

Motion and binocular disparity processing: Two sides of two different coins

11

Christian Quaia*, Edmond J. FitzGibbon, Lance M. Optican, Bruce G. Cumming

Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, Bethesda, MD, United States

**Corresponding author: Tel.: +1-301-496-9375, e-mail address: quaia@nei.nih.gov*

Abstract

From a mathematical point of view, extracting motion and disparity signals from a binocular visual stream requires very similar operations, applied over time for motion and across eyes for disparity. This similarity is reflected in the theories that have been proposed to describe the neural mechanisms used by the brain to extract these signals. At the behavioral level there are, however, several differences in how humans react to these stimuli, which presumably reflect differences in how these signals are processed by the brain. Here we highlight three such differences: the degree to which different axes of motion/disparity are treated isotropically, the importance of reference signals, and the rules that underlie the combination of 1D signals to extract 2D signals.

Keywords

Motion, Disparity, Pattern motion, Pattern disparity, Relative disparity

1 Introduction

In species endowed with high-acuity binocular vision, motion and binocular disparity signals play a critical role in behavior planning and execution. Since natural environments are usually quite cluttered, these signals also contribute to object segmentation, a key step toward object identification. The importance of these signals has been long recognized, and their neural substrate has accordingly been the object of intense investigation (Burr and Thompson, 2011; Cumming and DeAngelis, 2001; Nakayama, 1985; Parker, 2007).

While studies of motion and binocular disparity have largely proceeded independently, analogies have been highlighted. At a computational level, extraction of motion requires the detection of displacement over time (e.g., by finding the peak of a temporal cross-correlation), while extraction of binocular disparity involves

identifying an image displacement across eyes (e.g., by measuring interocular correlation). These similarities are reflected in the structure of two models commonly used to describe the neuronal algorithm that extracts motion and binocular disparity: the motion energy model (Adelson and Bergen, 1985) and the binocular energy model (Ohzawa et al., 1990). Furthermore, at least in primates, these two processes share a common neural substrate. In primary visual cortex (area V1) a minority of cells (about 20%) are selective for the direction of motion, but most of these are also tuned to binocular disparity (Prince et al., 2002; Read and Cumming, 2005). In the middle temporal area (MT), the next stage in motion processing, most neurons are tuned to both direction of motion and binocular disparity (DeAngelis and Newsome, 1999).

Based on these similarities, it is natural to see motion and binocular disparity processing as two sides of the same coin. In this chapter, we will, however, focus on several major differences that exist between how these two signals are processed in the brain.

2 Horizontal/vertical anisotropy

The first and most obvious difference between motion and disparity processing is that, whereas all directions of motion are equally processed (with just a minor preference for the cardinal directions), this does not hold true for binocular disparity. Because, to a first approximation, the eyes rotate around a fixed point, disparities fall along a line on each retina (the so-called epipolar line). With horizontally displaced eyes, in primary position these lines are horizontal, and so are all disparities (Read et al., 2009). Rotations of the eyes cause rotations of the epipolar lines, such that disparities gain a (usually relatively small) vertical component. Accordingly, stereopsis relies almost exclusively on horizontal disparities, and the primary visual cortex encodes a wider range of horizontal than vertical disparities (Cumming, 2002; Read and Cumming, 2004).

This does not, however, mean that vertical disparities are not important, as they play a vital role in maintaining the vertical alignment of the two eyes through vergence eye movements. Because binocular fusion tolerates larger horizontal than vertical misalignments, vertical vergence must actually be more tightly controlled than horizontal vergence (although which is cause and which is effect has not been firmly established). For vergence control, vertical disparities are thus as important as horizontal disparities.

This asymmetry between horizontal and vertical disparities is also reflected in the processing of disparities in 1D stimuli, such as large sinusoidal gratings or random line stimuli (RLS). Since these could in principle be produced by a translation in any direction the brain is forced to make an inference. With moving stimuli, both perception and eye movements respond to the direction of motion orthogonal to the orientation of the stimulus. The brain thus selects the direction associated with the smallest displacement (or speed). In Bayesian parlance, the brain uses a low-speed prior

(Stocker and Simoncelli, 2006). With disparity, things are not as straightforward. Not surprisingly, for both 1D and 2D stimuli perception responds only to the horizontal component of the disparity vector, and thus features across the two monocular images are matched along, or near, the horizontal axis (van Ee and Schor, 2000). For vergence the picture is more complex. With 2D stimuli (e.g., random dot stereograms), an oblique disparity vector is exactly matched by an oblique vergence movement, indicating that the vergence system has access to, and is perfectly capable of using, 2D disparity vectors. With 1D stimuli, even vergence movements are, however, driven more strongly by the horizontal component of the disparity vector, and alignment of the eyes on the images is mediated preferentially, but not entirely, by horizontal vergence (Rambold and Miles, 2008). The prior for vergence control thus appears to be biased toward a combination of small disparity vectors and horizontal disparities.

