

Pulfrich phenomena are coded effectively by a joint motion-disparity process

Ning Qian

Department of Neuroscience, Columbia University,
New York, NY, USA



Ralph D. Freeman

Group in Vision Science, School of Optometry,
Helen Wills Neuroscience Institute,
University of California, Berkeley, CA, USA



Pulfrich phenomena are a class of depth illusions generated by an interocular time delay. This may be demonstrated with continuously moving stimuli, stroboscopic displays undergoing apparent motion, or dynamic noise patterns. Previous studies suggest that neurons jointly tuned to motion and disparity may be responsible for the phenomena. Model cells with such joint coding can explain all Pulfrich phenomena in a unified way (N. Qian & R. A. Andersen, 1997). However, the joint-coding idea has been challenged by recent models (J. C. Read & B. G. Cumming, 2005a, 2005c) that focus on the S shaped functions of perceived disparity in stroboscopic Pulfrich effect (M. J. Morgan, 1979). Here we demonstrate fundamental problems with the recent models in terms of causality, physiological plausibility, and definitions for joint and separate coding, and we compare the two coding schemes under physiologically plausible assumptions. We show that joint coding of disparity and either unidirectional or bidirectional motion selectivity can account for the S curves, but unidirectional selectivity is required to explain direction-depth contingency in Pulfrich effects. In contrast, separate coding can explain neither the S curves nor the direction-depth contingency. We conclude that Pulfrich phenomena are logically accounted for by joint encoding of unidirectional-motion and disparity.

Keywords: Pulfrich effect, computational models, motion, disparity, stereovision, binocular vision

Citation: Qian, N., & Freeman, R. D. (2009). Pulfrich phenomena are coded effectively by a joint motion-disparity process. *Journal of Vision*, 9(5):24, 1–16, <http://journalofvision.org/9/5/24/>, doi:10.1167/9.5.24.

Introduction

Motion processing and stereovision are conceptually related. Motion is displacement over time and stereovision relies on image displacement between the two eyes. Numerous studies suggest that the link between the two visual functions is more than conceptual. Physiologically, many cells are tuned to various combinations of direction of motion and binocular disparity (Anzai, Ohzawa, & Freeman, 2001; Bradley, Qian, & Andersen, 1995; Grunewald & Skoumbourdis, 2004; Pack, Born, & Livingstone, 2003). Psychophysically, motion and stereo processing are strongly interactive. For instance, motion aftereffects are contingent on disparity (Anstis, 1974; Nawrot & Blake, 1989; Neri & Levi, 2008; Regan & Beverley, 1973), and disparity facilitates transparent motion perception (Adelson & Movshon, 1982; Qian, Andersen, & Adelson, 1994a). Computational models have been used to integrate motion and stereo vision into a unified framework (Fernández, Watson, & Qian, 2002; Qian, 1994; Qian & Andersen, 1997; Qian, Andersen, & Adelson, 1994b).

Pulfrich phenomena provide an intriguing case for the study of motion-stereo integration. In the Pulfrich effect, a pendulum moving back and forth continuously in a

frontoparallel plane appears to traverse an elliptical path in depth when an interocular time delay is introduced (Pulfrich, 1922). The classic explanation is that an interocular time delay corresponds to a spatial displacement (disparity) between the two eyes' views. This in turn causes an apparent continual shift in depth. However, a stroboscopic Pulfrich effect exists whereby each eye is presented a flashed stimulus (Burr & Ross, 1979; Lee, 1970; Morgan, 1979; Morgan & Thompson, 1975). The two eyes' flashes may have the same spatial positions but occur at different times. Therefore, at a given time, there is no conventionally defined spatial disparity. The Pulfrich depth effect is also present with dynamic noise patterns (Falk & Williams, 1980; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Ross, 1974; Tyler, 1974). An interocular time delay makes a dynamic noise pattern appear as a volume resolving in depth. The classic Pulfrich explanation fails here because there is no coherent motion to convert an interocular time delay into binocular disparity. These problems are resolved by an integrated motion-stereo model which can account for all three variants of the Pulfrich effect (Qian & Andersen, 1997). This model requires a mathematical demonstration that cells with plausible binocular spatiotemporal receptive fields treat an interocular time delay as an equivalent binocular disparity. For stroboscopic Pulfrich stimuli,

these cells interpolate across space and time (Morgan, 1979). The model is consistent with subsequent physiological data on motion-stereo integration (Anzai et al., 2001; Grunewald & Skoumbourdis, 2004; Pack et al., 2003).

The notion of joint coding has been recently challenged by two papers, the first (Read & Cumming, 2005c) including a conceptual model and the second a computational one (Read & Cumming, 2005a). The first paper states that: 1) The stroboscopic Pulfrich effect can be explained by a pure disparity model that does not require motion-stereo integration; and 2) joint motion-stereo models cannot explain the S shaped functions of perceived disparity versus interocular time delay that are found for stroboscopic Pulfrich stimuli (Morgan, 1979). The second paper asserts that; 1) it provides a physiological implementation which can, with additional assumptions, explain all Pulfrich phenomena; and 2) both separate and joint coding models can produce the S curves.

Here, we evaluate these models and compare separate with joint coding assumptions. We find that there are major flaws in the recent proposals and that the joint encoding idea is clearly advantageous. We show that the recent studies (Read & Cumming, 2005a, 2005c) used an inappropriate definition of joint coding and never really implemented joint coding. More importantly, the separate coding models in the recent studies (Read & Cumming, 2005a, 2005c) are either non-causal or non-physiological or both, and cannot explain the most basic aspect of Pulfrich effects, namely the direction-depth contingency. Finally, under physiologically plausible assumptions, joint coding models, but not separate coding models, can explain the S curves.

Methods

We simulated the predictions of joint-coding and separate-coding models to stroboscopic Pulfrich stimuli. The stroboscopic Pulfrich stimuli consisted of a single dot undergoing apparent motion. The speed of the apparent motion was 4°/sec. In our computer simulations, we let each pixel in space and time dimensions represent 0.02° and 5 ms, respectively, so that the speed was 1 space pixel/1 time pixel. We varied the interflash time interval (T) and interflash distance (X) in proportion to keep the speed constant. For each T , we varied the interocular time delay from 0 to T .

We used the spatiotemporal version of the disparity energy model (Ohzawa, DeAngelis, & Freeman, 1990; Qian, 1994) and considered the horizontal spatial dimension (x) and the temporal dimension (t) in all our simulations. The x dimension of the left and right

receptive fields of a binocular simple cell was modeled as Gabor functions (Ohzawa et al., 1990):

$$g_l(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos(\omega_x x + \phi_l), \quad (1)$$

$$g_r(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos(\omega_x x + \phi_r), \quad (2)$$

where σ determines the receptive field size, ω_x is the preferred horizontal spatial frequency, and ϕ_l and ϕ_r are the phase parameters that determine the shifts of the ON/OFF subregions within the Gaussian envelope. For our simulations, we let $\sigma = 0.32^\circ$ (16 pixels), $\omega_x/2\pi = 0.031$ cycles/deg (32 pixels/cycle), and we uniformly sampled 8 values of $(\phi_l - \phi_r)$ in $[-\pi, \pi)$ to cover the full range of preferred disparities under a given ω_x (Qian, 1994). The values for the ϕ_l and ϕ_r pairs were the same as those in Qian (1994).

For the temporal impulse responses (temporal kernels), we used the gamma-cosine function of Chen, Wang, and Qian (2001):

$$h(t) = \begin{cases} \frac{1}{\Gamma(\alpha)\tau^\alpha} t^{\alpha-1} \exp\left(-\frac{t}{\tau}\right) \cos(\omega_t t + \phi_t) & t \geq 0 \\ 0 & t < 0 \end{cases}, \quad (3)$$

where τ is the time constant for the gamma envelope and α determines its degree of skewness. The cosine term with frequency ω_t generates multiphasic temporal kernels and the phase ϕ_t can be adjusted to allow the first and second half cycles of the kernels to have different durations as illustrated in Figure 1. The kernels are zero for negative t to ensure causality. With appropriate choice of parameters, this function can closely mimic the multiphase temporal kernels of real visual cells. The green curve in Figure 1 was obtained with $\alpha = 2.5$, $\tau = 22.5$ ms (4.5 pixels), $\omega_t/2\pi = 8.3$ cycles/sec (24 pixels/cycle), and $\phi_t = -0.2\pi$. The solid red curve in the figure was obtained with the same parameters but the cosine was replaced by sine in the equation. This is very close to the Hilbert transform (dashed red curve) of the green curve so that the green and solid red curves together with the spatial Gabor filters are well suited for constructing spatiotemporally oriented (directionally selective) filters in Equations 4 and 5 below (Adelson & Bergen, 1985; Watson & Ahumada, 1985). The gamma envelope is shown as the blue dotted curve. To generate a range of speed preferences for our simulations, we scaled the green and solid red curves (without changing their shapes) by dividing τ and multiplying ω_t by a factor taken from the

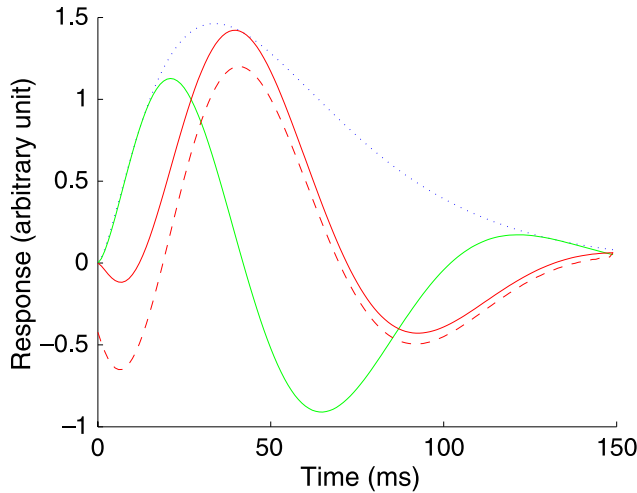


Figure 1. The gamma-cosine temporal kernels used in our simulations. The green curve was Equation 3 plotted with $\alpha = 2.5$, $\tau = 22.5$ ms, $\omega_t/2\pi = 8.3$ cycles/sec, and $\varphi_t = -0.2\pi$. The solid red curve was obtained from the green curve by replacing cosine by sine in Equation 3; it is very similar to the Hilbert transform of the green curve (red dashed). The gamma envelope is indicated by the dotted blue curve.

list (0.67, 1.0, 1.5, 2.25): these numbers form a geometric sequence with a common ratio of 1.5.

