



# How Invariant Feature Selectivity Is Achieved in Cortex

Tatyana O. Sharpee\*

Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA

Parsing the visual scene into objects is paramount to survival. Yet, how this is accomplished by the nervous system remains largely unknown, even in the comparatively well understood visual system. It is especially unclear how detailed peripheral signal representations are transformed into the object-oriented representations that are independent of object position and are provided by the final stages of visual processing. This perspective discusses advances in computational algorithms for fitting large-scale models that make it possible to reconstruct the intermediate steps of visual processing based on neural responses to natural stimuli. In particular, it is now possible to characterize how different types of position invariance, such as local (also known as phase invariance) and more global, are interleaved with nonlinear operations to allow for coding of curved contours. Neurons in the mid-level visual area V4 exhibit selectivity to pairs of even- and odd-symmetric profiles along curved contours. Such pairing is reminiscent of the response properties of complex cells in the primary visual cortex (V1) and suggests specific ways in which V1 signals are transformed within subsequent visual cortical areas. These examples illustrate that large-scale models fitted to neural responses to natural stimuli can provide generative models of successive stages of sensory processing.

## OPEN ACCESS

### Edited by:

Christian Gonzalez-Billault,  
University of Chile, Chile

### Reviewed by:

Taro Toyozumi,  
Riken Brain Science Institute, Japan  
Arianna Maffei,  
Stony Brook University, USA

### \*Correspondence:

Tatyana O. Sharpee  
sharpee@salk.edu

**Received:** 29 April 2016

**Accepted:** 05 August 2016

**Published:** 23 August 2016

### Citation:

Sharpee TO (2016) How Invariant  
Feature Selectivity Is Achieved in  
Cortex.  
*Front. Synaptic Neurosci.* 8:26.  
doi: 10.3389/fnsyn.2016.00026

**Keywords:** object recognition, Convolutional Neural Networks (CNN), visual system, area V4, phase invariance, quadrature model, curvature, auditory system

