

# Disparity-energy signals in perceived stereoscopic depth

**Seiji Tanabe**

Graduate School of Frontier Biosciences, Osaka University,  
Toyonaka, Osaka, Japan



**Satoko Yasuoka**

Graduate School of Frontier Biosciences, Osaka University,  
Toyonaka, Osaka, Japan



**Ichiro Fujita**

Graduate School of Frontier Biosciences, Osaka University,  
Toyonaka, Osaka, Japan



Stereopsis, the ability to sense the world in three dimensions (3D) from pairs of retinal images, functions when both images have corresponding elements. When observers view stereograms lacking a global match, they do not perceive 3D structure, whereas several cortical areas encode stereoscopic depth in the disparity energy. Whether these neural representations are exploited or ignored in perceptual decisions remains elusive. By combining contrast-reversal and delay between stereo images, we found that disparity-energy signals mediate the reversal of stereoscopic depth judgments. A crisp, adjacent plane of reference was crucial for the signal to be used in the judgments. Disparity discrimination relies on the disparity-energy signal when the stimulus has no global binocular match and is accompanied by a fixed surface of reference.

**Keywords:** stereopsis, disparity, random-dot stereogram, energy model

**Citation:** Tanabe, S., Yasuoka, S., & Fujita, I. (2008). Disparity-energy signals in perceived stereoscopic depth. *Journal of Vision*, 8(3):22, 1–10, <http://journalofvision.org/8/3/22/>, doi:10.1167/8.3.22.

## Introduction

The human visual system has a remarkable capability of deriving 3-dimensional (3D) surface structure from a pair of retinal images. The geometry of the visual projections poses a problem for the stereoscopic system to correctly match common features within both retinal images, known as the correspondence problem (Marr & Poggio, 1979). Binocular neurons in the primary visual cortex (V1) compute the disparity energy of the binocular images filtered through their receptive fields (RFs) (Ohzawa, DeAngelis, & Freeman, 1990). This computation is similar to calculating the cross-correlation of the binocular images and constitutes the initial encoding of binocular disparity. However, the encoded information disagrees with stereo percepts in some circumstances (Cumming & DeAngelis, 2001). When one of the paired images in a stereogram (a half-image) is the contrast-reversal of the other, the disparity information is encoded in the inverted, nonzero disparity energy (Cumming & Parker, 1997; Ohzawa et al., 1990), while sensation of 3D structure is abolished (Julesz, 1960). Similarly, when the sequence of an image for one eye is delayed relative to that for the other eye, the disparity energy is again inverted and nonzero for neurons with biphasic temporal RFs (Anzai, Ohzawa, & Freeman, 2001), whereas stereo performance only diminishes (Read & Cumming, 2005).

It has been controversial whether stereo percepts reflect these finite, inverted disparity-energy signals. The prevailing view asserts that the signals simply do not give rise to stereo percepts (Cumming & DeAngelis, 2001; Cumming & Parker, 1997). A handful of studies employing behavioral measures demonstrate that the signals do. However, solid evidence remains elusive because the depth direction can be either correct with respect to geometrical matches (Cogan, Kontsevich, Lomakin, Halpern, & Blake, 1995; Cogan Lomakin, & Rossi, 1993; Cumming, Shapiro, & Parker, 1998; Helmholtz, 1909) or reversed as encoded in the disparity energy (Hayashi, Miyawakai, Maeda, & Tachi, 2003; Masson, Bussetini, & Miles, 1997; Neri, Parker, & Blakemore, 1999; Read & Eagle, 2000; Rogers & Anstis, 1975).

A few confounds may underlie this marked discrepancy in the literature. First, to test the perceived stereoscopic depth of normal and inverted disparity-energy stereograms under equal condition, the distinction must be invisible to visual processes other than stereopsis. A contrast-reversed random-dot stereogram appears to have a texture of a lustrous aggregate of dots, while the original stereogram has a texture of a crisp surface. By randomly combining contrast-reversal with interocular delay, we can reduce the confounding effect of textural distinction because both manipulations to the stereogram eliminate the sensation of a crisp surface when applied individually (Cumming et al., 1998; Ross, 1974; Tyler, 1974). Another confound is the

fact that subjects need a reference for discrimination. Fixing the disparity energy of the reference helps anchor the subject's discrimination criterion across stimulus manipulations. Once the confounds are eliminated, the disparity-energy model predicts that perceived depth reverses with interocular delay even when the pair of stereo images have the same contrast, and that when a half-image is contrast-reversed, the perceived depth is the reversal of whatever the perceived depth was with the interocular delay of the original contrast. Although flashing stationary random dots are known to yield an interactive effect of contrast-reversal and delay on stereo percepts (Cogan et al., 1993), it is unclear whether this interaction agrees with the disparity-energy computation. We tested the predictions of the energy model on the stereo percepts of human subjects.

## Methods

### Depth discrimination experiment

Subjects were seated 57 cm in front of a computer display (Multiscan E230, Sony) with their heads restrained on a chin rest. On each trial, subjects viewed a bipartite RDS (Figure 1) through stereo shutter glasses (RE7-CANE, Elsa) for a period of 1 s and reported by clicking the right or left mouse-button whether the RDS center was seen nearer or farther than the annulus (Figures 2A and 2B). If the subject did not make a forced choice during the 1-s response epoch, the same trial was randomly inserted later. The center randomly varied from trial to trial in binocular disparity (crossed or uncrossed,  $0.32^\circ$ ) and interocular delay (12 to 156 ms) and was either binocularly

