5. Edgar, B.A. (2006). How flies get their size: genetics meets physiology. Nat. Rev. Genet. *7*, 907–916.

6. Campos, M., Surovtsev, I.V., Kato, S., Paintdakhi, A., Beltran, B., Ebmeier, S.E., and Jacobs-Wagner, C. (2014). A constant size extension drives bacterial cell size homeostasis. Cell *159*, 1433–1446.

7. Taheri-Araghi, S., Bradde, S., Sauls, J.T., Hill, N.S., Levin, P.A., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control and homeostasis in bacteria. Curr. Biol. *25*, 385–391.

8. Si, F., Li, D., Cox, S.E., Sauls, J.T., Azizi, O., Sou, C., Schwartz, A.B., Erickstad, M.J., Jun, Y., Li, X., and Jun, S. (2017). Invariance of initiation mass and predictability of cell size in Escherichia coli. Curr. Biol. *27*, 1278–1287.

9. Schreiber, G., Ron, E.Z., and Glaser, G. (1995). ppGpp-mediated regulation of DNA replication and cell division in Escherichia coli. Curr. Microbiol. *30*, 27–32.

10. Yao, Z., Davis, R.M., Kishony, R., Kahne, D., and Ruiz, N. (2012). Regulation of cell size in response to nutrient availability by fatty acid biosynthesis in Escherichia coli. Proc. Natl. Acad. Sci. USA *109*, E2561–E2568.

11. Hicks, D.L. (1967). Adipose tissue composition and cell size in fall migratory thrushes (Turdidae). The Condor *69*, 387–399.

12. Berleman, J., and Auer, M. (2013). The role of bacterial outer membrane vesicles for intra- and interspecies delivery. Environ. Microbiol. *15*, 347–354.

Visual Perception: Neural Networks for Stereopsis

Jenny C.A. Read[1] and Bruce G. Cumming[2]
[1]Newcastle University, Institute of Neuroscience, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK
[2]National Institutes of Health, National Eye Institute, Bldg 49 Room 2A50, Bethesda, Maryland 20892-4435, USA
Correspondence: jenny.read@ncl.ac.uk (J.C.A.R.), bgc@lsr.nei.nih.gov (B.G.C.)
http://dx.doi.org/10.1016/j.cub.2017.05.013

How does our brain use differences between the images in our two eyes, binocular disparities, to generate depth perception? New work shows that a type of neural network trained on natural binocular images can learn parameters that match key properties of visual cortex. Most information is conveyed by cells which sense differences between the two eyes' images.

Because our eyes are offset in the head, most points in a visual scene project to slightly different locations in the two retinae. These small differences are called binocular disparities, and they form the basis of our stereoscopic three-dimensional vision. The computation of binocular disparity begins in primary visual cortex, V1. A study [1] published recently in *Current Biology* has now shown that a neural network trained to discriminate depth in natural images learns parameters which match several properties of real V1 neurons.

V1 neurons can be broadly divided into two classes: simple cells and complex cells. Simple cells are characterised by a receptive field function, which specifies where in the image the cell responds to light. Typically, simple cell receptive fields consist of an ON region, where bright features in the image tend to excite the cell and dark features tend to inhibit its firing, and an OFF region, where the opposite is true — dark features excite the cell and bright ones inhibit it. A given simple cell behaves like a linear filter, representing the image at one location with a single number, although the firing rate is a nonlinear function of this number (negative firing rates are impossible, for example). Model binocular simple cells, with linear filters in each eye, are able to signal disparity [2].

Complex cells behave as if they receive input from several simple cells. They typically respond to both bright and dark features at a given location, with different simple cells contributing in each case. Many V1 complex cells are tuned to binocular disparity, thought to reflect the disparity tuning of their component simple cells. Traditional model complex cells are constructed in a very economical way with just four simple cells — this is known as the Binocular Energy Model [3] (Figure 1A). Mounting evidence in recent years suggests that real V1 complex cells are less economical, and receive input from a broader range of simple cells (Figure 1B; reviewed in [4]).

