frontiers in
COMPUTATIONAL NEUROSCIENCE

PERSPECTIVE ARTICLE
published: 09 March 2012
doi: 10.3389/fncom.2012.00012

Modeling invariant object processing based on tight integration of simulated and empirical data in a Common Brain Space

***Judith C. Peters*[1,2,3]*, *Joel Reithler*[1,2,3] *and Rainer Goebel*[1,2,3]**

[1] *Netherlands Institute for Neuroscience, Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, Netherlands*
[2] *Faculty of Psychology and Neuroscience, Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands*
[3] *Maastricht Brain Imaging Center (M-BIC), Maastricht University, Maastricht, Netherlands*

Edited by:
Evgeniy Bart, Palo Alto Research Center, USA

Reviewed by:
Peter Konig, University of Osnabrück, Germany
John Lisman, Brandeis University, USA

****Correspondence:***
Judith C. Peters, Department Neuroimaging and Neuromodeling, Netherlands Institute for Neuroscience, Meibergdreef 47, 1105 BA, Amsterdam, Netherlands.
e-mail: j.peters@nin.knaw.nl

Recent advances in Computer Vision and Experimental Neuroscience provided insights into mechanisms underlying invariant object recognition. However, due to the different research aims in both fields models tended to evolve independently. A tighter integration between computational and empirical work may contribute to cross-fertilized development of (neurobiologically plausible) computational models and computationally defined empirical theories, which can be incrementally merged into a comprehensive brain model. After reviewing theoretical and empirical work on invariant object perception, this article proposes a novel framework in which neural network activity and measured neuroimaging data are interfaced in a common representational space. This enables direct quantitative comparisons between predicted and observed activity patterns within and across multiple stages of object processing, which may help to clarify how high-order invariant representations are created from low-level features. Given the advent of columnar-level imaging with high-resolution fMRI, it is time to capitalize on this new window into the brain and test which predictions of the various object recognition models are supported by this novel empirical evidence.

Keywords: object perception, view-invariant object recognition, neuroimaging, large-scale neuromodeling, (high-field) fMRI, multimodal data integration

INTRODUCTION

One of the most complex problems the visual system has to solve is recognizing objects across a wide range of encountered variations. Retinal information about one and the same object can dramatically vary when position, viewpoint, lighting, or distance change, or when the object is partly occluded by other objects. In Computer Vision, there are a variety of models using alignment, invariant properties, or part-decomposition methods (Roberts, 1965; Fukushima, 1982; Marr, 1982; Ullman et al., 2001; Viola and Jones, 2001; Lowe, 2004; Torralba et al., 2008), which are able to identify objects across a range of viewing conditions.

Some computational models are clearly biologically inspired and take for example the architecture of the visual system into account (e.g., Wersing and Körner, 2003), or cleverly adapt the concept of a powerful Computer Vision algorithm (e.g., the Fourier-Mellin transform) to a neurobiologically plausible alternative (Sountsov et al., 2011). Such models can successfully detect objects in sets of widely varying natural images (Torralba et al., 2008) and achieve impressive invariance (Sountsov et al., 2011). In general however, computer vision models are developed for practical image analysis applications (handwriting recognition, face detection, etc.) for which fast and accurate object recognition and not neurobiological validity is pivotal. Therefore, these models are generally less powerful in explaining how object constancy arises in the human brain. Indeed, "Models are common; good theories are scarce" as suggested by Stevens (2000, p. 1177).

Humans are highly skilled in object recognition, and they outperform machines in object recognition tasks with great ease (Fleuret et al., 2011). This is partly because they are able to strategically use semantics and information from context or memory. In addition, they can direct attention to informative features in the image, while ignoring distracting information. Such higher cognitive processes are difficult to implement, but improve object recognition performance when taken into account (Lowe, 2000). Computer vision models might become more accurate in recognizing objects across a wide range of variations in image input, when implementing algorithms derived from neurobiological observations.

