Target tracking reveals the time course of visual processing with millisecond-scale precision

Johannes Burge[1.,2.,3.,*] & Lawrence K. Cormack[4.,5.,6.]

[1.] Department of Psychology, University of Pennsylvania
[2.] Neuroscience Graduate Group, University of Pennsylvania
[3.] Bioengineering Graduate Group, University of Pennsylvania
[4.] Department of Psychology, University of Texas at Austin
[5.] Center for Perceptual Systems, University of Texas at Austin
[6.] Institute for Neuroscience, University of Texas at Austin

[*] email for correspondence: jburge@psych.upenn.edu

Abstract

Image differences between the eyes can cause millisecond-scale interocular differences in processing speed. For moving objects, these differences can cause dramatic misperceptions of distance and 3D direction. Here, we develop a continuous target-tracking paradigm for measuring these processing differences. Human observers continuously tracked a target stimulus with various luminance differences across the eyes as it underwent Brownian motion in the horizontal plane. We show that suitable analysis recovers the time course of the visuomotor response, and comparisons across conditions reveal the temporal evolution of visual processing differences between the eyes. Next, we show analytically, and partially confirm experimentally, that the interocular difference between the temporal impulse response functions predicts how lateral target motion impacts the motion-in-depth response. Finally, using a direct within-observer comparison, we show that target tracking and traditional psychophysics provide scalar estimates of interocular delays that agree on average to within a fraction of a millisecond. Thus, target tracking accurately recovers millisecond-scale differences in processing speed while revealing the millisecond-by-millisecond time course of visual processing, all in a fraction of the time required by traditional methods. This paradigm provides the potential for new predictive power, the application of analytical techniques from computational neuroscience, and the rapid measurement of clinical and developmental populations in which traditional psychophysics might be impractical.

Keywords

motion-in-depth, binocular vision, Pulfrich effect, temporal impulse response, continuous psychophysics, temporal dynamics

Introduction

The binocular visual system combines spatial and temporal information from the eyes to estimate the structure of the three-dimensional environment, the three-dimensional motion of objects in the environment, and self-motion through the environment. A large body of research has focused on how spatial differences in the left- and right-eye images (i.e. binocular disparities) drive the estimation of 3D structure and motion [1-11]. A smaller but still substantial body of research has investigated how temporal differences in how the left- and right-eye images are processed impact the estimation of 3D motion [12-20]. Despite long-standing interest in the temporal properties of visual processing [21], there have been few psychophysical attempts to measure the time course of binocular visual processing for the perception of motion in depth [22]. Traditional psychophysical data collection techniques, which require an observer to view a series of individual trials and to respond to each trial with a binary choice, tend not to reveal continuous information about visual processing. Although recovering such information is possible in principle, traditional techniques are rather slow—they typically produce approximately no more than thirty data points per minute—so collecting sufficient data to recover time-course information is impractical. More generally, because thousands of data points are often required from each observer in any given study, the slow pace of data collection tends to place strong constraints on the scope of the experimental questions that can be practically addressed. Alternative techniques that yield information about temporal dynamics and that yield more data per unit time have the potential to assist vision science in the future.

Continuous target tracking, a new stimulus-response data collection technique, simultaneously provides information about the time course of visual processing and improves the rate of data collection [23]. With this technique, a property of a stimulus (e.g. position) is estimated or tracked via an input device, such as a computer mouse. There are many advantages of target tracking psychophysics. First, as we will show, target tracking and traditional button-press psychophysics make near-identical estimates of temporal processing delays. Second, target tracking provides time course information that button-press psychophysics typically does not; this additional information provides a richer picture of the temporal characteristics of the system and can lead to enhanced predictive power (see Discussion). Third, target tracking provides a direct continuous measure of behavioral performance; other non-invasive continuous measures (e.g. EEG) provide more direct measures of neural activity, but must be related to performance via other means. Fourth, target tracking is comparatively low-tech and does not need specialized equipment other than a display screen and a mouse. (Note that if target tracking in depth is desired, then a stereo-capable display is also needed). Fifth, target tracking requires little instruction or practice, and is therefore suitable for collecting large amounts of high-quality data from virtually any observer; this last advantage may make target tracking particularly advantageous for work with developmental, clinical, and otherwise non-traditional observer populations.

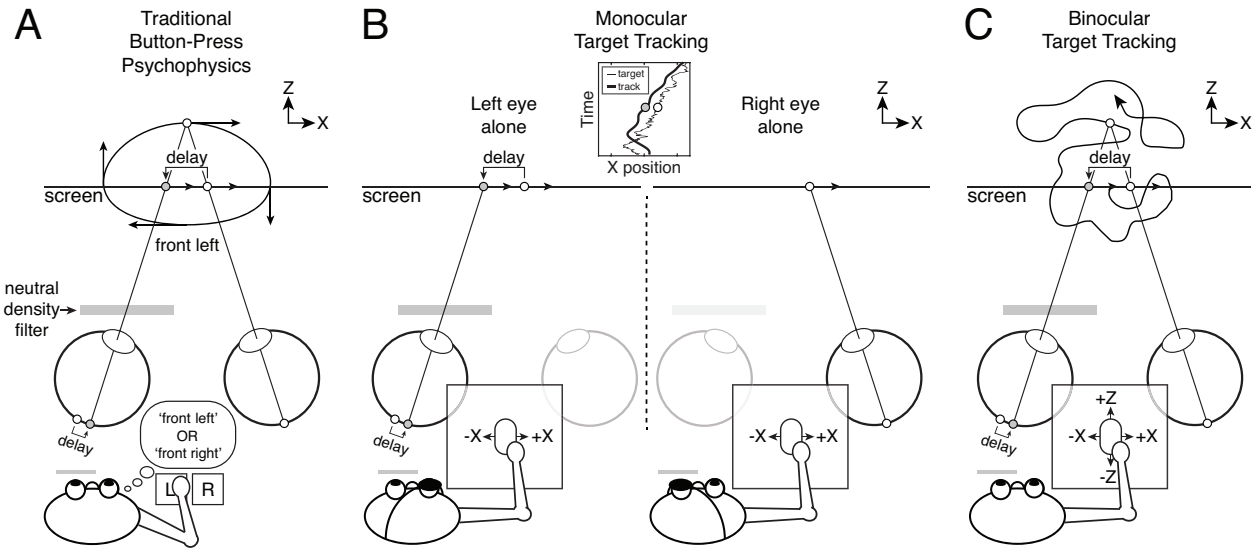


Figure 1. The classic Pulfrich effect, and three approaches to measuring interocular delays. **A** Classic Pulfrich effect. When a moving target is viewed with an unequal amount of light in the two eyes, the distance and 3D direction of horizontal motion is perceived incorrectly. With a neutral density filter in front of the left eye, sinusoidal target motion in the screen plane is misperceived as motion along a near-elliptical trajectory in depth. The image in the darker eye is processed with a delay relative to the brighter eye. For rightward motion, this interocular delay causes the effective target image position in the darker eye (gray dot) to be shifted leftward relative to the target image position in the brighter eye (white dot). For leftward motion, the target image position in the darker eye is shifted rightward (not shown). The binocular visual system computes the disparity from these effective left- and right-eye images, and the target is perceived behind the screen for rightward motion (and in front of the screen for leftward motion). In a traditional psychophysical experiment, observers report their percept (e.g. 'front left' or 'front left') with a button press. This method recovers estimates of interocular delay with sub-millisecond precision. **B** Monocular target tracking. A target undergoing a random walk in X on the screen plane is tracked when it is viewed with the left eye alone and also when it is viewed with the right eye alone. Although the target is always perceived in the plane of the screen (i.e. no illusory depth is perceived), comparing monocular tracking performance between the eyes can yield estimates of interocular delay that match those obtained with traditional psychophysics, assuming matched viewing conditions. **C** Binocular target tracking. A target undergoing a random walk in X and Z is tracked while being viewed with both eyes. When the left eye is dark, the interocular delay causes a target moving rightward to be perceived farther away than it is and vice versa. Analyzing binocular tracking performance can also reveal interocular differences in processing.

Here, to investigate the utility of target tracking for measuring the time course of visual processing, we make use of the Pulfrich effect, a well-known stereo-motion phenomenon [12,13]. When the image in one eye is darker than the image in the other, motion in the frontal plane like that of a clock pendulum is misperceived as near-elliptical motion in depth (Fig. 1A). The effect occurs because the image with less light is processed more slowly. For moving objects, the interocular mismatch (i.e. delay) in processing speed causes an effective neural disparity, which leads to illusory percepts of depth. With traditional psychophysical techniques, interocular delays can be measured with sub-millisecond precision by having observers report with a button press whether the stimulus appeared to move leftward or rightward when it appeared to be in front of the screen ("front left" in Fig. 1A). The Pulfrich effect therefore provides a stringent test for target tracking psychophysics. It is also convenient for present purposes because interocular delays of only a few milliseconds can cause large perceptual effects.

The goal of this paper is to determine whether continuous target tracking can be used to quantify millisecond-scale differences in visual processing between the eyes. We estimate interocular differences in visual processing with traditional button-press psychophysics (Fig. 1A), monocular target tracking (Fig. 1B), and binocular target tracking in depth (Fig. 1C), and then compare the estimates. We also prove a novel relationship between the temporal evolution of

left- and right-eye processing differences and key properties of binocular target tracking in depth, and then examine whether processing differences revealed by monocular target tracking in the frontoparallel plane predict these properties of tracking in depth. We find that target tracking has exquisite temporal sensitivity, provides estimates that are matched to those provided by traditional psychophysics with millisecond accuracy, and holds promise for a range of clinical and scientific applications.

Results

Monocular target-tracking psychophysics

First, we estimated the visual processing delays caused by luminance reductions with target tracking under monocular viewing conditions. Human observers tracked a white vertical target bar undergoing a random walk in X (i.e. horizontally) with a small mouse cursor dot (Fig. 2A). The task was to follow the target as accurately as possible. Tracking was performed monocularly under six viewing conditions: left eye alone and right eye alone at three different luminance levels each (see Methods). The task was performed without difficulty; the human response was a smoothed and delayed approximation of the target motion (Fig. 2B).

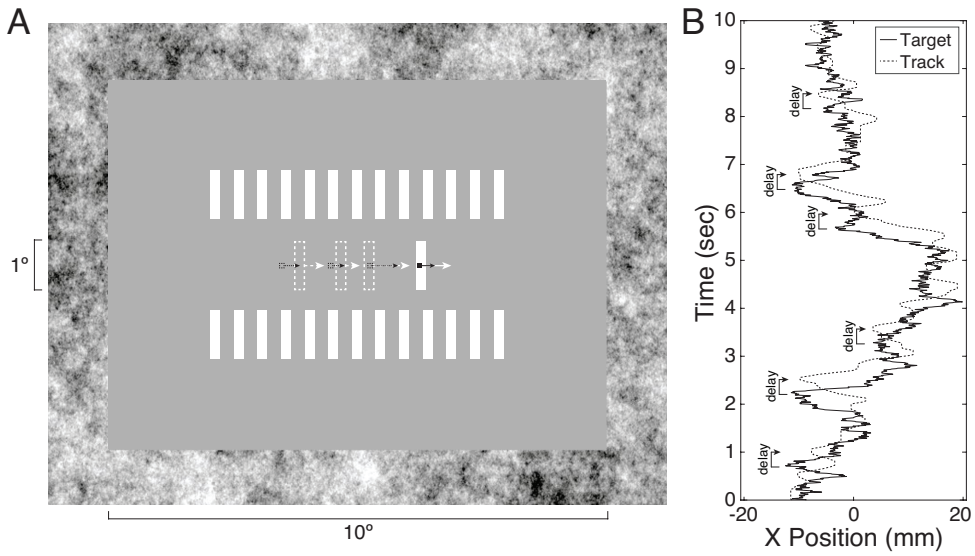


Figure 2. On-screen display and monocular x-position tracking performance. **A** Target tracking stimulus. The target bar underwent a horizontal random walk. The task was to track the target bar with a small dark mouse cursor. Motion direction and speed are indicated by arrows and dashed shapes; note that they are for illustrative purposes and were not present in the actual stimuli. **B** Target trajectory in X (solid curve) and tracked trajectory (dashed curve) across time. The response trajectory is a smoothed, delayed version of the target trajectory. Note that the delay in the human tracking response is approximately constant throughout the trial.