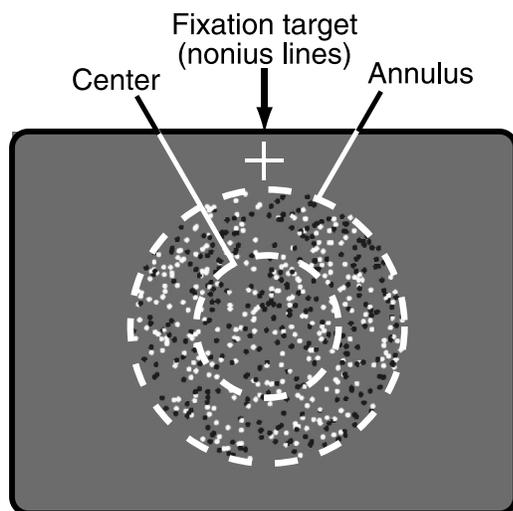


Figure 1. Configuration of visual stimulus. Dotted contours were added for illustration and were not visible to the subjects.

correlated or anticorrelated. The annulus was fixed at zero disparity and at minimal interocular delay and was always binocularly correlated, except for the experiments shown in Figure 5. The diameters of the center and the annulus were  $4.8^\circ$  and  $8.0^\circ$ , respectively. Dot size was  $0.14^\circ \times 0.14^\circ$ , dot density was 25%, and dot pattern was renewed every frame (43 Hz). Dot positions followed a uniform distribution, and their images were antialiased for sub-pixel resolution. A dot occluded another dot when they overlapped one another. The bright dots, the dark dots, and the background had luminance values of 1.90, 0.01, and  $0.95 \text{ cd/m}^2$ , respectively. Luminance measurements were made with lights passing through the filters of the dichoptic device. Computer graphics and data collection were performed with a custom-made program using the OpenGL Utility Toolkit (GLUT). The stereo shutter glasses alternated at 85 Hz. The cross-talk between the left- and right-eye images was 3%. Proportion of correct judgments were calculated from a block of 20 trials per delay value, in which half of the trials had crossed disparity and the other half had uncrossed disparity. Each subject performed at least three blocks. The experimental protocol was approved by the research ethics committee of Osaka University. Written informed consent was obtained from all subjects.

### Disparity-defined shape discrimination experiment

The same subjects discriminated the orientation of a “T”-shaped surface. All parameters of the dynamic RDS were the same as in the depth discrimination task. The “T”-shape was composed of two perpendicular bars that were  $1.6^\circ \times 0.4^\circ$ . For one of the subjects, a “T”-shape of  $3.2^\circ \times 0.8^\circ$  was also tested. The surface defining the “T”-shape was given a disparity of  $0^\circ$ , while the rest of the center region was given a crossed disparity of  $0.32^\circ$ . To vary the position of the “T”-shape, one location was randomly selected for each trial with an equal probability among five locations. The center randomly varied from trial to trial in the tilt direction of the “T”-shape (clockwise or counterclockwise,  $90^\circ$ ) and interocular delay (12 to 156 ms) and was either binocularly correlated or anticorrelated. Proportion of correct judgments was calculated from a block of 20 trials per delay value, in which half of the trials had clockwise tilt and the other half had counterclockwise tilt. Each subject performed three blocks.

### Model simulation

To address whether the depth discrimination follows the prediction of the disparity-energy model, we simulated the output of a model with spatiotemporal RFs. The RFs were space–time separable and expressed in terms of  $W(x, y, \tau) =$

$w_s(x, y) \times w_t(\tau)$ , in arbitrary units. The functions  $w_s(x, y)$  and  $w_t(\tau)$  describe the spatial and the temporal dimensions of the monocular RFs, respectively. The right eye's RF was modeled in the form of

$$w_s(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \cdot \cos(2\pi f_s x) \quad (1)$$

$$w_t(\tau) = A \cdot \tau \cdot \exp(-\alpha\tau) \cdot \cos(2\pi f_t x + \phi).$$

where  $\sigma = 2$  pix,  $f_s = 0.2$  cyc/pix,  $\alpha = 93.8$ ,  $A = 8789$ ,  $f_t = 10.8$  Hz, and  $\phi = 0.1\pi$ . The left eye's RF was given a  $90^\circ$  phase shift in  $w_s(x, y)$ . The mathematical expression of  $w_t(\tau)$  was originally modeled after typical cells in area 17 of cats (Chen, Wang, & Qian, 2001), but we modified the parameter values in order to match the sensitivity to higher temporal frequencies of typical cells in the V1 of macaque monkeys. The dip of  $w_t(\tau)$  has a typically 50-ms longer latency than the peak for monkey V1 cells (DeValois, Cottaris, Mahon, Elfar, & Wilson, 2000). The model parameters used in our simulations were chosen to capture these characteristics of monkey V1 cells.

A monocular image was a grid of dots consisting of 25% bright and 25% dark dots with 50% gray background. The contrast values of bright, dark, and gray dots were assigned, 1,  $-1$ , and 0, respectively, in arbitrary units. A sequence of 1875 images was delivered at 85 Hz. A blank screen always intervened between successive frames. All numerical calculations were run in the MATLAB environment (MATLAB, The Mathworks).

## Results

Human observers were asked to judge stereoscopic depth in binocularly correlated and anticorrelated random-dot stereograms (cRDSs and aRDSs, respectively) with various interocular delays. Subjects fixated on a small cross with nonius lines  $4.8^\circ$  above the center of the screen. While the subjects maintained alignment of the nonius lines, a bipartite RDS patch ( $8^\circ$  in diameter) was presented for 1 s at the center of the screen (see [Methods](#) section). Subjects reported by clicking a mouse button whether the center of the RDS was seen nearer or farther than its immediately surrounding dots. The surrounding annular region of the RDS was always binocularly correlated, had 0 disparity, and was presented with the minimum interocular delay (12 ms) available with our dichoptic device (see [Methods](#) section). The correct choice was defined as a “near” response for crossed and a “far” response for uncrossed disparities in both cRDS and aRDS trials ([Figures 2A](#) and [2B](#)). Correct choices were defined solely for off-line analysis, and no feedback was given to the subjects during the experiments.