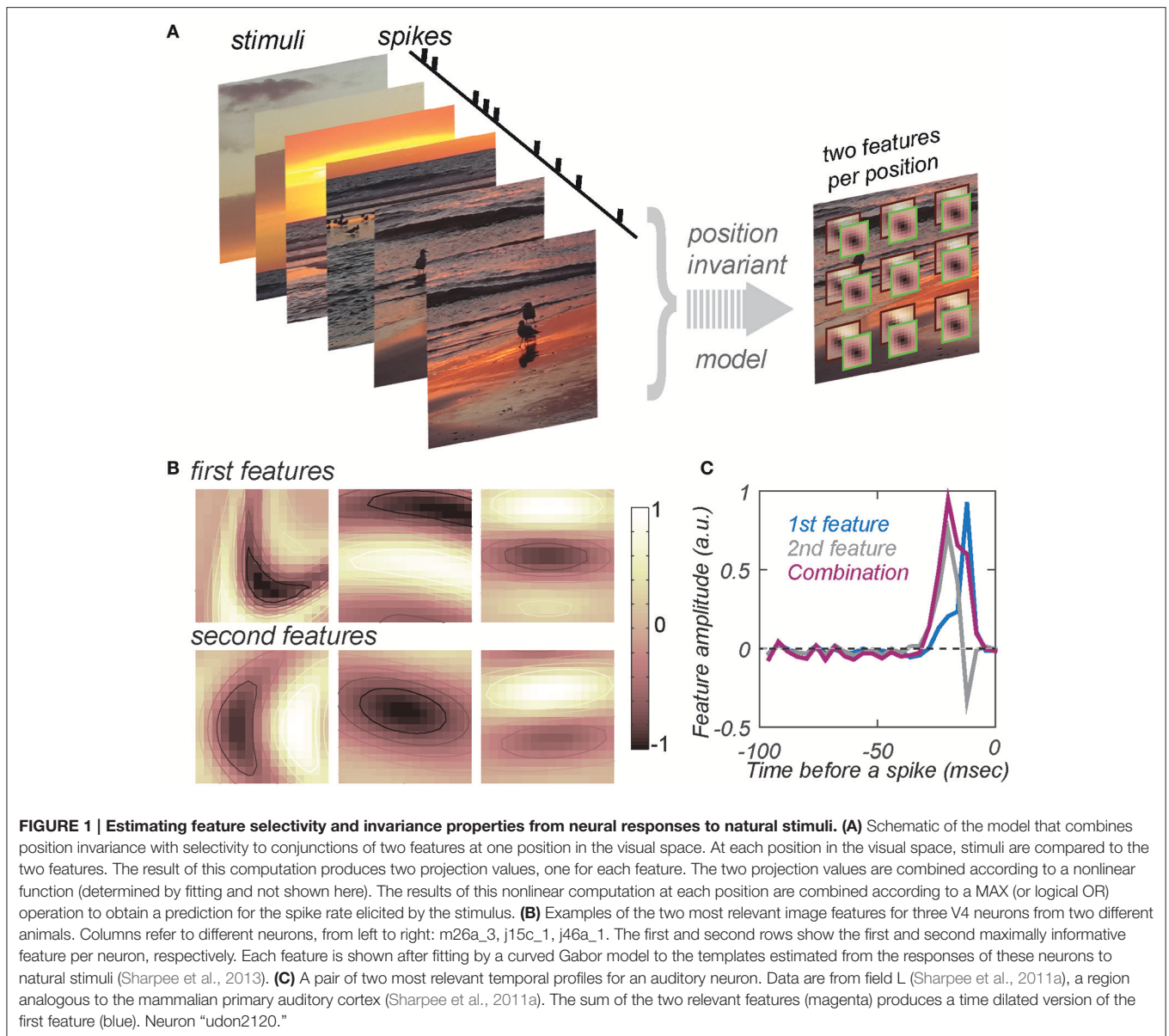
The current predominant hypothesis is that robust object recognition is made possible by transforming detailed signal representations to representations that encode objects independent of the viewing position (DiCarlo and Cox, 2007; Serre et al., 2007; DiCarlo et al., 2012). Such object-centered representations make it possible to perform fine discrimination, because these representations combine signals from viewing conditions wherein two objects might appear similar and where they are easily distinguishable. Achieving such object-centered representations is not trivial because in most cases integration across viewing positions destroys the specificity to configuration of object parts that is essential for correct identification (Ullman and Soloviev, 1999). Empirical studies in computer vision emphasize that increases in, for example, position tolerance have to be gradual and have to be interleaved with increases in specificity to the more complex features that will ultimately make it possible to distinguish between different objects (Ullman and Soloviev, 1999). How these representations are built in the visual system remains largely unknown. Similar computational tasks need to be solved by other sensory systems, including the somatosensory (Maravall and Diamond, 2014) and auditory systems (King and Nelken, 2009; Theunissen and Elie, 2014). Specifically, auditory perception includes a tolerance to changes in loudness, cadence and pitch (Trefethen and Embree, 2005). Again, however, the details of signal transformations within the auditory system remain to be worked out.

In this regard, large-scale models can provide vital information about how signals are transformed across their sensory processing pathways. So far, we know that neurons in early stages of cortical processing are primarily driven by simple stimulus features. Examples of such features include edges, in the case of neurons in the primary visual cortex (V1), and analogous features in the space of spectrotemporal modulations for neurons in primary auditory cortex or its analog in birds (Nagel and Doupe, 2008; Theunissen and Elie, 2014). Neurons at later stages tend to be selective for more complex combinations of stimulus features (Connor et al., 2007). For example, neurons in the mid-level visual area V4 exhibit selectivity for contour curvature (Connor et al., 2007). Neurons at subsequent stages of visual processing, such as in the inferotemporal (IT) cortex, exhibit selectivity for faces and their components (Tsao and Livingstone, 2008), as well as other objects of large biological significance (Desimone et al., 1984). Concomitant with the complexity of image features that drive the responses of visual neurons from V1 to V4 to IT, there is also an increase in the degree of tolerance that the responses of these neurons exhibit when relevant image features are displaced or scaled in size (DiCarlo et al., 2012; Roe et al., 2012). Importantly, artificial neural networks with this general structure can be optimized to reach human levels of categorization performance on a variety of visual recognition tasks (Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014). Thus, different sensory systems are all organized hierarchically with a progressive increase in the invariance and selectivity of neural responses to complex stimulus features (Connor et al., 2007; Meliza and Margoliash, 2010; DiCarlo et al., 2012; Roe et al., 2012). Nevertheless, the specific routes that signals take within the mid- and high-level sensory areas are difficult to characterize because they involve multiple intermediate nonlinear transformations and an incredible degree of convergence across brain regions. For example, some estimates suggest that a single neuron in area V4 can combine signals that originate from a substantial fraction of the V1 surface (Motter, 2009). If this pooling were indiscriminative, without any guiding principles, then this would seem to preclude any functional object recognition thought to be mediated by these brain regions (DiCarlo et al., 2012; Roe et al., 2012). Complicating the matter further, the process of feature extraction is a dynamic process (Olshausen et al., 1995) that is affected by neural adaptation to stimulus statistics (Sharpee et al., 2006; McManus et al., 2011) as well as by cognitive tasks, such as attention and perceptual learning (Ito et al., 1998; Ito and Gilbert, 1999).

