All Pulfrich-like illusions can be explained without joint encoding of motion and disparity

Jenny C. A. Read

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD, USA ✉

Bruce G. Cumming

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD, USA 🏠 ✉

In the Pulfrich effect, an interocular time delay results in the perception of depth. Two modified versions, the stroboscopic Pulfrich effect and dynamic visual noise with a delay, are generally explained by postulating an early stage of space/time-inseparable filtering, encoding motion and disparity jointly. However, most disparity sensors in monkey V1 do not show joint motion/disparity encoding, and we recently showed that depth perception in the stroboscopic Pulfrich effect is equally compatible with space/time-separable filtering. Here, we demonstrate that this filtering can be implemented with a population of physiologically plausible energy model units. Similar results are obtained whether the neurons are pure disparity sensors (like most V1 neurons) or joint motion/disparity sensors (like MT). We also demonstrate that the dynamic noise stimulus produces correlations between the activity in pure disparity sensors, and in a separate population of pure motion sensors. These correlations are sufficient to explain the percept. Thus, joint encoding of motion and disparity is not required to explain depth perception in Pulfrich-like stimuli: a brain which encoded motion and disparity in entirely separate neuronal pathways could still experience all of these illusions.

Keywords: binocular vision, computational modeling, interocular delay, primary visual cortex, Pulfrich effect, psychophysics

Introduction

The literature on stereopsis contains several cases in which an illusion of depth is caused by viewing stimuli with an interocular delay, the classic example being the Pulfrich effect (Pulfrich, 1922). When a pendulum is viewed swinging in the frontoparallel plane, the introduction of an interocular delay generates a sensation of depth, making the pendulum appear to follow an elliptical path in depth as it swings to and fro. While significant clinically—for example, patients whose optic neuritis causes a difference in conduction speeds between the optic nerves may experience a disconcerting Pulfrich effect—this can tell us little about how the brain works. As pointed out by Fertsch (Pulfrich, 1922), in the classic Pulfrich effect, the interocular delay introduces a real spatial disparity on the retina. Suppose the image reaching the right eye is delayed relative to the left eye by an amount Δt, and that the object, moving with speed v, has a position x when it is first seen by the left eye. By the time this same image reaches the right eye, the image in the left eye will have moved to a new position, $x + v\Delta t$. At this moment, the right eye's image is at x whereas the left eye's image is at $x + v\Delta t$, so there is a spatial disparity $v\Delta t$. Because any neuronal mechanism that produces depth from binocular disparity will also produce depth in the classic Pulfrich effect, we learn nothing new about brain mechanisms.

In the past thirty years, however, several modifications of this stimulus have been introduced, precisely to elucidate neuronal mechanisms for processing delay and disparity. The two most prominent are the stroboscopic Pulfrich effect (Burr & Ross, 1979; Lee, 1970a, 1970b; Morgan, 1979; Morgan & Thompson, 1975; Read & Cumming, 2005b) and dynamic visual noise (Falk & Williams, 1980; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Ross, 1974; Tyler, 1974, 1977). In the stroboscopic version of the Pulfrich effect, a target is presented in apparent motion, jumping from point to point across the screen instead of moving continuously, and is viewed with interocular delay. The space/time diagram for this stimulus is shown in Figure 1. The dotted lines represent the trajectory of the moving object; the stars represent its brief appearances. At any instant in time, the stimulus is visible in only one eye, so in this sense the stimulus has no spatial disparity. Yet this stimulus also gives rise to a perception of depth. The second stimulus, dynamic visual noise, resembles the "snowstorm" on an untuned television. When viewed with an interocular delay, the noise appears to swirl in depth, with points in front of the screen moving towards the delayed eye, and points behind it in the opposite direction. In these stimuli, unlike the classic Pulfrich effect, the depth percept is not a trivial consequence of stimulus geometry. For example, in the stroboscopic Pulfrich effect, the target is only ever visible monocularly at a given instant. In order for binocular matches to be made between left and right images, the brain must remember the image seen in the left eye to pair it with the delayed image seen in the right. Thus, this stimulus can tell us about the temporal integration properties of
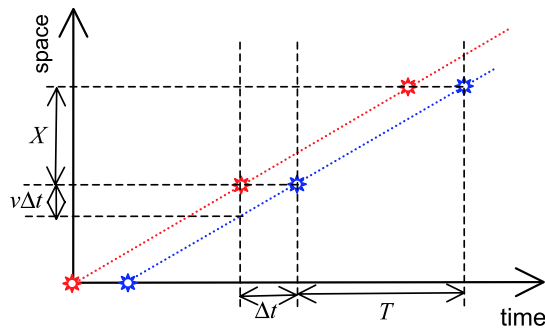
Figure 1. Space/time diagram for the stroboscopic Pulfrich stimulus. The dotted lines show the trajectory of the target (red = left eye; blue = right eye) defined by apparent motion. The "stars" show the appearances of the target, when it is briefly illuminated by the stroboscope. Because the right-eye image is artificially delayed, each appearance of the target occurs a time Δt later in the right eye than in the left eye. T is the interflash interval of the stroboscope; X is the interflash distance. $v = X/T$ is thus the apparent speed of the target. $v\Delta t$ is the virtual disparity between the apparent-motion trajectories.

the neuronal mechanisms subserving depth perception (Morgan, 1979).

Work with these stimuli, and similar experiments using vernier alignment tasks, has shown that what is perceived can best be explained by considering the effects of spatiotemporal filtering in early vision, before any attempt to extract information needed for a particular task (Morgan, 1975, 1976, 1979, 1980, 1992; Morgan & Watt, 1982, 1983; Read & Cumming, 2005b). Along with the development of explanations based on spatiotemporal filters, another idea has gained wide acceptance: that this spatiotemporal filtering is performed by direction-selective filters (*inseparable* functions of space and time); the logic behind this conclusion is laid out with particular clarity by Anzai, Ohzawa, & Freeman (2001). In this view, the receptive fields are tilted relative to the space/

time axes, so the neuron is sensitive to stimulus direction of motion (Anzai et al., 2001; Carney, Paradiso, & Freeman, 1989; Morgan & Fahle, 2000; Pack, Born, & Livingstone, 2003; Qian, 1997). Binocular neurons with such receptive fields would jointly encode both motion and disparity. A signature property of such joint motion/disparity sensors is their distinctive tilted tuning profile when probed with stimuli containing both interocular delay and binocular disparity (Figure 2A). Their preferred disparity changes as a function of interocular delay; they "cannot distinguish an interocular time delay from a binocular disparity" (Qian & Andersen, 1997).

If stereopsis was supported exclusively by joint motion/disparity sensors, then, as first proposed by Ross (1974), interocular delay would produce a depth percept in the same way as binocular disparity does. In stimuli with no interocular delay, a zero-disparity stimulus is presumably perceived as such because it elicits the strongest response in cells tuned to zero disparity. But if the disparity sensors are also sensitive to direction of motion, then when the right eye is delayed, rightwards-preferring cells shift their disparity tuning towards far disparities, and leftwards-preferring cells towards near (Figure 2A; see Figure 2 of Read & Cumming, 2005a, for an explanation of why the direction of the shift depends on the cell's preferred direction of motion.) Thus, the cells which respond best to the zero-disparity stimulus with interocular delay are not those which usually signal zero disparity in a conventional, nondelayed stimulus. Rather, the most responsive cells are those tuned to leftwards motion and far disparities (whose peak response now occurs at zero disparity, because the interocular delay has shifted the entire disparity tuning curve towards near disparities) and to rightwards motion/near disparities (whose tuning curve has been shifted towards far disparities). Because these cells usually respond only to nonzero disparities, the brain naturally interprets their activity as indicating that the stimulus contains depth: Leftward-moving objects are perceived as
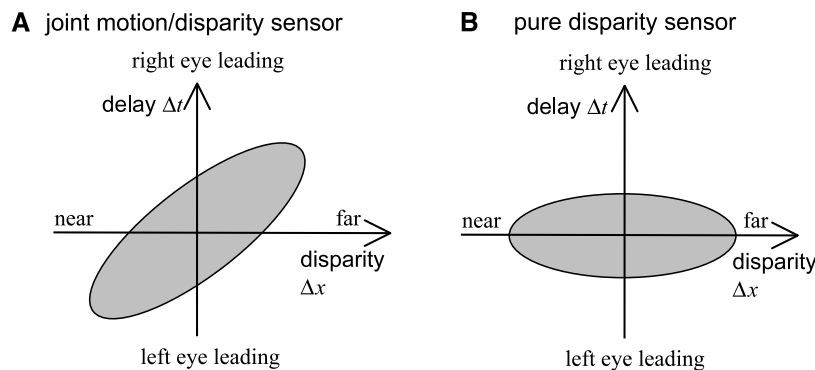

Figure 2. Delay/disparity tuning profile for (A) a joint motion/disparity sensor, and (B) a pure disparity sensor. Interocular delay is plotted on the vertical axis; disparity on the horizontal axis. The shaded region shows the combinations of delay and disparity that elicit a strong response from the cell. For a joint motion/disparity sensor, the preferred disparity depends on delay (e.g., with no interocular delay, the cell prefers zero disparity; when the left eye is leading it prefers near disparity; when the right eye is leading it prefers far disparity). For a pure disparity sensor, interocular delay merely weakens the disparity tuning, without altering the preferred disparity (here zero).

"far", and rightwards ones are perceived as "near". Thus, if one assumes that the brain is unaware of the interocular delay, and so reads out these filters as if there was no delay, the sensation of depth follows naturally.

In recent years, neurons with these properties have been reported in cat area 17/18 (Anzai et al., 2001; Carney et al., 1989) and in monkey V1 and MT (Bradley, Qian, & Andersen, 1995; DeAngelis, Cumming, & Newsome, 1998; DeAngelis & Newsome, 2004; DeAngelis & Uka, 2003; Maunsell & Van Essen, 1983; Pack et al., 2003; Roy, Komatsu, & Wurtz, 1992) and hailed as the neuronal basis for the Pulfrich effect. A mathematical analysis has confirmed that joint motion/disparity sensors can, as expected, give a depth percept in the stroboscopic Pulfrich effect and in dynamic visual noise (Qian & Andersen, 1997). A clear consensus now seems to have emerged in the scientific literature, and even in popular-science books, that joint motion/disparity encoding in early visual cortex is the neuronal basis for Pulfrich-like phenomena (Anzai et al., 2001; Carney et al., 1989; Morgan, 2003; Morgan & Castet, 1995; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Pack et al., 2003; Qian, 1997; Qian & Andersen, 1997).

But the physiological observations also raise important questions about this idea. Although joint motion/disparity sensors are common in cat area 17/18, they are much less common in monkey V1 (Pack et al., 2003; Read & Cumming, 2005a). This simply reflects the much lower incidence of direction selectivity in V1: a minority of disparity-selective cells in V1 do also encode direction of motion, as envisaged in the joint encoding theory, but the majority of disparity-selective cells are not sensitive to direction of motion. Thus, the standard view that only joint motion/disparity sensors contribute to Pulfrich illusions leads to a most surprising conclusion. In this view, the illusory depth is signaled only by the small minority of disparity sensors that are also direction selective, whereas the vast majority of non-direction-selective disparity sensors are signaling zero disparity. And yet the veridical perception of the majority is somehow overridden to create the illusory percept. That is, most disparity-selective cells in V1 do not contribute to depth perception, despite encoding useful information about disparity. This could certainly occur if perception depends upon neurons in extrastriate cortex where joint encoding is more common, for example, MT (Bradley et al., 1995; DeAngelis & Newsome, 2004; DeAngelis & Uka, 2003; Maunsell & Van Essen, 1983; Pack et al., 2003; Roy et al., 1992), but would still imply that information available in the population of V1 neurons has been lost. Before drawing this conclusion, it seems worth re-examining whether joint motion/disparity encoding is in fact the only explanation for Pulfrich-like illusions. We shall argue that, in fact, the illusory depth percept in all Pulfrich-like stimuli can be explained perfectly well in terms of pure disparity sensors. We conclude that, in Pulfrich-like illusions, both the direction-selective and the non-direction-selective disparity

sensors in V1 are signaling the illusory depth. Thus, there is no evidence that either group has a privileged role in perception.

In a recent paper (Read & Cumming, 2005b), we showed that the effective disparity produced by the stroboscopic Pulfrich effect (defined as minus the nulling disparity necessary to cancel the depth illusion) could be quantitatively explained in terms of spatial disparities present in the stimulus, if one allows for temporal integration over about 15 ms. In the model, the effective disparity is simply the weighted average of the disparities between all possible matches, with matches whose left and right members occur further apart in time given less weight. It seemed likely that this disparity-averaging algorithm could be simply implemented by a population of pure disparity sensors, much as Qian & Andersen (1997) did with a population of joint motion/disparity sensors. These pure disparity sensors would still implement an early stage of spatiotemporal filtering, as envisaged by Morgan and colleagues (Morgan, 1975, 1976, 1979, 1980, 1992; Morgan & Watt, 1982, 1983). However, we argued that the filters need not be spatiotemporally inseparable and hence direction-selective, as assumed in recent years (Morgan, 2003; Morgan & Castet, 1995; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Qian & Andersen, 1997); but that separable filtering would give similar results. The depth percept could be supported by pure disparity sensors (binocular neurons with spatiotemporally separable receptive fields), whereas the motion percept could be supported by a separate population of motion sensors with spatiotemporally inseparable receptive fields. Such a model, in which motion and disparity are encoded separately rather than jointly, still incorporates an early stage of spatiotemporal filtering and seems likely to be equally compatible with the psychophysical evidence. In this paper, we test this alternative way of implementing spatiotemporal filtering.

The stimulus which is most often adduced as compelling evidence for joint motion/disparity encoding is a dynamic visual noise viewed with an interocular delay (Morgan, 2003; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Ross, 1974). Yet immediately after Ross (1974) reported this illusion, Tyler (1974, 1977) explained it as due to spatial disparities actually present in the stimulus. Tyler noted that chance pairs of dots that happened to have a far disparity would also tend to have apparent motion to the left, whereas pairs which happened to have a near disparity would move to the right (if the right eye is delayed; vice versa otherwise). He argued that this association is sufficient to explain the swirling percept. In this view, the significance of the temporal delay is that it introduces spatial disparity into the stimulus, according to the geometrical relationship noted by Fertsch (Pulfrich, 1922); it is this disparity, rather than the delay itself, which gives rise to the perception of depth. Perhaps because this explanation was based on matching dots in the stimulus,

rather than spatiotemporal filtering, it has been neglected in recent years. Although it seems likely that the explanation could be updated to use spatiotemporal filtering performed by separate populations of disparity and motion sensors (Neill, 1981), no quantitative model demonstrating this has been produced. Consequently, the success of quantitative models based on spatiotemporally inseparable binocular filters has been taken as evidence in favor of joint motion/disparity encoding, in the absence of any clear demonstration that the alternative explanation fails.

In this paper, we address both issues by developing a quantitative model of depth perception in Pulfrich-like stimuli, using disparity sensors whose receptive fields are space/time-separable filters.

1. For the stroboscopic Pulfrich stimulus, we consider the response of pure disparity sensors (binocular neurons with space/time-separable receptive fields, built according to the energy model of Ohzawa, DeAngelis, & Freeman, 1990). These neurons are not sensitive to direction of motion, and their preferred disparity remains constant as interocular delay changes, as in Figure 2B. We show that our disparity-averaging model (Read & Cumming, 2005b) can be simply implemented by averaging the response of these neurons, weighted by their disparity preference. This produces a value for effective disparity which is in excellent agreement with psychophysics experiments.

2. For the dynamic visual noise stimulus, we examine the correlation between a population of pure disparity sensors and a population of pure motion sensors (monocular neurons with space/time-inseparable receptive fields). We show that the activity of pure disparity sensors is correlated with the activity of pure motion sensors. If the right eye experiences a delay Δt, then near-preferring disparity sensors are correlated with rightward-preferring motion sensors, whereas far disparity sensors are correlated with leftward motion sensors. Motion sensors tuned to speed v are most strongly correlated with disparity sensors tuned to a disparity of $\sim v\Delta t$. We argue that this correlation is sufficient to explain why motion is perceived in opposite directions on either side of the fixation plane, why speed increases with distance from fixation, and why the percept reverses when the noise is anti-correlated (Tyler, 1977).