The spatial receptive fields and the temporal kernels were then combined to generate binocular spatiotemporal receptive fields (Chen et al., 2001):

$$f_l(x, t) = g_l(x)h(t) + \eta \bar{g}_l(x)\bar{h}(t), \quad (4)$$

$$f_r(x, t) = g_r(x)h(t) + \eta \bar{g}_r(x)\bar{h}(t), \quad (5)$$

where \bar{g} and \bar{h} were obtained from the corresponding g and h functions by replacing all the cosine terms by the sine terms. The weighting factor η determines directional selectivity of the cells, with $\eta = 0$ for spatiotemporally separable receptive fields (bidirectional) and $\eta = \pm 1$ for spatiotemporally oriented receptive fields (unidirectional). For $|\eta|$ between 0 and 1, intermediate levels of directional selectivity can be created. For simulations in this paper, we use $\eta = 0, \pm 1$ to consider both unidirectional and bidirectional tuning. For each η , we generate four different speed preferences by scaling the same temporal kernels by the four factors noted above.

Although bidirectional cells are often viewed as motion insensitive, these cells do carry motion information although less than that of unidirectional cells. Unlike roughly isotropic cells in LGN or retina, bidirectional cells have a preferred motion axis (but not direction). Even when only the horizontal spatial dimension is considered, bidirectional cells have a preferred speed

range that depends on the preferred spatial and temporal frequencies. Regardless of whether the motion preference is unidirectional or bidirectional, Equations 4 and 5 combine selectivity for motion and disparity.

For our joint-coding model, at each spatial location, a set of complex cells covers ranges of disparity and motion preferences combinatorially. We always include 8 different disparity preferences. For the unidirectional case ($\eta = \pm 1$), we also have 8 motion preferences (4 speed preferences for each of the two opposite directions). Therefore, at each stimulus location there are a total of $8 \times 8 = 64$ complex cells jointly tuned to disparity and motion. For the bidirectional case ($\eta = 0$), the two opposite directions are combined yielding a total of 32 complex cells at each location. The responses of a complex cell are obtained from a quadrature pair of simple cells according to the standard disparity energy model (Ohzawa et al., 1990). The simple cell responses are obtained through spatial correlation and temporal convolution between the cell's binocular spatiotemporal receptive fields and the stimulus (Qian & Andersen, 1997).

Cells with dissimilar temporal kernels have different response time courses. To determine the equivalent disparity at each spatial location, we first integrate temporal responses at that location. Since the temporal responses last for only about 200 ms after the stimulation, it is sufficient to just integrate over this local time window. For disparity estimation, we also pool responses across different motion preferences. We then locate the peak along the disparity dimension and use it to represent the perceived equivalent disparity (Qian, 1994; Qian & Andersen, 1997). As the stimulus is flashed at successive locations, the equivalent disparity quickly (<200 ms) builds up to a steady value which is used for the plots presented here.

To evaluate the claim that a stroboscopic Pulfrich effect can be explained by a pure disparity model without joint motion-disparity coding (Read & Cumming, 2005c), we also consider separate coding with a population of cells that cover the same range of disparity as the joint-coding model above but all have identical motion selectivity chosen from one of the 8 motion preferences of the joint-coding model. By having exactly the same motion preference, these cells can code disparity but not motion. The recent separate-coding model (Read & Cumming, 2005a) used a population of cells covering a range of preferred disparity but all tuned to zero temporal frequency (and thus zero speed) because of the monophasic temporal kernel. It can be viewed as a special case of our definition.

To demonstrate that our conclusions do not depend on a specific form of the temporal kernels, we also use Gabor functions as temporal kernels in our simulations. The Gabor function is the same as Equation 1 except that space is replaced by time. We first let $\sigma = 40$ ms, $\omega/2\pi = 6.3$ cycles/sec, and $\varphi = 0$ to generate a temporal kernel and its sine counterpart. We then scale these kernels by the

same four factors as above to create four different speed preferences for each direction of motion. For each σ , we let the total kernel duration be 5σ . We then let the leftmost point of the kernels represent $t = 0$ so that the kernels are zero for negative t to ensure causality.

Results

We first examine in detail the recent conceptual and computational models for the Pulfrich phenomena (Read & Cumming, 2005a, 2005c). A main focus is to determine if these models are physiologically and logically plausible. We then simulate joint and separate coding models under physiologically plausible assumptions, to see if one or both models can explain the S curves in the stroboscopic Pulfrich effect (Morgan, 1979).

Evaluation of recent models

Figure 2 shows graphically the recent conceptual model (Read & Cumming, 2005c). Panel A is a standard spatiotemporal representation of a stroboscopic Pulfrich stimulus. The dots seen by the left and right eyes are represented by red and blue colors, respectively. In the model, the appearance of a given dot in the right eye is considered (the middle blue dot in Figure 2A). All matches between this one dot and all the left-eye dots (red dots) presented at all times are then determined. Three such matches are shown in Figure 2A; they are indicated by the solid green, brown, and pink arrows, and have positive, zero, and negative disparities, respectively. It is then assumed that the perceived disparity is the weighted average of all the disparities from all the matches between the single blue dot and all the red dots. The weighting factors are read from a Gaussian function of the time separation between the two dots in a match (Figure 2B). The assertion is made that when the width of the Gaussian is determined by physiological data, this conceptual model provides a parameter-free fit of psychophysical data on the stroboscopic Pulfrich effect (Read & Cumming, 2005c). Mathematically, the perceived disparity is assumed to be:

$$d = \frac{\sum_{j=-\infty}^{\infty} jX \cdot w(jT + \Delta t)}{\sum_{j=-\infty}^{\infty} w(jT + \Delta t)} \quad (6)$$

where X and T are the spatial and temporal step sizes of the apparent motion, Δt is the interocular time delay, jX is the disparity between the middle blue dot and the j th red dot, and $w(jT + \Delta t)$ is the weighting factor as a Gaussian

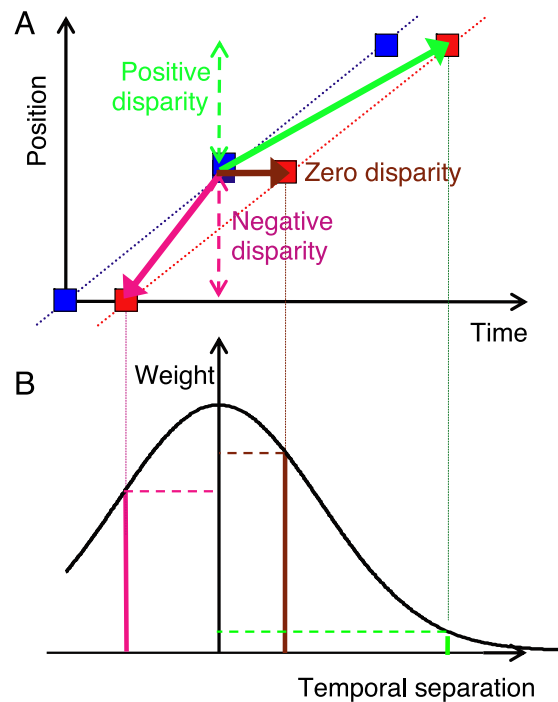


Figure 2. Conceptual model (Read & Cumming, 2005c). (A) The spatiotemporal representation of a stroboscopic Pulfrich stimulus. The red and blue squares represent dots presented to the left and right eyes, respectively. The arrows indicate matches between the two eyes. (B) The Gaussian weighting function for the matches. Copied from Read and Cumming (2005c) with permission from ARVO.

function of the temporal separation between the two dots in the match.

Causality analysis of the conceptual model

A main problem with this conceptual model is that it is non-causal and thus cannot be realized in any real system. This problem can be readily seen in Equation 6 which expresses the perceived disparity as the weighted average of all disparities between a single dot in one eye and all dots occurring at ALL times in the other eye. The summations in the equation go from infinite past to infinite future. However, if the current time is t , then the dots presented after time t have not yet appeared, and the disparities of the matches involving any future dots should not be included in the summation to estimate the currently perceived disparity.

