
Stereo motion transparency processing implements an ecological smoothness constraint

Finnegan J Calabro, Lucia M Vaina[¶]

Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA; [¶] also Department of Neurology, Harvard Medical School, 75 Francis Street, Boston, MA 02215, USA; e-mail: fcalabro@bu.edu, vaina@bu.edu
Received 28 January 2005, in revised form 19 September 2005; published online 26 July 2006

Abstract. Transparent motion stimuli allow us to investigate how visual motion is processed in the presence of multiple sources of information. We used stereo random-dot kinematograms to determine how motion processing is affected by the difference in direction and depth of two overlapping motion components. Observers judged whether a noise dot display contained one or two directions of motion. For all disparity differences, performance did not change among angles greater than 60°, but the ability to detect transparent motion fell dramatically as the direction difference decreased below 60°. When a disparity difference was added between the two motion components, detection became easier. We compared these results to an ideal-observer model limited by stimulus uncertainty and low-level sources of internal noise. The resulting measure of efficiency—the ratio of human to model performance—reflects changes in how motion stimuli are being processed. A decrease of both the direction and disparity differences had the effect of decreasing efficiency. These results suggest that the mechanism processing transparent motion may implement a smoothness constraint that tends to combine similar motions into a single percept.

1 Introduction

The perception of motion transparency can arise from physically transparent objects, such as windows or shadows, or from non-contiguously occluded objects, such as leaves or fences. In these cases, transparency is not indicated by the light reaching the eye from any one location in visual space, but rather by the pattern of light intensities across a wide area (Metelli 1974). Motion transparency provides a case that is incompatible with single-velocity-field mechanisms (Braddick and Qian 2001). A common feature of natural transparent scenes is that the two overlapping objects do not exist at the same place in three-dimensional space, but are generally separated in depth. Thus, depth separation may provide an additional solution to interpreting moving transparent scenes; if the visual system is able to separate the motion components into different depth planes, it may be able to recognize that they arise from different objects.

A number of psychophysical researchers have investigated motion transparency (Clifford et al 2004), and have shown that human observers are able to see multiple transparent-motion components (Stoner et al 1990), even without a depth separation. However, detection thresholds for transparency are higher than thresholds for coherent motion (one direction) (Hibbard and Bradshaw 1999). This suggests that independent motion detectors are required to detect each component of the transparent motion. When a disparity difference is applied to the two directions of motion, this performance penalty is removed, and subjects are able to detect transparency as easily as coherent motion, perhaps through scene segmentation. This effect of disparity is observed with only a few minutes of arc of disparity, but also appears to reverse soon after, causing thresholds to rise again after about 4 min of arc of disparity. Similarly, adding a disparity difference between two perpendicularly moving gratings favors the perception of transparent over coherent motion (Farell and Li 2004). Another recent study showed that a unidirectional motion aftereffect is seen after adapting to transparent motions, and provides evidence for two temporal channels (Alais et al 2005).

The effect on motion transparency of angular difference between two moving patterns was addressed by Braddick et al (2002) and Edwards and Nishida (1999). Their experiments have shown motion repulsion for angles greater than about 20° , making the directions appear further apart than they are. Below this limit the two motion components are more often seen as a single coherent motion, rather than as separate transparent-motion components (Edwards and Nishida 1999), suggesting that inhibition among detectors tuned to similar directions of motion may cause worse performance in a transparency detection task when the angle difference is less than 90° .

Transparent stimuli are processed in local to global processing stages in the motion pathway. Results from experiments in which opposite directions of motion occur within a few minutes of arc of each other demonstrate cancellation effect which prevents perception of transparency (Qian et al 1994). Yet, when the local dots are unpaired, either by a vertical displacement or a separation in depth, the perception of transparency is restored. Physiological studies with paired dots showed that neurons in the middle temporal (MT) area responded significantly better to unpaired stimuli (Qian and Andersen 1994). Furthermore, Snowden et al reported that V1 neurons continue to fire normally when their preferred direction is embedded in a transparent display, while MT neurons are suppressed (Snowden et al 1991). Taken together, these results suggest that global motion processing provides a plausible neural basis for motion transparency.

In this study, we address the joint effects of angle and disparity, which are of particular interest in elucidating the role of constraints on motion processing. Furthermore, we provide a quantitative analysis to account for stimulus ambiguity and characterize the behavior of the underlying motion processing mechanisms. Several existing tests of transparency have required subjects to detect transparency relative to an interval of pure noise (Hibbard and Bradshaw 1999). While this experimental design allows a comparison of detection difficulty between transparent and coherent motion stimuli, it may not necessarily require the perception of motion transparency (the tasks may be solved by simply detecting one of the motion components). In the experiments presented here, we aim to characterize the conditions necessary for the perception of motion transparency and to quantitatively analyze performance as a function of the angle and disparity differences between the two motion components. Additionally, we argue that performance under various depth and angle conditions presents psychophysical evidence for the implementation of an ecological smoothness constraint in the motion processing system. Part of this work was presented at the Vision Science Society 2004 Annual Meeting (Calabro et al 2004).

2 General methods

2.1 Stimulus

Observers were presented with stereo random-dot kinematograms. Stimuli consisted of 141 dots split between two depth planes, presented in a 6 deg diameter circular aperture (5 dots deg^{-2}). At each depth, one direction was chosen a priori for all signal dots occurring there. All signal dots moved in their specified direction at 4 deg s^{-1} . The directions of motion for each plane were specified as one of two angles independently. They could have either the same (coherent) or different (transparent) directions, but knowledge of one direction of motion gave no information as to the transparency of the display.

A fixation cross was shown at the center of the aperture, at a depth of 0 min of arc disparity. One of the depth planes was always presented at this disparity, while the second plane occurred at some depth. The stimulus was presented for 12 frames lasting 481 ms. Dots were square with 2 pixels per side ($20.37 \text{ min of arc}^2$) and were presented with antialiased subpixel resolution of $\frac{1}{20}$ th of a pixel (allowing disparity and displacement increments of 0.1 min of arc). Dot luminance was 79.55 cd m^{-2} with a background luminance of 42.3 cd m^{-2} and fixation luminance of 84.08 cd m^{-2} .