Recent computational work has provided an appealing account of why this might be the case. An important limitation of the traditional model is that the response depends on both the stimulus disparity and details of the monocular image (such as its contrast and spectral content). As a result, model complex cells frequently produce peaks in activity when the stimulus disparity is not the cell's preferred disparity. These *false matches* make it hard to decode disparity. Burge and Geisler [5] identified the small set (n = 8) of linear filters that provides the most information about disparity in natural images. Two key features of these 'ideal' filter sets seem to explain properties of real neurons. First, the filter shapes are more diverse than those used in a traditional complex cell model. Second, the filters often had different shapes in the two eyes (this is often called 'phase disparity' in neurons [6]). This means that the complex cells are most strongly activated, not by a single image feature that has been displaced by a disparity, but by different images in the two eyes. Although we had noted that such filters may help to recognize false matches [7], Burge and Geisler [5] proved that using such filters is optimal if the objective is to estimate disparity.

Welchman and Goncalves [1] used a similar approach, building a network

whose output layer consisted of a small number of complex cells, each tuned to a different disparity. These in turn were built from a large number of binocular simple cells. Their network included 28 different classes of simple cell, each with different binocular receptive fields and hence different disparity tuning. The model contains many different individual simple cells belonging to a given class, with identical disparity tuning but slightly different locations on the retina. This type of network, replicating identical receptive fields across the retina, is called 'convolutional'. Convolutional neural networks are very common in image processing. They exploit the fact that image statistics are very similar from place to place, so networks can learn much more efficiently if they tile the same receptive field across the whole retina, rather than trying to learn independently at each retinal location. This effectively exploits spatial pooling to make it easier to identify points in the two retinal images that correspond to a single point in visual space, an approach first explored by Qian [8].

Each complex cell in the Welchman–Goncalves model [1] receives input from the same set of 28 simple-cell classes. As each class is tuned to a different disparity, the weights from the different classes had to be different for each complex cell, in order to make complex cells be tuned to different disparities. The network learned a beautifully simple rule for achieving this: to make a complex cell tuned to disparity δ, have each simple cell project to it with a synaptic weight proportional to its mean response at disparity δ (normalized by the mean response to all stimuli). Welchman and Goncalves [1] show that this rule produces complex cells like those in [5], whose firing rate approximates the log-likelihood that the stimulus disparity is the cell's preferred value of δ.

Welchman and Goncalves' [1] complex cells also exhibit a property of real neurons that is not captured by the traditional energy model. When stimulated with patterns of random dots, both the model and real neurons preserve their disparity tuning. If the dot colors are reversed in one eye ('anticorrelated'), the resulting stimulus is something that cannot be produced in natural binocular viewing. Energy-model