Causal versions of the conceptual model

The conceptual model picks a single dot in one eye (the middle blue dot in Figure 2) and matches it to all the dots

in the other eye. The first dot picked is termed the special dot below. We consider two causal versions of the conceptual model. 1) The current time is aligned with the special dot, and 2) the current time leads the special dot. If current time lags the special dot, it has not appeared, so we do not consider this case.

The first causal version of the model is illustrated in Figure 3. Assume that the middle blue dot (the special dot) appears, and that the current time is marked by the vertical dashed line going through this dot (Figure 3A). The dots on the right of the current time have not yet appeared and these future dots are represented by open squares. To determine the currently perceived disparity, one should only look for the matching red dots on the left side of the current time. One such match is indicated by the pink arrow in Figure 3A. Since all the matches have a negative disparity (left eye positions minus the right eye position are all negative), the perceived disparity predicted by the model is negative.

A moment later, the middle red dot appears as shown in Figure 3B, and is the new special dot that aligns with the new current time. Again, one should only look for matching blue dots on the left side of the new current time. Two such matches are indicated by the brown and green arrows in Figure 3B. Since all the matches have either a zero or a positive disparity (left eye positions minus the right eye position are zero or positive), the perceived disparity predicted by the model is positive. Therefore, this causal version of the conceptual model incorrectly predicts periodically alternating disparities as each new dot appears even when the apparent motion of the stroboscopic stimulus has a constant velocity.

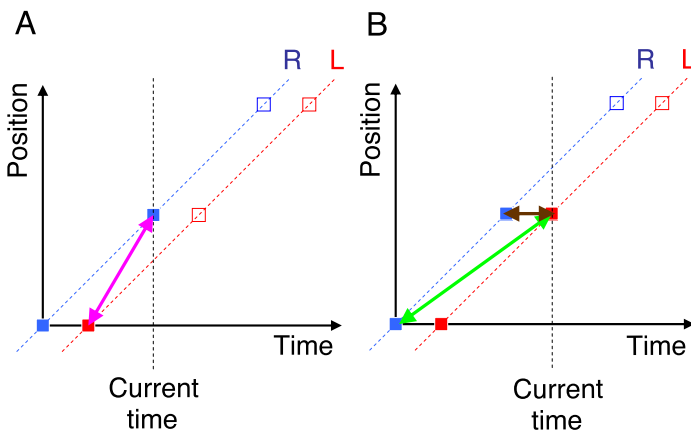


Figure 3. A causal version of the conceptual model (Read & Cumming, 2005c). It predicts oscillating disparities as new dots appear. (A) The current time aligns with the second blue dot. (B) A moment later, the current time aligns with the second red dot which just appears. Dots that have and have not appeared at the current time are represented by solid and open squares, respectively.

Can the above problem be avoided by assuming that the current time does not align with the special dot but instead leads the special dot by an amount S ? This second causal version of the model is shown in Figure 4A. If S is much larger than the time constant τ for determining the Gaussian weighting factors in Equation 6, then contributions of the future matches are negligible. Thus, summation to the infinite future in the conceptual model (Equation 6) can be replaced by summation up to the current time, and all past dots' positions and times can be assumed to be stored in memory and are available for disparity calculations. Unfortunately, when this solution to the causality problem is applied, it immediately creates a new problem. The model only considers matches involving a special dot (the middle blue dot in Figure 4A). Other dots seen by the same eye as the special dot (other blue dots in Figure 4A) are ignored. The new problem occurs when the current time is not aligned with the special dot: The special dot is just one of many that appeared in the past and there should not be anything special about it. Therefore, one has to include all the previously ignored matches (Figure 4B).

The question then is how these additional matches should be weighted together with the previous ones. For example, the three brown arrows in Figure 4B all have the same time separation between left and right eyes (S_{LR}) but their separations from the current time (S_{CT}) are very different. It certainly does not make sense to weight them equally regardless of how large S_{CT} is. The weighting factor in Equation 6 of the conceptual model proposal (Read & Cumming, 2005c) is only a function of S_{LR} . The effect of S_{CT} is not considered.

The weighting function used in the conceptual model

The conceptual model (Read & Cumming, 2005c) uses a Gaussian function of $S_{LR} = \Delta t$ to weight different matches (Figure 2B). We now examine the validity of this choice.

Cogan, Lomakin, and Rossi (1993) and others (see Howard and Rogers, 2002 for a review) have studied how perceived disparity degrades with interocular time separation Δt . If the two eyes' patterns have the same contrast polarity, one does not see depth for $\Delta t > 50$ ms. Interestingly, for larger values of Δt , depth perception can be restored when the two eyes' patterns have opposite polarities (Cogan et al., 1993). This observation may be explained computationally (Grunewald & Grossberg, 1998) by use of typical *biphasic* temporal kernels of visual cells (DeAngelis, Ohzawa, & Freeman, 1993a, 1993b; Hawken, Shapley, & Groszof, 1996). A flash of light in a V1 cell's ON region, for example, generates an initial excitatory response, followed by a longer inhibitory response (cf. the green and solid red curves in Figure 1). The full temporal response lasts for about 100 to 200 ms. When the two retinal images have the same contrast

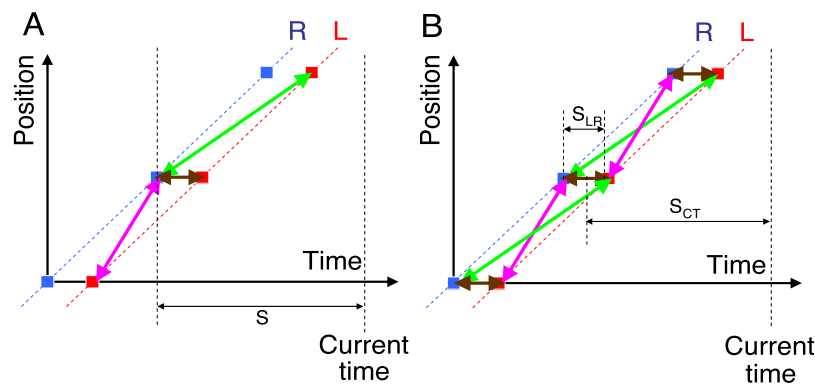


Figure 4. Another causal version of the conceptual model (Read & Cumming, 2005c). It misses many matches (A), and if they are included (B), it is not clear how they should be weighted. It could be according to the time separation between the two dots in a match between the eyes (S_{LR}) or the time separation between the match and the current time (S_{CT}) or both.

polarity, Δt has to be less than 50 ms to allow an overlap between the same-signed responses evoked through the two eyes and thus enable stereo matching. If Δt is greater than 50 ms, there is only an interocular overlap between the opposite-signed responses and stereovision fails. When the two retinal images have opposite polarities and Δt is greater than 50 ms, the overlapping responses evoked through the two eyes have the same sign again, and stereovision is restored.

This set of consistent psychophysical, physiological and computational studies strongly indicates that the weighting factor in the conceptual model (Read & Cumming, 2005c) as a function of interocular time separation (Δt) should drop to zero at about $\Delta t = 50$ ms and then reverse its sign. Obviously, the model will not work with such a dramatic change of the weighting function (Read & Cumming, 2005a). Even if we assume that the weighting function stays at zero for $\Delta t > 50$ ms, the model will still not work. For example, if the sum of the interflash interval T and the interocular delay Δt exceeds 50 ms, the match indicated by the green arrow in Figure 2 will have zero weight.

The most direct prediction of the conceptual model

The main claim of the model is that it does not depend on motion processing (Read & Cumming, 2005c). Therefore, it is most straightforward to probe the model using stimuli without motion. Consider a single dot (or other pattern) flashed to both the left and right eyes. Assume dichoptic spatial and temporal offsets of D and Δt , respectively. Since there is only one match between the eyes, Equation 6 reduces to:

$$d = \frac{D \cdot w(\Delta t)}{w(\Delta t)} = D. \quad (7)$$

In other words, regardless of the magnitudes of Δt and D , the predicted disparity is always D . This prediction is clearly incorrect because as we mentioned above, the perceived disparity should decrease to zero when Δt increases to 50 ms. Note that the conceptual model critically depends on this prediction. If the model is revised to eliminate this incorrect prediction, it will no longer explain the perceived disparity in stroboscopic Pulfrich effect.

Temporal kernel in the computational implementation of the conceptual model

In the second paper, a computational model is proposed to implement the conceptual one by use of binocular receptive fields for cells in primary visual cortex (Read & Cumming, 2005a). However, as the authors note, the model does not work when the cells in the model have a biphasic temporal kernel as is typically found in real V1 neurons (DeAngelis et al., 1993a, 1993b; Hawken et al., 1996). The model requires a mono-phasic temporal kernel. In other words, the model V1 cells have to be low-pass temporal filters which prefer zero temporal frequency and zero speed. This is contradicted by both physiological and psychophysical evidence as discussed above.

An attempt to justify the non-physiological assumption is made by stating that according to the energy model, “band-pass temporal kernels would generate a biphasic response to interocular delay, yet this is not observed in the responses of V1 neurons” (Read & Cumming, 2005a). However, multi-phasic responses to interocular delay have been observed so there is no problem with the energy model in this aspect (Anzai et al., 2001). One study failed to show multi-phasic responses to interocular delay (Read & Cumming, 2005b) but in that case, a much narrower range of interocular delay was used which limits the ability

to address this point. In any case, a non-physiological assumption is not justified even if it is made to accommodate a specific model.