Methods

Stroboscopic Pulfrich stimulus

Receptive fields

The space/time receptive field function $\rho(x, t)$ represents the response to a stimulus at retinal position x that

occurred at time t relative to the present moment. We adopt the convention that negative values of t represent times before the present moment. In accordance with causality, we set $\rho(x,t) = 0$ for all $t > 0$ because the cell cannot be influenced by the stimuli that have not yet occurred. Because the experimental stimulus contains only horizontal motion, we need include only one spatial dimension. For our model disparity sensors, except where otherwise specified, we use space/time-separable receptive fields, where the function $\rho(x, t)$ can be expressed as the product of a spatial component $\rho_x(x)$ and a temporal component $\rho_t(t)$. Neurons with space/time-separable receptive fields are not sensitive to direction of motion.

We consider a population of binocular disparity sensors, whose receptive fields are identical in all respects except their positions on the retina. Differences in the position of the receptive fields in left and right eyes result in a range of disparity tuning within the population (position disparity). We take the receptive field centered on the origin, $\rho_0(x,t)$, as a template; a receptive field at position x_{L0}, for example, can be written as $\rho_0(x-x_{L0}, t)$. Except where otherwise noted, we model the spatial component of the receptive field profile as a Gabor function:

$$\rho_{0x}(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)\cos(2\pi f x), \tag{1}$$

where $\sigma = 0.1°$ and $f = 2$ cycles per degree, corresponding to a full-width half-maximum power bandwidth of about 2.3 octaves. In fact, this choice is irrelevant because we prove in the Appendix that, with the read-out rule of Equation 15, the same effective disparity is obtained whatever function is chosen for the spatial component. Except where otherwise stated, the temporal component is modeled as a Gaussian:

$$\rho_{0t}(t) = \exp\left(-\frac{(t + t_{\text{lag}})^2}{2\tau^2}\right) \text{ for } t < 0;$$
$$= 0 \text{ otherwise,} \tag{2}$$

where the standard deviation τ is 10 ms, and the time between stimulus onset and peak response, t_{lag}, is 50 ms. This means that the cell's response gradually rises after the appearance of a stimulus, reaching a peak 50 ms after stimulus onset, and decaying thereafter. This receptive field is shown in Figure 3A. A more realistic temporal kernel would be biphasic, reflecting the band-pass temporal tuning of most real V1 cells. However, this would generate problems with the binocular temporal integration. The energy model predicts that the response to interocular delay should be governed by the cross-correlation of the temporal kernels. Band-pass temporal kernels would therefore generate a biphasic response to interocular delay, yet this is not observed in the responses of V1 neurons (Anzai
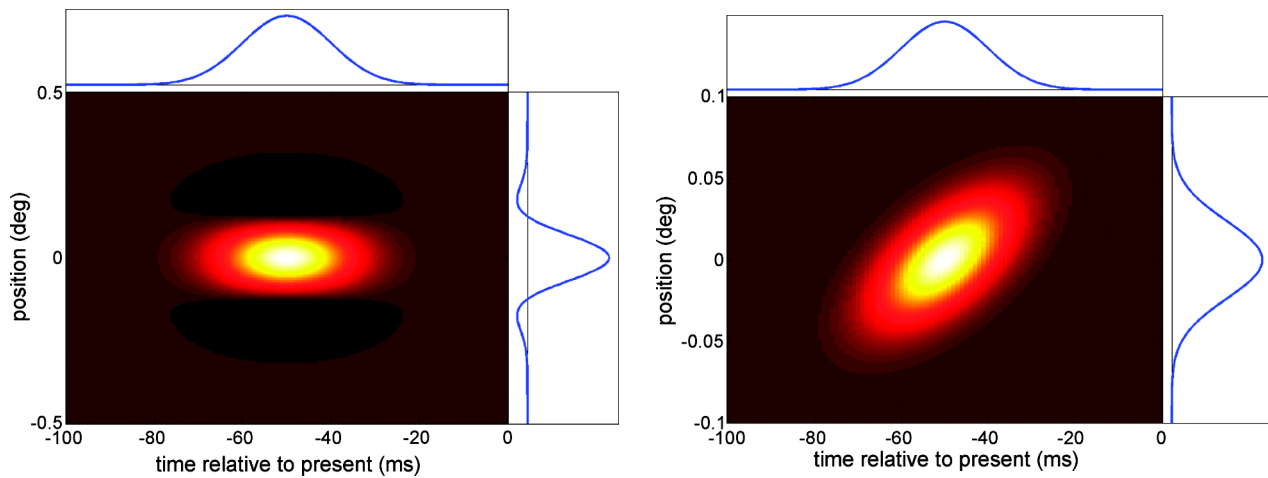
Figure 3. Receptive field profiles used for the strobe Pulfrich simulation. The color plot shows the receptive field function; the marginal plots show its projections on the space and time axes. Only negative times are shown because in our convention positive times represent stimuli which will occur in the future, and so the receptive field function is zero for all $t > 0$. (A) Standard space/time-separable receptive field, spatial component given by Equation 1 and temporal component by Equation 2 results shown in Figure 8. (B) Tilted (space/time inseparable) receptive field given by Equation 4, results shown in Figure 11. Note different vertical axes.

et al., 2001; Read & Cumming, 2005a). This is a known problem of the binocular energy model, which has yet to be addressed. It causes particular difficulties for our model of the strobe Pulfrich effect. Here, the cross-correlation of the temporal kernels also acts as a weight function controlling the weight given to different interocular delays when disparities are averaged (see Appendix). If this weight function is biphasic, then matches at some interocular delays have the effect of repelling the effective disparity away from the disparity of the match, a phenomenon with no psychophysical support. For all these reasons, we restricted ourselves to monophasic temporal kernels when modeling the strobe Pulfrich effect. In the dynamic noise simulation, we do also consider Gabor receptive fields with band-pass temporal frequency tuning.

In Figure 10, we show results when the temporal receptive fields are exponential impulse functions:

$$\rho_{0t}(t) = \exp\left(-\frac{(t + t_{\text{lag}})}{\tau}\right) \text{ for } -t < t_{\text{lag}};$$

$$= 0 \text{ otherwise.} \tag{3}$$

In Figure 11, we show results when the receptive field is an inseparable function of space and time tuned to the apparent velocity of the strobe stimulus, Figure 3B, and compare these to the results obtained with the space/time-separable receptive field of Figure 3A. For this comparison to be valid, it is essential that the temporal extent of the two receptive fields should be the same. The marginal plots along the top of Figure 3A and B show that the projections of the two receptive fields onto the time axis are the same. However, note that the spatial extent of the tilted receptive field is smaller (note different vertical axis scales in Figure 3A

and B). This is because, given the constraint that the temporal projections should be the same, the spatial extent has to be narrow to obtain meaningful velocity tuning. To see this, consider how to increase the spatial extent of the RF in Figure 3B. If the ellipsoid were expanded along all axes equally, the temporal extent would increase along with the spatial extent. If the RF were stretched only along the vertical axis, this would make the ellipsoid more circular and thus weaken the velocity tuning. It might be argued that, for a fair comparison, we should make the spatial extent of the space/time-separable receptive field similarly narrow. In fact, this is not necessary because, as noted above, for space/time-separable receptive fields the results are independent of the spatial component. The receptive field function in Figure 3B is

$$\rho_0(x,t) = \exp\left(-\frac{(x \sin\theta + (t + t_{\text{lag}})\cos\theta)^2}{2\sigma_1^2}\right.$$
$$\left. -\frac{(x\cos\theta - (t + t_{\text{lag}})\sin\theta)^2}{2\sigma_2^2}\right), \tag{4}$$

where $\tan\theta = 3.6$ deg/s, the apparent velocity of the strobe stimulus. $\sigma_1 = 0.025$ and $\sigma_2 = 0.008$, where t is in seconds and x in degrees.

Energy model

Our disparity sensors are constructed according to the stereo energy model (Ohzawa et al., 1990) with position disparity (e.g., Anzai, Ohzawa, & Freeman, 1997). The energy model was chosen because it is mathematically tractable and enabled us formally to prove the results in the Appendix; however, this choice is not critical. Qualitatively similar results are obtained with, for example, the modified version of the energy model

proposed in Read et al. (2002). All these models begin with the response of each receptive field at time t:

$$v(t) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{\infty} dt' \rho(x,t') I(x, t+t'). \qquad (5)$$

The function $I(x,t)$ represents the image. $I(x,t)$ is the luminance at retinal position x and time t, relative to the mean luminance. Thus, values of 0 represent gray, positive values represent bright features and negative values represent dark features. The function $\rho(x,t)$ represents the space/time receptive field, as described in the preceding section.

Each model neuron has two receptive fields, one in each eye. In the energy model, the neuron's response at time t is the square of the sum of the inputs from the two eyes:

$$C(t) = [v_L(t) + v_R(t)]^2. \qquad (6)$$

The energy model response can be divided into the sum of the monocular terms, $M = v_L^2 + v_R^2$, which is insensitive to the binocular correlation of the stimulus, and a binocular component

$$B = 2v_L v_R, \qquad (7)$$

which makes the energy model sensitive to disparity, even in random-dot stereograms. We assume that the effective disparity of the stimulus depends only on the binocular component of the population response.

Population response to the stroboscopic Pulfrich stimulus

We simulate the response of each neuron, as a function of time, to a stroboscopic Pulfrich stimulus (Figure 1), in which the left and right images are taken to be

$$I_L(x,t) = \sum_{j=-\infty}^{\infty} \delta(x - jX, t - jT);$$
$$I_R(x,t) = \sum_{j=-\infty}^{\infty} \delta(x - jX, t - \Delta t - jT). \qquad (8)$$

This equation assumes that the moving target is so small and so briefly illuminated that the stimulus may be described as a series of Dirac delta functions, δ. Without loss of generality, we have also assumed that one of the flashes occurs at $t = 0$, and that the image in the left eye is then at position $x = 0$. T is the interflash interval of the stroboscope; X is the distance the target moves in this period; Δt is the interocular delay. Positive values of Δt means that the right eye's image is delayed relative to the left eye's; negative Δt means that left is delayed relative to right.

Substituting these images (Equation 8) into Equation 5, we find that the response from the left receptive field is, at time t:

$$v_L(t) = \sum_{j=-\infty}^{\infty} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dt' \rho_L(x,t') \delta(x - jX, t + t' - jT).$$

After integrating over x and t, this becomes

$$v_L(t) = \sum_{j=-\infty}^{\infty} \rho_L(jX, jT - t). \qquad (9)$$

For simplicity, we have written the summation as over all values of j, although terms with $j > t/T$, representing appearances of the target which have not yet occurred, make no contribution (recall that the receptive field function is zero for positive values of its time argument).

Our model includes a population of neurons distinguished only by the position of their receptive fields on the left and right retinae, x_{L0} and x_{R0}. The difference between these two defines the preferred disparity $\Delta x_{pref} = x_{L0} - x_{R0}$, controlling the distance from the observer of stimuli which optimally drive the cell. Their mean value gives the neuron's preferred cyclopean position $x_{pref} = (x_{L0} + x_{R0})/2$, controlling the visual direction of optimal stimuli. Thus, we can write each neuron's left- and right-eye receptive fields, ρ_L, ρ_R, as a shifted version of the reference receptive field ρ_0, which is centered on the origin. We write $\rho_L(x,t) = \rho_0(x - x_{L0}, t), \rho_R(x,t) = \rho_0(x - x_{R0}, t)$. In terms of the neuron's preferred disparity Δx_{pref} and cyclopean position x_{pref}, we have $\rho_L(x,t) = \rho_0(x - x_{pref} - \Delta x_{pref}/2, t), \rho_R(x,t) = \rho_0(x - x_{pref} + \Delta x_{pref}/2, t)$. Substituting into Equation 9, we find that, for the neuron tuned to disparity Δx_{pref} and cyclopean position x_c, the response from the left eye at time t is

$$v_L(t, x_{pref}, \Delta x_{pref}) =$$
$$\sum_{j=-\infty}^{\infty} \rho_0\left(jX - x_{pref} - \frac{\Delta x_{pref}}{2}, jT - t\right). \qquad (10)$$

For the right eye, the expression is

$$v_R(t, x_{pref}, \Delta x_{pref}) =$$
$$\sum_{j=-\infty}^{\infty} \rho_0\left(jX - x_{pref} + \frac{\Delta x_{pref}}{2}, jT + \Delta t - t\right). \qquad (11)$$

We use these expressions in Equation 6 to calculate the response, as a function of time, of a population of binocular neurons tuned to different cyclopean position x_{pref} and disparities $\Delta x_{pref}, C(t, x_{pref}, \Delta x_{pref})$. We now need a read-out rule relating the activity of this population

to perceptual judgments performed in psychophysics experiments.

Extracting a single disparity judgment from population activity

In the psychophysical experiments whose results we are seeking to model, the subject was asked to find a disparity that nulled the disparity introduced by the Pulfrich effect (the effective disparity of the Pulfrich effect). A single effective disparity was found for the entire stimulus duration. It is thus natural to assume that, in making this judgment, subjects averaged across time and position. We therefore sum across cyclopean position and time to obtain total activity in the population as a function of disparity only:

$$A(\Delta x_{\mathrm{pref}}) = \int_0^T dt \int_{-\infty}^{+\infty} dx_{\mathrm{pref}} C(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}}). \qquad (12)$$

$A(\Delta x_{\mathrm{pref}})$ is the (unnormalized) time-averaged activity of the pool of neurons with preferred disparity Δx_{pref}. Note that, in computing the time average, we need only integrate over one strobe interflash interval T because the integral over cyclopean position is periodic with period T. As the target moves, the activity moves across the population: If at time t the most active cells are those tuned to some particular cyclopean position x_{pref}, then at time $t + T$ the most active cells will be those tuned to $x_{\mathrm{pref}} + X$, but the distribution of activity across sensors tuned to different disparities will be the same.

As noted above, the energy model response can be divided into monocular and binocular components M and B (Equation 7). In this stimulus, when averaged over a population of cells tuned to different cyclopean positions but the same disparity, the sum of the monocular terms is independent of the cells' preferred disparity: It simply indicates the presence of a stimulus somewhere in the visual field. Equation 12 can thus be rewritten as

$$A(\Delta x_{\mathrm{pref}}) = \mathcal{M} + \int_0^T dt \int_{-\infty}^{+\infty} dx_{\mathrm{pref}} B(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}}), \qquad (13)$$

where $\mathcal{M}$ is the "baseline" contribution from the monocular components M, which is independent of the preferred disparity, and the integral represents the contribution from the binocular component B, which does depend on the preferred disparity. It is the binocular component that endows the energy model with its key property of disparity tuning even for stimuli which contain no monocular cues to disparity, such as random-dot patterns; the monocular terms contribute only a baseline response that is observed even with binocularly uncorrelated patterns. In the simpler stimuli considered here (bars), the distinct image features that carry the

disparity are visible monocularly. However, the monocular stimulus location gives no reliable information about the disparity of the target, so the binocular component of the response is the only part that is useful for estimating disparity. We therefore examine the disparity-dependent term in Equation 13:

$$D(\Delta x_{\mathrm{pref}}) = \int_0^T dt \int_{-\infty}^{+\infty} dx_{\mathrm{pref}} B(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}}). \qquad (14)$$

$D(\Delta x_{\mathrm{pref}})$ is the amount by which the total response of the pool of neurons tuned to disparity Δx_{pref}, averaged over time, exceeds the baseline response of all pools. Obviously this will be larger for pools whose preferred disparity, Δx_{pref}, corresponds to a disparity present in the stimulus. We now wish to choose a neuronal read-out rule that implements disparity averaging, because this is what appears to happen psychophysically. We shall use the set of responses $D(\Delta x_{\mathrm{pref}})$ as if it were a probability distribution. For example, if there were two pools whose responses were above baseline, disparity averaging means that the effective disparity lies between the preferred disparities of the two pools. This can be achieved by postulating that the effective disparity is the mean of the disparity distribution implied by $D(\Delta x_{\mathrm{pref}})$:

$$\Delta x_{\mathrm{eff}} = \frac{\int_{-\infty}^{+\infty} d\Delta x_{\mathrm{pref}} \Delta x_{\mathrm{pref}} D(\Delta x_{\mathrm{pref}})}{\int_{-\infty}^{+\infty} d\Delta x_{\mathrm{pref}} D(\Delta x_{\mathrm{pref}})}. \qquad (15)$$

In the Appendix, we show that this read-out rule gives the same results as the weighted disparity averaging considered in Read & Cumming (2005b).

