Spatiotemporal relationships in a dynamic scene: stereomotion induction and suppression

Lora T. Likova The Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA

Christopher W. Tyler The Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA

We establish the existence of purely stereoscopic motion induction, i.e., perceived depth motion induced into a fixed-disparity target by disparity changes in a surround region. The stimuli were dynamic autostereograms consisting of a target and a surround, both consisting of horizontal lines of discs. We explored the stereomotion induction process by (i) direct estimation of the perceived distance moved, (ii) a cancellation technique with compensatory target motion, and (iii) extension of the compensatory motion into the zone beyond the null point. Adding compensatory stereomotion to the target reduced the induced motion experience to a null point. Beyond the cancellation point, two surprising results were obtained; perceived motion in the target increased, while the surround stereomotion perception was almost suppressed over a wide range of disparity changes (reciprocal stereomotion suppression). A model of the target/surround interactions was developed in the context of dynamic organization principles operating in stereomotion perception and misperception.

Keywords: stereomotion, induced motion, 3D induction, dynamic perceptual organization, relative motion, frame of reference, rigidity

Introduction

A basic problem for the visual system is interpreting the spatiotemporal events in the three-dimensional (3D) world constructed from two-dimensional (2D)-retinal images. Differences between the locations of matching features on the retina are termed binocular disparities, and the ability to perceive depth from these disparities is stereopsis. Binocular disparity is one of the most powerful sources of 3D information, providing the signal for perceiving stereoscopic depth. It is well established that discrete changes in disparity can elicit a continuous sense of stereoscopic motion in depth (Corbin, 1942; Attneave & Block, 1973).

Although the static aspects of stereovision have been intensively studied, its dynamic aspects have received less attention, despite their high ecological importance. However, the world is a complex net of dynamic long-range relationships. One might expect these influences to be reflected evolutionarily in the way the brain interprets the global visual-input information. In other words, the perception of every object or event is likely to be influenced by its spatial and temporal context. Here we approach the complexity of the real world by focusing on the interaction of stereomotion in spatiotemporal relationships with its global surroundings.

Many motion phenomena have been found and explored only for movements in the frontoparallel plane. The phenomenon of induced motion was first studied by Duncker (1929/1937), who found that a stationary dot surrounded by a moving background will appear to be moving. He inferred that induced motion perception was dependent on the global relationship among points, rather than the absolute velocities of isolated points. Later, Nakayama & Tyler (1978) used paired opposite motions to establish that lateral induced motion exists as a retinal phenomenon in the absence of eye movements (which could have contaminated Duncker's paradigm). Farné (1972) described induced motion in the third dimension, but it was based only on monocular cues such as size change, so it was not induced stereomotion in the sense of induction by disparity change.

Gogel & Griffin (1982) found that induced motion is not limited to continuous motion presented on a frontoparallel plane. They studied continuously and apparently moving inducers that generated the perception of both lateral and depth motion into a dot undergoing vertical apparent motion. An alternative interpretation of that phenomenon in terms of an apparent vergence for the two images of the test point was considered to be unlikely. However, it is unclear whether the induced motion they observed was purely stereoscopic or could be explained by lateral induction in opposite directions in the two eyes.

We first asked whether the stereomotion of a surround can induce depth motion into a static target. Dynamic stereoscopic displays consisting of a target and a surround undergoing binocular disparity shifts were used to establish and quantify the extent of perceived motion induction in the stereodomain. A stereomotion cancellation experiment further validated the perceptual reports and investigated its properties under a range of

DOI 10:1167/3.4.5 Received August 2, 2002; published June 10, 2003 ISSN 1534-7362 © 2003 ARVO

spatiotemporal configurations. The results show that induced stereomotion has a wealth of curious properties that are difficult to predict from the simple ecological hypothesis of the lateral motion case.

Methods

Stimuli

The stereoscopic stimuli were generated in the form of repetitive autostereograms (Figure 1) consisting of five horizontal rows of disks with the central line of disks specified as the target (Minev & Likova, 1999). Vertical distance between the rows was 65 arcmin. The diameter of each disk was 26 arcmin, and their luminance was 52 cd/m^2 against a background of 0.31 cd/m^2. The vertical as well as the horizontal extend of the whole image in all of the experiments did not change (the rows extended across the 29° width of the monitor screen).

