Report

Current Biology

Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex

Highlights

- We developed a population receptive field (pRF) model for ventral temporal cortex (VTC)
- Model accurately predicts responses to novel stimulus positions and sizes
- Attention increases pRF gain, eccentricity, and size in high- but not low-level areas
- Attention improves the quality of spatial representations in VTC

Authors

Kendrick N. Kay, Kevin S. Weiner, Kalanit Grill-Spector

Correspondence

kendrick@post.harvard.edu

In Brief

Kay et al. tackle the problem of modeling the final stage in the ventral visual pathway. They do so with a model of visual space, which is surprising since this stage is traditionally viewed as coding form, not space. They show that attention reduces uncertainty in the peripheral visual field where humans and primates are the worst at seeing.

Kay et al., 2015, Current Biology *25*, 1–6
March 2, 2015 ©2015 Elsevier Ltd All rights reserved
http://dx.doi.org/10.1016/j.cub.2014.12.050

Current Biology 25, 1–6, March 2, 2015 ©2015 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.12.050

Report

Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex

Kendrick N. Kay,[1,*] **Kevin S. Weiner,**[2]
and Kalanit Grill-Spector[2,3]
[1]Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA
[2]Department of Psychology, Stanford University, Stanford, CA 94305, USA
[3]Stanford Neurosciences Institute, Stanford University, Stanford, CA 94305, USA

Summary

Ventral temporal cortex (VTC) is the latest stage of the ventral "what" visual pathway, which is thought to code the identity of a stimulus regardless of its position or size [1, 2]. Surprisingly, recent studies show that position information can be decoded from VTC [3–5]. However, the computational mechanisms by which spatial information is encoded in VTC are unknown. Furthermore, how attention influences spatial representations in human VTC is also unknown because the effect of attention on spatial representations has only been examined in the dorsal "where" visual pathway [6–10]. Here, we fill these significant gaps in knowledge using an approach that combines functional magnetic resonance imaging and sophisticated computational methods. We first develop a population receptive field (pRF) model [11, 12] of spatial responses in human VTC. Consisting of spatial summation followed by a compressive nonlinearity, this model accurately predicts responses of individual voxels to stimuli at any position and size, explains how spatial information is encoded, and reveals a functional hierarchy in VTC. We then manipulate attention and use our model to decipher the effects of attention. We find that attention to the stimulus systematically and selectively modulates responses in VTC, but not early visual areas. Locally, attention increases eccentricity, size, and gain of individual pRFs, thereby increasing position tolerance. However, globally, these effects reduce uncertainty regarding stimulus location and actually increase position sensitivity of distributed responses across VTC. These results demonstrate that attention actively shapes and enhances spatial representations in the ventral visual pathway.

Results

Does a pRF Model Predict Responses in VTC?

To develop a model of how spatial information is encoded in ventral temporal cortex (VTC), we measured fMRI responses (3T, 2-mm voxels) in a series of face-selective regions [13] while subjects fixated centrally and viewed images of faces that varied systematically in position and size (Figure 1A). We used face-selective regions as a model system as they are a highly studied subsystem of VTC [3, 14, 15] with a well-understood functional organization that is anatomically consistent across subjects [13, 16]. After estimating and denoising stimulus-evoked responses [17], we modeled responses in each voxel using the compressive spatial summation (CSS) model [12]. The CSS model characterizes the population receptive field (pRF) [11] of a voxel and predicts the response to a face by first computing the spatial overlap between the face and an isotropic 2D Gaussian and then applying a compressive nonlinearity (Figure 1B). Cross-validation analyses demonstrate that the CSS model accurately characterizes responses of individual voxels in face-selective regions located on the inferior occipital gyrus (IOG), posterior fusiform gyrus (pFus), and mid-fusiform gyrus (mFus) [13] and successfully predicts responses to faces at novel positions and sizes (Figures 1C and S1A). To assess whether these results are specific to face stimuli, we also performed measurements using phase-scrambled faces. Although phase-scrambled faces evoke weaker responses and produce noisier pRF estimates, pRF properties are largely invariant to stimulus type (Figures S1B and S1C).

What Is the Nature of pRFs in VTC?

Similar to early and intermediate visual areas [11, 12], pRF size increases with eccentricity in face-selective regions within VTC (Figures 2A and S2B), suggesting that size-eccentricity scaling is a pervasive organizing principle across the ventral visual pathway. However, different from earlier visual areas, pRFs in face-selective regions are quite large compared to their eccentricity. Consequently, these pRFs extend substantially into the ipsilateral visual field (Figure 2B). Also, unlike pRFs in earlier areas, pRFs in face-selective regions are consistently centered near the fovea, producing a representational scheme in which nearly all neural resources are dedicated to the central portion of the visual field (approximately the central 7°; see Figures 2B and S2A). This convergence of spatial coverage is consistent with the foveal bias of face-selective regions [14, 15]. Notably, this organization is different from the distributed tiling of visual space in earlier retinotopic visual regions [18], suggesting unique computational strategies in VTC. Interestingly, pRF properties vary hierarchically across face-selective regions: anterior regions in VTC generally have larger and more foveal pRFs than posterior regions (Figures 2A and S2C), features also observed in monkey inferotemporal cortex (IT) [19–22].

How Are pRF Properties Affected by Attention?

To understand the contribution of top-down attentional signals to the observed results, we measured pRFs under different attentional states. While maintaining central fixation, subjects performed one of three tasks: digit task (one-back task on rapid serial presentation of digits at fixation), dot task (detection of a red dot appearing on the faces; same as the first experiment), and face task (one-back task on the identity of the faces; see Supplemental Experimental Procedures and Figure S3A). In the digit task, attention is directed toward fixation, whereas in the dot and face tasks, attention is directed toward the faces.

Comparing pRF properties across tasks, we find no substantial changes in pRF properties in early visual areas V1–V3 (Figures 3A and S3C). However, in hV4 and, more substantially, in face-selective regions, voxel responses are strongly modulated by the task (Figure S3B). In these regions,

*Correspondence: kendrick@post.harvard.edu

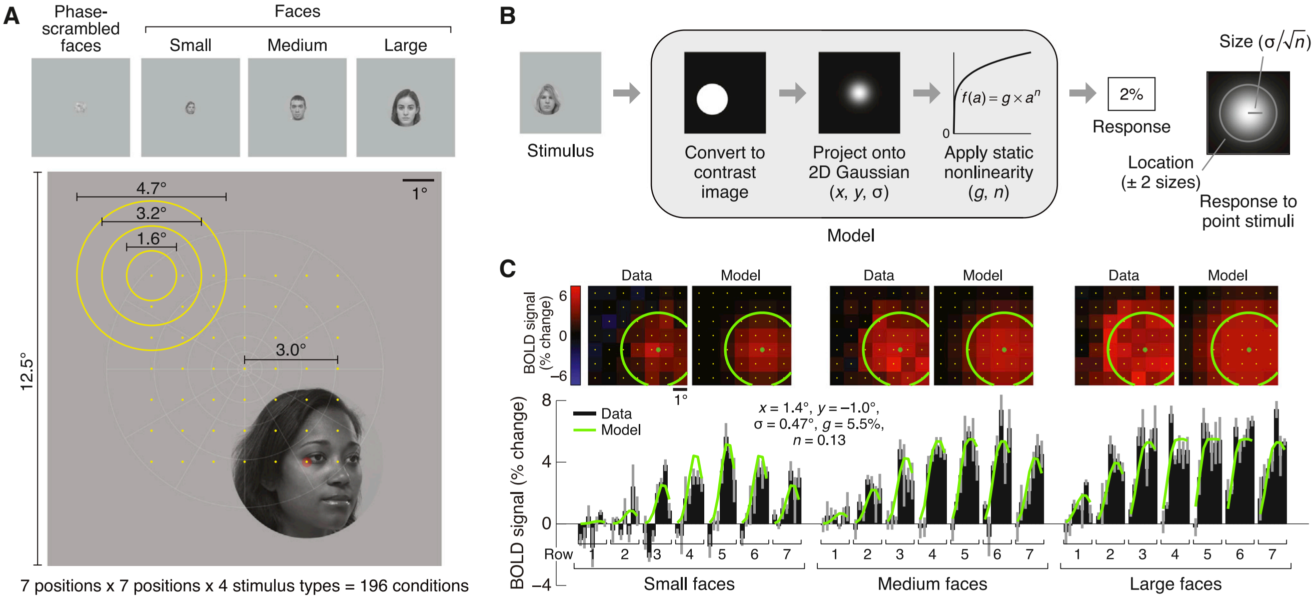


Figure 1. Compressive Spatial Summation Accurately Models Responses in VTC

(A) Stimuli. Subjects viewed faces while fixating centrally. Faces varied systematically in position (centers indicated by yellow dots) and size (sizes indicated by yellow circles). During each trial, face position and size were held constant while face identity and viewpoint were dynamically updated.

(B) Compressive spatial summation (CSS) model. The response to a face is predicted by computing the spatial overlap between the face and a 2D Gaussian and then applying a compressive power-law nonlinearity. The model includes two parameters (x, y) for the position of the Gaussian, a parameter (σ) for the size of the Gaussian, a parameter (n) for the exponent of the nonlinearity, and a parameter (g) for the overall gain of the predicted responses.

(C) Example voxel (left IOG, subject 1). Top row: responses arranged spatially according to face position. Bottom row: responses arranged serially for better visualization of measurement reliability and goodness of fit of CSS model. Blood-oxygenation-level-dependent (BOLD) response magnitudes (black bars; median across trials ± 68% confidence interval [CI]) are accurately fit by the model (green line). Note that a single set of model parameters accounts for the full range of the data.

See also Figure S1.

pRFs exhibit increased eccentricity, size, and gain when subjects attend to the faces (dot and face tasks) compared to when they attend to fixation (digit task) (Figures 3A–3C). These effects are consistent with the concept of response enhancement at the attended location [23], and the effects are large in size: for example, in mFus, comparing pRF properties across the digit and face task, respectively, the median pRF eccentricity increases from 1.3° to 1.9°, the median pRF size increases from 1.8° to 3.4°, and the median pRF gain increases from 0.83% to 1.32%.

Control experiments reveal that changes in pRF properties are observed even if tasks are interleaved on a trial-by-trial basis (Figure S3G), indicating that the changes cannot be attributed to variation in general subject arousal across tasks. Furthermore, performing the digit task on digits presented to the left of fixation produces leftward shifts of pRFs in hV4, IOG, and pFus compared to performing the digit task on central digits (Figure S3H). This indicates that even though attention is drawn away from faces during the digit task, pRF modulations occur in a manner consistent with response enhancement at the attended location, irrespective of the content of the attended stimulus.

Interestingly, attentional effects in face-selective regions are stronger for the face task, which specifically requires perceptual processing of the faces, compared to the dot task ($p < 10^{-9}$, two-tailed sign test in each region for each pRF property). Increases in pRF size under the dot and face tasks relative to the digit task are particularly intriguing as they indicate that locally, at the voxel level, attention to the stimulus increases the position tolerance of the neural representation.