Monocular cross-correlograms—the cross-correlation of the target and response motions—are shown for all observers in the highest and lowest luminance conditions (Fig. 3A). The cross-correlogram yields an estimate of the temporal impulse response function of the visuo-motor system, assuming the system is linear. The latency of the initial response, as quantified by the first point that the cross-correlogram rises out of the noise, ranges between 150-200ms. The rise is steep such that the peak correlation occurs 50-75ms after the cross-correlation becomes significantly different from zero. The temporal integration period, as quantified by the bandwidth (i.e. full-width at half-height), ranges between 100-300ms across observers.

Importantly, the visuomotor response of the observer in any given condition will be affected by the characteristics of the visual system, the motor system, and the input device. However,

because the properties of the motor system and the input device should remain constant across visual conditions within a given observer, comparisons across visual conditions should reveal the associated changes in visual processing. On quick examination, the monocular cross-correlograms appear to be essentially identical for the high- and low-luminance conditions. However, closer examination reveals a small but systematic shift between the cross-correlograms in the two luminance conditions. This shift is most clear and consistent in the rising edges of the cross-correlograms (Fig. 3A, insets). The visuomotor responses in the high luminance conditions are faster by several milliseconds than the responses in the low luminance conditions. Thus, assuming that the dynamics of the motor system itself are unchanged by the visual stimulus, the results imply that the speed of visual processing is faster in high than in low luminance conditions.

Figure 3B shows these differences in processing speed (i.e. delays) as a function of the luminance difference (i.e. optical density difference; see Methods) in the left and right eyes. (Note that we use the term 'delay' loosely to refer to any change in temporal processing in the more slowly processed eye that shifts the effective position of a moving target to lag that in the other eye.) We estimated the delays by computing the cross-correlation of the cross-correlograms in two luminance conditions (e.g. high luminance vs. low luminance). This method makes use of all of the raw data, is robust, and is relatively unladen by assumptions. Across all observers, lower luminance causes slower visuomotor processing. In the conditions with the largest luminance differences, the delays ranged between 4 and 10 milliseconds. Interestingly, the estimated delays reported in Fig. 3B are in general agreement with previous psychophysical and neurophysiological investigations of the Pulfrich effect, which have shown that every log-unit reduction of luminance reduces the speed of processing by approximately 10 milliseconds [15,17]. Moreover, the fact that these results are systematic and regular suggests that continuous target tracking is sensitive enough to estimate relative delays on the order of milliseconds.

The ultimate question, however, is whether these measured differences in visuomotor processing speed reflect differences in *visual* processing speed that are associated with stimuli of different luminances. To answer this question, we examined how estimates of processing speed differences based on target tracking are related to estimates of processing speed differences obtained with traditional psychophysical techniques. With these techniques, observer responses do not depend at all on the dynamics of the motor system, so a quantitative comparison of delays in the two paradigms is a stringent test of whether target tracking can be used to measure millisecond-scale differences in visual processing.

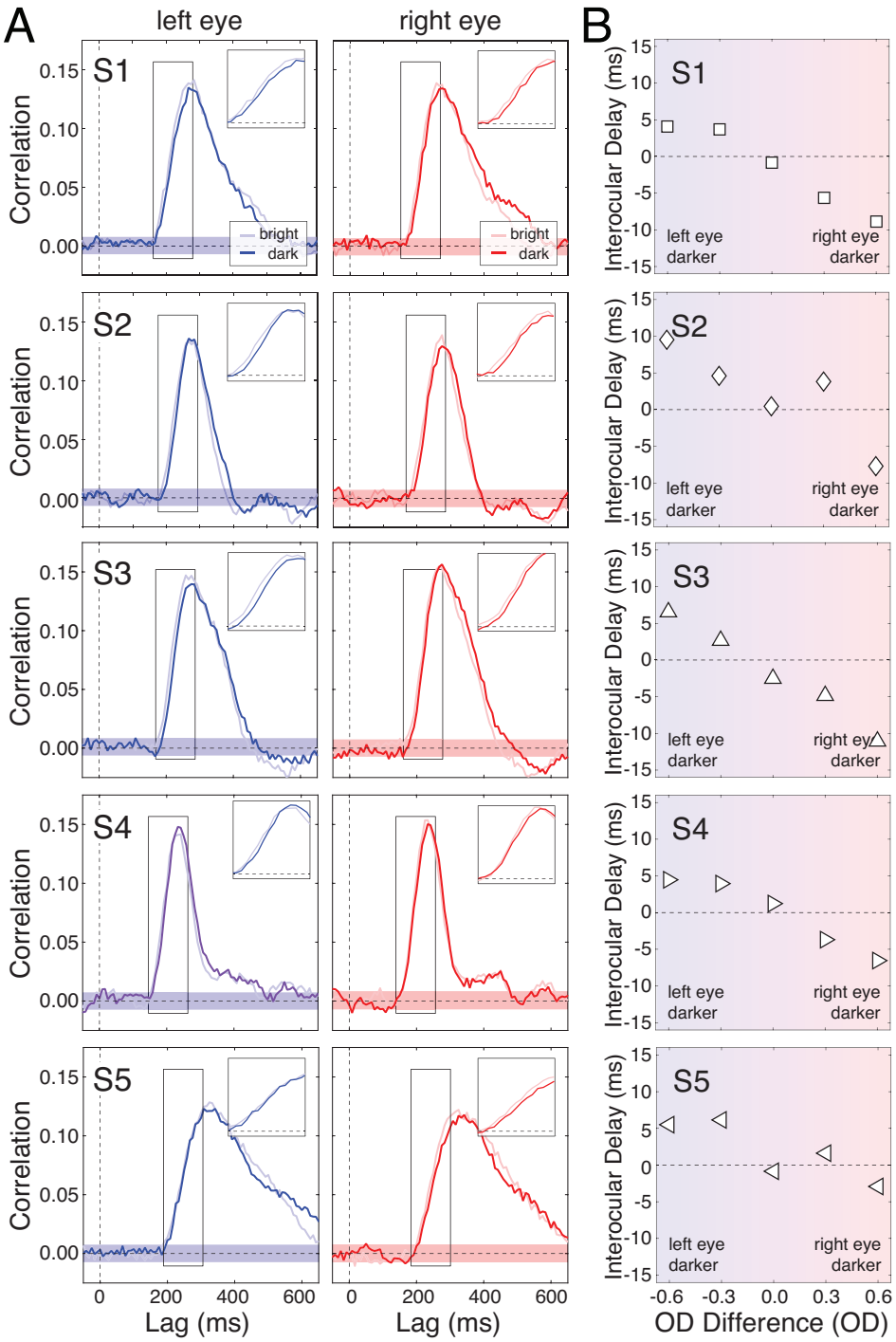


Figure 3. Monocular target tracking results. Monocular temporal cross-correlograms and estimated interocular differences in processing speed at different luminance levels. **A** Monocular target tracking in X for all observers. The cross-correlograms reflect target tracking performance when the image was bright (i.e. maximum luminance: OD=0.0; light curve) and when the image was dark (i.e. 75% less light than maximum luminance: OD=0.6; dark curve) for the left eye (blue) and the right eye (red). Although the cross-correlograms are nearly overlapping, the dashed curve is shifted rightward by a small but consistent amount, indicating an increased tracking lag for the darker image. Insets show the systematic delay between the rising edges of the cross-correlograms. The colored band represents +1SD of the baseline noise, computed from lags less than 0ms. **B** Interocular tracking delays as a function of optical density difference in the two eyes ΔO for maximum luminance in the right eye vs. 75% and 50% less light in the left eye (ΔO=-0.6 and -0.3) and maximum luminance in the left eye vs. 50% and 75% less light in the right eye (ΔO=+0.3 and +0.6). Interocular delay is an approximate linear function of the optical density difference between the eyes.

Traditional button-press psychophysics

To directly compare estimates of processing speed from tracking and traditional psychophysics, we used a standard paradigm to measure interocular delays associated with the Pulfrich effect [20]. The luminance levels in each eye were matched to those used in the monocular tracking experiment. The primary differences between the experiments were the motion trajectories followed by the stimulus and the method of response. Rather than a random walk, the target stimulus followed a sinusoidal trajectory on each eye's monitor (Fig. 4ABC). Rather than continuously tracking the target with a cursor, observers viewed the target and then made a binary response reporting an aspect of the 3D motion percept.

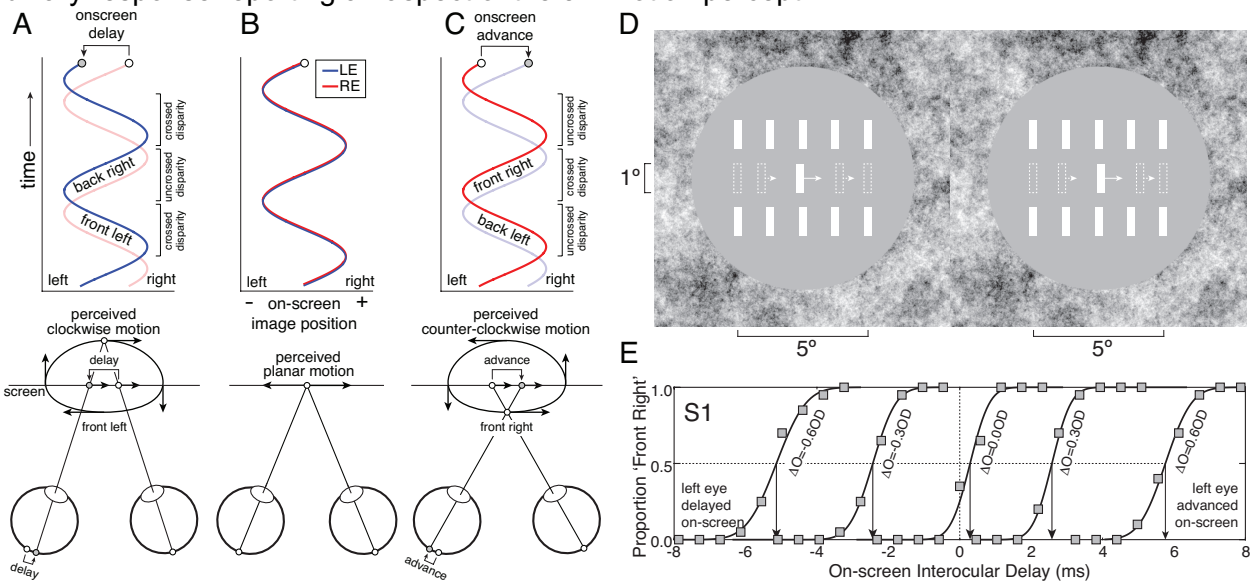


Figure 4. Binocular target stimulus, on-screen positions, and disparity-specified 3D target trajectories for button-press psychophysics experiment. **A** On-screen left- and right-eye image positions over time. When the left-eye image is delayed on-screen relative to the right-eye image (top), disparity specifies clockwise motion (i.e. 'front left' motion) when viewed from above (bottom). **B** When no on-screen delays are present (top), disparity specifies motion in the plane of the screen (bottom). **C** When the left-eye image is advanced on-screen relative to the right-eye image, disparity specifies counter-clockwise motion (i.e. 'front right' motion) when viewed from above (bottom). **D** Binocular target stimulus. Arrows and dashed bars show motion direction and speed; they are for illustrative purposes and were not present in the actual stimuli. Free-fuse to see stimulus to see in 3D. Cross-fusers will see a depiction of 'front right' 3D motion, which is consistent with counter-clockwise motion when viewed from above. Divergent-fusers will see a depiction clockwise motion when viewed from above. **E** Example psychometric functions from the first human observer for interocular differences in optical density (i.e. ΔO={-0.6OD, -0.3OD, 0.0OD, +0.3OD, +0.6OD}) corresponding luminance differences ranging from the left eye having 75% less light than the right eye to the right eye having 75% less light than the left eye. To cancel the neural delays, the required on-screen interocular delays (arrows) change systematically with the luminance differences.