Despite the difficulties, some progress can be made by fitting neural responses with multi-scale computational models that use built-in constraints to reduce the number of parameters incurred when characterizing the feature selectivity of mid- and high-level sensory neurons. For visual neurons, position invariance is one of the dominant constraints (Bouvier et al., 2009; Lee et al., 2009). Models that incorporate position invariance explicitly are known as hierarchical convolution networks (Le Cun et al., 1989; Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014; Vintch et al., 2015). Such models achieve good performance

on the object recognition task. However, the computations performed by the optimized models are difficult to interpret (Yamins and DiCarlo, 2016). To circumvent this problem, one can develop methods that explicitly determine the features that drive the responses of any given neuron while simultaneously taking into account position invariance (Eickenberg et al., 2012; Sharpee et al., 2013; Zeiler and Fergus, 2014; Vintch et al., 2015). The corresponding model is schematically depicted in **Figure 1A**. Compared to standard models that estimate relevant features without position invariance (de Boer and Kuyper, 1968; Victor and Shapley, 1980; Chichilnisky, 2001; Nykamp and Ringach, 2002; Bialek and de Ruyter van Steveninck, 2005; Schwartz et al., 2006; Fitzgerald et al., 2011), convolutional models drastically reduce the number of independent parameters when they require the relevant image features to be the same for each position within the neuron's response field. Although the model estimates only two relevant image features per position, when pooling across positions is taken into account, this is equivalent to estimating models with as many as ~50–100 relevant image features. With such a reduction in model complexity, it becomes feasible to begin deciphering how neural circuits simultaneously achieve invariance and selectivity for complex stimulus features.

Models of this structure have recently been used to describe how neurons in the mid-level visual area V4 encode natural stimuli (Sharpee et al., 2013). One of the concerns when fitting such models is that the stimulus set needs to be diverse enough to probe different aspects of the neural response. Stimuli from the natural sensory environment fulfill this requirement. Natural stimuli also elicit robust responses of neurons at different stages of sensory processing. In particular, mid- and high-level sensory neurons exhibit stronger responses when exposed to natural stimuli as compared to randomized inputs (Sen et al., 2001). Historically, randomized stimuli have primarily been used to characterize neural feature selectivity because they allow for computationally simpler estimation procedures (Bialek and de Ruyter van Steveninck, 2005; Gollisch, 2006; Schwartz et al., 2006; Dimitrov et al., 2009; Samengo and Gollisch, 2012). However, the increased availability of computing resources now makes estimation procedures tenable with natural stimuli. A typical dataset of responses from an individual neuron includes ~300 movie segments as stimuli, each containing ~100 frames updated at 30 Hz. (The duration of individual movie segments is limited by how long animals can maintain fixation in awake experiments). Thus, models are typically fit using ~30,000 stimulus/response associations. Such large numbers of stimuli are needed in order to probe the neural response function under a broad range of conditions, and because it is not known *a priori* which movie segment will elicit a high firing rate from a given neuron. The fitting of the model used in **Figure 1A** to the responses of individual neurons produces (1) a pair of most relevant image features for a neuron, (2) the nonlinear function describing how these two features jointly affect its neural response, and (3) the range of position invariance, defined as the range of positions in the visual space across which signals are combined according to logical OR or Max operations (**Figure 1A**). All of these parameters can be estimated through