What Is the Benefit of Attentional Modulation of pRFs?

Although we have demonstrated local changes in pRF properties as a result of attention, an open question is whether these attention-induced changes are beneficial to the global, or distributed, representation of the stimulus. Specifically, we ask the following question: does attention affect the ability of a collection of pRFs to discriminate the location of the stimulus? This question cannot be answered through simple summary statistics of pRF properties (such as the ones in Figure 3A) because discrimination performance depends not only on the properties of individual pRFs but also on how the pRFs collectively tile the visual field. For example, large but overlapping pRFs might discriminate stimulus locations better than small, non-overlapping pRFs [24]. We therefore designed a model-based decoding analysis that quantifies the spatial discrimination performance of a collection of pRFs. In this analysis, we calculate spatial uncertainty, that is, the distance over which changes in stimulus position cannot be well discriminated based on the distributed responses across the pRFs (thus, low spatial uncertainty indicates good discrimination performance). We applied this analysis separately to each region, analyzing the pRFs observed under each task.

As expected from the stability of pRF properties in early visual areas, there is little change in spatial uncertainty in these areas across tasks (Figure 4A, top). In all tasks, spatial uncertainty in V1–V3 is less than 0.5° near the fovea (1° eccentricity) and less than 1.5° in the periphery (5° eccentricity) (Figures 4B and 4C). In contrast, there are large changes in spatial uncertainty in face-selective regions across tasks. In the periphery, spatial uncertainty is substantially reduced under the dot task

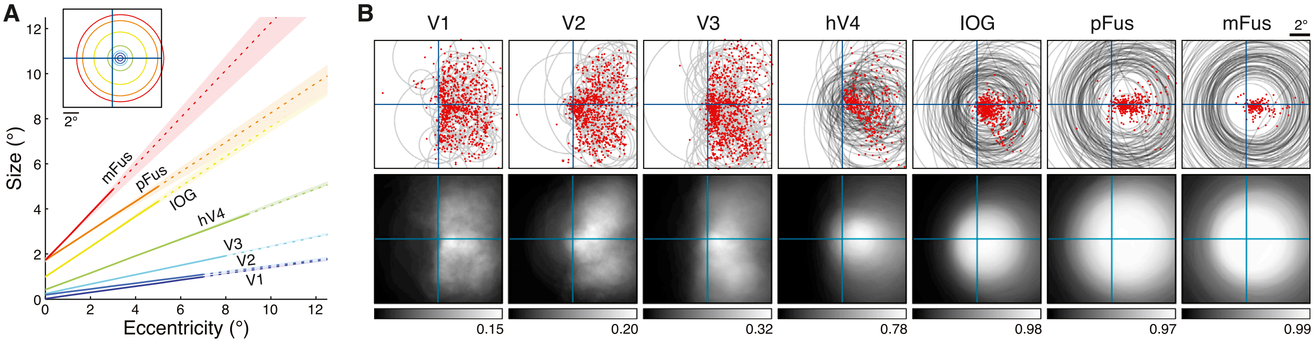


Figure 2. Systematic Organization of pRF Properties across the Ventral Visual Pathway

(A) pRF size versus eccentricity. Each line represents a region (median across voxels ± 68% CI). Dotted lines indicate eccentricity ranges containing few voxels. The inset shows a schematic of pRF sizes at 1° eccentricity. IOG, inferior occipital gyrus; pFus, posterior fusiform; mFus, mid-fusiform.

(B) pRF locations and visual field coverage in the left hemisphere. Top row: pRF centers (red dots) and pRF sizes for 100 randomly selected voxels from each region (gray circles). Bottom row: visual field coverage, computed as the proportion of pRFs covering each point in the visual field. Each image is normalized such that black corresponds to 0, and white corresponds to the maximum value. IOG, pFus, and mFus contain large pRFs centered near the fovea.
See also Figure S2.

(1.9-fold reduction, on average, across face-selective regions) and the face task (2.7-fold reduction, on average, across regions) compared to the digit task (Figure 4A, bottom; Figure 4C). For example, in mFus, uncertainty in the periphery is more than 3° under the digit task but only about 1° under the face task (Figure 4C). Importantly, these improvements are not simply due to increased pRF gain: improvements are still observed if pRF gain is held constant and only the task-induced changes in pRF location and size are considered (Figure S4A). These results indicate that attending the stimulus either explicitly (face task) or implicitly (dot task) reduces uncertainty with respect to the location of the stimulus. As a complement to our model-based decoding analysis, we also performed direct decoding of the distributed response patterns evoked by faces with no intervening modeling step. Results are consistent with our model-based analysis: in face-selective regions, there is improved decoding of face position in the periphery under the face task compared to the digit and dot tasks (Figures S4B and S4C).

Discussion

The experiments in the present study reveal that spatial representations are prevalent in the ventral “what” visual pathway. First, we have shown that responses in VTC are modulated by changes in the position and size of the stimulus. These modulations are systematic and are accurately characterized by a pRF model utilizing spatial summation and a compressive nonlinearity. Second, spatial representations within VTC are actively shaped by top-down task demands. Specifically, attention modulates pRFs in high and intermediate levels of the ventral pathway, but not early visual regions. While prior research has shown that spatial attention shifts receptive fields in the dorsal “where” visual pathway [6–9, 25], as well as intermediate visual areas in the ventral pathway [23, 26–28], we extend these results to high-level areas in the ventral pathway for the first time.

Attentional Effects in the Ventral Visual Pathway

The observed attentional modulations of pRFs are consistent with the theory that neural responses in visual cortex reflect the combination of bottom-up stimulus drive and a top-down attentional field that enhances responses to stimuli at the current locus of attention [23, 26, 29]. While both implicit (dot task) and explicit (face task) attention toward faces lead to response enhancement, we find that explicit attention toward faces produces larger modulations (see Figure 3A). This suggests that responses in the ventral visual pathway are modulated by both spatial and object-based attention, consistent with recent demonstrations of category-based attentional effects in the ventral pathway [30]. An interesting subject for future work is examining whether the attentional modulations observed here can be quantitatively described as an interaction between a global attentional field [7, 10, 23] and local classical receptive fields. Recent data suggest that the effect of a global attentional field on pRFs depends on pRF size, with larger effects obtained for larger pRFs [10]. Thus, these models predict larger attentional shifts of pRFs at higher stages of the visual processing hierarchy. We facilitate efforts to examine such questions and to further develop attentional models by making our data publicly available (http://kendrickkay.net/vtcdata/).

One question that stems from our findings is whether the demonstrated impact of attention on cortical responses has behavioral consequences. We hypothesize that attention-induced changes in the representation of spatial information in VTC may affect behavioral judgments of spatial position. Specifically, reduction of neural spatial uncertainty during the dot and face tasks compared to the digit task suggests that behavioral judgments of face position would be more accurate during the dot and face tasks. This hypothesis can be tested in future behavioral studies.

Another open question is exactly how the attentional modulations measured with fMRI manifest at the level of individual neurons. As prior electrophysiological studies have demonstrated that attention modulates neuronal firing rates in monkey IT [26, 28], we hypothesize that similar attentional modulations of receptive fields (RFs) occur for individual neurons in the ventral visual pathway. Notably, our observation of task-dependent pRFs might explain the variability of previous reports of neuronal RF sizes in monkey IT: RFs were largest during passive viewing [31, 32] and anesthesia [19], whereas RFs were smallest during demanding discrimination tasks near the fovea [33].

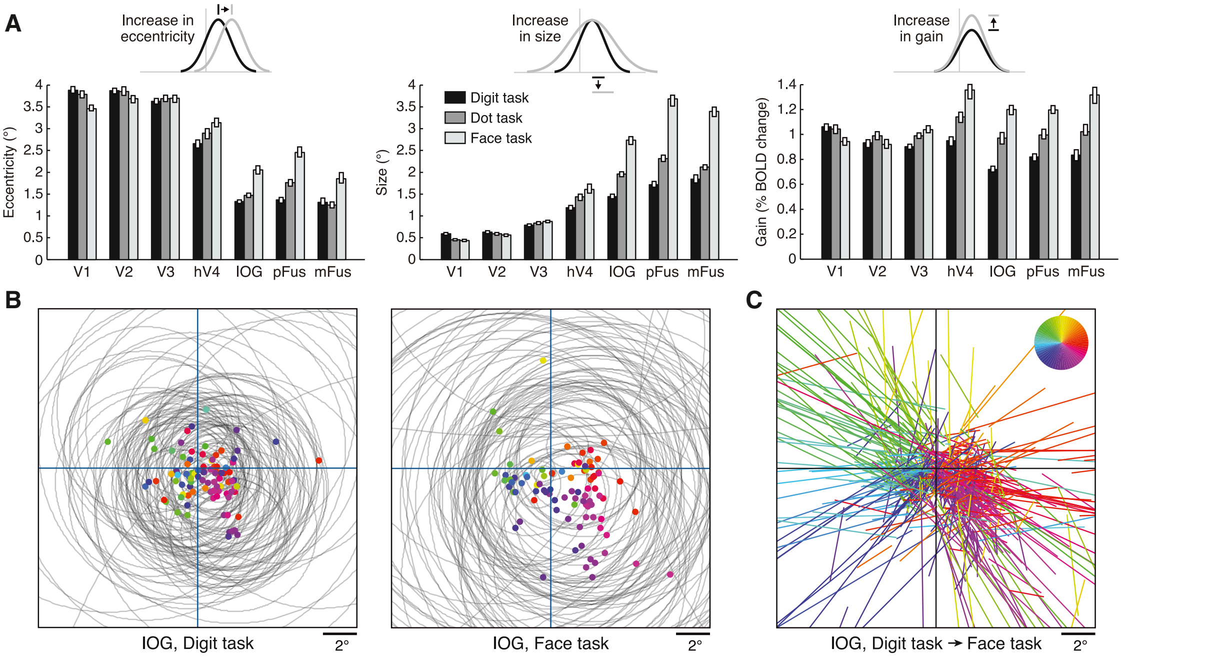


Figure 3. Attention Modulates pRF Properties in VTC

pRFs were measured under three tasks using the same stimulus. While maintaining fixation, subjects performed a one-back task on centrally presented digits (digit task), detected a small dot superimposed on the faces (dot task), or performed a one-back task on face identity (face task).

(A) Summary of results. Each bar represents a region under a single task (median across voxels ± 68% CI). In hV4, and more so in IOG, pFus, and mFus, attending to the stimulus (dot task, face task) causes an increase in pRF eccentricity, size, and gain compared to attending to fixation (digit task). These effects are larger for the face task than for the dot task.

(B) Visualization of pRFs for 100 randomly selected voxels from an example region, bilateral IOG (colored dots represent pRF centers; gray circles represent pRF sizes). In the digit task, pRFs are small and cluster near the fovea, whereas in the face task, pRFs are large and spread out into the periphery.