Implementation details

The time-averaged disparity-dependent activity D (Equation 14) was evaluated at 151 different values of preferred disparity Δx_{pref} between $\pm(4\sigma + X(4\sigma_t + |\Delta t|)/T)$. The resulting distribution D was used to calculate effective disparity as in Equation 15. The limits, notionally infinite, were chosen to make sure of including all neuronal pools whose time-averaged activity is above baseline. The integration limits on x_{pref} were set to $\pm(4X + \sigma)$, centered on the most active neuronal population, again to make sure of including all the members of the neuronal population which would be activated above baseline during one stimulus temporal period. All integrals were performed by the rectangle rule, using 61 steps in the integral over cyclopean position and 151 steps in the integral over time. The sums in Equations 10 and 11 were evaluated by initially performing the sum from $j = -15$ to $j = 15$, and then continuing to add pairs of j

on either side of zero until the fractional change was less than 2 parts in a million. To check that these accuracy parameters were fine enough, we redid the simulation using 101 values of disparity, 41 steps in cyclopean position, 101 in time and evaluated the sums in Equations 10 and 11 to an accuracy of 5 parts in a million. The results did not change appreciably.

Dynamic visual noise with an interocular delay

In the previous section, we were interested in explaining the depth percept, so we modeled only pure disparity sensors, assuming that the motion percept was supported by a separate population of motion sensors that we did not model. In the dynamic visual noise stimulus, we are interested in the relationship between depth and motion to understand why this stimulus produces a sensation of opposite directions of motion on opposite sides of the fixation plane. Here, therefore, we need to include both motion and disparity sensors in the simulation. In our simulation, these populations are entirely separate: We model a population of disparity sensors which are not sensitive to stimulus direction of motion, and a population of motion sensors which are not sensitive to stimulus disparity.

Disparity sensors

Our disparity sensors are binocular energy model units with space/time-separable receptive fields, like those in the previous section. Again, the positions of the receptive field centers differ between left and right eyes, giving a range of disparity tuning. However, there are differences between the simulation needed for the dynamic noise stimulus and that needed for the stroboscopic Pulfrich stimulus. First, the dynamic noise stimulus does not contain a moving target, so there is no need to include neurons tuned to a range of cyclopean positions. For simplicity, therefore, we only consider neurons whose preferred cyclopean position is zero. That is, although neurons with different preferred disparities have receptive fields with different positions, the mean of the receptive field centers in left and right eyes is zero for all neurons. Second, because the dynamic noise stimulus contains motion energy in all directions, we use two spatial dimensions in our simulation. We include neurons tuned to four different orientations, assuming that orientation tuning is always the same in both eyes (Bridge & Cumming, 2001).

Spatially, the receptive fields are two-dimensional Gaussians with a long axis of $\sigma_1 = 0.06°$ and a short axis of $\sigma_2 = 0.02°$, centered on $x = \Delta x_{\text{pref}}/2$ in the left eye

and $x = -\Delta x_{\text{pref}}/2$ in the right. That is, the left-eye receptive field is

$$\rho_L(x,y,t) = \exp\left(-\frac{(x-\Delta x_{\text{pref}}/2)^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right). \quad (16)$$

For neurons tuned to horizontal orientations, $\sigma_x = \sigma_1$ and $\sigma_y = \sigma_2$, whereas for vertical orientations $\sigma_x = \sigma_2$ and $\sigma_y = \sigma_1$. The right-eye receptive field is similar with Δx_{pref} replaced by $-\Delta x_{\text{pref}}$. The standard deviation along the temporal axis, σ_t, is 10 ms.

Motion sensors

Our motion sensors differ in only two respects from the disparity sensors. First, their receptive fields are inseparable in space and time, making them sensitive to the direction of stimulus motion even when there is no interocular delay. Second, they are monocular, so they cannot sense disparity. In fact, essentially the same results are obtained with binocular motion sensors, which square the input from each eye before combining them. Such binocular motion sensors, although they would be sensitive to the disparity of a stimulus such as a bar (an inevitable consequence of having receptive fields of finite extent in the two eyes), would not be joint motion/disparity sensors in the usual sense because they would not be sensitive to disparity in cyclopean stimuli such as random-dot patterns. However, we used monocular sensors to make it clear that they are not disparity tuned. We include a population of motion sensors tuned to different orientations θ_{pref}. The single receptive field, in the left eye only, is centered on the cyclopean location of the disparity sensors' receptive fields, that is, the origin.

The motion sensors are tuned to a velocity along an axis orthogonal to their preferred orientation. Thus, sensors tuned to horizontal orientations are tuned to upwards or downwards motion. The receptive field for these sensors is

$$\rho_{\substack{\text{up} \\ \text{down}}}(x,y,t) = \exp\left(-\frac{(y\cos\varpi \mp t\sin\varpi)^2}{2\sigma_3^2} - \frac{x^2}{2\sigma_1^2} - \frac{(t\cos\varpi \pm y\sin\varpi)^2}{2\sigma_4^2}\right), \quad (17)$$

where

$$\sigma_3^2 = \sigma_t \frac{\sigma_t - \sqrt{\sigma_t^2 - \sigma_2^2\sin^2 2\varpi}}{2\sin^2\varpi};$$

$$\sigma_4^2 = \sigma_t \frac{\sigma_t + \sqrt{\sigma_t^2 - \sigma_2^2\sin^2 2\varpi}}{2\cos^2\varpi}; \tan\varpi = v$$
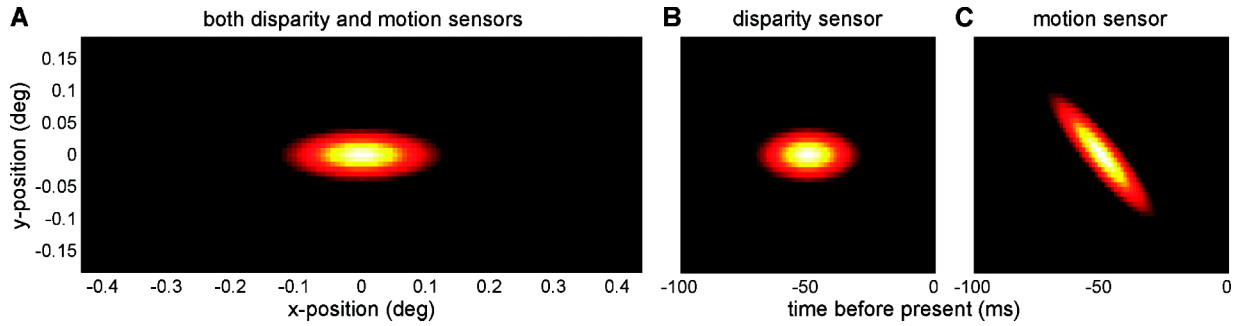
Figure 4. Receptive field profile used for the dynamic visual noise simulation. This simulation uses two spatial dimensions and includes receptive fields tuned to a range of orientations. This example shows a horizontally oriented receptive field ($\theta_{pref}= 0°$). (A) The spatial profile of the receptive field at the time of its peak response. The axes are horizontal and vertical position on the retina. This spatial profile is the same for both disparity sensors and motion sensors, except that for the disparity sensors, the profile is not necessarily centered on $x = 0$, as here. The scatter in the horizontal position of the receptive field means that the population includes sensors tuned to a range of horizontal disparities. (B and C) The vertical space/time profile of the receptive field, showing its dependence on time and on vertical retinal position, at the horizontal retinal position $x = 0$. (B) Separable space/time profile, used for the disparity sensors. (C) Inseparable space/time profile, used for the motion sensors.

and the $\pm$ determines whether the sensor prefers upward or downward motion. Similarly, sensors tuned to vertical orientations are

$$\rho_{\substack{\text{left} \\ \text{right}}}(x,y,t) = \exp\left(-\frac{(x\cos\varpi \mp t\sin\varpi)^2}{2\sigma_3^2}\right.$$

$$\left.-\frac{y^2}{2\sigma_1^2}-\frac{(t\cos\varpi \pm x\sin\varpi)^2}{2\sigma_4^2}\right).$$

σ_1 and σ_2 are the same as for the disparity sensors: $0.06°$ and $0.02°$, respectively. We show results for $v = 5$ deg/s ($\sigma_3 = 0.004, \sigma_4 = 0.046$ in units where time is in seconds and distance in degrees), and $v = 10$ deg/s ($\sigma_3 = 0.002, \sigma_4 = 0.098$).

Example receptive fields are shown in Figure 4. Because the receptive fields in the dynamic noise simulation are three dimensional, depending on x, y, and t, only slices through the receptive fields can be shown. Figure 4A shows the spatial profile of the receptive field at the moment of the cell's peak response (50 ms after the onset of a stimulus). This is the same for both disparity and motion sensors. The example shown here is tuned to horizontal orientations. Figure 4B–C show the vertical space/time profile of the receptive field, at the retinal position $x = 0$. Figure 4B is for a disparity sensor. The receptive field is space/time separable, as for the disparity sensor in Figure 3. Figure 4C is for a downward motion sensor. Here, the receptive field is space/time inseparable, meaning that the cell is tuned to a particular speed and direction of motion. The pixelation in this figure reflects the detail with which the receptive fields were sampled in the simulation: receptive field functions were evaluated on a grid of $117x$, $49y$, and $80t$ values. Due to the horizontal position disparity, the population of disparity sensors included members whose receptive

fields were centered on a range of x positions (whereas the receptive fields for the motion sensors were all centered on $x = 0$, like the example in Figure 4A). This is why the grid had to extend further in the x direction than in the y direction. The sampling was the same for both x and y: one pixel represented 0.45 arcmin in both directions.

We also performed simulations using Gabor receptive fields, with bandpass spatial and temporal frequency tuning. Here, the disparity sensors are

$$\rho_L(x,y,t) = \exp\left(-\frac{(t+t_{lag})^2}{2\sigma_t^2}\right)$$

$$\times \exp\left(-\frac{(x-\Delta x_{pref}/2)^2 + y^2}{2\sigma^2}\right)$$

$$\times \cos\left(2\pi f\left\{\left[x-\frac{\Delta x_{pref}}{2}\right]\sin\theta_{pref} + y\cos\theta_{pref}\right\}\right)$$

$$\times \cos\left(2\pi v(t+t_{lag})\right) \tag{18}$$

(the right eye's is similar with Δx_{pref} replaced by $-\Delta x_{pref}$), and the motion sensors are

$$\rho(x,y,t) = \exp\left(-\frac{(t+t_{lag})^2}{2\sigma_t^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

$$\cos\left(2\pi v(t+t_{lag}) + 2\pi f\{x\sin\theta_{pref} + y\cos\theta_{pref}\}\right), \tag{19}$$

where θ_{pref} is the preferred orientation of each neuron and also defines the preferred direction of motion for motion sensors. All neurons, both disparity and motion sensors,

had the same tuning to spatial and temporal frequency: $f = 2$ cycles per degree, $\nu = 10$ Hz, $\sigma = 0.1°$, $\sigma_t = 10$ ms. The results were essentially the same as those as shown in Figure 12 for the Gaussian receptive fields.

Images

Our dynamic visual noise stimulus consists of patterns of 117×49 pixels, in which each pixel is colored either black or white at random. As shown in Figure 4, each "time-pixel" in the simulation lasted 1.3 ms. To simulate the patterns used in experiment, a new pattern was generated every 10 time-pixels, corresponding to a simulated video frame of 13 ms. The simulated monitor was assumed to display each frame for exactly one time-pixel. Thus, a receptive field experienced each pattern for 1.3 ms, then experienced 11.4 ms of blank screen before the next pattern appeared. This was not critical to our results; essentially the same results were obtained when the simulated monitor was assumed to display each frame for a full 13 ms. The image presented to the right eye lagged one frame (13 ms) behind the left. A sequence of 50 random patterns was generated over 500 time-pixels, simulating a 633-ms presentation of visual noise. The response of the disparity and motion sensors was calculated at each of the 500 time-pixels. The input ν_L, ν_R from each eye's receptive field in response to an image $I(x, y, t)$ was calculated as in Equation 5, with an additional integral over all values of vertical retinal position y. The response of each disparity sensor, $C_D(t, \Delta x_{\text{pref}}, \theta_{\text{pref}})$, is given by the squared sum of the inputs from the two eyes (Equation 6), whereas the response of each motion sensor $C_M(t, \theta_{\text{pref}})$ is given by the squared input from the left eye. We then calculated the correlation coefficient r between the 500 responses of the motion sensor tuned to an orientation θ_{pref}, and the corresponding responses of disparity sensors tuned to the same orientation θ_{pref} and different disparities Δx_{pref}:

$$
r(\theta_{\text{pref}}, \Delta x_{\text{pref}}) = \left\{ \sum_{j=1}^{500} [C_M(t_j, \theta_{\text{pref}}) - \overline{C}_M(\theta_{\text{pref}})] \right.
$$

$$
\left. \times [C_D(t_j, \Delta x_{\text{pref}}, \theta_{\text{pref}}) - \overline{C}_D(\Delta x_{\text{pref}}, \theta_{\text{pref}})] \right\} \Bigg/
$$

$$
\left\{ \sqrt{\sum_{j=1}^{500} [C_M(t_j, \theta_{\text{pref}}) - \overline{C}_M(\theta_{\text{pref}})]^2} \right.
$$

$$
\left. \times \sqrt{\sum_{k=1}^{500} [C_D(t_k, \Delta x_{\text{pref}}, \theta_{\text{pref}}) - \overline{C}_D(\Delta x_{\text{pref}}, \theta_{\text{pref}})]^2} \right\},
$$

$$(20)$$

where the bar indicates the average over all times j. A single 633 ms presentation yields curves with the same features as

are visible in Figure 12, but with noise. To obtain the smooth curves shown in Figure 12, we repeated this process 500 times and took the average correlation coefficient.

Results

The stroboscopic Pulfrich effect

The classic Pulfrich effect has traditionally been explained by noting the geometrical equivalence of spatial and temporal disparity. However, this stimulus equivalence does not hold for the stroboscopic version of the effect. It is often argued that in this stimulus there is no spatial disparity in the images presented to the two eyes, only interocular delay (Burr & Ross, 1979; Lee, 1970a; Morgan & Thompson, 1975; Qian, 1997; Qian & Andersen, 1997). Of course, this argument depends critically on the assumption that each appearance of the stimulus in the left eye is paired with that appearance in the right eye, which occurs closest together in time. When the interocular delay Δt is less than half the strobe period T, then this match has zero spatial disparity.

However, even when $\Delta < T/2$, there are other possible matches, separated by longer periods of time, which do contain spatial disparity. We have previously developed a simple quantitative model that, while granting that matches separated by the shortest amount of time have the greatest influence on perception, also allows more widely separated matches to influence perception (Morgan, 1979; Read & Cumming, 2005b; Tyler, 1977). We refer to this as the disparity-averaging model. This model assumes that the disparity assigned to an object is made up of a weighted average of all possible matches between appearances of the target in the left and right eyes. The disparity of each match is weighted by the time delay between the left- and right-eye image in each match, so that matches between appearances which occur at nearly the same time in the two eyes influence perception more than matches between appearances which occur at very different times. The effective disparity in the stroboscopic stimulus is:

$$
\Delta x_{\text{perceived}} = \frac{\sum_{j=-\infty}^{\infty} jX \, w(jT + \Delta t)}{\sum_{j=-\infty}^{\infty} w(jT + \Delta t)},
$$

$$(21)$$

where w is the weight function describing how the weight given to a potential match falls off as a function of the interocular delay between the left and right members of the pair, T is the interflash interval of the stroboscope, and X is the distance traveled by the target during this period.