The definition of joint motion-disparity coding

We have defined joint motion-disparity coding as a representation by cells tuned to a range of motions and a range of disparities combinatorially (see [Methods](#)) because these cells can indeed encode both motion and disparity (Qian & Andersen, 1997). In contrast, the recent computational study (Read & Cumming, 2005a) implemented a “joint” coding model by using a set of cells that cover a range of disparity preferences but all have *identical* motion tuning (we thank Dr. Read for clarifying this implementation detail). The problem is that cells with identical motion tuning cannot encode motion just like cells all with identical orientation tuning curves cannot encode orientation and a visual system with only red (L) cones is color blind. To encode a stimulus property, cells preferring a *range* of that property are required. Thus, the recent “joint” motion-disparity coding model (Read & Cumming, 2005a) cannot encode motion and does not qualify as a joint coding model. Since both the “joint” and separate coding models in the recent study (Read & Cumming, 2005a) can only code disparity but not motion, they are really different versions of separate coding according to our definition (see [Methods](#)).

Spatial kernel in the recent computational study

Both the “joint” and separate coding models in the recent computational study (Read & Cumming, 2005a) rely on monophasic temporal kernel tuned to zero temporal frequency. Since cells’ preferred speed approximately equals the preferred temporal to spatial frequency ratio, the only way to have a non-zero preferred speed in the “joint” model is to set the cells’ preferred spatial frequency to zero as well. This is indeed the case: a low-pass Gaussian function, instead of a band-pass Gabor function, was used to implement spatial receptive fields (see Equation 4 and Figure 3B of Read and Cumming, 2005a). This constitutes another non-physiological aspect of the study.

Examination of the separate-coding explanation of dynamic-noise Pulfrich effects

An interocular time delay can make a dynamic noise pattern appear as a cylindrical volume rotating in depth (Falk & Williams, 1980; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Ross, 1974; Tyler, 1974). When the left eye image lags behind that of the right eye, dots in the near half of the cylinder appear to move from right to left. When the time delay is reversed, the perceived direction of rotation is also reversed. Since

there is no net physical motion in dynamic noise patterns, the dynamic-noise Pulfrich effect constitutes strong evidence for a mechanism of joint motion-disparity coding. A model for this dynamic-noise Pulfrich effect, using joint coding, has been elaborated previously (Qian & Andersen, 1997).

It has been pointed out that a dynamic-noise Pulfrich stimulus contains a correlation between random motion and disparity signals (Tyler, 1974). This correlation can be encoded by a joint motion-disparity process to explain the dynamic-noise Pulfrich effect (Qian & Andersen, 1997). However, in the computational implementation of separate coding (Read & Cumming, 2005a), it is stated that joint coding is not necessary to account for the dynamic-noise Pulfrich effect. A key new assumption is that the correlation between the firing rates of *separate* motion and disparity cells is sufficient to account for the dynamic-noise Pulfrich effect (Read & Cumming, 2005a). But how is this correlation computed? In the model (Read & Cumming, 2005a), this correlation is computed artificially and not by the separate motion and disparity cells. If the brain computes this correlation, then the cells involved must link motion and disparity responses in some way (Spang & Morgan, 2008). Although, a priori, the linkage could take different forms, the presence of jointly tuned cells in V1 and MT makes it unnecessary to propose other mechanisms.

It should be noted that the correlation between two firing rates is very different from timing correlation or synchronization between individual spikes from different cells (see below). Timing correlation between spike trains implies an underlying anatomical substrate that could form the physical basis of encoding. The firing-rate correlation is simply a reflection of the same correlation in the raw stimulus.

Therefore, the assertion that separate coding accounts for the dynamic-noise Pulfrich effect (Read & Cumming, 2005a) amounts to an assumption that the brain does not have to actually compute the motion-disparity correlation. This is analogous to an assertion that all vision problems are solved by photoreceptors and the rest of the brain is superfluous. For example, one could state that orientation detection is accomplished by photoreceptors because when an oriented bar is presented, there is correlation among the responses of the specific set of photoreceptors covered by the bar. Likewise, one could assert that binocular disparity is detected by photoreceptors because when a stimulus of disparity d excites photoreceptors at x on the left retina and at $x + d$ on the right retina, there is correlation between the responses of these two sets of photoreceptors. The logical flaw here is that although one can compute these correlations artificially, the photoreceptors cannot. All the information the visual system receives, including various correlations across space, time, and eyes, is already present at the retina. The presence of a correlation provides an opportunity for encoding but that is not the same as actual encoding.

Spike-timing synchronization and the separate-coding model

We note above that the presence of jointly tuned cells in V1 and MT makes it unnecessary to propose other mechanisms to encode motion-disparity correlation for Pulfrich stimuli. Nevertheless, one may ask whether it is possible, in principle, to support a separate-coding idea by use of spike synchronization between motion and disparity cells. First, an increase of firing rates of two cells tends to cause a spurious increase of spike synchronization: spikes from two cells are more likely to occur together *by chance* when there are more spikes. For this reason, studies on synchronization remove spurious synchronization via shuffle correction or randomization between data and experimental conditions (Castelo-Branco, Neuenschwander, & Singer, 1998; Fries, Womelsdorf, Oostenveld, & Desimone, 2008). If spurious synchronization were allowed as a coding mechanism, one would end up with a logical fallacy similar to the one discussed above, namely that all vision problems are solved by retinal ganglion cells.

After the deduction of spurious synchronization, the resulting real spike synchronization has an anatomical basis and could represent coding. Synchronization between disparity and motion cells can occur when they receive inputs (excitatory or inhibitory) from a common set of cells, or from each other (directly or indirectly). In the case of common inputs, the cells providing the inputs should already be selectively responsive to the motion-disparity correlation in the Pulfrich stimuli; otherwise, the synchronization they produce between the downstream motion and disparity responses is unrelated to the Pulfrich effects. In the case of mutual inputs, disparity cells have motion driven spikes and/or motion cells have disparity driven spikes or both. Therefore, in both cases, motion and disparity responses have to be linked to code motion-disparity correlation. Incorporation of spike synchronization does not increase the likelihood of a separate-coding model.

The motivation of separate-coding models

Most monkey V1 cells are tuned to disparity but only a small fraction of them are tuned to direction of motion (Hubel & Wiesel, 1968; Poggio & Fischer, 1977). It is thus not surprising that only a small fraction of monkey V1 cells are jointly tuned to both disparity and motion (Grunewald & Skoumbourdis, 2004). It has been argued that if the Pulfrich phenomena require joint motion-disparity coding, then it is “puzzling” why the majority of non-joint-coding V1 cells do not contribute (Read & Cumming, 2005c). Thus the main motivation to propose a separate-coding model is to include most V1 cells in the Pulfrich phenomena (Read & Cumming, 2005c). However, with this reasoning, one could also question why retinal and LGN cells do not represent orientation and

propose a model for orientation detection based on correlation among firing rates of non-oriented retinal or LGN cells. This line of argument is not compelling if one assumes that the brain simply uses the most relevant information to solve a given problem. Thus, orientation detection is best done in V1, and LGN cells contribute by providing organized inputs to V1. Likewise, for monkeys, the Pulfrich phenomena are best explained by MT where most cells are jointly tuned to motion and disparity (Bradley et al., 1995; Pack et al., 2003), and V1 contributes by providing inputs to MT. For cats, many V1 cells are already jointly tuned to motion and disparity and this can explain the Pulfrich phenomena (Anzai et al., 2001).

The relationship between conceptual and computational models of separate coding

A mathematical proof is presented of the equivalence between the conceptual model and its computational implementation (Read & Cumming, 2005a). However, the conceptual model is non-causal while the computational model uses causal temporal filters. Therefore, there cannot be an exact mathematical equivalence between the two models in general. A close examination reveals that the proof depends on the assumption that the energy responses be integrated over exactly one interflash interval T (see Equation 12 of Read and Cumming, 2005a), or equivalently an integer multiple of T . A key step of the proof, Equation A3 of Read and Cumming (2005a), fails without this assumption. However, the assumption is non-physiological because it implies that neural integration time has to equal to arbitrary interflash intervals of stroboscopic Pulfrich stimuli. If the assumption is replaced by a generally applicable and physiologically plausible one, such as integration over a 200 ms window of V1 responses (as we did in our simulations; see [Methods](#)), the proof fails and it is not clear whether the simulations in that study (Read & Cumming, 2005a) can still work. A related ambiguity with the study (Read & Cumming, 2005a) is that although it claims to use the disparity energy model (Ohzawa et al., 1990), there is no indication of using the standard quadrature pair construction, or its equivalent (Qian & Mikaelian, 2000), in either the proof or the simulations.