2.2 Procedure

Observers were asked to make a two-temporal-alternative forced-choice judgment to discriminate between spatially superimposed dots moving in a single direction of motion and two sets of dots moving transparently. In one interval, the two sets of dots had the same direction of motion, while in the other interval they moved at some angle relative to each other (transparency). The observers were asked to identify which interval contained the transparent motion (figure 1). Both intervals contained the same number of dots at each depth, so the transparent interval could be distinguished only by a comparison of the directions of motion at each depth.

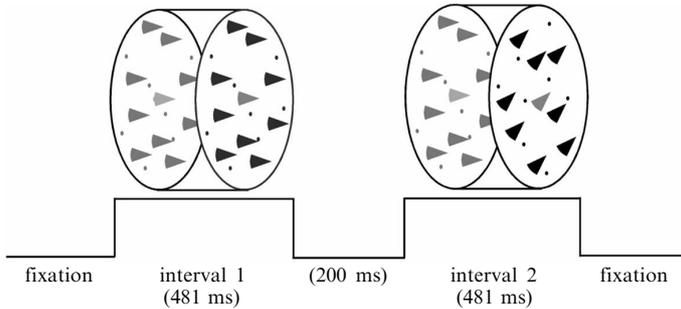


Figure 1. Paradigm for the stereomotion transparency task. Subjects were presented sequential stimuli both containing motion at two depth planes. The task was to identify the interval which contained motions in two different directions.

Data were collected with an adaptive staircase (Vaina 2003), by varying the proportion of signal in the display. In each trial, a proportion of the dots in each frame were randomly pre-selected as signal dots, and moved in the specified direction. The remaining noise dots were replaced randomly from one frame to the next (Newsome and Paré 1988). During a staircase, the coherence was varied to find the level at which the observer correctly identified the transparent stimulus 79% of the time. Step sizes were initially large in the adaptive phase of the staircase, to get a rough estimate of the threshold. After the first three reversals, the test switched to a classical three-up one-down staircase (Levitt 1970). In this phase, step sizes were equal within each decade. Thresholds were calculated on the basis of the final six reversals. Standard deviations were measured across the independent threshold estimates and reflected the variance of the threshold measurements.

2.3 Apparatus

Stimuli were generated by a Power Mac G4 and presented in 8-bit gray-scale mode on a calibrated Apple monitor. A pair of stereo glasses (Crystal Eyes 2, StereoGraphics Corporation, San Rafael, CA) was used to temporally interlace the left-eye and right-eye images, and was synchronized with LCD shutter glasses allowing each eye to see the specified image. Dots were displaced between the images for each eye to give the perception of depth.

2.4 Subjects

Eight observers were tested in the first experiment and three of them also took the second test. All had normal or corrected-to-normal vision, and were screened with a stereoacuity test, described below. One author was used as a subject; the other seven subjects were naive as to the purposes of the experiments. All participants gave informed written consent before the start of the experimental sessions, and the Boston University, Charles River Campus, approved all the procedures under the Protocol #989E.

2.5 Control task: Assessment of stereoacuity

Prior to the experimental sessions, we measured subjects' ability to perceive depth from disparity. Two dots were presented sequentially and the subjects were asked to respond which dot was nearer to them in depth (dots were presented randomly: both in front, both behind, or on either side of fixation in this case). Data were collected by a staircase procedure varying the depth difference between the two dots. The results were used as a screening test to show that all subjects were able to determine depth from disparity. The test also served as a quantitative measure of the subjects' stereoacuity, used in an ideal-observer model described below.

3 Experiment 1: Effects of angle and disparity

3.1 Stimuli

The stimuli were generated according to the description in section 2. During each staircase, angles and disparities were selected from a predefined set. The disparity and angle were changed between staircases in a predetermined, randomized sequence. A threshold estimate was obtained for each staircase, and data points show the mean of 4–8 staircases on a given condition. Coherence thresholds for detecting the interval containing transparent motion were measured for disparities of 0, 4, and 12 min of arc, and for angles of 15°, 30°, 45°, 60°, 90°, and 180° at each disparity. Outliers were identified and eliminated if greater than 5 median absolute deviations from the median value (Sheskin 2000).

3.2 Results

Performance in this task when there was no disparity varied among observers, but all showed the same general trends: thresholds were similar for angles of 60° and larger, then rose sharply to the minimum level tested (15°). Subjects were still able to perform the task at the smallest angular separations, but often required >70% coherence.

Figure 2a shows the results averaged across observers when the dot planes were separated by various depths. Disparities of 0, 4, and 12 min of arc were collected for all observers. Disparity caused a general decrease of thresholds at every angle tested. This is seen by a clear downward progression in thresholds from 0 to 4 to 12 min of arc seen at nearly all direction differences (excepting only the 4, 12 min of arc reversal at 90°). Figures 2b and 2c show the mean effects of angle and disparity on this task (threshold change for each observer relative to the 0 min of arc case, averaged across angles for the effect of disparity, and relative to 15° for the effect of angle). These data illustrate the overall effect of angle and disparity on the detection of motion transparency. Performance increases as both direction and disparity separations increase. The mean threshold change across all eight observers was a decrease of 3.0% coherence for 4 min of arc of disparity, and 5.4% coherence at 12 min of arc (figure 2c). This effect was much smaller than the effect of angle, which caused threshold changes of over 30% coherence (figure 2b).

We used a generalized linear model to determine whether there were significant effects of angle and disparity on performance. We included simple nonlinear (exponential) terms to model the effects of angle and disparity since we expected that these effects would saturate at some level. Mean thresholds from each observer for each disparity–angle combination were used in the model, and variances reflect differences among observers. Significant effects were found for disparity ($t = 2.62$, $p < 0.01$) and angle ($t = 6.77$, $p < 0.0001$), but there was no significant interaction ($t = 0.31$, $p > 0.75$) between them. This shows that increased disparity and direction differences cause improvements in performance for our subjects, and suggests a nonlinear effect of each factor, as expected (figure 2a).