When the interflash interval t is short enough, the illumination is effectively continuous, so the strobe

Pulfrich stimulus must produce the same depth as the classic Pulfrich effect. That is, the effective disparity becomes equal to $v\Delta t$, the "virtual disparity" between the apparent motion trajectories of the target in the two eyes (Figure 1; Burr & Ross, 1979). It is easy to verify that Equation 21 satisfies this. As the interflash interval increases, Equation 21 correctly predicts that the effective disparity will fall below the virtual disparity as the interflash interval increases. When the weight function is a Gaussian with mean 0 and standard deviation ~15–20 ms, Equation 21 provides an excellent account of human perception (Read & Cumming, 2005b).

This disparity-averaging model (Equation 21) assumes that the appearances of the strobe stimulus in each eye have been identified and paired. It might therefore appear that it could be implemented only at a very high level, after the stereo correspondence problem has been solved. It turns out, however, that this is not the case. At any one-time separation, there is at most one possible match, so the spatial correspondence problem is trivial. The model of Equation 21 can be very straightforwardly read out from the population activity of disparity-tuned neurons in V1. In the next section, we show how this can be done.

Population response

We postulate an ensemble of disparity-sensitive units described by the stereo energy model (Ohzawa et al., 1990). Although this model does not capture all aspects of the responses of real disparity-sensitive cells (Read & Cumming, 2003), it has the advantage of mathematical simplicity. Similar results were obtained with the modified energy model units of Read, Parker, & Cumming (2002). The neurons in the ensemble are identical in all respects except for the position of their receptive fields on the retina, which gives them different preferences for stimulus disparity, Δx_{pref} and cyclopean position, x_{pref}.

Figure 5 shows how this population responds to one example stroboscopic Pulfrich stimulus, with a negative interocular delay Δt equal to 40% of the interflash interval. The stimulus is represented by the space/time diagrams along the top row (A–D). Dots indicate appearances of the stimulus in the left (red) and right (blue) eyes. Some of these are labeled for convenience in discussing the stimulus. The four columns show the response of the population at four different times in one period of the stimulus. The current time in each column is indicated by the yellow vertical line in the space/time plots A–D. A small complication is that because the neurons have a temporal lag in their response, they are not driven by the stimulus currently displayed, but rather by the stimulus as it was at previous times. The background of the space/time plot is shaded to show the temporal kernel of the neurons. The darker the shading, the less responsive the neurons are to stimuli at that time. The

maximum responsiveness, indicated by the bright region, occurs 50 ms before the current time.

Figure 5E–H shows the response of the population at the different times. Each pixel in the plot represents one neuron; the color shows its current firing rate (black = silent, white = maximal; the color scale is the same for all panels in a row). The neuron's position in the plot indicates its tuning: Preferred disparity is indicated by position along the horizontal axis and preferred cyclopean position by position along the vertical axis.

The features of this population response can be understood in relation to the stimulus. Column 1 (Figure 5AEIM) shows the situation at a time when the target has just made an appearance at $x = 0$ in the left eye and will shortly make an appearance in the right. However, these appearances have not yet begun to influence the neurons. The neurons are responding optimally to the second-to-last appearance of the target in the left eye (L1, at $x = -X, t = -T$), as shown by the fact that L1 falls in the middle of the bright band indicating the temporal kernel in the space/time diagram. This appearance has activated all the neurons with a left-eye receptive field close to $-X$. These neurons lie along a downwards diagonal stripe in the population plots, because the preferred disparity Δx_{pref} and cyclopean position x_{pref} compatible with a particular left-eye location x_L are given by $x_{pref} + \Delta x_{pref}/2 = x_L$, which defines a downward diagonal stripe on axes of $(\Delta x_{pref}, x_{pref})$.

In Column 2 (BFJN), time has moved on, and the neurons are no longer responding so strongly to L1. However, neurons with a receptive field at $x = -X$ in the right eye are now responding to the most recent appearance of the target in the right eye, R1 (at $x = -X$ and $t = -0.6T$; due to the interocular delay, this is $0.4T$ later than the corresponding appearance in the left eye). This response shows in Figure 5F as an upward diagonal stripe, $x_{pref} - \Delta x_{pref}/2 = x_R$. Naturally, the neurons which are firing most are those whose receptive fields are at $x = -X$ in both eyes because these receive excitation from both eyes. This explains the peak in the population activity at cyclopean position $x_{pref} = -X$ and $\Delta x_{pref} = 0$. Figure 5J shows only the binocular component (Equation 7) of the cells' response. This has removed the stripes due to monocular activation and focuses attention on the peak. Figure 5N shows the disparity distribution at this moment, that is, the binocular component averaged across cyclopean position. The distribution is symmetric and centered on $\Delta x_{pref} = 0$.

In Column 3 (CGKO), the neurons have almost entirely stopped responding to L1, and they have hardly begun yet to respond to L2, so there is no downward diagonal stripe of activity. They are still responding well to R1, so the most prominent feature in the population response is an upwards diagonal stripe, corresponding to the activation of neurons with a receptive field at $x = -X$ in the right eye. Because the input is essentially monocular at this moment, the binocular component in Figure 5K is very
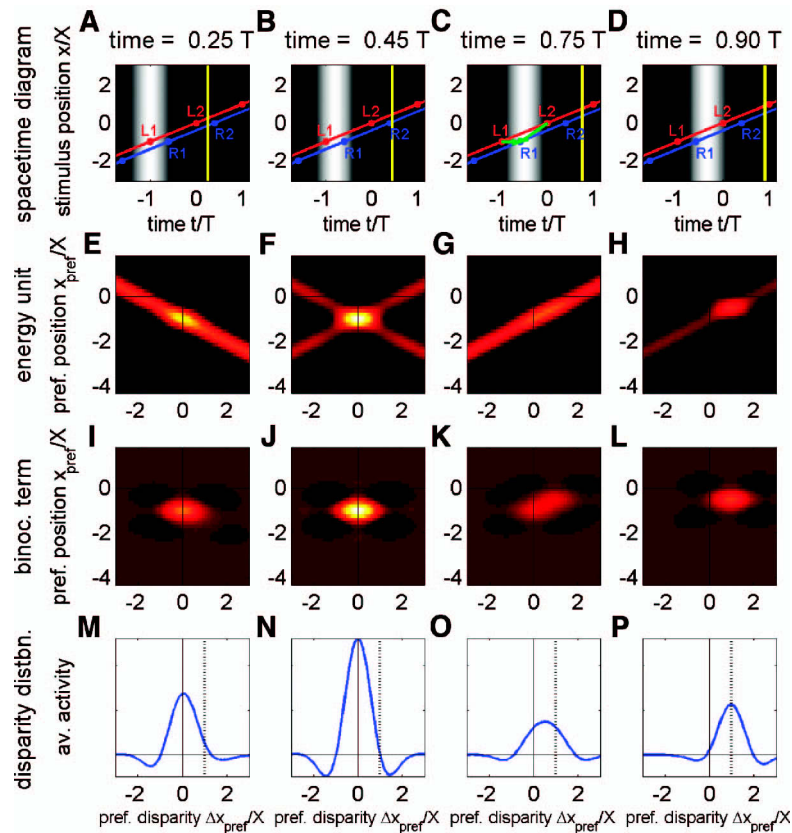
Figure 5. The response of a population of disparity-tuned neurons to a stroboscopic Pulfrich stimulus with period four times larger than RF time constant ($T = 4\tau = 40$ ms), shown at four different times. The delay is 0.4 times the strobe period ($\Delta t = 0.4T = 16$ ms). (A–D) Space/time diagrams for this stroboscopic Pulfrich stimulus. The yellow vertical lines show the current time; stimuli to the right of this line have not yet appeared. The background shading shows the temporal kernel of the receptive fields, which is Gaussian with a lag of 50 ms and an *SD* τ of 10 ms. Dark = least response; white = most responsive. Thus, the center of the bright vertical bar marks the stimulus that is most strongly driving the present response; this is always 50 ms before the present time. (E–H) Instantaneous response of a population of disparity-tuned energy model neurons to this stimulus. Each pixel in the plot represents a neuron; the color is the neuron's current firing rate. The preferred disparity of each neuron is plotted on the horizontal axis, while its preferred cyclopean position is on the vertical axis. The color scale is the same in each plot (black = silent, white = maximal firing). The cross-hairs show the axes. (I–L) Binocular component. As the middle row, except that the color shows only that component of the firing rate which is contributed by the binocular term in the energy model, Equation 7. (M–P) Binocular component averaged over RF cyclopean position. These are the plots in I–L, averaged down each column. The dashed line marks a disparity of one strobe interflash distance *X*. Note that all times are expressed as a fraction of the strobe interflash interval *T* and all distances are expressed as a fraction of the strobe interflash distance *X*.

weak. However, weak activation is visible at disparities of *X* and 0, corresponding to the pairings L1↔R1 and L2↔R1, marked with green arrows in Figure 5C. The disparity distribution, Figure 5O, represents the average of these, and peaks at $\Delta x_{\text{pref}} = X/2$.

In Column 4, the neurons are still responding weakly to R1 and are also beginning to respond weakly to L2. Its other appearances are either too recent to have yet influenced the neurons, or are too far in the past. The only stimulus disparity visible to the population is therefore *X*, corresponding to the match L2↔R1. This is visible in Figure 5LP, where the binocular component shows a peak for detectors tuned to a disparity of *X*, not zero. Any sensible read-out rule will therefore predict the perception of a nonzero disparity at this moment. However, this peak at disparity *X* is weaker than the

response at zero disparity in the second column. Consequently, one would expect a single disparity judgment based on this activity over time to lie close to zero than to *X*. The exact value of the disparity judgment will depend on what rule is used to combine these population distributions over time, which we explore below.

Figure 6 shows a similar set of results for a shorter strobe period. The strobe interflash interval *T* is now 20 ms, which is only twice the *SD* of the receptive field temporal kernel. The delay is once again 0.4*T*, that is, 8 ms. Because the stimulus period is now so much shorter relative to the neurons' temporal integration period, the population response varies very little with time. Instead of seeing strong peaks at zero disparity at some times, and weak peaks at disparity = *X* at others, as in Figure 5, the long integration time averages these peaks out. That
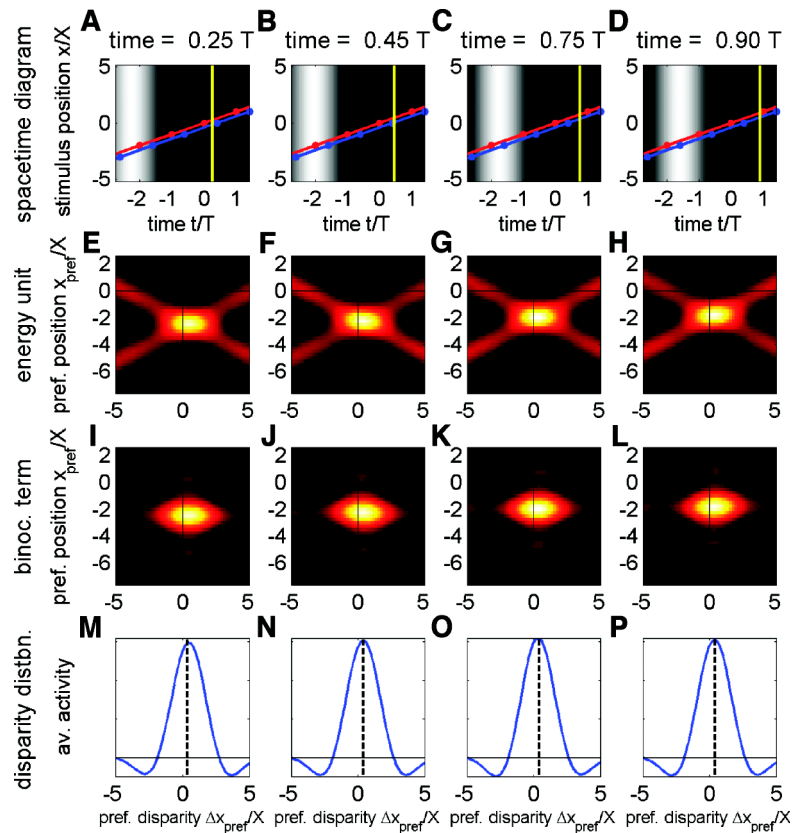
Figure 6. The response of a population of disparity-tuned neurons to a stroboscopic Pulfrich stimulus with $T = 2\tau$, shown at four different times. As Figure 5, except the strobe period T is 20 ms and the delay is 8 ms. Note that here the peak of the binocular response is at $0.4X$ (marked with a dashed line) for all time intervals.

is, the time averaging performed by the spatiotemporal filters themselves is sufficient that the peak in the population activity is nearly constant, located slightly to one side of zero disparity, at a preferred disparity of $0.4X$. It drifts up the vertical axis over time (Figure 6I–L), reflecting the apparent motion of the target, which stimulates neurons with different preferred cyclopean positions each time it appears. However, the disparity distribution remains constant (Figure 6M–P).

Read-out rule

The results in Figures 5 and 6 show the activity of a population of disparity detectors at different cyclopean locations, as a function of time. We wish to compare these results with psychophysical data in which subjects provided a single judgment of disparity over a whole trial, during which the moving target appeared at many different locations. To do this we must combine all the disparity signals over space and time to yield a single disparity value. One can think of this as implemented by a higher brain area pooling inputs over time from V1 cells tuned to many different cyclopean positions but the same disparity. This results in a measure of "total support" for each possible stimulus disparity. For simplicity, we assume that this estimate is performed simply by calculating the mean of all responses.

To understand the results of this averaging, it is useful to consider the results at each time instant (so the mean is calculated only over all cyclopean positions). Figure 7 shows this average disparity response as a function of time, for the two different strobe periods illustrated in Figures 5 and 6. The color of each pixel in Figure 7 represents the total activity of a pool of V1 neurons tuned to the same disparity, but different cyclopean positions. Each pixel row in Figure 7 shows the same data as one of the panels M–P of Figures 5 and 6, here represented in pseudocolor. Each pixel row represents the distribution of activity across the whole population of disparity detectors at a given time. Each pixel column shows how the total activity in all neurons with a given preferred disparity varies over time. The stimulus is periodic, and so is the steady-state response of the neuronal pools; therefore, only one period is shown.

Figure 7A is for the stimulus illustrated in Figure 5, where the strobe interflash interval is four times the time constant of the V1 neurons. The blue line traces the peak, that is, the preferred disparity of the currently most active neuronal pool. For part of each stimulus period, as in Figure 5IJ, there is a strong peak of activity in the neuronal pool tuned to zero disparity. For the rest, there is a weaker peak in the pool of neurons whose preferred disparity is the strobe interflash distance X, as in Figure 5L.
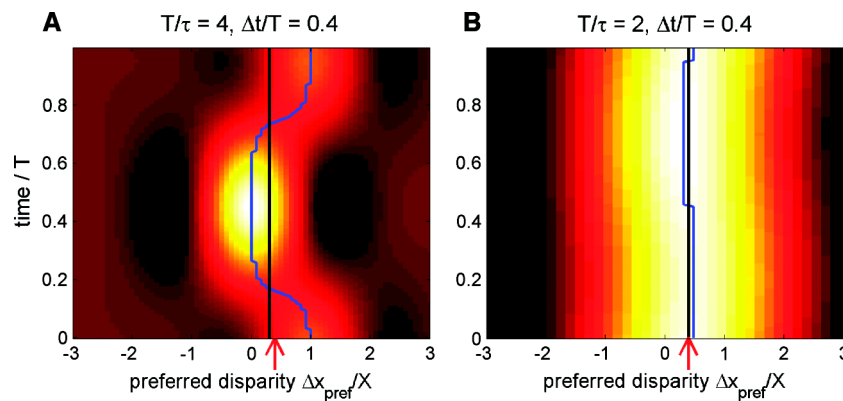
Figure 7. Total activity as a function of time for neuronal pools with different preferred disparities. The pseudocolor shows $\int_{-\infty}^{+\infty} dx_{\mathrm{pref}} B(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}})$: the binocular component of neuronal firing, averaged across neurons with different RF locations x_{pref}, as a function of preferred disparity and time. (A) Strobe interflash interval is four times the neuronal integration period τ, as in Figure 5; (B) strobe interflash interval is 2τ, as in Figure 6. The blue line traces the peak of the population activity at each moment of time. The straight black line shows the disparity judgment made for the stimulus, extracted from this activity according to Equation 14. For comparison, the virtual disparity $v\Delta t$ is indicated with a red arrow.