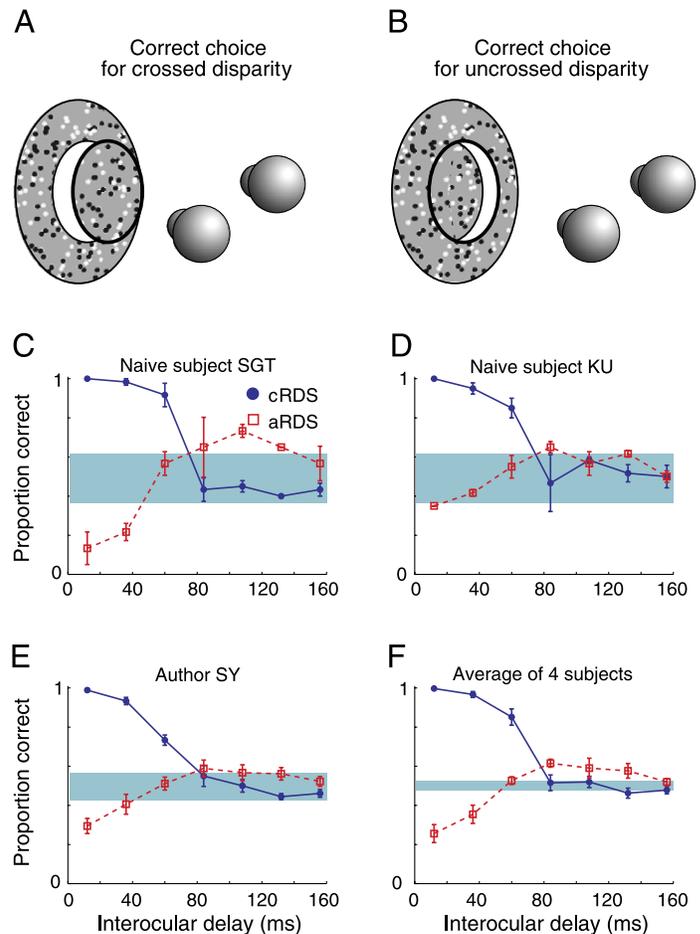


Figure 2. Depth judgments of cRDSs and aRDSs with interocular delay. (A, B) Regardless of cRDS or aRDS, choices were considered as correct if a subject responded as seeing “near” for crossed disparity trials, or if the subject responded as seeing “far” for uncrossed disparity trials. (C, D, E) Proportion of correct choices is plotted against interocular delay in two naives and one author. The error bars depict the *SEM*. Panels C and D are based on data collected from three blocks of 20 trials. Panel E is based on nine blocks. Gray area indicates the 95% confidence interval of chance performance following a binomial distribution. A performance below chance indicates reversed sensation of depth. (F) Average proportion of correct choices across four subjects. One more subject was included in addition to the three individuals shown in panels C–E. The added subject did nine blocks.

## Perception of stereoscopic depth

[Figures 2C–2E](#) show the proportion of correct choices as a function of interocular delay for three subjects. For cRDSs, performance decreased with interocular delay until it reached chance level at roughly 80 ms.

For aRDSs with the minimum delay, the proportion of correct choices fell below chance in all the subjects ([Figures 2C–2F](#)). This indicates a reversal, not absence, of stereoscopically perceived depth. As interocular delay was added, the proportion correct for aRDSs increased toward

chance. It reached a maximum at 108 ms where the average across the subjects was significantly above chance ( $p < 0.05$ , binomial test), indicating “correct” depth judgments with an aRDS (Figure 2F). The proportion correct as a function of interocular delay was inverted in shape relative to the function for cRDSs about the horizontal line representing chance performance. Given that the binocular correlations for cRDSs and aRDSs are sign-inversions of each other, the proportion correct can be described as a product of one function of binocular correlation and another function of interocular delay. Note, however, that the separability was not perfect because the dashed curve for aRDSs was shallower than the solid curve for cRDSs.

Surprisingly, but not necessarily contradictory to the significant depth judgments, the subjects verbally reported that they did not see a planar surface in most trials. The subjects reported that the center of the RDS looked like a lustrous cloud of dots spread in 3D, which they managed to locate as either near or far relative to the annular plane. We asked for the verbal report only after a whole session was finished. In this experiment, there was no record of the conditions or the trials on which the subjects saw a surface. The perception of a surface was tested in the next experiment.

## Perception of cyclopean surface

Presumably, the subjects discriminated stereoscopic depth without perceiving a “cyclopean” plane when the stereogram either was anticorrelated or had interocular delay. The term “cyclopean” refers to properties of visual percepts mediated only by stereoscopic mechanisms (Julesz, 1971; Tyler, 1990). To test the perception of cyclopean surfaces, we conducted a second psychophysical experiment. Three subjects from the preceding experiment were asked to discriminate a disparity-defined shape that was embedded in the center portion of the RDS. The cyclopean shape was a “T”-oriented  $90^\circ$  clockwise or counterclockwise (Figures 3A and 3B). Dots in the region defining the “T”-shape had zero disparity, while those in the rest of the center region had a crossed disparity of  $0.32^\circ$  (see Methods section). With the oriented “T”-shapes, the cross-correlation between a series of contrast values along the horizontal dimension of the left- and right-eye images is equal between the two shapes at any vertical position. The subjects could not discriminate the two shapes based on the disparity energy of the entire center disk. In addition, the position of the oriented “T”-shape was scattered across five locations from trial to trial such that even local disparity-energy signals from a fixed retinal position would not provide information for discriminating the two shapes.