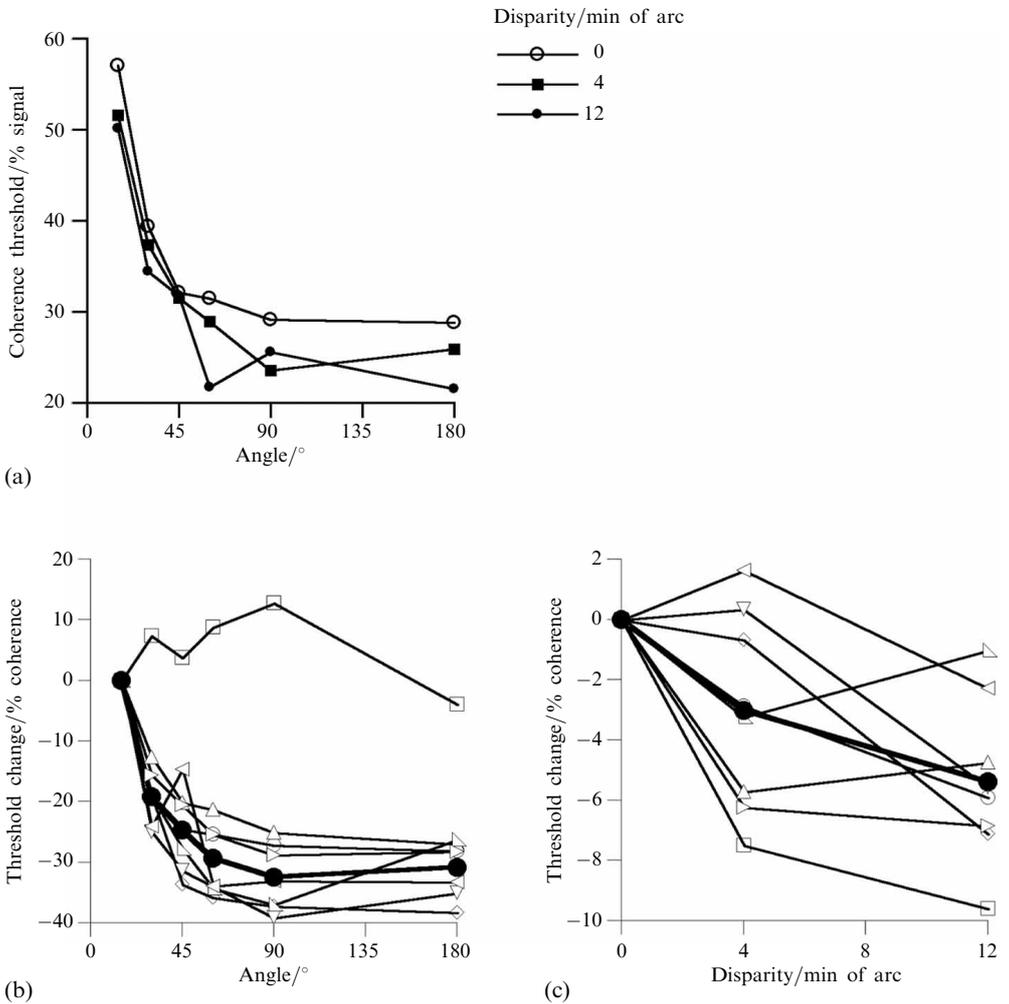


Figure 2. Effects of direction and disparity on transparency detection. (a) Mean results for all eight observers, shown for each disparity. (b) Effect of angle on thresholds, averaged across disparities. Values are the change in threshold from the 15° condition. (c) Effect of disparity averaged across angles. Values are the change in threshold from 0 min of arc disparity. In (b) and (c), open symbols show each observer, solid circles are the mean of all observers.

3.3 Discussion

The results shown in figure 2c are at odds with previous studies of the effect of disparity on motion transparency detection (Hibbard and Bradshaw 1999). Hibbard and Bradshaw showed that as the disparity separation increased from 0, thresholds decreased, but that this continued only for 2–4 min of arc. Beyond this value, thresholds rose back to the 0 min of arc separation levels. In contrast, our results show a monotonically decreasing trend in thresholds as disparity is increased, up to 12 min of arc. A further data point (16 min of arc) was collected on one subject (FC) and showed a continuation of the trend. One possible explanation might be that the higher density used (12.5 dots deg^{-2}) by Hibbard and Bradshaw might have led to stereo mismatches or stereo correspondence problems. This would occur when dots are near enough to each other so that the correct stereo pair becomes ambiguous. This mismatch may happen when stereo separation (the amount of displacement necessary for the specified disparity) exceeds half the distance to the nearest dot. In this case, the nearest dot will not

be the correct stereo match, but rather one of the dots from another stereo pair. Thus, to avoid stereo correspondence problems, the stereo pair separation (δ_{cm}) must be less than half the average dot spacing (d). The stereo pair separation is a function of the disparity used and viewing distance, given as

$$\delta_{\text{cm}} = v \frac{\delta_{\text{min}}}{60} \frac{\pi}{180}$$

where δ_{min} is the disparity (in min of arc) and v is the viewing distance in cm. For a dot density N (N dots deg^{-2}) and a uniform distribution, the dot spacing, d , is given by

$$d_{\text{deg}} = \frac{\sqrt{1}}{\sqrt{N}} = \frac{1}{\sqrt{N}},$$

$$d_{\text{cm}} = \frac{1}{\sqrt{N} \tan^{-1}(1/v)}.$$

Thus, we can find the densities and disparities at which stereo correspondence begins to become a problem:

$$\text{dist}(a_{\text{R}}, b_{\text{L}}) < \text{dist}(a_{\text{R}}, a_{\text{L}}),$$

$$d - \delta < \delta$$

$$\frac{1}{2}d < \delta$$

$$\frac{1}{\sqrt{N} \tan^{-1}(1/v)} < v \frac{\delta_{\text{min}}}{60}$$

$$N > \left[\frac{1}{2} \frac{1}{\tan^{-1}(1/v)} \frac{60}{\delta_{\text{min}}} (1/v) \right]^2 \approx \left[\frac{30}{\delta_{\text{min}}} \right]^2.$$

