The stroboscopic Pulfrich effect is not evidence for the joint encoding of motion and depth

Jenny C. A. Read
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD

Bruce G. Cumming
Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, MD

In the Pulfrich effect, an illusion of depth is produced by introducing differences in the times at which a moving object is presented to the two eyes. In the classic form of the illusion, there is a simple explanation for the depth percept: the interocular delay introduces a spatial disparity into the stimulus. However, when the moving object is viewed stroboscopically, this simple explanation no longer holds. In recent years, depth perception in the stroboscopic Pulfrich effect has been explained by invoking neurons that are sensitive both to stereo disparity and to direction of motion. With such joint motion/disparity encoders, interocular delay causes a perception of depth by causing a shift in each neuron's preferred disparity. This model has been implemented by N. Qian and R. A. Andersen (1997). Here we show that this model's predictions for perceived disparity are quantitatively at odds with psychophysical measures. The joint-encoding model predicts that the perceived disparity is the virtual disparity implied by the apparent motion; in fact, the perceived disparity is smaller. We show that the percept can be quantitatively explained on the basis of spatial disparities present in the stimulus, which could be extracted from pure disparity sensors. These results suggest that joint encoding of motion and depth is not the dominant neuronal basis of depth perception in this stimulus.

Introduction

The detection of image motion on the retina and the detection of binocular disparities present similar computational problems: Object motion gives rise to images that occupy different retinal locations at different times; in binocular viewing, the images occupy different retinal locations in the two eyes. The most famous demonstration of a link between the processing of motion and stereoscopic depth is the Pulfrich effect (Burr & Ross, 1979; Lee, 1970a, 1970b; Morgan, 1976; Morgan & Thompson, 1975; Pulfrich, 1922). This is a depth illusion, created when the image from one eye reaches the brain later than from the other. The delay may arise clinically (e.g., in multiple sclerosis if optic nerve damage slows transmission of the image from one eye; Rushton, 1975) or can be introduced artificially. Before the advent of electronic stimuli, this was done by placing a neutral density filter over one eye to dim the image; dim images require more retinal processing time than bright images, so the image from the filtered eye reaches the brain later (Julesz & White, 1969; Ross & Hogben, 1975). Nowadays, computer-generated stimuli can be presented with a real interocular delay between the times at which the images are presented to the two eyes, allowing the interocular delay to be precisely controlled, and its effect studied without confounding effects due to luminance differences. If images of a target moving horizontally back and forth in the frontoparallel plane are viewed with interocular delay, an illusion of depth is created: The target appears first in front of, then behind, the fixation plane, depending on its direction of motion.

It was pointed out many years ago (Pulfrich, 1922) that there is a simple geometrical explanation for this: A moving object changes its position over time, so interocular delay gives rise to a real spatial disparity on the retina. Suppose the image reaching the left eye is delayed relative to the right eye, and that the object, moving at speed v, has position x when it is first seen by the right eye. By the time this same image reaches the left eye, the image in the right eye will have moved to a new position, $x + v\Delta t$. At this moment, the left eye's image is at x, while the right eye's image is at $x + v\Delta t$, so there is a spatial disparity, $\Delta x = v\Delta t$, which is interpreted as depth in the usual way. Thus, the illusion of depth in the classic Pulfrich effect is a consequence of the stimulus and – while confirming that spatial disparity results in a percept of depth – does not inform us further about the neuronal mechanisms of depth perception.

However, this geometrical explanation does not hold for the stroboscopic version of the Pulfrich effect (Burr & Ross, 1979; Lee, 1970b; Morgan, 1975, 1976, 1979; Morgan & Thompson, 1975), in which the moving target is illuminated only intermittently. A space-time diagram of this stimulus is shown in Figure 1. The squares represent appearances of the target; blue indicates its appearances in the right eye, and red those in the left. The horizontal axis represents the time of each appearance; the vertical axis is position. Note that the two eyes never receive simultaneous images of the target. If the strobe flashes at time t, both eyes

see an image of the target in the position it occupied at time t, but the right eye sees this image immediately, while the left eye does not see this image until time $t + \Delta t$. The difference between this and the classic stimulus is that here the right eye is not presented with an image at time $t + \Delta t$. The brain must therefore "remember" the right eye's image from time t to match it with the image that occurs later in the left eye. This stimulus does have the potential to inform us about the neuronal mechanisms involved. For example, the perception of depth in this stimulus depends on the temporal integration period over which stereo matches are possible (Lee, 1970a; Morgan, 1979).

The early literature on Pulfrich-like stimuli contained two major explanations for depth perception in Pulfrich-like stimuli. Ross (Ross, 1974, 1976; Ross & Hogben, 1974) suggested that interocular delay might per se produce a perception of depth, given that moving objects with non-zero disparity stimulate corresponding points on the two retinas at different times. Tyler (1974, 1977) and Morgan (1979) suggested that spatial disparities physically present in the stimulus, after temporal filtering by the visual system, might suffice to explain the depth percept. In recent years, these two explanations have merged. Depth perception is now assumed to depend on neuronal mechanisms that apply joint spatiotemporal filtering, making them sensitive to direction of motion as well as to disparity (Anzai, Ohzawa, & Freeman, 2001; Carney, Paradiso, & Freeman, 1989; Morgan & Castet, 1995; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Pack, Born, & Livingstone, 2003; Qian, 1997; Qian & Andersen, 1997). These joint motion/disparity sensors are characterized by receptive field profiles that are tilted relative to the space-time axes ("space/time-inseparable") (Adelson & Bergen, 1985; DeAngelis, Ohzawa, & Freeman, 1995), so their preferred disparity changes as a function of interocular delay. Joint motion/disparity sensors "cannot distinguish an interocular time delay from a binocular disparity" (Qian & Andersen, 1997), so they represent a modern version of Ross's suggestion that interocular delay directly causes a depth percept. To see how they explain depth perception in the stroboscopic Pulfrich stimulus, note that the flashed stimuli define an apparent motion in both eyes. The interocular delay means that the trajectory of this apparent motion has a disparity, even when the individual flashed images do not. Filters that encode disparity and motion simultaneously would be sensitive to the virtual disparity defined by these apparent motion paths. This joint-encoding model provides the modern, unified explanation of all Pulfrich-like phenomena, including both the classic Pulfrich effect and the stroboscopic version, as well as other depth illusions, such as dynamic noise viewed with an interocular delay (Falk & Williams, 1980; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Morgan & Ward, 1980; Ross, 1974; Tyler, 1974, 1977). Recent studies finding joint motion/disparity sensors in areas 17/18 of the cat (Anzai et al., 2001) and in area MT/MST of the monkey (Pack et al., 2003) have

therefore hailed them as the physiological substrate underlying depth perception in these stimuli.

However, while plausible, this model has never been thoroughly tested. A detailed comparison between the predictions of this model and psychophysical data has been hampered by the conflicting reports concerning disparity perception in the strobe Pulfrich effect. Lee (1970b) reported that depth in the stroboscopic Pulfrich effect was considerably smaller than in the classic case, and did not classify it as a "true" Pulfrich effect at all. In contrast, Burr and Ross (1979) reported that the amount of depth perceived was exactly the same as for the "classic" Pulfrich effect where the object is continuously visible: The object is seen at the virtual disparity implied by its apparent motion (see Figure 1). Morgan and colleagues (Morgan, 1979; Morgan & Thompson, 1975) found that the amount of depth depended on the time separating successive appearances of the stroboscope. When this inter-flash interval was short (10 ms), the perception was indeed that of the virtual disparity. However, at the longer inter-flash interval used by Burr and Ross (50 ms), Morgan found that – while depth was still perceived – the perceived disparity was less than the virtual disparity.
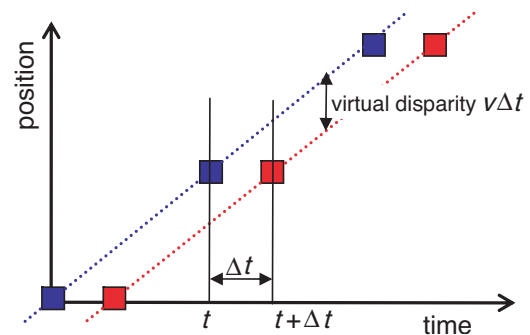

Figure 1. Space-time diagram of stroboscopic Pulfrich stimulus. The squares represent appearances of the stroboscopically illuminated target in the two eyes: blue for appearances in the right eye and red for those in the left eye. The target appears at the same position in each eye, but there is an interocular delay such that the target appears a time Δt later in the left eye than it does in the right. The dotted lines indicate the trajectory implied by the apparent motion. The "virtual disparity" $v\Delta t$ is defined to be the spatial separation between these two lines (Burr & Ross, 1979).

Both groups noted that this discrepancy could potentially be explained by subjects' eye movements. In the experiments of Burr and Ross, subjects were free to track the target. The target moved across a random-dot background, and subjects judged the depth of the target relative to the background. Because the interocular delay was applied only to the target and not to the background, tracking the target would introduce a real spatial disparity, equal to the virtual disparity, between target and background. Thus, the perception of the virtual disparity might reflect this real

spatial disparity, rather than depth induced by the interocular delay. In contrast, in Morgan's experiments, subjects were asked to maintain fixation on fixation bars, and the target was presented without a background (subjects had to judge the sense of its apparent rotation in depth as its direction of motion alternated). However, differences in the stimulus (the presence/absence of a background, the different target speeds, and possible differences in luminance or adaptation) might also have accounted for the different results, and the discrepancy remains unresolved.

Morgan (1979) suggested that the reduction in perceived depth he demonstrated for large inter-flash intervals reflected finite temporal integration early in visual processing. In principle this explanation might be applied either to direction-selective filters (joint encoding of motion and depth) or to nondirectional filters (separate encoding of motion and depth). Although this distinction was not drawn by Morgan (1979), subsequent publications have tacitly assumed the joint-encoding model (Anzai et al., 2001; Morgan & Castet, 1995; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Pack et al., 2003; Qian, 1997; Qian & Andersen, 1997). However, there is no clear reason for this assumption. If spatiotemporal filtering by non–motion-sensitive disparity sensors would explain the depth percept, then joint motion/disparity sensors would not be necessary to explain depth perception in the Pulfrich effect, potentially leading to very different conclusions about the neuronal substrate.