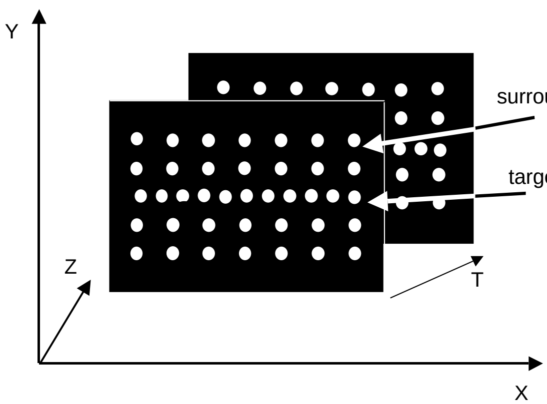

Figure 1. Schematic diagram of the stimuli (not to scale), consisting of a target and a surround whose depth was defined solely by binocular disparity. The surround consisted of four rows of disks switching back and forth between two disparity planes every 600 ms to produce surround stereomotion. The target was a similar row of disks, disparate from the surround. The disparity of the target was either constant (Experiments 1 and 2) or alternated to produce disparity-defined stereomotion (Experiments 3 and 4).

The stimuli were typically presented in two frames (Figure 1), where disparity in the surround disks was changed in order to provide them with alternating stereomotion. The target was either presented in one unchanging disparity plane, or given a disparity alternation that was varied over a wide range in some experiments. The principal frame duration used was 600 ms for each frame (0.83 Hz). However, the phenomena described are not restricted to this frame duration or to the particular sizes and disparities used in the measurements presented here.

The concept of the disparity images generated in space by an autostereogram is depicted in Figure 2 (Tyler & Clarke, 1990). The physical autostereogram plane is located at the solid line. To view an autostereogram, the eyes do not converge at the plane of the screen image, but at some other point in space indicated by the intersection of the lines of sight at another location in space. The dotted-line shapes represent the disparity images provided by this particular example of repeat periods. (For this application, the pattern was always uniform in the horizontal direction, with the repeat period varying only in the vertical direction.) These disparity images project into the eyes and up to the visual cortex to generate the corresponding depth percept.

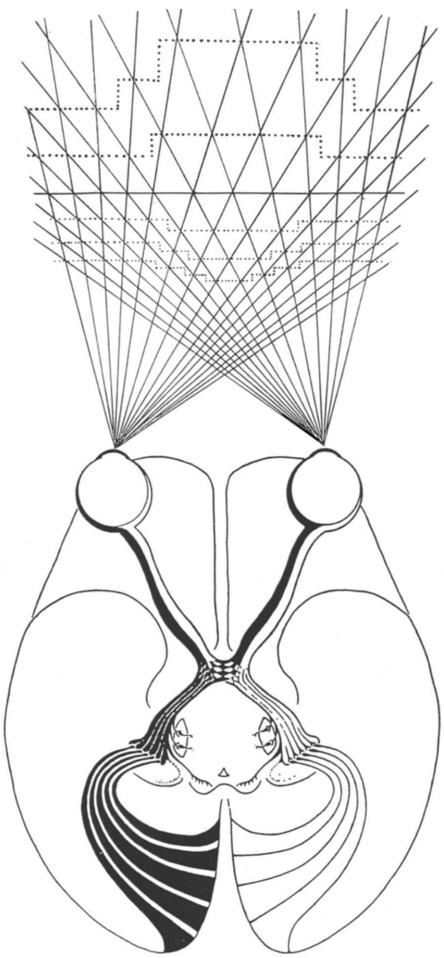

Figure 2. Depiction of the disparity images generated in space by an autostereogram. The autostereogram plane is shown by the thick line. The intersecting lines of sight along this thick line represent the repetition period of the autostereogram texture. The lines of sight further intersect at multiple locations in space. The dotted-line shapes represent the disparity images provided by this particular pattern of repeat periods. These disparity images project into the eyes and up to the visual cortex to generate the corresponding depth percept. The convergence of the eyes along a particular pair of lines of sight determines which of the disparity patterns will predominate.