For example, for 16 min of arc of disparity and a viewing distance of 120 cm, stereo correspondence should become a problem for densities greater than 3.5 dots deg^{-2} . Reformulating to solve for δ , we find that for a density of 12.5 dots deg^{-2} , stereo mismatches begin to occur at a disparity of 8.5 min of arc. This is higher than the 2–4 min of arc Hibbard and Bradshaw found as their peak disparities, and thus cannot explain the threshold reversal they reported.

Another possible explanation for the discrepancy between our results and those of Hibbard and Bradshaw is that a time-dependent process (such as a time requirement to shift attention in depth) prevents the observers from fully processing both planes (their display time—150 ms—was significantly shorter than ours, 481 ms). To test this possibility, we reduced our stimulus presentation to 3 frames (120 ms) and collected data for motions 180° apart for various disparities. The results are shown in figure 3. The same trend we observed in the long-duration displays persists for the short-duration display, suggesting that the temporal properties of the stimuli do not explain the discrepancy in results.

A third possible explanation for the difference between the results of the two groups could be the nature of the task. In Hibbard and Bradshaw's task, subjects were asked to judge which interval contained transparency compared to an interval of pure noise. We suggest that, in their task, observers do not have to actually detect transparency, but that it is sufficient that they segregate the two planes and perform a motion-detection task on one of them. Thus, the U-shaped function (Hibbard and Bradshaw 1999, their figure 3) may be a combination of two factors: a decrease in thresholds over the first few minutes as the segregation of the planes becomes easier, and a slow, steady rise in thresholds as the motion-detection task gets harder (both planes

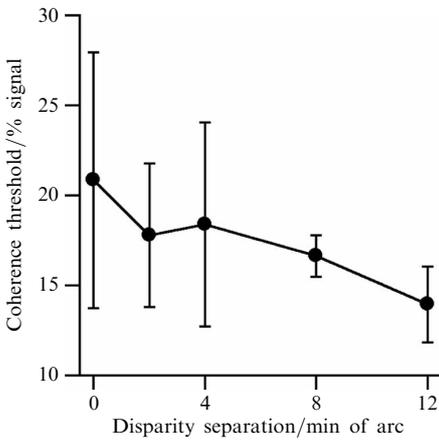


Figure 3. Threshold as a function of disparity for short (120 ms) stimulus presentations. To determine whether long display times enabled subjects to process both planes when separated by large disparities, we repeated the experiment on subject FC for opposing motions and various depth separations. Although thresholds are higher than before, the steady decline in thresholds as disparity is added persists.

were being moved away from the fixation depth, so the motion detection must be done at larger disparities, which may be underrepresented or underused). Data presented at the Vision Sciences Society Annual Meeting (2006) by our laboratory suggest that the segregation occurs at 2–6 min of arc of disparity separation (varying by observer), which would agree with the location of the peak of the U-shape function in the Hibbard and Bradshaw experiment. On this basis we suggest that the task used by Hibbard and Bradshaw does not necessarily require the detection of multiple directions of motion, and the difference between our results and theirs stems from the nature of the tasks.

4 Ideal-observer model

4.1 Methods

Ideal-observer models have been productive in motion (Crowell and Banks 1996; Watamaniuk 1993) and motion-transparency (Wallace and Mamassian 2003) experiments in determining whether psychophysical results can be explained simply by stimulus properties. Generally, ideal-observer models use a Bayesian statistics approach to solve the psychophysical task and use exact knowledge of the stimulus, thus setting the upper limit of performance, and providing a reference for comparison with human results.

We developed an ideal-observer model to determine whether performance changes reported in experiment 1 could be explained by changes in the information available in the stimulus and noise in early visual processing stages. Our model is limited by subjects' error in determining the direction and disparity of a single dot in control experiments, similar to an ideal-observer model proposed for solving heading direction from optic flow (Crowell and Banks 1996). It encapsulates sources of local noise, such as that arising from early visual processing areas, including the retina and area V1. In experiment 1, direction and disparity are the only relevant variables for solving the task; therefore, we can assume that error in judging any other characteristics of the stimulus (eg speed, luminance, or size) will not affect performance. We chose as input to the model a noisy-motion vector field, with each dot perturbed in direction and depth by Gaussian noise. Error ranges reported by Crowell and Banks for direction discrimination are used here, and a similar task (which was used as the stereoacuity screening test) was implemented to determine error in representing the depth of a single dot. Error values were found by independently estimating the direction and disparity which would give $d' = 1$, in which case the internal representation of the parameter is equal to the variance of the representation.

The ideal observer has knowledge of the potential directions and disparities used in a given trial, and compares the input to four templates (either of two motions presented at each disparity). A probability is calculated for each dot as having arisen from

each of the four templates. Probabilities are combined by Bayes's rule (see appendix: Ideal-observer formulation) to give the probability that the stimulus presented was generated by each of the underlying distributions, and from there an overall probability of transparency can be calculated.

By comparing the probability of transparency for the two intervals, the model is able to determine which interval is more likely to contain transparency. As in the case of the human observers in experiment 1, here too noise was added to the stimulus (again restricted to the two available disparities), and coherence thresholds were measured for various angle and depth parameters. We used Monte Carlo simulations of the model to determine the 79%-correct level for comparison to psychophysical results.