3 Absolute vs relative signal detection

Another important difference between motion and disparity processing relates to the extent of signal elaboration that precedes perceptual detection. With motion, both perception and tracking eye movements are exquisitely sensitive to the physical motion of a single object relative to the background: Even a small displacement (or low speed) of an object is highly salient and immediately perceived. The local motion signal extracted in primary visual cortex (and in lower animals in the retina) thus seems to be sufficient to mediate perception. Vergence eye movements are similarly controlled by the physical disparity of a stimulus (the so-called absolute disparity). However, stereopsis is most sensitive to the difference in disparity between two nearby objects (relative disparity), a signal that is not present in area V1, and thresholds for detecting absolute disparities are at least an order of magnitude larger than those for relative disparities (Norcia et al., 2017; Westheimer, 1979). It thus appears that perception does not use the disparity signal extracted in primary visual cortex. Instead, it seems to rely on a relative disparity signal, a signal which is first computed (Clery et al., 2017; Thomas et al., 2002) in the secondary visual cortex (area V2).

4 Pattern computation

Because neurons in early visual cortex are selective for the orientation and spatial frequency of sinusoidal gratings (Hubel and Wiesel, 1968; Maffei and Fiorentini, 1973), these have been widely used to study vision. However, most natural visual stimuli have energy over multiple orientations at a single location (e.g., a corner). To internally represent such patterns the 1D signals extracted in area V1 must somehow be combined. Understanding the rules that govern the combining of 1D signals into pattern signals using arbitrary 2D stimuli, in which the stimulus energy is spread

evenly across all orientations, is challenging. Stimuli obtained by summing two 1D gratings having different orientations (plaids) provide a simpler approach. Plaids come in different flavors. If the two components have similar speed (disparity), the pattern speed (disparity) vector falls between those of the two components (type I plaid). If the two components have very different speed (disparity), it is possible to construct plaids in which the pattern speed (disparity) vector falls outside those of the two components (type II plaid). When one of the two components has zero speed (disparity), a so-called unikinetic (unidisparsity) plaid, a special case of a type II plaid, is obtained; with these stimuli the pattern speed (disparity) vector is parallel to the static (zero disparity) component, while its sign and magnitude are determined by the other component. Thus, if, for example, a vertical grating moving horizontally is paired with an oblique static grating, the speed vector of the resulting 2D pattern will have a vertical component, although neither of the two components does.

Most studies have employed plaids composed of two sinusoidal gratings. When the 1D components have very dissimilar spatial frequency (SF) or contrast, two 1D gratings superimposed on each other (transparency) are usually perceived. When the components differ only in orientation and speed, a 2D pattern is, however, usually perceived, with 1D motion and disparity signals being combined in a manner compatible with the so-called Intersection-Of-Constraints (IOC) rule (Adelson and Movshon, 1982).

To get a more detailed view of this neural computation, we have exploited two reflexive eye movements: ocular following and disparity vergence. These are ultra-short latency responses (70 ms in humans, 45 ms in monkeys) that in primates are mediated by a cortical pathway from V1, through MT, and to MST (Takemura et al., 2007). Projections from MT/MST to the pontine nuclei, and from there to the cerebellum, cause the eyes to move, either conjugately (ocular following) or in opposite directions (disparity vergence). Studying these eye movements has three important advantages over studying perception. First, they can be used to directly quantify the strength of a response to various supra-threshold stimuli, as opposed to finding detection thresholds. Second, their temporal evolution can be used to monitor the unfolding of neural computations, as opposed to just evaluating their outcome. Finally, they occur at a subconscious level, and are thus less susceptible to cognitive biases and strategies.