Reciprocally, our interpretation of such neurobiological findings might be greatly improved by insights in the underlying computational mechanisms. Humans can identify objects with great speed and accuracy, even when the object percept is degraded, occluded or presented in a highly cluttered visual scene (e.g., Thorpe et al., 1996). However, which computational mechanisms enable such remarkable performance is not yet fully understood. To create a comprehensive theory of human object recognition and how it achieves invariant object recognition, computational mechanisms derived from modeling efforts should be incorporated in neuroscientific theories based on experimental findings.

In the current paper, we highlight recent developments in object recognition research and put forward a "Common Brain Space" framework (CBS; Goebel and De Weerd, 2009; Peters et al.,

2010) in which empirical data and computational results can be directly integrated and quantitatively compared.

EXPLORING INVARIANT OBJECT RECOGNITION IN THE HUMAN VISUAL SYSTEM

Object recognition, discrimination, and identification are complex tasks. Different encounters with an object are unlikely to take place under identical viewing conditions, requiring the visual system to generalize across changes. Information that is important to retrieve object identity should be effectively processed, while unimportant view-point variations should be ignored. That is, the recognition system should be stable yet sensitive (Marr and Nishihara, 1978), leading to inherent tradeoffs. How the visual system is able to accomplish this task with such apparent ease is not yet understood. There are two classes of theories on object recognition. The first suggests that objects can be recognized by cardinal ("non-accidental") properties that are relatively invariant to the objects' appearance (Marr, 1982; Biederman, 1987). Thus, these invariant properties and their spatial relations should provide sufficient information to recognize objects regardless of their viewpoint. However, how such cardinal properties are defined and recognized in an invariant manner is a complex issue (Tarr and Bülthoff, 1995). The second type of theory suggests that there are no such invariants but that objects are stored in the view as originally encountered (which, in natural settings encompasses multiple views being sampled in a short time interval), thereby maintaining view-dependent shape and surface information (Edelman and Bülthoff, 1992). Recognition of an object under different viewing conditions is achieved by either computing quality matches between the input and stored presentations (Perrett et al., 1998; Riesenhuber and Poggio, 1999) or by transforming input to match the view specifications of the stored representation (Bülthoff and Edelman, 1992). The latter normalization can be accomplished by interpolation (Poggio and Edelman, 1990), mental transformation (Tarr and Pinker, 1989), or alignment (Ullman, 1989).

These theories make very different neural predictions. View-invariant theories suggest that the visual system recognizes objects using a limited library of non-accidental properties, and neural representations are invariant. Evidence for such invariant object representations have been found at final stages of the visual pathway (Quiroga et al., 2005; Freiwald and Tsao, 2010). In contrast, the second class of theories assumes that neural object representations are view-dependent, with neurons being sensitive to object transformations. Clearly, the early visual system is sensitive to object appearance: the same object can elicit completely different, non-overlapping neural activation patterns when presented at different locations in the visual field. So, object representations are input specific at initial stages of processing, whereas invariant representations emerge in final stages. However, how objects are represented by intermediate stages of this processing chain is not yet well understood. Likely, multiple different transforms are (perhaps in parallel) performed at theses stages. This creates multiple object representations, in line with the various types of information (such as position and orientation) that have to be preserved for interaction with objects. Moreover, position information aids invariant object learning (Einhäuser et al., 2005; Li and DiCarlo, 2008, 2010) and representations can reflect view-dependent and view-invariant information simultaneously (Franzius et al., 2011).

The following section reviews evidence from monkey neurophysiology and human neuroimaging on how object perception and recognition are implemented in the primate brain. As already alluded to above, the visual system is hierarchically organized in more than 25 areas (Felleman and Van Essen, 1991) with initial processing of low-level visual information by neurons in the thalamus, striate cortex (V1) and V2; and of more complex features in V3 and V4 (Carlson et al., 2011). Further processing of object information in the human ventral pathway (Ungerleider and Haxby, 1994), involves higher-order visual areas such as the lateral occipital cortex (LOC; Malach, 1995) and object selective areas for faces ("FFA"; Kanwisher et al., 1997), bodies ("EBA"; Downing et al., 2001), words ("VWFA"; McCandliss et al., 2003), and scenes ("PPA"; Epstein et al., 1999).