To generate a single depth judgment from this population response, we simply take the mean of the whole distribution (Equation 15, disparity averaging). Thus, for example, in Figure 7A, the time constant of the V1 neurons means that they respond to two of the disparities present in the stimulus: neurons tuned to a disparity of 0 respond strongly, and neurons tuned to a disparity of X respond weakly. (The stimulus contains other possible matches, with disparities of $2X$, $3X$, etc., but the halves of these matches are separated by so long a time that the neurons do not respond to them.) Under our disparity-averaging read-out rule, the effective disparity lies in between the two peaks, but closer to the stronger peak. This is shown with the black line in Figure 7. The effective disparity is thus a weighted average of the disparities present in the stimulus, with the weight depending on the temporal delay between the different possible matches. In Figure 7B, the strobe interflash interval is short relative to the integration time of the neurons, so the most active pool is always the same, namely, the pool with preferred disparity $0.4X$.

It can be proved (see Appendix) that taking the mean of the population disparity activity in this way yields the same results as the weighted disparity-averaging equation, Equation 21, with weight function equal to the cross-correlation of the temporal components of the receptive fields in each eye. Figure 8 demonstrates this. The crosses in Figure 8 show the effective disparity obtained by reading out the activity of a simulated neuronal population, for strobe Pulfrich stimuli with different interocular delays Δt and interflash intervals T. The solid curves show the effective disparity obtained with the original disparity-averaging equation, Equation 21, when the weight function is a Gaussian with SD equal to $\tau\sqrt{2}$ (inset in Figure 8; this is the cross-correlation of the temporal receptive fields, which are Gaussians

with SD τ). The results are the same. Thus, the read-out rule presented here represents a simple way to implement the weighted disparity-averaging equation with a population of physiologically plausible model neurons.

Effect of different read-out rules

The read-out rule discussed so far (Equation 15) was chosen to implement disparity averaging (Equation 21), in which the disparities of the possible matches present in the stimulus are averaged after weighting by the interocular delay of each match. Figure 8 shows that this equation can be implemented simply by averaging the output of V1 disparity sensors. But of course, there are many other ways in which population activity in V1 might be processed to arrive at a subjective report of disparity. We have experimented with various read-out rules and found that they generally fall into two classes. Either the effective disparity is generally less than the virtual disparity, as in Figure 8 (except that the precise position of each curve depends on the read-out rule), or the effective disparity is always equal to the virtual disparity (that is, plots of $\Delta x/X$ vs. $\Delta t/T$ lie along the identity line). Figure 9 shows an example of the latter case. In this simulation, instead of averaging V1 activity over time and then extracting a single disparity for the whole stimulus, an instantaneous disparity, Δx_{inst}, is assigned at every moment based on the preferred disparity of the most active V1 cells (winner takes all). The resulting disparities were then averaged over time. Formally, this rule is

$$\Delta x_{\mathrm{eff}} = \frac{1}{T} \int_0^T dt \, \Delta x_{\mathrm{inst}}(t), \text{ where}$$

$$\Delta x_{\mathrm{inst}}(t) = \frac{\arg \max}{\Delta x_{\mathrm{pref}}} \int_{-\infty}^{+\infty} dx_{\mathrm{pref}} \, B(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}}). \qquad (22)$$

The instantaneous disparity corresponds to the blue line in Figure 7. Looking at Figure 7A, for example, the read-out

rule of Equation 22 notes that the peak of the activity is at zero disparity 60% of the time and at disparity X 40% of the time, and so assigns effective disparity $0.4X$ (the virtual disparity, Figure 9). With this read-out rule, the effective disparity is always the virtual disparity, even for long interflash intervals where human subjects report disparities much closer to zero. Although this rule does not therefore match experimental data, it is nevertheless of interest. It demonstrates that it is possible for a population of pure disparity sensors to encode the virtual disparity implicit in the apparent motion of the strobe stimulus, although the sensors do not respond to motion. As we shall see below (Figure 14), the reason for this paradoxical result is that even pure disparity sensors become sensitive to direction of motion in stimuli with an interocular delay, due to the geometrical equivalence of motion and disparity in such stimuli (Pulfrich, 1922).

To recover the sigmoid pattern which is in fact obtained with human observers, we need to take into account that the zero-disparity peak not only lasts longer, but is of larger amplitude. We can modify Equation 20 to achieve this by weighting the instantaneous disparity Δx_{inst} (Equation 22) by the height of the peak:

$$\Delta x_{\mathrm{eff}} = \frac{\int_0^T dt\, B_{\max}(t)\Delta x_{\mathrm{inst}}(t)}{\int_0^T dt\, B_{\max}(t)},\ \text{where}$$

$$B_{\max}(t) = \max_{\Delta x_{\mathrm{pref}}} \int_{-\infty}^{+\infty} dx_{\mathrm{pref}} B(t, x_{\mathrm{pref}}, \Delta x_{\mathrm{pref}}). \qquad (23)$$

This read-out rule (results not shown) gives results that are similar to those of Equation 15 (Figure 8): The effective disparity is in general less than the virtual disparity, although the magnitude of the effective disparity is slightly different from the read-out rule of Equation 15.

Effect of different receptive field functions

The receptive fields used in this model were the product of a Gaussian temporal kernel and a Gabor spatial kernel. What are the effects of different choices for the spatial and temporal kernels? For the tuned-excitatory neurons considered here (neurons with position disparity but no phase disparity), the effective disparity specified by Equation 15 is independent of the spatial component of the receptive field (see Appendix). Thus, the choice of Gabor functions with a particular phase, spatial frequency and bandwidth was immaterial to the results. In particular, the same results would have been obtained with Gabor functions of any phase (provided that the phase was the same in both eyes' receptive fields, to make the neurons tuned-excitatory). It follows that although the results displayed in this paper are for binocular simple cells, with only one receptive field in each eye, the same results would be obtained for model complex cells receiving input from pairs of receptive fields in quadrature in each eye (Ohzawa et al., 1990), or from many receptive fields with random phases.
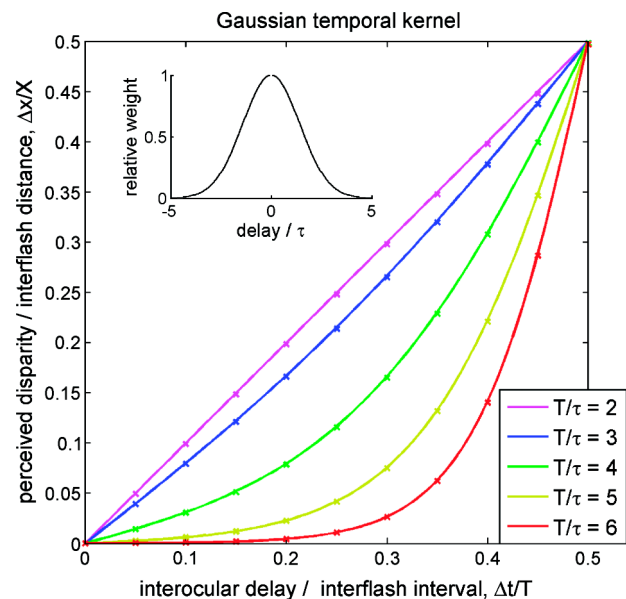

Figure 8. Effective disparity as a function of interocular delay in a simulation of the stroboscopic Pulfrich effect based on pure disparity sensors. Effective disparity Δx, as a proportion of the interflash distance X, is plotted as a function of interocular delay Δt as a proportion of the interflash interval T. Crosses show the results from a population of energy model units, where the temporal component of the receptive field function is a Gaussian with standard deviation τ, and the effective disparity is calculated according to Equation 15. Solid curves are for the model of Equation 21, where the weight function is the auto-correlation of the receptive field temporal kernel, that is, a Gaussian with standard deviation $\tau\sqrt{2}$ (inset). The different colors are for different stimulus interflash intervals (see legend).

The temporal component of the receptive field does affect the effective disparity. As already noted, a population of model V1 neurons, with the read-out rule specified by Equation 15, yields the same effective disparity as the disparity-averaging equation, Equation 21, when the temporal weight function is the cross-correlation of the V1 temporal kernels in the two eyes. Figure 8 showed the results for model neurons with a Gaussian temporal kernel (Equation 2). As another example, the crosses in Figure 10 show results from a neuronal simulation when the temporal receptive fields are exponential impulse functions (Equation 3). Once again, the same results are obtained using Equation 21 (solid curves in Figure 10), when the weight function is the auto-correlation of the exponential kernel (inset).

The receptive fields used so far in this model are separable in space and time. However, around 20–30% of V1 neurons are direction-selective (DeValois, Yund, & Hepler, 1982; Hawken, Parker, & Lund, 1988; Orban, Kennedy, & Bullier, 1986; Schiller, Finlay, & Volman, 1976), implying that they have space/time-inseparable receptive fields. Binocular direction-selective neurons would jointly encode motion and disparity (Anzai et al., 2001; Pack et al., 2003; Qian & Andersen, 1997; Read & Cumming, 2005a). What percept
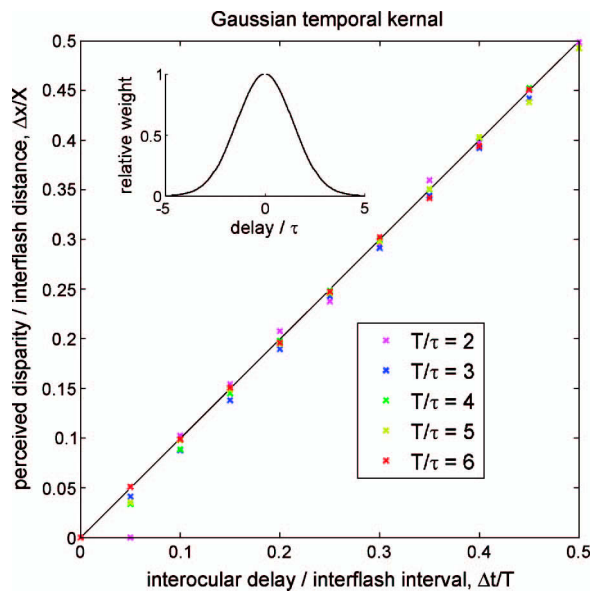
Figure 9. Effective disparity as a function of interocular delay for an alternative read-out rule, in which disparity is extracted by a winner-take-all rule prior to averaging over time (Equation 22). As in Figure 8, the receptive fields are space/time separable, with Gabor spatial kernels and Gaussian temporal kernels.

would the disparity-averaging read-out rule, Equation 15, imply if it were applied to the response of such joint motion/disparity sensors? To address this, we ran the simulation using neurons whose receptive fields were oriented Gaussians in space and time (Figure 3B), tuned to the apparent velocity of the moving stimulus. The results are shown in Figure 11. The solid curves show the results that would have been obtained with space/time-separable receptive fields with the same temporal extent. The simulation results are slightly different quantitatively, but essentially the same as those generated by space/time-separable filters. In particular, the same sigmoid pattern is observed, showing that this feature is largely determined by the temporal integration of the receptive fields and the read-out rule used to determine effective disparity. Whether the receptive fields are space/time separable or not is of little importance. Qian & Andersen (1997) used joint motion/disparity sensors like those used in Figure 11 and obtained effective disparity equal to the virtual disparity (identity line in Figure 11). There are several differences between their simulation and ours that account for the different results. Most importantly, Qian and Andersen used a winner-take-all read-out rule, and we have already seen in Figure 9 that using a winner-take-all rule to arrive at instantaneous disparity produces effective disparity equal to the virtual disparity.

Dynamic noise with an interocular delay

Having shown that realistic pure-disparity sensors can explain the depth percept in the stroboscopic Pulfrich effect,

we now move on to another important stimulus, dynamic visual noise with an interocular delay but no spatial disparity. Joint encoding of motion and disparity has frequently been invoked to explain the "swirling" perception of depth in this stimulus (Morgan, 2003; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Qian & Andersen, 1997). In this section, we show that the illusion can equally well be explained using initial sensors that encode only motion and only disparity. We show that—if the right eye's image is delayed relative to the left—there is a positive correlation between the activity of the near disparity sensors and an independent population of rightward motion sensors, and between the activity of the far disparity sensors and the leftward motion sensors. We argue that this correlation could, in itself, give rise to the perception of swirling depth motion, along the lines proposed by Tyler (1977).

To implement these ideas quantitatively, we considered two populations of model neurons based on the energy model: (1) a population of disparity-selective neurons, like those in the previous simulation, which encode disparity but are not sensitive to the direction of stimulus motion; (2) a population of monocular direction-selective neurons, which are not sensitive to binocular disparity. Because the experimental stimulus we are investigating contained motion energy in all directions, we use two spatial dimensions in our simulations and include neurons tuned to horizontal and vertical stimuli. This means there are four types of motion sensor: preferring upward, downward, leftward, or rightward motion. The disparity-selective population includes neurons tuned to different horizontal
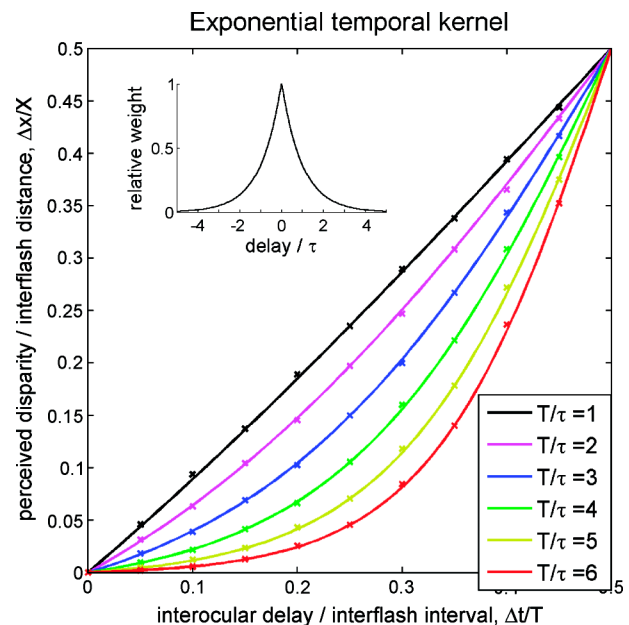

Figure 10. Effective disparity as a function of interocular delay for exponential temporal kernels. As Figure 8, except that the temporal kernel of neurons in the simulation decayed exponentially. The weight function used in Equation 21 to obtain the solid curves is the auto-correlation of this temporal kernel (inset).
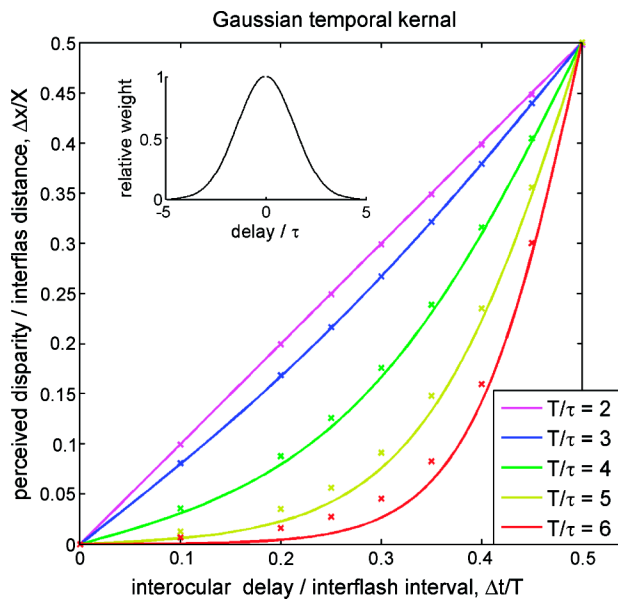
Figure 11. Effective disparity as a function of interocular delay for joint motion/disparity sensors. As Figure 8, except that the receptive fields of the neurons in the simulation were 2D Gaussians oriented in space/time along the apparent velocity of the stimulus (Figure 3B). To obtain an equivalent temporal kernel for comparison with the space/time-separable case, the receptive field was integrated over space, yielding a Gaussian function of time with standard deviation τ. The auto-correlation of this temporal kernel, that is, a Gaussian with standard deviation $\tau\sqrt{2}$ (inset), was used with Equation 21 to generate the solid curves.

stimulus disparities Δx_{pref}, because the receptive fields in the two eyes are centered on different positions. Note that the preferred disparity was always horizontal, irrespective of the orientation preference of the disparity sensor. In contrast, all the motion sensors are monocular, with a single receptive field centered on the origin of the left retina. All neurons in the simulation have identical spatial and temporal frequency tuning.