maximum likelihood fitting (or related methods) based on neural responses to a large set of stimuli. The position invariance can be modeled either with uniform (Sharpee et al., 2013) or graded (Vintch et al., 2012) contributions across positions to the measured neural response. Once a convolutional model is fitted to the responses of a set of recorded neurons, the distribution of its parameters also produces a so-called generative model (Yamins and DiCarlo, 2016). Generative models are those that can recreate a set of responses across a neural population, yielding a distribution of parameters that best characterize the feature selectivity and invariance ranges. This distribution is obtained by fitting the model to a set of recorded neurons. With this setup, we can now discuss what convolutional models have revealed about the distribution of these parameters in mid-level visual area V4.

## CURVATURE SELECTIVITY

The reconstructions of V4 neural responses to natural stimuli indicate selectivity to segments of curve contours (Sharpee et al., 2013); examples of the relevant features are given in **Figure 1B**. Obtained with natural stimuli, this observation extends previous reports of curvature selectivity obtained with curved parametric stimuli (Gallant et al., 1996; Pasupathy and Connor, 1999, 2001, 2002; David et al., 2006; Connor et al., 2007) to the case of more diverse stimulus conditions. It is worth noting that natural stimuli were optimized neither for curvature selectivity nor for a particular neuron or area. In fact, the same set of stimuli when used in V1 yields selectivity to straight contours (Sharpee et al., 2006, 2008). This insures that any curvature selectivity obtained by analyzing V4 responses reflects genuine aspects of

their feature selectivity that are not influenced by the stimulus properties themselves. Furthermore, the fact that curvature selectivity also appears for mid-level units in artificial networks after optimization to maximize natural stimuli classification (Cadieu and Olshausen, 2012; Zeiler and Fergus, 2014) reinforces the notion that neural circuits are optimized for the structure of the natural sensory environment (Bialek, 2013).

## INVARIANCE-COMPLEXITY TRADE-OFF

The tightness of the preferred contour's curvature (in what follows we will refer to it as the preferred curvature value) varies substantially across neurons (Sharpee et al., 2013; **Figure 1B**). Intuitively, tighter curvatures can be viewed as describing more complex relevant image features compared to more shallow or straight contours. Supporting this intuition, tighter curvatures are also less frequently observed in natural scenes as compared to contours with more shallow curvatures (Lawlor and Zucker, 2013; Sharpee et al., 2013). Given that invariance and complexity concomitantly increase from stage to stage within the ventral visual pathway that performs visual object recognition, one might have expected that contours with tighter curvatures would be associated with larger ranges of position invariance. In this regard, the second observation came as a surprise because the opposite trend was observed by (Sharpee et al., 2013): neurons with smaller ranges of position invariance had preferred image features with tighter curvatures. This trend was reproduced in experiments with parametric stimuli (Nandy et al., 2013). It is also congruent with recent reports on the trade-off between position invariance and selectivity (Cadieu et al., 2007; Zoccolan et al., 2007; Rust and Dicarlo, 2010; Rust and DiCarlo, 2012). The preference of neurons with smaller ranges of position invariance for more tightly curved contours, together with the reduced frequency of curved contours in the natural environment, could explain at least in part the observed trade-off between invariance and selectivity (Rust and Dicarlo, 2010; Rust and DiCarlo, 2012).

## PHASE OR LOCAL POSITION INVARIANCE IN V4

Some convolutional models make it possible to estimate conjunctions of features that simultaneously affect the neural response at each retinotopic position (Eickenberg et al., 2012; Sharpee et al., 2013; Vintch et al., 2015). Applying these methods to V4 responses to natural stimuli, one finds that the two most relevant features of a given neuron often formed a pair of odd- and even-symmetric functions in the direction perpendicular to the preferred contour. This type of selectivity is reminiscent of the selectivity established for V1 complex cells (Adelson and Bergen, 1985). In V1, the so-called energy model accounts for the responses of V1 neurons as a quadratic function of the output of two relevant features: an even (cosine) and odd (sine) function in the direction perpendicular to the preferred orientation of the V1 neuron. With probed with grating stimuli, the output of this model does not vary with the

phase of the grating. For this reason, this type of selectivity to combinations of even- and odd-symmetric functions also became known as phase invariance. However, phase invariance also corresponds to local position invariance. This is because the odd-symmetric function can be well approximated as the difference of two slightly displaced even-symmetric functions (think of an edge as the difference between two bars). Thus, a model that allows for multiple relevant features can account for local position invariance even if does not have explicit convolutional architecture.