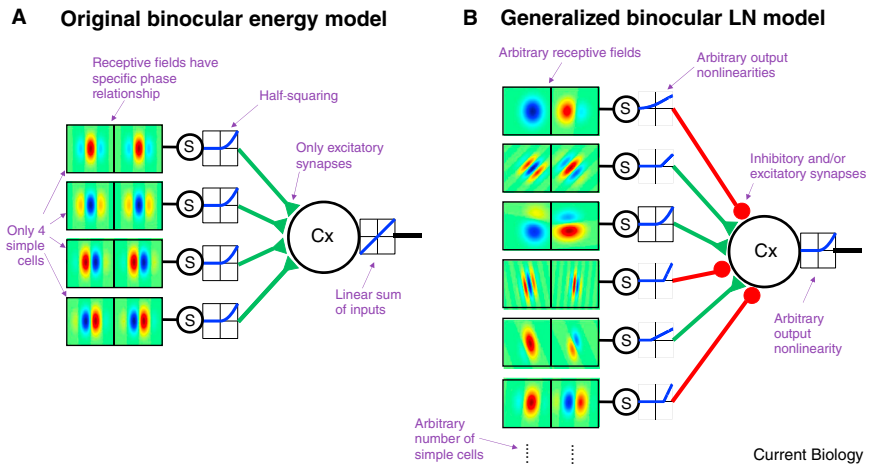

Figure 1. Network architecture for V1 disparity detectors.
(A) The original energy model [3] and (B) the generalisation proposed by [12]; the Welchman and Goncalves [1] network is of this form. Both of these are examples of 'LN' (linear/nonlinear) cascades. In both models, a disparity-tuned complex cell (Cx) receives input from several binocular simple cells (S), characterised by a linear binocular receptive field (pseudocolour). In the original model (A), the parameters of the model are severely constrained. The complex cell has no output nonlinearity and receives only excitatory input from the simple cells; the simple cells have halfwave rectification followed by squaring. The simple cells come in pairs with opposite receptive fields, each pair equivalent to a simple cell with a full squaring nonlinearity. This ensures that the disparity tuning curve has the same amplitude for both correlated and anti-correlated stimuli. Phase relations between the pairs ensure that the complex cell computes contrast energy, hence the name. In the generalised model (B), each simple cell may either excite or inhibit the complex cell; the receptive fields can have any form, and their output nonlinearities can be more general. In the Welchman-Goncalves model, the complex cell has no output nonlinearity while the simple cells have halfwave rectification with a threshold individual to each cell. One subtlety: this description is for a convolutional network with an average-pooling stage, whereas the Welchman-Goncalves [1] model uses max-pooling, adding a winner-take-all stage between groups of four adjacent simple cells. The authors do not attribute their results to the use of max-pooling vs average-pooling, and [15] found similar results with both, so for simplicity we neglect the distinction.

neurons produce inverted tuning curves to these stimuli. Although real neurons also display this inversion, they also show a weaker modulation to anticorrelated stimuli — the responses seem to sense the unnatural property of the stimulus, in a way that the traditional energy model does not. The complex cells produced by Welchman and Goncalves' [1] model reproduced this property of real neurons.

Welchman and Goncalves' [1] model cells are able to do this because (unlike in [5]) each simple cell is allowed a unique threshold. Previous models have exploited similar nonlinearities to achieve the same effect [9–13], but those models were constructed by hand precisely in order to explain the phenomenon. The remarkable feature of Welchman and Goncalves' [1] work is that the thresholds and the weights are learned simply by optimizing performance on a set of natural images. The resulting model cells then correctly predict how cortical neurons respond to a completely different, unnatural, stimulus. This suggests that the effect seen in real neurons results from a computation optimized to reduce the problem of false matches.

An interesting feature of the receptive fields learnt from the natural images concerns the type of disparity tuning they produce. After training, only seven out of the 28 classes of binocular simple cell show the type of disparity tuning known as *tuned-excitatory*, where the cell fires most to one particular preferred disparity [14]; 19 of the 28 (68%) are *tuned-inhibitory*, where the cell is silenced by one particular 'null' disparity and responds roughly equally to everything else. (Two classes in [1] have *odd-symmetric* tuning, where the tuning curve has a peak and trough of roughly equal amplitudes.) Welchman and Goncalves [1] point out that there is a good reason for this: tuned-inhibitory neurons convey more information than tuned-excitatory neurons. This is because of the confound between contrast and disparity, mentioned above. When a tuned-excitatory neuron sees its preferred disparity, its response will depend on the contrast of the image

feature there. A low-contrast feature at the preferred disparity may elicit less response than a higher-contrast feature at a different disparity, for example. When an ideal tuned-inhibitory neuron sees its null disparity, however, it is always silenced. It has receptive fields with opposite profiles in left and right eyes, so no matter how large or small the monocular responses are, they always cancel out perfectly. For this reason, a lack of response from a tuned-inhibitory cell is much stronger evidence that the stimulus is at a given disparity than a strong response from a tuned-excitatory cell, even if both responses are equally different from the cell's mean firing rate. This property, previously highlighted by [5], is presumably why the network so favoured learning opposite receptive fields in the two eyes. Hunter and Hibbard [15], who used independent subspace analysis to encode natural binocular images efficiently but without extracting disparity, found only 37% of their learnt cells were tuned-inhibitory.