In most conditions, the stimulus appeared to be following a near-elliptical motion trajectory that exited the plane of the screen. The task was to report whether the target appeared to be moving leftward or rightward when it appeared to be in front of the screen. To change the stereoscopically-defined 3D motion trajectory, on-screen interocular delays were manipulated (see Methods). When the left-eye image was delayed relative to the right-eye image, disparity specified that the target was undergoing 'back right / front left' motion (i.e. clockwise motion when viewed from above; Fig. 4A). When there was no delay, disparity specified that the target was moving in the plane of the screen (Fig. 4B). And when the left-eye image was advanced relative to the right-eye image, disparity specified that the target was undergoing 'front right / back left' motion (i.e. counter-clockwise motion when viewed from above; Fig. 4C). Nearly all aspects of the display—the size and shape of the target bar, the size and shape of the picket fence reference bars, the luminance levels, the peripheral 1/f noise—were identical to those

presented in the tracking experiment (Fig. 4D). The only differences were the shape of the mean luminance gray region (circular vs. rectangular) and the number of picket fence bars (five vs. thirteen).

For a given interocular luminance difference, we measured a psychometric function: the proportion of trials that observers reported 'front right' as a function of the on-screen interocular delay (Fig. 4E). The goal of the experiment was to find, for each luminance difference, the on-screen delay that made the target appear to move in the plane of the screen: the point of subjective equality (PSE). Each PSE should be equal in magnitude and opposite in sign to the neural delay caused by the luminance difference between the eyes in each visual condition. We found that the on-screen delays required to null the neural delays (i.e. the PSEs) change systematically with the associated luminance differences in all five observers (Fig. 5).

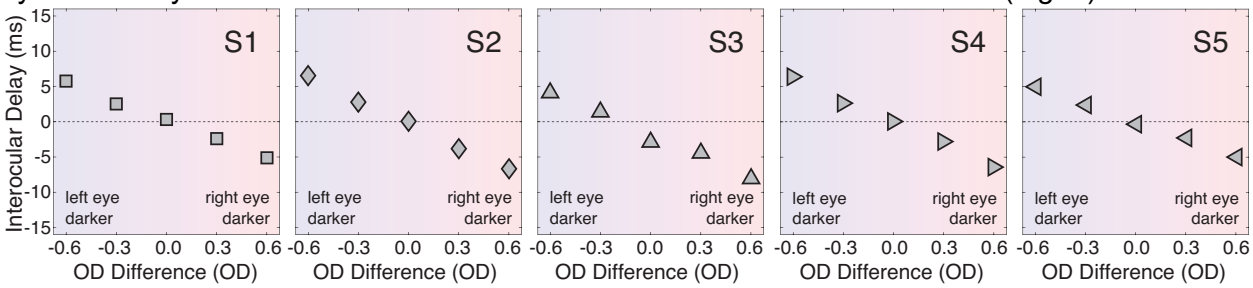


Figure 5. Traditional button-press psychophysics results. On-screen interocular delays that resulted in percepts of zero motion in depth as a function of the interocular difference in optical density (ΔO ={-0.6OD, -0.3OD, 0.0OD, +0.3OD, +0.6OD}). These optical density differences correspond to the left-eye stimulus having 75% and 50% lower luminance than the right-eye stimulus, the stimuli in both eyes having the same luminance, and the right-eye stimulus having 50% and 75% lower luminance than the left-eye stimulus, respectively. Positive on-screen interocular differences indicate that the left-eye image was delayed neurally relative to the right. Negative on-screen interocular differences indicate that the left-eye image was advanced neurally relative to the right.

To examine whether target tracking and traditional button-press psychophysics are measuring the same underlying quantity, we plotted the estimated interocular differences in processing speed from the two experiments (see Figs. 3B & 5) against the luminance difference in the two eyes (Fig. 6A). The agreement between the two methods is good.

To quantitatively assess the agreement, we plotted the estimates from the traditional button-press psychophysics experiment directly against the estimates from the tracking experiment. The data are tightly clustered about the unity line, indicating millisecond-scale agreement between estimates of interocular delay provided by the two experiments (Fig. 6B). Across conditions and observers, the differences between the estimated delay with traditional and tracking psychophysics were very small. The mean difference in delay was -0.16ms (-1.04ms to 0.71ms 95% confidence interval) with a standard deviation of 2.06ms (Fig. 6B, inset).

The level of agreement between the estimates derived from the two methods is striking, especially given the enormous differences between them. One method—traditional button-press psychophysics—presented a target stimulus following a stereotyped motion trajectory (i.e. a near-elliptical path through a 3D volume of space) and obtained a binary response. This binary response reflected an aspect of the observer's percept, and the response is essentially independent of the temporal properties of the motor system. The other method—target-tracking psychophysics—presented a target stimulus following an unpredictable motion trajectory (i.e. a random walk in the 2D plane of the display monitor) and obtained the continuous motor response of the observer. This continuous response is fundamentally constrained to reflect the temporal properties of both the visual and motor systems. And yet, the estimates of interocular

delay from the two methods agree in each individual condition to within a few milliseconds, and on average to within a fraction of a millisecond. The fact that substantially different stimuli and substantially different experimental paradigms yield near-identical estimates of interocular delay strongly suggests that both paradigms are measuring the same underlying quantity.

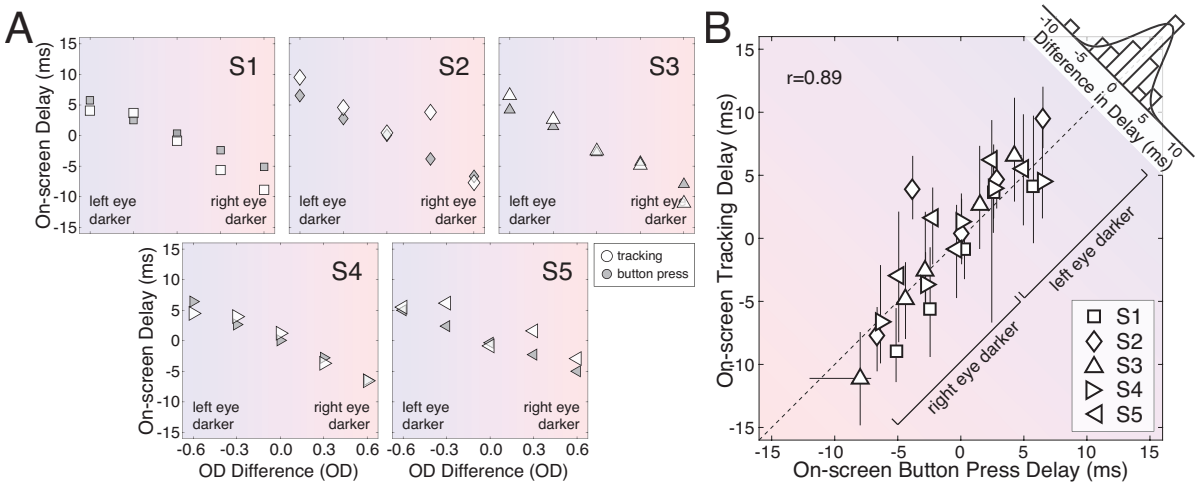


Figure 6. Interocular delays from target tracking vs. traditional button-press psychophysics. **A** Interocular delay as a function of interocular difference in optical density as measured with target tracking (white symbols) and button-press psychophysics (gray symbols) for individual observers in each of five different conditions (i.e. ΔO ={-0.6,-0.3,0.0,+0.3,+0.6}). Positive on-screen delays indicate that the left eye was processed more slowly. **B** Interocular delays measured with target tracking plotted against interocular delays measured with button-press psychophysics in equivalent conditions. Red tint indicates conditions in which the right eye is darker. Blue tint indicates conditions in which the left eye is darker. Error bars indicate 68% bootstrapped confidence intervals; if error bars are not visible, they are smaller than the size of the data symbol. The histogram shows the distribution of differences between the button-press delays and the tracking delays; the mean difference was only -0.16ms (SD=2.06ms).

Binocular target-tracking psychophysics

The preceding experiments show that monocular target tracking and traditional button-press psychophysics yield similar estimates of millisecond-scale processing differences between the eyes. Does binocular target tracking afford similarly precise measurements of interocular differences in visual processing? In this experiment, human observers binocularly viewed and tracked a target bar with a small cursor undergoing a random walk in the XZ plane (see Fig. 1C). The visual conditions were matched to a subset of those used in the monocular tracking experiment: either both images were equally bright or the image for one eye was substantially darker than the other (i.e. ΔO={-0.6OD,0.0OD,0.6OD}). First, we examine the cross-correlation of the target and response velocities when both eyes were equally bright. Then, we examine the results with unequal brightness in the two eyes.

Binocular tracking of horizontal (i.e. X) target motion is similar to monocular tracking of horizontal motion (Fig. 7; X vs. X). The latency of the initial response ranges between 150-200ms across observers, and the temporal integration period ranges between 100-300ms. Binocular target tracking in depth is uniformly more sluggish (Fig. 7; Z vs. Z). In each observer, the latency of the initial response in Z occurs approximately 50ms later than the initial response in X, and the period of temporal integration in Z is nearly double the temporal integration period in X. These results hold when both eyes have the same luminance (Fig. 7A), when the left eye is darker than the right (Fig. 7B), and when the right eye is darker than the left (Fig. 7C). These results also replicate the primary finding from Bonnen et al. (2017) and are generally consistent with the many results showing that changes in depth are processed more slowly than changes in horizontal position. However, neither the X vs. X nor the Z vs. Z cross-correlation provides information about differences in temporal processing between the eyes.