In the simulation, we calculated the response of both neuronal populations at 500 moments 1.3 ms apart, while the eyes viewed fifty 13 ms frames of a binocular random noise stimulus, with the right eye's image delayed by 1 frame relative to the left. Then, for each preferred orientation θ_{pref} and disparity Δx_{pref} in turn, we calculated the correlation $r(\theta_{\mathrm{pref}}, \Delta x_{\mathrm{pref}})$ between the 500 successive responses of the motion sensor tuned to θ_{pref} and the 500 successive responses of the disparity sensor tuned to θ_{pref} and Δx_{pref} (Equation 20). To remove noise, we repeated this process 500 times with different random noise stimuli and averaged the correlation coefficient. The curves in Figure 12A shows this average correlation coefficient as a function of preferred orientation and disparity. The four colors correspond to the four motion sensors, tuned to motion leftward (red), rightward (blue), downward (green dashed), or upward (pink, largely obscured under the green curve). Each point on the curve

shows the correlation between the activity of that motion sensor, and the activity of the disparity sensor with the same preferred orientation and with preferred disparity indicated on the horizontal axis.

First, we consider the curves for the two motion sensors tuned to upward and downward motion. These sensors have the highest correlation with disparity sensors tuned to zero disparity. This reflects the simple fact that when a stimulus strongly activates the monocular RF of a disparity detector, it will also strongly activate a motion sensor with the same monocular receptive field. In our simulation, all motion sensors have receptive fields at the origin of the left eye. The only disparity sensors to have receptive fields in this position are those tuned to zero disparity. (The mean position of the receptive fields in the two eyes is zero for all disparity sensors, and so for those tuned to nonzero disparities, the receptive fields lie to one side or the other of the origin.) A pair of images that happens to have a lot of
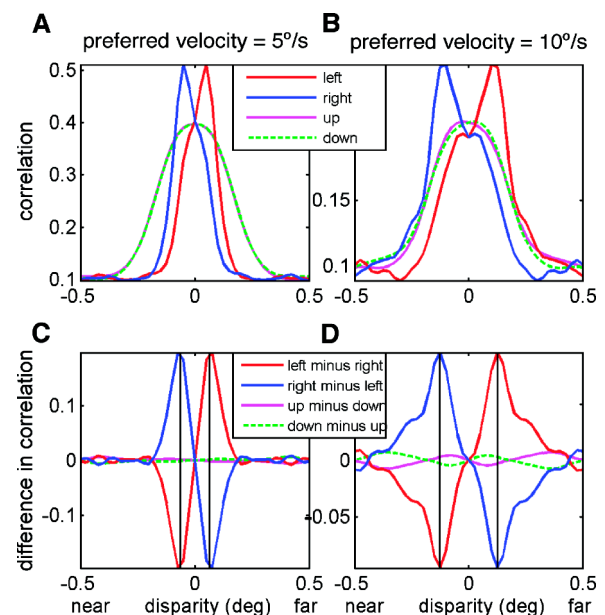

Figure 12. Correlation between the responses of disparity sensors and motion sensors. The interocular delay was a single frame, $\Delta t = 13$ ms. Each curve shows the correlation between the activity of one motion sensor (as indicated in the legend) and the activity of a population of disparity sensors, with preferred disparity shown on the horizontal axis and the same preferred orientation as the motion sensor. The two sharper-peaked curves (red, blue) show the correlation for motion sensors tuned to leftward and rightward motion (note that the preferred orientation is therefore vertical). The two broader curves (green, pink obscured underneath) show the correlation for horizontally oriented sensors, where the motion sensors prefer upwards or downwards motion. These results are for the sensors defined in Equations 16 and 17. The smaller, broader peaks obtained for the 10 deg/s case occur because of the way we defined our Gaussian receptive fields: the 10 deg/s sensors have a broader velocity bandwidth than the 5 deg/s sensors. The black vertical lines in panels C and D mark disparities $\pm v\Delta t$.

white pixels close to the origin in both eyes will drive all the motion sensors strongly but will preferentially drive the disparity sensors with $\Delta x_{\text{pref}} = 0$, resulting in a strong correlation between the activity of the motion sensor and the zero-disparity sensor. Importantly, this correlation is identical for motion detectors sensing opposite directions (up vs. down), so it provides no basis for any sensation of motion in any one direction.

This situation is very different when considering motion sensors tuned to rightward and leftward motion (note that these sensors have vertically oriented receptive fields). These curves have narrower peaks as a function of horizontal disparity tuning, as expected because their receptive field functions have a smaller horizontal standard deviation. However, the most noticeable feature is that these curves are displaced from the origin. The activity of the rightward motion sensor is most strongly correlated with the activity of disparity sensors tuned to near disparities, while the leftward motion sensor is correlated with far disparities. As Figure 13 illustrates, this occurs for a very simple reason (Tyler, 1977). Figure 13 shows two successive frames in each eye (left eye, top row; right eye, bottom row). The two successive images presented to the left eye (top row) happen to contain motion energy to the left, so activating the leftward motion sensor. But the stimulus is presented with an interocular delay of one frame, such that the image presented to the right eye on the second frame is the same as the image presented to the left eye on the previous frame (labeled "Image 1" in Figure 13). This means that the stimulus on the second frame must necessarily contain far disparity. Now consider a bank of motion sensors tuned to different velocities, and a bank of disparity sensors tuned to different disparities. We can group these into fours: a leftward motion sensor tuned to *v*, a rightward sensor tuned to −*v*, a far disparity sensor tuned to disparity *v*Δ*t*, and a near disparity sensor tuned to −*v*Δ*t*. Whatever the precise form of the spatiotemporal filters, provided that the image pairs 1 and 2 activate the leftward motion sensor more than its rightward partner, then they must activate the far disparity sensor more than its near disparity partner. This is the origin of the correlation in Figure 12 between activity in leftward motion sensors and in far disparity sensors.

Another way of looking at this is to note that when there is an interocular delay, our disparity sensors are actually also sensitive to direction of motion, although they have space/time-separable receptive fields. The cartoon in Figure 14 explains why. The red and blue ellipses represent space/time-separable receptive fields in the left and right eyes, respectively. The receptive fields have a position disparity, meaning that the binocular neuron receiving input from both receptive fields is tuned for that disparity. The upper panels show the trajectory of a zero-disparity object moving to the left (Figure 14A) and to the right (Figure 14B), with no interocular delay. The receptive fields are identically activated by both trajectories, so the binocular neuron gives the same response for
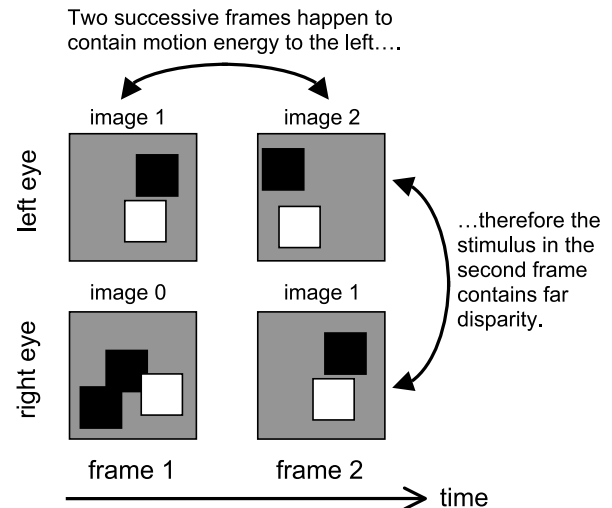

Figure 13. Geometrical relationship between motion and disparity energy in noise with an interocular delay. This shows two frames from a small region of a hypothetical random noise stimulus. The top row shows the two frames presented to the left eye, and the bottom row those presented to the right eye. The stimulus has an interocular delay of one frame, such that the image presented to the right eye on the second frame is the same as the image presented to the left eye on the previous frame. The two frames presented to the left eye happen to contain motion to the left. As a consequence, the second frame contains far disparity.

leftward and rightward motion. This is obvious—because the receptive field is space/time separable, it cannot encode direction of motion.

Now consider what happens when an interocular delay is added to the stimulus, so that the right eye (blue line) leads the left (red). This is shown in the lower panels (Figure 14C–D). Now, the binocular neuron responds much more strongly to the rightward (D) than the leftward trajectory (C). This is a consequence of the geometric equivalence of spatial disparity and interocular delay noted by Fertsch (Pulfrich, 1922). The same interocular delay corresponds to opposite disparities for opposite directions of motion. Thus, the direction selectivity which this cell manifests for stimuli with an interocular delay is really just a consequence of its disparity selectivity for normal stimuli. When the stimulus has an interocular delay, all disparity sensors must necessarily become selective for direction of motion as well disparity. When the right eye is delayed, near disparity sensors effectively become rightward motion sensors, explaining why their activity is correlated with that of cells whose space/time-inseparable receptive fields make them rightward motion sensors to begin with. This is a property of the stimulus, not the particular way in which the sensors encode disparity. Note that to make use of this property to identify motion direction, the readout mechanism would have to know the sign of the interocular delay, so this property may not be a useful way for the brain to encode movement direction (it has motion sensors
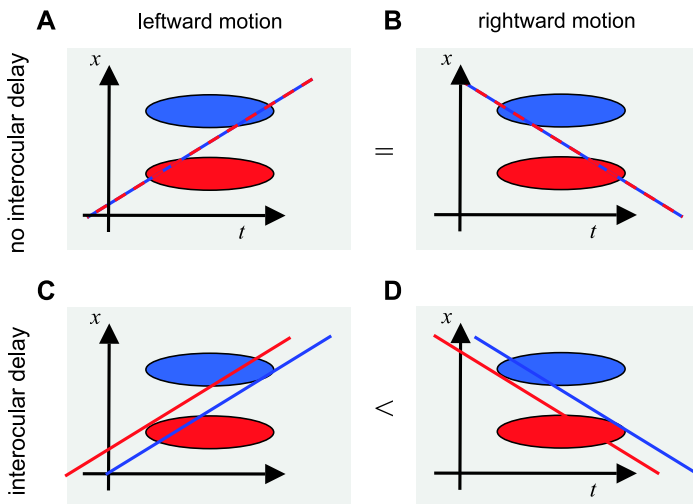
Figure 14. Disparate space/time-separable receptive fields become sensitive to the direction of motion when an interocular delay is added. The ellipses represent the left-eye (red) and right-eye (blue) receptive fields of a binocular neuron tuned to a near disparity. The lines represent the trajectory of a moving stimulus in the left eye (red) and right eye (blue). In AC the stimulus is moving from right to left; in BD, it is moving from left to right. The dots in the lower plots represent the appearances of a stimulus under stroboscopic illumination. (A and B) Zero disparity stimulus with no interocular delay. Left- and right-eye trajectories superimpose. The distance the left-eye trajectory spends within the left-eye receptive field is the same in both panels A and B and similarly for the right eye. The binocular neuron is therefore stimulated identically by both stimulus directions. (C and D) Stimulus with an interocular delay, such that the right-eye image is delayed (equivalent to a disparity). Now, the binocular neuron is stimulated much more by the rightwards-moving stimulus (D), where the left-eye trajectory substantially overlaps the left-eye receptive field, and similarly for the right eye, than by the leftwards-moving stimulus (C), where the left-eye trajectory barely intersects the left-eye receptive field. The same argument applies for a stroboscopic stimulus, in which the lines, representing a moving object under continuous illumination, are replaced by the dots, representing the object's brief appearances under stroboscopic illumination. In panel D, one of the left-eye (red) dots falls squarely within the left-eye receptive field; in panel C, all left-eye dots fall outside the left-eye receptive field. This explains qualitatively how disparity sensors with space/time-separable receptive fields can respond differently depending on the direction of motion of a stimulus with interocular delay, which is the basis for the illusion of depth due to interocular delay in our model.

for that). Its importance lies in understanding why motion and disparity become inextricably linked in the presence of an interocular delay.

Figure 12 demonstrates a correlation between the disparity sign and the motion direction of local energy variations in dynamic noise with an interocular delay. The simulations also predict the observed relationship between disparity magnitude and motion speed. Figure 12A–C

showed results where all motion sensors were tuned to a speed of 5 deg/s, while Figure 12B–D shows analogous results for 10 deg/s. In each case, the greatest difference in correlation between leftward and rightward motion sensors occurs at a disparity of $v\Delta t$, marked with vertical lines in Figure 12C–D. Similar results were also obtained with Gabor receptive fields, Equations 18 and 19, comparing channels tuned to different spatial frequencies and hence different speeds (not shown). Again, the reason for this is clear from Figure 13. A pair of frames which optimally stimulates a leftward motion sensor tuned to speed v must also stimulate the disparity sensor tuned to disparity $v\Delta t$. This means that larger disparities will be associated with higher speeds, in accordance with the reported shearing percept of this stimulus (Tyler, 1974, 1977).

Tyler (1977) also studied the effect of introducing an interocular delay into binocularly anti-correlated noise. In this stimulus, the contrast polarity is inverted in one eye, so that (before the interocular delay) each white pixel in the left eye corresponds to a black pixel in the right, and vice versa. Tyler found that there was still a perception of depth in this stimulus, although considerably weaker than in the original correlated stimulus. However, the perceived association of motion and depth was reversed. Tyler was able to explain both the reversal and the weaker percept with his random spatial disparity hypothesis. Similarly, our model shows a reversal in the correlations between disparity and motion sensors. Whether the model also predicts weaker depth percept depends on what population of disparity detectors are used (in particular the range of phase disparities). In practice, however, it is already known that real V1 neurons respond more weakly to anti-correlated stimuli than the energy model (Cumming & Parker, 1997; Read et al., 2002), so any attempt to model the strength of the depth percept would need to accommodate this fact first.