4.2 Efficiency

We compared the model with the data from the three observers with an efficiency measure (Crowell and Banks 1996) defined as the ratio of thresholds (or, equivalently, as human sensitivity over ideal-observer sensitivity). An efficiency of 1.0 indicates identical performance, while smaller efficiencies mean that the human observer is unable to perform as well as the ideal observer. Since the model is subject to the same sources of local noise as the human observers, changes in efficiency reflect changes in higher-level areas, possibly including area MT. We therefore suggest that efficiency is a metric of performance independent of the information available to the mechanism, and that changes in efficiency should reflect changes in the processing mechanism itself.

4.3 Results

Overall, the model performs much better than the human observers did in experiment 1, by a factor of roughly 10, which is within the range of other ideal-observer models (Crowell and Banks 1996). Figure 4 shows the changes in efficiency caused by direction differences (figure 4a, relative to the 15° condition) and disparity differences (figure 4b, relative to the 0 min of arc condition). In both cases, efficiencies increase as the separations increase, before saturating among the highest levels of each variable.

We again modeled both angle and disparity as exponential terms in our statistical model (efficiencies are modeled as the ratio of two exponential functions, which produces another exponential). Both angle ($t = -4.10$, $p < 0.0001$) and disparity ($t = -1.96$, $p = 0.05$) were significant factors when controlling for differences in baseline performance

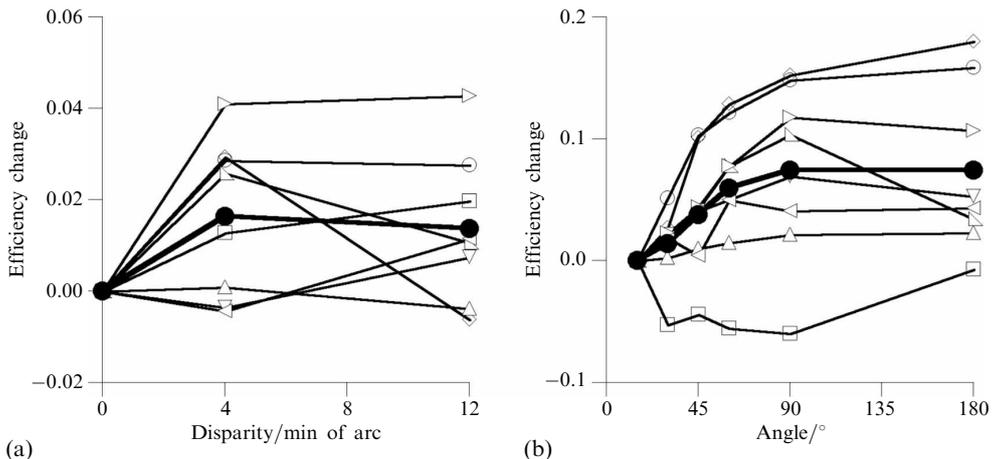


Figure 4. Effects of angle and disparity on efficiencies. (a) Effect of angle on efficiencies, averaged across disparities. Values are the change in efficiency from the 15° condition. (b) Effect of disparity on efficiencies, averaged across angle, relative to the 0 min of arc disparity case. Open symbols show each observer, solid circles are the mean of all observers.

by subject, but there was no interaction between the two ($t = 0.56$, $p > 0.5$). The results show a large drop in efficiency for angles below 60° (figure 4a), despite not changing among the larger angles tested (linear model for angles 60° and larger gives $p = 0.88$, $t = -0.14$ for the effect of angle). For these small angles, human performance begins to suffer, while the model has not yet been noticeably affected. Thus, efficiencies drop as the angle decreases, and this continues even when the model performance is no longer constant (below 30°). This shows that the angle separating the motions affects observers on a motion-transparency task at much larger angles than predicted by a model limited by observers' resolution in making motion-direction judgments. The effect of disparity shows a similar trend, with efficiencies decreasing when the depth separation is removed. There is little to no effect of disparity on efficiency between the 4 and 12 min of arc conditions.

4.4 Discussion

The results reported here suggest that as both the disparity and direction differences between two components of motion decrease, not only do subjects' performances suffer, but efficiencies drop as well. Since our ideal-observer model is limited by the direction and disparity resolution of the observers who performed the task, the efficiency results demonstrate that performance changes are not fully governed by the ability of these subjects to determine the depth and direction of the individual dots in the stimulus. The results demonstrate that variations in the stimulus uncertainty as limited by local noise cannot completely account for the performance changes reported in experiment 1. This suggests the existence of an additional effect on the mechanism, either as an external effect or as a consequence of the properties of the processing mechanism.

A direct comparison of the ideal-observer model as applied to experiment 1 might be flawed if human observers were not using all of the stimulus area available to them. Even if they were, it is known that foveal resolution is much better than in the periphery, causing some parts of the visual field to have a disproportionate representation. The ideal-observer uses the entire stimulus available and makes no distinction based on visual field location. Furthermore, the ideal-observer performance is relatively constant until very small angles, and coherence thresholds are very low since it is able to utilize all information maximally. We extended the ideal observer by varying the available stimulus size in order to determine whether trends in ideal-observer performance and human efficiency depend on the summation area. As the stimulus area is decreased, down to a radius of 0.5 deg (area of 0.78 deg²), the baseline performance of the ideal observer becomes more similar to that of the human observers. However, the drops in efficiency for smaller direction and disparity differences are maintained for all summation areas tested.

5 Experiment 2: Is the perception of depth necessary for depth segmentation?

5.1 Stimuli

Using the test methodology described in section 2, we used anticorrelated dots to determine whether the perception of depth was a necessary condition for performance improvements in the transparency task. This method alters a corresponding pair of dots to have opposite contrasts relative to the background. Each opposite-contrast dot is presented to one eye. Although early visual areas respond to this phenomenon in a similar way as they do to veridical disparity (Cumming and Parker 1997), the display is not perceived to have any depth. This allows the investigation of the requirements and processing stages of stereo transparency.

Anticorrelated dot luminances were calculated by computing Michelson contrasts of equal magnitude but opposite sign, and presenting each contrast to one eye. This gave dots of luminances 79.55 cd m⁻² and 22.5 cd m⁻² on a background of 42.3 cd m⁻².