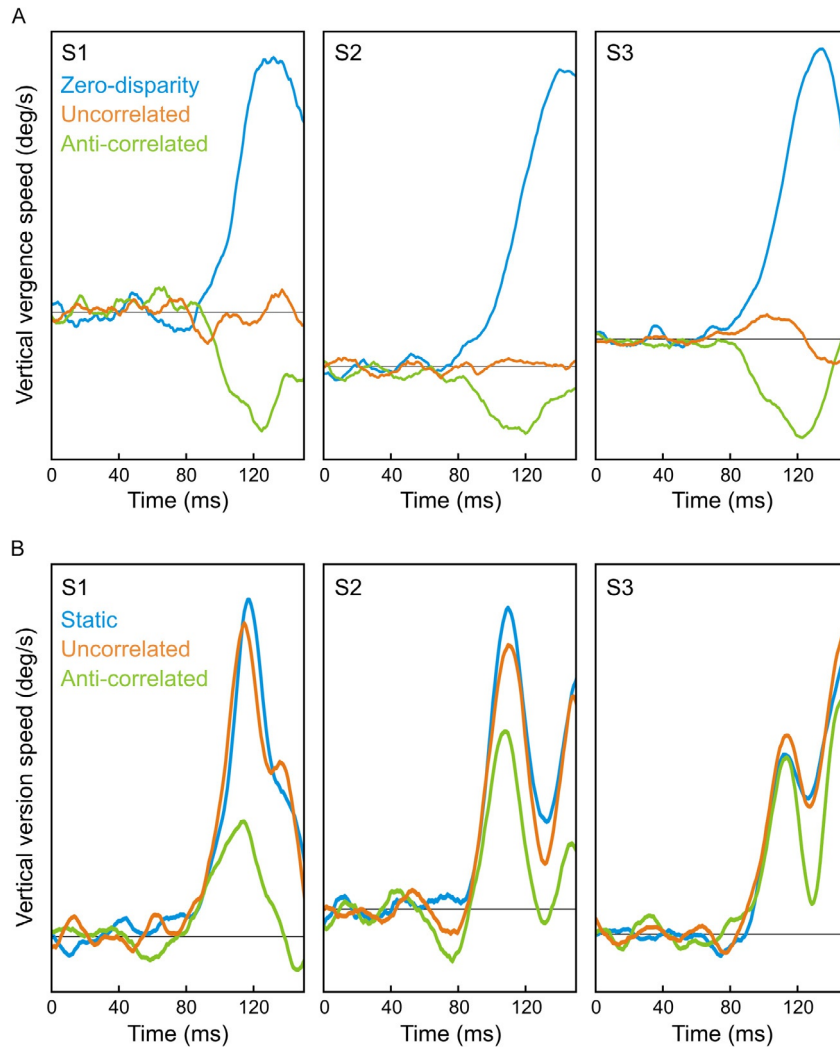
At first sight, unikinetic and unidisparsity sinusoidal plaids induce analogous reflexive eye movements. The ocular following response (OFR) to unikinetic sinusoidal plaids starts off in the direction orthogonal to the drifting component, but 15–20 ms later it turns parallel to the orientation of the static grating and starts tracking pattern motion (Masson and Castet, 2002; Quaia et al., 2016). Similarly, the disparity vergence response (DVR) to unidisparsity sinusoidal plaids starts off in the direction orthogonal to the component with disparity, but 15–20 ms later it turns parallel to the orientation of the zero-disparity grating, tracking pattern disparity (Quaia et al., 2013, 2017). However, beneath the surface important differences emerge. First, strong pattern DVRs are observed only when the two components have similar SF (Quaia et al., 2013). In contrast, pattern OFRs are maximal when the SF of the

static grating is around 0.7–0.8 cpd, regardless of the SF of the moving grating (Quaia et al., 2016). It thus appears that, while the DVRs are shaped by a computation that combines 1D disparity signals within the same SF channel, OFRs are obtained by combining motion signals across the SF spectrum with static signals in a narrower range (generally having higher SFs, given the preference of the OFR for low SFs). By systematically varying the delay between the onset of the two components, we also found that the 15–20 ms delay with which an OFR in the pattern direction emerges is not due to the time it takes to compute this direction, as it was originally proposed. This delay seems instead to be attributable simply to a longer latency of the static signal in reaching the area where pattern motion is extracted (presumably MT): If the static component is presented before the moving component, or if it has higher contrast, the movement starts off from its onset in the pattern direction (Quaia et al., 2016). With DVRs this is not the case: The pattern disparity response always trails the presentation of the plaid by a constant delay, even when one of the two components is presented in advance of the other (our unpublished observations).