In fact, it has not been demonstrated that joint spatiotemporal filtering is even capable of explaining Morgan's (1979) results. The only explicit quantitative model using directional spatiotemporal filters (Qian & Andersen, 1997) produced results in accordance with Burr and Ross (1979). When the inter-flash interval of the strobe stimulus is sufficiently short compared to the integration time of the neuronal sensors, allowing them to respond to the apparent motion in the stimulus, the perceived disparity predicted by their model equals the virtual disparity. At long inter-flash intervals, the depth percept breaks down, because "cells tuned to different velocities report different equivalent disparities, and no particular disparity dominates perception" (Qian & Andersen, 1997, p. 1690). This model does not produce results like those of Morgan (1979), in which, at intermediate inter-flash intervals, there is a reliable percept of a disparity less than the virtual disparity but greater than zero. Thus, at the moment it is unclear whether the joint-encoding model produces the correct predictions or not. If we accept the results of Burr and Ross (1979), then the model is successful. If, however, Morgan is correct in attributing the results of Burr and Ross (1979) to eye movements, then the depth due to interocular delay in the strobe Pulfrich effect is in general *less* than the virtual disparity, and the results of the only extant joint-encoding model are at odds with the psychophysics, a discrepancy that has been universally ignored. Before we can accept the joint-encoding model as the unique explanation of depth perception in Pulfrich-like phenomena, it is clearly essential that this discrepancy be resolved.

As well as being important in its own right, this is a prerequisite for understanding the neuronal substrates supporting depth perception in different stimuli. If joint encoding turns out to be the only sustainable explanation for depth perception in the Pulfrich effect, this points to some rather surprising conclusions. Although joint motion/disparity encoding has been reported in cat area 17/18 (Anzai et al., 2001) and monkey MT/MST (Bradley, Qian, & Andersen, 1995; DeAngelis & Newsome, 2004; DeAngelis & Uka, 2003; Maunsell & Van Essen, 1983; Pack et al., 2003; Roy, Komatsu, & Wurtz, 1992), two recent studies have found that it is rather rare in monkey V1 (Pack et al., 2003; Read & Cumming, 2003, 2005). The data appear consistent with the suggestion that most cells that encode motion also encode disparity, whereas many cells encode disparity without encoding motion. Thus, areas with a high number of direction-selective cells (monkey MT/MST and cat area 17/18) also display a high proportion of joint motion/disparity encoding, whereas areas where direction-selectivity is less common (monkey V1) display a higher proportion of pure disparity encoding. It is, of course, possible that the depth percept in the strobe Pulfrich effect is supported entirely by joint motion/disparity sensors in areas like MT, with the pure disparity sensors in V1 playing no role. This would be surprising, because it would mean that the disparity signal carried by pure disparity sensors in V1 is ignored by the stereo system. Before drawing such a novel conclusion, it is obviously important to be sure that joint motion/disparity encoding is the only way to explain depth perception in the Pulfrich effect.

To clarify these issues, we felt it important to revisit the psychophysical data that have led so many workers to conclude that joint encoding of motion and disparity is the only way to explain depth perception in stimuli with an interocular delay. In this study, we examine perceived disparity and stereoacuity in the stroboscopic Pulfrich effect. We document how these vary as a function of interocular delay and inter-flash interval. As noted, previous studies have been hampered by eye movements, which can convert interocular delay into a spatial disparity on the retina. We devise a stimulus that removes the effect of eye movements, and enables us to study the depth percept induced by interocular delay alone. We find that the perceived disparity is less than the virtual disparity, in agreement with Morgan (1979) and in disagreement with the predictions of the joint-encoding model (Qian & Andersen, 1997). We develop an alternative model, based purely on the spatial disparities present in the stimulus (Tyler, 1974, 1977; Morgan, 1979), which quantitatively accounts for our results. We argue that this could be implemented by spatiotemporal filtering in pure disparity sensors, with no need to invoke filtering in joint motion-disparity sensors.

Methods

Experimental stimuli

Stimuli were generated on a Silicon Graphics Octane workstation and presented on two gamma-corrected Sony GDM F520 monitors (mean luminance 43.8 cd/m², contrast 99%, frame rate 96 Hz) viewed via a Wheatstone stereoscope. At the viewing distance used (89 cm), each pixel in the 1280×1024 display subtended 1.1 arcmin, and anti-aliasing was used to render with sub-pixel accuracy. The subjects were all experienced in stereo psychophysics; two were authors and one was a colleague without detailed knowledge of the experiments. No specific instructions were given; specifically, subjects were not instructed to avoid tracking the target.

Experiment 1

The target was a white square (0.5°×0.5°) moving horizontally against a black background with an apparent motion of 3.6°/s, and a disparity that varied between trials. The regions immediately above and below the path of the target were filled with a static zero-disparity background pattern (4°×4°) of randomly placed white dots (each 0.09°×0.09°). The moving target started at one side of the background region, and moved with a constant velocity until it reached the opposite boundary of the background region, at which point it was replotted at its starting location and repeated the same motion. This pattern repeated until the subject responded or 2 s elapsed, whichever came sooner. Both target and background were presented stroboscopically. The background dots were plotted simultaneously in each eye, but the target was plotted with an interocular delay. This stimulus reproduces the important features of the stimuli used by Burr and Ross (1979): The moving target is presented stroboscopically with a delay, while no delay is used for the stationary background. In a forced-choice task, subjects reported whether the target appeared to be in front of or behind the background pattern. The disparity of the target was varied according to a staircase procedure to determine the point where disparity nulled the depth induced by the interocular delay. Multiple interleaved staircases were used so that a block of trials contained stimuli of both directions, and both signs of interocular delay, randomly interleaved. The magnitude of the delay was constant within each block.

Experiment 2

Experiment 2 was the same as Experiment 1 except that the same interocular delay was applied to both target and background. This cancelled out the effect of eye tracking: Whether or not the subject tracked the moving target with their eyes, the relative spatial disparity between target and background was the same.

This is illustrated in Figure 2. The large squares indicate the moving target, while the small squares show one of the background dots. In the top row, the background dots are presented synchronously, as in Experiment 1; in the bottom, they are presented with the same interocular delay as the target, as in Experiment 2. The left column shows the stimulus as it would appear on the retina if the subjects kept their eyes still; the right column shows the retinal position if the subjects move their eyes with the apparent motion of the target.
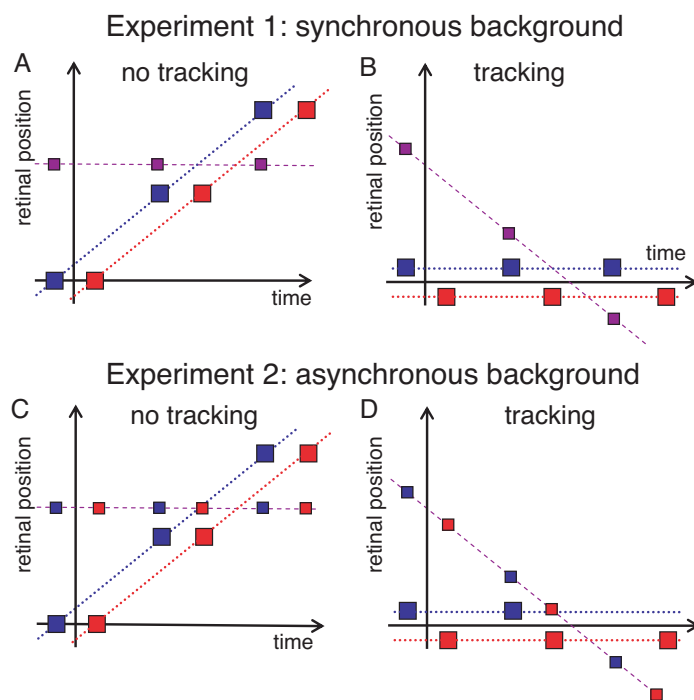

Figure 2. Tracking eye movements introduce relative disparity with a synchronous background, but have no effect when the background is asynchronous. As before, the large squares represent the target (red = left eye, blue = right). The smaller squares represent one dot from the random background pattern (shown in purple [blue+red] when it appears simultaneously in left and right eyes). The dotted lines represent the apparent motion of the target, and the dashed lines those of the background dot. A and B. Background has no interocular delay. C and D. Background has same interocular delay as the target. A and C. No tracking eye movements. The background dot is stationary on the retina, while the target moves. There is no spatial disparity between the target matches that are closest together in time. B and D. This shows the retinal position of the images when both eyes move together at the stimulus velocity (tracking, but no change in convergence). This is generated by displacing each dot downward by a distance *vt*, where *v* is the velocity of the eye movement. There is now spatial disparity between the closest matches of the target. If the background is synchronous (B), this results in a relative spatial disparity between target and background. If the background is asynchronous (D), it has the same spatial disparity as the target, so there is no relative spatial disparity.

Figure 2B shows the effect of tracking in Experiment 1. Because the eyes move during the interocular delay period, the target now has a spatial disparity on the retina. The background dots have retinal velocity in the opposite direction, and no spatial disparity on the retina. There is thus a relative spatial disparity between the target and background. If the subject tracked while also verging (not shown) to keep the target's image in each eye foveated, the target would have no spatial disparity but the vergence would cause the background to have the opposite spatial disparity, so the same relative disparity would exist between target and background.

When the background has an interocular delay, tracking eye movements do *not* introduce a relative spatial disparity between the background and the target. This is shown in Figure 2D. Tracking eye movements now introduce the same spatial disparity into the background as into the target. If the subject verged as well as tracking (not shown), neither background nor target would have spatial disparity. Thus, with this stimulus, tracking does not change the relative spatial disparity between target and background.

Experiment 3

Here the stimulus consisted of a random-dot pattern divided into upper and lower halves, distinguished by their direction of movement. Each half moved with a constant speed of 1.8°/s. The pattern occupied a window of size 4×4°; as dots left on one side of the window, they were replaced with new random dots on the other side of the window. The pattern consisted of randomly scattered black and white dots on a gray background. Each dot was 0.09°×0.09°; the dot density covered 50% of the screen, although because dots were allowed to overlap, the actual area covered was slightly smaller than this. A fixation dot was placed in the center of the pattern. The same interocular delay was applied to the whole pattern. Because the top and bottom halves were moving in opposite directions, this caused them to appear opposite in depth. Disparity was applied with the same magnitude but opposite signs in upper and lower halves. The magnitude of disparity was varied randomly according to a staircase procedure to find the relative disparity, which nulled the perception of relative depth between upper and lower halves.