All three subjects easily discriminated the cyclopean shape for cRDSs at the minimum interocular delay ( $p < 0.05$ , binomial test), but not for an aRDS at this delay or either of

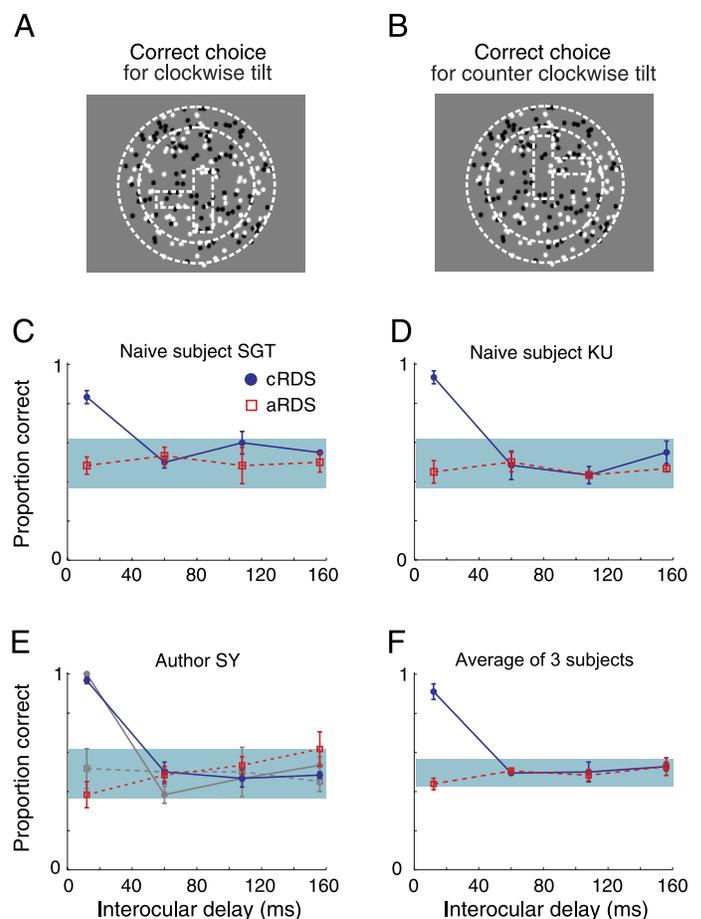


Figure 3. Disparity-defined shape discrimination. (A) A “T”-shape with a clockwise,  $90^\circ$  tilt is illustrated inside the center disk. Dotted contours are added for illustration and were not present in the actual stimulus. (B) The “T”-shape with a counterclockwise,  $90^\circ$  tilt is illustrated in a position different from the one in panel A. (C, D, E) Proportion of correct choices is plotted against interocular delay for two naives and one author. Plotting conventions are the same as in Figures 2C–2E. The gray curves in E are the results of a control experiment where the sizes of the patch and T-shape were doubled. (F) Average proportion of correct choices across the three subjects.

the two RDSs with any other interocular delay ( $p > 0.05$ , binomial test) (Figures 3C–3F). To check whether the shape was too small for discriminating the orientation, we performed the same experiment with a larger stimulus with one of the subjects. The diameters of the center and the annular patches were doubled. The “T”-shape was also doubled, bringing the bar width to  $0.8^\circ$ . Essentially, the same results were obtained with the enlarged stimulus (Gray curves in Figure 3E). The smallness of the “T” region does not explain the lack of cyclopean surfaces for inverted disparity-energy stimuli. This verifies the verbal responses by the subjects and demonstrates that humans are not able to see a cyclopean surface of a shape even in the stereo images that produce stereoscopic depth.

## The disparity-energy computation

The psychophysical performance of depth discrimination (Figure 2) strikingly resembled the output of a disparity-energy model incorporating a temporal filter. A sequence of RDSs displaced along time between the two eyes was fed into the model as an input. We simulated our psychophysical experiment where a blank screen intervened successive random patterns (Figure 4A). On each frame, one eye was presented with random dots while the other eye was presented with the blank screen. The output of the filtering stage at a given time was the spatiotemporal sum of the monocular image sequence weighed according to the RF profile ( $X_R$  and  $X_L$ ; Figure 4B). The RFs were modeled after the properties of typical neurons in monkey V1 (see Methods section). The RF structure

was space–time separable, in that it was expressed in terms of multiplication of one function of space and another function of time. The spatial dimension of the RF was a Gabor function, while the time dimension of the RF was biphasic with a primary peak (15 ms) followed by a slower trough (69 ms).

The disparity modulation decreased at an interocular delay of 24 ms and inverted with an interocular delay of 60 ms (Figure 4C). We assessed the tuning amplitude as the response to the preferred disparity minus the response to the null disparity at each interocular delay, in which the preference was based on the tuning for cRDSs with the minimum delay. As the interocular delay increased, the response amplitude for cRDSs gradually decreased, then it reached a minimum at 60 ms (filled circles in Figure 4D). The sensitivity of the model to interocular delay closely