Each eye randomly saw either the positive or negative contrast dot, so each eye saw an approximately equal number of each contrast level. Other than changing the contrast of one of each pair of dots, the stimuli from experiment 1 were used.

5.2 Results

Three disparities (0, 4, 12 min of arc) were tested at five angles (15°, 30°, 45°, 90°, 180°) for observer FC and two angles (45°, 180°) for observers EC and AG. An N -way ANOVA test showed that the effects of angle, as seen in experiment 1, persist ($p < 0.0001$), but there is no longer any change in performance caused by disparity ($p = 0.848$). Further, there is no effect of disparity or angle for large angles (90° and 180°, $p = 0.09$ for disparity, $p = 0.35$ for angle). For angles of 45° and larger, the effect of angle is nearly significant ($p = 0.059$), although that of disparity is not ($p = 0.368$). This cutoff of significance agrees with the results of experiment 1, suggesting that anticorrelation did not affect the role that angular separation plays in motion-transparency perception, but did remove all effects of disparity. The results for one observer (FC) are shown in figure 5.

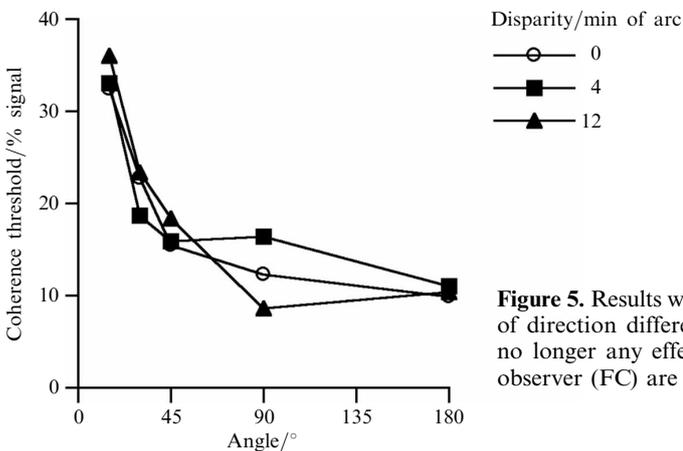


Figure 5. Results with anticorrelated dots. The effect of direction difference is unchanged, but there is no longer any effect of disparity. Data from one observer (FC) are plotted.

5.3 Discussion

The result of this experiment supports the hypothesis that the perception of disparity is a necessary condition for the improvement in performance created by a depth separation between the transparent planes. Neurophysiological experiments with anticorrelated dots suggest that early stages of the visual processing pathway (ie V1) pair these dots, but that no depth is perceived. This suggestion, along with the results of this experiment, supports the hypothesis that segregation between transparent motions is processed further along the motion processing pathway, perhaps in area MT as has been suggested (Snowden et al 1991).

6 General discussion

In this study we have shown how psychophysical performance in a motion-transparency detection task varies as a function of the depth and angular separations between the motion components. Experiment 1, figure 2 demonstrates that motion transparency becomes easier to detect as the angle and disparity differences between the motion components increase. To explain these results we proposed an ideal-observer model which suggests that the experimental outcomes reflect true changes in the underlying motion-processing mechanisms. Not only do subjects' performances drop for small disparity and direction differences, but stimulus ambiguity and low-level sources of noise cannot fully account for the experimental results, which indicates an additional negative effect on the processing of motion transparency.

Perception of depth is required to gain any performance benefit from increased disparity in detecting motion transparency, as indicated by experiment 2, figure 5. This suggests that depth segmentation for this task occurs beyond areas that respond to anticorrelated disparities, such as V1 (Cumming and Parker 1997). This is consistent with previous neurophysiological studies which indicate that neurons in area MT underlie the perception of motion transparency (Qian and Andersen 1994; for a review, see Braddick and Qian 2001).

Taken together, the psychophysical and ideal-observer model results presented here indicate that the mechanism processing motion transparency attempts to implement a smoothness constraint (Hildreth 1983; Marr 1982) by grouping across similar features (eg depth and direction). When local motions are observed at very similar depths, or moving in very similar directions, it is more likely that they actually arise from the same object, and external noise or measurement error caused slight variations in the perceived features. It would be advantageous for the visual system to exploit this constraint in order to form the most probable scene interpretation from the available visual information. The results presented here provide psychophysical evidence that the visual motion system obeys such a smoothness constraint by combining similar features into a single percept.

Several physiological implementations are plausible for such a constraint. One possibility would be the use of detectors with large tuning widths that are unable to make fine discriminations, as tuning widths for direction (Gattass and Gross 1981) and disparity (Maunsell and Van Essen 1983) are large relative to the ranges tested in our experiments. However, population responses are known to be capable of discriminating on a much finer scale than the individual detectors by looking for peaks or by the overall shape of the activity profile (Treue et al 2000). Furthermore, psychophysical discrimination limits reported for direction (Ball and Sekuler 1979) and disparity (see section 2.5) are well below the level at which effects are observed in our experiments. Edwards and Nishida (1999) suggest another potential implementation, where inhibition among similarly tuned detectors may favor the perception of a single direction of motion and disparity.

About 70% of MT neurons are tuned to disparity in addition to motion (Gonzalez and Perez 1998), about 84% are strongly directionally selective (Maunsell and Van Essen 1983; Snowden et al 1992), and MT activity is strongly correlated with psychophysical performance on motion tasks (Newsome et al 1986). This is consistent with the likelihood that the nervous system would find it ecologically advantageous to implement a smoothness constraint at the level of area MT. MT is the first level in the visual motion hierarchy where individual neural responses are consistent with the perception of motion, and hence should accurately reflect real-world visual events.

Acknowledgments. Many thanks to Scott Beardsley for his ideas and to Greg DeAngelis for suggestions regarding the anticorrelated dot experiment. Stimulus code was contributed by Martin Kopcik and Mike Lekyr. Thanks to all our subjects. This work was funded by NIH grant EY-2RO1-07861 to LMV.