Notation

We write T for the inter-flash interval of the notional stroboscope, and X for the inter-flash distance. The speed of the apparent motion is therefore $v = X/T$. It will be convenient to express the interocular delay Δt as a fraction of the inter-flash interval T, and the perceived disparity Δx as a fraction of the inter-flash distance X; thus, the results (Figure 5-7) show $\Delta x/X$ as a function of $\Delta t/T$. We define positive interocular delay to mean that the right eye sees a given stimulus first. We define positive disparity to be far (uncrossed).

Psychometric functions and fitting

Psychometric functions are fitted with a cumulative Gaussian according to the maximum-likelihood method. The mean of the cumulative Gaussian represents the point of subjective equivalence (PSE), and its standard deviation represents the threshold. The 68% confidence intervals on the estimates of PSE and threshold are obtained by finding the most extreme values, which reduce the log likelihood to 0.5 below the maximum (Watson & Pelli, 1983).

Modeling

The joint-encoding model (Qian & Andersen, 1997) predicts that the perceived disparity in a strobe Pulfrich stimulus is given by $\Delta x/X = \Delta t/T$. In contrast, we postulate that the perceived disparity is the weighted average of all disparities physically present in the stimulus, considering all appearances of the stimulus in the left eye as possible matches for a given appearance of the stimulus in the right eye. For example, consider the three possible matches in the left eye for the appearance of one target in the right eye shown in Figure 3A. The zero-disparity match (brown) has the shortest interocular time difference, and hence is given greatest weight. As shown in Figure 3B, we assign a weight that falls off as a Gaussian function of the temporal separation:

$$w(\Delta t) = \exp(-\Delta t^2/2\tau^2), \tag{1}$$

where the time-constant τ represents the binocular integration time. This represents a simple and reasonably realistic approximation to the autocorrelation function of the temporal impulse function, which is what should control the binocular response. In the example of Figure 3, the match with a negative disparity (pink) has smaller temporal separation and hence greater weight than the match with a positive disparity (green), hence the weighted average of all three matches will yield a negative disparity. Averaging over all possible matches, we predict that the disparity perceived in a strobe Pulfrich stimulus is, as a fraction of the strobe inter-flash distance X,

$$\frac{\Delta x_{perc}}{X} = \frac{\sum_{j=-\infty}^{\infty} j.w(jT + \Delta t)}{\sum_{j=-\infty}^{\infty} w(jT + \Delta t)}, \tag{2}$$

where the index j describes each possible match, and $j = 0$ corresponds to the match that is most nearly simultaneous. In Figure 3, the brown, pink, and green matches are $j = 0$, $j = -1$, and $j = +1$, respectively. Note that because this equation results in a negative (near) perceived disparity when the interocular delay is positive (right eye experiences a given stimulus first), it applies when the target is moving to the left. For target motion to the right, a similar derivation results in the same equation with an overall minus sign.
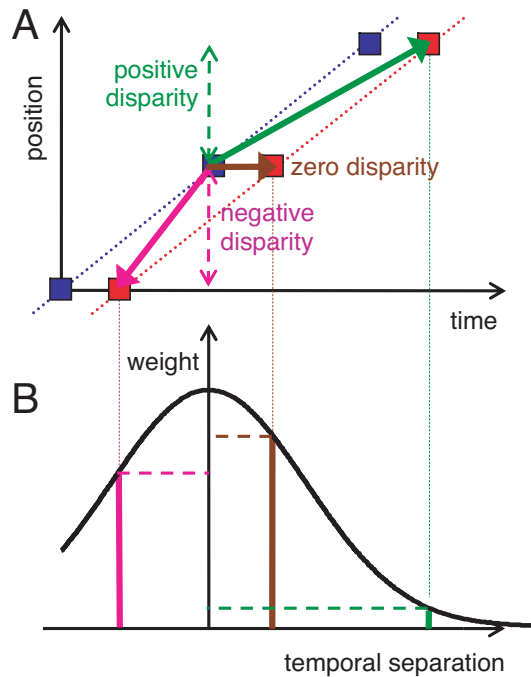
Figure 3. Different possible matches in the stroboscopic Pulfrich stimulus. A. As in Figure 1, the squares show the stroboscopic appearances of the target, and the dotted lines show the apparent motion of the target (red = left-eye; blue = right-eye). Consider the second appearance of the target in the right eye. The arrows indicate three possible matches for this in the left eye. Because position is plotted on the vertical axis, the vertical component of each arrow indicates the spatial disparity of that match, while the horizontal component indicates the temporal separation between left and right half-images of the match. The match with the smallest temporal separation has zero spatial disparity (brown). However, at larger temporal separations, matches exist with positive (green) or negative (pink) spatial disparities. B. The weight assigned to each match has a Gaussian dependence on the temporal separation (black curve). The heights of the bars indicate the weights given to the three matches shown in A. Most weight is given to the match whose left and right images are closest together in time (brown). The match with a negative disparity (pink) has a shorter temporal separation than the match with a positive disparity (green), and hence a greater weight. As a result, the weighted sum has a negative disparity.

We also considered a linear combination of the joint-encoding and separate-encoding models, where λ represents the relative contribution from joint encoding. Here the perceived disparity is

$$\frac{\Delta x}{X} = \lambda \frac{\Delta t}{T} + (1-\lambda) \frac{\sum_{j=-\infty}^{\infty} j.w(jT+\Delta t)}{\sum_{j=-\infty}^{\infty} w(jT+\Delta t)} . \quad (3)$$

Results

Psychophysics

Experiment 1

In this experiment, the target square moved horizontally against a static background of random dots. In each run with a given magnitude of interocular delay, both directions of target motion (left/right) and both signs of delay (left/right eye leading) were randomly interleaved. Example psychometric functions for stimuli with and without an interocular delay are shown in Figure 4. When the target is moving to the right (Figure 4A, top panel), then presenting the target in the right eye before the left (positive delay, green ▲) makes the target appear in front of the background in the absence of spatial disparity. A far disparity (positive) must be added to make the target appear in the same plane as the background. This PSE, obtained from the fitted curve, is shown with the red line. If the target appears in the left eye first (negative delay, purple ▼), then the target appears behind the background when its spatial disparity is zero, and the PSE is negative (near). If the target is moving to the left (Figure 4B, bottom), then a given sign
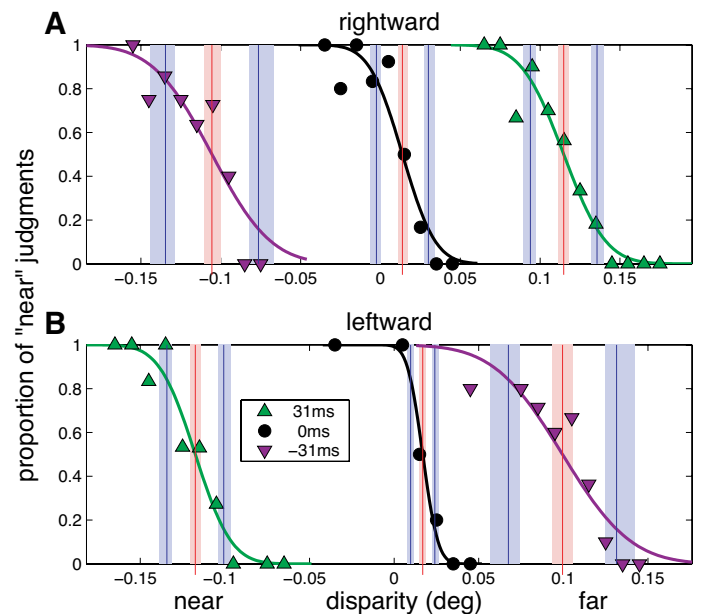

Figure 4. Example of psychometric functions for subject JR. A. Target motion rightward. B. Target motion leftward. Symbols show the proportion of "near" judgments at a given disparity, for the interocular delay specified in the legend. The curves show the cumulative Gaussian function fitted to the data. The vertical red lines show the mean of each fitted Gaussian, which is the point of subjective equivalence (PSE); the blue lines show ±1 *SD*, which is the threshold. The shaded regions around the PSE and threshold show the 68% confidence intervals estimated from the log likelihood ratio.

of delay has the opposite effect. We take the disparity threshold to be the change in disparity necessary to take the subject from the PSE, where they report "near" or "far" at random, to performing with 84% consistency. The location of this 84% performance level is indicated on each plot with blue lines. The shaded regions around the red and blue lines indicate the 68% confidence interval on the PSE and threshold, respectively.

It is clear from Figure 4 that the subject has a bias toward near disparities; when there is no interocular delay and no spatial disparity, she tends to judge that the target is in front of the background, meaning that a far disparity must be applied to the target to make her perceive it in the same plane as the background. We can remove the effect of this bias by looking at the difference in her judgments when the target is moving left versus when it is moving right. From this, we extract a perceived disparity independent of any such constant bias.

Figure 5 summarizes the resulting estimates of perceived disparity for three subjects, three inter-flash intervals and a range of interocular delays. The perceived disparity is plotted as if the target was moving to the right (i.e., positive delays produce a positive perceived disparity [negative PSE]). The axes show perceived disparity Δx as a fraction of inter-flash distance X, and interocular delay Δt as a fraction of inter-flash interval T. The dots lie close to the identity line $\Delta x/X = \Delta t/T$, indicating that the perceived disparity Δx is close to the virtual disparity $v\Delta t$ implied by the apparent motion $v = X/T$ of the target, just as would be the case in the classic Pulfrich effect if the target were illuminated continuously. These results are in agreement with those reported by Burr and Ross (1979), and with the predictions of the joint-encoding model (Qian & Andersen, 1997).

However, as noted by Burr and Ross themselves, these results are also exactly what would be expected if the subjects were tracking the target, their eyes moving smoothly with the apparent speed of the target. Suppose the target appears in the right eye first, and appears a time Δt later in the left eye. Although both appearances are presented at the same position on the screen, because the eyes have moved through an angle $v\Delta t$ in this time, the target appears at different positions in the two retinas (Figure 2B). Eye movements would thus introduce a real relative disparity of $v\Delta t$ between target and background, which would naturally be perceived as a difference in depth. Thus, the perceived disparity would equal the virtual disparity $v\Delta t$, independently of the neuronal mechanisms encoding depth, and in particular independently of whether interocular delay can influence depth perception. Burr and Ross felt that it was unlikely that subjects could track a stroboscopically presented target with the required accuracy. However, detailed recording of eye movements by Morgan and colleagues (Morgan & Turnbull, 1978; Morgan & Watt, 1982, 1983; Ward & Morgan, 1978) suggest that, in fact, such accuracy may be possible.