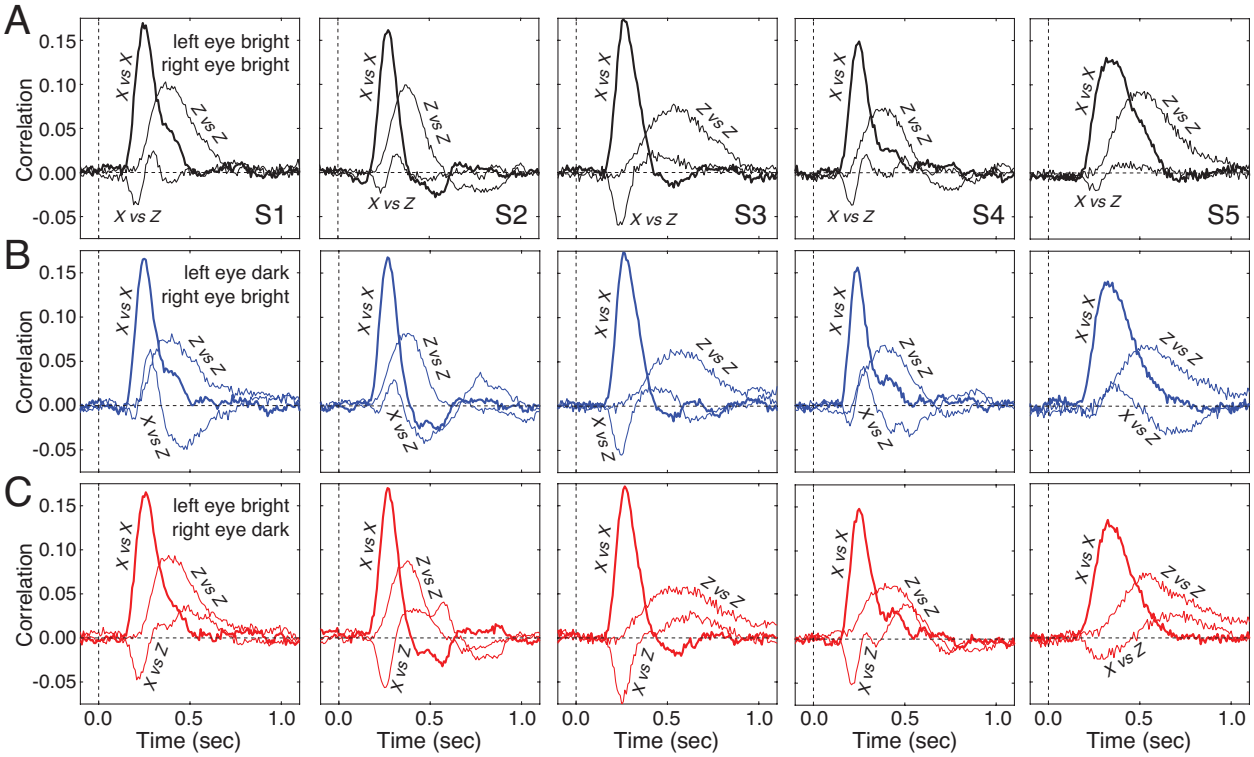


Figure 7. Binocular tracking. Temporal cross-correlograms between X target motion and X response motion (X vs. X), Z target motion and Z response motion (Z vs. Z), and X target motion and Z response motion (X vs. Z) for all five human observers for three visual conditions: **A** Both on-screen images are equally bright (ΔO= 0.0OD). **B** The left-eye image is dark (i.e. 75% of maximum luminance) and the right-eye image is bright (ΔO=-0.6OD). **C** The left-eye image is bright and the right-eye image is dark (i.e. 75% of maximum luminance; ΔO=+0.6OD). Tracking in X is comparatively swift, tracking in Z is more sluggish, and the impact of horizontal target motion on the depth response depends systematically on the luminance differences between the eyes.

Recall that, in the context of the Pulfrich effect (see Fig. 4), interocular differences in processing speed cause target objects that are moving horizontally to be misperceived as moving in depth. To examine whether this signature of the Pulfrich effect is present in binocular tracking in depth, we computed the cross-correlation between X target motion and Z response motion. The X vs. Z cross-correlogram clearly depends on the visual condition (Fig. 7ABC; X vs. Z). When the left eye is dark (Fig. 7B), there tends to be an initial positive lobe, followed by a second negative lobe. When the right eye is dark (Fig. 7C), the shape of the cross-correlogram is approximately mirror reversed. The dependence of the X vs. Z cross-correlograms on the visual condition suggests that it may be useful for recovering differences in the time course of visual processing between the eyes. Note also that when both eyes are bright (Fig. 7A), there are small but systematic deviations from zero, which implies a small baseline asymmetry in left- and right-eye processing. These small deviations will become important for a subsequent analysis.

Interestingly, as we show in the Supplement, the cross-correlation of target X-velocity and response Z-velocity is proportional to the difference between the temporal impulse response functions in the left and right eyes

$$\dot{x}_T \otimes \dot{\tilde{z}}_T \propto h_R - h_L \qquad (1)$$

where $\dot{x}_T$ is the X velocity of the target, $\dot{\hat{z}}_T$ is the Z velocity of the human response, and h_L and h_R are the left- and right-eye impulse response functions. Thus, it may be possible to use 3D target tracking to directly estimate the time course of interocular differences in visual processing in a single condition.

Simulations confirm the mathematical result. The simulations were performed as follows. First, for a given random walk in X and Z, we calculated the left- and right-eye image positions of the target on each time step using projective geometry. Next, we convolved the image positions with model left- and right-eye impulse response functions to obtain the effective left- and right-eye image positions. Finally, we back projected the effective left- and right-eye target image positions into 3D space (see Methods). The 3D tracking response was assumed to equal the position of the back-projected location in space.

Simulated tracking performance for an observer with different impulse response functions for the two eyes is shown in Fig. 8A. The simulated observer smoothly tracks the horizontal target motion component with a consistent delay, just as in the monocular tracking experiment. Tracking performance in depth, however, is distinguished by obvious inaccuracies (Fig. 8B). The over- and under-shooting of the response in depth relative to the target depth depends systematically on horizontal target motion (Fig. 8B, rectangular boxes). This dependence is the hallmark of the Pulfrich effect, and can be quantified by the cross-correlogram of target X-motion with response Z-motion. The thick curves in Fig. 8C show this cross-correlogram. Importantly, the difference between the left- and right-eye temporal impulse response functions, shown by the thin curves in Fig. 8C, beautifully predicts the X vs. Z cross-correlogram, as is indicated by Equation 1.

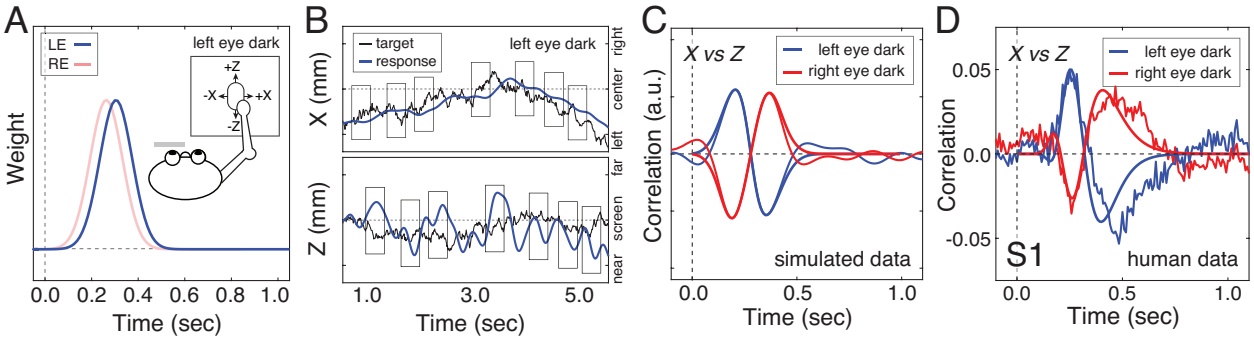


Figure 8. Simulated binocular tracking through depth with interocular luminance differences. **A** Impulse response functions for the model observer's left and right eyes when the left eye is dark and the right eye is bright. **B** Resulting tracking performance in X and Z when the left eye is dark and processed more slowly than the right eye. Target motion in X is tracked smoothly and accurately, but large tracking inaccuracies are apparent in Z. Large over- and under-estimates of depth (i.e. Z position) are caused by rightward and leftward target movements, respectively, the hallmark of the classic Pulfrich effect. The vertical rectangles highlight regions of the time-series where these effects are most noticeable. **C** The cross-correlation of X target motion and Z response motion (thick curves) from the simulated data and the difference between the left- and right-eye impulse response functions (thin curves) are plotted for two different conditions with one eye dark and the other eye bright (colors). The simulated data agrees with the mathematical prediction in Eq. 1. **D**. Predicted differences vs. the actual data for the first observer.

Data from the first human observer has clear similarities to the simulated data (Fig. 8D, noisy vs. smooth curves). But the strongest test of whether the binocular target tracking data is consistent with the other data thus far collected is to examine whether human binocular tracking performance is predicted by discrepancies between the left- and right-eye impulse response functions. To find out, we made use of the cross-correlograms from the monocular tracking

experiment (see Fig. 3A). Recall that these monocular left- and right-eye cross-correlograms reflect the processing dynamics of both the visual and the motor systems. However, if the dynamics of the motor system are stable across two monocular conditions (e.g. left eye dark, right eye bright), then the difference between the corresponding cross-correlograms will reflect only differences in visual processing (i.e. the differences between the left- and right-eye impulse response functions; see Eq. 1). Thus, if there are no significant linear filtering operations and there are no significant non-linearities associated with binocular combination, then the difference between the left- and right-eye cross-correlograms should predict the cross-correlogram between target X-motion and response Z-motion during binocular tracking, just as it does in the simulated dataset.

The differences between fits to the monocular cross-correlograms (see Methods) nicely predict the binocular X vs. Z cross-correlograms in the first human observer (Fig. 8D, smooth curves). (The baseline X vs. Z asymmetry in the left-eye bright, right-eye bright condition was subtracted off before making the comparison; see Fig. 7A.) In this observer, binocular tracking performance is predicted by discrepancies between the left and right-eye impulse response functions. Thus, when the left- and right-eye images have different luminances, the impact of horizontal target motion on the human response in depth is well-accounted for by differences in the monocular cross-correlograms.

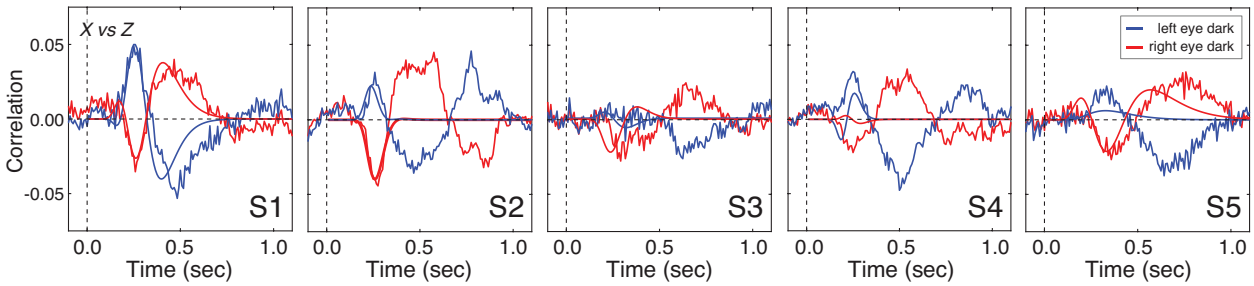


Figure 9. Binocular tracking in depth for all human observers (Note: data in the left-most panel is identical to the data in Fig. 8D). Horizontal target motion systematically affects human tracking responses in depth curves (noisy curves). The blue and red curves indicated the change from baseline (i.e. both eyes bright) when one eye is dark (OD=0.6) and the other eye is bright (OD=0.0; see Methods). Blue curves indicate the X vs. Z cross-correlograms when the left eye is dark and the right eye is bright. Red curves indicate the X vs. Z cross-correlograms when the left eye is bright and the right eye is dark. Predictions from the difference between the monocular cross-correlograms (smooth curves) in the corresponding visual conditions (dark eye: OD=0.6, bright eye: OD=0.0) account for the first lobe of the binocular X vs Z cross-correlograms in half of the conditions (smooth curves). The second lobes and third lobes of the cross-correlograms are poorly accounted for and must be due to other factors.

However, prediction quality is mixed across the set of human observers (Fig. 9). Across the five observers, the first lobe of the X vs. Z cross-correlograms is nicely predicted by differences between the monocular cross-correlograms in half of the conditions. There are obvious prediction failures in the other half of the conditions. The prediction quality (and the raw data) is particularly bad for observer S3, who reported struggling to track the target in depth. Also, note that the other lobes of the cross-correlograms are poorly accounted for in all observers (except perhaps for observer S1). We speculate that these additional lobes, which tend to occur half a second or more after the driving stimulus motion are driven by fundamentally different processes (e.g. higher-level cognitive processes) than those that are measured in the monocular tracking and traditional button-press psychophysics experiments. The analytical result, simulations, and partial experimental validation are intriguing, but the discrepancies between the predictions and results suggest that more work must be done. More responsive effectors (e.g. the eyes instead of the hand; see Discussion) and higher fidelity input devices may improve the match between prediction and data in the first lobe of each cross-correlogram.

The fact that the additional lobes are unaccounted for by the predictions suggests that other processes not addressed here are involved in binocular combination or in binocular tracking in depth. These phenomena are potentially quite interesting, but a full analysis of them must await future targeted experiments and analyses.

