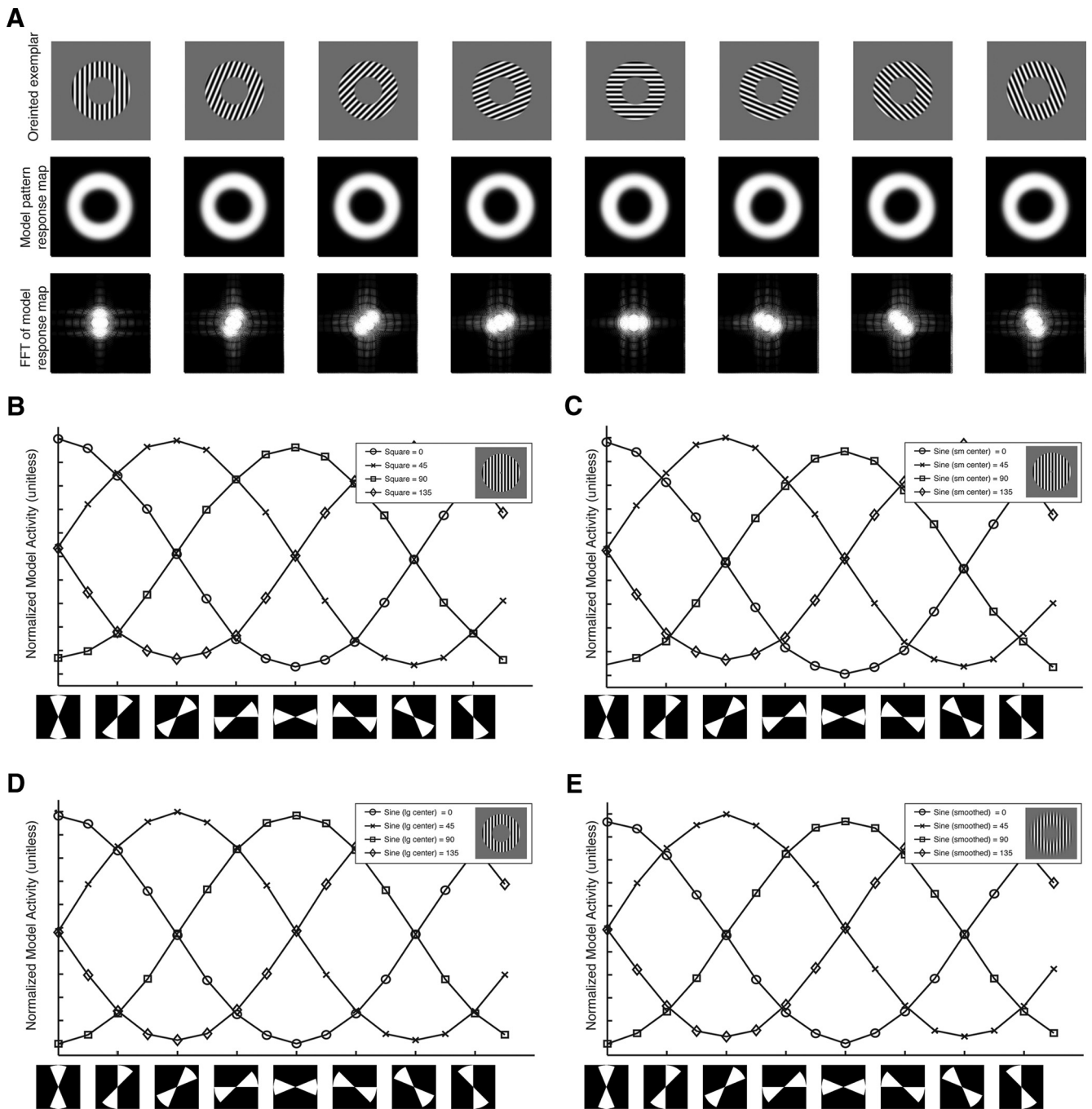
Figure 4. Model activity mimics a radial bias. *A*, Top, Orientation exemplars from the sine-wave stimulus set with a large inner edge radius. Middle, PCM's pattern response. Bottom, Log amplitude of the 2D Fourier analysis on the model response. Spatial frequency is lowest in the center of the image and increases from the center. The highlighted central region indicates frequencies below the Nyquist sampling rate. *B–E*, PCM activity as a function of polar angle. The *x*-axis is a double wedge region of the PCM's pattern response (diagrammatically shown below the axis). The wedge has a width of 45° and its angular position shifts stepwise in 22.5° steps. The *y*-axis is the mean PCM activation within the defined region. The lines are different stimulus orientations. *B*, Square-wave grating with a small inner radius edge. *C*, Sine-wave grating with a small inner radius edge. *D*, Sine-wave grating with a large inner radius edge. *E*, Sine-wave grating with a small inner radius edge and heavy blur function.