(C) Visualization of pRF shifts for region IOG. For each voxel, a line is drawn that connects the pRF center under the digit task to the pRF center under the face task; color indicates the direction of the shift (see legend), and the same color is used for the corresponding dots in (B). In general, pRFs shift away from the center. Although it appears as if there are many shifts to far eccentricities, the majority (81%) of pRF centers under the face task are actually located within 5° eccentricity.

See also Figure S3.

Rethinking Position Tolerance in the Ventral Visual Pathway

Position and size tolerance are considered key features of the ventral visual pathway, useful for object and face recognition. Tolerance indicates reduced sensitivity to incidental properties of a stimulus, such as the specific position or size at which it is viewed [34, 35]. Prevailing theories suggest that tolerance is achieved by systematic increase in RF sizes across processing stages in the ventral visual pathway [19, 20, 22, 36, 37]. Intuitively, a large RF implies that a wide range of stimulus positions and sizes drives the neural response [12].

Although our pRF measurements are consistent with this account and reveal a hierarchy of pRF sizes within VTC, there are two aspects of our data that prompt a rethinking of position tolerance in VTC. First, we show that position tolerance at the level of individual voxels is partially the result of top-down attentional mechanisms and not simply due to static RF properties (see also [38]). Specifically, we find that when subjects attend to the stimulus, pRFs enlarge, thereby increasing position tolerance. Second, we show that the common intuition that larger pRFs degrade spatial information may be misleading. Despite the enlargement of pRFs when subjects attend the stimulus, the spatial precision with which the location of the stimulus is represented in VTC improves, rather than worsens.

At first glance, these observations seem inconsistent: how can attention increase spatial tolerance while also increasing spatial precision? The answer lies in the distinction between the local scale (i.e., information carried by a single voxel) and the global scale (i.e., information carried by distributed responses across voxels). At the local scale, each individual voxel shows reduced sensitivity to stimulus location due to increased pRF size. However, at the global scale, sensitivity to stimulus location improves due to increased pRF coverage and scatter in the periphery (Figures 3B and 3C), which together provide a better tiling of the visual field.

Conclusions

We have used a model-based approach to understand how attention influences representation in visual cortex. Our approach consisted of measuring responses to a wide range of stimulus conditions [39, 40], developing an encoding model that describes how stimulus information is represented locally [11, 41, 42], and using decoding analyses to quantify the information present in distributed responses [43]. Importantly, although we implemented this approach with fMRI, the approach is general and can be applied to other experimental techniques, such as electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), and electrophysiology. Comparing results from different experimental techniques in a common model-based framework may help elucidate the neural signals measured by different techniques [44] and may help resolve discrepancies in the sizes

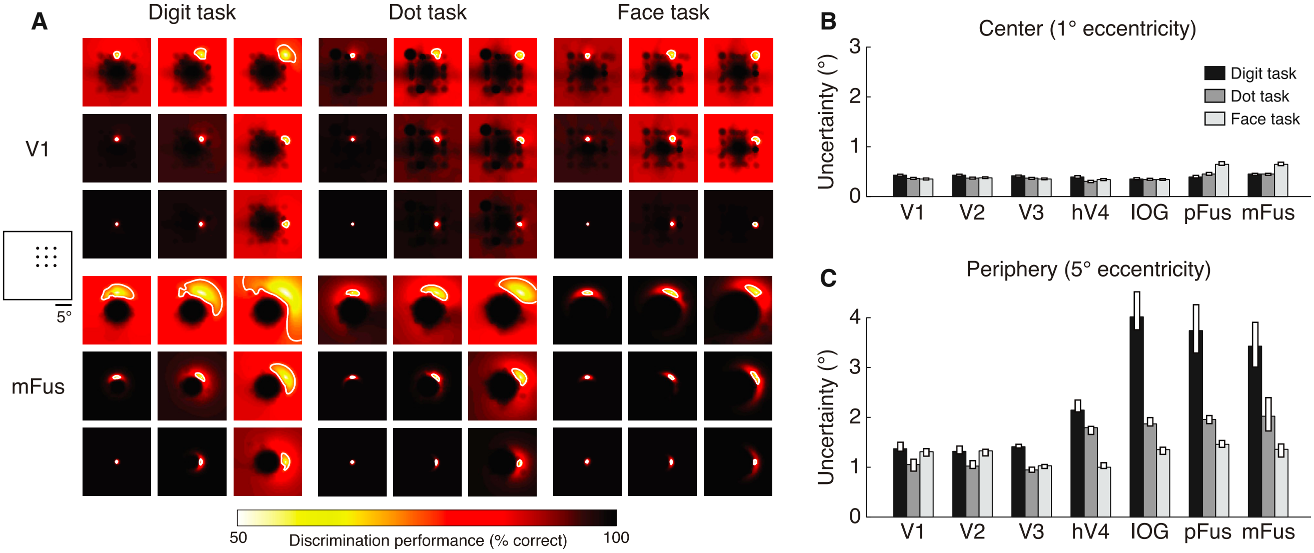


Figure 4. Attention Reduces Spatial Uncertainty in VTC

For each region and task, we assess the quality of the representation of spatial information using a model-based decoding analysis. This analysis quantifies how well a linear classifier can discriminate stimuli at different visual field positions from a stimulus at a reference position.

(A) Example results for a 3 × 3 grid of reference positions in the upper-right visual field (left inset). Each image is a map of discrimination performance for one reference position (indicated by the relative position of the image). We define spatial uncertainty as the square root of the area of the 75% correct contour (white line).

(B) Uncertainty at 1° eccentricity. Each bar represents uncertainty in a region under a single task (median across angular positions ± 68% CI). All regions exhibit low uncertainty, irrespective of the task.

(C) Uncertainty at 5° eccentricity. Face-selective regions IOG, pFus, and mFus exhibit high uncertainty under the digit task. However, this uncertainty is dramatically reduced under the dot and face tasks.

See also Figure S4.

of attentional effects found by different techniques [45]. Overall, our study reveals that spatial information is systematically represented in the ventral visual pathway and that attention modifies and enhances this spatial representation. These results provide important insights into how position coding is implemented in the ventral visual pathway.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.12.050.

Author Contributions

K.N.K. conducted the experiment and analyzed the data. K.S.W. performed localizer experiments and assisted with data collection. K.N.K., K.S.W., and K.G.-S. conceived and designed the experiments. K.N.K., K.S.W., and K.G.-S. wrote the paper.

Acknowledgments

We thank T. Naselaris, A. Rokem, J. Winawer, and E. Zohary for helpful discussions. We also thank A. Stigliani for assistance with behavioral experiments; J. Winawer for providing retinotopic mapping data; and N. Witthoft for assisting in the collection of face photographs. This work was supported by the McDonnell Center for Systems Neuroscience and Arts & Sciences at Washington University (K.N.K.), NEI grant 1R01EY02391501A1 (K.G.-S.), and NEI grant RO1 EY03164 (to Brian Wandell). Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially provided through grant NCRR 1S10RR022984-01A1.

Received: September 7, 2014
Revised: November 17, 2014
Accepted: December 18, 2014
Published: February 19, 2015

References

1. Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds. (MIT Press), pp. 549–586.
2. Goodale, M.A., Milner, A.D., Jakobson, L.S., and Carey, D.P. (1991). Object awareness. Nature *352*, 202.
3. Schwarzlose, R.F., Swisher, J.D., Dang, S., and Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. Proc. Natl. Acad. Sci. USA *105*, 4447–4452.
4. Kravitz, D.J., Kriegeskorte, N., and Baker, C.I. (2010). High-level visual object representations are constrained by position. Cereb. Cortex *20*, 2916–2925.
5. Carlson, T., Hogendoorn, H., Fonteijn, H., and Verstraten, F.A.J. (2011). Spatial coding and invariance in object-selective cortex. Cortex *47*, 14–22.
6. Silver, M.A., Ress, D., and Heeger, D.J. (2005). Topographic maps of visual spatial attention in human parietal cortex. J. Neurophysiol. *94*, 1358–1371.
7. Sprague, T.C., and Serences, J.T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat. Neurosci. *16*, 1879–1887.
8. Szczepanski, S.M., Konen, C.S., and Kastner, S. (2010). Mechanisms of spatial attention control in frontal and parietal cortex. J. Neurosci. *30*, 148–160.
9. Saproo, S., and Serences, J.T. (2010). Spatial attention improves the quality of population codes in human visual cortex. J. Neurophysiol. *104*, 885–895.
10. Klein, B.P., Harvey, B.M., and Dumoulin, S.O. (2014). Attraction of position preference by spatial attention throughout human visual cortex. Neuron *84*, 227–237.
11. Dumoulin, S.O., and Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. Neuroimage *39*, 647–660.
12. Kay, K.N., Winawer, J., Mezer, A., and Wandell, B.A. (2013). Compressive spatial summation in human visual cortex. J. Neurophysiol. *110*, 481–494.

13. Weiner, K.S., and Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage *52*, 1559–1573.
14. Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. Nat. Neurosci. *4*, 533–539.
15. Yue, X., Cassidy, B.S., Devaney, K.J., Holt, D.J., and Tootell, R.B.H. (2011). Lower-level stimulus features strongly influence responses in the fusiform face area. Cereb. Cortex *21*, 35–47.
16. Weiner, K.S., Golarai, G., Caspers, J., Chuapoco, M.R., Mohlberg, H., Zilles, K., Amunts, K., and Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. Neuroimage *84*, 453–465.
17. Kay, K.N., Rokem, A., Winawer, J., Dougherty, R.F., and Wandell, B.A. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci *7*, 247.
18. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.
19. Gross, C.G., Bender, D.B., and Rocha-Miranda, C.E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science *166*, 1303–1306.
20. Boussaoud, D., Desimone, R., and Ungerleider, L.G. (1991). Visual topography of area TEO in the macaque. J. Comp. Neurol. *306*, 554–575.
21. Op De Beeck, H., and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. J. Comp. Neurol. *426*, 505–518.
22. Issa, E.B., and DiCarlo, J.J. (2012). Precedence of the eye region in neural processing of faces. J. Neurosci. *32*, 16666–16682.
23. Reynolds, J.H., and Heeger, D.J. (2009). The normalization model of attention. Neuron *61*, 168–185.
24. Snippe, H.P., and Koenderink, J.J. (1992). Discrimination thresholds for channel-coded systems. Biol. Cybern. *66*, 543–551.
25. Treue, S., and Maunsell, J.H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature *382*, 539–541.
26. Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science *229*, 782–784.
27. Connor, C.E., Preddie, D.C., Gallant, J.L., and Van Essen, D.C. (1997). Spatial attention effects in macaque area V4. J. Neurosci. *17*, 3201–3214.
28. Richmond, B.J., Wurtz, R.H., and Sato, T. (1983). Visual responses of inferior temporal neurons in awake rhesus monkey. J. Neurophysiol. *50*, 1415–1432.
29. Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.
30. Çukur, T., Nishimoto, S., Huth, A.G., and Gallant, J.L. (2013). Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. *16*, 763–770.
31. Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. *4*, 2051–2062.
32. Tovee, M.J., Rolls, E.T., and Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J. Neurophysiol. *72*, 1049–1060.
33. DiCarlo, J.J., and Maunsell, J.H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J. Neurophysiol. *89*, 3264–3278.
34. DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434.
35. Poggio, T., and Ullman, S. (2013). Vision: are models of object recognition catching up with the brain? Ann. N Y Acad. Sci. *1305*, 72–82.
36. Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA *111*, 8619–8624.
37. Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. USA *104*, 6424–6429.
38. Olshausen, B.A., Anderson, C.H., and Van Essen, D.C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J. Neurosci. *13*, 4700–4719.
39. Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.
40. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron *60*, 1126–1141.
41. Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. Neuroimage *56*, 400–410.
42. Kay, K.N. (2011). Understanding visual representation by developing receptive-field models. In Visual Population Codes: Towards a Common Multivariate Framework for Cell Recording and Functional Imaging, N. Kriegeskorte and G. Kreiman, eds. (MIT Press), pp. 133–162.
43. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.
44. Winawer, J., Kay, K.N., Foster, B.L., Rauschecker, A.M., Parvizi, J., and Wandell, B.A. (2013). Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. Curr. Biol. *23*, 1145–1153.
45. Boynton, G.M. (2011). Spikes, BOLD, attention, and awareness: a comparison of electrophysiological and fMRI signals in V1. J. Vis. *11*, 12.