Methods

Participants

Five human observers participated in the experiment; four males and one female. Two were authors and the rest were naïve to the purposes of the experiment. All had normal or corrected to normal visual acuity (20/20), and normal stereoacuity as determined by the Titmus Stereo Test. The observers were aged 26, 31, 36, 41, and 55 years old at the time of the measurements. Four observers were males; the other was female. All human observers provided informed consent in accordance with the Declaration of Helsinki using a protocol approved by the Institutional Review Board at the University of Pennsylvania.

Apparatus

Stimuli were displayed on a custom-built four-mirror stereoscope. Left- and right-eye images were presented on two identical Vpixx VIEWPixx LED monitors. The gamma function of each monitor was linearized using custom software routines. The monitors were 52.2x29.1cm, with a spatial resolution of 1920x1080 pixels, a refresh rate of 120Hz, and a maximum luminance of 105.9cd/m^2. After light loss due to mirror reflections, the maximum luminance was 93.9cd/m^2. A single AMD FirePro D500 graphics card with 3GB GDDR5 VRAM controlled both monitors to ensure that the left and right eye images were presented simultaneously. To overcome bandwidth limitations of the monitor cables, custom firmware was written so that a single color channel drove each monitor; the red channel drove the left monitor and the green channel drove the right monitor. The single-channel drive to each monitor was then split to all three channels for gray scale presentation.

Observers viewed the monitors through a pair of mirror cubes positioned one inter-ocular distance apart (Supplementary Fig. S1). Heads were stabilized with a chin and forehead rest. The mirrors were adjusted such that the vergence distance matched the distance of the monitors. The monitors were positioned at a distance of 100cm. This distance was confirmed both by a laser ruler measurement and by a visual comparison with a real target at 100cm. The mirror cubes had 2.5cm openings which yielded an approximate field of view of approximately 15x15º. At this distance, each pixel subtended 1.09arcmin. Stimulus presentation was controlled via the Psychophysics Toolbox-3 [24]. Anti-aliasing enabled sub-pixel resolution permitting accurate presentations of disparities as small as 15-20arcsec.

Neutral density filters

To control the stimulus luminance for each eye and to induce luminance differences between the eyes we placed 'virtual' neutral density filters in front of the eyes. First, we converted optical density to transmittance, the proportion of incident light that is passed through the filter, using the standard expression $T = 10^{-OD}$ where T is transmittance and OD is optical density. Then, we reduced the monitor luminance by a scale factor equal to the transmittance. We have previously verified that real and virtual neutral density filters with equivalent strengths yield identical performance [20]. The interocular difference in optical density

$$\Delta O = OD_R - OD_L \qquad (2)$$

quantifies the luminance difference between the eyes. Human observers ran in five conditions with virtual neutral density filters, with equally spaced interocular differences in optical density ranging from -0.6 and 0.6; in all conditions, at least one eye had an optical density of 0.0. The extremes of this range correspond to one eye having only 25% the luminance of the stimulus to the other eye.

Procedure

Tracking data was collected from each human observer in individual runs. The human observer initiated each run by clicking the mouse, which caused the target bar to appear in the center of the screen. After a stationary period of 500ms, the target bar followed a random-walk trajectory for the next eleven seconds. The task was to track the target bar as accurately as possible with a small dark mouse cursor.

In the monocular tracking experiment, data was collected in each of six visual conditions: left eye alone and right eye alone at 75% less light than maximum luminance, at 50% less light than maximum luminance, and at maximum luminance. These values are equivalent to placing neutral density filters with optical densities of 0.6, 0.3, and 0.0, respectively, in front of the viewing eye. The non-viewing eye was occluded with an eye patch. Data was collected in twenty intermixed blocks of twelve runs each for a total 40 runs per condition.

In the binocular target tracking experiment, data was collected in each of four visual conditions. Three of the conditions were the analogs of a subset of the conditions in the monocular tracking experiment: left eye at 75% maximum luminance and right eye at maximum luminance (i.e. left eye dark & right eye bright; ΔO=-0.6OD), left and right eye at maximum luminance (ΔO=-0.0OD), and left eye at maximum luminance and right eye at 75% maximum luminance (i.e. left eye bright and right eye dark; ΔO=+0.6OD). In each block, visual conditions were intermixed; a total of 40 runs per condition were collected. Left-eye-alone monocular blocks, right-eye-alone blocks, and binocular blocks were interleaved and counterbalanced.

Tracking stimuli and tracking response

For the tracking experiments, the target bar was a moving white vertical bar on a gray background. The target bar, which subtended 0.25x1.00º of visual angle, was vertically flanked by two stationary sets of thirteen picket fence bars (Fig. 2A). The gray background subtended 10.0x7.5º of visual angle.

The target bar performed a random walk in X only or in X and Z. The X and Z positions of the target on each time step $t+1$ were generated as follows

$$x(t+1) = x(t) + \varepsilon_x \; ; \quad \varepsilon_x \sim N(0,Q) \tag{3a}$$

$$z(t+1) = z(t) + \varepsilon_z ; \quad \varepsilon_z \sim N(0,Q) \tag{3b}$$

where ε_x and ε_z are samples of Gaussian noise that determine the change in target position from the current to the next time step, and Q is the so-called drift variance. The drift variance controls the mean magnitude of the position change on each time step, and hence the overall variability of the random walk. The variance of the walk positions across multiple walk instantiations $\sigma^2(t) = Qt$ is equal to the product of the drift variance and the number of elapsed

time steps. The value of the drift variance in our task (0.8mm per time step) was chosen to be as large as possible such that each walk would traverse as much ground as possible while maintaining the expectation that less than one walk out of 500 (i.e. less than one per human observer throughout the experiment) would escape the horizontal extent of the gray background area (176x131mm) before the 11 second trial completed.

In many experiments involving stereopsis, the distance of the virtual stimulus is specified by retinal disparity, which is directly determined by the stimulus position on each of two eye-specific monitors. The on-screen stimulus positions corresponding to a particular target position in stereoscopic space can be obtained using similar triangles (Supplementary Fig. S2). The left- and right-eye image positions corresponding to particular target position, are given by

$$x_L = x_T - \left(\frac{z_T - z_S}{z_T}\right)\left(x_T + I/2\right) = x_T\left(\frac{z_S}{z_T}\right) + \frac{I}{2}\left(\frac{z_S}{z_T}\right) - \frac{I}{2} \tag{4a}$$

$$x_R = x_T - \left(\frac{z_T - z_S}{z_T}\right)\left(x_T - I/2\right) = x_T\left(\frac{z_S}{z_T}\right) - \frac{I}{2}\left(\frac{z_S}{z_T}\right) + \frac{I}{2} \tag{4b}$$

where z_S is the screen distance and I is the interocular distance.

The effective left- and right-eye images are obtained by convolving the on-screen left and right eye target images with eye-specific temporal impulse response functions

$$\tilde{x}_L(t) = x_L(t) * h_L(t) \tag{5a}$$

$$\tilde{x}_R(t) = x_R(t) * h_R(t) \tag{5b}$$

where $h_L(t)$ and $h_R(t)$ are the left- and right-eye temporal impulse response functions, respectively. Convolving the left- and right-eye target velocities with the impulse response function of each respective eye gives the velocities of the effective left- and right-eye images.

The predicted 3D response position is obtained by back-projecting the effective left- and right-eye images into 3D space. The X and Z positions of the 3D response are given by

$$\tilde{x}_T = \frac{I}{2}\left(\frac{\tilde{x}_L + \tilde{x}_R}{\tilde{x}_L - \tilde{x}_R + I}\right) \tag{6a}$$

$$\tilde{z}_T = I\left(\frac{z_S}{\tilde{x}_L - \tilde{x}_R + I}\right) \tag{6b}$$

For the monocular tracking conditions, the stimuli only changed position in X and the observers wore an eye patch over one eye (see Fig. 1B).

The response velocities in X and Z are obtained by differentiating the X and Z response positions with respect to time

$$\dot{\tilde{x}}_T = \frac{I}{2}\left[\frac{2\left(\tilde{x}_L\dot{\tilde{x}}_R - \dot{\tilde{x}}_L\tilde{x}_R\right) + I\left(\dot{\tilde{x}}_L + \dot{\tilde{x}}_R\right)}{\left(\tilde{x}_L - \tilde{x}_R + I\right)^2}\right] \tag{7a}$$

$$\dot{\tilde{z}}_T = Iz_S\left(\frac{\dot{\tilde{x}}_R - \dot{\tilde{x}}_L}{\left(\tilde{x}_L - \tilde{x}_R + I\right)^2}\right) \tag{7b}$$

where $\dot{\tilde{x}}_L$ and $\dot{\tilde{x}}_R$ are the velocities of the effective target images in the left and right eyes, respectively. The target velocities in X and Z are given by functions having an identical form, provided that the velocities of the left- and right-eye target images are substituted for the velocities of the effective target images in the two eyes.

To determine the impulse response function relating the target and response, we computed the zero-mean normalized cross-correlations between the target and response velocities

$$\rho\left(\tau;\dot{x},\dot{\tilde{x}}\right) = \frac{1}{\left\|\dot{x}(t)\right\|\left\|\dot{\tilde{x}}(t)\right\|}\left[\sum_{t=1}^{N}\left(\dot{x}(t)-\bar{\dot{x}}\right)\left(\dot{\tilde{x}}(t+\tau)-\bar{\dot{\tilde{x}}}\right)\right] \tag{8a}$$

$$\rho\left(\tau;\dot{z},\dot{\tilde{z}}\right) = \frac{1}{\left\|\dot{z}(t)\right\|\left\|\dot{\tilde{z}}(t)\right\|}\left[\sum_{t=1}^{N}\left(\dot{z}(t)-\bar{\dot{z}}\right)\left(\dot{\tilde{z}}(t+\tau)-\bar{\dot{\tilde{z}}}\right)\right] \tag{8b}$$

where τ is the lag, $\dot{x}$ and $\dot{\tilde{x}}$ are the target and response velocities in X, $\dot{z}$ and $\dot{\tilde{z}}$ are the target and response velocities in Z. We also computed the zero-mean normalized cross-correlation between the X target velocity and the Z response velocity

$$\rho\left(\tau;\dot{x},\dot{\tilde{z}}\right) = \frac{1}{\left\|\dot{x}(t)\right\|\left\|\dot{\tilde{z}}(t)\right\|}\left[\sum_{t=1}^{N}\left(\dot{x}(t)-\bar{\dot{x}}\right)\left(\dot{\tilde{z}}(t+\tau)-\bar{\dot{\tilde{z}}}\right)\right] \tag{9}$$

to determine how target X motion impacts response Z motion. The influence of X target motion on Z response motion is the hallmark of the Pulfrich effect.

The normalized cross-correlation can be interpreted as a series of normalized dot products between the two time-series at each of many different lags. Thus, it tells one how the similarity of two time-series changes as a function of time lag. Assuming a linear system, when the input time series (i.e. the target velocities) is white, as it is here, the cross-correlation with the response gives the impulse response function of the system.

When computing the normalized cross-correlations, we excluded the first second of each eleven second tracking run so that observers reached steady state tracking performance. First, we computed the normalized cross-correlation in each run (Eqs. 8 & 9). Then, we averaged these cross-correlograms across runs in each condition. It is these mean cross-correlation functions (i.e. the cross-correlograms) that are presented in the Results section figures.

Predicting Binocular from Monocular Tracking Performance

Predicting binocular tracking performance from monocular tracking performance was a three-step process. To determine how lateral target motion influences the depth response we first fit

the monocular cross-correlograms with log-Gaussian shaped functions using least squared regression. In nearly all cases, the log-Gaussian provided an excellent fit to the monocular cross-correlograms once the curves exceeded two standard deviations of the baseline noise. Second, we computed the difference between the fits in the bright condition (OD=0.0) in one eye and the dark condition (OD=0.6) in the other eye (e.g. left eye dark, right eye bright). Third, we subtracted the X vs. Z cross-correlogram in the condition where both eyes are bright from the cross-correlograms in the conditions of interest (i.e. one eye dark, one eye bright). This subtraction removes each observer's baseline asymmetry, and isolates the changes caused by the luminance differences between the eyes.