Simulations of S curves for a stroboscopic Pulfrich effect

For stroboscopic Pulfrich stimuli with relatively large interflash time intervals, the perceived disparity as a function of interocular time delay has an S shape (Morgan, 1979). It is asserted that joint coding models cannot explain these S curves while separate coding

notions can (Read & Cumming, 2005c). This claim is later retracted implicitly by the observation that under the identical set of assumptions made, the joint-coding and separate-coding models make similar predictions (Read & Cumming, 2005a). However, as detailed above, these assertions are based on models with serious problems. In particular, the “joint” motion-disparity coding model (Read & Cumming, 2005a) cannot encode motion and is actually a version of separate coding model. Furthermore, these separate-coding models do not work when realistic multiphasic temporal kernels are used. If non-physiological assumptions are excluded, can any model explain the S curves? We have performed extensive computer simulations to address this question.

The joint-coding model can explain the S curves

We considered stroboscopic Pulfrich stimuli containing a dot undergoing apparent motion. We constructed a set of simple and complex cells with binocular spatiotemporal receptive fields, and computed responses to a stimulus according to the disparity energy model (see [Methods](#)). At each location, the set of cells covers a range of disparity and motion preferences combinatorially (Qian & Andersen, 1997). The range of disparity preference is obtained by sampling 8 phase differences between the left and right eyes’ receptive fields over the full 2π range. The range of motion preference is obtained by scaling the same temporal kernels by four different factors so that there are four different preferred temporal frequencies and thus four different preferred speed ranges. In addition, we include preferences to opposite directions of motion so that there are 8 motion preferences (2 directions each with 4 speed ranges). Overall, there are a total of $8 \times 8 = 64$ complex cells at each location to jointly code disparity and motion. The responses of the underlying simple cells at all locations as a function of time were obtained via spatial correlation and temporal convolution (Qian & Andersen, 1997), and the responses of the complex cells were computed via the energy method (Ohzawa et al., 1990). Cells with different temporal kernels have different response time courses. To determine the equivalent disparity at each spatial location, we first integrate temporal responses of each complex cell at that location over the past 200 ms. For disparity estimation, we also pool responses across different motion preferences. We then locate the peak along the disparity dimension and used it to represent the perceived equivalent disparity in exactly the same way as for previous models (Qian, 1994; Qian & Andersen, 1997). As the stimulus is flashed at successive locations, the equivalent disparity quickly (<200 ms) builds up to a steady state value which is used in our plots.

An example of our simulations using joint-coding is shown in [Figure 5](#). The equivalent disparity (d) as a fraction of the interflash distance (X) is plotted against

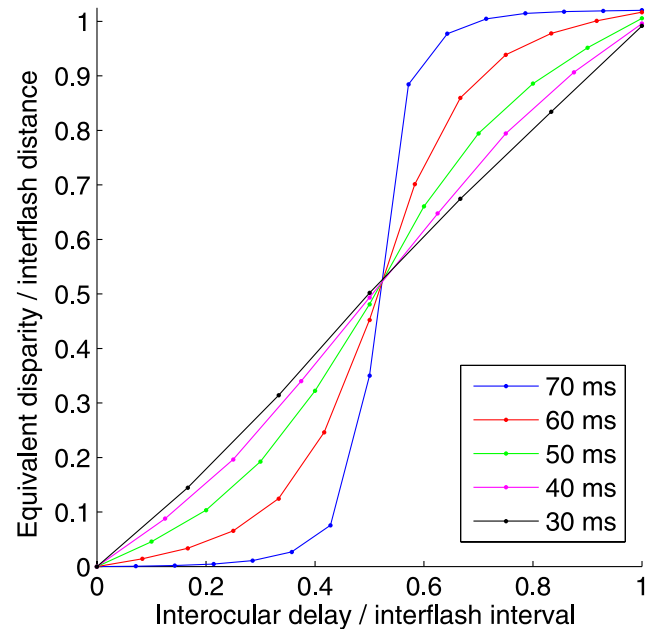


Figure 5. Simulations of the S curves in stroboscopic Pulfrich effect using the joint-coding model and the gamma-cosine temporal kernels. The equivalent disparity as a fraction of the interflash distance is plotted as a function of interocular time delay as a fraction of the interflash time interval. The five curves are for interflash intervals of 30, 40, 50, 60, and 70 ms, respectively, as labeled.

interocular time delay (Δt) as a fraction of the interflash time interval (T) for several T 's. To understand what is presented here, first note that for continuously moving stimuli, $d = (X/T) \Delta t$, or $d/X = \Delta t/T$; this corresponds to the diagonal line in the figure. When T is 30 ms (or smaller), the results follow the diagonal line indicating that the equivalent disparities are equal to the values for a stimulus moving continuously. When T is larger than 30 ms, the model reproduces the S curves (Morgan, 1979).

The essence of the S curves is that when T is relatively large and Δt is between 0 and $T/2$, the equivalent disparity is smaller than expected from the continuous motion case. To understand the reason that the joint-coding model can explain this finding, first consider the case when T is much smaller than cells’ temporal response durations so that the apparent motion of the stimulus is almost as strong as the continuous motion case. Here, the cells whose motion preference matches the stimulus motion will respond far more vigorously than other cells. In this case, pooling across different motion preferences can be well approximated by use only of cells tuned to the stimulus motion. This simplification is used in the original Pulfrich model (Qian & Andersen, 1997). The computed equivalent disparity is equal to that of the continuous motion case. When T is comparable to the cells’ temporal response durations, the apparent motion is very weak and the Pulfrich effect will disappear. In this case, cells with

difference motion preferences will respond equally, and since different motion preferences predict different equivalent disparities (including different signs), the pooled response predicts a near zero equivalent disparity (Qian & Andersen, 1997). For intermediate T values, the equivalent disparity is not zero but smaller than that for the continuous motion case, as shown in Figure 5.

To explain further, consider the 5th point along the blue curve (interflash interval of 70 ms) in Figure 5; the ratio of interocular delay to interflash interval for the point is 0.36. We show in Figure 6A the time-averaged, normalized

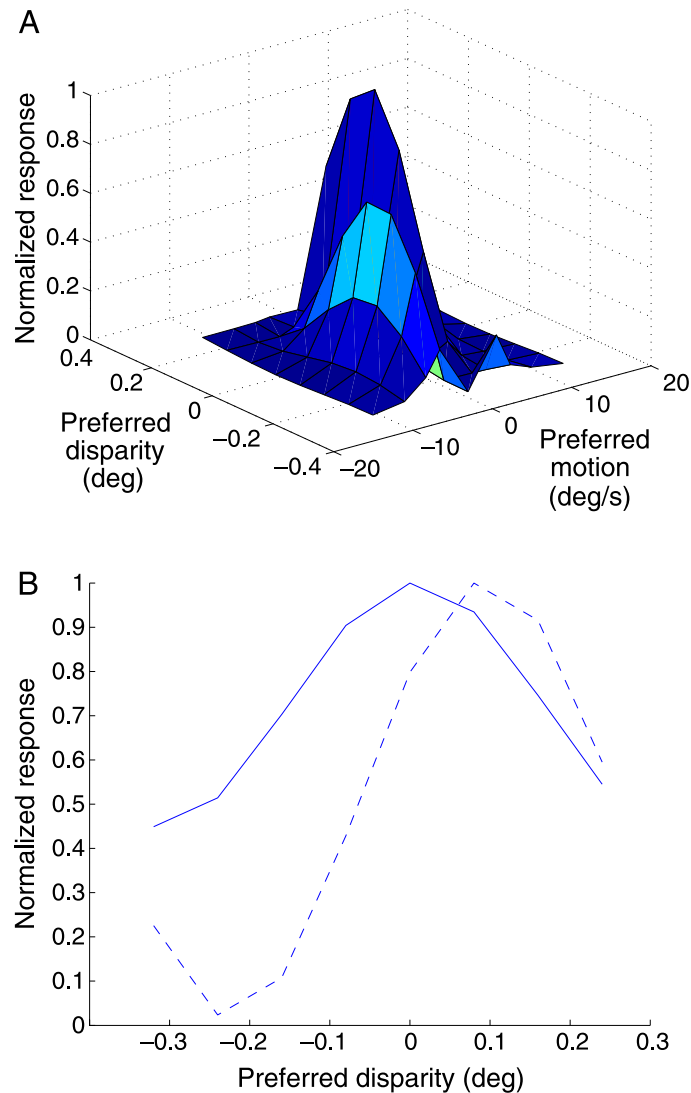


Figure 6. Population response of the model cells. The results are for the 5th point along the blue curve in Figure 5 whose interocular delay to interflash interval ratio is 0.36. (A) The time-averaged, normalized response at a fixed spatial location as a function of the preferred disparity and preferred motion of all cells for that location. (B) The response as a function of the preferred disparity if the best-matched motion preference is considered (dashed curve) or if the different motion preferences are pooled (solid).

response at a fixed spatial location as a function of the preferred disparity and preferred motion of all cells for that location. Notice that for different motion preferences, the peak locations along the disparity dimension are different. This is because different motion preferences convert a given interocular time delay into different equivalent disparities (Qian & Andersen, 1997). If we use the cells with the best-matched motion preference, we get the dashed disparity curve in Figure 6B. On the other hand, if we pool across all motion preferences (as we did in Figure 5), we get the solid curve in Figure 6B. The disparity estimated from the solid curve is much smaller than that from the dashed curve. Thus, when the interflash interval is large, motion pooling reduces the equivalent disparity, leading to the S curve in Figure 5.