The first studies on the neural mechanisms of object recognition were neurophysiological recordings in monkeys. In macaque anterior inferotemporal (IT) cortex, most of the object-selective neurons are tuned to viewing-position (Logothetis et al., 1995; Booth and Rolls, 1998), in line with viewpoint-dependent theories. On the other hand, IT neurons also turned out to be more sensitive to changes in "non-accidental" than to equally large pixel-wise changes in other shape features ("metric properties"; Kayaert et al., 2003), providing support for structural description theories (Biederman, 1987). Taken together, these studies provide neural evidence for both theories (see also Rust and Dicarlo, 2010). However, to which degree object representations are stored in an invariant or view-dependent manner across visual areas, and how these representations arise and are matched to incoming information, remains elusive.

Also human neuroimaging studies have not provided conclusive evidence. In fMRI studies, the BOLD signal reflects neural activity at the population rather than single-cell level. The highest functional resolution provided by standard 3 Tesla MRI scanners is around $2 \times 2 \times 2\,\mathrm{mm}^3$, which is too coarse to zoom into the functional architecture within visual areas. However, more subtle information-patterns can be extracted using multi-voxel pattern analysis (MVPA; Haynes et al., 2007) or fMRI-adaptation (fMRI-A; Grill-Spector and Malach, 2001). MVPA can reveal subtle differences in distributed fMRI patterns across voxels resulting from small biases in the distributions of differentially tuned neurons that are sampled by each voxel. By using classification techniques developed in machine learning, distributed spatial patterns of different classes (e.g., different objects) can be successfully discriminated (see Fuentemilla et al., 2010 for a temporal pattern classification example with MEG). For example, changing the position of an object significantly changes patterns in LOC, even more than replacing an object (at the same position) by an object of a different category (Sayres and Grill-Spector, 2008). Rotating the object (up to 60°) did not change LOC responses however (Eger et al., 2008) suggesting that LOC representations might be view-dependent in only some aspects. fMRI-A exploits the fact that the neuronal (and the corresponding hemodynamic) response is weaker for repeated compared to novel stimuli (Miller and Desimone, 1994). Thus, areas are sensitive to view-dependent

changes when their BOLD response returns to its initial level for objects that are presented a second time, but now from a different view-point. This technique revealed interesting and unexpected findings. For example, a recent study observed view-point and size dependent coding at intermediate processing stages (V4, V3A, MT, and V7), whereas responses in higher visual areas were view-invariant (Konen and Kastner, 2008). Remarkably, these view-invariant representations were not only found in the ventral (e.g., LOC), but also in the dorsal pathway (e.g., IPS). The dorsal "where/how" or "perception-for-action" pathway is involved in visually guided actions toward objects rather than in identifying objects—which is mainly performed by the ventral or "what" pathway (Goodale and Milner, 1992; Ungerleider and Haxby, 1994). For this role, maintaining view-point dependent information in higher dorsal areas seems important, which however was thus not confirmed by the view-invariant results in IPS (but see James et al., 2002). Likewise, another recent study (Dilks et al., 2011) revealed an unexpected tolerance for mirror-reversals in visual scenes in a parahippocampal area thought to play a key role in navigation (e.g., Janzen and van Turennout, 2004) and reorientation (e.g., Epstein and Kanwisher, 1998), functions for which view-dependent information is essential. Furthermore, mixed findings have been reported for the object-selective LOC. For example, different findings on size, position, and viewpoint-invariant representations in different subparts of the LOC have been found (Grill-Spector et al., 1999; James et al., 2002; Vuilleumier et al., 2002; Valyear et al., 2006; Dilks et al., 2011). These divergent findings might be partly related to intricacies inherent to the fMRI-A approach (e.g., Krekelberg et al., 2006), and its sensitivity to the design used (Grill-Spector et al., 2006) and varying attention (Vuilleumier et al., 2005) and task demands (e.g., Ewbank et al., 2011). The latter should not be regarded as obscuring confounds however, since they appear to strongly contribute to our skilled performance. Object perception is accompanied by cognitive processes supporting fast (e.g., extracting the "gist" of a scene, attentional selection of relevant objects) and accurate (e.g., object-verification, semantic interpretation) object identification for subsequent goal-directed use of the object (e.g., grasping; tool-use). These processes engage widespread memory- and frontoparietal attention-related areas interacting with object processing in the visual system (Corbetta and Shulman, 2002; Bar, 2004; Ganis et al., 2007). As the involvement of such top-down processes might be particularly pronounced in humans—and weaker or even absent in monkeys and machines respectively—efforts to integrate computational modeling with *human* neuroimaging remain essential (see Tagamets and Horwitz, 1998; Corchs and Deco, 2002 for earlier work).