Experiment 2

The most reliable way to eliminate this problem is to devise a stimulus in which eye movements do not introduce a relative disparity between the target and background. This can be achieved very simply by applying the interocular delay not only to the target but also to the background. Because the background is stationary, the delay does not alter the perceived depth of the background when the eyes are fixating. However, if the eyes move, then the *same* retinal disparity is applied to both target and background (because both the retinal slip caused by the eye movement, and the
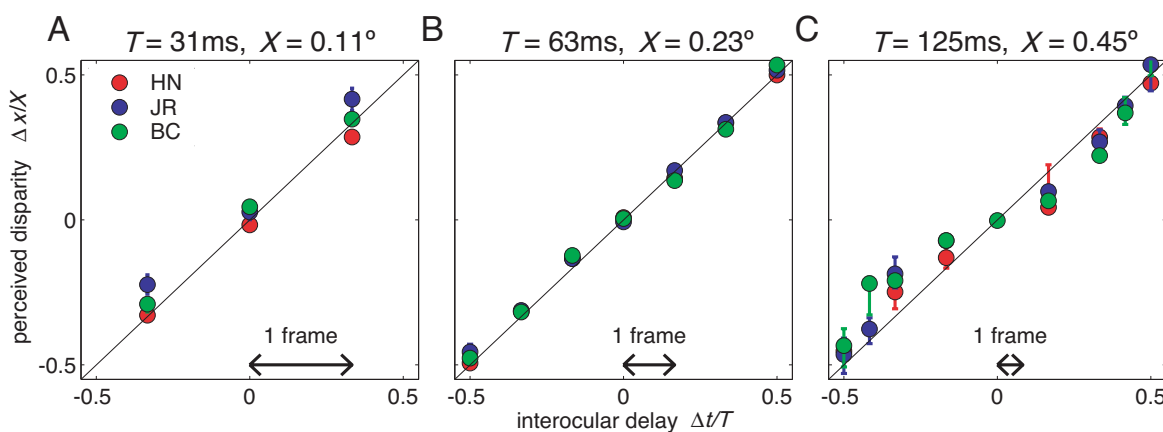

Figure 5. Results of Experiment 1. The dots represent the perceived disparity (i.e., the negative of the point of subjective equivalence) for stroboscopic Pulfrich stimuli with different inter-flash intervals T (A: T = 31 ms; B: T = 63 ms; and C: T = 125 ms) and interocular delays. The interocular delay Δt is plotted as a fraction of the strobe inter-flash interval T, so the smallest increment in Δt, namely 1 frame, is represented by a different distance on each of the three axes, as indicated by the arrow. The perceived disparity is displayed as a fraction of the strobe inter-flash distance X. Because the apparent speed was always the same, the inter-flash distance X increases with the inter-flash time interval T. The 68% confidence intervals on perceived disparity (see Figure 4) are shown with error bars, although in almost all cases these are smaller than the symbols.

interocular delay in the stimulus, are the same for both target and background). The relative spatial disparity between target and background thus remains zero (Figure 2D). Any perceived difference in their depth is due only to the movement of the target across the static background. Thus, the results are unaffected by eye movements. This stimulus therefore allows us to measure the depth that would be perceived in the stroboscopic Pulfrich stimulus if the eyes remained stationary.

The results are shown in Figure 6. Clearly, the pattern is very different from Experiment 1 (Figure 5). At the shortest interocular delay (T = 31 ms, Figure 6A), the perceived disparity is the virtual disparity, so points lie on the identity line. However, at the longer interocular delays, the perceived disparity falls below the virtual disparity, so points fall on a sigmoid curve. This strongly suggests that the results of Experiment 1 were due to tracking eye movements. Traces of this sigmoid pattern are visible in Experiment 1 for T = 125 ms (Figure 5C), perhaps because subjects had difficulty in smoothly tracking the apparent motion of the target at this long inter-flash interval. The results of Experiment 2 suggest that if subjects could keep their eyes perfectly still while viewing the strobe Pulfrich stimulus, they would in general perceive less than the virtual disparity. Note that the difference between the stimuli in Experiments 1 and 2 is very subtle: Only the relative timing (between the eyes) for the appearance of stationary background dots is altered. It is therefore hard to see any explanations for the different results, other than the effects of tracking eye movements.

Conclusion

Early studies reported that the depth perceived in the stroboscopic Pulfrich effect was the virtual disparity implied by the apparent motion of the target. This has been explained by invoking joint motion/disparity sensors that are sensitive to this apparent motion. Subsequent studies, especially Morgan (1979), suggested that this result was the result of tracking eye movements. Here, for the first time, we directly compare two stimuli that differ only in the possible effects of eye movements, and find that the virtual disparity is perceived only in stimuli where eye movements can convert virtual disparities into physical disparities. This strongly supports the conclusion of Morgan (1979): If the eyes remain still, or if the effect of eye movements is removed by an appropriate stimulus, then the depth percept is less than expected from the virtual disparity. We now examine how this depth percept can be explained in terms of spatial disparities present in the stimulus.

Computational modeling

The sigmoid pattern of results in Experiment 2 was first found by Morgan (1979), who pointed out that it can be understood in terms of the temporal window over which the brain is able to combine inputs from the two eyes. Figure 3A shows three potential matches between appearances of the target in the two eyes. The left and right appearances, which are separated by the shortest time (brown arrow in Figure 3A), have zero disparity, but other appearances, more widely separated in time, do have disparity (pink and green arrows in Figure 3A). Because cortical neurons have finite integration times, their response is influenced by these more distant matches. We can distinguish two helpful limiting cases. (1) If the inter-flash interval T of the stroboscope is very short compared to the binocular integration time, then the apparent motion of the flashing target is indistinguishable from true continuous motion. The perceived disparity Δx is thus equal to $v\Delta t$, where v is
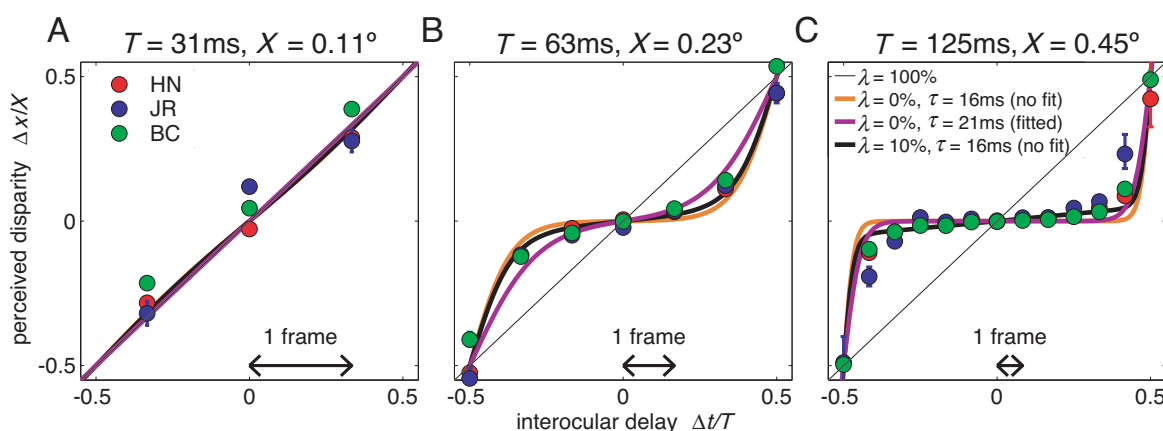

Figure 6. Results of Experiment 2 along with the predictions of our model. Details are as for Figure 5, except the interocular delay was applied to the background as well as to the target. The curves show the predictions of the disparity-averaging model in Equation 3, with different values for the integration time, τ, and the contribution from joint encoding, λ. The colored curves have λ = 0% (no joint encoding). For the orange curve, τ was set equal to 16 ms, the value obtained from V1 physiology. For the purple curve, τ was chosen to obtain the maximum-likelihood fit to the data (assuming all points were subject to the same error); this was 21 ms. For the black curve, τ was again set equal to 16 ms, but this time λ was 10%, the approximate proportion of joint motion/disparity sensors found in V1. The identity line is the prediction for 100% joint encoding (where τ is irrelevant).

the speed of the apparent motion, as in the classic Pulfrich effect. (2) If, on the other hand, the inter-flash interval is much longer than the integration time, then a given appearance of the target in the delayed eye will either go unmatched, or will be matched only with the immediately preceding appearance in the leading eye, all previous appearances having been "forgotten"; either way, no depth will be perceived and $\Delta x = 0$. From Figure 6, $T = 31$ ms and $T = 125$ ms seem to be examples of these respective extremes. We sought to develop this argument to see whether it can quantitatively explain the perceived disparity.

We hypothesize that perception reflects a weighted average of the disparities of all possible matches in the stimulus (Morgan, 1979; Tyler, 1977), with most weight being given to matches whose left and right members appear simultaneously. We assume that the weight given to any match decays as a Gaussian function of the temporal separation between left and right members of the match (Figure 3B). Together, these assumptions yield the model of perceived disparity given in Equation 2. The standard deviation τ of the Gaussian weight function represents the integration time of binocular disparity sensors. Our experiments in macaque V1 (Read & Cumming, 2005) suggest that this is around 16 ms.

How well does this model account for the data? The orange curves in Figure 6 show the predictions of Equation 2, with the value $\tau = 16$ ms derived from the physiology, for the dependence of perceived disparity on interocular delay and inter-flash interval. With no free parameters, this model successfully captures the change from a linear relationship between perceived disparity Δx and interocular delay Δt at short inter-flash intervals, to a sigmoid relationship at long intervals. In contrast, the joint-encoding model (Qian & Andersen, 1997) predicts that the perceived disparity, Δx, is always proportional to the interocular delay, Δt (identity lines in Figure 6), which is clearly not supported by the data.

The predictions of Equation 2 are not perfect. For the longer inter-flash intervals, the disparities predicted by Equation 2 are smaller than those actually perceived. One possibility is that the integration time is longer in these human subjects than expected from the monkey physiology. Fitting τ as a free parameter results in a marginally better fit with $\tau = 21$ ms, shown with the purple curves in Figure 6; the improvement in log likelihood was not significant given the extra parameter.