Current Biology

Supplemental Information

Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex

Kendrick N. Kay, Kevin S. Weiner, and Kalanit Grill-Spector

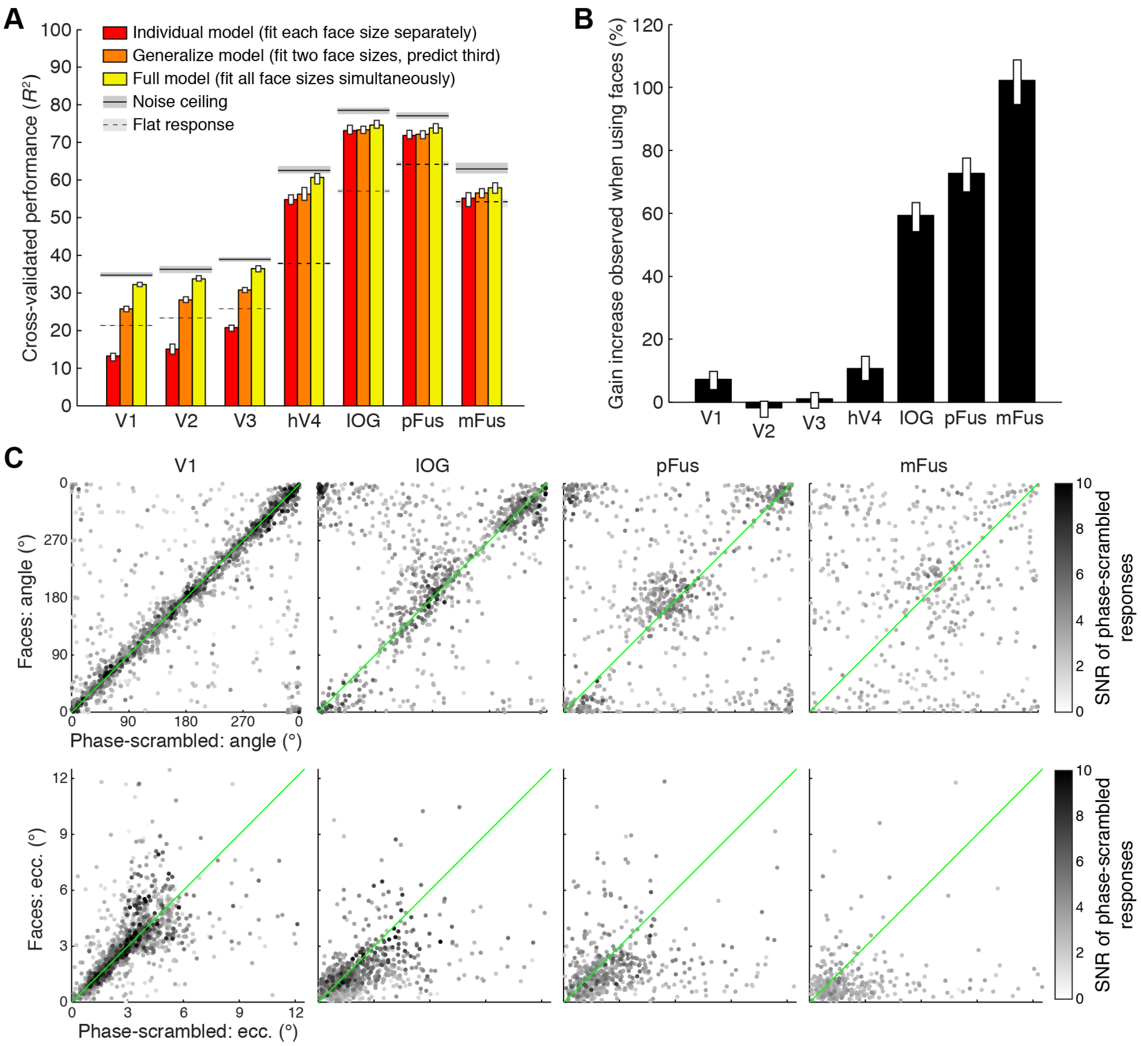


Figure S1, Related to Figure 1. Examination of cross-validation performance and stimulus dependence of CSS model. (A) Cross-validation performance. The *Full model* is fit to all face sizes simultaneously and is asked to predict a few held-out face positions on each cross-validation iteration. This model performs quite well and approaches the noise ceiling in each region. The *Generalize model* is fit to two face sizes and is asked to predict the held-out face size. Performance is again quite high, indicating the model's ability to predict responses to entirely novel face sizes. Finally, the *Individual model* is fit and cross-validated on each face size separately. The performance of this model is lower than that of the Full model, indicating that a single model (the Full model) is sufficient to account for responses to different face sizes. The low noise ceilings in V1–V3 are due to the fact that voxels in these areas have small pRFs and respond to few stimulus conditions. Bars indicate median across voxels ± 68% CI. (B) Faces elicit substantially stronger responses compared to phase-scrambled faces in face-selective regions, but not early visual areas. Increase in gain is quantified as 100 x (*F* – *P*) / *P* where *F* and *P* refer to the gain observed for faces and phase-scrambled faces, respectively. Bars indicate median across voxels ± 68% CI. (C) Faces and phase-scrambled faces produce similar angle and eccentricity estimates. Each dot represents a voxel, and the gray-level of each dot indicates the signal-to-noise ratio (SNR) of the phase-scrambled responses, defined as the maximum response amplitude divided by the average error (dark indicates high SNR). When SNR is high, similar angle and eccentricity estimates are obtained. This suggests that dissimilarities in estimates are simply due to low SNR. These results indicate that pRFs in face-selective regions can, in theory, be estimated using stimuli other than faces but that, in practice, responses may be weak, leading to noisy pRF estimates.