With the advent of ultra-high field fMRI ($\geq$7 Tesla scanners), both the sensitivity (due to increases in signal-to-noise ratio linearly dependent on field strength) and the specificity (due to a stronger contribution of gray-matter microvasculature compared to large draining veins and less partial volume effects) of the acquired signal improves significantly, providing data at a level of detail which previously was only available via invasive optical imaging in non-human species. The functional visual system can be spatially sampled in the range of hundreds of microns, which is sufficient to resolve activation at the cortical column (Yacoub et al., 2008; Zimmermann et al., 2011) and layer (Polimeni et al., 2010) level. Given that cortical columns are thought to provide the organizational structure forming computational units involved in visual feature processing (Hubel and Wiesel, 1962; Tanaka, 1996; Mountcastle, 1997), the achievable resolution at ultra-high fields will therefore not only produce more detailed maps, but really has the potential to yield new vistas on within-area operations.

INTEGRATION OF COMPUTATIONAL AND EXPERIMENTAL FINDINGS IN CBS

The approach we propose is to project the predicted activity in a modeled area onto corresponding cortical regions where empirical data are collected (**Figure 1**). By interfacing empirical and simulated data in one anatomical "brain space", direct and quantitative mutual hypothesis testing based on predicted and observed spatiotemporal activation patterns can be achieved. More specifically, modeled units (e.g., cortical columns) are 1-to-1 mapped to corresponding neuroimaging units (e.g., voxels, vertices) in the empirically acquired brain model (e.g., cortical gray matter surface). As a result, a running network simulation creates spatiotemporal data directly on a linked brain model, enabling highly specific and accurate comparisons between neuroimaging and neurocomputational data in the temporal as well as spatial domain. Note that in CBS (as implemented in Neurolator 3D; Goebel, 1993), computational and neuroimaging units can flexibly represent various neural signals (e.g., fMRI, EEG, MEG, fNIRS, or intracranial recordings). Furthermore, both hidden and output layers of the neural network can be projected to the brain model, providing additional flexibility to the framework as predicted and observed activations can be compared at multiple selected processing stages simultaneously (see **Figure 2** for an example).

To model the human object recognition system, we developed large-scale networks of cortical column units, which dynamics can either reflect the spike activity, integrated synaptic activity, or oscillating activity (when modeled as burst oscillators), resulting from excitatory and inhibitory synaptic input. To create simulated spatiotemporal patterns, each unit of a network layer (output and/or hidden) is linked to a topographically corresponding patch on a cortical representation via a so-called Network-to-Brain Link (NBL). Via this link, activity of modeling units in the running network is transformed into timecourses of neuroimaging units, spatially organized in an anatomical coordinate system. Importantly, when simulated and measured data co-exist in the same representational space, the same analysis tools (e.g., MVPA, effective connectivity analysis) can be applied to both data sets allowing for quantitative comparisons (**Figure 2**). See Peters et al. (2010) for further details.