region of the stimulus, but did respond to the edges. Similar to the 22.5° OD filter, the peak of the response to the edge is tilted away from the orientation of the stimulus in the direction of the filter's orientation. A dashed white line denotes the orientation of the peak edge response for ODS greater than zero in the figure. Numerically, the orientation of the peak edge response was 11.25°, 22.5°, 33.75°, and 45° for the 22.5°, 45°, 67.5°, and 90° OD filters, or the average of the filter and stimulus orientation. Importantly this means that the edge response of the non-zero OD filters will

be shifted toward locations in the map corresponding to the stimulus orientation (the radial bias prediction). Although the exact location of the edge is not perfect, when the positive (clockwise rotation) and negative (counterclockwise rotation) OD filter offsets are pooled, there is more activity in the response map in locations corresponding to the stimulus orientation (Fig. 6B). In particular, note that the filters with shallow ODS (22.5° and 45°) generate an edge response in the upper and lower visual field for a vertical stimulus, matching the radial bias prediction. Activity
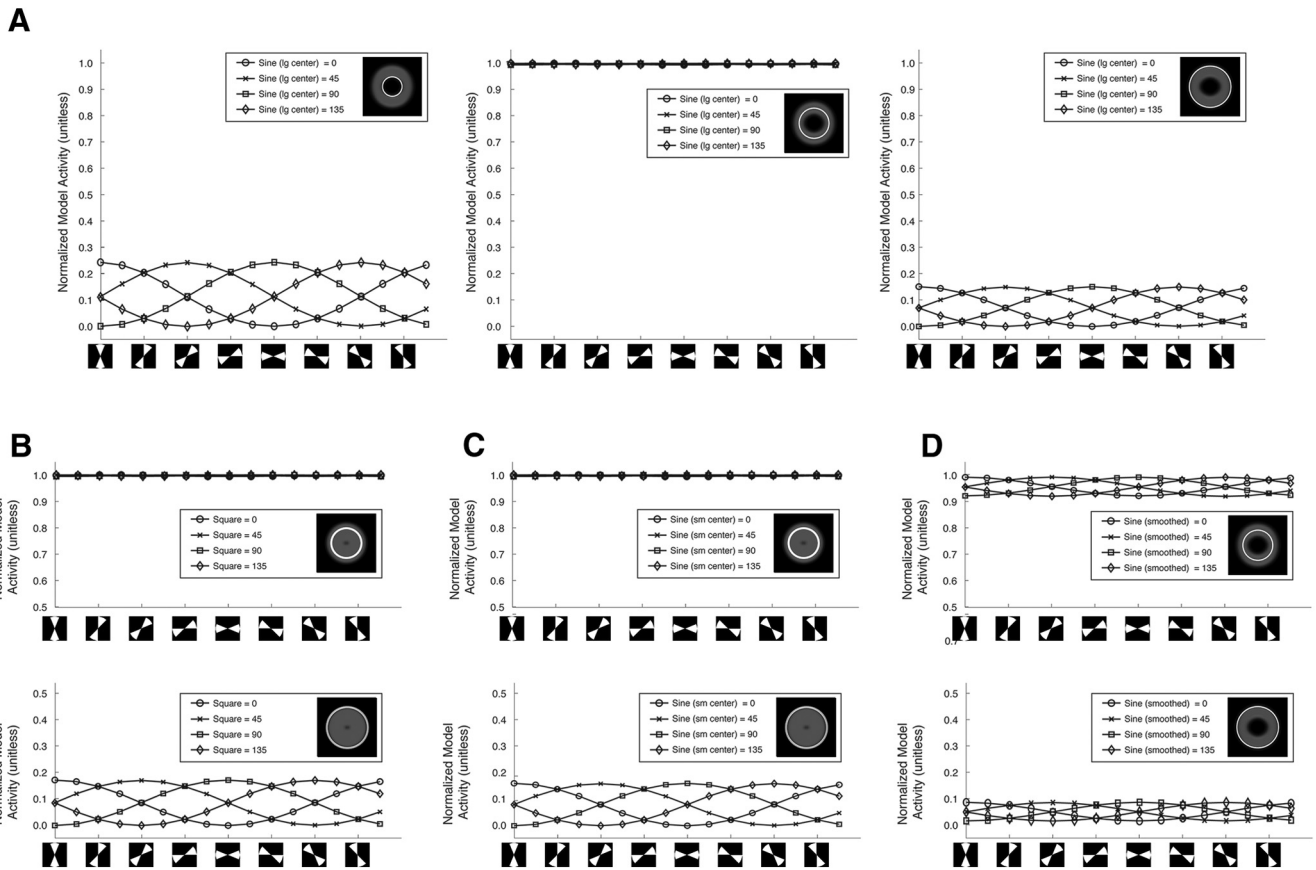
Figure 5. Edge related activity mimics a radial bias. **A**, Mean activation of the PCM as a function of polar angle for 3 radiuses: the inner edge (left), the center of the annulus (middle), and the outer edge (right). The image in the inset of the legends graphically shows the radius. The x-axis is a defined double wedge region with a width of 45° that shifts stepwise in 22.5° steps (diagrammatically shown below the axis). The y-axis is the mean PCM activation within the defined region. The lines are different stimulus orientations. Activity was normalized collectively over the three radius conditions (inner edge, middle of the annulus, and outer edge) to range from 0 to 1 for the purposes of comparison. **B–D**, Mean activation of the PCM as a function of polar angle for two radiuses: the center of the annulus (top), and the outer edge (bottom) for the square-wave stimulus (**B**), the sine-wave grating stimuli with a small inner radius (**C**), and the heavily smoothed sine-wave grating stimuli with a large inner radius (**D**).