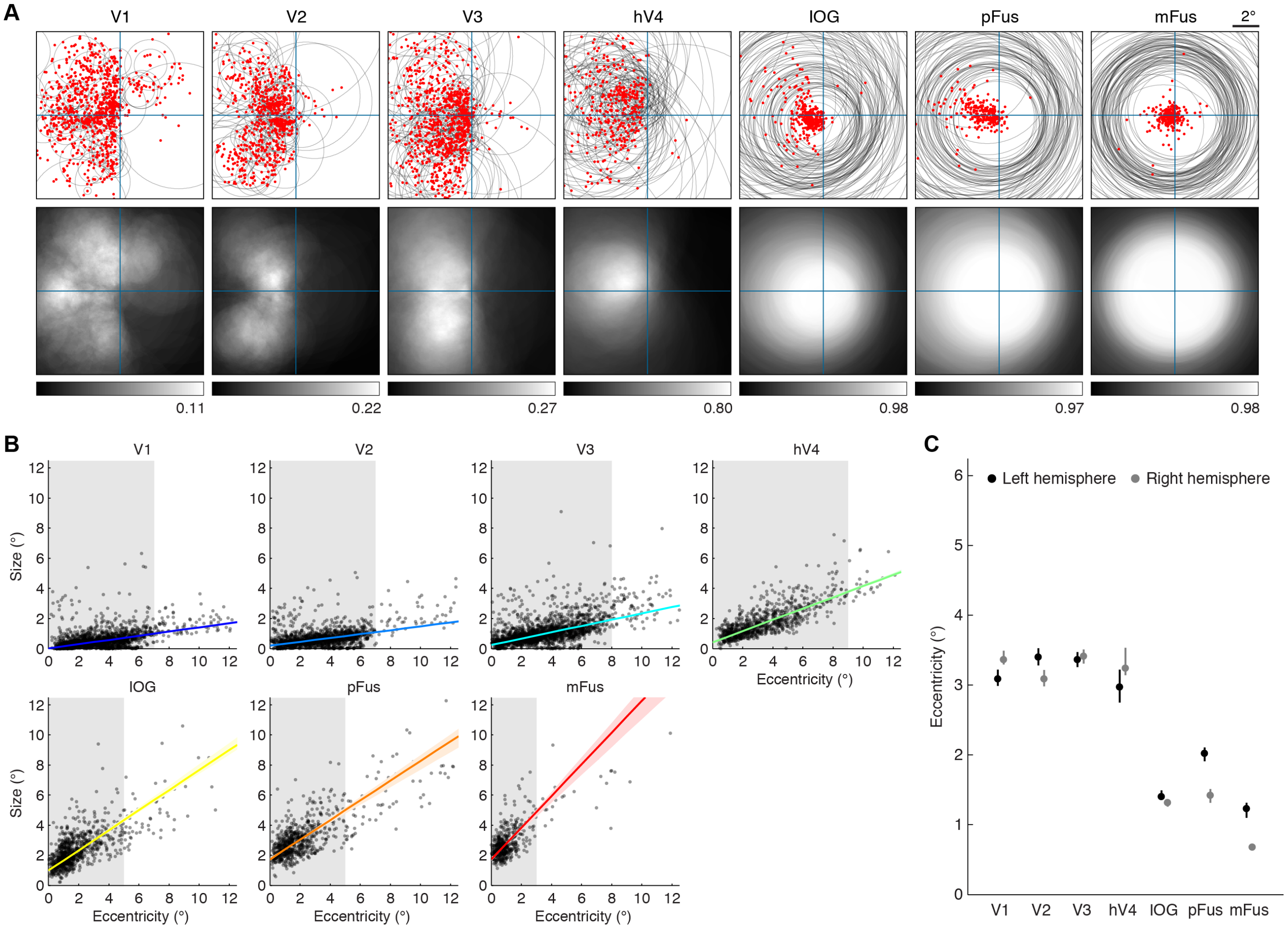
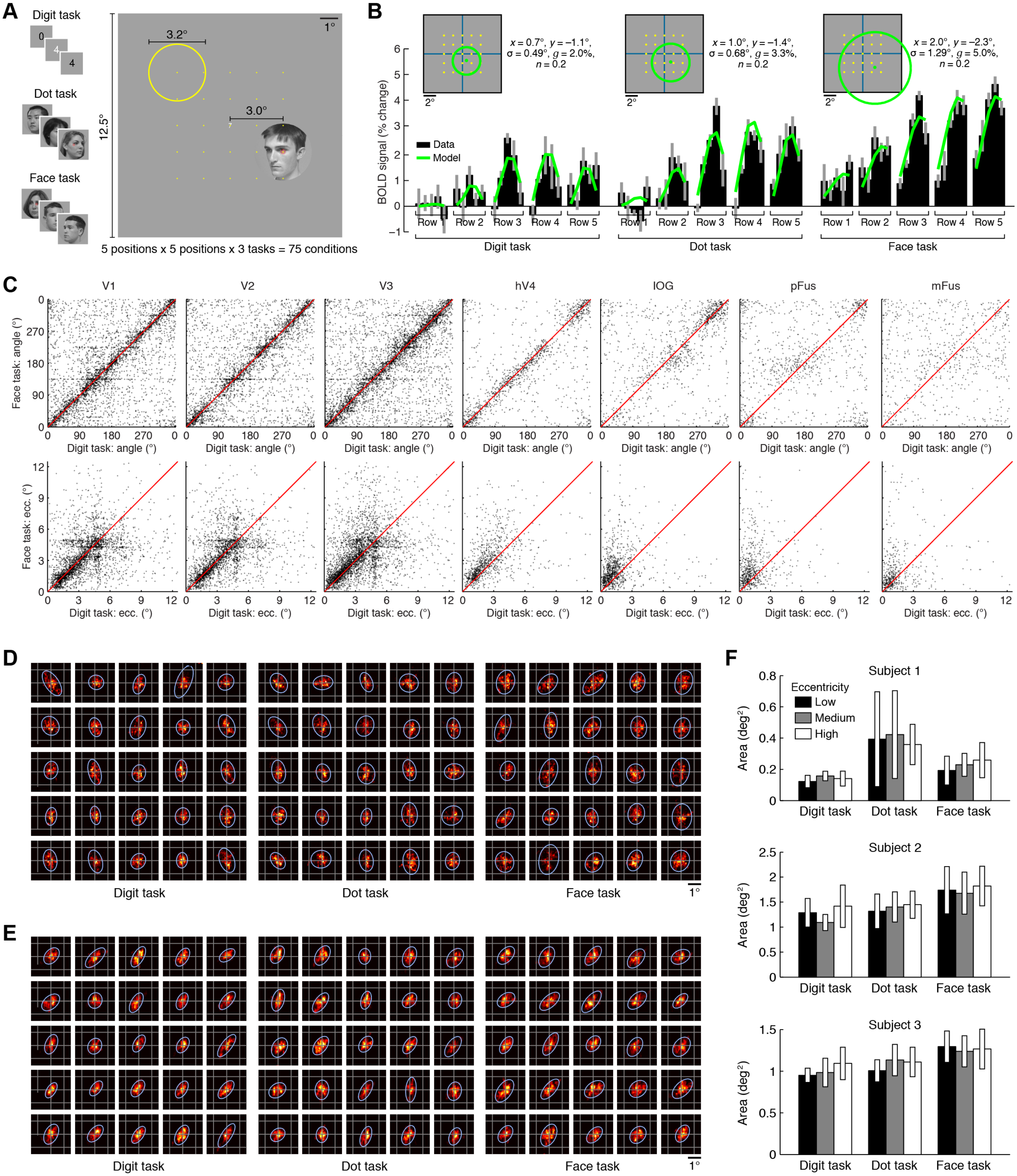


Figure S2, Related to Figure 2. Additional quantification of pRF properties. (A) pRF locations and visual field coverage in the right hemisphere. Same format as Figure 2B. pRFs centers are located in the contralateral visual field and become larger and more foveal in anterior regions. (B) Scatterplots of pRF eccentricity versus size. Individual voxels underlying the lines relating pRF eccentricity and size in Figure 2A are plotted for each region (combining across hemispheres). Each dot corresponds to one voxel, and dots are partially transparent. For each region, we fit a line minimizing the sum of the perpendicular distances between the voxels and the line. The fitting procedure is bootstrapped to estimate error (68% CI), indicated by the band around each line (some bands are too small to be visible). To summarize the range of eccentricities observed in each region, we count voxels in 1° eccentricity bins (0–1°, 1–2°, etc.) and identify bins with at least 10% of the number of voxels in the largest bin. These bins are indicated by gray shading and correspond to the solid lines in Figure 2A. (C) Summary of pRF eccentricity. Dots indicate the median eccentricity across voxels in each region ± 68% CI. In face-selective regions, pRF centers are typically located at foveal eccentricities. Right pFus and mFus exhibit a stronger foveal bias compared to left pFus and mFus.


A
Digit task
0
4
4
Dot task
Face task
3.2°
1°
3.0°
12.5°
5 positions x 5 positions x 3 tasks = 75 conditions
B
$x = 0.7°$, $y = -1.1°$, $\sigma = 0.49°$, $g = 2.0\%$, $n = 0.2$
$x = 1.0°$, $y = -1.4°$, $\sigma = 0.68°$, $g = 3.3\%$, $n = 0.2$
$x = 2.0°$, $y = -2.3°$, $\sigma = 1.29°$, $g = 5.0\%$, $n = 0.2$
2°
2°
2°
BOLD signal (% change)
6
5
4
3
2
1
0
−1
Data
Model
Row 1
Row 2
Row 3
Row 4
Row 5
Digit task
Dot task
Face task
C
V1
V2
V3
hV4
IOG
pFus
mFus
Face task: angle (°)
0
270
180
90
0
Digit task: angle (°)
0
90
180
270
0
Face task: ecc. (°)
12
9
6
3
0
Digit task: ecc. (°)
0
3
6
9
12
D
Digit task
Dot task
Face task
1°
E
Digit task
Dot task
Face task
1°
F
Subject 1
Eccentricity
Low
Medium
High
Area (deg²)
0.8
0.6
0.4
0.2
0
Digit task
Dot task
Face task
Subject 2
Area (deg²)
2.5
2
1.5
1
0.5
0
Digit task
Dot task
Face task
Subject 3
Area (deg²)
1.5
1
0.5
0
Digit task
Dot task
Face task