These arguments can now help interpret the results obtained in V4 using convolutional models (Sharpee et al., 2013). The convolutional model used in that study included only one explicit pooling stage. However, at each position, the two estimated most relevant features turned out to form a quadrature pair. This type of local feature selectivity indicates the presence of a local position invariance that is in addition to the more global position invariance captured by the convolutional part of the model. One important aspect of the quadrature pair selectivity observed locally in V4 is that it occurs with respect to curved contours, whereas in V1 it is observed with respect to straight contours. The most straightforward way of connecting these observations to the circuitry of the ventral visual pathway is to suppose that local position invariance corresponds to a summation of subunits representing V1 complex cells. This summation first takes place across different orientations, giving rise to curve contours, and then across positions, giving rise to positional invariance. Knowing that signals reach area V4 primarily through area V2, one could associate the second summation with a pooling of signals across V2 subunits. Further, the observed trade-off between preferred curvature and (global) invariance range (Sharpee et al., 2013) suggests that, for individual neurons, either a summation across orientations or across positions dominates.

The concept of local and global invariance is also directly applicable to other sensory circuits. For example, for the case of motion perception, neurons that project from V1 to MT are predominantly complex and orientation tuned (Movshon and Newsome, 1996). Given that MT neurons have 10 times larger receptive fields than the V1 complex cells whose responses they integrate (Simoncelli and Heeger, 1998), the responses of MT neurons would also be well described by a combination of local and global invariance.

In the auditory system, recent psychophysical studies found that birds attend to a mixture of local and global rhythmic features (Ten Cate et al., 2016). A re-examination of published neurophysiological data (Sharpee et al., 2011a) from field L, an area in bird's brain analogous to the mammalian primary auditory cortex, provides evidence for local invariance with respect to changes in cadence or time dilation in the responses of these neurons. Previous analyses showed that for neurons tuned to a specific frequency, the temporal profiles of the two most relevant features form a quadrature pair in a sense that they are described by a pair of integration/differentiation features (Sharpee et al., 2011a). This type of selectivity could be consistent with shifts in temporal offsets or temporal jitter (Aldworth et al., 2005; Dimitrov and Gedeon, 2006; Gollisch, 2006; Dimitrov



et al., 2009) as well as with changes in cadence or time dilation. However, detailed statistical analysis ruled out temporal jitter as the cause underlying integration/differentiation pair of features for that dataset (Sharpee et al., 2011a). Furthermore, for auditory signals, integration over temporal latencies would only be relevant in the context of binaural time differences, which were not analyzed in Sharpee et al. (2011a). On the other hand, an integration over different time dilation would be perceptually relevant (Nagel et al., 2010) and would result in pairs of features that could also be approximated as integration/differentiation. Indeed, in the case of selectivity to the temporal profile  $F(t)$  and its time-dilated version  $F(t/\tau)$ , for  $\tau \sim 1$ , one would expect to find a combination of features  $F(t)$  and  $F'(t)t$ . If  $F(t)$  has unimodal shape, the second feature would approximate a time derivative. **Figure 1C** shows how a pair of integration/differentiation relevant temporal profiles for an auditory neuron can produce two unimodal features, one of which is a time-dilated version of the other. It is worth noting that similar types of selectivity were observed in the peripheral olfactory system (Kim et al., 2011) as well as in the granular layers of A1 cortex (Atencio et al., 2008, 2009). Neurons in the infragranular of A1 cortex exhibited more complex forms of selectivity, potentially analogous to the curvature selectivity discussed here for V4 neurons (Atencio et al., 2009; Sharpee et al., 2011b). The relationship between the two most relevant features of A1 neurons with such complex forms of auditory selectivity could potentially be consistent with a model of local invariance with respect to dilation in time or frequency, but this hypothesis would need to be quantitatively tested in future work.