Another possibility is that there is some contribution from joint-encoding mechanisms. In our previous study (Read & Cumming, 2005), we found that, while most cells encoded only spatial disparity, a subpopulation representing <10% of V1 cells did jointly encode both disparity and motion, as envisaged by Qian and Andersen (1997). If these cells contribute to perception, it is possible that subjects' performance reflects the virtual disparity as well as the spatial disparities actually present in the stimulus. To investigate this, we adopted a very simplistic model in which the weight given to the virtual disparity was assumed to reflect

the proportion of neurons in V1 that were sensitive to the virtual disparity (Equation 3 with $\lambda = 10\%$ and $\tau = 16$ ms). This prediction is shown with the black curve in Figure 6. Allowing for a contribution from joint encoding has slightly improved the match to experimental data, especially at the longest inter-flash interval $T = 125$ ms. Note that again the black curve has not been fitted to the experimental data; the parameters τ and λ in Equation 3 were taken from the physiology. If we do fit λ and τ as free parameters, the improvement from the original values ($\lambda = 0$, $\tau = 16$ ms, orange curve) is not worth the additional two free parameters (χ^2 log likelihood test); similarly, if we keep either λ or τ at the original values and fit the other as a free parameter, there is no significant improvement in fit.

Experiment 3

A potential problem with using the integration time derived from the physiology is that the stimuli used in the physiology experiments (Read & Cumming, 2005) differed in several respects from the stimuli used in the psychophysics presented so far. In Experiments 1 and 2, the moving target and background dots were all white, and the background was black; the target dot was large and the background dots were small. In contrast, the stimuli used in the physiology experiments were random-dot patterns made up of equal-sized white and black dots on a gray background. These differences in dot size and (particularly) mean luminance might have resulted in a different measure of integration time. We therefore studied the strobe Pulfrich effect with stimuli designed to be as similar as possible to those used in the physiology experiments, to see whether this changed the estimate of integration time. In this experiment, the stimuli were patterns of black and white random dots on a gray background, just as in the physiology experiments. The dots moved horizontally with an apparent speed of 1.8°/s, with dots in the top and bottom halves of the pattern moving in opposite directions. The relative speed between the two halves was thus 3.6°/s, the same as the speed of the target relative to the background in Experiments 1 and 2. The same interocular delay was applied to all dots in the pattern. Thus, because of the opposite directions of motion, the two halves of the pattern appeared at different depth planes. Subjects had to judge whether the top or bottom half appeared in front. We added a physical relative disparity between the top and bottom halves, and varied this in a staircase procedure to null the depth produced by the Pulfrich effect. The results are shown in Figure 7.

Subjects found this stimulus harder than the previous ones, and the subjective impression was reflected in higher thresholds. Subject JR could not perform reliably on the $T = 63$ ms stimulus. At higher interocular delays, subject BC showed a direction-dependent bias: That is, he was more likely to judge the rightward-moving half of the pattern to be in front, and the leftward-moving half behind. Our trick of removing a constant bias by looking at the difference in subjects' judgments when the target is moving

left versus when it is moving right obviously fails to deal with this direction-dependent bias. This is why BC's results lie above the identity line for the largest interocular delays in Figure 7. Fortunately, this problem has little effect on the estimates of integration time (the largest interocular delays for $T = 63$ ms, $\Delta t/T = 0.5$, have no effect on the fit at all, because the fit always intersects the identity line at that point, regardless of the parameters τ and λ). Because the depth percept was weaker, it was impossible to get reliable results at $T = 125$ ms, so the experiment was performed only at inter-flash intervals of 31 ms and 63 ms. These are in any case the most informative in constraining the integration time.

As in Figure 6, the orange and black curves in Figure 7 show the disparity predicted from Equation 3 using the integration time obtained from physiology (16 ms), first with no contribution from joint motion/disparity encoding ($\lambda = 0\%$, orange curve) and second with a contribution at the level suggested by the physiological incidence of joint motion/disparity sensors ($\lambda = 10\%$, black curve). Once again, the agreement is excellent. The estimates of binocular integration time obtained from the present study are also in reasonable agreement with the results of a previous study (Read & Cumming, 2005). There, we measured the decline in stereoacuity as interocular delay was added to dynamic random-dot stereograms with the same spatial structure as the stimuli used in Experiment 3. Stereoacuity declined as a Gaussian function of interocular delay, with a *SD* of 11 ms (mean for the two subjects BC and HN). Thus, the two independent psychophysical measures in humans are in reasonable agreement with each other and with the results of monkey physiology: All suggest a binocular integration time of around ~15 ms.

Conclusion

This suggests that the disparity perceived in the stroboscopic Pulfrich effect can be understood simply in terms of the known physiological mechanisms of V1. Perception is largely due to the spatial disparities present in the stimulus, weighted according to the interocular delay between left and right members of each match, but the minority of cells that jointly encode motion and depth may cause a small shift in perception toward the virtual disparity implied by the apparent motion.

Stereoacuity

According to our description of how depth is generated in stroboscopic Pulfrich displays, there should also be a systematic relationship between the interocular delay, inter-flash interval, and the reliability of depth judgments about the PSE (stereoacuity). As the delay increases, the visual responses in the two eyes overlap less, so the signal that gives rise to a depth sensation gets weaker. Thus, our explanation suggests that stereoacuity should increase with interocular delay in the stroboscopic Pulfrich effect. This is not a property of explanations based on joint encoding of
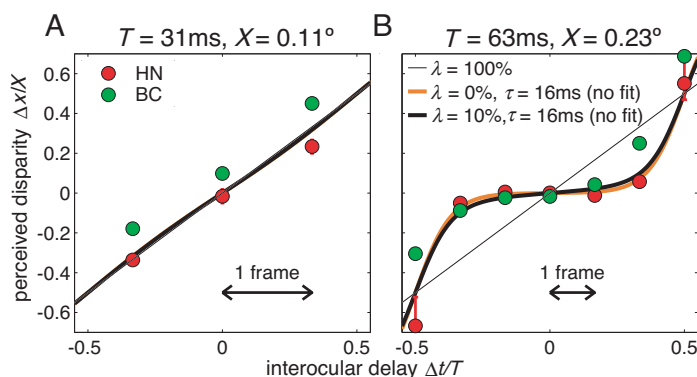

Figure 7. Results of Experiment 3 using the random-dot stimulus. For explanation of symbols and axes, see legends of Figures 5 and 6.

motion and depth, where the apparent motion trajectory is defined in each eye with equal precision at all delays. (Stereoacuity may increase with the inter-flash interval in these schemes, but it need not increase as a function of interocular delay.) We therefore examined stereo disparity thresholds, defined as the standard deviation of the cumulative Gaussian fitted to the psychometric function (see Figure 4). Figure 8 shows the thresholds measured in Experiment 2. Because the thresholds were more variable between subjects than perceived disparity was, the data for each subject are plotted in a separate panel, with colors now indicating the inter-flash interval *T*. Note that thresholds in degrees are plotted as a function of interocular delay in milliseconds, rather than normalized by the inter-flash interval as in previous figures.

Three key features of the data are apparent:

1. Threshold increases with interocular delay.

2. The threshold at zero interocular delay is largest for the shortest inter-flash interval.

3. The rate of increase in threshold with interocular delay is more rapid for $T = 63$ ms than for the other inter-flash intervals.

In the next several paragraphs, we examine each feature in turn, and show that each can be understand within the terms of our disparity-averaging model.

1. Threshold increases with interocular delay

Within our model, the increase of threshold with interocular delay is entirely expected. Cortical neurons have a finite integration time, so as the separation between left and right matches increases, the extent of overlap between incoming signals from the two eyes diminishes. Consequently, disparity tuning weakens and disparity thresholds increase. Thus, this effect is predicted by our disparity-averaging model. However, it is not clear that it can be explained by existing joint-encoding models. In the model of
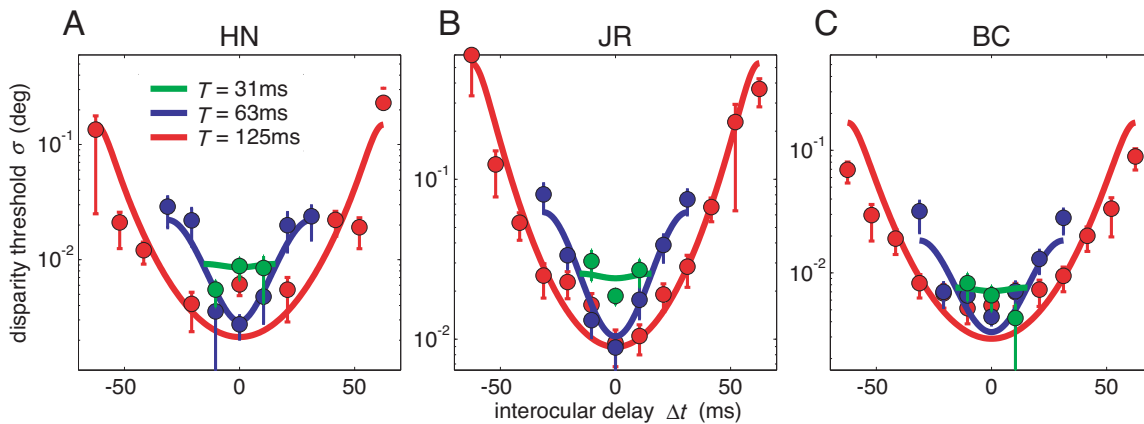
Figure 8. Dependence of disparity threshold on interocular delay and inter-flash interval. The symbols show the disparity thresholds for Experiment 2; the corresponding PSEs were plotted in Figure 6. Error bars show 68% confidence intervals estimated from the log likelihood ratio. The three panels show results for the different subjects; colors show results for different inter-flash intervals. The curves show fits from the model developed in the Appendix.

Qian and Andersen (1997), stereo matches are not made between individual appearances of the strobe target, but between the interpolated apparent motion paths. Because at the nulling disparity these apparent motion paths coincide, it is not clear why the size of the interocular delay should affect the stereoacuity.