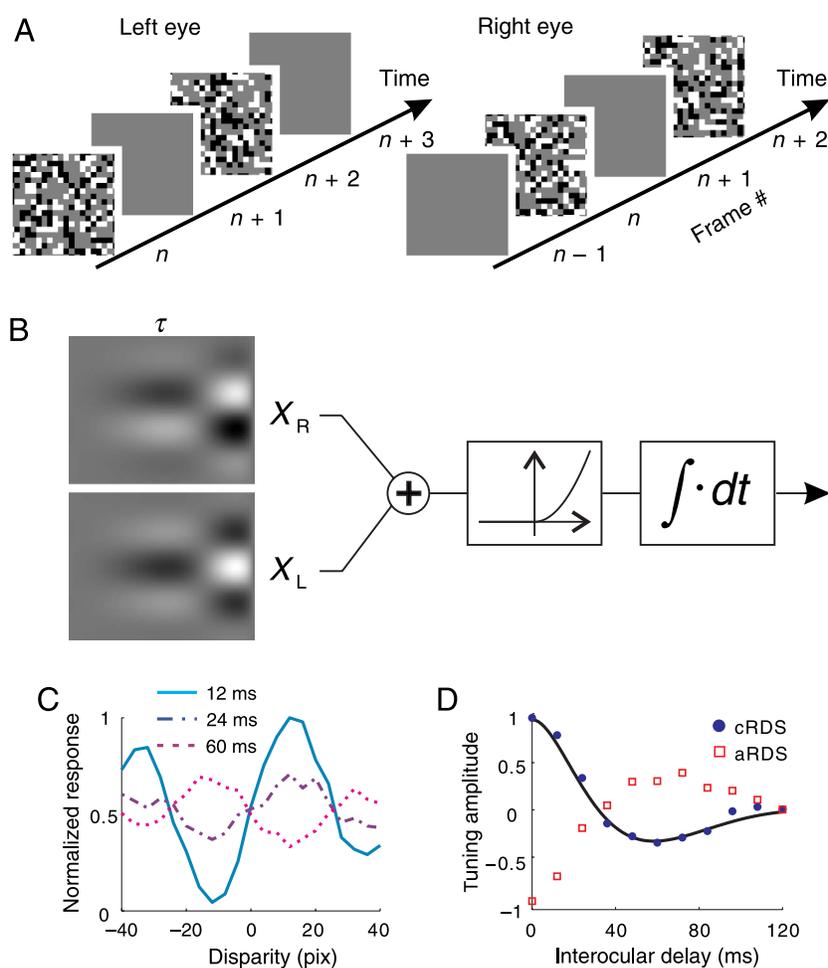


Figure 4. Disparity-energy model with spatiotemporal RFs. (A) The right-eye stimulus images in this example are delayed by one frame relative to the left-eye stimulus images. The blank screen intervening successive frames simulates the image sequence delivered in the psychophysical experiments. (B) The RF of a model neuron is defined in spatial ( $X$ ,  $Y$ ) and temporal ( $\tau$ ) coordinates for left-eye (L) and right-eye (R) inputs (indicated here in grey scale for  $X$  and  $\tau$  only). The model outputs the temporal integration of the signal that passes through the static nonlinear filter. (C) The disparity tuning of the model in response to cRDSs plotted for three interocular delays (12, 24, 60 ms). The response magnitude is normalized to the maximum at 12 ms delay. (D) The amplitude of disparity tuning plotted against interocular delay for cRDSs (filled circles) and aRDSs (open squares). The solid curve indicates the cross-correlation of the left and right temporal filters normalized to a maximum value of one.

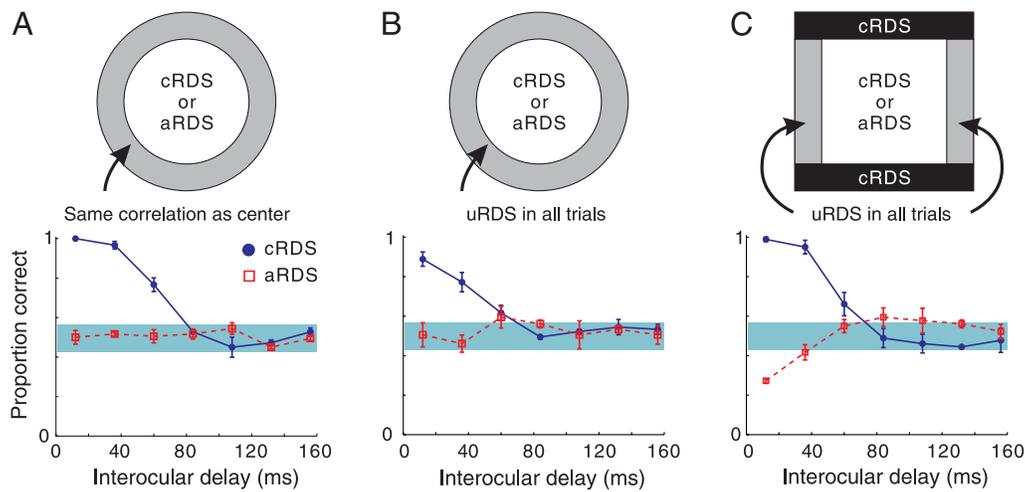


Figure 5. Influence of surround properties on stereo discrimination. (A) Surround dots were switched to aRDSs when the center portion being tested was an aRDS. The results replicate previous studies showing abolished stereo performance for aRDSs. Plotting conventions are the same as in Figures 2C–2E. (B) Surround dots were binocularly uncorrelated in all trials. (C) The RDS patch was rectangular. The top and bottom portions of the surround were correlated, while the left and the right portions were uncorrelated. In all three panels, the proportion of correct choices is plotted against interocular delay for two naïves and one author.

matched with the cross-correlation of the left and right temporal filters with appropriate scaling (solid curve). This close match indicates that the intervening blank screens did not confound the predictions of the model. The amplitude for aRDSs was exactly the sign inversion of that for cRDSs at all interocular delays (open squares).

The model output was in accordance with the main features of the psychophysical data. For any interocular delay, if the percentage correct for cRDSs was above chance level, the percentage for aRDSs was below chance level, and vice versa. There were, however, two notable deviations of the model from psychophysical data. One was the time scale, and the other was the attenuated depth discrimination for stimuli with inverted disparity-energy signals. In the psychophysical data, the distance from chance level was smaller in the bottom half than in the top half in both functions representing cRDSs and aRDSs, respectively. The attenuation is instantiated at two interocular delay values. With the minimum interocular delay 12 ms, the percentage of correct depth was 50% above chance for the cRDSs but only 20% below chance for the aRDSs. With an interocular delay of 84 ms, the percentage of correct depth was not lower than chance level for the cRDSs, but 10% above chance for the aRDSs.