References

- Alais D, Verstraten F, Burr D, 2005 "The motion aftereffect of transparent motion: Two temporal channels account for perceived direction" *Vision Research* **45** 403–412
- Ball K, Sekuler R, 1979 "Masking of motion by broadband and filtered directional noise" *Perception & Psychophysics* **26** 206–214
- Braddick O, Qian N, 2001 "The organization of global motion and transparency", in *Motion Vision—Computational, Neural, and Ecological Constraints* Eds J M Zanker, J Zeil (Berlin: Springer) pp 85–112
- Braddick O J, Wishart K A, Curran W, 2002 "Directional performance in motion transparency" *Vision Research* **42** 1237–1248

- Calabro F J, Beardsley S A, Vaina L M, 2004 "Effects of disparity and noise on motion transparency" *Journal of Vision* **4** 857a (abstract)
- Clifford C W, Spehar B, Pearson J, 2004 "Motion transparency promotes synchronous perceptual binding" *Vision Research* **44** 3073–3080
- Crowell J A, Banks M S, 1996 "Ideal observer for heading judgments" *Vision Research* **36** 471–490
- Cumming B, Parker A, 1997 "Responses of primary visual cortical neurons to binocular disparity without depth perception" *Nature* **389** 280–283
- Edwards M, Nishida S, 1999 "Global-motion detection with transparent-motion signals" *Vision Research* **39** 2239–2249
- Farell B, Li S, 2004 "Seeing depth coherence and transparency" *Journal of Vision* **4** 209–223
- Gattass R, Gross C G, 1981 "Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque" *Journal of Neurophysiology* **46** 621–638
- Gonzalez F, Perez R, 1998 "Neural mechanisms underlying stereoscopic vision" *Progress in Neurobiology* **55** 191–224
- Hibbard P B, Bradshaw M F, 1999 "Does binocular disparity facilitate the detection of transparent motion?" *Perception* **28** 183–191
- Hildreth E C, 1983 *The Measurement of Visual Motion* (Cambridge, MA: MIT Press)
- Levitt H, 1970 "Transformed up-down methods in psychoacoustics" *Journal of Acoustical Society of America* **49** 467–477
- Marr D, 1982 *Vision* (New York: W H Freeman)
- Maunsell J H, Van Essen D C, 1983 "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation" *Journal of Neurophysiology* **49** 1127–1147
- Metelli F, 1974 "The perception of transparency" *Scientific American* **230**(4) 90–98
- Newsome W T, Mikami A, Wurtz R H, 1986 "Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion" *Journal of Neurophysiology* **55** 1340–1351
- Newsome W T, Paré E B, 1988 "A selective impairment of motion perception following lesions of the middle temporal visual area (MT)" *Journal of Neuroscience* **8** 2201–2211
- Qian N, Andersen R A, 1994 "Transparent motion perception as detection of unbalanced motion signals. II. Physiology" *Journal of Neuroscience* **14** 7367–7380
- Qian N, Andersen R A, Adelson E H, 1994 "Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics" *Journal of Neuroscience* **14** 7357–7366
- Sheskin D J, 2000 *Handbook of Parametric and Nonparametric Statistical Procedures* (New York: Chapman and Hall/CRC)
- Snowden R J, Treue S, Andersen R A, 1992 "The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns" *Experimental Brain Research* **88** 389–400
- Snowden R J, Treue S, Erickson R G, Andersen R A, 1991 "The response of area MT and V1 neurons to transparent motion" *Journal of Neuroscience* **11** 2768–2785
- Stoner G R, Albright T D, Ramachandran V S, 1990 "Transparency and coherence in human motion perception" *Nature* **344** 153–155
- Treue S, Hol K, Rauber H-J, 2000 "Seeing multiple directions of motion—physiology and psychophysics" *Nature Neuroscience* **3** 270–276
- Vaina L M, Grzywacz N M, Saiviroonporn P, LeMay M, Bienfang D C, Cowey A, 2003 "Can spatial and temporal motion integration compensate for deficits in local motion mechanisms?" *Neuropsychologia* **41** 1817–1836
- Wallace J M, Mamassian P, 2003 "The efficiency of speed discrimination for coherent and transparent motion" *Vision Research* **43** 2795–2810
- Watamaniuk S N J, 1993 "Ideal observer for discrimination of the global direction of dynamic random-dot stimuli" *Journal of the Optical Society of America A* **10** 16–28

Appendix: Ideal-observer formulation

The ideal-observer model receives a noisy vector field, shown in figure A1, as an input and must integrate the local motions (where each moving dot comprises one local motion) and make a decision regarding the global transparency of the stimulus. There are four possible templates of motions. These are defined by direction and disparity, and consist of two Gaussian peaks (the two motions at each of two disparities, either in the same or different directions) and a Gaussian ridge centered at each of the two disparities but flat for all directions to represent the probability of a noise dot. These probability density functions are illustrated in figure A2.

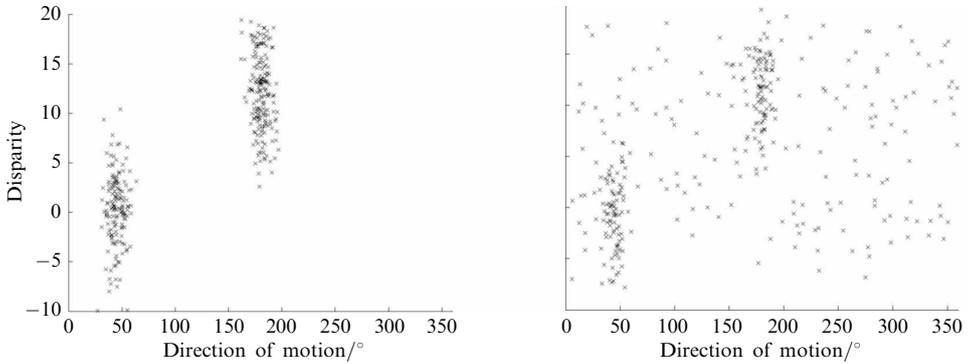


Figure A1. Example inputs to the ideal-observer model. The signal dots originate from two centers (at either the same or different angle, for two depths), but are perturbed by noise in both of these dimensions. Noise dots occur at a random direction at one of the two specified depths. Example stimuli without (left) and with (right) external noise are shown.