Button-press psychophysics

For the button press psychophysics, we simulated the classic pendulum Pulfrich stimulus on the display. For each trial, the left- and right-eye on-screen bar positions in degrees of visual angle were given by

$$x_L(t) = E\cos\left(2\pi\omega\cdot(t+\Delta t)+\phi_0\right) \tag{10a}$$

$$x_R(t) = E\cos\left(2\pi\omega\cdot(t)+\phi_0\right) \tag{10b}$$

where E is the target movement amplitude in degrees of visual angle, ω is the temporal frequency of the target movement, ϕ_0 is the starting phase, t is time in seconds, and Δt is the on-screen delay between the left- and right-eye target images. When the interocular delay equals zero, the virtual bar moves in the frontoparallel plane at the distance of the monitors. When the interocular delay is non-zero, a spatial binocular disparity results, and the virtual bar follows a near-elliptical trajectory of motion in depth. Negative values indicate the left-eye on-screen image is delayed relative to the right; positive values indicate the left eye on-screen image is advanced relative to the right. Note that we did not temporally manipulate when left- and right-eye images were presented on-screen; both eyes' images were presented coincidently on each monitor refresh. Rather, we calculated the disparity $\Delta x = \dot{x}\Delta t$ given the target velocity and the desired on-screen delay on each time step, and appropriately shifted the spatial positions of the left- and right-eye images.

The on-screen binocular disparity associated with a given interocular delay as a function of time was given by

$$\Delta x(t) = x_R(t) - x_L(t) = 2E\sin\left(\pi\omega\cdot(\Delta t)\right)\sin\left(\pi\omega\cdot(2t+\Delta t)+\phi_0\right) \tag{11}$$

where negative disparities are crossed (i.e. nearer than the screen) and positive disparities are uncrossed (i.e. farther than the screen). The disparity takes on its maximum magnitude when the perceived stimulus is directly in front of the observer and the lateral movement is at its maximum speed. When the stimulus is moving to the right, the maximum disparity in visual angle is given by $\Delta x_{\max} = 2E\sin(\pi\omega)$.

The movement amplitude was 2.5º of visual angle (i.e. 5.0º total change in visual angle in each direction; see Fig. 4D), the temporal frequency was 1 cycle per second, and the starting phase ϕ_0 was randomly chosen to be either 0 or π. Restricting the starting phase to these two values forced the stimuli to start either 2.5º to the right or 2.5º to the left of center on each trial. The on-screen interocular delays ranged between $\pm$10.0ms at maximum. The range and particular values were adjusted to the sensitivity of each human observer.

Two sets of five vertical 0.25x1.00º bars in a 'picket fence' arrangement flanked the region of the screen traversed by the target bar (Fig. 4D). The picket fences were defined by disparity to be at the screen distance, and served as a stereoscopic reference of the screen distance for the observer. A 1/f noise texture, also defined by disparity to be at the screen distance, covered the periphery of the display to help anchor vergence and serve as a further stereoscopic reference to the screen distance. A small fixation dot was located at the center of the screen, at which observers were instructed to fixate.

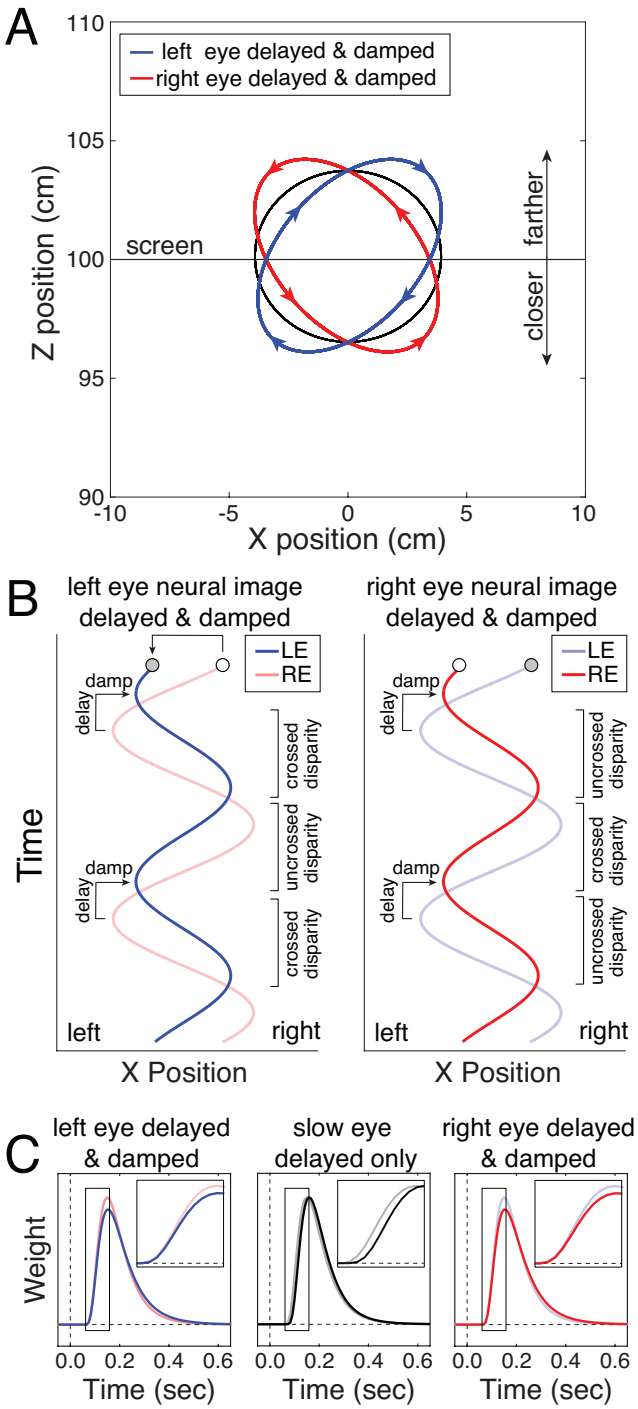
The observer's task was to report whether the target bar was moving leftward or rightward when it appeared to be nearer than the screen on its virtual trajectory in depth. Using a one-interval two-alternative forced choice procedure, psychometric functions were collected in each condition using the method of constant stimuli. Nine different levels of interocular delay were collected for each function. The psychometric functions were fit with a cumulative Gaussian via maximum likelihood methods. The 50% point on the psychometric function—the point of subjective equality (PSE)—indicates the on-screen interocular delay needed to null the interocular difference in processing speed. Observers ran 180 trials per condition in counter-balanced blocks of 90 trials each.

Discussion

We have shown that continuous target tracking in two and three dimensions can provide estimates of the time course of visual processing. We have also shown that millisecond-scale differences in visual processing between the eyes can be estimated with precision that is comparable to traditional psychophysics. Here, we discuss some of the advantages of the target tracking paradigm over traditional psychophysics, consider future methodological challenges, speculate about the potential for clinical application, and reflect on how the current results fit in the historical development of the science.

Time-course information provides potential for enhanced predictive power

The ability to recover the time course of information processing affords us new predictive and explanatory power. For example, in the button-press psychophysics task, when sinusoidal target motion was presented in the plane of the screen and one eye was darker than the other, rather than the expected trajectory (Fig. 10A, black), one observer spontaneously reported perceiving elliptical trajectories in depth that were not aligned with the screen (Fig. 10A, colors). Upon follow-up debriefing, two additional observers reported perceiving misaligned elliptical trajectories. The most likely proximate cause of these anomalous Pulfrich percepts is that the amplitude of the effective image motion in the more slowly processed eye is not merely delayed but also damped relative to the effective image motion in the other eye (Fig. 10B). Shape differences in the impulse response functions in the two eyes can lead to these differences in the effective neural image positions (Fig. 10C). Consider the specific case when the left eye is darker than the right eye, and the impulse response function in the left eye has both a longer latency and a longer temporal integration period than the right eye (Fig. 10C, left). In this case, the effective position of the left-eye image will be delayed and damped relative to the right-eye image (Fig. 10B, left). As a consequence, observers will perceive the target stimulus undergoing 'front left' motion along an elliptical trajectory that is rotated such that the primary axis is askew to the screen (Fig. 10A, blue). The fact that subtle changes in the shape of the impulse response functions (Fig. 10C) can cause dramatic changes in the predicted percept (Fig. 10A) may prove useful for measuring small stimulus-specific changes in visual processing. A systematic investigation of these effects will be an interesting direction for future work.



Figures 10. Anomalous Pulfrich percepts. **A** Some observers spontaneously reported perceiving anomalous near-elliptical motion trajectories that were not aligned with the screen (blue or red). Other observers perceived trajectories that were aligned with the screen (black). **B** Effective neural image positions as a function of time that can elicit the anomalous perceived motion trajectories in A. Left-eye (blue) and right-eye (red) neural image positions when the processing is delayed and damped in the left eye (left subplot) or right eye (right subplot). **C** Temporal impulse response functions that have different temporal integration periods can account for neural image positions in B. Impulse response functions with longer temporal integration periods tend to dampen the amplitude of the effective image position in that eye. Left subplot: Right-eye processing is fast and left-eye processing is delayed (i.e. has a longer latency) with a longer temporal integration period. Middle subplot: Slow-eye processing is delayed but not damped relative to the fast-eye processing. Right subplot: Left-eye processing is fast and right-eye processing is delayed with a longer temporal integration period.

Preservation of visual processing differences in the visuo-motor response

The millisecond-scale differences in visual processing that underlie Pulfrich-like phenomena have most likely have arisen by early visual cortex [25,26]. The fact that these millisecond-scale differences are preserved at the level of the visuo-motor response is remarkable. To effect a motor response, electrical impulses must travel down multiple myelinated axon sheaths, and chemical communication must occur at multiple synaptic junctions as signals move from cortex, to the brain stem, and down the arm to the hand. The motion of the hand represents the culmination of all of these processes. Despite the myriad signal transformations and delays that occur after early visual cortex, the tiny interocular delays that cause the Pulfrich effect are preserved in the motor response. There is a clear evolutionary advantage to being able to effect motor responses as rapidly and consistently as visual processing allows, but there is no logical guarantee that this will happen [27]. The delay associated with a 'low strength' signal at one stage of processing, for example, may be magnified as the signal proceeds through subsequent stages of processing. The fact that measurements of manual tracking can reflect subtle changes in signal strength bodes well for its applicability to a wide array of experimental questions.

Implications and applications

Traditional psychophysical techniques are the current gold standard for minimizing measurement error in sensory-perceptual experiments. The price one pays for using traditional psychophysics, however, is that experimental data take a long time to collect, and the experiments themselves are somewhat tedious for observers. While merely inconvenient in many laboratory settings, time and tedium become very real problems if one wants to make measurements in children or in clinical populations that might be unwilling or unable to produce a satisfactory number of trials. Moreover, instructions that seem simple for experienced psychophysical observers, such as "press the right arrow key when the stimulus appears to be moving anti-clockwise when viewed from above, and the left arrow if is going the other way" might be confusing to young children, and unintelligible to pre-verbal infants. Indeed, because infants naturally follow moving visual targets, visual function could be assessed behaviorally early in the developmental arc. Finally, if one wishes to study effects that have large individual differences, then a technique is needed in which it is possible to obtain a satisfactory amount of data from a single observer in a short amount of time so that it is, in turn, possible to collect good data from a very large number of non-traditional observers.

In clinical settings, functional vision is measured using simple psychophysical techniques. Multiple visual functions–acuity, binocular function, and color vision, for example–are assessed with different tests requiring different instructions. It may be simpler to have patients perform only a single task (i.e. track a target) and to have the target be defined by fine spatial detail, or binocular disparity, or color on different tracking runs. Moreover, it is likely that many pathologies—psychophysical, physical, developmental—have sensory-motor processing delays as an indicating factor. A sensitive clinical test for these processing delays that would easy, and perhaps fun, for a wide range of patient populations could be quite useful.