We propose that such a tight integration of neuroimaging and modeling data allows reciprocal fine-tuning and facilitates hypothesis testing at a mechanistic level as it leads to falsifiable predictions that can subsequently be empirically tested. Importantly, there is a direct topographical correspondence between computational (cortical columnar) units at the model and brain level. Moreover, comparisons between simulated and empirical data are not limited to activity patterns in output stages

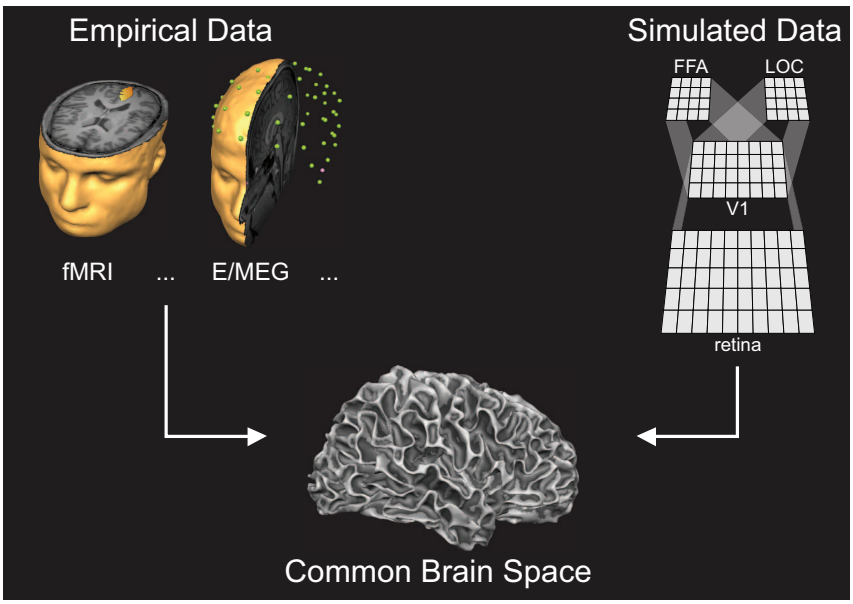

FIGURE 1 | The Common Brain Space framework: measured neuroimaging data (left panel) and simulated data (right panel) are projected to the same anatomical brain space via network-brain links (see **Figure 2**).