from non-zero disparity filters, particularly those with shallow ODS, is thus the likely source of the edge activity masquerading as a radial bias.

Before examining the contribution of both the zero and non-zero OD filters to the edge response, we sought an explanation for why the non-zero OD filters respond to the edges. Figure 6C shows four panels displaying the upper right quadrant of the vertical stimulus with semi transparent diagrammatic filters superimposed over the stimuli. To best explain the edge response, the filters have been placed at the location on the stimulus where the filter generates the maximum edge response (peak location denoted by the dashed line; Fig. 6A). On the edge, the receptive field of the filter is occupied partially by the stimulus and partially by the empty background. It is the interaction between the stimulus region and the empty background that is the source of the edge response. For example, for the 22.5° OD filter (Fig. 6A, top) the filter's receptive field is partially occupied by the stimulus and in the region there is a relatively good correspondence with the pattern, which would drive a strong response. The other part of the filters receptive field is occupied by the background, which has no effect on the response. The net effect is a positive response (strong response from the stimulus + no response from the background). This is even more pronounced in the filters with ODS >22.5° (Fig. 6B, bottom three panels). Each of these filters does not respond to the body of the stimulus (Fig. 6A) because the

misalignment between the stimulus orientation and filter orientation is sufficiently large that the positive and negative components perfectly cancel one another. At the edges, however, these filters net a positive response because of the imbalance created by the empty background region.

We examined each of the different OD filters' contribution to the edge response to better determine the source of the edge effect. Figure 6D–G shows the cumulative response of the OD filters to the edge for a vertical stimulus at different polar angles. The "hump" in the middle of the plot corresponds to the activity mimicking the radial bias, in this case activity in the upper and lower visual field in response to a vertical stimulus. For the zero OD filter, there is a large amount of activity contributed to the overall edge response, but little or no variation as a function of polar angle. The zero OD filter thus contributes little to the response that is mimicking the radial bias. The source is clearly the non-zero OD filters, and particularly the filters with smaller disparities. Activity for the 22.5° and −22.5° OD filters, in fact accounts for >90% of the activity mimicking the radial bias. It is the response of the low OD filters that best accounts for the activity masquerading as a radial bias.