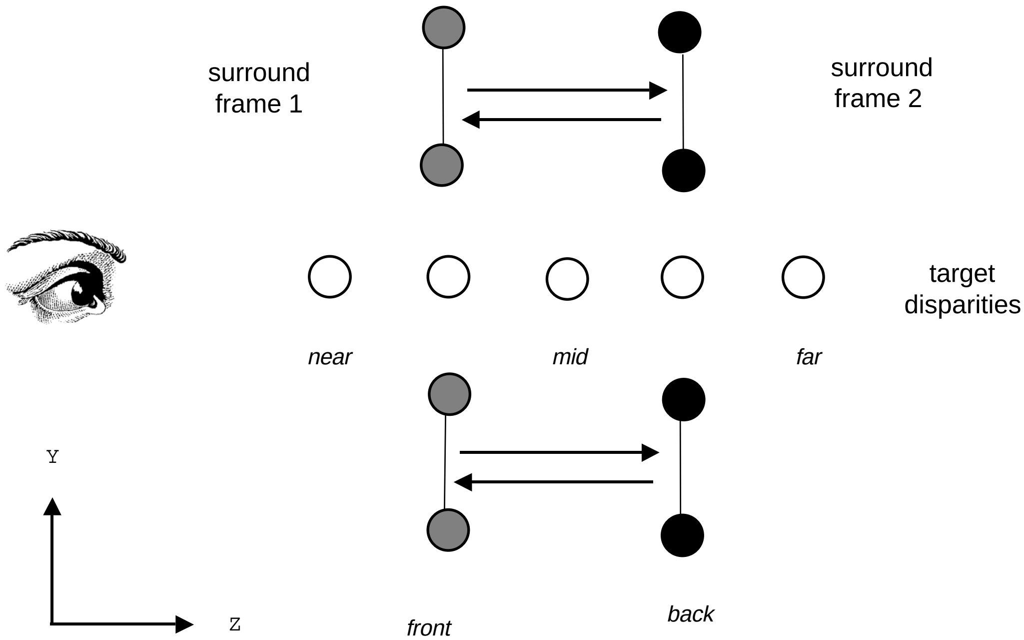

Figure 3. A diagram depicting the side view of the basic TSD-configurations (configurations between the target and surround disparities): near TSD, the target was before the front surround plane; front TSD, the target was in the front disparity plane of the surround; mid TSD, the target was between the two surround planes; back TSD, the target was in the back surround plane; far TSD, the target was behind the back surround plane. No stereomotion was present in the target, whereas 7 cm of stereomotion was presented in the surround in all TSD configurations.

Thus, in the autostereogram presentation technique, the change in surround disparity depicted in Figure 1 was actually achieved by changing the spacing between each pair of dots from 70 pixels center-to-center to 80 pixels center-to-center (in 1.3' pixels). The default spacing of the target line was 75 pixels, halfway between these two surround spacings (although other conditions were used in some experiments). At the screen distance of 70 cm, the viewing distance of the target line with uncrossed convergence was 105 cm in terms of its optical geometry. The observer was thus viewing a stereoscopic space behind the computer screen, apparently inside the monitor. Its properties will be specified subsequently in terms of the absolute dot disparities in arcmin.

To evaluate the role of static disparity in the stereomotion induction, we varied the spatiotemporal configuration between the target and the surround disparities (Figure 3). Absolute surround disparity was varied between 202.5 and 189.5 arcmin, whereas the target row was set at one of five disparities at 6.5 arcmin intervals around the mean surround disparity, giving absolute target disparities of 209, 202.5, 196, 189.5, or 183 arcmin (corresponding to the five locations specified as near, front, mid, back and far in Figure 3). Thus, surround always jumped back or forth by 13 arcmin of disparity; expressed in terms of optical depth-distances, the two surround distances were 101.6 cm and 108.6 cm, giving a simulated stereomotion in the surround of 7 cm in magnitude. The optical distances of the target locations were 98.4, 101.6, 105, 108.6, and 112.5 cm.