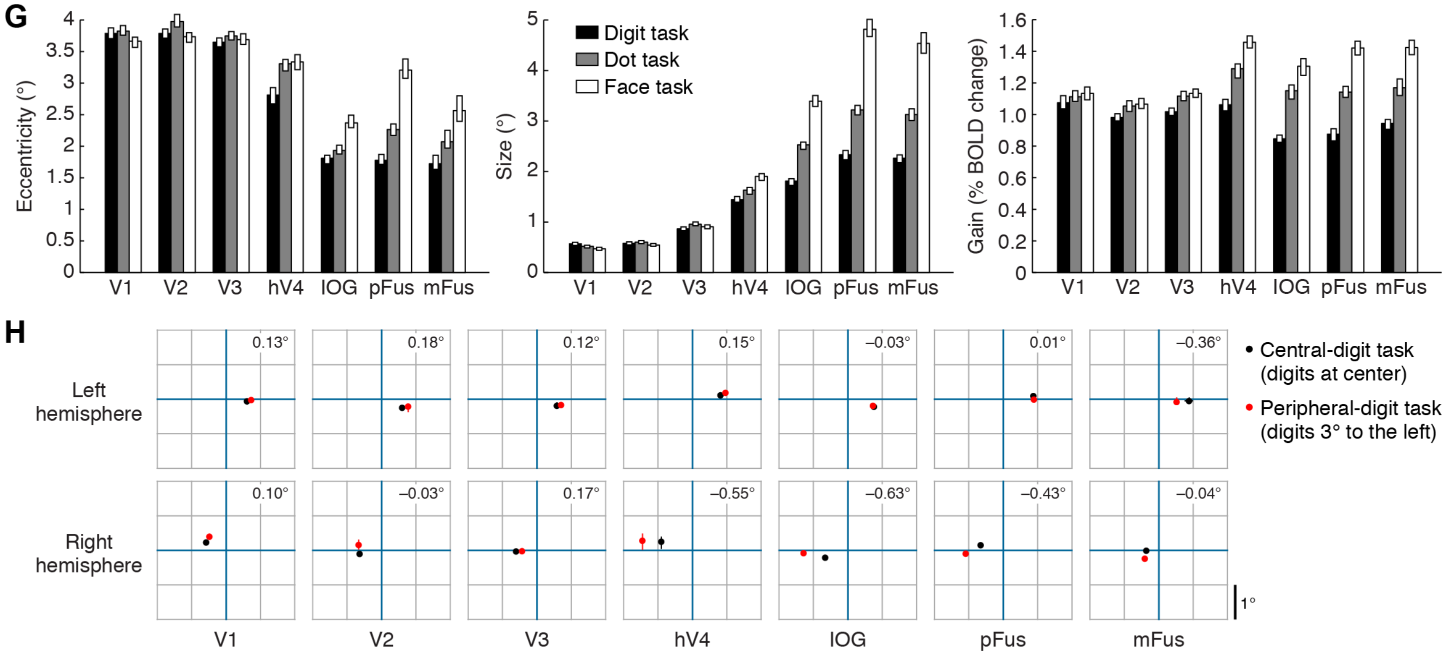


Figure S3, Related to Figure 3. Additional details on the task experiment. (A) Schematic of task experiment. The design is similar to that of the pRF-estimation experiment (Figure 1A), except that faces are presented on a coarser grid and at a single size. In different runs, the subject performs one of three tasks while maintaining central fixation: press a button when the fixation digit repeats (digit task), when a red dot appears on top of the face (dot task), or when face identity repeats (face task). (B) Example voxel from task experiment (left IOG, subject 3). A separate pRF is fit to responses measured under each task. Data (black bars; median across trials ± 68% CI) and pRF estimates (green lines) are shown for each task. Compared to the pRF obtained under the digit task (left), the pRFs obtained under the dot task (middle) and face task (right) exhibit increased eccentricity, size, and gain. (C) pRF angle and eccentricity are stable across tasks in V1–V3. Each dot represents a single voxel. Dots lie near the line of unity in V1–V3, indicating that the task does not substantially change angle and eccentricity tuning in these regions. (D) Eye tracking results (subject 2). Each image is a 2D histogram of eye positions measured during the presentation of faces at a single location (the complete 5 x 5 grid of locations is shown for each task). Horizontal and vertical gray lines indicate 1° increments. Eye positions were summarized by fitting a 2D Gaussian probability distribution, and a contour that contains 95% of the fitted distribution is indicated by a blue line. Eye positions remain consistently near the center of the display. (E) Eye tracking results (subject 3). (F) Summary of eye tracking results. For each face location, we compute the area of the 95%-contour that summarizes the range of eye positions observed (see panels D–E). We then divide face locations into three eccentricity bins (Low: 0–1.6°, Med: 1.6–3.2°, High: 3.2–4.8°) and plot the mean and standard deviation of the contour areas observed in each bin. (Due to calibration error, the scale of the results for Subject 1 may be inaccurate.) Although contour area varies significantly across tasks (p < 0.0001, one-way ANOVA in each subject), the size of the variation is quite small. To better understand the size of the variation, we compute the average contour area across face positions (separately for each task) and use a circle to approximate the contour shape. This analysis reveals that for subject 2, eye positions remain within a circle of radius 0.64° during the digit task and this radius increases by 0.03° during the dot task and by 0.11° during the face task. For subject 3, eye positions remain within a circle of radius 0.57° during the digit task and this radius increases by 0.02° during the dot task and by 0.06° during the face task. The increase in scatter of eye positions during the dot and face tasks is quite small (~0.1° or less) and cannot account for the sizes of the attentional effects that we observe (in face-selective regions, the median increase in pRF eccentricity from the digit task to the face task is ~0.6° and the median increase in pRF size from the digit task to the face task is ~1.5°; see Figure 3A). (G) Results of interleaved-task experiment (same format as Figure 3A). In this experiment, the same three tasks (digit, dot, face) were randomly interleaved within each run. We observe the same task-related changes in pRF properties as in the original task experiment (compare to Figure 3A). Thus, pRF changes cannot be explained by differences in subject arousal across runs. (H) Results of peripheral-digits experiment. In this experiment, pRFs were measured while subjects performed the digit task at fixation (central-digit task) or 3° left of fixation (peripheral-digit task). To examine whether this attentional manipulation causes pRFs to be shifted, we selected pRFs positioned near the center-of-gaze under the central-digit task (eccentricity less than 2°), and then computed the centroid of the pRF centers separately for each task (median x- and y-coordinate across voxels ± 68% CI; some error bars are not visible due to their small size). In right-hemisphere hV4, IOG, and pFus, the centroid computed for the peripheral-digit task is shifted leftward relative to the centroid computed for the central-digit task, indicating that pRFs in these regions shift towards the locus of attention (size of change in x-coordinate indicated at top-right of each plot). The absence of a leftward shift in right-hemisphere mFus suggests that attentional effects in mFus may specifically require attention to face features (which is present during the face task but not during the peripheral-digit task). Overall, the leftward shift of pRF centers is consistent with the influence of a top-down attentional field in ventral temporal cortex.

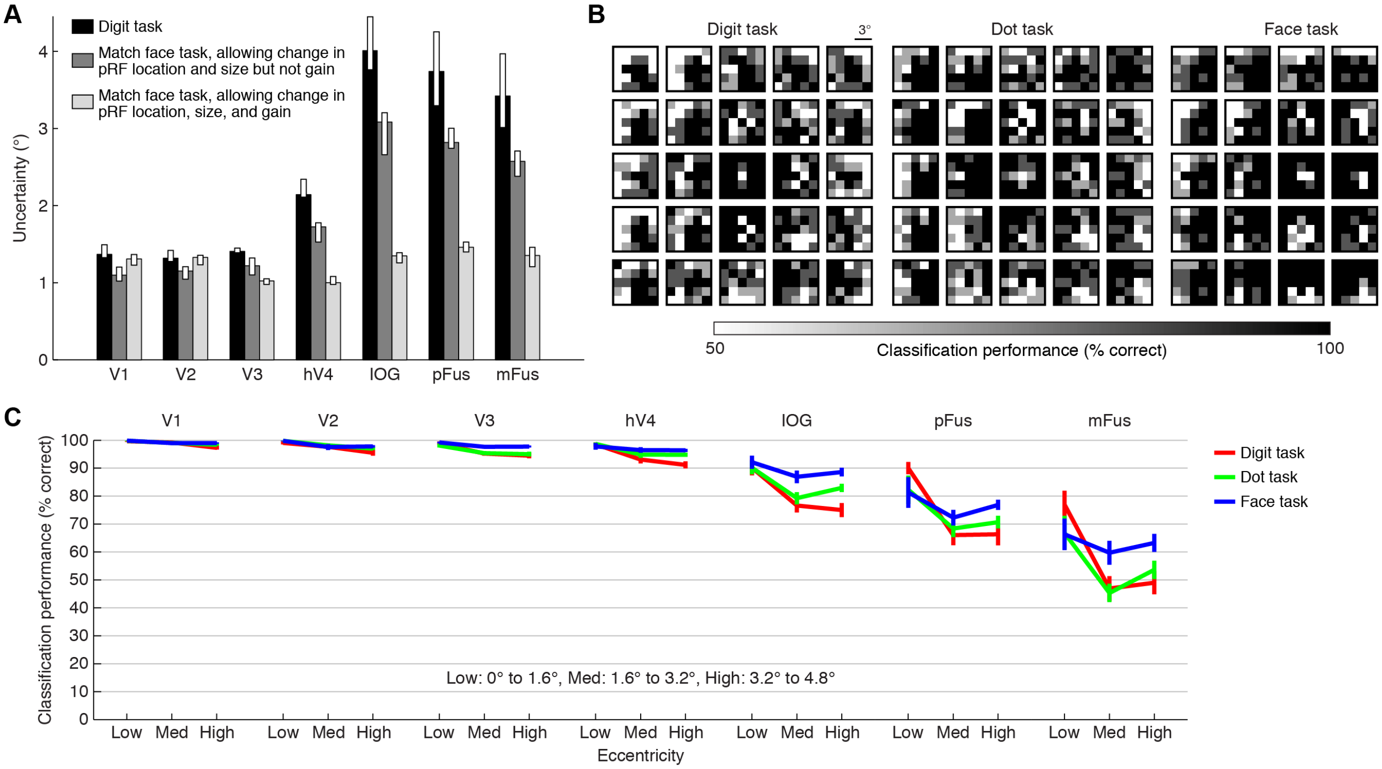


Figure S4, Related to Figure 4. Effect of pRF parameters on decoding and results of direct-decoding analysis. (A) Changes in pRF location and size reduce spatial uncertainty in distributed responses. We compared the uncertainty of pRFs at 5° eccentricity under the digit task (black) and face task (light gray) to the uncertainty of pRFs with same location and size measured in the face task, but with the gain measured in the digit task (dark gray). We reasoned that if the reduction in uncertainty under the face task is due solely to increased pRF gain, then uncertainty under the location-and-size-matched condition should be similar to uncertainty under the digit task. The results show that there are reductions in uncertainty when location and size (but not gain) are matched to the face task. This indicates that reductions in uncertainty under the face task are due not only to changes in pRF gain, but also to changes in pRF location and size. (B) Visualization of direct decoding results for an example region, IOG. Each image is a 5 x 5 map of classification performance, where the shading of each pixel indicates how well a given face position can be discriminated from the reference face position (which is indicated by the relative position of the image within the overall grid). (C) Summary of direct decoding results. For each 5 x 5 map, we compute the average classification performance excluding the face position corresponding to the reference. Then, we bin the results according to the eccentricity of the reference face position. The mean of each bin (± SEM) is plotted. In face-selective regions, performance substantially improves under the face task compared to the digit task for peripheral reference face positions, with the dot task yielding intermediate performance.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Subjects

Three experienced fMRI subjects (the authors; two males, one female) participated in this study. All subjects had normal or corrected-to-normal visual acuity. Informed written consent was obtained from all subjects, and the experimental protocol was approved by the Stanford University Institutional Review Board. Each subject participated in 8–9 scanning sessions: 3–4 scanning sessions to develop and test the model (*pRF-estimation experiment*), 3 scanning sessions to examine the effects of attention on pRF properties (*task experiment*, *interleaved-task experiment*, *peripheral-digits experiment*), and 2 scanning sessions in which retinotopic mapping [S1, S2] and functional localizers [S3, S4] were conducted in order to define regions of interest.

Visual stimuli

Stimuli were presented using a Samsung SyncMaster 305T LCD monitor positioned at the head of the scanner bed. Subjects viewed the monitor via a mirror mounted on the RF coil. The monitor operated at a resolution of 1280 x 800 at 60 Hz. Stimuli subtended 12.5–12.6° of visual angle (viewing distance 182–184 cm). A MacBook Pro computer controlled stimulus presentation using code based on Psychophysics Toolbox [S5, S6]. Behavioral responses were recorded using a button box.

Experimental design

pRF-estimation experiment

The goal of this experiment was to develop a population receptive field (pRF) model [S7, S8] that describes VTC responses. We therefore designed a novel protocol optimized for estimating pRFs in high-level visual areas. In this protocol, subjects maintained central fixation while faces and phase-scrambled faces were presented at various positions and sizes. To construct the face stimuli, we photographed individuals in a controlled laboratory environment and generated a stimulus set of 95 individuals x 7 viewpoints (0°, ±15°, ±30°, ±45°) = 665 faces. The faces were converted to grayscale, cropped with a circular mask, rescaled to three different sizes (1.6°, 3.2°, 4.7° diameter), and placed on a gray background (Figure 1A). The positions of the faces were varied according to a 7 x 7 grid with 1° spacing (with the central position coincident with the center-of-gaze). Phase-scrambled faces were constructed for the smallest face size (1.6° diameter). This was done by randomizing the phase spectrum of each face (excluding the DC component), cropping the resulting images with a circular mask to match the spatial extent of the original faces, and then reducing image contrast by 30% to avoid clipping of luminance values. The pRF-estimation experiment included a total of 7 positions x 7 positions x 4 stimulus types (small faces, medium faces, large faces, phase-scrambled small faces) = 196 conditions.

Stimuli were presented in 4-s trials, one condition per trial. In a trial, 7 images lasting 0.5 s each were sequentially presented and were followed by a 0.5-s delay before the next trial. The images presented in a trial were drawn from the same position and stimulus type but were randomized with respect to identity and viewpoint in order to reduce adaptation effects [S9, S10] and increase response magnitudes. A white semi-transparent (20% opacity) fixation grid was superimposed on top of the stimuli and was present throughout the duration of the experiment [S11].

The 196 conditions were randomly split into 4 groups containing 49 stimuli each. Stimulus types were equally distributed across groups (e.g. each group contained 12–13 small faces, 12–13 medium faces, etc.). In each run, conditions from one of the groups were presented once and in random order. To establish the baseline signal level, each run also included null trials in which no stimuli were presented. Four null trials were inserted at the beginning and end of each run, and ten null trials were randomly intermixed with the 49 stimulus trials under the constraint that null trials could not occur first nor last and null trials could not occur back-to-back. Each run lasted ~4.5 min. A total of 4 groups x 3 repetitions = 12 runs were collected in a scanning session. Thus, each of the 196 conditions was presented 3 times over the course of the session.