In summary, (1) filters (neurons) of all levels of OD generate a response that mimics the radial bias, (2) edge related activity for non-zero disparity filters (neurons) is brought about from an interaction between the stimulus and background (nonstimulus)
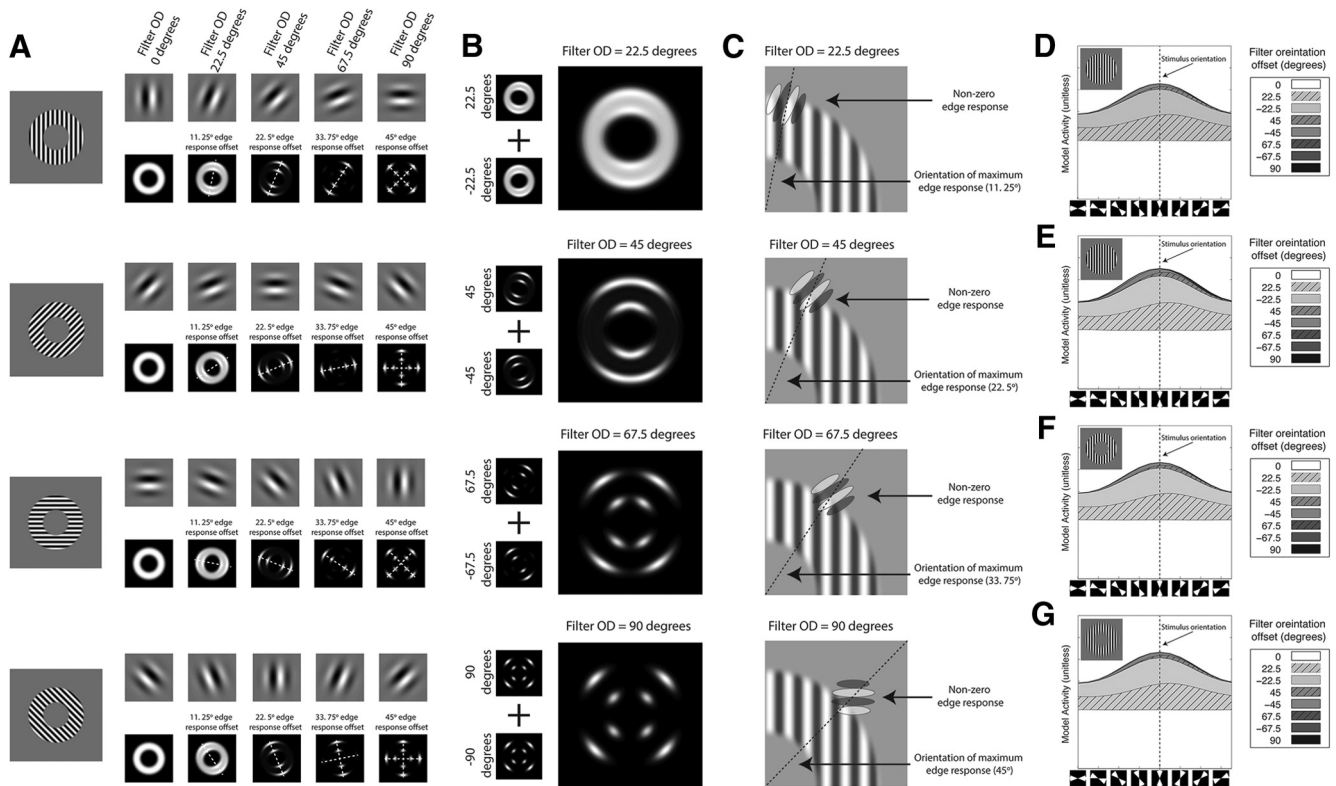
Figure 6. Orientation coding and edge activity. **A**, The response to oriented gratings by filters with different levels of OD. Shown are four oriented grating exemplars with 0°, 45°, 90°, and 135° orientations. To the right of the exemplars is the response of filters with varying levels of OD. The orientation disparity of the filters increases left to right with 0°, 22.5°, 45°, 67.5°, and 90° offsets (top row). Below each filter the filter's pattern response. The dashed line placed over the pattern response indicates the orientation of the peak response to the edge. **B**, Summed response of the positive (clockwise) and negative (counterclockwise) OD filter responses a vertical grating pattern. **C**, Graphical depiction of interaction between filter disparity and the stimulus edge. Shown is the upper right quadrant of a vertically sine-wave stimulus. Semitransparent filters with varying OD are superimposed over the stimulus in the location corresponding to the peak edge response (indicated by the dashed line) for the filter. From the top to the bottom, plots show filters with increasing OD. **D**–**G**, Individual filters with varying OD contributions to the outer edge response. Shown are area plots showing the cumulative response of the filters to the outer edge response to a vertical grating. Inset, The grating exemplar from each stimulus set. The x-axis is a double wedge with a width of 45° region constrained to the outer edge that shifts stepwise in 22.5° steps (diagrammatically shown below the axis). The y-axis is the filter's response within the wedge. The filled regions are the response of individual filters. Filter orientation is described relative to the stimulus orientation. **D**, Sine-wave grating with a small inner radius edge. **E**, Square-wave grating with a small inner radius edge. **F**, Sine-wave grating with a large inner radius edge. **G**, Sine-wave grating with a small inner radius edge and heavy blur function.

region of the figure, and (3) the edge related activity mimicking the radial bias is best accounted for by filters (neurons) with small ODS.