### Cyclopean surface as reference

In the previous studies where aRDSs abolished rather than reversed stereoscopic depth, both the test and the surrounding regions of the RDS patch were anticorrelated (Cumming & DeAngelis, 2001; Cumming & Parker,

1997; Julesz, 1960). When we tested our subjects with this same configuration, we replicated the abolished stereo judgments (Figure 5A). Subjects performed similarly when the surround was uncorrelated (Figure 5B). We hypothesized that a crisp reference surface in the vicinity is a crucial factor that determines the abolishment or reversal of stereoscopic depth in aRDSs. We tested subjects with a rectangular patch of RDS whose top and bottom portions of the surround were correlated and left and right portions were uncorrelated. The uncorrelated dots in the surround prevent inference of depth based on monocular portions occluded from the other eye (Nakayama & Shimojo, 1990). The subjects reported reversed stereoscopic depth for aRDSs (Figure 5C) in a manner similar to the main test (shown in Figures 2C–2F). Thus, we conclude that, at least in a two-alternative forced-choice task, a crisp cyclopean surface of reference is crucial for the reversed depth in aRDSs.

## Discussion

We demonstrated that observers could discriminate stereoscopic depth in aRDSs when the annular surround of the RDS is correlated and provides a crisp reference. We also revealed that the sensitivity of stereo discrimination to interocular delay and binocular correlation are both accounted for by the disparity-energy signal with biphasic temporal filters. The disparity-energy signal mediated the reversal of stereoscopic depth when the stereoscopic depth was dissociated from the perception of cyclopean surface.

## Relationship with previous studies

First, we wish to set aside the studies using solid-figure stereograms (Helmholtz, 1909; Cogan et al., 1995). Although anticorrelated solid figures are consistently perceived in the correct depth, there is no widely accepted model that can explain this percept. Therefore, we overview only the studies that employ random-dot stereograms. Among these studies, there is only one that reported correct depth with aRDSs (Cumming et al., 1998). In their study, subjects were presented with only aRDSs and were reinforced to report the correct depth by receiving feedback. This reinforcement most likely affected the sensory-motor transformation of the subjects. The remaining studies either report no depth perception or reversed depth perception with aRDSs.

The appearance of a crisp surface for cRDSs and the lack thereof for aRDSs may undermine previous psychophysical experiments. Subjects could have reported depth randomly without trying to determine the depth of the stimulus when they did not see a crisp surface. If subjects did not use the available disparity-energy signal, then the loss of performance in aRDS trials could result from a switch to ignoring the signal. Novice subjects are particularly liable to switch, because they judge no depth in aRDSs, while experienced subjects report reversed depth (Hayashi et al., 2003). Studies that have proposed a loss of perception for aRDSs have mainly relied on demonstrations (by having readers view a stereogram printed on the paper) and have put less weight on experimental evidence (Julesz, 1960; Cumming & Parker, 1997; Masson et al., 1997; Janssen, Vogels, Liu, & Orban, 2003). The paucity of quantitative evidence reflects the inherent flaw that the lack of perception can only be indicated by negative data. Our study mitigates the flaw by distinguishing the divide of cRDS/aRDS stimuli from the divide of surface/cloud percepts. This design prevented the subjects from switching their perceptual strategies between cRDS and aRDS trials and thereby enabled us to assess the encoded disparity equally.

Some studies have shown a similar reversal of stereoscopic depth with aRDSs. In an early study, as an anticorrelated image with disparity was blended into a correlated image without disparity, the stereoscopic depth appeared in the reversed direction (Rogers & Anstis, 1975). More recently, when subjects detected signal-dots that shared one disparity among noise-dots whose disparities were randomized, the psychophysical reverse-correlation of noise-dots revealed kernels resembling the disparity energy. Binocularly correlated noise-dots help the detection of signal-dots, while anticorrelated noise-dots suppress the detection (Neri et al., 1999). In addition, when subjects discriminated the disparity of two successive aRDSs, the stereoscopic depth was weakly reversed when there was no surround (Read & Eagle, 2000). Thus, the reversal of depth perception that we found with anticorrelated stereograms is consistent with these studies.

The reversal of perceived depth with aRDSs is only one feature of our data. One of the most important results is the dissociation between depth perception and cyclopean-surface perception. The dissociation implies that there are two mechanisms for mediating the respective perception. The striking similarity between the depth perception and the disparity-energy model suggests that one of the two mechanisms mediate disparity-energy signals. Others have reported another form of dissociation in stereo processing (Ziegler & Hess, 1999), but their interpretation contrasts with ours. They suggest that the mechanism detects features in the monocular image prior to binocular combination. In addition to the dissociation, we revealed that the binocular correlation of the surround plays a crucial role in mediating those signals. When the surround was either anticorrelated (Figure 5A; Read & Eagle, 2000) or uncorrelated (Figure 5B), the perceived depth no longer reflected the disparity energy. The perception of a cyclopean surface in the background is necessary for the disparity-energy signal to be used in stereo discrimination.