Determining the exact neural computation that underlies pattern motion/disparity extraction is far from trivial. Unfortunately, with sinusoidal plaids several alternative neural computations (such as tracking 2D features, zero-crossings, contrast modulations, detecting the orientation of 2D features, and even vector average for type I plaids) predict similar outcomes, limiting their discriminating power. Plaids composed of random line stimuli (RLS), which we have used extensively, do not have clear 2D features or contrast variations that can be tracked, and thus eliminate many of these complications. Recently, we have introduced a new type of RLS plaid, which we called a flicker-plaid (Quaia et al., 2016), that has allowed us to uncover additional differences about how pattern speed and disparity are extracted.

A flicker-plaid is similar to a unikinetic plaid, but the static (zero-disparity) component is replaced with a different RLS in each frame (eye). Importantly, this stimulus does not define a unique pattern motion direction, as the 2D pattern is by construction different on each frame. Nonetheless, with OFRs, the pattern response to a flicker-plaid is just as strong as it is with a unikinetic plaid (Quaia et al., 2016), implying that what is being computed is not strictly a pattern motion direction (undefined for flicker plaids). In fact, just showing one frame of the flickering component (a condition in which a 2D pattern is only present in one frame) induces a response that is almost as strong as that observed with a unikinetic plaid (Quaia et al., 2016). Based on these data we proposed that the motion signal from the drifting component is combined with a form (i.e., orientation) signal from the non-moving one. Again, this is not a proper pattern motion computation, but might get the job done in a natural environment, where flickering or briefly flashed components are not the norm.

The results are, however, very different with DVRs. When a unidisparsity plaid composed of a vertical RLS with disparity and an oblique RLS with zero-disparity is presented, a strong vertical vergence component emerges ~ 80 ms after stimulus presentation (Fig. 1A, blue). When the zero-disparity RLS is replaced with an uncorrelated RLS (the binocular equivalent of a flicker-plaid), the latter is completely ignored, and the DVR is purely horizontal, as it is driven simply by the disparity

**FIG. 1**

(A) Vertical component of the disparity vergence response induced in three subjects by plaids composed of a vertical RLS with disparity and an oblique RLS with zero disparity (blue), uncorrelated across eyes (orange), or anti-correlated across eyes (green). (B) Vertical component of the 2-frame ocular following response induced in three subjects by plaids composed of a vertical RLS moving to the right and an oblique RLS that was static (blue), uncorrelated in the two frames (orange), or anti-correlated across frames (green).

of the vertical RLS (Fig. 1A, orange). We took this a step further by replacing the uncorrelated oblique RLS with one that is anti-correlated, i.e., it is contrast-reversed across the eyes (but has otherwise zero-disparity). With this stimulus, which also lacks a well-defined pattern disparity, we observed a reversal of the vertical