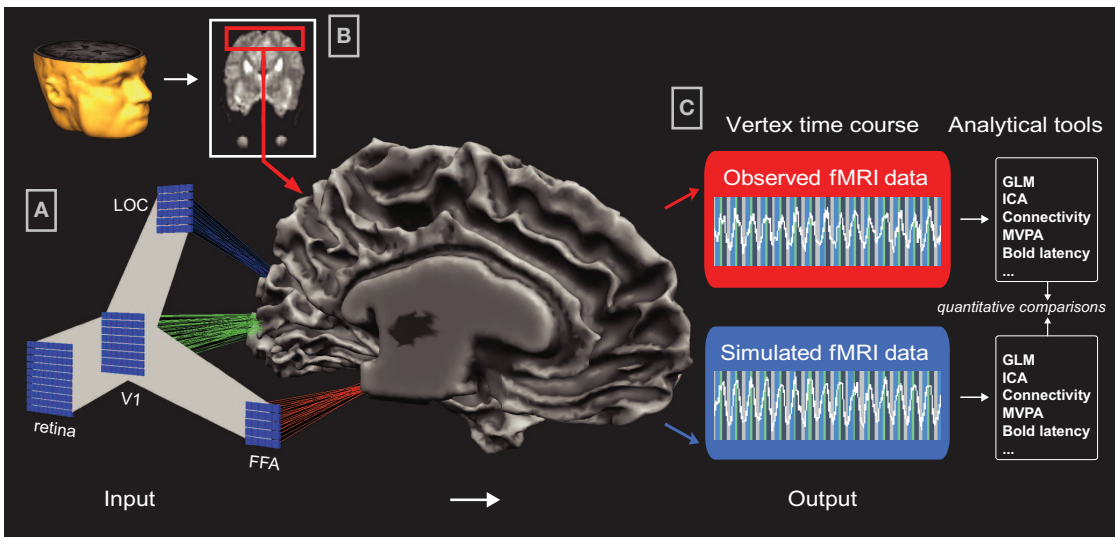

FIGURE 2 | Data Integration in Common Brain Space. Input: **(A)** Visualization of Common Brain Space (CBS) in Neurolator: Each computational unit of a neural network layer is separately connected to a topographically corresponding location on the cortical sheet via a Network–Brain Link (NBL). In this example, model layers V1, LOC, and FFA are connected to the corresponding brain regions V1, LOC, and FFA on a mesh reconstruction of an individual's gray-white matter boundary. For this participant, V1, LOC, and FFA were localized using standard retinotopy and related fMRI Region-of-Interest mapping techniques. By connecting a running neural network, activity in the connected layers is projected to the cortical sheet via the NBLs, creating spatially specific timecourses. **(B)** In Neurolator, functional MRI data can be projected on the cortical mesh, similar to the standard functional-anatomical data co-registration applied in fMRI analyses. Output: **(C)** Depending on display mode, cortical patches (i.e., vertices) either represent the empirical or the simulated fMRI data. Since the observed and simulated datasets are in the same anatomical space, identical fMRI analyses tools can be used to analyze observed and simulated timeseries.

(i.e., object-selective areas in anterior IT such as FFA or even more anterior in putative "face exemplar" regions; Kriegeskorte et al., 2007), but also at intermediate stages (such as V4 and LOC). Interpreting the role of feature representations at intermediate stages may be essential for a comprehensive brain model of object recognition (Ullman et al., 2002).

Studying several stages of the visual hierarchy simultaneously, by quantitatively comparing ongoing visual processes across

stages both within and between the simulated and empirically acquired dataset, may help to clarify how higher-order invariant representations are created from lower-level features in several ways. Firstly, this may reveal how object-coding changes along the visual pathway. Incoming percepts might be differently transformed and matched to stored object representations at several stages, with view-dependent matching at intermediate stages and matching of only informative properties (Biederman, 1987; Ullman et al., 2001) at later stages. Secondly, monitoring activity patterns at multiple processing stages simultaneously is desirable, given that early stages are influenced by processing in later stages. To facilitate object recognition, invariant information is for example fed back from higher to early visual areas (Williams et al., 2008), suggesting that object perception results from a dynamic interplay between visual areas. Finally, it is important to realize that such top-down influences are not limited to areas within the classical visual hierarchy, but also engage brain-wide networks involved in "initial guessing" (Bar et al., 2006), object selection (Serences et al., 2004), context integration (Graboi and Lisman, 2003; Bar, 2004), and object verification (Ganis et al., 2007). Such functions should be incorporated in computational brain models to fully comprehend what makes human object recognition so flexible, fast, and accurate. Modeling higher cognitive functions is in general challenging, but may be aided by considering empirical observations in object perception studies where the level of top-down processing varies (e.g., Ganis et al., 2007). The interactions between the visual pathway and frontoparietal system revealed by such fMRI studies can be compared at multiple processing stages to simulations, allowing a more subtle, process-specific fine-tuning of the modeled areas.

A number of recent fMRI studies applied en- and decoding techniques developed in the field of Machine Learning and Computer Vision, to interpret their data (Kriegeskorte et al., 2008; Miyawaki et al., 2008; Haxby et al., 2011; Naselaris et al., 2011; see LaConte, 2011 for an extention to Brain-Computer-Interfaces), showing that both fields are starting to approach each other. For example, by summarizing the complex statistical properties of natural images using a computer vision technique, a visual scene percept could be successfully reconstructed from fMRI activity (Naselaris et al., 2009). The trend to investigate natural vision is noteworthy, given that processing cluttered and dynamic natural visual input rather than artificially created isolated objects poses additional challenges to the visual system (Einhäuser and König, 2010). We believe that now columnar-level imaging is in reach with the advent of high-resolution fMRI (in combination with the recently developed en- and decoding fMRI methods) the time has come to more directly integrate computational and experimental neuroscience, and test which predictions of the various object recognition models are supported by this new type of empirical evidence.

ACKNOWLEDGMENTS

This work received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 269853.

REFERENCES

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., and Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.

Booth, M. C., and Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* 8, 510–523.

Bülthoff, H., and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. U.S.A.* 89, 60–64.

Carlson, E. T., Rasquinha, R. J., Zhang, K., and Connor, C. E. (2011). A sparse object coding scheme in area V4. *Curr. Biol.* 21, 288–293.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Corchs, S., and Deco, G. (2002). Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. *Cereb. Cortex* 12, 339–348.