## Vergence eye movement

In this study, we did not measure eye position. Potentially, the effect of the surround in Figure 5 is explained by the vergence being anchored better with a cRDS reference than without. This would be a simpler explanation than the presence of the reference per se. The vergence explanation is likely only if both of the following were true: (A) stereoscopic depth is encoded in an absolute reference frame, and (B) trial-to-trial variability of the vergence angle was larger when the vicinity of fixation was an aRDS than when it was a cRDS. Although we are uncertain about what reference frame was used in this task, we suspect that an absolute reference frame was used because of the similarity between the task in Figure 5 and a study that directly tested the reference frame (Uka & DeAngelis, 2006). In their “coarse task”, as well as in ours, the strength of the disparity-energy signal was varied across trials, while keeping the encoded disparity constant, apart from a sign change. As the disparity in their “coarse task” was encoded in an absolute reference frame, we deduced that the disparity in our task is similarly encoded in an absolute reference frame. Thus, the condition (A) holds. However, while we do not have eye position data of the experiment in Figure 5, we think that the variability of the vergence angle was not different whether an aRDS or a cRDS was in the vicinity of fixation. This is based on a previous study that directly measured eye position while subjects viewed a similar stimulus (Masson et al., 1997). In this study, the variance of vergence for a zero-disparity aRDS was not different from that for a zero-disparity cRDS. It is also worth noting that the variance is expected to be smaller in our experiment than theirs because the vergence was deliberately fixed with a nonius line in our

experiment, while it was open to involuntary movement in theirs. In summary, condition (A) holds but (B) is unlikely to hold; therefore, we do not think that the results in [Figure 5](#) were confounded by vergence posture.

## Comparison between model and data

Our model simulation captured the main feature of the psychophysical data; for any interocular delay, the percentage correct on cRDS trials and the percentage correct on aRDS trials were on opposite sides of 50%. Although the qualitative profile of the feature agreed between the model and the data, there were some quantitative differences. The model predicted a mirror-symmetry about a horizontal line at 50% correct ([Figure 4D](#)), while the data revealed compression toward the 50% line below the horizontal in both curves ([Figure 2F](#)). This discrepancy between model and data can be reconciled if we add “tuned-excitatory (TE)”-type cells to the “near”- and “far”-type cells in our simulation and then passing the disparity-energy signals through an additional expansive nonlinearity (data not shown). The added expansive nonlinearity has no effect on the signal amplitudes carried by the “near”- and “far”-type cells while it reduces the signal amplitude of the “TE”-type cell only if the disparity energy is inverted. This modification only works if the disparity is decoded from the population activity. A single odd-symmetric cell does not attenuate the inverted signal even with the additional nonlinearity.

How the stereoscopic system attenuates the inverted signal in odd-symmetric cells is still an active topic of research. One study concludes that rectification prior to binocular summation explains the attenuation (Read, Parker, & Cumming, 2002), while another study proposes that a three-layered network model explains it (Lippert, Fleet, & Wagner, 2000). Physiological studies offer evidence that the attenuation is caused by integration of cells with different preferred disparities (Haefner & Cumming, 2008; Tanabe & Cumming, 2006) or integration of SF channels (Kumano, Tanabe, & Fujita, 2008).

## Cortical mechanisms

Many binocular properties of V1 neurons agree with the prediction of the disparity-energy model. However, quantitative examinations show systematic deviations, which explain the discrepancy between the model and our psychophysical data. Real V1 cells show weaker inverted responses to aRDS, as well as to interocular delay, than the model predicts (Cumming & Parker, 1997; Read & Cumming, 2005). If stereo discrimination uses the signals generated in V1, we expect degradation in performance. This is exactly what our psychophysical data show. It is unclear, though, how attenuated the

inverted responses of real V1 cells are to interocular delay because the range of interocular delay used by Read and Cumming (2005) might have been too narrow. Our simulation suggests that interocular delay of at least 100 ms is needed to distinguish the dip from the asymptote, while the maximum delay tested by Read and Cumming was only 42 ms.

The present results provide an account for a puzzling discrepancy in the physiology literature. Neurons in the middle temporal area (MT/V5) of the monkey cortex have been most strongly linked to stereoscopic depth judgments (DeAngelis, Cumming, & Newsome, 1998; Uka & DeAngelis, 2004) whereas they encode the disparity of contrast-reversed stereograms just as V1 neurons (Krug, Cumming, & Parker, 2004). This poses the question whether neurons can ever contribute to stereopsis without solving the stereo correspondence problem. Our results demonstrate psychophysically that perceptual-decision processes can indeed exploit the disparity-energy signals. Local disparity-energy signals conveyed from V1 to MT and then to the medial superior temporal area may suffice to discriminate coarse disparity tested in previous studies (DeAngelis et al., 1998; Uka & DeAngelis 2003, 2004) or to guide vergence eye movement (Takemura, Inoue, Kawano, Quaia, & Miles, 2001). On the other hand, perception of a cyclopean surface and discrimination of fine disparity may be mediated by ventral pathway areas where responses to contrast-reversed stereograms are eliminated (Janssen et al., 2003; Tanabe, Umeda, & Fujita, 2004; Uka, Tanabe, Watanabe, & Fujita, 2005; for reviews, see Neri, 2005; Parker, 2007).

## Conclusion

We conclude that there are at least two complementary subsystems underlying stereopsis: One mediates disparity-energy signals for immediate behavioral responses, and another evokes a percept of a cyclopean surface only in the presence of a global match. This study demonstrates that the mere ability for observers to discriminate stereoscopic depth is not a direct indicator of the subjective sensation of cyclopean structures. Application of our psychophysical assessment of shape discrimination based on stereo cues to physiological recordings may help distinguish the neural computation of cyclopean structures from disparity energies.

## Acknowledgments

We thank Bruce Cumming, Okihide Hikosaka, Ikuya Murakami, and Satoshi Shimegi for helpful comments on an earlier version of the manuscript. Supported by grants

from MEXT (13308046, 00190003), CREST of the Japan Science and Technology Agency, and the Takeda Science Foundation. Current address of ST is Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD, USA. Current address of SY is Graduate School of Medicine, Nagoya Univ, Furoh-cho, Chikusa-ku, Nagoya, Japan.

Commercial relationships: none.

Corresponding author: Ichiro Fujita.

Email: [fujita@fbs.osaka-u.ac.jp](mailto:fujita@fbs.osaka-u.ac.jp).

Address: Graduate School of Frontier Biosciences, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka, 560-8531, Japan.