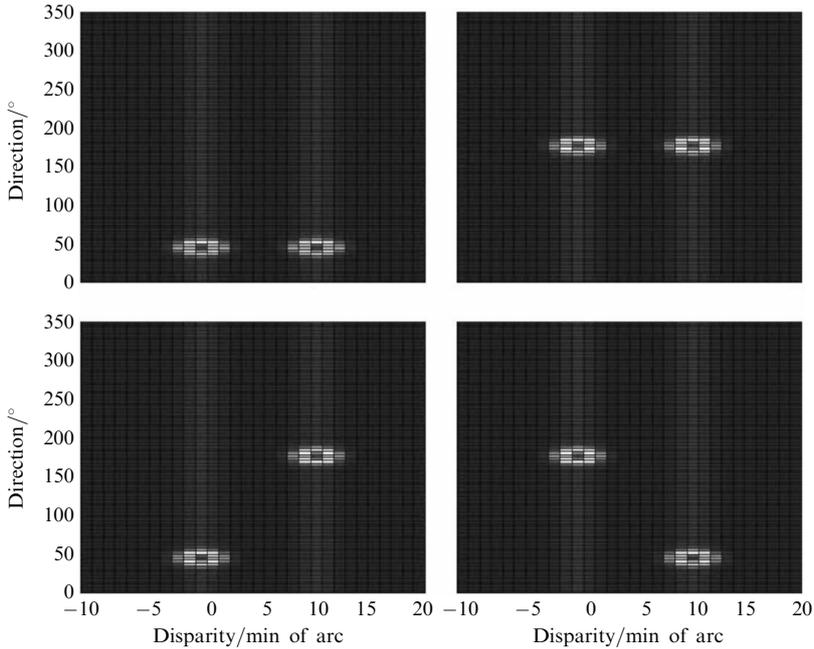


Figure A2. Probability density functions for ideal-observer model. The model uses these as the four possible templates that could have generated the input display. By calculating the probability that each of these gave rise to the display, the model determines whether the stimulus was more likely to have contained transparent motion.

Each dot has a probability for having come from each distribution ($\Theta_1, \dots, \Theta_4$), or for the noise distribution. The signal and noise distributions are given by:

$$p_s(x_n | \Theta_i) = \frac{1}{2\pi\sigma_d\sigma_z} \exp \left\{ -\frac{1}{2} \left[\frac{(z - \mu_{za})^2}{\sigma_z^2} + \frac{(d - \mu_{da})^2}{\sigma_d^2} \right] \right\} \\ + \exp \left\{ -\frac{1}{2} \left[\frac{(z - \mu_{zb})^2}{\sigma_z^2} + \frac{(d - \mu_{db})^2}{\sigma_d^2} \right] \right\},$$

$$p_n(x_n | \text{noise}) = \frac{1}{360\sqrt{(2\pi)\sigma_z}} \exp \left\{ -\frac{1}{2} \left[\frac{(z - z_1)^2}{\sigma_z} + \frac{(z - z_2)^2}{\sigma_z} \right] \right\},$$

where $\mu_{d,n}$ and $\mu_{z,n}$ are the centers of Gaussians for direction (d) and disparity (z) for distributions n . Each distribution Θ_i is given by the sum of Gaussians for the two appropriate ns . The total PDF for a given dot, then, is given by:

$$p(x_n|\Theta_i) = cp_s(x_n|\Theta_i) + (1-c)p_n(x_n|\text{noise}),$$

where c is the coherence. From these four probabilities for each dot, we find the probability of the entire dot distribution, X , having arisen from each of the four distributions:

$$P(X|\Theta_i) = \prod_{n=1}^N p(x_n|\Theta_i).$$

Using Bayes's rules, we may reformulate this as the probability of each of the four distributions being the underlying distribution, given the exact display observed:

$$P(\Theta_i|X) = \frac{P(X|\Theta_i)P(\Theta_i)}{P(X)},$$

where $P(\Theta_i)$ is the prior for each of the four distributions [since all four are equally likely, $P(\Theta_i) = 0.25$ for all i] and $P(X)$ is the probability of observing the exact dot distribution. This is derived as follows:

$$\begin{aligned} P(X) &= \int P(\Theta') \prod_{n=1}^N p(x_n|\Theta') d\Theta' = \frac{1}{4} \int \prod_{n=1}^N p(x_n|\Theta') d\Theta' \\ &= \frac{1}{4} \sum_{i=1}^4 \prod_{n=1}^N p(x_n|\Theta_i) = \frac{1}{4} \sum_{i=1}^4 P(X|\Theta_i). \end{aligned}$$

Thus,

$$P(\Theta_i|X) = \frac{P(X|\Theta_i)P(\Theta_i)}{P(X)},$$

$$P(\Theta_i|X) = \frac{P(X|\Theta_i)}{\sum_{i=1}^4 P(X|\Theta_i)}.$$

If we revisit the four distributions, the third and fourth represent the stimuli containing transparency motion, so the probability for transparency in the given interval, i , is:

$$P_t(i) = \frac{P(\Theta_3|X) + P(\Theta_4|X)}{\sum_i P(\Theta_i|X)}.$$

Finally, from this we can compute a likelihood ratio between the probability of the transparency occurring in the first and second interval:

$$L = \frac{P_t(1)}{P_t(2)}.$$

For $L > 1$, we select the first interval as that which contains transparency. For $L < 1$, we choose the second interval.

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 35 2006

www.perceptionweb.com

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.