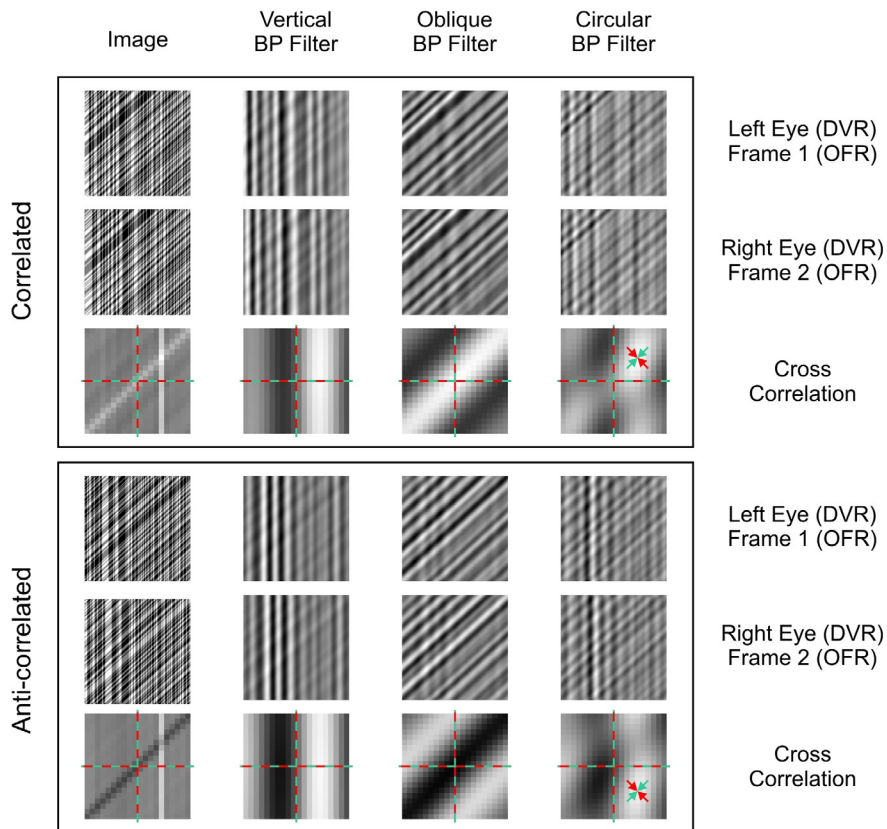


FIG. 2

Prediction of cross-correlation models applied to RLS plaids. In all cases, a vertical RLS moves to the right, or has uncrossed disparity. *Top*: The oblique RLS is static, or has zero-disparity, and the pattern motion/disparity vector points right and up. The peak of the cross-correlation of the images (third row, left column) matches the pattern vector, and so does the cross-correlation applied after passing the images through a circular band-pass filter (right column). If the images are passed through vertical and oblique band-pass filters (second and third column), their cross-correlations (bottom row) also leads to the same solution. *Bottom*: Now the oblique RLS is anti-correlated across images, and again the pattern vector is undefined. Cross-correlation of the images fails to extract a pattern signal, but band-pass filtering the images predicts a pattern vector pointing right and down. Note that if the oblique RLS is uncorrelated, all three methods (cross-correlation of the image, product of oriented band-pass filters cross-correlations, and cross correlation on circular band-pass filtered images) would extract the same signal (motion/disparity of the vertical RLS).

vergence component observed under the zero-disparity condition (Fig. 1A, green). The behavior observed with these three stimuli is compatible with a mechanism that performs either a 2-D cross-correlation between band-pass filtered images, or the product of 1-D cross-correlations between band-passed filtered 1D signals (Fig. 2).

Given what we know about early visual processing, a neural implementation of the latter is more realistic. In fact, it is difficult to imagine any other computation that would produce such a result, making this a very strong test.

This last stimulus is harder to use in a motion setting, as the oblique component would be “washed out” by temporal averaging. We thus used two-frame sequences, taking the two monocular images from a DVR experiment, and showing them to both eyes sequentially (Fig. 1B). In all cases, a vertical component emerges approximately 80ms from stimulus presentation, always with a positive sign, regardless of whether the oblique grating was static (blue), uncorrelated (orange) or anti-correlated (green). The magnitude of the response was reduced with the uncorrelated and anti-correlated RLS, but this can be simply a consequence of a reduced effective contrast of the oblique RLS due to temporal averaging in the visual system. Again, this behavior is compatible with a motion-from-form mechanism, in which only the orientation of the non-moving grating is used, but not with a proper IOC or cross-correlation operation.

From all these experiments, we conclude that, at least as far as eye movements are concerned, neuronal estimates of pattern motion and disparity are based on quite different computations. Motion processing appears to rely on a bag-of-tricks approach (Ramachandran, 1985), whereas disparity processing seems to perform a proper cross-correlation between band-passed signals (indistinguishable from an IOC computation for stimuli for which a pattern disparity exists).

5 Discussion

Early in processing, motion and disparity appear to be extracted using similar mechanisms. These early signals do not directly inform perception nor drive behavior. Further processing steps intervene, and, as we described above, motion and disparity computations here follow different paths.

The reasons underlying this split are most likely ecological: Motion and disparity signals have different patterns of occurrence in the environment. Horizontal/vertical asymmetries in disparity processing can be easily explained based on the larger range of horizontal disparities present in the environment, and their straightforward correlation with distance-in-depth. Their neural over-representation, and the consequent dominant role they play over vertical disparities, is thus not surprising.

In a similar vein, the absolute speed of an object in the world is of primary importance, both for estimating time to impact, and for conveying information about the nature of the object itself. Small differences in absolute speed matter: To a hedgehog, a snail is food, but an acorn isn't, and vice-versa for a squirrel. In contrast, absolute disparity does not tell us anything about the nature of an object; it only informs us about its relationship to our current state of vergence. Relative disparities are more important, as they encode the relative distance of objects in the world: Knowing which of two preys (or predators) is closer is obviously of great importance.