A single task was used throughout the pRF-estimation experiment: the subject was instructed to fixate the center of the grid and to press a button whenever a red dot (0.3° x 0.3°, 25% opacity) appeared on top of the stimulus. The red dot was always positioned at the center of the stimulus. The dot appeared with probability 0.5 on each trial, lasted 0.5 s, and coincided with the presentation of one of the 7 images from that trial.

To increase the signal-to-noise ratio (SNR) in the pRF-estimation experiment, each subject participated in 3–4 scanning sessions of the experiment. The stimuli and their temporal ordering were matched across scanning sessions. Data from different sessions were pre-processed (see *Data pre-processing*) and then averaged together before subsequent analysis.

Task experiment

The goal of this experiment was to characterize the effects of attention on pRF properties. The task experiment was similar to the pRF-estimation experiment, except that the fixation grid was removed, the number of stimuli was reduced, and responses were measured for three different tasks (digit task, dot task, face task). To reduce the number of stimuli, only medium-sized faces were used (3.2° diameter) and a coarser grid was used (5 x 5 grid, 1.5° spacing). The task experiment included a total of 5 positions x 5 positions x 3 tasks = 75 conditions (Figure S3A). In each trial, subjects viewed a sequence of 7 faces presented at a single position.

To support the tasks, additional characteristics were added to the stimulus design: (1) We placed a red dot on top of one of the faces in a trial at a probability of 0.5 per trial (same as in the pRF-estimation experiment). (2) For each 4-s trial, we generated a random sequence of 7 faces satisfying the constraint that both identity and viewpoint change from one face to the next. Then, with probability 0.5, we manipulated this sequence to contain a repetition of face identity. This was done by randomly selecting one of the faces (excluding the first) and forcing that face to repeat the identity of the previous face (thus, the repeated face shows a different viewpoint of the same individual). (3) We placed a stream of small digits (0.3° x 0.3°) at the center-of-gaze. The identity of the digit (0–9) changed every 0.5 s: each digit was presented for 0.25 s and was followed by a delay of 0.25 s. To minimize visual adaptation, the digit color alternated between black and white on successive presentations. Digit repetitions occurred with a probability of 0.052, with a maximum of two successive identical digits allowed (this matches the overall frequency of digit repetitions to the overall frequency of dot occurrences and the overall frequency of face-identity repetitions).

At the beginning of each run, the subject was instructed to perform one of three tasks while viewing the stimuli. In the *digit task*, the subject was instructed to fixate the central digit and to press a button whenever the same digit repeated. In the *dot task*, the subject was instructed to fixate the central digit and to press a button whenever the red dot appeared. In the *face task*, the subject was instructed to fixate the central digit and to press a button whenever the same face identity was repeated within a trial.

In each run, each of the 25 stimuli was presented twice and in random order. Null trials were included just as in the pRF-estimation experiment. Each run lasted ~4.5 min. The first run involved the dot task, the second run the digit task, the third run the face task, and this cycled until a total of 3 tasks x 4 repetitions = 12 runs were collected. Thus, each of the 75 conditions was presented 8 times over the course of the session. To ensure that the only difference across tasks is the subject's attentional state, the exact same physical stimulus (including faces, dots, digits, and their temporal ordering) was used for each task. This was accomplished by generating four distinct stimulus sequences and repeating them over the course of the session (i.e., the stimuli are given by ABCD ABCD ABCD, where each letter corresponds to a distinct stimulus sequence, while the tasks are given by DGF DGF DGF DGF, where D, G, and F indicate the dot, digit, and face tasks, respectively). To verify accurate fixation, eye tracking was performed using a scanner-compatible SR Research EyeLink 1000 eye tracker.

Interleaved-task experiment

To control for potential differences in subject arousal across tasks, we conducted an experiment in which tasks were randomly interleaved on a trial-by-trial basis within each run. This experiment was identical to the original task experiment except for the following: (1) A central red letter (0.3° x 0.3°) presented at the beginning of each trial served as a cue for which task to perform ('G': digit, 'D': dot, 'F': face). (2) Trials lasted 6 s and were structured as follows: 0.5-s cue presentation, 0.5-s delay, 1.0-s of digits, 3.5-s of faces and digits, and 0.5-s delay. (3) The 25 face locations were randomly split into 2 groups (13 and 12 locations each). In each run, face locations from one of the groups were presented three times, once for each task. Each run also included three cue-only trials for each task as well as three null trials at the beginning and end of the run and three null trials intermixed during the run. During cue-only trials, the cue and digits were presented but faces were omitted; during the null trials, just the digits were presented (no cue, no faces). The total number of trials was 3 null + ((13 (or 12) conditions + 3 cue-only) x 3 tasks + 3 null) + 3 null = 57 (or 54) trials, lasting 5.7 (or 5.4) min. (4) Digit repetitions occurred with probability 0.5 on each cued trial. (5) A total of 2 groups x 6 repetitions = 12 runs were collected.

Peripheral-digits experiment

This experiment was similar to the task experiment, except that the fixation grid was present and responses were measured for two tasks. The *central-digit task* was identical to the original digit task: the subject performed a 1-back task on centrally positioned digits of size 0.3° x 0.3°. In the *peripheral-digit* task, the subject also performed a 1-back task on digits, but the digits were positioned 3.0° left of center and were enlarged to 0.75° x 0.75° to compensate for lower reading acuity in the periphery. The two tasks were performed in alternating runs until 10 runs were collected. The same physical stimulus (up to the location and size of the digits) was used for the two tasks.

MRI data acquisition

Functional MRI data were collected at the Stanford Center for Cognitive and Neurobiological Imaging using a 3T GE Signa MR750 scanner and a Nova 16-channel visual RF coil. In each scanning session, 26 oblique slices covering occipitotemporal cortex were defined: slice thickness 2 mm, slice gap 0 mm, field-of-view 160 mm x 160 mm, phase-encode direction anterior-posterior. A T2*-weighted, single-shot, gradient-echo EPI pulse sequence was used: matrix size 80 x 80, TR 2006.553 ms, TE 33 ms, flip angle 77°, nominal spatial resolution 2 mm x 2 mm x 2 mm. The TR was matched to the refresh rate of the display such that there are exactly 2 TRs for each 4-s trial (i.e. 240 refreshes). Measurements of the B_0 magnetic field were performed for post-hoc correction of EPI spatial distortion.

Data analysis

Region-of-interest (ROI) definition

All ROIs were defined in individual subjects based on each subject's native anatomical space and without spatial smoothing as in our prior studies [S3, S12]. Visual field maps were defined using retinotopic mapping scans. Subjects participated in 4–8 runs in which they viewed sweeps of bar apertures that contained a flickering checkerboard pattern while fixating on a central point [S4, S7]. These data were used to define V1, V2, V3, and hV4 [S13]. Face-selective regions were defined using functional-localizer scans. Subjects participated in 2 runs of a block-design experiment in which images of faces, limbs, flowers, houses, cars, guitars, and scrambled objects were presented [S3]. Using these data, three face-selective regions were defined based on their significantly higher responses to faces compared to other stimuli ($t > 3$, voxel level, uncorrected) and their anatomical location and topological relationship to retinotopic areas and other high-level visual regions [S3, S12, S14]. The defined regions are the inferior occipital gyrus, IOG-faces/OFA (abbreviated IOG); posterior fusiform gyrus, pFus-faces/FFA-1 (abbreviated pFus); and middle fusiform gyrus, mFus-faces/FFA-2 (abbreviated mFus). Whether these face-selective regions in humans are retinotopically organized (as has been suggested for macaques [S15]) is an important question and is outside the scope of the present study.

Data pre-processing

The fMRI data were pre-processed by dropping the first five volumes of each run (to allow magnetization to reach steady-state) and performing slice time correction, spatial undistortion, and motion correction (both within and across scanning sessions). The combined effects of distortion and motion were corrected using a single cubic interpolation of the slice-time corrected volumes (for further details of pre-processing, see [S16]). No spatial smoothing was performed. Data were aligned to each subject's native anatomical volume in order to identify voxels in each ROI.

GLM analysis

The pre-processed fMRI data were analyzed using GLMdenoise [S17] (MATLAB implementation available at http://kendrickkay.net/GLMdenoise/), a data-driven denoising method that derives estimates of correlated noise from the data and incorporates these estimates as nuisance regressors in a general linear model (GLM) analysis of the data. GLMdenoise uses polynomials to model the baseline signal level in each run and provides an estimate of the BOLD response amplitude of each voxel to each condition. Response amplitudes are converted to units of percent BOLD signal change by dividing by the mean signal intensity in each voxel, and error bars on response amplitudes are obtained by bootstrapping (resampling runs with replacement). Subsequent analyses involved analyzing these response amplitudes. For the interleaved-task experiment, the GLM included a set of finite impulse response regressors for each cue type (extending 0–20 s after cue onset) in order to account for any non-specific effects related to the cue or task preparation.

Population receptive field (pRF) analysis

pRF analysis was performed independently for each voxel. We modeled the response amplitudes from each voxel using the compressive spatial summation (CSS) model [S8] (MATLAB implementation available at http://kendrickkay.net/socmodel/), which is an extension of standard pRF analyses [S7]. The CSS model predicts voxel responses to stimuli presented at arbitrary locations in the visual field. Each voxel's pRF is modeled using an isotropic 2D Gaussian combined with a static power-law nonlinearity. By fitting the model, we determine where and how large the pRF must be in order to generate the response amplitudes observed for a given voxel. The model includes two parameters (x, y) for the position of the Gaussian, a parameter (σ) for the size of the Gaussian, a parameter (n) for the exponent of the power-law nonlinearity, and a parameter (g) for the overall gain of the predicted responses. In the CSS model, stimuli are represented as contrast images (see Figure 1B) and the predicted response is obtained by computing a weighted sum of the stimulus with the Gaussian and then applying the power-law nonlinearity.

Intuitively, the CSS model predicts the response to a stimulus based solely on the spatial extent of the stimulus and its spatial overlap with the pRF (if the stimulus completely covers the pRF, it will produce the largest response; if it partially overlaps the pRF, it will produce an intermediate response; and if it does not overlap the pRF, it will produce no response). The power-law nonlinearity, which is typically compressive ($n < 1$), allows the model to exhibit tolerance to a range of changes in the position and size of the stimulus [S8]. In our data, the exponent of the power-law nonlinearity (n) is 0.53, 0.32, 0.22, 0.19, 0.20, 0.16, and 0.23 in V1, V2, V3, hV4, IOG, pFus, and mFus, respectively (median across voxels). Note that the CSS model predicts larger responses for larger faces, since larger faces cover more of the pRF. Also, note that the CSS model does not account for response variations driven by stimulus type—for example, stronger responses to faces compared to phase-scrambled faces in face-selective regions (Figure S1B)—but extensions of the CSS model (as in [S16]) may be able to do so. In general, expanding the range of stimuli for which the model makes accurate predictions—such as a model that is fully computable for arbitrary images [S18]—is an important direction for future research.