The arms and hands are large, heavy, and sluggish. The eyes are smaller, lighter, more responsive. The eyes therefore have the potential to increase the fidelity of tracking performance. When the eyes are tracking random walk motion, for example, the latency of smooth pursuit eye movements tends to be in the range of 80-120ms [28,29]. Although smooth pursuit latencies depend on the properties of the target being tracked, these latencies are considerably less than the 150-200ms latencies of the motor response of the hand in the current tracking experiments (see Fig. 3). There also tends to be considerably less inter-subject variability with smooth pursuit [28]. Finally, by using eye movements, it is possible to adapt

continuous target tracking methods to young children and animal models. Young children and many animals reflexively follow moving targets with their eyes, potentially obviating the need for verbal instruction (in the case of children) or extensive training (in the case of animals). Indeed, continuous target tracking with eye movement monitoring has recently been used to measure the ocular following responses of macaques and marmosets [30], and to demonstrate both vergence and accommodative responses in infants as young as five weeks old [31].

Binocular eye tracking, in particular, might have interesting applications for assessing processing differences between the eyes. First, binocular target tracking is potentially more efficient than monocular tracking; a single binocular run can yield information about interocular processing differences, whereas at least two distinct runs for each eye must be performed with monocular tracking. Second, with binocular tracking, the interocular comparison is performed automatically by the binocular visual system before the motor system is engaged; with monocular target tracking, the comparison is performed post-hoc by the experimenter after the motor response has been collected in two different visual tracking conditions. Third, binocular tracking eliminates the possibility that state changes (e.g. alertness, motivation) between runs could corrupt estimates of interocular delay. Finally, as with the Pulfrich effect, small changes in one eye's processing (e.g. due to a pathology) might produce detectable interocular differences even though the change in the monocular response per se would be impossible to detect without some previously-recorded baseline. Using eye instead of hand movements as the response measure in continuous target tracking experiments may therefore have a number of advantages, and these probable advantages will only increase as eye tracking technology continues to improve.

The continuous nature of the stimulus and the response also afford new possibilities for the application of analytical methods from systems neuroscience to psychophysical data. In research on human perception and behavior, it is common to compare human performance to that of a normative model of the task[32-35]. But normative models in these domains are rarely constructed to predict continuous responses over time. In systems neuroscience research, popular models for neural systems identification are designed to recover computational-level descriptions of how stimuli drive neural response. These models (e.g. the generalized linear model; GLM) are commonly applied to continuous time stimuli and responses. But they have rarely been applied to human psychophysical data, perhaps because human datasets collected with traditional button-press methods have insufficient data for the method's full statistical power to be realized[36-38]. The continuous target-tracking paradigm, paired with recent developments linking normative models to methods for systems identification[39] and human behavior[40,41], provides an exciting direction for future work in the analysis of human perception and behavior.

The reverse Pulfrich effect

It has long been known that interocular differences in luminance cause interocular differences in processing speed; the darker image is processed more slowly [12,13]. More recently, it was discovered that interocular differences in blur also cause interocular differences in processing speed: the blurrier image is processed more quickly [20,42]. These phenomena are not simply laboratory curiosities. Monovision, a common prescription lens correction for presbyopia, intentionally induces interocular blur differences, and monovision corrections are currently worn by millions of people. The induced blur differences have the advantage of increasing the viewer's depth of field, enabling them to see clearly at both near and at far distances. But if the correction induces an interocular mismatch in processing speed of only a few milliseconds—as it can—dramatic misperceptions of motion and depth, similar to those described here, can result. In some real-world circumstances (e.g. driving), these misperceptions have the potential

to impact public safety. For example, a motorist may misperceive the distance to a cyclist in cross-traffic by the width of an entire lane of traffic [20]. Fortunately, simple but novel combinations of existing ophthalmic interventions hold promise for eliminating these misperceptions [20,42]. The continuous target tracking paradigm developed here may prove useful for developing methods for measuring interocular delays in a clinical setting, where screening a large volume of unpracticed observers would be necessary.

Given that blur decreases contrast and given that decreases in contrast generally decrease processing speed [19,26,43-47], it may at first seem surprising that blur increases rather than decreases processing speed. However, it is known that higher spatial frequencies are processed more slowly than lower spatial frequencies [19,26,43-48]. Hence, because blur reduces the contrast of high spatial frequencies (i.e. fine detail) more than low spatial frequencies (i.e. coarse detail) [49-51], a blurrier image, having fewer fine details, is processed more quickly. Experiments have confirmed that this explanation explains the reverse Pulfrich effect [20]. However, the computational rules relating image properties to processing speed remain to be worked out. The tracking paradigm, because of its fine temporal sensitivity, may be well suited for investigating this problem as well.

Conclusion

In this paper, we have used continuous target tracking to explore how image differences between the eyes impact the time course of visual processing in the context of the Pulfrich effect, a classic stereo-motion phenomenon. We have shown that the small (i.e. less than 10ms) interocular delays underlying the Pulfrich effect are preserved through the entire sensory motor loop and can thus be recovered from tracking data. Moreover, because the time course of information processing is recovered (unlike traditional psychophysics), predictions can be made about the entire perceived trajectories of moving targets. The fact that target tracking can recover subtle differences in early visual processing suggests that the technique could have applications as a screening tool in clinical optometry and ophthalmology.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01- EY028571 from the National Eye Institute and the Office of Behavioral and Social Science Research to JB, and startup funds from the University of Pennsylvania to JB.

AUTHOR CONTRIBUTIONS

JB conceived the project, collected and analyzed data, performed simulations, and wrote the paper. JB and LKC developed concepts and edited the paper.

REFERENCES

1. Wheatstone, C. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* **128,** 371–394 (1838).
2. Ogle, K. N. On the limits of stereoscopic vision. *Journal of Experimental Psychology* **44,** 253–259 (1952).
3. Julesz, B. Binocular depth perception without familiarity cues. *Science* **145,** 356–362 (1964).
4. Tyler, C. W. & Julesz, B. Binocular cross-correlation in time and space. *Vision Research* **18,** 101–105 (1978).
5. Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* **249,** 1037–1041 (1990).

6. Cormack, L. K., Stevenson, S. B. & Schor, C. M. Interocular correlation, luminance contrast and cyclopean processing. *Vision Research* **31,** 2195–2207 (1991).
7. DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* **352,** 156–159 (1991).
8. Cumming, B. G. & DeAngelis, G. C. The physiology of stereopsis. *Annu. Rev. Neurosci.* **24,** 203–238 (2001).
9. Banks, M. S., Gepshtein, S. & Landy, M. S. Why is spatial stereoresolution so low? *J. Neurosci.* **24,** 2077–2089 (2004).
10. Burge, J. & Geisler, W. S. Optimal disparity estimation in natural stereo images. *Journal of Vision* **14,** 1–18 (2014).
11. Cormack, L. K., Czuba, T. B., Knöll, J. & Huk, A. C. Binocular Mechanisms of 3D Motion Processing. *Annu Rev Vis Sci* **3,** 297–318 (2017).
12. Pulfrich, C. Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Naturwissenschaften* **10,** 553–564 (1922).
13. Lit, A. The magnitude of the Pulfrich stereophenomenon as a function of binocular differences of intensity at various levels of illumination. *Am J Psychol* **62,** 159–181 (1949).
14. Wilson, J. A. & Anstis, S. M. Visual delay as a function of luminance. *Am J Psychol* **82,** 350–358 (1969).
15. Rogers, B. J. & Anstis, S. M. Intensity versus adaptation and the Pulfrich stereophenomenon. *Vision Research* **12,** 909–928 (1972).
16. Morgan, M. J. & Thompson, P. Apparent motion and the Pulfrich effect. *Perception* **4,** 3–18 (1975).
17. Carney, T., Paradiso, M. A. & Freeman, R. D. A physiological correlate of the Pulfrich effect in cortical neurons of the cat. *Vision Research* **29,** 155–165 (1989).
18. Lages, M., Mamassian, P. & Graf, E. W. Spatial and temporal tuning of motion in depth. *Vision Research* **43,** 2861–2873 (2003).
19. Reynaud, A. & Hess, R. F. Interocular contrast difference drives illusory 3D percept. *Sci Rep* **7,** 5587 (2017).
20. Burge, J., Rodriguez-Lopez, V. & Dorronsoro, C. Monovision and the Misperception of Motion. *Curr. Biol.* **29,** 2586–2592.e4 (2019).
21. Watson, A. in (eds. Boff, K., Thomas, J. & Kaufman, L.) **1,** (Handbook of perception and human performance, 1986).
22. Bonnen, K., Huk, A. C. & Cormack, L. K. Dynamic mechanisms of visually guided 3D motion tracking. *J. Neurophysiol.* **118,** 1515–1531 (2017).
23. Bonnen, K., Burge, J., Yates, J., Pillow, J. & Cormack, L. K. Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision* **15,** 1–16 (2015).
24. Brainard, D. H. The Psychophysics Toolbox. *Spat Vis* **10,** 433–436 (1997).
25. Wolpert, D. M., Miall, R. C., Cumming, B. & Boniface, S. J. Retinal adaptation of visual processing time delays. *Vision Research* **33,** 1421–1430 (1993).
26. Vassilev, A., Mihaylova, M. & Bonnet, C. On the delay in processing high spatial frequency visual information: reaction time and VEP latency study of the effect of local intensity of stimulation. *Vision Research* **42,** 851–864 (2002).
27. Sternberg, S. & Knoll, R. L. in *Attention and Performance IV* (ed. Kornblum, S.) 629–685 (Academic Press, 1973).
28. Tavassoli, A. & Ringach, D. L. Dynamics of smooth pursuit maintenance. *J. Neurophysiol.* **102,** 110–118 (2009).
29. Mulligan, J. B., Stevenson, S. B. & Cormack, L. K. Reflexive and voluntary control of smooth eye movements. in (eds. Rogowitz, B. E., Pappas, T. N. & de Ridder, H.) **8651,** 86510Z1–22 (SPIE, 2013).

30. Knöll, J., Pillow, J. W. & Huk, A. C. Lawful tracking of visual motion in humans, macaques, and marmosets in a naturalistic, continuous, and untrained behavioral context. *Proc. Natl. Acad. Sci.* **115,** E10486–E10494 (2018).
31. Downey, C., Pace, G., Seemiller, E., Candy, R. & Cormack, L. Dynamic Characteristics of 5 to 22 week-old Infants' Accommodation and Vergence Tracking Responses. *Journal of Vision* **17,** 443–443 (2017).
32. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415,** 429–433 (2002).
33. Kording, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427,** 244–247 (2004).
34. Geisler, W. S. Contributions of ideal observer theory to vision research. *Vision Research* **51,** 771–781 (2011).
35. Burge, J. Image-Computable Ideal Observers for Tasks with Natural Stimuli. *Annu Rev Vis Sci* (2020). doi:10.1146/annurev-vision-030320-041134
36. Knoblauch, K. & Maloney, L. T. Estimating classification images with generalized linear and additive models. *Journal of Vision* **8,** 1–19 (2008).
37. Macke, J. H. & Wichmann, F. A. Estimating predictive stimulus features from psychophysical data: The decision image technique applied to human faces. *Journal of Vision* **10,** 1–24 (2010).
38. Murray, R. F. Classification images and bubbles images in the generalized linear model. *Journal of Vision* **12,** 1–8 (2012).
39. Jaini, P. & Burge, J. Linking normative models of natural tasks to descriptive models of neural response. *Journal of Vision* **17,** 1–26 (2017).
40. Burge, J. & Geisler, W. S. Optimal speed estimation in natural image movies predicts human performance. *Nat Commun* **6,** 7900 (2015).
41. Chin, B. M. & Burge, J. Predicting the Partition of Behavioral Variability in Speed Perception with Naturalistic Stimuli. *J. Neurosci.* **40,** 864–879 (2020).
42. Rodriguez-Lopez, V., Dorronsoro, C. & Burge, J. Contact lenses can cause the reverse Pulfrich effect and anti-Pulfrich monovision corrections can eliminate it. *bioRxiv* 1–22 (2020). doi:10.1101/2020.04.05.026534
43. Nachmias, J. Effect of Exposure Duration on Visual Contrast Sensitivity with Square-Wave Gratings. *J Opt Soc Am* **57,** 421–427 (1967).
44. Shapley, R. M. & Victor, J. D. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.* **285,** 275–298 (1978).
45. Levi, D. M., Harwerth, R. S. & Manny, R. E. Suprathreshold spatial frequency detection and binocular interaction in strabismic and anisometropic amblyopia. *Invest. Ophthalmol. Vis. Sci.* **18,** 714–725 (1979).
46. Albrecht, D. G. Visual cortex neurons in monkey and cat: effect of contrast on the spatial and temporal phase transfer functions. *Visual Neurosciences* **12,** 1191–1210 (1995).
47. Bair, W. & Movshon, J. A. Adaptive temporal integration of motion in direction-selective neurons in macaque visual cortex. *J. Neurosci.* **24,** 7305–7323 (2004).
48. Min, S. H., Reynaud, A. & Hess, R. F. Interocular Differences in Spatial Frequency Influence the Pulfrich Effect. *Vision* **4,** 1–13 (2020).
49. Campbell, F. W. & Green, D. G. Optical and retinal factors affecting visual resolution. *J. Physiol.* **181,** 576–593 (1965).
50. Navarro, R., Artal, P. & Williams, D. R. Modulation transfer of the human eye as a function of retinal eccentricity. *J Opt Soc Am A* **10,** 201–212 (1993).
51. Burge, J. & Geisler, W. S. Optimal defocus estimation in individual natural images. *Proc. Natl. Acad. Sci.* **108,** 16849–16854 (2011).