Spiral sense is decodable from an unbiased visual representation too
Spiral sense (i.e., direction of rotation; Fig. 7A) is decodable from activity in the visual cortex measured using fMRI (Mannion et al., 2009; Alink et al., 2013; Freeman et al., 2013). This observation has been used to argue that the radial bias is insufficient to explain orientation decoding (Clifford et al., 2011; but see Freeman et al., 2013). We next asked whether spiral sense could be decoded from an unbiased representation. Because the spatial frequency of spiral stimuli changes as a function of eccentricity, we constructed a variant of the PCM called msPCM that encodes the stimuli at multiple scales using six spatial frequency channels.

We first examined msPCM's differential response to spiral sense in the six spatial frequency channels individually (Fig. 7B). The differential spatial response varied across the spatial frequency channels. The high-frequency channels (top row) showed a larger differential response to the interior portion of the spiral; while the low spatial frequency channels (bottom row) showed a larger differential response to the outer region. This is expected as the center of the spiral has high spatial frequencies and the outer region has low frequencies. Importantly each of the six channels

produces a differential pattern response, thus spiral sense is decodable from msPCM's representation.

We next examined how the differential activation patterns between spiral sense changes as a function of polar angle (Fig. 7C). Here we observed a relationship between location in the response map and the differential response. All of the channel's differential responses waxed and waned as a function of polar angle, but phase differed depending on spatial frequency. Channels encoding high spatial frequencies (i.e., spatial frequencies less than or equal to 0.033 cycles/pixel) showed the largest differential response in the upper visual field and the smallest in the lower visual field; the intermediate spatial frequency channel (0.025 cycles/pixel) showed the largest differential response in the upper left visual field and the smallest in the lower right visual field; and the channels encoding low spatial frequency patterns (i.e., spatial frequencies greater than or equal to 0.02 cycles/pixel) showed the largest differential response in the upper right visual field and the smallest in the lower left visual field.

At the level of typical measurement in fMRI (3 mm³ voxel), the response of neurons encoding different spatial frequencies (and orientations) will be pooled. We next examined the pooled response of msPCM. Figure 7D shows the differential activation between spiral senses pooled over the six spatial frequency channels. As there is a differential response at the pooled pattern level,
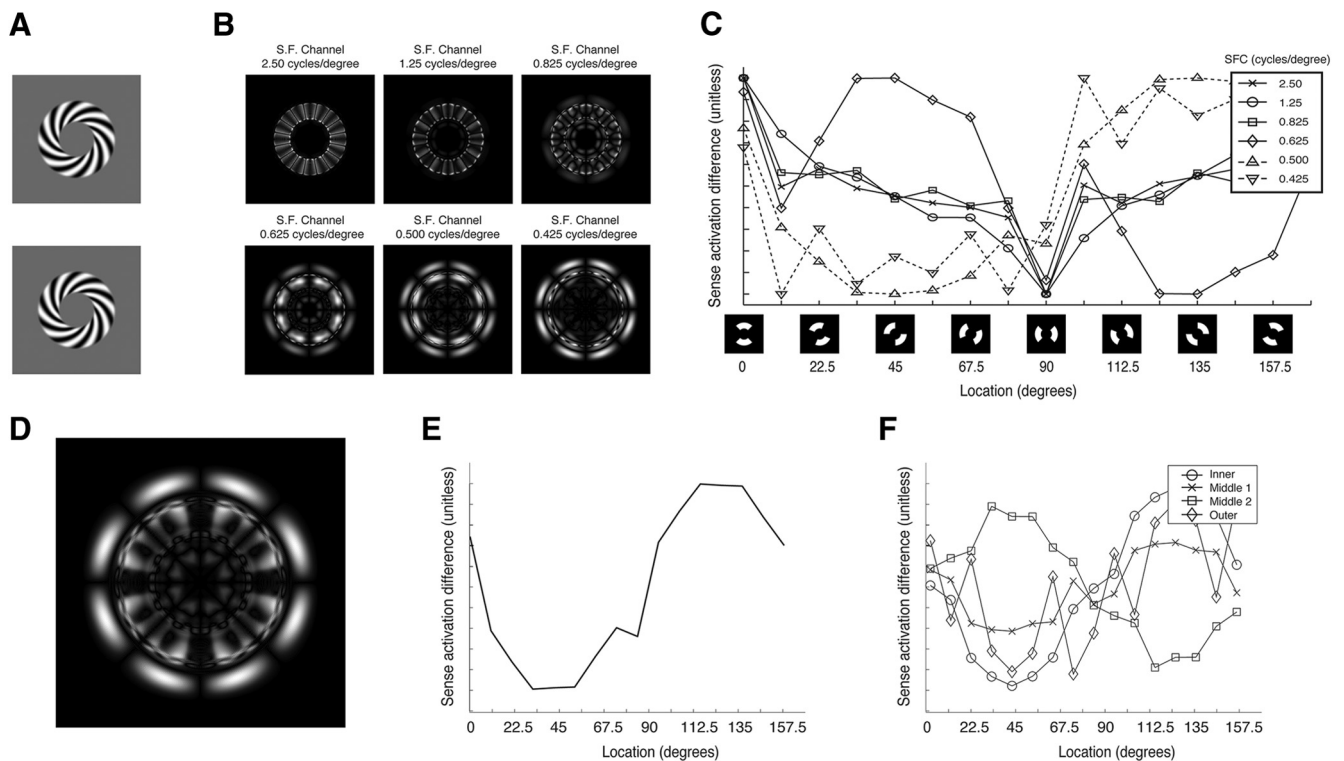
Figure 7. Decoding spirals. **A**, Spiral sense exemplars. **B**, Multiscale PCM's spatial pattern response output difference between spiral sense for the six different spatial frequency channels (SFC): 2.50, 1.25, 0.8250, 0.6250, 0.5000, and 0.4250 cycles/°. **C**, Mean activation as a function of polar angle for the different spatial frequency channels. The *x*-axis is a double wedge region with a width of 90° that shifts stepwise in 22.5° steps (diagrammatically shown below the axis). The *y*-axis is the mean layer activation within the defined region for the spatial frequency channel. **D**, Spatial difference between senses after combining the six multiscale PCM spatial frequency channels. **E**, Mean activation as a function of polar angle. **F**, Mean activation as a function of polar angle for four radiuses: the inner edge (circles), the outer edge (diamonds), and two radiuses inside the stimulus (squares and crosses).