Overall, recent progress in experimental and computational methods for fitting large-scale models to neural responses to natural stimuli offers the hope of reconstructing detailed transformations that make biological vision so much more

efficient than machine vision. Clearly, the present models lack many of the important aspects of visual processing, including various forms of gain control (Carandini and Heeger, 2013), adaptive properties (Olshausen et al., 1995; Wark et al., 2007; McManus et al., 2011), and modulation by attention and cognitive tasks (Koch and Ullman, 1985; Olshausen et al., 1993; Ito et al., 1998; Ito and Gilbert, 1999). Increasingly more sophisticated models has been built for the retina that can relate better to the underlying neural circuitry (Kaardal et al., 2013; Freeman et al., 2015). Further, improvements in computational methods are needed to be able to scale and fit such detailed models to cortical responses.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

This research was supported by the National Science Foundation (NSF) CAREER award number 1254123 and IOS-1556388, the National Eye Institute of the National Institutes of Health under Award Number R01EY019493, NEI Core grant P30EY019005, and T32EY020503, McKnight Scholarship, Ray Thomas Edwards Career Award, and the Janelia Visitor Scientist Program.

## ACKNOWLEDGMENTS

The author also would like to acknowledge discussions with William Bialek, Oleg Barabash, Minjoon Kouh, John Reynolds, Ryan Rowekamp, and Terrence Sejnowski.

## REFERENCES

- Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284–299. doi: 10.1364/JOSAA.2.000284
- Aldworth, Z. N., Miller, J. P., Gedeon, T., Cummins, G. I., and Dimitrov, A. G. (2005). Dejittered spike-conditioned stimulus waveforms yield improved estimates of neuronal feature selectivity and spike-timing precision of sensory interneurons. *J. Neurosci.* 25, 5323–5332. doi: 10.1523/JNEUROSCI.0359-05.2005
- Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2008). Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58, 956–966. doi: 10.1016/j.neuron.2008.04.026
- Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2009). Hierarchical computation in the canonical auditory cortical circuit. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21894–21899. doi: 10.1073/pnas.0908383106
- Bialek, W. (2013). *Biophysics: Searching for Principles*. Princeton, NJ; Oxford, UK: Princeton University Press.
- Bialek, W., and de Ruyter van Steveninck, R. R. (2005). Features and dimensions: motion estimation in fly vision. arXiv:q-bio/0505003, 1–18.
- Bouvier, J., Rosasco, L., and Poggio, T. (2009). “On invariance in hierarchical models,” in *Advances in Neural Information Processing Systems*, Vol. 22, eds Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, A. Culotta (Red Hook, NY: Curran Associates, Inc.), 162–170.
- Cadieu, C. F., and Olshausen, B. A. (2012). Learning intermediate-level representations of form and motion from natural movies. *Neural Comput.* 24, 827–866. doi: 10.1162/NECO\_a\_00247
- Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., and Poggio, T. (2007). A model of V4 shape selectivity and invariance. *J. Neurophysiol.* 98, 1733–1750. doi: 10.1152/jn.01265.2006
- Carandini, M., and Heeger, D. J. (2013). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3424
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network* 12, 199–213. doi: 10.1080/713663221
- Connor, C. E., Brincat, S. L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Curr. Opin. Neurobiol.* 17, 140–147. doi: 10.1016/j.conb.2007.03.002
- David, S. V., Hayden, B. Y., and Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *J. Neurophysiol.* 96, 3492–3505. doi: 10.1152/jn.00575.2006
- de Boer, E., and Kuypers, P. (1968). Triggered Correlation. *IEEE Trans. Biomed. Eng.* 15, 169–179. doi: 10.1109/TBME.1968.4502561
- 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.
- DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341. doi: 10.1016/j.tics.2007.06.010
- DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