2. The threshold at zero interocular delay is largest for the shortest inter-flash interval

Looking at Figure 8, for all three subjects the threshold at zero interocular delay is largest for the shortest inter-flash interval ($T = 31$ ms, green). This at first seems paradoxical: Surely perception should be clearest when the stimulus is

presented most frequently. To understand this effect, we need to consider the activity in a population of disparity sensors for different inter-flash intervals (Figure 9). At zero interocular delay, the most strongly activated sensors are those tuned to zero disparity. At long inter-flash intervals (Figure 9C and 9F), these are essentially the *only* active sensors, because the other disparities present in the stimulus are so widely separated in time that they do not cause significant activation. At short inter-flash intervals (Figure 9A and 9D), sensors tuned to many different disparities – integer multiples of the inter-flash distance – on either side of zero become activated as well, although less strongly.
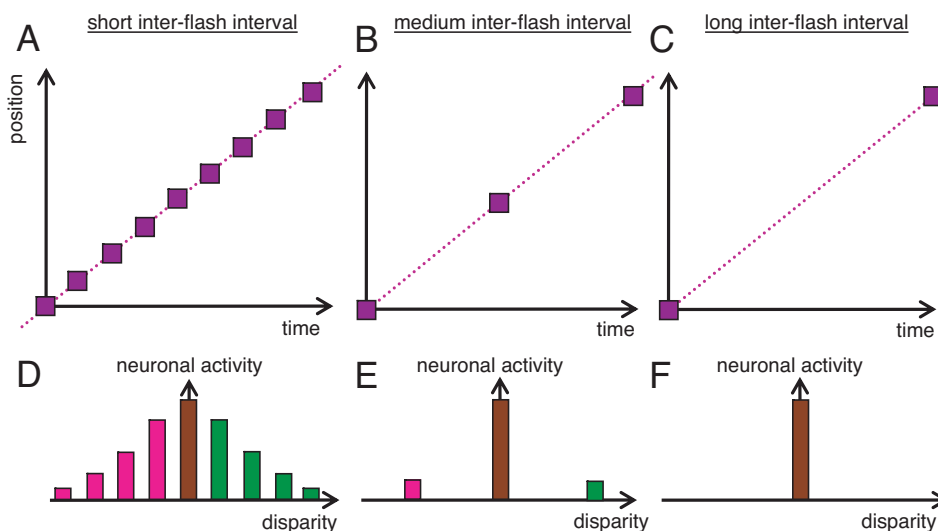

Figure 9. Population neuronal activity for zero interocular delay, at three different inter-flash intervals. A, B, and C: Space-time diagrams for three strobe Pulfrich stimuli with zero interocular delay and different inter-flash intervals. The appearances of the target are shown in purple, because the appearances in left and right eye (shown in red and blue in previous figures) now coincide. D, E, and F. Cartoons of neuronal activity (green = neurons tuned to positive disparity, brown = zero, and pink = negative). In D, the inter-flash interval is short compared to the neurons' integration time, so the population responds even to matches that are separated by several strobe inter-flash intervals, although the response is still strongest to the matches that are closest together in time. In F, the inter-flash interval is long compared to the integration time, so the only match detected by the population is the one between the coincident target appearances.

What does this imply for stereoacuity? It seems plausible that the perception of zero depth should be clearest when only zero-disparity sensors are activated, and thus that thresholds should be lower for longer inter-flash intervals, as observed. Yet because our disparity-averaging model relies on calculating the center of an activity distribution, one might also argue that the recruitment of additional sensors tuned to equal and opposite disparities about the center should help to reduce the overall noise level and hence make perception more accurate. The simple idea of disparity averaging does not, on its own, indicate what to expect here; one needs some additional assumptions about how it is implemented in the brain.

The curves fit to the data in Figure 8 show the results of one particular set of assumptions, described in the Appendix. This model assumes that the brain implements disparity averaging by summing the activity of all disparity sensors, weighted by their preferred disparity. The model takes into account that the noise on neuronal activity increases with mean firing rate (Dean, 1981), which turns out to be important in enabling the model to explain the higher zero-delay threshold at the shortest inter-flash interval. For simplicity, this model assumes no contribution from joint encoding, so for consistency, we use the integration time estimated previously by fitting the PSEs for all subjects together with Equation 2 (i.e., $\tau = 21$ ms, purple curve in Figure 6). Two further parameters, representing the amount of noise and its dependence on neuronal activity, are fitted to the threshold data for each subject individually. This model gives a good fit to the observed thresholds for each subject, demonstrating that a reasonably simple and plausible way of implementing disparity averaging with noisy neurons can account for the stereoacuity data.

Without performing a similar implementation of noise into the model of Qian and Andersen (1997), it is unclear what one would expect from joint encoding. The cartoons in Figure 9D, 9E, and 9F make it seem plausible that the winner-take-all model of Qian and Andersen (1997) could explain the observed increase in threshold at short inter-flash intervals for zero interocular delay: At short inter-flash intervals, when nonzero-disparity sensors are also activated, there is more chance that random noise fluctuations will boost one of them above the activity of the zero-disparity sensor, even though that has the strongest signal. This would predict higher thresholds for short inter-flash intervals, as observed. However, what the cartoon fails to take into account is that, in the model of Qian and Andersen, the neurons are also more strongly activated by the stimulus with the short inter-flash interval, because it supplies more power to the tilted receptive fields on which the model is based. Thus, for the model of Qian and Andersen, the activity of the zero-disparity sensor in Figure 9F would be much less than that in Figure 9D. For constant noise, this effect would predict *lower* thresholds for short inter-flash interval, contrary to what is observed. Which of these opposing effects wins out would presumably depend on the

details of how one incorporated noise into the model (the published version is noise-free).

3. The increase in threshold with interocular delay is more rapid for T=63 ms than for the other inter-flash intervals

The final effect noted in Figure 8 is that the rate at which the thresholds increase as a function of interocular delay is different for the three inter-flash intervals ($T = 31$ ms, 63 ms, and 125 ms, shown in green, blue, and red, respectively). Interestingly, the rate of increase seems to be steeper for the middle value, $T = 63$ ms, than for the shorter and longer inter-flash intervals. As indicated by the curves in Figure 8, this effect can also be explained by the implementation of disparity averaging developed in the Appendix. Here we give an intuitive account of how this works.

As we have seen, for short inter-flash intervals at zero interocular delay, many disparity sensors are activated (Figure 9D). In the noise model we have developed (see Appendix), this causes an increase in threshold. However, when an interocular delay is introduced, there is rather little change in the distribution of activity (Figure 10D). Once again, many disparity sensors are activated, and the perceived disparity is just the average of these. Although the distribution is no longer quite symmetric about zero, this has rather little effect on the reliability of the average. Thus, for the shortest inter-flash interval ($T = 31$ ms, Figure 8), the threshold is already high for zero interocular delay, but does not increase much further when an interocular delay is introduced.

Similarly for long inter-flash intervals, the situation for nonzero interocular delays (Figure 10C and 10F) is again rather similar to the situation for an interocular delay of zero (Figure 9C and 9F). In both cases, the only active sensor is that tuned to zero disparity, resulting in a percept of zero disparity. The percept is most reliable when the interocular delay is zero, because then the zero-disparity sensor is most strongly activated (compare Figure 9B and 9D with Figure 10C and 10F). Thus, the threshold increases with the magnitude of interocular delay.

For intermediate inter-flash intervals, however, the situation for zero interocular delay (Figure 9B and 9D) is quite different from that for finite interocular delays (Figure 10B and 10E). Because the stimuli are shown at the nulling disparity, in both cases the percept is of zero disparity. However, whereas for zero interocular delay this is supported by activity in the zero-disparity sensor (Figure 9B and 9E), for nonzero interocular delays, the zero-disparity percept is due to opposing activity in sensors tuned to *non*-zero disparities (Figure 10B and 10E). In Figure 10B, the positive-disparity match (green arrow) is stronger than the negative-disparity match (purple arrow) because the members of the match are closer together in time. However, it also has a smaller disparity, so the two cancel out to give a perceived disparity of zero. Thus, the perception of zero disparity is mediated by sensors tuned to *non*-zero disparity,
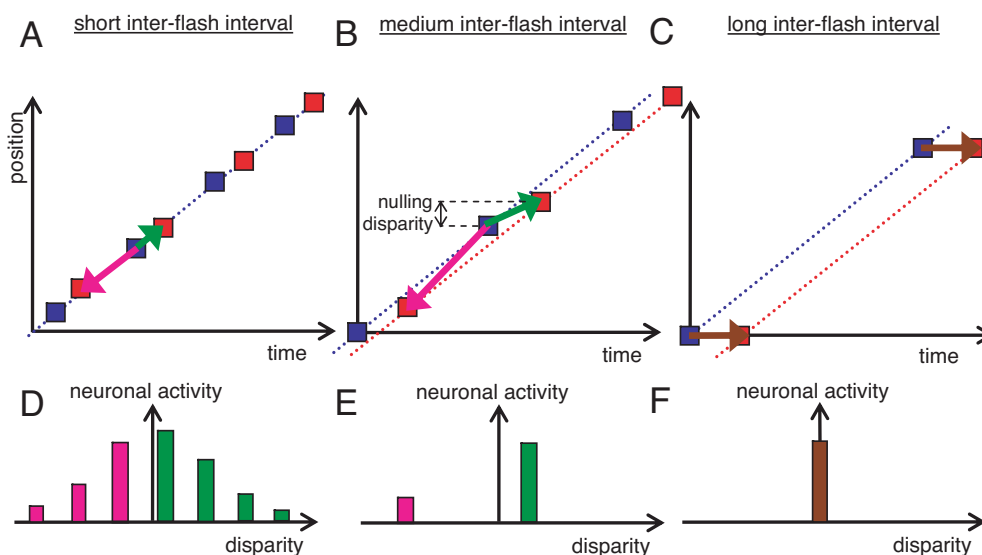
Figure 10. Why thresholds are smaller as inter-flash interval increases at a given interocular delay. A, B, and C. Space-time diagrams (see Figure 1) for three strobe Pulfrich stimuli with different inter-flash intervals. The stimuli in A and B have the spatial disparity necessary to null the perception of depth that would otherwise result. D, E, and F. Cartoons of activity level in a neuronal population tuned to different disparities. The active neurons are those tuned to the spatial disparities present in the stimulus (green = positive, pink = negative, and brown = zero disparity; see Figure 3). The activity level is higher when the two members of the match are close together in time.

as indicated in the cartoon of neural activity in Figure 10E. This in itself could explain the larger thresholds for short T, because two conflicting signals might be expected to give a weaker percept than single signal. In addition, disparity thresholds rise exponentially with the absolute disparity of the stimulus (Blakemore, 1970; Ogle, 1953), suggesting that the effective noise level may be higher when sensation is supported by sensors tuned to nonzero disparity. For both these reasons, thresholds rise most steeply as a function of interocular delay for the medium inter-flash interval.