this shows that spiral sense is decodable from msPCM's unbiased representation. Figure 6E shows the differential response as a function of polar angle. Here, the cyclical relationship is still present. In interpreting this pooled response, it is important to point out that the pooled response will be dependent on the spatial scales chosen to be included in the model. Including more layers representing high spatial frequencies, for example, would increase differential activation near the center of the map and would generate a greater differential response in the upper visual field. We selected six scales that would evenly cover the range of spatial frequencies in the spiral stimuli. This selection may not be representative of the actual distribution in visual cortex. What is important is that different scales evoke different patterns of differential activation, and it would be very unlikely that the different scales' patterns would cancel one another. Thus at the pooled map level, there would likely exist a decodable pattern for recovering spiral sense.

Finally, we examined whether or not decodable activation patterns could be ascribed to the edges. Figure 7D shows the differential activation for sense as a function of polar angle for the two edges (inner and outer) and two radiuses in the center of the pattern. Unlike the grating stimuli, differential activity for spiral sense was present throughout the annulus. This indicates that spiral decoding can rely both on the edges and the body the pattern.

In summary, (1) spiral sense can be decoded from an unbiased representation, (2) spiral sense can be recovered at multiple spatial scales (spatial frequencies) and from the integrated multiscale representation, and (3) decodable information for spiral sense is not confined to the edges.

Discussion

We implemented a version of the ice cube model of visual cortex (Hubel and Wiesel, 1972; Hubel, 1988) to study the nature of decodable information for orientation decoding in fMRI. Using stimuli from existing studies with the same aim (Kamitani and Tong, 2005; Freeman et al., 2011, 2013), we tested whether or not it was possible to decode orientation from the unbiased representation using the PCM. Contrary to the predictions of the two prominent theories put forth to explain orientation decoding, both of which assume a biased representation is the source of decodable information, we found orientation was decodable. Further analysis showed the source of information for gratings to be a pattern level response originating from stimulus edges. We then used a multiscale implementation of the PCM to show that spiral sense is also decodable from an unbiased representation. Our instantiation of the model thus has broad explanatory power for a range of studies examining the nature of decodable information in human visual cortex.

MVPA is rapidly becoming a standard tool for the analysis of fMRI data. In many studies, MVPA constitutes a black box, allowing researchers to answer the question of whether or not a representation contains decodable information, while obscuring the information source. Using models that incorporate knowledge of cortical organization and neuronal tuning can reduce the opacity of these methods. For example, progress has been made by using existing knowledge of the receptive field properties of visual neurons to model the response of voxels in visual cortex, yielding more in-depth knowledge of the underlying representa-

tion and also the capacity to decode novel stimuli (Thirion et al., 2006; Dumoulin and Wandell, 2008; Kay et al., 2008; Naselaris et al., 2011). By explicitly characterizing the underlying neural representation, these biologically inspired model-based approaches clarify the relationship between stimulus, representation, and decodable activity. The current study similarly models the neuronal representation to study the nature of the decodable information in visual cortex.