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., and Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *J. Neurosci.* 31, 11305–11312.

Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473.

Edelman, S., and Bülthoff, H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Res.* 32, 2385–2400.

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R. J., and Rees, G. (2008). fMRI activity patterns in human LOC carry information about object exemplars within category. *J. Cogn. Neurosci.* 20, 356–370.

Einhäuser, W., Hipp, J., Eggert, J., Körner, E., and König, P. (2005). Learning viewpoint invariant object representations using a temporal coherence principle. *Biol. Cybern.* 93, 79–90.

Einhäuser, W., and König, P. (2010). Getting real-sensory processing of natural stimuli. *Curr. Opin. Neurobiol.* 20, 389–395.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23, 115–125.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601.

Ewbank, M. P., Lawson, R. P., Henson, R. N., Rowe, J. B., Passamonti, L., and Calder, A. J. (2011). Changes in 'top-down' connectivity underlie repetition suppression in the ventral visual pathway. *J. Neurosci.* 31, 5635–5642.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cereb. Cortex* 1, 1–47.

Fleuret, F., Li, T., Dubout, C., Wampler, E. K., Yantis, S., and Geman, D. (2011). Comparing machines and humans on a visual categorization test. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17621–17625.

Franzius, M., Wilbert, N., and Wiskott, L. (2011). Invariant object recognition and pose estimation with slow feature analysis. *Neural. Comput.* 23, 2289–2323.

Freiwald, W. A, and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., and Düzel, E. (2010). Theta-coupled periodic replay in working memory. *Curr. Biol.* 20, 606–612.

Fukushima, K. (1982). Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recogn.*15, 455–469.

Ganis, G., Schendan, H. E., and Kosslyn, S. M. (2007). Neuroimaging evidence for object model verification theory: role of prefrontal control in visual object categorization. *Neuroimage* 34, 384–398.

Goebel, R. (1993). "Perceiving complex visual scenes: an oscillator neural network model that integrates selective attention, perceptual organisation, and invariant recognition," in *Advances in Neural Information Processing Systems,*

Vol. 5, eds J. Giles, C. Hanson, and S. Cowan (San Diego, CA: Morgan Kaufmann), 903–910.

Goebel, R., and De Weerd, P. (2009). "Perceptual filling-in: from experimental data to neural network modeling," in *The Cognitive Neurosciences* Vol. 6, ed M. Gazzaniga (Cambridge, MA: MIT Press), 435–456.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Graboi, D., and Lisman, J. (2003). Recognition by top-down and bottom-up processing in cortex: the control of selective attention. *J. Neurophysiol.* 90, 798–810.

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.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23.

Grill-Spector, K., and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta. Psychol. (Amst.)* 107, 293–321.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., and Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Curr. Biol.* 17, 323–328.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A. (2002). Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron* 35, 793–801.

Janzen, G., and van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nat. Neurosci.* 7, 673–677.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kayaert, G., Biederman, I., and Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *J. Neurosci.* 23, 3016–3027.

Krekelberg, B., Boynton, G. M., and van Wezel, R. J. (2006). Adaptation: from single cells to BOLD signals. *Trends Neurosci.* 29, 250–256.

Konen, C. S., and Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat. Neurosci.* 11, 224–231.

Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20600–20605.

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.

LaConte, S. M. (2011). Decoding fMRI brain states in real-time. *Neuroimage* 56, 440–454.

Li, N., and DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321, 1502–1507.

Li, N., and DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron* 67, 1062–1075.

Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563.

Lowe, D. G. (2000). "Towards a computational model for object recognition in IT cortex," *First IEEE International Workshop on Biologically Motivated Computer Vision* (Seoul, Korea), 1–11.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* 60, 91–110.

Malach, R. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8135–8139.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman.

Marr, D., and Nishihara, K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond. B Biol. Sci.* 200, 269–294.

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299.

Miller, E. K., and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science* 263, 520–522.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., Sadato, N., and Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60, 915–929.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120, 701–722.