As demonstrated by the curves in Figure 8, our model of neuronal noise is able to account for this behavior. Specifically, it captures the counter-intuitive result that, for a fixed interocular delay, stereoacuity *improves* as the inter-flash interval gets *longer*. It seems unlikely that this effect could be produced by the joint-encoding model of Qian and Andersen (1997), where the perceived depth corresponds to the preferred disparities of the most active cells in the population. In this model, as the inter-flash interval decreases, this peak activity increases, which would be expected to lead to higher stereoacuity – not lower, as observed. In the disparity-averaging model, this effect is countered by the fact that, as inter-flash interval decreases, perception starts to reflect the average of two competing groups of neurons. This does not apply to the model of Qian and Andersen (1997).

Conclusion

The increase of threshold with interocular delay is straightforwardly predicted by our disparity-averaging model, and is hard to reconcile with joint encoding. Two

other aspects of the data can also be accounted for by disparity averaging, with some other plausible assumptions. Together, these data strengthen the case that depth perception in the stroboscopic Pulfrich effect is best explained without invoking joint encoding of motion and depth. The quantitative fits shown in Figure 8 depend on a number of assumptions and fitted parameters, and are therefore less compelling evidence than the prediction curves shown in Figure 6 and Figure 7, where even with no free parameters at all our model produces a very good account of the data. Nevertheless, this demonstrates that the stereoacuity data are consistent with disparity averaging. In contrast, neither perceived disparity nor stereoacuity can be well described by existing joint-encoding models.

Discussion

When a moving object is viewed with an interocular delay, it appears with an illusory depth. This effect, named after Carl Pulfrich, is straightforwardly explained in terms of the geometry of the stimulus: For a moving object, the delay causes a spatial displacement between the images of the object in each eye, which is naturally interpreted as depth. The same illusion has been reported when the object appears only intermittently, as if viewed under stroboscopic illumination. Here the illusion is no longer a simple consequence of stimulus geometry, but reflects the properties of the neural mechanisms supporting depth perception. A full understanding of the neuronal mechanisms responsible would produce a quantitative account of both the amount of depth perceived and the strength of the depth

percept as indicated by the acuity. Yet the stereoacuity for the stroboscopic Pulfrich effect has never been examined in detail, while the published literature contains conflicting reports regarding the magnitude of the perceived depth (Burr & Ross, 1979; Lee, 1970b; Morgan, 1979; Morgan & Thompson, 1975).

The need to resolve this discrepancy has become greater in recent years, following an upsurge of interest in the neuronal basis of depth perception in stimuli with interocular delay (Anzai et al., 2001; Carney et al., 1989; Morgan & Castet, 1995; Morgan & Fahle, 2000; Morgan & Tyler, 1995; Pack et al., 2003; Qian, 1997; Qian & Andersen, 1997). Several recent studies have explained this in terms of binocular neurons with space/time-inseparable receptive fields, which are sensitive to both disparity and to direction of motion. Indeed, many of these articles imply that this joint encoding of disparity and motion is the only viable explanation. The characteristic property of such joint disparity/motion sensors is that their preferred disparity changes as a function of interocular delay. Interocular delay therefore shifts the peak of activity in a population of such neurons, leading to an altered perception of depth. Qian and Andersen (1997) have implemented a realistic neuronal model of this scheme, and found that the perceived depth is predicted to be the virtual disparity, as reported by Burr and Ross (1979). To assess the validity of the modern joint-encoding explanation for Pulfrich-like phenomena, it is therefore important to understand whether the virtual disparity really is what is perceived in the stroboscopic Pulfrich effect. This was the motivation for the present study.

Our results strongly suggest that the strobe Pulfrich stimulus does *not* induce the same perception of depth as the classic Pulfrich effect. Using a stimulus in which tracking eye movements converted interocular delay into a real spatial disparity on the retina, subjects did perceive the virtual disparity, reproducing the results of Burr and Ross (1979). But when we removed the effect of eye movements (by applying interocular delay to the background as well as to the target), we found the sigmoid pattern reported by Morgan (1979), in which the perceived disparity was equal to the virtual disparity only for short inter-flash intervals, and falls well below the virtual disparity when the inter-flash interval is increased. This is at odds with the modern joint-encoding explanation of Pulfrich-like phenomena.

We therefore reexamined an earlier suggestion that the disparity experienced by subjects in the strobe Pulfrich illusion represents an average of the disparities physically present in the stimulus (Morgan, 1979;Tyler, 1974, 1977), after filtering by space/time-*separable* receptive fields. Because disparity tuning in real neurons becomes weaker when the stimulus is presented with an interocular delay, potential matches in which the left-eye and right-eye appearances are widely separated in time are expected to have less influence on perception than matches in which the two images occur nearly simultaneously. Quantifying these arguments, we were able to write down an equation that predicts the amount of disparity subjects should perceive at any stimulus interocular delay and inter-flash interval (Equation 2). Because this model explains the depth percept in terms of spatial disparities physically present in the stimulus, it could be implemented in the brain by "pure" disparity sensors, without the need for sensors that encode both motion and disparity jointly. The only parameter in the model is the binocular integration time. When this was set equal to the average binocular integration time of disparity-sensitive neurons in macaque V1 (Read & Cumming, 2005), the predictions of the equation were in good agreement with the psychophysics (Figure 6). The agreement became even better when we allowed for a small contribution from joint-encoding mechanisms (Equation 3), again guided by the known incidence of joint motion/disparity encoding in primate V1 (Pack et al., 2003; Read & Cumming, 2005).

Disparity averaging and joint encoding also make different predictions concerning disparity discrimination thresholds as a function of inter-flash interval and interocular delay. Disparity averaging straightforwardly predicts the main observation that thresholds increase with interocular delay, whereas this is not a property of joint-encoding models. Secondly, the rate of the increase with interocular delay was steeper for intermediate inter-flash intervals. This can be explained if perception reflects the average of disparities present in the stimulus, but it is not at all clear how this latter effect could be reconciled with the joint-encoding model. In the disparity-averaging model, the effect occurs because at shorter inter-flash intervals, the perception of zero depth at the nulling disparity reflects the averaging of far and near disparities in the stimulus, whereas at long inter-flash intervals, the stimulus effectively contains only zero disparity, resulting in a sharper perception. The third effect noticeable in our data concerned perception for stimuli with no interocular delay. Naturally, no depth was perceived here for any inter-flash interval. However, thresholds were significantly higher for the shortest inter-flash interval, despite the fact that here the target makes the most appearances per unit time and so the stimulus contains most power. This again is consistent with our disparity-averaging model. The Appendix shows how the disparity-averaging model, combined with a few realistic assumptions such as the signal-dependence of neuronal noise, yields predictions for disparity threshold as a function of interocular delay and inter-flash interval, which are in excellent agreement with the psychophysics (Figure 8). These observations on threshold values replicate for stereo the observations of Morgan and Watt (1983) for vernier acuity. We concur with their conclusion that such effects can be explained by spatiotemporal filters, and show that by using space/time separable filters we can produce a quantitative account of the major features in the data. Current models based on inseparable filters (joint encoding of motion and depth) would require modification to explain these data.

There are two main ways in which the model developed here is not complete. First, it is not a detailed physiological model such as that presented by Qian and Andersen (1997). Equation 2 predicts perceived disparity as the weighted average of each potential match between appearances of the target, but does not explain how the disparities of the potential matches might be extracted from the activity of early disparity sensors. Nevertheless, it seems plausible that something close to this model could be encoded in the activity of pure disparity sensors, which make up the majority of disparity-tuned neurons in V1. The temporal weight function would be implemented by the temporal filtering applied by these neurons. In this respect, our model is the same as previous explanations of interpolation based on spatiotemporal filtering. The difference is that the currently accepted model assumes that this spatiotemporal filtering must be applied by motion sensors (i.e., with space/time-inseparable receptive fields), whereas our results suggest that space/time-separable, non–direction-selective filters would suffice.

Second, our model currently applies only to the stroboscopic Pulfrich effect. A great strength of the joint-encoding model is that it is able to explain the existence of depth perception in a wide variety of stimuli with an interocular delay (Qian, 1997; Qian & Andersen, 1997). For example, it can explain the perception of depth in dynamic noise viewed with an interocular delay (Falk & Williams, 1980; Ross, 1974; Tyler, 1974, 1977). Earlier studies suggested that depth perception here could also be explained in terms of spatial disparities present in the stimulus (Tyler, 1974, 1977). Although no explicit models have been implemented, there has been no refutation of this idea in principle. Therefore it is not necessary to invoke joint encoding of disparity and motion to explain depth perception in dynamic noise; it may be possible to explain the effects of this stimulus, too, in terms of the activity of pure disparity sensors. We are currently developing a neuronal population model to explore responses to both stroboscopic and dynamic noise stimuli. If it does prove possible to explain all forms of the Pulfrich effect with space/time separable spatiotemporal filters, it seems likely that the classical form of the Pulfrich effect is also best explained in this way. Under continuous illumination, this percept is dominated by instantaneous disparities generated on the retina, much as originally described by Fertsch (Pulfrich, 1922).

Conclusion

We have shown that spatio-temporal filtering in pure disparity sensors (space-time separable filters) provides an excellent quantitative account of perceived depth in the stroboscopic Pulfrich effect. The depth primarily reflects the average of disparities present in the stimulus. Most weight is given to stereo matches between appearances of the target that occur simultaneously, or nearly so, in both eyes. Less weight is given to potential matches in which

there is a delay between the target's appearance in each eye. This fall-off in weight is well described by a Gaussian function with a standard deviation of ~15 ms, and probably reflects the binocular integration time of disparity-tuned neurons in V1. Consequently, the perception of depth in Pulfrich-like stimuli cannot be taken as evidence that space-time inseparable filters play a special role. Indeed, the results we present here, and those of Morgan (1979), are at odds with current models based on such filters (Qian & Andersen, 1997). These data therefore provide clear evidence against currently published interpretations based on joint encoding of motion and depth. We note, however, that these data do not exclude all possible models based on joint encoding: They only exclude joint encoding in its currently published forms. Indeed, it seems inevitable that one could devise a rule for combining the outputs of directional spatiotemporal filters that exactly reproduces the information available in nondirectional filters, rendering the two hypotheses indistinguishable. Our main contribution is to point out that, while the stroboscopic Pulfrich effect may be compatible with joint encoding of motion and disparity, it is not – contrary to the commonly made claim – evidence for it.