When Δt is larger than $T/2$, the S curves of Figure 5 are above the diagonal line. This is because the n th dot in the delayed eye image is temporally closer to the $(n + 1)$ st dot than to the n th dot in the other eye so that the retinal disparity of this match is equal to the interflash distance instead of 0.

To make sure that the S curves are a general property of the joint-coding model and do not depend on the specific form of the multiphasic temporal kernels, we also did simulations using a Gabor function as temporal kernels. Like the simulations with the gamma-cosine kernel noted above, we let the leftmost point of the temporal kernel represent zero time to ensure causality, scale the same temporal kernels to generate four preferred speed ranges, and include two opposite preferred directions. An example of our simulations is shown in Figure 7. The results are similar to those in Figure 5, showing S curves for large interflash intervals. We have also run simulations with a doubled number of disparity- and motion-tuned cells in the model or differently sampled preferred speed ranges, and again obtain similar results (not shown). We therefore conclude that the S curves are a robust property of the joint-coding model.

Separate-coding models cannot explain the S curves

For comparison, we also simulated the stroboscopic Pulfrich effect under the separate coding assumption. Separate coding of motion and disparity means that there are two separate populations of cells with one population encoding motion but not disparity and the other encoding disparity but not motion. Since the S curves are about disparity perception, we consider separate coding for disparity. We converted our joint-coding model to that of separate coding by use of cells covering the same disparity range as the joint-coding model but all having identical motion preferences (and thus unable to encode motion). We fixed the motion tuning of all cells to each one of the 8 motion preferences in the joint-coding model above. None of the resulting 8 versions of the separate coding model could explain the S curves. Examples are shown in

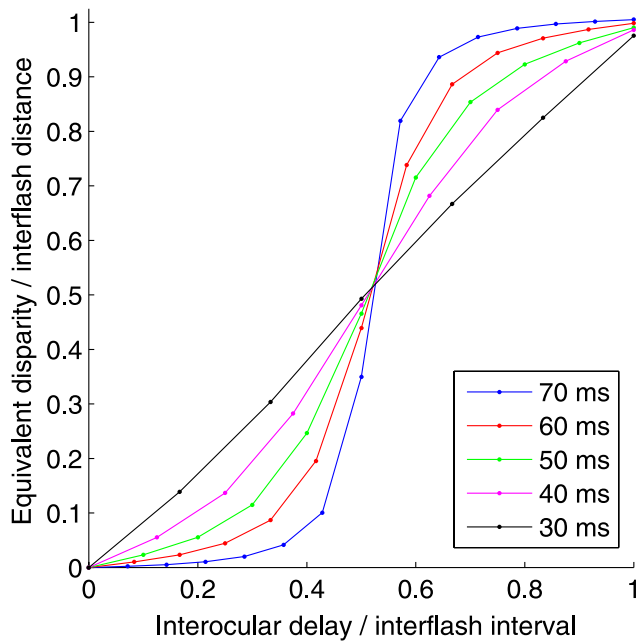


Figure 7. Simulations of the S curves in stroboscopic Pulfrich effect using the joint-coding model and the Gabor temporal kernels. The equivalent disparity as a fraction of the interflash distance is plotted as a function of interocular time delay as a fraction of the interflash time interval. The five curves are for interflash intervals of 30, 40, 50, 60, and 70 ms, respectively, as labeled.

Figure 8. Here we consider an interflash interval of 70 ms. The blue and red curves correspond to the simulation results when all cells prefer leftward and rightward motion, respectively, with identical temporal scaling factor (1.5) for speed preference (see [Methods](#)). Neither curve shows an S shape while the joint coding model predicts a strong S curve (blue curve in [Figure 5](#)). This confirms that a separate coding system does not work with realistic multiphasic temporal kernels (Read & Cumming, 2005a).

We next consider a special version of a separate-coding model (Read & Cumming, 2005a) to determine if it is possible to obtain S curves as in [Figure 5](#). It is identical to the 8 failed versions above except that all cells prefer zero speed (Read & Cumming, 2005a) instead of non-zero speed. To do this, we replace the gamma-cosine function by a gamma envelope for the temporal kernel so that the kernel is monophasic and tuned to zero temporal frequency and speed. We find that this specific version of the separate-coding model can indeed generate S curves like those in [Figure 5](#) (results not shown), but only when the total duration of the monophasic temporal kernel is as short as about 50 ms. This simulation also suggests that the biphasic temporal kernels combined with separate coding are responsible for the peculiar shapes of the curves in [Figure 8](#).

As we emphasize above, most visual cortical cells have a multiphasic temporal kernel (DeAngelis et al., 1993a, 1993b; Hawken et al., 1996). Consequently, they prefer non-zero temporal frequency and speed of motion, and are most easily activated by a time-varying stimulus. However, for the separate-coding model to explain the S curves, one has to assume that visual cortical cells have a monophasic temporal kernel, prefer zero temporal frequency and speed, and are more easily activated by a static than by a time-varying stimulus. The 8 versions of the separate-coding model using multiphasic temporal kernels cannot reproduce the S curves. We conclude that in general, the separate coding model cannot explain the S curves.

Joint coding of disparity and bidirectional motion can also explain the S curves

In our joint-coding model we use cells with uni-directional motion as well as disparity preference. Such cells are found in cat V1 (Anzai et al., 2001), monkey V1 (Grunewald & Skoumbourdis, 2004), and monkey MT (Bradley et al., 1995; Pack et al., 2003). In monkey V1, however, most cells are bidirectional: they prefer a motion axis but respond to the two opposite directions

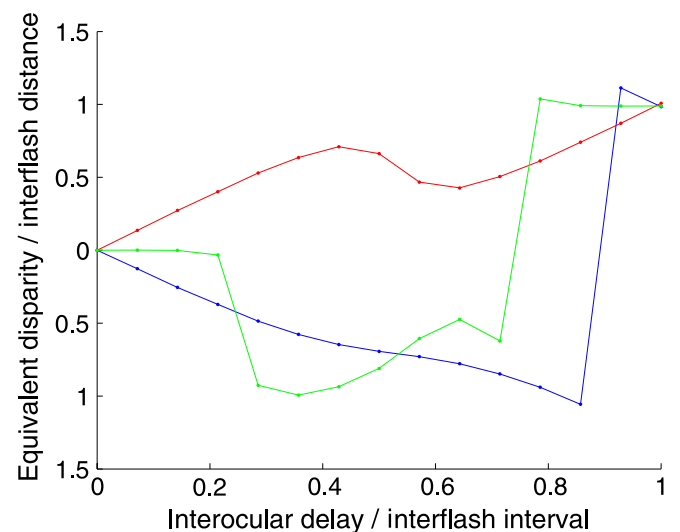


Figure 8. Simulations of the S curves in the stroboscopic Pulfrich effect using the separate-coding model and the gamma-cosine temporal kernels. The equivalent disparity as a fraction of the interflash distance is plotted as a function of interocular time delay as a fraction of the interflash time interval. The interflash interval was fixed at 70 ms. The speed preferences of all cells were identical and corresponded to a temporal kernel scaling factor of 1.5 (see [Methods](#)). The red and blue curves are results from the separate coding models whose cells are all tuned to rightward and leftward directions of motion, respectively. The green curve is from the separate coding model whose cells all have identical bidirectional motion preference.

about equally well. Although traditionally called non-directional, these cells do not respond equally to all directions around the clock, and as a population they do carry information about motion axis and speed. Since the S curves do not involve perception of a specific motion direction, we naturally have constructed a joint-coding model with bidirectional motion preferences. It is identical to the joint-coding model above except for the motion part. Each cell has spatiotemporally separable receptive fields ($\eta = 0$ in Equations 4 and 5) and the 8 motion preferences used above are collapsed into 4 motion preferences because each cell treats the two opposite directions equally.

The simulation results with bidirectional motion preferences are similar to those in Figure 5 with unidirectional characteristics. This is easy to understand intuitively. A bidirectional (spatiotemporally separable) cell is equivalent to the sum of two unidirectional (spatiotemporally oriented) cells. Therefore, pooled responses across 4 bidirectional and 8 unidirectional motion preferences are similar. Formally, it can be shown that the derivation for the relationship between an interocular time delay Δt and its equivalent disparity d (Equation 10 in Qian & Andersen, 1997)

$$d \approx \frac{\omega_t}{\omega_x} \Delta t \approx -v \Delta t, \quad (8)$$

for the unidirectional case also holds for the bidirectional condition. The only difference is that a unidirectional cell is approximately tuned to velocity $v = -\omega_t/\omega_x$, whereas a corresponding bidirectional cell is tuned to two opposite velocities ($\pm v$). When the Pulfrich stimulus contains a strong motion signal, then either the v or the $-v$ component of the bidirectional cell will be strongly activated and an interocular time delay Δt will be treated as an equivalent disparity of $v \Delta t$ or $-v \Delta t$. As the motion signal gets weaker, the dominance of one component over the other in the bidirectional cell will become weaker and the equivalent disparity will be between $v \Delta t$ or $-v \Delta t$.

Although joint coding of disparity and *bidirectional* motion can also produce the S curves, we emphasize that Pulfrich phenomena are best explained by joint coding of disparity and *unidirectional* motion. The reason is that for Pulfrich phenomena, the perceived depth and direction of motion are contingent (e.g., when the left eye's view is delayed, dots moving to the left and right have near and far disparity, respectively). Bidirectional cells cannot determine the direction of motion and thus cannot explain the direction-depth contingency while unidirectional cells can (Qian & Andersen, 1997).