That 1D disparity signals appear to be more faithfully combined than 1D motion signals is more surprising, but only at first sight. A critical issue in both cases is that of avoiding falling for false matches, i.e., solutions driven by spurious correlations. In motion, because the stimuli are extended in time, a relatively simple algorithm may suffice to avoid false matches and correctly identify the motion of objects. Because disparity is computed with just two images, a more complete, and possibly more computationally costly, solution may be required to converge the eyes on the object of interest.

References

- Adelson, E.H., Bergen, J.R., 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284–299.
- Adelson, E.H., Movshon, J.A., 1982. Phenomenal coherence of moving visual patterns. *Nature* 300, 523–525.
- Burr, D., Thompson, P., 2011. Motion psychophysics: 1985–2010. *Vision Res.* 51, 1431–1456.
- Clery, S., Cumming, B.G., Nienborg, H., 2017. Decision-related activity in macaque V2 for fine disparity discrimination is not compatible with optimal linear readout. *J. Neurosci.* 37, 715–725.
- Cumming, B.G., 2002. An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature* 418, 633–636.
- Cumming, B.G., DeAngelis, G.C., 2001. The physiology of stereopsis. *Annu. Rev. Neurosci.* 24, 203–238.
- DeAngelis, G.C., Newsome, W.T., 1999. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* 19, 1398–1415.
- Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.
- Maffei, L., Fiorentini, A., 1973. The visual cortex as a spatial frequency analyser. *Vision Res.* 13, 1255–1267.
- Masson, G.S., Castet, E., 2002. Parallel motion processing for the initiation of short-latency ocular following in humans. *J. Neurosci.* 22, 5149–5163.
- Nakayama, K., 1985. Biological image motion processing: a review. *Vision Res.* 25, 625–660.
- Norcia, A.M., Gerhard, H.E., Meredith, W.J., 2017. Development of relative disparity sensitivity in human visual cortex. *J. Neurosci.* 37, 5608–5619.
- Ohzawa, I., DeAngelis, G.C., Freeman, R.D., 1990. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249, 1037–1041.
- Parker, A.J., 2007. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* 8, 379–391.
- Prince, S.J., Pointon, A.D., Cumming, B.G., Parker, A.J., 2002. Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J. Neurophysiol.* 87, 191–208.
- Quaia, C., Sheliga, B.M., Optican, L.M., Cumming, B.G., 2013. Temporal evolution of pattern disparity processing in humans. *J. Neurosci.* 33, 3465–3476.
- Quaia, C., Optican, L.M., Cumming, B.G., 2016. A motion-from-form mechanism contributes to extracting pattern motion from plaids. *J. Neurosci.* 36, 3903–3918.

- Quaia, C., Optican, L.M., Cumming, B.G., 2017. Combining 1-D components to extract pattern information: it is about more than component similarity. *J. Vis.* 17, 21.
- Ramachandran, V.S., 1985. The neurobiology of perception. *Perception* 14, 97–103.
- Rambold, H.A., Miles, F.A., 2008. Human vergence eye movements to oblique disparity stimuli: evidence for an anisotropy favoring horizontal disparities. *Vision Res.* 48, 2006–2019.
- Read, J.C.A., Cumming, B.G., 2004. Understanding the cortical specialization for horizontal disparity. *Neural Comput.* 16, 1983–2020.
- Read, J.C., Cumming, B.G., 2005. Effect of interocular delay on disparity-selective V1 neurons: relationship to stereoacuity and the Pulfrich effect. *J. Neurophysiol.* 94, 1541–1553.
- Read, J.C., Phillipson, G.P., Glennerster, A., 2009. Latitude and longitude vertical disparities. *J. Vis.* 9, 11.1–11.37.
- Stocker, A.A., Simoncelli, E.P., 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9, 578–585.
- Takemura, A., Murata, Y., Kawano, K., Miles, F.A., 2007. Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. *J. Neurosci.* 27, 529–541.
- Thomas, O.M., Cumming, B.G., Parker, A.J., 2002. A specialization for relative disparity in V2. *Nat. Neurosci.* 5, 472–478.
- Van Ee, R., Schor, C.M., 2000. Unconstrained stereoscopic matching of lines. *Vision Res.* 40, 151–162.
- Westheimer, G., 1979. Cooperative neural processes involved in stereoscopic acuity. *Exp. Brain Res.* 36, 585–597.

Further reading

- Nienborg, H., Cumming, B.G., 2007. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat. Neurosci.* 10, 1608–1614.