While the Welchman-Goncalves [1] model is remarkably successful at accounting for a range of psychophysical and physiological data, it is not yet clear how literally we should interpret it as a description of neurophysiology. Their model complex cells receive input from over 4000 simple cells with different retinal locations, so have receptive fields with diameters 50% larger than those of simple cells. Both numbers are large for real V1 complex cells, so the model complex cells may best be viewed as representing an additional level of abstraction, perhaps beyond V1. More problematically, only around 16% of real neurons are tuned-inhibitory [6,14,16], raising the question of why the brain has not exploited the full potential of 'what not' detectors. One explanation could be that the model was trained only to encode disparity, whereas V1 is simultaneously encoding disparity, contrast, orientation and many other image properties. Conceivably, a network forced to encode these image properties as well might produce even more realistic solutions.

REFERENCES

1. Welchman, A.E., and Goncalves, N.R. (2017). "What not" detectors help the brain see in depth. Curr. Biol. *27*, 1403–1412.
2. Ohzawa, I., and Freeman, R.D. (1986). The binocular organization of simple cells in the cat's visual cortex. J. Neurophysiol. *56*, 221–242.
3. Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science *249*, 1037–1041.
4. Henriksen, S., Tanabe, S., and Cumming, B. (2016). Disparity processing in primary visual cortex. Philos. Trans. R. Soc. Lond. B. *371*, 20150255.
5. Burge, J., and Geisler, W.S. (2014). Optimal disparity estimation in natural stereo images. J. Vis. *14*, 1–1.
6. DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. (1991). Depth is encoded in the visual cortex by a specialised receptive field structure. Nature *352*, 156–159.
7. Read, J.C.A., and Cumming, B.G. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. Nat. Neurosci. *10*, 1322–1328.
8. Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. Neural Comput. *6*, 390–404.
9. Tanabe, S., Haefner, R.M., and Cumming, B.G. (2011). Suppressive mechanisms in monkey V1 help to solve the stereo correspondence problem. J. Neurosci. *31*, 8295–8305.
10. Lippert, J., and Wagner, H. (2001). A threshold explains modulation of neural responses to opposite-contrast stereograms. Neuroreport *12*, 3205–3208.
11. Read, J.C.A., Parker, A.J., and Cumming, B.G. (2002). A simple model accounts for the reduced response of disparity-tuned V1 neurons to anti-correlated images. Vis. Neurosci. *19*, 735–753.
12. Haefner, R.M., and Cumming, B.G. (2008). Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. Neuron *57*, 147–158.
13. Tanabe, S., and Cumming, B.G. (2014). Delayed suppression shapes disparity selective responses in monkey V1. J. Neurophysiol. *111*, 1759–1769.
14. Poggio, G.F., Gonzalez, F., and Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. J. Neurosci. *8*, 4531–4550.
15. Hunter, D.W., Hibbard, P.B., Levy, M., Marre, O., Sári, K., and Kisvárday, Z. (2016). Ideal binocular disparity detectors learned using independent subspace analysis on binocular natural image pairs. PLoS One *11*, e0150117.
16. Prince, S.J., Cumming, B.G., and Parker, A.J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. J. Neurophysiol. *87*, 209–221.

Evolution: Hearing and Feeding in Fossil Whales

Jonathan H. Geisler[1,2]
[1]Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Northern Boulevard, Old Westbury, NY, USA
[2]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
Correspondence: jgeisler@nyit.edu
http://dx.doi.org/10.1016/j.cub.2017.05.007

The evolution of whales marks one of the major transitions in the history of mammals. Two new studies provide key insights into the evolution of hearing specializations and feeding strategies in early whales.

The past 30 years have witnessed a revolution in our understanding of the evolution of cetaceans (whales, dolphins and porpoises). Molecular phylogenies have demonstrated that cetaceans are most closely related to hippos, and together with other even-hoofed mammals, form the clade Artiodactyla [1]. Cetaceans and close relatives first