In general, we are using the autostereogram technique as a convenient method of presenting wide-field stereograms without extra hardware devices (for future application in functional magnetic resonance imaging studies). The quality of the depth from autostereograms is equivalent to that from the best dichoptic stereoscopes, so there is good reason to expect that the results should generalize to other methodologies. The advantages of some possible shortcomings are evaluated next.

An important property of these spatially repetitive stimuli is that the monocular motions do not have the right structure to produce coherent monocular-induced motion. The basic manipulation that varies disparity in the autostereogram is a uniform change in the spacing between the dots in a horizontal row without varying the overall width of the display. However, this uniform change generates a variety of monocular local dot motions in the surround. Therefore, if the depth motion were a result of lateral motion induction into the target in each eye, it would be expected to be different at each location along the row. The observers reported, however, that the depth motion seemed quite uniform, having the same amplitude with fixation at any point along the target row. This uniformity is one line of evidence that the target stereomotion was induced by the disparity changes per se, rather than indirectly by summation of the lateral induction effects in the two eyes. A second line of evidence is considered under Experiment 1.

Task

The observers' task was to estimate (i) the direction of the target and the surround motions, and (ii) depth motion magnitudes in centimeters. The display was viewed by free-fusion with uncrossed vergence. To assess the role of vergence tracking in the percepts, the task was performed using controlled fixation in two different conditions: (i) target fixation and (ii) surround fixation on the row of disks immediately above the target row. With a static target, vergence eye movements should be minimized for target fixation but maximized for surround fixation. If depth motion perception was mainly based either on retinal disparity or on vergence movements, perceived target motion should therefore be minimal with target fixation and much increased with surround fixation. If, on the other hand, depth motion perception was mainly based on either relative disparity change in the configuration or interactions within a global depth representation, the estimated depth motion magnitudes should be essentially independent of fixation position.

Observers and Equipment

Five observers with normal or corrected-to-normal vision participated in the experiments. The stimulus patterns were produced using custom software implemented on a Macintosh G3 and displayed on a monitor subtending 29° x 22° at 70-cm viewing distance in a dark room.

Experiment 1. Induced Stereomotion

Stimuli

Stereomotion was produced in the four rows of surround disks by alternating their binocular disparity between the 189.5 and 202.5 arcmin disparity planes, while motion induction was estimated for five fixed target disparity positions as shown in Figure 3.

Results

Figure 4 illustrates the resulting percept of depth motion. Because the target was not changed in disparity and had no other cue for depth motion, it should not be perceived as moving in depth. However, a profound depth motion backward and forward was experienced in the stationary target. The basic pattern of results was the same for the five observers under both fixation conditions. The perceived induced magnitude varied with the target/surround disparity configuration (TSD-configuration), but its direction was always opposite to the surround direction.

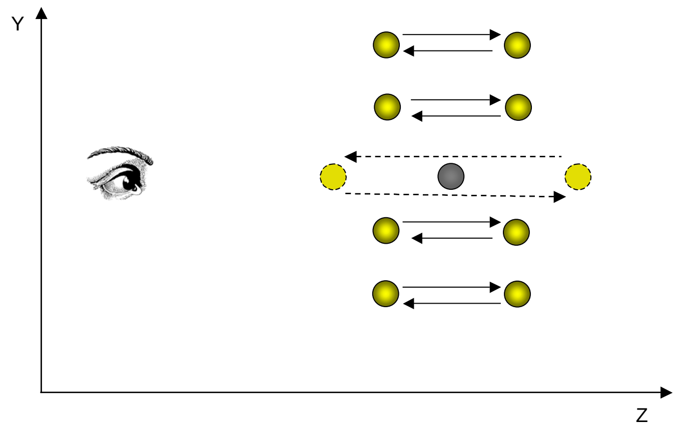

Figure 4. Induced stereomotion. Directions of the surround stereomotion and induced illusory stereomotion in the target were opposite. The upper arrow in the target row indicates the perceived direction of the induced stereomotion when the surround switched from the near to the far depth plane (upper surround arrows), and the lower target arrow indicates its perceived direction for the reverse surround motion (lower surround arrows).