We define pRF size as $\sigma/\sqrt{n}$, which is the standard deviation of a 2D Gaussian that characterizes the response of the model to point stimuli [S8]. We visualize the location of a pRF by plotting a contour at ±2 pRF sizes away from the pRF center (Figure 1B).

Finally, we quantify pRF gain by computing the maximum predicted response for the stimuli considered in the model fitting (this empirically observed gain is more robust than the raw gain parameter).

Model fitting was performed using nonlinear optimization (MATLAB Optimization Toolbox). For the purposes of modeling, we rectified the response amplitudes observed at each voxel (negative response amplitudes were set to zero). In our experiments, negative BOLD responses can be found in V1–V3, indicating that the presentation of a face can cause the BOLD signal in a voxel to drop below baseline. Such suppression may reflect early attentional filtering and may have a distinct physiological source [S19] compared to positive BOLD responses which are the focus of the present study.

For the pRF-estimation experiment, several versions of the CSS model were fit. The *Full* model involved fitting the model to all three face sizes simultaneously; the *Individual* model involved fitting the model to each face size separately as well as the phase-scrambled faces separately; and the *Generalize* model involved fitting the model to two face sizes and predicting responses to the third face size (details below). The final parameter estimates for each voxel were those obtained from the *Full* model; the purpose of the other versions of the model was to assess the cross-validation accuracy of the model (Figure S1A) and to examine whether pRF estimates depend on the type of stimulus used (Figures S1B, S1C; comparison of pRFs obtained using small faces and pRFs obtained using phase-scrambled small faces).

For the three experiments in which task was manipulated, the CSS model was fit separately to the response amplitudes measured under each task. Since the reduced number of stimuli in these experiments provide insufficient data to reliably estimate the exponent parameter (n), the exponent parameter was set to a fixed value (0.2). Note that any task-induced changes in response amplitudes reflect changes in evoked activity and not changes in baseline activity. This is because in the task experiment and peripheral-digits experiment, any shifts in the baseline signal level due to the task are explicitly modeled in the GLM (using a separate set of polynomial regressors for each run), and in the interleaved-task experiment, any non-specific effects related to task preparation are explicitly modeled in the GLM (using finite impulse response regressors).

Cross-validation was used to estimate the accuracy of the CSS model. For the *Full* and *Individual* models, the response amplitudes for each face size were randomly split into ten groups, and each group was systematically left out and used as the testing set (thus, a total of 30 cross-validation iterations were performed). A different cross-validation scheme was used for the *Generalize* model: this model involved three cross-validation iterations in which each face size was left out and used as the testing set. In all cases, accuracy was quantified as the percentage of variance explained (R^2) between the cross-validated predictions of the response amplitudes and the measured response amplitudes (variance was defined with respect to 0% BOLD signal change; see [S8, S16]). For comparison, we computed the noise ceiling for model predictions, i.e. the maximum possible performance given the level of noise in the data. This was accomplished through Monte Carlo simulations (see [S8] for details). We also computed, for comparison purposes, the accuracy of a flat-response (invariant or fully tolerant) model that simply predicts the same response regardless of the position and size of the face. Results of the cross-validation analysis are shown in Figure S1A.

Model-based decoding analysis

The purpose of the model-based decoding analysis is to assess how well a collection of pRFs represent spatial information. The decoding analysis is applied separately to pRFs observed under different tasks; this helps us understand why task-induced changes in pRF properties might be useful. In the decoding analysis, we quantify how well a simple linear classifier can discriminate faces at two different positions based on the distributed pattern of responses across the collection of pRFs in a region of interest.

First, we used the pRFs to predict the response patterns that would be produced by a small-sized face (1.6° diameter) at different visual field positions. Then, for each position and a given reference position, we calculated how well an optimal linear decoder can discriminate the two face positions based on the response patterns. We assumed that response patterns are affected by additive, independent Gaussian noise with standard deviation $0.05*\sqrt{n}$ where n is the number of pRFs (scaling by the square root of the number of pRFs compensates for differences in the number of pRFs in different regions). Example maps of discrimination performance are shown in Figure 4A. Finally, to summarize discrimination performance, we calculated the square root of the area of the 75%-correct contour of each map. We refer to the result as the *spatial uncertainty*, as it indicates the distance below which shifts in face position cannot be well discriminated. Intuitively, the larger the contour, the more uncertain the position of the face.

To disregard differences in gain across voxels, we fixed the gain of all pRFs observed under the digit task to 1. We then set the gains of pRFs observed under the dot and face tasks according to the scaling that was observed (e.g., if the gain under the digit task for a given voxel was 0.8% BOLD change and the gain under the face task was 1.2% BOLD change, then the gain of the pRF under the face task would be set to 1.5). To avoid the influence of outliers, gain scalings were subjected to 50% Winsorization (gain scalings below the 25th percentile were set to the 25th percentile and gain scalings above the 75th percentile were set to the 75th percentile).

In Figures 4B, 4C, and S4A, we show results for reference positions located at 16 equally spaced angles at two different eccentricities, 1° and 5°. The median and error across angular positions is plotted.

Direct decoding analysis

The model-based decoding analysis described above provides the ability to control the amount and type of noise present, to explore stimulus configurations beyond the particular ones used in the experiment, and to identify the specific characteristics of representation that influence decoding performance. However, the analysis relies on the validity of the models that are fit to the data. We performed an alternative decoding analysis in which decoding is performed directly from the data, without an intervening modeling step. First, we analyzed the time-series data from the task experiment using a GLM in which a separate set of BOLD response amplitudes is estimated for each run. This produced four independent sets of response amplitudes for each face position and task. Next, response amplitudes were converted into *t* units. Finally, for each region and task, we quantified how well response patterns discriminate each face position from all other face positions. This was done by performing nearest-centroid classification using a split-half cross-validation scheme. For example, response patterns from runs 1 and 2 are averaged together, producing one centroid for the reference face position and one centroid for the target face position. Then, response patterns from runs 3 and 4 for the target face position are averaged together and classified by determining the nearest centroid. All possible training/testing splits of the runs were performed (a total of six), and classification performance was aggregated across splits.

Voxel selection

Voxels in each region of interest (ROI) were pooled across subjects. Unless otherwise indicated, all figure plots use median as a measure of central tendency and error bars (68% confidence intervals) were obtained using bootstrapping. Error bars reflect variability across voxels in each ROI (Figures 2A, 3, S1, S2), across trials (Figures 1C, S3B), or across angular position (Figures 4B, 4C, S4A). For quantification of pRF properties, to avoid noisy unreliable voxels, we selected voxels with a goodness-of-fit R^2 of at least 50%. For quantification of task effects, we selected voxels with a goodness-of-fit R^2 of at least 50% for at least one of the tasks (with the exception of Figures 3B–3C where a cutoff of 80% was used in order to improve visibility).

Behavioral analysis

Button responses within two seconds of each behavioral event of interest were analyzed. Behavioral performance was quantified using the *d′* sensitivity index [S20]. In some cases, no misses (all hits) or no false alarms (all correct rejections) were observed. To ensure finite *d′* values for these cases, the minimum number of misses or false alarms was set to 1. In the pRF-estimation experiment, *d′* was 2.9, 3.1, and 5.1 for Subjects 1, 2, and 3, respectively. In the task experiment, *d′* for the digit, dot, and face tasks was [2.3, 2.6, 1.3], [2.0, 3.8, 1.2], and [3.1, 4.7, 2.5] for Subjects 1, 2, and 3, respectively (mean across subjects: [2.5, 3.7, 1.7]).

Eye-tracking analysis

Eye tracking was performed during the task experiment. For each run, time-series of horizontal and vertical eye positions were obtained. Blinks were detected based on outlier values, and data within ±0.1 seconds of each blink were excised. A line was fit and removed from each time-series to compensate for instrumental drift. Eye-tracking results show that all subjects were able to maintain central fixation in each of the three tasks (Figures S3D–S3F). Due to calibration error, the absolute units for horizontal and vertical eye positions may be inaccurate for Subject 1.

SUPPLEMENTAL REFERENCES

S1. Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., and Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

S2. Engel, S. A., Rumelhart, D. E., Wandell, B., Lee, A. T., Glover, G. H., Chichilnisky, E. J., and Shadlen, M. N. (1994). fMRI of human visual cortex. Nature *369*, 525.

S3. Weiner, K. S., and Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. NeuroImage *52*, 1559–1573.

S4. Weiner, K. S., and Grill-Spector, K. (2011). Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. NeuroImage *56*, 2183–2199.

S5. Brainard, D. H. (1997). The Psychophysics Toolbox. Spat Vis *10*, 433–436.

S6. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis *10*, 437–442.

S7. Dumoulin, S. O., and Wandell, B. (2008). Population receptive field estimates in human visual cortex. NeuroImage *39*, 647–660.

S8. Kay, K. N., Winawer, J., Mezer, A., and Wandell, B. (2013). Compressive spatial summation in human visual cortex. Journal of neurophysiology *110*, 481–494.

S9. Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

S10. Weiner, K. S., Sayres, R., Vinberg, J., and Grill-Spector, K. (2010). fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. Journal of neurophysiology *103*, 3349–3365.

S11. Schira, M. M., Tyler, C. W., Breakspear, M., and Spehar, B. (2009). The foveal confluence in human visual cortex. J. Neurosci. *29*, 9050–9058.

S12. Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., Amunts, K., and Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. NeuroImage *84*, 453–465.

S13. Brewer, A. A., Liu, J., Wade, A. R., and Wandell, B. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. Nature neuroscience *8*, 1102–1109.

S14. Weiner, K. S., and Grill-Spector, K. (2012). The improbable simplicity of the fusiform face area. Trends in cognitive sciences *16*, 251–254.

S15. Janssens, T., Zhu, Q., Popivanov, I. D., and Vanduffel, W. (2014). Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. J. Neurosci. *34*, 10156–10167.

S16. Kay, K. N., Winawer, J., Rokem, A., Mezer, A., and Wandell, B. (2013). A two-stage cascade model of BOLD responses in human visual cortex. PLoS computational biology *9*, e1003079.

S17. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F., and Wandell, B. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci *7*, 247.

S18. Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences of the United States of America *111*, 8619–8624.

S19. Gouws, A. D., Alvarez, I., Watson, D. M., Uesaki, M., Rogers, J., and Morland, A. B. (2014). On the role of suppression in spatial attention: evidence from negative BOLD in human subcortical and cortical structures. J. Neurosci. *34*, 10347–10360.

S20. Macmillan, N. A., and Creelman, C. D. (2004). Detection Theory (Psychology Press).