Discussion

The stroboscopic Pulfrich stimulus and dynamic noise are two intriguing stimuli in which an illusion of depth is created by introducing a time delay between the images presented to the two eyes. A number of previous studies have argued that, to understand these illusions, it is necessary to consider the responses of spatiotemporal filters (Anzai et al., 2001; Carney et al., 1989; Morgan, 1979, 1992; Morgan & Fahle, 2000; Morgan & Thompson, 1975; Morgan & Tyler, 1995; Morgan & Watt, 1982; Pack et al., 2003; Qian & Andersen, 1997). Until recently, such explanations have depended upon the activation of sensors that jointly encode motion and disparity, most fully elaborated in the work of Qian & Andersen (1997). This theory received support with the finding that joint motion/disparity encoding was common in cat striate cortex when disparity was applied orthogonal to the receptive field orientation (Anzai et al., 2001). However, two recent

reports have found that joint motion/disparity encoding is rare in monkey V1 (Pack et al., 2003; Read & Cumming, 2005a). If this is also true in humans, the joint-encoding theory would imply that the vast majority of disparity-sensitive cells in V1 do not contribute to depth perception in Pulfrich-like stimuli. This led us to consider the possibility that joint motion/disparity encoding may not, after all, be the primary neuronal substrate for Pulfrich-like effects. We find that the depth percept both in the stroboscopic Pulfrich effect and in dynamic visual noise can be explained equally well with separate encoding of motion and disparity, based on distinct populations of disparity sensors and motion sensors. Physiologically, motion and disparity are known to be encoded both jointly and separately in the striate cortex. Our results demonstrate that there are no grounds to suppose that any one class of these filters plays a privileged role in perception.

Stroboscopic Pulfrich effect

For the strobe Pulfrich effect, we had previously put forward an explanation based on averaging the spatial disparities present in the stimulus (Read & Cumming, 2005b). We have now done for our disparity-averaging explanation what Qian & Andersen (1997) did for the joint-encoding explanation of the Pulfrich effect: we demonstrated that it can be implemented as a detailed physiological model, based on the activity of a population of model neurons with a realistic integration time. Unlike the model of Qian and Andersen, our model gives quantitatively correct predictions for the magnitude of disparity perceived at different interflash intervals and interocular delays (Morgan, 1979; Read & Cumming, 2005b). The key part of the model that generates these predictions is its rule for "reading out" the population activity of early disparity sensors to produce a single disparity judgment (Equation 15). We postulate that, in the strobe Pulfrich stimulus, effective disparity reflects the average activity of sensors tuned to different disparities. With this read-out rule, it is immaterial whether the neuronal disparity sensors are also sensitive to direction of motion or not. Virtually the same results were obtained from a population of joint motion/disparity sensors tuned to the apparent velocity of the stimulus, as from a population of pure disparity sensors insensitive to direction of motion (Figures 8 and 11). Conversely, with a different read-out rule, a population of pure disparity sensors could yield effective disparity equal to the virtual disparity (Figure 9), previously assumed to be a unique signature of joint motion/disparity sensors. Crucially, it is changes in the read-out rule that generate the important differences in predicted percepts for various stimuli; the choice of initial encoding has very little influence. The only significant property of the receptive fields is their temporal extent: their finite integration time allows them to respond to disparities of binocular matches whose left and right

members occur at different times (providing a degree of time-averaging that precedes any read-out rule). Qian and Andersen, in common with other authors such as Morgan & Fahle (2000), recognized that extending the energy model to include time was critical to the understanding of Pulfrich-like phenomena, but all these workers have assumed that the receptive fields are inseparable functions of space and time, leading to joint motion/disparity encoding. Our simulations show that separable functions of space and time work just as well in explaining the psychophysics. Thus, we cannot conclude that this percept is supported by joint motion/disparity sensors, as opposed to separate populations of motion and disparity sensors.

The read-out rule specified in Equation 15 is not the only one compatible with the psychophysics. Several other read-out rules also give a pattern of results similar to that observed psychophysically, in which the effective disparity is zero for long interflash intervals and approaches the virtual disparity implied by the apparent motion of the target as the interflash interval is reduced. As noted above, there are also fairly simple and plausible read-out rules in which the effective disparity is always the virtual disparity (Figure 9). The different results obtained with different read-out rules demonstrate that quantitative measurements of depth perception in this stimulus potentially carry information not only about the temporal integration time of the disparity sensors (Morgan, 1979; Read & Cumming, 2005b), but also about the way in which their activity is read out to yield perception. The read-out rules of Equations 15 and 23 are compatible with psychophysical data; that of Equation 22 is not.

One could object to the way Equation 15 averages neuronal activity over time to produce a single depth percept from the varying population activity. If the response of early disparity detectors varies with time, perhaps we should expect a time-varying depth percept too. This is allowed for in the read-out rule of Equation 23, which extracts an instantaneous disparity from the neuronal activity at each moment. This instantaneous disparity is then averaged over time (weighted by the strength of the percept at each moment) to give a single depth judgment. To date, all studies of the stroboscopic Pulfrich effect have measured only the perceived mean disparity, and in general the modulations in instantaneous disparity produced by our model are beyond the temporal limits of human stereopsis (Norcia & Tyler, 1984; Regan & Beverley, 1973). Thus, the psychophysical data are not sufficiently detailed to permit us to choose between the read-out rules of Equations 15 and 23, or indeed the other possibilities which certainly exist.

Dynamic visual noise

To explain the illusory perception in the dynamic noise stimulus, we revisit an earlier explanation of Tyler's (1974; 1977), which pointed out that random pairings of dots in

this stimulus will tend to associate disparity and direction. Although the original form of this idea (in terms of individual dots) has been discarded, we develop a modified version based on spatiotemporal filters. We model the response of separate populations of disparity sensors and motion sensors to dynamic noise with an interocular delay and show that the activity in these populations is systematically correlated. If it is the right eye which is delayed by Δt, then far disparity sensors are more strongly correlated with leftward motion sensors than rightward ones (and vice versa for near disparity sensors). If we consider only motion sensors tuned to speed v, then this difference between leftward and rightward is most pronounced for disparity sensors tuned to $v\Delta t$. The reason for this was explained by Tyler (1974; 1977). Suppose the right eye is delayed by frame, and suppose two successive frames presented to the left eye happen to contain motion to the left. But then, when the left eye is viewing the second frame, the right eye is viewing the first frame, which must therefore contain far disparity. The very structure of the stimulus means that rightward motion inevitably co-occurs with near disparity, and leftward motion with far disparity (Figure 13).

It seems entirely plausible that these correlations might generate the observed percept of a cloud of noise moving in all directions, with a tendency for leftward-moving points to appear beyond the screen and rightward-moving points to be in front, with the distance from the screen proportional to their speed. This explanation does not, in itself, require the correlation between far disparity sensors and leftward motion sensors to be explicitly represented by neural activity anywhere in the brain. The binding together of the motion signal with the depth signal is an example of the general "binding problem", whose physiological substrate is poorly understood. The ability to perceive a distant object moving to the left does not necessarily imply the existence of a joint "leftward, far" detector any more than the ability to perceive its color implies a joint "leftward, red" detector. However, it would also be trivial to construct neurons that would explicitly represent these correlations, even if the initial encoding of motion and disparity were entirely separate. A simple output nonlinearity suffices to make a neuron sensitive to correlation between its inputs; this is why energy model neurons, with a half-squaring nonlinearity, are sensitive to the correlation between inputs from left and right eyes. Many neurons in area MT are selective for both disparity and motion (Bradley et al., 1995; DeAngelis et al., 1998; DeAngelis & Newsome, 2004; DeAngelis & Uka, 2003; Maunsell & Van Essen, 1983; Pack et al., 2003; Roy et al., 1992), while many neurons in MST show an interdependence of disparity and direction selectivity (Roy et al., 1992; Roy & Wurtz, 1990). Joint motion/disparity tuning in MT may be inherited from the minority of joint motion/disparity sensors in V1, just as direction selectivity in MT seems to be largely inherited from the minority of direction-selective cells in V1 (Movshon & Newsome, 1996).

Implications

The two stimuli simulated in this paper are not the only ones in which interocular delay leads to a perception of depth, but they are those most often cited in support of the view that joint encoding of motion and depth is required to explain the Pulfrich effect. We believe that the principles of the explanation we develop here can account for depth perception in all other Pulfrich-like stimuli, without invoking space/time-inseparable binocular filters. This implies that there is currently no evidence suggesting a privileged role for joint encoding of motion and disparity in depth perception. Although space does not permit us to demonstrate the model's behavior to all possible variants of the Pulfrich effect, it is instructive to consider briefly one or two additional examples. (1) Applying interocular delay to dynamic random-dot stereograms results in a decrease in stereoacuity, with no shift in the point of subjective equivalence, on a forced-choice front/back discrimination task. Both features are predicted by the behavior of pure disparity sensors in V1; the decrease in stereoacuity with interocular delay agrees well with the binocular integration time estimated physiologically (Read & Cumming, 2005a). (2) Morgan & Ward (1980) argued against Tyler's (1977) disparity-averaging hypothesis based on a variant of the dynamic noise stimulus, in which dots moved horizontally for n frames before being replaced by a new dot in a random position. They reported that perceived depth increases with n. Many authorities still consider this compelling evidence in favor of joint motion disparity encoding. However, when one considers the changes that occur between any pair of frames in this stimulus, it is easy to see why our model accounts for the results equally well. On each frame only $100/n\%$ of the dots are replaced, whereas in a standard noise stimulus 100% of the dots are replaced on each frame. The dots that are not replaced move coherently and hence produce a spatial disparity as in the classic Pulfrich effect. The dots that are replaced behave just like the dynamic noise we simulated above. Thus, the stimulus of Morgan and Ward is a simple sum of a standard interocularly delayed noise stimulus (to which it reduces when $n = 1$) and a random-dot strobe Pulfrich stimulus (to which it asymptotes as $n\rightarrow\infty$, Figure 15). Morgan and Ward asked their subjects to match a probe to the depth of the dots. They do not report results for $n = 1$, but it seems clear that in this case, the matching depth must be zero because there is symmetry about fixation. Our own results with this stimulus confirm that interocular delay in zero-disparity noise does not bias depth perception (see Figure 10 of Read & Cumming, 2005a). As n increases above 1, the symmetry about fixation is broken, because there is now more power in horizontal motion to the right than in other directions. In terms of depth, this places more power at the virtual disparity (blue arrows in Figure 15). It therefore does not seem surprising that the matching depth reported by Morgan and Ward's subjects moved
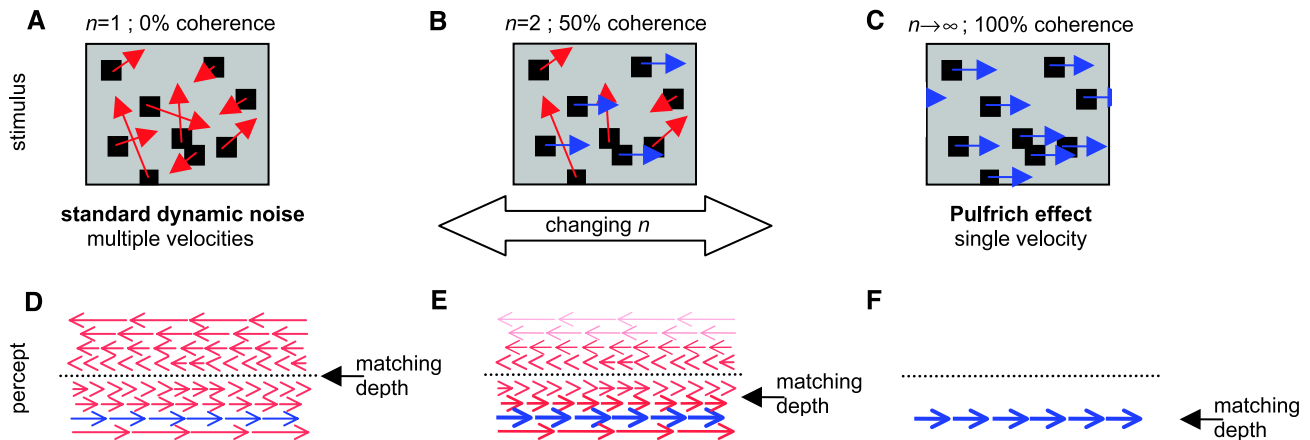
Figure 15. Stimulus and percept for the Morgan–Ward stimulus with different values of *n*. (A–C) Random-dot motion stimulus with varying percentages of coherent motion. (D–F) Sketch of percept, after Tyler (1977). Dotted line indicates fixation plane; arrows represent motion (length of arrow indicates speed, position relative to fixation indicates depth). Blue represents the coherent motion component. As *n* is increased, more power is placed into this velocity. Red represents the random motions also present in the stimulus for low *n*. The black arrow ← indicates the depth that would be reported in a matching task. For *n* = 1, the stimulus is dynamic visual noise, with motion in all directions. The depth reported in a depth-matching task must be zero because the stimulus is symmetric on either side of fixation. As *n* tends to infinity, the stimulus becomes the classic Pulfrich stimulus (strictly, a strobe Pulfrich stimulus, but the frame rate is high) that contains only a single speed. The depth-matching task must report the corresponding virtual disparity. For intermediate *n*, Morgan & Ward (1980) reported that the reported matching depth was intermediate between 0 and the virtual disparity.

away from zero towards this virtual disparity. For large *n*, the matching disparity was equal to the virtual disparity implied by the apparent motion of the stimulus, as expected because the stimulus is now a strobe Pulfrich stimulus with a short interflash interval (25 ms) (Morgan, 1979; Read & Cumming, 2005b). Thus, the results of Morgan and Ward follow naturally from the percepts elicited by dynamic noise and the Pulfrich effect, both of which can be explained with separate motion/disparity encoding. We are not aware of any stimuli that require joint motion/disparity encoding to explain them.

Why then have Pulfrich-like stimuli been so widely accepted as evidence for joint motion/disparity encoding? One way of understanding the confusion is to notice that, although by definition pure disparity sensors are not sensitive to direction of motion in normal stimuli, they are sensitive to direction of motion in stimuli with an interocular delay (Figure 14). This is a consequence of the geometrical equivalence of disparity and motion in such stimuli (Pulfrich, 1922): Even pure disparity sensors effectively turn into joint motion/disparity sensors once an interocular delay is introduced into the stimulus.

As noted in the Introduction section, this recognition has substantial implications for our understanding of stereopsis. The existing theory of Pulfrich-like phenomena assumes that only joint motion/disparity sensors can support the illusory depth percept. If so, then pure disparity sensors (never explicitly considered in these theories, but constituting a majority in V1) are presumably signaling zero disparity. The veridical zero-disparity percept signaled by this majority of disparity sensors is

apparently overridden by the illusory percept signaled by a small minority of direction-selective disparity sensors. If true, this would require re-evaluation of the relationship between perception and neuronal activity in V1 and extrastriate cortex. However, our results show that in fact, the illusory depth percept in Pulfrich-like stimuli can be equally well encoded by pure disparity sensors (and the motion percept by pure motion sensors). Once the finite temporal integration time of cortical neurons is taken into account, Pulfrich-like stimuli contain real spatial disparities, which suffice to explain the depth percept they cause. Thus, even for a hypothetical brain in which motion and stereo were processed in completely separate neural pathways, the stroboscopic Pulfrich stimulus and dynamic visual noise could still produce the same depth percept as they do for us. Of course, real brains are known to contain some filters which do encode motion and disparity jointly. Our work does not imply that these filters are not used, only that there is no evidence that they have a privileged role in perception. The extent to which disparity and motion are encoded jointly or separately therefore remains a question for physiology rather than psychophysics, unless a stimulus can be devised which genuinely separates interocular delay and disparity. Current physiological evidence suggests that disparity is encoded largely separately from motion in V1 (Pack et al., 2003; Read & Cumming, 2005a), while both are encoded jointly in MT and MST. Our results suggest that all these disparity signals—irrespective of whether they are encoded jointly with motion, or separately—contribute to stereo depth perception.

Appendix

In this Appendix, we prove that a population of space/time-separable tuned-excitatory energy-model neurons, with the read-out rule specified in Equation 15, gives the same effective disparity as the disparity-averaging equation postulated in Read & Cumming (2005b), with the temporal weight function given by the cross-correlation of the temporal components of the receptive fields in the two eyes. The spatial component of the receptive fields is immaterial.