Naselaris, T., Kay, K. N., Nishimoto, S., and Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage* 56, 400–410.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., and Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 902–915.

Perrett, D. I., Oram, M. W., and Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67, 111–145.

Peters, J. C., Jans, B., van de Ven, V., De Weerd, P., and Goebel, R. (2010). Dynamic brightness induction in V1: analyzing simulated and empirically acquired fMRI data in a "common brain space" framework. *Neuroimage* 52, 973–984.

Poggio, T., and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature* 343, 263–266.

Polimeni, J. R., Fischl, B., Greve, D. N., and Wald, L. L. (2010). Laminar analysis of 7T BOLD using an imposed activation pattern in human V1. *Neuroimage* 52, 1334–1346.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107.

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

Roberts, L. G. (1965). "Machine perception of 3-D solids," in *Optical and Electro-Optical Information Processing*, ed J. T. Tippet (Cambridge, MA: MIT Press), 159–197.

Rust, N. C., and Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30, 12978–12995.

Sayres, R., and Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *J. Neurophysiol.* 100, 249.

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., and Yantis, S. (2004). Control of object-based attention in human cortex. *Cereb. Cortex* 14, 1346–1357.

Sountsov, P., Santucci, D. M., and Lisman, J. E. (2011). A biologically plausible transform for visual recognition that is invariant to translation, scale, and rotation. *Front. Comput. Neurosci.* 5:53. doi: 10.3389/fncom.2011.00053

Stevens, C. (2000). Models are common; good theories are scarce. *Nature* 3, 92037.

Tagamets, M. A, and Horwitz, B. (1998). Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cereb. Cortex* 8, 310–320.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.

Tarr, M. J., and Bülthoff, H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple-views? *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1494–1505.

Tarr, M. J., and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cogn. Psychol.* 21, 233–282.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.

Torralba, A., Fergus, R., and Freeman, W. T. (2008). 80 million tiny images: a large data set for nonparametric object and scene recognition. *IEEE. Trans. Pattern. Anal. Mach. Intell.* 30, 1958–1970.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition* 32, 193–254.

Ullman, S., Sali, E. and Vidal-Naquet, M. (2001). "A fragment-based approach to object representation and classification," in *International Workshop on Visual Form*, eds A. Arcelli, L. P. Cordella, and G. Sanniti di Baja (Berlin: Springer), 85–100.

Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat. Neurosci.* 5, 682–687.

Ungerleider, L. G., and Haxby, J. V. (1994). "What" and "where" in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.

Valyear, K. F., Culham, J. C., Sharif, N., Westwood, D., and Goodale, M. A. (2006). A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia* 44, 218–228.

Viola, P., and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Comput. Vis. Pattern Recog.* 1, I-511–I-518.

Vuilleumier, P., Henson, R. N., Driver, J., and Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat. Neurosci.* 5, 491–499.

Vuilleumier, P., Schwartz, S., Dolan, R. J., and Driver, J. (2005). Selective attention modulates neural substrates of repetition priming and "implicit" visual memory: suppressions and enhancements revealed by fMRI. *J. Cogn. Neurosci.* 17, 1245–1260.

Wersing, H., and Körner, E. (2003). Learning optimized features for hierarchical models of invariant object recognition. *Neural. Comput.* 15, 1559–1588.

Williams, M. A, Baker, C. I., Op de Beeck, H. P., Shim, W. M., Dang, S., Triantafyllou, C., and Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nat. Neurosci.* 11, 1439–1445.

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

Zimmermann, J., Goebel, R., De Martino, F., van de Moortele, P.-F., Feinberg, D., Adriany, G., Chaimow, D., Shmuel, A., Uğurbil, K., and Yacoub, E. (2011). Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PLoS One* 6:e28716. doi: 10.1371/journal.pone.0028716

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2011; accepted: 24 February 2012; published online: 09 March 2012.

Citation: Peters JC, Reithler J and Goebel R (2012) Modeling invariant object processing based on tight integration of simulated and empirical data in a Common Brain Space. Front. Comput. Neurosci. ***6****:12. doi: 10.3389/fncom.2012.00012*