Separate coding of disparity and bidirectional motion cannot explain the S curves

We also have considered separate coding of disparity and bidirectional motion. Once again, we let the cells

cover the same range of disparity as in the joint coding model above but all have identical bidirectional motion preference. We again find that the separate coding model is unable to explain the S curves. A simulation result for an interflash interval of 70 ms is shown as the green curve in Figure 8. This means that to explain the S curves, it is not sufficient to just have two opposite directions. The model also has to cover a range of different speed preferences.

Discussion

Summary of main findings

We have examined here the issue of whether joint coding of motion and disparity is needed to explain Pulfrich phenomena. We first critically evaluate recent conceptual and computational models (Read & Cumming, 2005a, 2005c) that code motion and disparity separately. We find that the conceptual model (Read & Cumming, 2005c) is non-causal as it sums matches between a special dot in one eye and all dots in the other eye from the infinite past to the infinite future. We then show that causal versions of the conceptual model have new problems, including oscillating disparities and many missing matches. In addition, the weighting function used for combining different matches is at odds with both physiological and psychophysical data. The most direct prediction of the model under the condition of a single flash in each eye is clearly wrong but the model depends on this prediction. Subsequent computational implementation of the conceptual model (Read & Cumming, 2005a) relies on the non-physiological assumption that temporal kernels of V1 cells are monophasic and thus prefer zero temporal frequency. The implementation fails if realistic multiphasic temporal kernels are used. The implementation further assumes that neural integration time is exactly equal to one interflash interval of the stroboscopic stimuli. The computational study (Read & Cumming, 2005a) also includes a “joint” motion-disparity coding model. However, the cells in the model all have exactly the same motion selectivity and therefore cannot encode motion. This model is really just a version of separate coding model. In addition, the cells in the model have to prefer zero spatial frequency in order to produce a non-zero preferred speed. Finally, we show that with physiologically plausible assumptions, joint coding but not separate coding of motion and disparity can explain S curves for the stroboscopic Pulfrich effect (Morgan, 1979).

Definitions of joint and separate coding

It is worth revisiting the definitions of joint coding and separate coding of motion and disparity. We define joint

coding as a combinatorial representation of a range of disparities and a range of motions (Qian & Andersen, 1997). In our implementation, we use $8 \times 8 = 64$ complex cells at each location to represent combinatorially 8 disparity preferences and 8 motion preferences. The model can pool across different motion preferences to estimate stimulus disparity as we do in this study, or pool across different disparity preferences to estimate stimulus motion. Alternatively, the model may avoid pooling and estimate, for example, stimulus disparity for a specific direction and speed of motion. Whether pooling should be done depends upon the psychophysical condition to which the model is applied. In the S curve measurements (Morgan, 1979) subjects were asked to report only stimulus disparity, but not motion, in a trial. It is reasonable to pool across motion to estimate disparity.

In contrast, separate coding means that disparity and motion estimation are done independently by two separate populations of cells. The separate-coding disparity model contains cells that cover a range of disparity but all have the same (unidirectional or bidirectional) motion preference. These cells can encode disparity but not motion. Similarly, the separate-coding motion model contains cells that cover a range of motion but all have the same disparity preference. These cells can encode motion but not disparity. When there is a real, conventionally defined spatial disparity in the stimulus, the separate-coding disparity model should be able to extract it. However, for stroboscopic Pulfrich stimuli with large interflash intervals, we show here that the separate-coding disparity model fails to explain the S curves when realistic, multiphasic temporal kernels are used. The joint-coding model can be viewed as a collection of several separate-coding disparity models, each with a different motion preference. By pooling across motion, the joint-coding model can explain the S curves.

Note that separate coding and separable filters are very different concepts. As mentioned above, separate coding of motion and disparity means that there are separate populations of cells coding motion and disparity, respectively, and there is no interaction between the populations. However, in both our separate-coding and joint-coding models, we used filters that are separable in motion and disparity, meaning that each cell's response can be expressed as a product of a motion term and a disparity term. As demonstrated previously, this separability of motion and disparity results from assuming identical motion selectivity for the left and right receptive fields of each binocular cell (Chen et al., 2001; Qian, 1994; Qian & Andersen, 1997), and this is supported by experimental data (Grunewald & Skoumbourdis, 2004; Ohzawa, DeAngelis, & Freeman, 1996). In addition, in our models, unidirectional motion tuning is created via filters inseparable in space and time while bidirectional motion tuning is achieved via spatiotemporally separable filters. The recent computational study (Read & Cumming, 2005a) equates spatiotemporally inseparable filters with joint

motion-disparity coding. However, since the cells in the “joint” model all have identical motion preference, the model cannot encode motion and therefore is not a joint motion-disparity coding model.

Unidirectional and bidirectional motion selectivity

Most monkey V1 cells are disparity tuned but only a small fraction is directionally selective (Hubel & Wiesel, 1968; Poggio & Fischer, 1977). A separate coding model is motivated by the idea that all V1 disparity cells should be responsible for Pulfrich phenomena (Read & Cumming, 2005c). Ironically, a separate coding model only works with monophasic temporal kernels with brief total response duration (<50 ms), a property not shared by the majority of V1 cells (DeAngelis et al., 1993a, 1993b; Hawken et al., 1996). Thus, the effort to include most V1 cells ends up excluding them. Note that many of the non-directional V1 cells should really be viewed as bidirectional. These cells have a preferred speed and motion axis but not that of direction along the axis. We show in the current study that joint but not separate coding, between disparity and bidirectional motion, can also explain the S curves in stroboscopic Pulfrich stimuli (Morgan, 1979).

The importance of unidirectional motion selectivity in explaining Pulfrich effects

It is critical to note that the S curves are about perceived depth but not motion direction. Since the percept of a Pulfrich stimulus contains both depth and direction and the perceived depth is contingent on perceived direction, they are best explained by monkey MT cells (and a small fraction of monkey V1 cells) that are tuned to both disparity and *unidirectional* motion. Joint coding between disparity and bidirectional motion in monkey V1 can only extract stimulus disparity and speed but not direction.

Joint coding between disparity and *unidirectional* motion is also needed to explain the dynamic noise Pulfrich effect (Qian & Andersen, 1997). Here the percept is a volume revolving in depth. When the left eye's view is delayed with respect to that of the right eye, the near and far halves of the volume rotate to the left and right, respectively. When the right eye's view is delayed, the directions of motion reverse. Bidirectional motion preferences cannot distinguish between the two opposite directions of motion and thus cannot fully explain the perception. An assertion in the recent computational model (Read & Cumming, 2005a) is that dynamic noise Pulfrich effects can be explained via a correlation between pure disparity and pure motion responses. However, this correlation has to be computed by joint motion-disparity

cells as was done in a previous Pulfrich model (Qian & Andersen, 1997). Otherwise, such correlation is simply a reflection of the same correlation in the stimulus. It does not represent coding in the same sense that correlation among photoreceptors in response to a bar does not represent orientation coding. The joint coding of disparity and unidirectional motion by MT cells and a small fraction of V1 cells naturally represents the correlation between disparity and motion in dynamic noise Pulfrich stimuli and explains the perception (Qian & Andersen, 1997).

Overview of key models for Pulfrich effects

Finally, we summarize key models for Pulfrich effects. Although this summary repeats some of the discussions above, we hope that collecting relevant information in one place helps clarify the literature. An early integrated motion-stereo model (Qian & Andersen, 1997) used cells covering a range of disparity preferences and a range of motion preferences combinatorially. This model can encode both motion and disparity in stimuli, and has since been called the joint motion-disparity coding model. The model used multi-phasic temporal kernels and included only unidirectional motion preferences (spatiotemporally inseparable filters) but not bidirectional motion preferences (spatiotemporally separable filters). This is a natural choice because directional processing is an essential part of motion processing, and is required to explain the direction-depth contingency in Pulfrich effects. The model was applied to uniformly explain Pulfrich stimuli with continuous motion, stroboscopic motion, and dynamic noise. However, for stroboscopic stimuli, only small interflash intervals were considered and consequently, the S curves were not simulated. Under this condition, it is sufficient to consider, at a given time, cells whose motion preference matches the stimulus motion since their responses dominate over other cells' responses. As the stimulus goes through different directions and speeds of motion, cells with different motion preferences dominate the responses and explain the perception (Qian & Andersen, 1997).

Subsequent single-unit recordings from cat V1 (Anzai et al., 2001) and monkey V1 and MT (Grunewald & Skoumbourdis, 2004; Pack et al., 2003) are consistent with the joint motion-disparity coding model. Again, to account for motion perception in general and the direction-depth contingency in Pulfrich effects in particular, these studies naturally focused on cells that are jointly tuned to unidirectional motion and disparity. Different cells are tuned to different combinations of motion and disparity, and collectively, they cover a range of motion preferences and a range of disparity preferences combinatorially, as required by the joint coding model.