- Dimitrov, A. G., and Gedeon, T. (2006). Effects of stimulus transformations on estimates of sensory neuron selectivity. *J. Comput. Neurosci.* 20, 265–283. doi: 10.1007/s10827-006-6357-1
- Dimitrov, A. G., Sheiko, M. A., Baker, J., and Yen, S. C. (2009). Spatial and temporal jitter distort estimated functional properties of visual sensory neurons. *J. Comput. Neurosci.* 27, 309–319. doi: 10.1007/s10827-009-0144-8
- Eickenberg, M., Rowekamp, R. J., Kouh, M., and Sharpee, T. O. (2012). Characterizing responses of translation-invariant neurons: maximally informative invariant dimensions. *Neural Comput.* 24, 2384–2421. doi: 10.1162/NECO\_a\_00330
- Fitzgerald, J. D., Rowekamp, R. J., Sincich, L. C., and Sharpee, T. O. (2011). Second order dimensionality reduction using minimum and maximum mutual information models. *PLoS Comput. Biol.* 7:e1002249. doi: 10.1371/journal.pcbi.1002249
- Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., et al. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. *Elife* 4:e05241. doi: 10.7554/eLife.05241
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., and Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* 76, 2718–2739.
- Gollisch, T. (2006). Estimating receptive fields in the presence of spike-time jitter. *Network* 17, 103–129. doi: 10.1080/09548980600569670
- Ito, M., and Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* 22, 593–604. doi: 10.1016/S0896-6273(00)80713-8
- Ito, M., Westheimer, G., and Gilbert, C. D. (1998). Attention and perceptual learning modulate contextual influences on visual perception. *Neuron* 20, 1191–1197. doi: 10.1016/S0896-6273(00)80499-7
- Kaardal, J., Fitzgerald, J. D., Berry, M. J. 2nd., and Sharpee, T. O. (2013). Identifying functional bases for multidimensional neural computations. *Neural Comput.* 25, 1870–1890. doi: 10.1162/NECO\_a\_00465
- Khaligh-Razavi, S. M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10:e1003915. doi: 10.1371/journal.pcbi.1003915
- Kim, A. J., Lazar, A. A., and Slutskiy, Y. B. (2011). System identification of *Drosophila* olfactory sensory neurons. *J. Comput. Neurosci.* 30, 143–161. doi: 10.1007/s10827-010-0265-0
- King, A. J., and Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat. Neurosci.* 12, 698–701. doi: 10.1038/nn.2308
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.
- Lawlor, M., and Zucker, S. W. (2013). “Third-order edge statistics: contour continuation, curvature, cortical connections,” in *Advances in Neural Information Processing Systems* 26, Vol. 26, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Red Hook, NY: Curran Associates, Inc.), 1763–1771.
- Le Cun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541–551. doi: 10.1162/neco.1989.1.4.541
- Lee, H., Grosse, R., Ranganath, R., and Ng, A. (2009). “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in *Proceedings of the Twenty-Sixth International Conference on Machine Learning* (New York, NY: Association for Computing Machinery), 873–880.
- Maravall, M., and Diamond, M. E. (2014). Algorithms of whisker-mediated touch perception. *Curr. Opin. Neurobiol.* 25, 176–186. doi: 10.1016/j.conb.2014.01.014
- McManus, J. N., Li, W., and Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9739–9746. doi: 10.1073/pnas.1105855108
- Meliza, C. D., and Margoliash, D. (2010). Emergence of selectivity and tolerance in the avian auditory cortex. *J. Neurosci.* 32, 15158–15168. doi: 10.1523/JNEUROSCI.0845-12.2012
- Motter, B. C. (2009). Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface. *J. Neurosci.* 29, 5749–5757. doi: 10.1523/JNEUROSCI.4496-08.2009
- Movshon, J. A., and Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* 16, 7733–7741.
- Nagel, K. I., and Doupe, A. J. (2008). Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron* 58, 938–955. doi: 10.1016/j.neuron.2008.04.028
- Nagel, K. I., McLendon, H. M., and Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J. Neurophysiol.* 104, 1426–1437. doi: 10.1152/jn.00028.2010
- Nandy, A. S., Sharpee, T. O., Reynolds, J. H., and Mitchell, J. F. (2013). The fine structure of shape tuning in area V4. *Neuron* 78, 1102–1115. doi: 10.1016/j.neuron.2013.04.016
- Nykamp, D. Q., and Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *J. Vis.* 2, 1–11. doi: 10.1167/2.1.1
- Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representations. *J. Comput. Neurosci.* 2, 45–62. doi: 10.1007/BF00962707
- Olshausen, B., Andersen, C., and Essen, D. V. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* 13, 4700–4719.
- Pasupathy, A., and Connor, C. E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* 82, 2490–2502.
- Pasupathy, A., and Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86, 2505–2519.
- Pasupathy, A., and Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/972
- Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., et al. (2012). Toward a unified theory of visual area v4. *Neuron* 74, 12–29. doi: 10.1016/j.neuron.2012.03.011
- 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. doi: 10.1523/JNEUROSCI.0179-10.2010
- Rust, N. C., and DiCarlo, J. J. (2012). Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J. Neurosci.* 32, 10170–10182. doi: 10.1523/JNEUROSCI.6125-11.2012
- Samengo, I., and Gollisch, T. (2012). Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond Gaussian stimuli. *J. Comput. Neurosci.* 34, 137–161. doi: 10.1007/s10827-012-0411-y
- Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J. Vis.* 6, 484–507. doi: 10.1167/6.4.13
- Sen, K., Theunissen, F. E., and Doupe, A. J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86, 1445–1458. Available online at: <http://jn.physiology.org/content/jn/86/3/1445.full.pdf>
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56
- Sharpee, T. O., Atencio, C. A., and Schreiner, C. E. (2011b). Hierarchical representations in the auditory cortex. *Curr. Opin. Neurobiol.* 21, 761–767. doi: 10.1016/j.conb.2011.05.027
- Sharpee, T. O., Kouh, M., and Reynolds, J. H. (2013). Trade-off between curvature tuning and position invariance in visual area V4. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11618–11623. doi: 10.1073/pnas.1217479110
- Sharpee, T. O., Miller, K. D., and Stryker, M. P. (2008). On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. *J. Neurophysiol.* 99, 2496–2509. doi: 10.1152/jn.01397.2007
- Sharpee, T. O., Nagel, K. I., and Doupe, A. J. (2011a). Two-dimensional adaptation in the auditory forebrain. *J. Neurophysiol.* 106, 1841–1861. doi: 10.1152/jn.00905.2010
- Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrink, S. P., Stryker, M. P., and Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature* 439, 936–942. doi: 10.1038/nature04519
- Simoncelli, E. P., and Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res.* 38, 743–761. doi: 10.1016/S0042-6989(97)00183-1
- Ten Cate, C., Spierings, M., Hubert, J., and Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* 7:730. doi: 10.3389/fpsyg.2016.00730