## References

- Anzai, A., Ohzawa, I., & Freeman, R. D. (2001). Joint-encoding of motion and depth by visual cortical neurons: Neural basis of the Pulfrich effect. *Nature Neuroscience*, *4*, 513–518. [[PubMed](#)] [[Article](#)]
- Chen, Y., Wang, Y., & Qian, N. (2001). Modeling V1 disparity tuning to time-varying stimuli. *Journal of Neurophysiology*, *86*, 143–155. [[PubMed](#)] [[Article](#)]
- Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L., & Blake, R. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception*, *24*, 33–47. [[PubMed](#)]
- Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms: Effects of spatial density and interocular delay. *Vision Research*, *33*, 1959–1975. [[PubMed](#)]
- Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, *24*, 203–238. [[PubMed](#)]
- Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, *389*, 280–283. [[PubMed](#)]
- Cumming, B. G., Shapiro, S. E., & Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception*, *27*, 1367–1377. [[PubMed](#)]
- DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*, 677–680. [[PubMed](#)]
- De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D., & Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Research*, *40*, 3685–3702. [[PubMed](#)]
- Haefner, R. M., & Cumming, B. G. (2008). Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron*, *57*, 147–158. [[PubMed](#)]
- Hayashi, R., Miyawaki, Y., Maeda, T., & Tachi, S. (2003). Unconscious adaptation: A new illusion of depth induced by stimulus features without depth. *Vision Research*, *43*, 2773–2782. [[PubMed](#)]
- Helmholtz, H. L. (1909). *Treatise on physiological optics*. New York: Dover.
- Janssen, P., Vogels, R., Liu, Y., & Orban, G. A. (2003). At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, *37*, 693–701. [[PubMed](#)] [[Article](#)]
- Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, *39*, 1125–1162.
- Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.
- Krug, K., Cumming, B. G., & Parker, A. J. (2004). Comparing perceptual signals of single V5/MT neurons in two binocular depth tasks. *Journal of Neurophysiology*, *92*, 1586–1596. [[PubMed](#)] [[Article](#)]
- Kumano, H., Tanabe, S., & Fujita, I. (2008). Spatial frequency integration for binocular correspondence in macaque area V4. *Journal of Neurophysiology*, *99*, 402–408. [[PubMed](#)]
- Lippert, J., Fleet, D. J., & Wagner, H. (2000). Disparity tuning as simulated by a neural net. *Biological Cybernetics*, *83*, 61–72. [[PubMed](#)]
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B: Biological Sciences*, *204*, 301–328. [[PubMed](#)]
- Masson, G. S., Busetini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, *389*, 283–286. [[PubMed](#)]
- Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*, 1811–1825. [[PubMed](#)]
- Neri, P. (2005). A stereoscopic look at visual cortex. *Journal of Neurophysiology*, *93*, 1823–1826. [[PubMed](#)] [[Article](#)]
- Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*, 695–698. [[PubMed](#)] [[Article](#)]
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, *249*, 1037–1041. [[PubMed](#)]
- Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews, Neuroscience*, *8*, 379–391. [[PubMed](#)]

- Read, J. C., Parker, A. J., & Cumming, B. G. (2002). A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. *Visual Neuroscience*, *19*, 735–753. [[PubMed](#)]
- Read, J. C., & Cumming, B. G. (2005). Effect of interocular delay on disparity-selective v1 neurons: Relationship to stereoacuity and the pulfrich effect. *Journal of Neurophysiology*, *94*, 1541–1553. [[PubMed](#)] [[Article](#)]
- Read, J. C., & Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research*, *40*, 3345–3358. [[PubMed](#)]
- Rogers, B. J., & Anstis, S. M. (1975). Reversed depth from positive and negative stereograms. *Perception*, *4*, 193–201.
- Ross, J. (1974). Stereopsis by binocular delay. *Nature*, *248*, 363–364. [[PubMed](#)]
- Takemura, A., Inoue, Y., Kawano, K., Quaia, C., & Miles, F. A. (2001). Single-unit activity in cortical area MST associated with disparity-vergence eye movements: Evidence for population coding. *Journal of Neurophysiology*, *85*, 2245–2266. [[PubMed](#)] [[Article](#)]
- Tanabe, S., & Cumming, B. G. (2006). *Coding of position disparity with odd-symmetric disparity tuning in macaque V1 and V2* (Program No. 407.4). Washington, DC: Society for Neuroscience Abstract Viewer/Itinerary Planner.
- Tanabe, S., Umeda, K., & Fujita, I. (2004). Rejection of false matches for binocular correspondence in macaque visual cortical area V4. *Journal of Neuroscience*, *24*, 8170–8180. [[PubMed](#)] [[Article](#)]
- Tyler, C. W. (1974). Stereopsis in dynamic visual noise. *Nature*, *250*, 781–782. [[PubMed](#)]
- Tyler, C. W. (1990). A stereoscopic view of visual processing streams. *Vision Research*, *30*, 1877–1895. [[PubMed](#)]
- Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: Comparison of neuronal and psychophysical sensitivity. *Journal of Neuroscience*, *23*, 3515–3530. [[PubMed](#)] [[Article](#)]
- Uka, T., & DeAngelis, G. C. (2004). Contribution of area MT to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron*, *42*, 297–310. [[PubMed](#)] [[Article](#)]
- Uka, T., & DeAngelis, G. C. (2006). Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *Journal of Neuroscience*, *26*, 6791–6802. [[PubMed](#)] [[Article](#)]
- Uka, T., Tanabe, S., Watanabe, M., & Fujita, I. (2005). Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *Journal of Neuroscience*, *25*, 10796–10802. [[PubMed](#)] [[Article](#)]
- Ziegler, L. R., & Hess, R. F. (1999). Stereoscopic depth but not shape perception from second-order stimuli. *Vision Research*, *39*, 1491–1507. [[PubMed](#)]