Unfortunately, the recent computational study (Read & Cumming, 2005a) changed the definition of joint coding. For both separate and “joint” coding models, the study

(Read & Cumming, 2005a) used a set of cells with a range of disparity preferences but all having exactly identical motion preference. For the separate coding model, all cells prefer zero velocity whereas for the “joint” coding model, all cells prefer a non-zero velocity and are thus spatiotemporally inseparable. The study (Read & Cumming, 2005a) incorrectly assumes that a model jointly codes motion and disparity when spatiotemporal inseparable filters are used. Although the original Pulfrich model (Qian & Andersen, 1997) also used spatiotemporally inseparable filters, an essential requirement of joint coding is that cells cover a range of motion preferences as well as a range of disparity preferences so that they can actually encode both motion and disparity. In contrast, the recent “joint” coding model (Read & Cumming, 2005a) can encode only disparity but not motion and is thus not a joint motion-disparity coding model.

Also note that the recent Pulfrich models (Read & Cumming, 2005a, 2005c) focus on simulating the S curves in stroboscopic Pulfrich effect (Morgan, 1979) under non-physiological assumptions such as monophasic temporal kernels. The S curves, although important, are about depth perception only and do not concern the most basic aspect of Pulfrich phenomena—the direction-depth contingency. A model for Pulfrich effects should be able to explain both.

In this study, we show that the joint coding model can reproduce the S curves whereas the corresponding separate coding models cannot, when realistic multi-phasic temporal kernels are used. Together with previous explanation of direction-depth contingency for various Pulfrich stimuli (Qian & Andersen, 1997), the joint-coding model provides the most complete account of Pulfrich effects to date. To simulate the S curves, we need to consider *large* interflash intervals in stroboscopic Pulfrich stimuli. Since under this condition, no motion preference dominates the responses, we pooled across different motion preferences before estimating disparity. We further found that either unidirectional or bidirectional motion selectivity can be used in the joint coding model to explain the S curves. This is not surprising because the S curves do not involve perception of motion direction. However, one may debate whether joint coding of bidirectional motion and disparity could also be viewed as a version of separate coding. We do not think so because the model covers a range of motion-speed preferences and a range of disparity preferences combinatorially, and thus can jointly encode speed and disparity. By reducing the range of speed preferences to a single speed preference, we produced truly separate coding models and showed that they cannot explain the S curves. In any case, the debate is not particularly interesting because only joint coding of *unidirectional* motion and disparity can explain both the S curves and the direction-depth contingency in Pulfrich effects.

In conclusion, Pulfrich phenomena are best explained by joint coding of unidirectional-motion and disparity, but

not by separate coding. These phenomena thus provide strong evidence of joint processing of motion and disparity in the brain.

Acknowledgments

This work was supported by NIH Grants EY016270 (NQ) and EY01175 (RDF).

Commercial relationships: none.

Corresponding author: Dr. Ning Qian; Dr. Ralph D. Freeman.
Email: nq6@columbia.edu; freeman@neurovision.berkeley.edu.

Address: Department of Neuroscience, Columbia University/NYSPI, 1051 Riverside Drive, Kolb Annex 519, New York, NY, 10032, USA; School of Optometry, University of California, 360 Minor Hall, Berkeley, CA 94720, USA.

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, Optics and Image Science*, 2, 284–299. [[PubMed](#)]
- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523–525. [[PubMed](#)]
- Anstis, S. M. (1974). Movement aftereffects contingent on binocular disparity. *Perception*, 3, 153–168. [[PubMed](#)]
- 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, 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, 609–611. [[PubMed](#)]
- Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research*, 19, 523–532. [[PubMed](#)]
- Castelo-Branco, M., Neuenschwander, S., & Singer, W. (1998). Synchronization of visual responses between the cortex, lateral geniculate nucleus, and retina in the anesthetized cat. *Journal of Neuroscience*, 18, 6395–6410. [[PubMed](#)] [[Article](#)]
- Chen, Y., Wang, Y., & Qian, N. (2001). Modeling V1 disparity tuning to time-varying stimuli. *Journal of Neurophysiology*, 86, 143–155. [[PubMed](#)] [[Article](#)]
- Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms: Effects of spatial density and interocular delay. *Vision Research*, 33, 1959–1975. [[PubMed](#)]
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993a). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex: 1. General characteristics and postnatal development. *Journal of Neurophysiology*, 69, 1091–1117. [[PubMed](#)]
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993b). Spatiotemporal organization of simple-cell receptive-fields in the cat's striate cortex: 2. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69, 1118–1135. [[PubMed](#)]
- Falk, D. S., & Williams, R. (1980). Dynamic visual noise and the stereophenomenon: Interocular time delays, depth, and coherent velocities. *Perception & Psychophysics*, 28, 19–27. [[PubMed](#)]
- Fernández, J. M., Watson, B., & Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research*, 42, 883–898. [[PubMed](#)]
- Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *Journal of Neuroscience*, 28, 4823–4835. [[PubMed](#)] [[Article](#)]
- Grunewald, A., & Grossberg, S. (1998). Self-organization of binocular disparity tuning by reciprocal corticogeniculate interactions. *Journal of Cognitive Neuroscience*, 10, 199–215. [[PubMed](#)]
- Grunewald, A., & Skoumbourdis, E. K. (2004). The integration of multiple stimulus features by V1 neurons. *Journal of Neuroscience*, 24, 9185–9194. [[PubMed](#)] [[Article](#)]
- Hawken, M. J., Shapley, R. M., & Gross, D. H. (1996). Temporal-frequency selectivity in monkey visual cortex. *Visual Neuroscience*, 13, 477–492. [[PubMed](#)]
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth*. Thornhill, Ont.: I. Porteous.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195, 215–243. [[PubMed](#)] [[Article](#)]
- Lee, D. N. (1970). A stroboscopic stereophenomenon. *Vision Research*, 10, 587–593. [[PubMed](#)]
- Morgan, M. J. (1979). Perception of continuity in stroboscopic motion: A temporal frequency analysis. *Vision Research*, 19, 491–500. [[PubMed](#)]
- Morgan, M. J., & Fahle, M. (2000). Motion-stereo mechanisms sensitive to inter-ocular phase. *Vision Research*, 40, 1667–1675. [[PubMed](#)]

- Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception, 4*, 3–18. [PubMed]
- Morgan, M. J., & Tyler, C. W. (1995). Mechanisms for dynamic stereomotion respond selectively to horizontal velocity components. *Proceedings of the Royal Society of London B: Biological Sciences, 262*, 371–376. [PubMed]
- Morgan, M. J., & Ward, R. (1980). Interocular delay produces depth in subjectively moving noise patterns. *Quarterly Journal of Experimental Psychology, 32*, 387–395. [PubMed]
- Nawrot, M., & Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science, 244*, 716–718. [PubMed]
- Neri, P., & Levi, D. M. (2008). Evidence for joint encoding of motion and disparity in human visual perception. *Journal of Neurophysiology, 100*, 3117–3133. [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]
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. *Journal of Neurophysiology, 75*, 1779–1805. [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*, 525–535. [PubMed] [Article]
- Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology, 40*, 1392–1405. [PubMed]
- Pulfrich, C. (1922). Die Stereoskopie im Dienste der isochromem und herterochromem Photometrie. *Naturwissenschaft, 10*, 553–564.
- Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6*, 390–404.
- Qian, N., & Andersen, R. A. (1997). A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research, 37*, 1683–1698. [PubMed]
- Qian, N., Andersen, R. A., & Adelson, E. H. (1994a). Transparent motion perception as detection of unbalanced motion signals: 1. Psychophysics. *Journal of Neuroscience, 14*, 7357–7366. [PubMed] [Article]
- Qian, N., Andersen, R. A., & Adelson, E. H. (1994b). Transparent motion perception as detection of unbalanced motion signals: 3. Modeling. *Journal of Neuroscience, 14*, 7381–7392. [PubMed] [PubMed]
- Qian, N., & Mikaelian, S. (2000). Relationship between phase and energy methods for disparity computation. *Neural Computation, 12*, 279–292. [PubMed]
- Read, J. C., & Cumming, B. G. (2005a). All Pulfrich-like illusions can be explained without joint encoding of motion and disparity. *Journal of Vision, 5*(11):1, 901–927, <http://journalofvision.org/5/11/1/>, doi:10.1167/5.11.1. [PubMed] [Article]
- Read, J. C., & Cumming, B. G. (2005b). Effect of interocular delay on disparity-selective V1 neurons: Relationship to stereoacuity and the Pulfrich effect. *Journal of Neurophysiology, 94*, 1541–1553. [PubMed] [Article]
- Read, J. C., & Cumming, B. G. (2005c). The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth. *Journal of Vision, 5*(5):3, 417–434, <http://journalofvision.org/5/5/3/>, doi:10.1167/5.5.3. [PubMed] [Article]
- Regan, D., & Beverley, K. I. (1973). Disparity detectors in human depth perception: Evidence for directional selectivity. *Science, 181*, 877–879. [PubMed]
- Ross, J. (1974). Stereopsis by binocular delay. *Nature, 248*, 363–364. [PubMed]
- Spang, K., & Morgan, M. (2008). Cortical correlates of stereoscopic depth produced by temporal delay. *Journal of Vision, 8*(9):10, 1–12, <http://journalofvision.org/8/9/10/>, doi:10.1167/8.9.10. [PubMed] [Article]
- Tyler, C. W. (1974). Stereopsis in dynamic visual noise. *Nature, 250*, 781–782. [PubMed]
- Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, Optics and Image Science, 2*, 322–341. [PubMed]