In our disparity-averaging read-out rule, the effective disparity is given by Equation 15:

$$\Delta x_{\text{eff}} = \frac{\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \Delta x_{\text{pref}} D(\Delta x_{\text{pref}})}{\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} D(\Delta x_{\text{pref}})}.$$

If we substitute for D from Equations 7, 10, 11, and 14, we obtain the expanded form:

$$\Delta x_{\text{eff}} = \frac{\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \Delta x_{\text{pref}} \int_0^T dt \int_{-\infty}^{+\infty} dx_{\text{pref}} \sum_{j=-\infty}^{+\infty} \rho_{0x}\left(jX - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty}\rho_{0x}\left(kX - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0t}(kT + \Delta t - t)}{\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \int_0^T dt \int_{-\infty}^{+\infty} dx_{\text{pref}} \sum_{j=-\infty}^{+\infty} \rho_{0x}\left(jX - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty}\rho_{0x}\left(kX - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0t}(kT + \Delta t - t)},$$

where we have assumed that the receptive field function $\rho_0(x, t)$ is space/time separable, that is, it can be written as the product of a spatial component $\rho_{0x}(x)$ and a temporal component $\rho_{0t}(t)$. Note that, for convenience, the sums over j and k are written as extending to infinity, rather than terminating at $j = \lfloor t/T \rfloor$ etc. This does not alter the result, because, due to causality, $\rho_{0t}(t)$ is zero for $t > 0$ (future time). Thus, terms for which j exceeds $\lfloor t/T \rfloor$ contribute nothing to the sum anyway.

Next, we rearrange the order of summation and integration so that the integral over preferred cyclopean position, x_{pref}, is performed first:

$$\Delta x_{\text{eff}} = \frac{\int_0^T dt \sum_{j=-\infty}^{+\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty} \rho_{0t}(kT + \Delta t - t) \int_{-\infty}^{+\infty} \Delta x_{\text{pref}} \Delta x_{\text{pref}} \int_{-\infty}^{+\infty} dx_{\text{pref}} \rho_{0x}\left(jX - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0x}\left(kX - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}\right)}{\int_0^T dt \sum_{j=-\infty}^{+\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty} \rho_{0t}(kT + \Delta t - t) \int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \int_{-\infty}^{+\infty} dx_{\text{pref}} \rho_{0x}\left(jX - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0x}\left(kX - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}\right)}.$$

(A1)

Let us write S for the cross-correlation of the spatial components of the receptive fields in the two eyes: $S(d) = \int_{-\infty}^{+\infty} dx \rho_{0x}(x)\rho_{x0}(x-d)$. Then the integrals over x_{pref} in Equation A1 are simply

$$\int_{-\infty}^{+\infty} dx_{\text{pref}} \rho_{0x}\left(jX - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}\right)\rho_{0x}\left(kX - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}\right) = \int_{-\infty}^{+\infty} dx_{\text{pref}} \rho_{0x}(x)\rho_{0x}(x-(j-k)X + \Delta x_{\text{pref}}) = S((j-k)X - \Delta x_{\text{pref}}).$$

The effective disparity is then

$$\Delta x_{\text{eff}} = \frac{\int_0^T dt \sum_{j=-\infty}^{+\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty} \rho_{0t}(kT + \Delta t - t) \int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \cdot \Delta x_{\text{pref}} \cdot S((j-k)X - \Delta x_{\text{pref}})}{\int_0^T dt \sum_{j=-\infty}^{+\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{+\infty} \rho_{0t}(kT + \Delta t - t) \int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \cdot S((j-k)X - \Delta x_{\text{pref}})}.$$

Changing the integration variable from Δx_{pref} to $\Delta x_{\text{pref}}' = (j-k)X - \Delta x_{\text{pref}}$, the integral over Δx_{pref} can be rewritten as

$$\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \Delta x_{\text{pref}} S((j-k)X - \Delta x_{\text{pref}}) = \int_{-\infty}^{+\infty} d\Delta x'_{\text{pref}} ((j-k)X - \Delta x'_{\text{pref}}) S(\Delta x'_{\text{pref}}).$$

If the receptive field profiles are identical in the two eyes, as assumed in this model, then their cross-correlation $S(d)$ is even-symmetric about zero. The second term therefore vanishes, and we have

$$\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} \Delta x_{\text{pref}} S((j-k)X - \Delta x_{\text{pref}}) = (j-k)X \int_{\infty}^{+\infty} d\Delta x'_{\text{pref}} S(\Delta x'_{\text{pref}})$$

while

$$\int_{-\infty}^{+\infty} d\Delta x_{\text{pref}} S((j-k)X - \Delta x_{\text{pref}}) = \int_{-\infty}^{+\infty} d\Delta x'_{\text{pref}} S(\Delta x'_{\text{pref}}).$$

The term $\int_{-\infty}^{+\infty} d\Delta x'_{\text{pref}} S(\Delta x'_{\text{pref}})$ therefore cancels out between the numerator and denominator, and we are left with

$$\frac{\Delta x_{\text{eff}}}{X} = \frac{\int_0^T dt \sum_{j=-\infty}^{\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{\infty} \rho_{0t}(kT + \Delta t - t)(j-k)}{\int_0^T dt \sum_{j=-\infty}^{\infty} \rho_{0t}(jT-t) \sum_{k=-\infty}^{\infty} \rho_{0t}(kT + \Delta t - t)}.$$

This proves that, for tuned-excitatory neurons where the receptive field spatial profiles are identical in the two eyes, the shape of the spatial profile does not influence the effective disparity obtained with this read-out rule.

To proceed further, we take the integral over time t within the summation over j, and replace the integration variable t with $t' = jT - t$:

$$\frac{\Delta x_{\text{eff}}}{X} = \frac{\sum_{j=-\infty}^{\infty} \int_{(j-1)T}^{jT} dt' \rho_{0t}(t') \sum_{k=-\infty}^{\infty} \rho_{0t}((k-j)T + \Delta t + t')(j-k)}{\sum_{j=-\infty}^{\infty} \int_{(j-1)T}^{jT} dt' \rho_{0t}(t') \sum_{k=-\infty}^{\infty} \rho_{0t}((k-j)T + \Delta t + t')}. \tag{A2}$$

We now replace the summation variables j and k in both numerator and denominator with sums over $n = (j+k)/2$ and $m = (j-k)/2$. To do this, we note that for any function f,

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(j,k) \equiv \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n+m, n-m) + \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n+m+1, n-m).$$

Therefore, the numerator of the right-hand side of Equation A2 may be rewritten as

$$\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\int_{(j-1)T}^{jT}dt\rho_{0t}(t)\rho_{0t}((k-j)T+\Delta t+t)(j-k)=\sum_{m=-\infty}^{\infty}2m\sum_{n=-\infty}^{\infty}\int_{(n+m-1)T}^{(n+m)T}dt\rho_{0t}(t)\rho_{0t}(-2mT+\Delta t+t)$$

$$+\sum_{m=-\infty}^{\infty}(2m+1)\sum_{n=-\infty}^{\infty}\int_{(n+m)T}^{(n+m+1)T}dt\rho_{0t}(t)\rho_{0t}(-(2m+1)T+\Delta t+t)$$

(where for simplicity we have dropped the prime on t' in Equation A2). Now, different values of n simply select different ranges of time to integrate over. Thus, we can replace the summation over n and finite integral over t with a single infinite integral over t:

$$\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\int_{(j-1)T}^{jT}dt\rho_{0t}(t)\rho_{0t}((k-j)T+\Delta t+t)(j-k)=\sum_{m=-\infty}^{\infty}2m\int_{-\infty}^{\infty}dt\rho_{0t}(t)\rho_{0t}(-2mT+\Delta t+t)$$

$$+\sum_{m=-\infty}^{\infty}(2m+1)\int_{-\infty}^{\infty}dt\rho_{0t}(t)\rho_{0t}(-(2m+1)T+\Delta t+t). \qquad (A3)$$

Having done this, we can now merge the two sums over m into a single sum over a variable p:

$$\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\int_{(j-1)T}^{jT}dt\rho_{0t}(t)\rho_{0t}((k-j)T+\Delta t+t)(j-k)=\sum_{p=-\infty}^{\infty}p\int_{-\infty}^{\infty}dt\rho_{0t}(t)\rho_{0t}(-pT+\Delta t+t);$$

the first term on the right-hand side of Equation A3 summed over even values of p, and the second over odd values, and we have now unified these.

Carrying out a similar procedure for the denominator of Equation A2, we obtain

$$\frac{\Delta x_{\mathrm{eff}}}{X}=\frac{\displaystyle\sum_{p=-\infty}^{\infty}p\int_{-\infty}^{\infty}dt\rho_{0t}(t)\rho_{0t}(t+\Delta t-pT)}{\displaystyle\sum_{p=-\infty}^{\infty}\int_{-\infty}^{\infty}dt\rho_{0t}(t)\rho_{0t}(t+\Delta t-pT)}\;.$$

Now let us write $\mathcal{T}$ for the cross-correlation of the temporal components of the receptive fields in the two eyes: $\mathcal{T}(\tau)=\int_{-\infty}^{+\infty}dx\rho_{0t}(t)\rho_{0t}(t-\tau)$. Thus,

$$\frac{\Delta x_{\mathrm{eff}}}{X}=\frac{\displaystyle\sum_{p=-\infty}^{\infty}p\mathcal{T}(pT-\Delta t)}{\displaystyle\sum_{p=-\infty}^{\infty}\mathcal{T}(pT-\Delta t)}.$$

This is simply the disparity-averaging equation of Read & Cumming (2005b), with the weight function given by the cross-correlation of the temporal components of the receptive fields in the two eyes. This proves that the read-out rule of Equation 15 provides a simple way of implementing disparity averaging within a population of realistic model neurons.

Acknowledgments

References

Anzai, A., Ohzawa, I., & Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: Position vs. phase. *Proceedings of the National Academy of Sciences of the United States of America, 94,* 5438–5443. [PubMed] [Article]

Anzai, A., Ohzawa, I., & Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: Neural basis of the Pulfrich effect. *Nature Neuroscience, 4*(5), 513–518. [PubMed]

Bradley, D. C., Qian, N., & Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature, 373*(6515), 609–611. [PubMed]

Bridge, H., & Cumming, B. G. (2001). Responses of macaque V1 neurons to binocular orientation differences. *Journal of Neuroscience, 21*(18), 7293–7302. [PubMed] [Article]

Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research, 19*(5), 523–532. [PubMed]

Carney, T., Paradiso, M. A., & Freeman, R. D. (1989). A physiological correlate of the Pulfrich effect in cortical neurons of the cat. *Vision Research, 29*(2), 155–165. [PubMed]

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389,* 280–283. [PubMed]

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394*(6694), 677–680. [PubMed]

DeAngelis, G. C., & Newsome, W. T. (2004). Perceptual "Read-Out" of conjoined direction and disparity maps in extrastriate area MT. *PLoS Biology, 2*(3), E77. [PubMed] [Article]

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89*(2), 1094–1111. [PubMed] [Article]

DeValois, R., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22,* 531–544. [PubMed]

Falk, D. S., & Williams, R. (1980). Dynamic visual noise and the stereophenomenon: Interocular time delays, depth, and coherent velocities. *Perception & Psychophysics, 28*(1), 19–27. [PubMed]

Hawken, M. J., Parker, A. J., & Lund, J. S. (1988). Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. *Journal of Neuroscience, 8*(10), 3541–3548. [PubMed]

Lee, D. N. (1970a). Spatio-temporal integration in binocular-kinetic space perception. *Vision Research, 10*(1), 65–78. [PubMed]

Lee, D. N. (1970b). A stroboscopic stereophenomenon. *Vision Research, 10*(7), 587–593. [PubMed]

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*(5), 1148–1167. [PubMed]

Morgan, M. J. (1975). Stereoillusion based on visual persistence. *Nature, 256*(5519), 639–640. [PubMed]

Morgan, M. J. (1976). Pulfrich effect and the filling in of apparent motion. *Perception, 5*(2), 187–195. [PubMed]

Morgan, M. J. (1979). Perception of continuity in stroboscopic motion: A temporal frequency analysis. *Vision Research, 19*(5), 491–500. [PubMed]

Morgan, M. J. (1980). Analogue models of motion perception. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 290*(1038), 117–135. [PubMed]

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature, 355*(6358), 344–346. [PubMed]

Morgan, M. J. (2003). *The space between our ears.* London: Weidenfeld & Nicolson.

Morgan, M. J., & Castet, E. (1995). Stereoscopic depth perception at high velocities. *Nature, 378*(6555), 380–383. [PubMed]

Morgan, M. J., & Fahle, M. (2000). Motion-stereo mechanisms sensitive to inter-ocular phase. *Vision Research, 40*(13), 1667–1675. [PubMed]

Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception, 4*(1), 3–18. [PubMed]

Morgan, M. J., & Tyler, C. W. (1995). Mechanisms for dynamic stereomotion respond selectively to

horizontal velocity components. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 262*(1365), 371–376. [PubMed]

Morgan, M. J., & Ward, R. (1980). Interocular delay produces depth in subjectively moving noise patterns. *Quarterly Journal of Experimental Psychology, 32*(3), 387–395. [PubMed]

Morgan, M. J., & Watt, R. J. (1982). Mechanisms of interpolation in human spatial vision. *Nature, 299*(5883), 553–555. [PubMed]

Morgan, M. J., & Watt, R. J. (1983). On the failure of spatiotemporal interpolation: A filtering model. *Vision Research, 23*(10), 997–1004. [PubMed]

Movshon, J. A., & Newsome, W. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Journal of Neuroscience, 16*(23), 7733–7741. [PubMed]

Neill, R. A. (1981). Spatio-temporal averaging and the dynamic visual noise stereophenomenon. *Vision Research, 21*(5), 673–682. [PubMed]

Norcia, A. M., & Tyler, C. W. (1984). Temporal frequency limits for stereoscopic apparent motion processes. *Vision Research, 24*(5), 395–401. [PubMed]

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. [PubMed]

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. *Journal of Neurophysiology, 56*(2), 462–480. [PubMed]

Pack, C. C., Born, R. T., & Livingstone, M. S. (2003). Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. *Neuron, 37*(3), 525–535. [PubMed]

Pulfrich, C. (1922). Die Stereoscopie im Dienste der isochromen und heterochromen Photometrie. *Naturwissenschaft, 10,* 553–564.

Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron, 18*(3), 359–368. [PubMed]

Qian, N., & Andersen, R. A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research, 37*(12), 1683–1698. [PubMed]

Read, J. C. A., & Cumming, B. G. (2003). Testing quantitative models of binocular disparity selectivity in primary visual cortex. *Journal of Neurophysiology, 90*(5), 2795–2817. [PubMed] [Article]

Read, J. C. A., & Cumming, B. G. (2005a). The effect of interocular delay on disparity selective V1 neurons: Relationship to stereoacuity and the Pulfrich effect. *Journal of Neurophysiology, 94,* 1541–1553 [PubMed]

Read, J. C. A., & Cumming, B. G. (2005b). The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *Journal of Vision, 5*(5), 417–434, http://journalofvision.org/5/5/3/, doi:10.1167/5.5.3. [PubMed] [Article]

Read, J. C. A., Parker, A. J., & Cumming, B. G. (2002). A simple model accounts for the reduced response of disparity-tuned V1 neurons to anti-correlated images. *Visual Neuroscience, 19,* 735–753. [PubMed]

Regan, D., & Beverley, K. I. (1973). Some dynamic features of depth perception. *Vision Research, 13*(12), 2369–2379.

Ross, J. (1974). Stereopsis by binocular delay. *Nature, 248*(446), 363–364. [PubMed]

Roy, J. P., Komatsu, H., & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *Journal of Neuroscience, 12*(7), 2478–2492. [PubMed]

Roy, J. P., & Wurtz, R. H. (1990). The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature, 348*(6297), 160–162. [PubMed]

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *Journal of Neurophysiology, 39*(6), 1288–1319. [PubMed]

Tyler, C. W. (1974). Stereopsis in dynamic visual noise. *Nature, 250*(5469), 781–782. [PubMed]

Tyler, C. W. (1977). Stereomovement from interocular delay in dynamic visual noise: A random spatial disparity hypothesis. *American Journal of Optometry and Physiological Optics, 54*(6), 374–386. [PubMed]