In taking this approach, our study identifies a new source of decodable information for orientation decoding. To date, studies investigating mechanisms of orientation decoding have focused on biases and inhomogeneities in the cortical representation. We found that the edges of stimuli create distortions at the map level that could be used by a classifier to decode orientation. Freeman et al. (2011) was sensitive to the possibility that edges might be a source of decodable information. The authors, however, ruled out edge related activity based on their qualitative observations. They first noted that the heavily blurred stimulus with very weak edges produced similar maps to stimuli with sharper edges, albeit with a reduced effect size. Our modeling results are in agreement with these observations. The heavily blurred stimuli produced distortions similar to the other stimuli (see Fig. 4B–E) and were less decodable (i.e., had lower between stimulus correlations; Fig. 2). Second the authors noted that the relocation of the inner edge of the annulus did not qualitatively change the maps. More specifically they did not see more activation near the fovea corresponding to the radial bias when using a smaller radius inner edge stimulus. The relocation of the edge in the present study clearly shifted the activation patterns (Fig. 3). One reason for this discrepancy might be signal-to-noise is generally lower near the fovea (Schira et al., 2009), thus the effect of the inner edge manipulation might have been obscured in the fMRI data. Another is it might just be a matter of qualitative perspective. Our own qualitative view of their maps is that there is more activity close to the fovea for stimuli with a smaller inner radius.

Several fMRI experiments have described an increase in activity at retinotopic locations corresponding to the radial stimulus orientation, i.e., the radial bias (Sasaki et al., 2006; Mannion et al., 2009; Alink et al., 2013). Our analysis showed that edge related activity generates distortions at the map level that is qualitatively similar to the radial bias. This is potentially a serious concern, as it is possible that the postulation of a radial bias, at least from fMRI studies alone, might entirely be based on edge related activity. At present it's unclear to what extent the fMRI data are a result of a bias in the proportion of neurons representing radial orientations (i.e., a true radial bias) or edge related activity. In support of the radial bias interpretation are several animal studies (Levick and Thibos, 1982; Leventhal, 1983; Schall et al., 1986). The bias observed in these studies may, however, not be observable with fMRI. Oddly enough, this is one of the arguments against the hyperacuity account; that is, fMRI voxels assuredly will have small biases from irregular sampling of orientation columns but that does not necessarily mean we can measure these biases and decode from orientation from them. The individual contribution of a true radial bias and edge related activity to the observed radial bias in fMRI could be resolved in future research by studying the radial bias as function of eccentricity more carefully (Fig. 5) and parceling out the effect that might be ascribed to the edges.

Clifford et al. (2011) argued that the radial bias could not solely explain orientation coding, as log spiral stimuli are radially balanced and decodable (Mannion et al., 2009). Subsequent research has shown that a radial bias may also be sufficient for decoding spiral sense (Freeman et al., 2013). In the present study, we showed that spiral sense is decodable from an unbiased representation, and this could be done at multiple scales and also from an integrated multi scale representation. We additionally found that the deferential response between sense patterns, i.e., decodable information for spiral sense, modulated cyclically with polar angle and that its modulatory nature varied with spatial scale. Freeman et al. (2013) also observed cyclical modulation for spiral sense as a function of polar angle, but the modulation frequency was twice that of the model. They found a global preference for clockwise in the top-left and lower-right visual-field quadrants; and a global preference for counterclockwise in the top-right and lower-left visual-field quadrants. Presently it is not clear to us whether or not there is a relationship between these findings, although the cyclical nature of both results is suggestive. That aside, it is noteworthy that spiral sense can be decoded from an unbiased representation, and thus there is no need to posit a bias either at fine grain or coarse scale representation to account for decoding spiral sense.