Supplement

Target tracking reveals the time course of visual processing with millisecond-scale precision

Johannes Burge[1.,2.,3.] & Lawrence K. Cormack[4.,5,6.]

[1.] Department of Psychology, University of Pennsylvania
[2.] Neuroscience Graduate Group, University of Pennsylvania
[3.] Bioengineering Graduate Group, University of Pennsylvania
[4.] Department of Psychology, University of Texas at Austin
[5.] Center for Perceptual Systems, University of Texas at Austin
[6.] Institute for Neuroscience, University of Texas at Austin

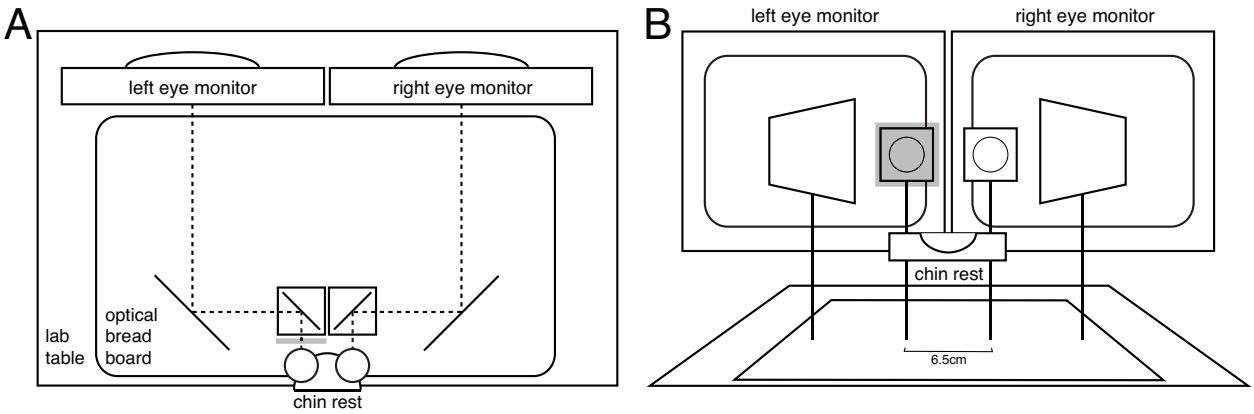


Figure S1. Haploscope apparatus for stimulus delivery. Each eye has its stimulus delivered by a dedicated monitor. The light bounces off two front-surface mirrors on its way to each eye (dashed lines) **A** Top down view. The dashed lines indicate the light path from the monitors to the eyes. A neural density filter is shown in front of the left eye. **B** Head on view. A neutral density filter is depicted in front of the left-eye viewport.

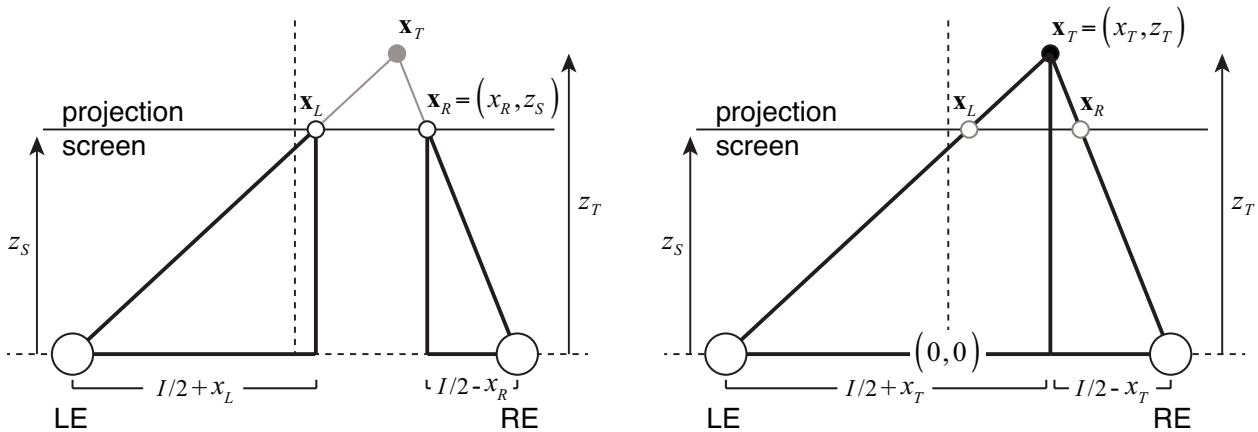


Figure S2. Projection geometry, similar triangles, and rendering positions. For a given target position $\mathbf{x}_T = (x_T, z_T)$ in stereoscopic space, the left- and right-eye stimuli must be rendered on the projection screen at positions $\mathbf{x}_L = (x_L, z_S)$ and $\mathbf{x}_R = (x_R, z_S)$, respectively. These positions are easily obtained using similar triangles (bold line segments in left and right subplots). Left-eye image position is obtained by solving the expression $(I/2 + x_L)/(I/2 + x_T) = z_S/z_T$ for x_L (Eq. 4a). Right-eye image position is obtained by solving the expression $(I/2 - x_R)/(I/2 - x_T) = z_S/z_T$ for x_R (Eq. 4b). In the actual apparatus, the left-eye stimulus and the right eye stimulus were displayed on separate monitors; but the geometry remains the same. Note that the drawings are not to scale.

Supplementary Derivation

In the main text, we asserted in Eq. 1 that the difference between the left- and right-eye impulse response functions are proportional to the cross-correlation between the target velocity in X and the response velocity in Z

$$\dot{x}_T \otimes \dot{\tilde{z}}_T \propto h_R - h_L \tag{S1}$$

where $\dot{x}_T$ is the X velocity of the target, $\dot{\tilde{z}}_T$ is the Z velocity of the human response, and h_L and h_R are the left- and right-eye impulse response functions. Here, we provide the derivation. Throughout the derivation (and in the main text), variables with dots denote velocities and variables with tildes indicate that they have been acted on by impulse response functions.

First, we substitute an expression for $\dot{x}_T$ that has the form of Eq. 7a and substitute the expression for $\dot{\tilde{z}}_T$ in Eq. 7b to obtain

$$\dot{x}_T \otimes \dot{\tilde{z}}_T = \frac{I}{2}\left[\frac{2\left(x_L\dot{x}_R - \dot{x}_L x_R\right) + I\left(\dot{x}_L + \dot{x}_R\right)}{\left(x_L - x_R + I\right)^2}\right] \otimes Iz_S\left(\frac{\dot{\tilde{x}}_R - \dot{\tilde{x}}_L}{\left(\tilde{x}_L - \tilde{x}_R + I\right)^2}\right) \tag{S2}$$

Given that the display screen was far from the observer (z_s =1000mm) and that the virtual position of the target in depth was never far from the display screen, the on-screen disparities (see Eq. 11) were always small relative to the interocular distance (i.e. a typical interocular distance I is 65mm, whereas 68% of the values assumed by $x_L - x_R$ and $\tilde{x}_L - \tilde{x}_R$ are between $\pm$1.2mm and 99% are between $\pm$3.9mm). Hence, $\left(x_L - x_R + I\right)$ and $\left(\tilde{x}_L - \tilde{x}_R + I\right)$ are approximately equal to I, allowing us to write

$$\cong \frac{I}{2}\left[\frac{2\left(x_L\dot{x}_R - \dot{x}_L x_R\right) + I\left(\dot{x}_L + \dot{x}_R\right)}{I^2}\right] \otimes Iz_S\left(\frac{\dot{\tilde{x}}_R - \dot{\tilde{x}}_L}{I^2}\right) \tag{S3}$$

Grouping all the constant factors on the left-hand side and simplifying yields

$$\cong \frac{z_S}{2I^2}\left[2\left(x_L\dot{x}_R - \dot{x}_L x_R\right) + I\left(\dot{x}_L + \dot{x}_R\right)\right] \otimes \left(\dot{\tilde{x}}_R - \dot{\tilde{x}}_L\right) \tag{S4}$$

Now, note that the first term in square brackets on the left-hand side has a positive and a negative term that are approximately equal because x_L and x_R are approximately equal to x_T and because $\dot{x}_L$ and $\dot{x}_R$ are approximately equal to $\dot{x}_T$. (Note that these approximate equalities again hold because the on-screen disparities were small.) These terms approximately cancel, allowing us to write

$$\cong \frac{z_S}{2I}\left[\left(\dot{x}_T + \dot{x}_T\right)\right] \otimes \left(\dot{\tilde{x}}_R - \dot{\tilde{x}}_L\right) \tag{S5}$$

Expanding the second term by plugging in the effective left- and right-eye image velocities from the velocity equivalents of Eqs. 5a and 5b and simplifying gives

$$= \frac{z_S}{I}\left(\dot{x}_T\right) \otimes \left(\dot{x}_R * h_R - \dot{x}_L * h_L\right) \tag{S6}$$

Substituting $\dot{x}_T$ for $\dot{x}_L$ and $\dot{x}_R$ gives

$$\cong \frac{z_S}{I}\left(\dot{x}_T\right) \otimes \left(\dot{x}_T * h_R - \dot{x}_T * h_L\right) \tag{S7}$$

Using the distributive property of convolution

$$= \frac{z_S}{I}\left(\dot{x}_T\right) \otimes \left(\dot{x}_T * \left(h_R - h_L\right)\right) \tag{S8}$$

Using the associative property of cross-correlation and convolution

$$= \frac{z_S}{I}\left(\dot{x}_T \otimes \dot{x}_T\right) * \left(h_R - h_L\right) \tag{S9}$$

The autocorrelation of $\dot{x}_T$ is a delta function centered at zero because the horizontal target velocities are distributed as white Gaussian noise by experimental design

$$= \frac{z_S}{I}\left(\delta_0\right) * \left(h_R - h_L\right) \tag{S10}$$

Convolving a delta function centered at zero with an arbitrary function yields the arbitrary function

$$= \frac{z_S}{I}\left(h_R - h_L\right) \tag{S11}$$

Finally, dropping the scale factors yields the proportionality asserted in the main text

$$\dot{x}_T \otimes \dot{\tilde{z}}_T \propto h_R - h_L \tag{S12}$$