Appendix: Derivation of predicted disparity threshold

Here we derive an expression for the dependence of disparity threshold on interocular delay and inter-flash interval, subject to some reasonable assumptions. For simplicity, we ignore the contribution from joint encoding, and assume that the perception of depth relies entirely on the activity of pure disparity sensors. According to the disparity-averaging model of Equation 2, the perceived disparity is simply the average of all the disparities present in the stimulus, weighted by the time separating the two members of the match. A very simple model of how this could be achieved in the brain is to imagine that a population of pure disparity sensors each feeds into an output "perception" neuron. Three assumptions are needed to make this system implement Equation 2: (1) The activity of each sensor decreases monotonically with the magnitude of the match's temporal separation; (2) the synaptic weight given to each sensor reflects the sensor's preferred disparity; and (3) the activity of the output neuron is normalized by the total activity of the population. Then the activity of the output neuron simply encodes the perceived disparity according to the disparity-averaging model:

$$\Delta x_{perc} = \frac{\sum_k \Delta_k \, r_k}{\sum_k r_k} \tag{4}$$

where the sum k is over all members of the population; r_k represents the firing rate of the kth sensor and Δ_k its preferred disparity.

Because sensors are affected by noise, the firing rate r includes some noise. In real neurons, the variance of firing is not constant, but increases with the mean firing rate (Dean, 1981). We model this by taking the noise on each disparity sensor to be a Gaussian random variable with variance equal to ($b_k + c\, r_k^{\,p}$), where b_k is the sensor's baseline noise level, Δ_k is its preferred disparity, r_k is its mean firing rate for this stimulus, and c, p, d are constants (the same for all sensors and stimuli). We assume that $p > 0$, so that the second term in the variance represents noise that increases with neuronal firing.

As we shall see below, for the strobe Pulfrich stimulus, the noise increases with the preferred disparity of the sensor, so it will have more effect on the denominator of Equation 4, where terms are weighted according to preferred disparity, than on the numerator. We therefore neglect the noise in the numerator of Equation 4. The denominator is the sum of many Gaussian random variables; its variance V is given by the sum of the variances of each variable:

$$V = \sum_k \Delta_k^{\,2}\left(b_k + c r_k^{\,p}\right) = \sum_k \Delta_k^{\,2} b_k + c\sum_k \Delta_k^{\,2} r_k^{\,p} \ .$$

The first term is a constant representing the effect of the baseline noise from all neurons in the population; it is independent of the stimulus, and we lump it together into a new constant B^2. In contrast, the second term depends on the stimulus; it is contributed to only by those disparity sensors that are activated by a particular stimulus, because $r_k = 0$ for the other sensors. For a strobe Pulfrich stimulus presented with a real spatial disparity Δx_{expt}, as in our nulling experiments, the most active disparity sensors are those tuned to the stimulus disparity Δx_{expt} itself. However, because of the periodic nature of the stimulus, sensors tuned to Δx_{expt} plus or minus a whole number of strobe inter-flash distances X are also active, although more weakly because the left-eye and right-eye members of the matches they are responding to are separated by longer in time. Thus, the active sensors are those for which $\Delta_k = (jX + \Delta x_{\text{expt}})$, and for these $r_k = w(jT + \Delta t)$, where w is the weight function describing how the activity of disparity sensors decays as a function of interocular delay (Equation 1). Note that this formulation implicitly assumes that each sensor responds only to its own preferred disparity. A more realistic model would allow neurons to respond less strongly to disparities close to its preferred disparity, implementing spatial as well as temporal filtering. This would be important for tracking the response of the population as the inter-flash interval decreased toward the continuous limit, when the stimulus would contain many closely spaced disparities. However, because in our stimuli

the disparity spacing X is at least $0.1°$, the simplification is acceptable. The variance V on the denominator of Equation 4 therefore reduces to

$$V = B^2 + c\sum_j \left(jX + \Delta x_{\text{expt}}\right)^2 w^p\left(jT + \Delta t\right),$$

and the perceived disparity itself is a Gaussian random variable with mean

$$\left\langle \Delta x_{perc} \right\rangle = \frac{\sum_j \left(jX + \Delta x_{\text{exp}t}\right).w\left(jT + \Delta t\right)}{\sum_j w\left(jT + \Delta t\right)} \qquad (5)$$

and standard deviation

$$\sigma = \frac{\sqrt{B^2 + c\sum_j \left(jX + \Delta x_{\text{expt}}\right)^2 w^p\left(jT + \Delta t\right)}}{\sum_j w\left(jT + \Delta t\right)} \ . \qquad (6)$$

We assume that, in our psychophysical experiments, subjects answer "far" or "near" according to whether the random variable Δx_{perc} is positive or negative. In our nulling paradigm, we find how much disparity it is necessary to apply to make the subject answer "near" as often as "far" (i.e., the value of Δx_{expt} for which $\langle\Delta x_{\text{perc}}\rangle=0$). Equation 5 shows that, as required, this nulling disparity is minus the perceived disparity given in Equation 2. We are now finally in a position to extract the disparity threshold θ. This is the amount by which disparity must be increased beyond the nulling value to make the subject answer "far" 84% of the time. But this is just the standard deviation of the perceived disparity when the stimulus disparity Δx_{expt} is $\Delta x_{\text{null}}+\theta$. For simplicity, we shall ignore the small change in standard deviation caused by changing the stimulus disparity from Δx_{null} to $(\Delta x_{\text{null}}+\theta)$, and take the disparity threshold to be the standard deviation of the perceived disparity when the stimulus disparity Δx_{expt} is Δx_{null}. Thus, finally, the disparity threshold is approximately

$$\theta = \frac{\sqrt{B^2 + c\sum_j \left(jX + \Delta x_{\text{null}}\right)^2 w^p\left(jT + \Delta t\right)}}{\sum_j w\left(jT + \Delta t\right)} \qquad (7)$$

where

$$\Delta x_{null} = -\frac{\sum_j jX.w\left(jT + \Delta t\right)}{\sum_j w\left(jT + \Delta t\right)} \ .$$

The parameter B represents the value of the threshold when the interocular delay is zero and the inter-flash interval T is long compared to the integration time (to see this, note that under these conditions, the $j \neq 0$ terms in the sums in Equation 7 become negligible).

Note that the dependence of noise on signal, encoded by the parameters c and p, is critical in making the threshold at zero interocular delay increase as the inter-flash interval is reduced, in accordance with the psychophysics. If the noise level were independent of neuronal activity, $c = 0$, then Equation 7 for zero delay would reduce to $\theta = B/\Sigma_j w(jT)$. As the inter-flash interval T shortens, the sum in this expression increases, and so threshold reduces. If the noise were entirely dependent on signal ($B = 0$), then the threshold would be zero for large T, and would increase above zero as T decreased. Thus, the dependence of the noise on neuronal firing is critical in enabling the model to reproduce the increase in zero-delay threshold as inter-flash interval decreases.

The curves in Figure 8 show the disparity thresholds predicted by Equation 7 with suitable parameters. The parameters needed are B, c, p, and d, plus the integration time τ implicit in the function w (Equation 1). The value of τ was chosen by fitting the perceived disparity for all subjects simultaneously ($\tau = 21$ ms, purple curve in Figure 6). Thus, τ was set without reference to the threshold data shown in Figure 8, and was the same for all subjects. The value of p was set equal to 1.5. This was designed to capture experimental findings that the variance of neuronal spike counts increases in proportion to the mean. The exponent of 1.5 is a little higher than experimental evidence suggests, but was chosen because it gave a better fit to the threshold data. Again, this value was the same for all subjects. The remaining parameters B, c, and d were fitted to the observed thresholds, for each subject independently. B comes from the baseline noise in neuronal firing; as noted above, it sets the minimum value of the threshold for long inter-flash intervals and no interocular delay. B was 8 arcsec for HN, 11 arcsec for BC, and 32 arcsec for JR. c represents the amount of signal-dependent noise relative to baseline; c was 0.043 for HN, 0.028 for BC, and 0.326 for JR (the units of c are not meaningful because they are in terms of the "firing rate" of the notional sensors).

Acknowledgments

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*(2), 284-299. [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*(5), 513-518. [PubMed]

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology, 211*(3), 599-622. [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]

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]

Dean, A. F. (1981). The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research, 44*(4), 437-440. [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., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neuroscience, 18*(10), 451-458. [PubMed]

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]

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

Julesz, B., & White, B. (1969). Short-term visual memory and the Pulfrich phenomenon. *Nature, 222*(194), 639-641. [PubMed]

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

Lee, D. N. (1970b). A stroboscopic stereo phenomenon. *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.

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., & 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., & Turnbull, D. F. (1978). Smooth eye tracking and the perception of motion in the absence of real movement. *Vision Research, 18*(8), 1053-1059. [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 B, 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). Effect of motion sweep duration and number of stations upon interpolation in discontinuous motion. *Vision Research, 22*(10), 1277-1284. [PubMed]

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

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America, 43*(10), 907-913. [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). The neural basis of the Pulfrich effect in the monkey. Program No. 339.7. *Abstract Viewer/Itinerary Planner.* Washington, DC: Society for Neuroscience.

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

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

Ross, J. (1976). The resources of binocular perception. *Scientific American, 234*(3), 80-86. [PubMed]

Ross, J., & Hogben, J. H. (1974). Short-term memory in stereopsis. *Vision Research, 14*(11), 1195-1201. [PubMed]

Ross, J., & Hogben, J. H. (1975). Letter: The Pulfrich effect and short-term memory in stereopsis. *Vision Research, 15*(11), 1289-1290. [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]

Rushton, D. (1975). Use of the Pulfrich pendulum for detecting abnormal delay in the visual pathway in multiple sclerosis. *Brain, 98*(2), 283-296. [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]

Ward, R., & Morgan, M. J. (1978). Perceptual effect of pursuit eye movements in the absence of a target. *Nature, 274*(5667), 158-159. [PubMed]

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*(2), 113-120. [PubMed]