Our model results are also compatible with several other more fine-grained observations in the literature. First, a number of studies have taken a filtering approach to investigate the nature of decodable information (Kamitani and Sawahata, 2010; Kriegeskorte et al., 2010; Op de Beeck, 2010; Swisher et al., 2010; Freeman et al., 2011; Alink et al., 2013). Generally these studies found decodable information for orientation is broadband, which has been taken to favor to the global map interpretation (Op de Beeck, 2010; Freeman et al., 2011). Here we have shown evidence that edge related activity is also broadband, which opens up an alternative interpretation of these findings. Second is the observation that the radial bias is more dominant in the periphery (Sasaki et al., 2006; Freeman et al., 2011). Most of these studies have used stimuli with small radius inner annuluses. This, and the low signal-to-noise in the fovea (Schira et al., 2009) would allow the edge related activity in the periphery to dominate. Even if one accepts the radial bias as a dominant factor, edge related activity (mimicking the radial bias) would increase the effect size in the periphery and account for the observed dominance of the radial bias in the periphery. The bias for spiral sense, in contrast, was found to be more dominant closer to the fovea (Freeman et al., 2013). The model might also account for these results. We found that decodable information for spirals was distributed throughout the stimulus. This would predict that decodable information for spirals would be closer to the fovea, at least when referenced to decodable information for gratings.

In summary, the present study has shown that a classic model of the organization of V1 (Hubel and Wiesel, 1972; Hubel, 1988) provides a robust account of the existing research on orientation decoding in fMRI. To be clear, our findings do not directly address the contribution of the imperfect sampling of orientation columns in fMRI voxels (i.e., the hyperacuity account) or the contribution of global biases in retinotopic maps to orientation decoding. Our study identifies a new source of decodable information (edge-related activity) for decoding, and shows that the assumption of biases in fMRI voxels is not needed to account for the capacity to decode gratings and spiral sense.

References

Alink A, Krugliak A, Walther A, Kriegeskorte N (2013) fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. Front Psychol 4:493. CrossRef Medline

Boynton GM (2005) Imaging orientation selectivity: decoding conscious perception in V1. Nat Neurosci 8:541–542. CrossRef Medline

Clifford C, Mannion DJ, Seymour KJ, McDonald JS, Bartels A (2011) Are coarse-scale orientation maps really necessary for orientation decoding? J Neurosci available at: http://www.jneurosci.org/content/31/13/4792/reply

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647–660. CrossRef Medline

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not columns. J Neurosci 31:4792–4804. CrossRef Medline

Freeman J, Heeger DJ, Merriam EP (2013) Coarse-scale biases for spirals and orientation in human visual cortex. J Neurosci 33:19695–19703. CrossRef Medline

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691. CrossRef Medline

Hubel DH (1988) Eye, brain, and vision. New York: WH Freeman.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148:574–591. Medline

Hubel DH, Wiesel TN (1972) Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. J Comp Neurol 146:421–450. CrossRef Medline

Kamitani Y, Sawahata Y (2010) Spatial smoothing hurts localization but not information: pitfalls for brain mappers. Neuroimage 49:1949–1952. CrossRef Medline

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. CrossRef Medline

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352–355. CrossRef Medline

Kriegeskorte N, Cusack R, Bandettini P (2010) How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatio-temporal filter? Neuroimage 49:1965–1976. CrossRef Medline

Kriegeskorte N, Kreiman G, eds (2011) Visual population codes: toward a common multivariate framework for cell recording and functional imaging. Cambridge, MA: MIT.

Leventhal AG (1983) Relationship between preferred orientation and receptive-field position of neurons in cat striate cortex. J Comp Neurol 220:476–483. CrossRef Medline

Levick WR, Thibos LN (1982) Analysis of orientation bias in cat retina. J Physiol 329:243–261. Medline

Mannion DJ, McDonald JS, Clifford CW (2009) Discrimination of the local orientation structure of spiral glass patterns early in human visual cortex. Neuroimage 46:511–515. CrossRef Medline

Mannion DJ, McDonald JS, Clifford CW (2010) Orientation anisotropies in human visual cortex. J Neurophysiol 103:3465–3471. CrossRef Medline

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage 56:400–410. CrossRef Medline

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430. CrossRef Medline

Op de Beeck HP (2010) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage 49:1943–1948. CrossRef Medline

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025. CrossRef Medline

Sasaki Y, Rajimehr R, Kim BW, Ekstrom LB, Vanduffel W, Tootell RB (2006) The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. Neuron 51:661–670. CrossRef Medline

Schall JD, Vitek DJ, Leventhal AG (1986) Retinal constraints on orientation specificity in cat visual cortex. J Neurosci 6:823–836. Medline

Schira MM, Tyler CW, Breakspear M, Spehar B (2009) The foveal confluence in human visual cortex. J Neurosci 29:9050–9058. CrossRef Medline

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T (2007) Robust object recognition with cortex-like mechanisms. Trans Pattern Anal Mach Intell 29:411–426. CrossRef Medline

Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon CH, Kim SG, Tong F (2010) Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. J Neurosci 30:325–330. CrossRef Medline

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S (2006) Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33:1104–1116. CrossRef Medline