- Theunissen, F. E., and Elie, J. E. (2014). Neural processing of natural sounds. *Nat. Rev. Neurosci.* 15, 355–366. doi: 10.1038/nrn3731
- Trefethen, L. N., and Embree, M. (2005). *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton, NJ: Princeton University Press.
- Tsao, D. Y., and Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.* 31, 411–437. doi: 10.1146/annurev.neuro.30.051606.094238
- Ullman, S., and Soloviev, S. (1999). Computation of pattern invariance in brain-like structures. *Neural Netw.* 12, 1021–1036. doi: 10.1016/S0893-6080(99)00048-9
- Victor, J., and Shapley, R. (1980). A method of nonlinear analysis in the frequency domain. *Biophys. J.* 29, 459–483. doi: 10.1016/S0006-3495(80)85146-0
- Vintch, B., Movshon, J. A., and Simoncelli, E. P. (2015). A convolutional subunit model for neuronal responses in macaque V1. *J. Neurosci.* 35, 14829–14841. doi: 10.1523/JNEUROSCI.2815-13.2015
- Vintch, B., Zaharia, A., Movshon, J. A., and Simoncelli, E. P. (2012). “Efficient and direct estimation of a neural subunit model for sensory coding,” in *Advances in Neural Information Processing Systems 25*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Red Hook, NY: Curran Associates, Inc.), 3104–3112.
- Wark, B., Lundstrom, B. N., and Fairhall, A. (2007). Sensory adaptation. *Curr. Opin. Neurobiol.* 17, 423–429. doi: 10.1016/j.conb.2007.07.001
- Yamins, D. L., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244
- Yamins, D. L., 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. U.S.A.* 111, 8619–8624. doi: 10.1073/pnas.1403112111
- Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. *Lect. Notes Comput. Sc.* 8689, 818–833. doi: 10.1007/978-3-319-10590-1\_53
- Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J. Neurosci.* 27, 12292–12307. doi: 10.1523/JNEUROSCI.1897-07.2007

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

Copyright © 2016 Sharpee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.