Figure 5 compares the perceived motion magnitudes in the target and in the surround obtained under two different fixation conditions: fixation on the stationary target or on the stereomoving surround. Note that there was no significant change with fixation on the static target versus fixation on the moving surround. The plotted magnitudes are measured for 600-ms frame duration throughout, but a similar pattern of results was observed with variation of the duration over a wide range.

As discussed in “Methods,” a wide variety of monocular local apparent motion extents is generated in the surround rows by a uniform change in the dot spacing. This variety of lateral shifts allows a test of the idea that the induced stereomotion may arise from induced lateral motion in opposite directions in the two eyes. By viewing the display monocularly, we verified that the net result of local motions in the surround gives no perception of a pattern of induced lateral motions in the target dots (such as would be expected by local monocular induction effects). This control observation implies that the perceived stereomotion induced by the surround could *not* have been generated by a combination of monocular lateral inductions in opposite directions in the two eyes. The induction had no monocular correlate, so it must have been generated through a purely stereoscopic induction process.

However, we found that, at least in the range of long-range interactions explored, a small lateral shift in the target (at least 5 arcmin) in synchrony with the surrounding disparity change was critical for inducing

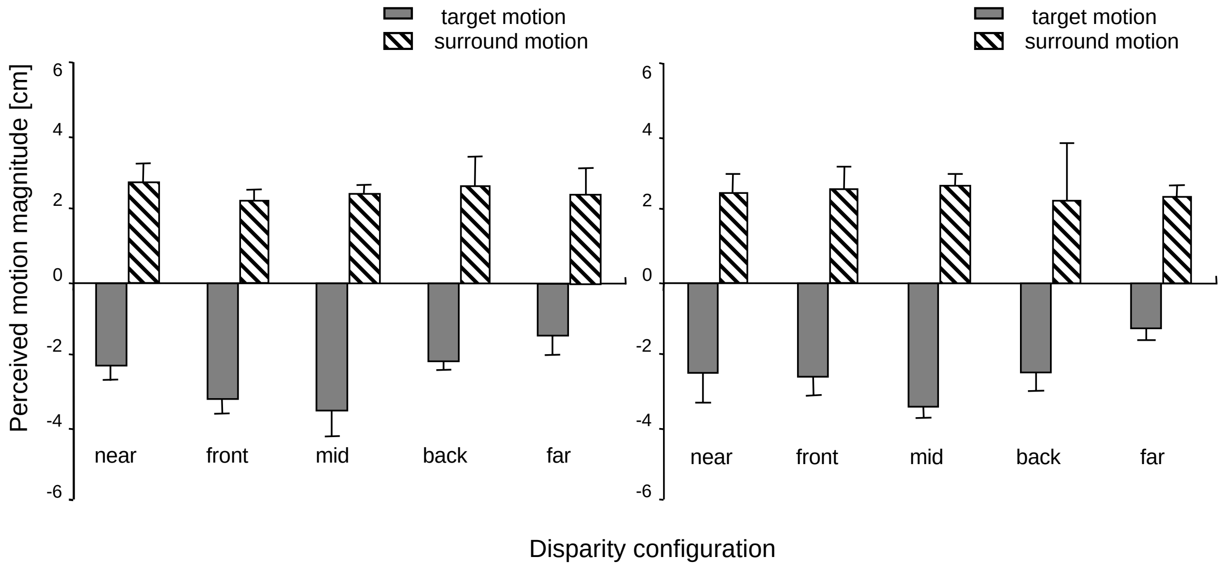

Figure 5. Stereomotion induction. Effect of the stereomoving surround on the stationary target: the target was perceived moving (gray bars) in a direction opposite to the disparity-simulated surround motion (hatched bars). The magnitudes of the induced target motion varied in a function of the TSD configuration. Error bars represent +/- SEM. Left panel. Perceived motion magnitudes in target fixation mode plotted against the TSD. Right panel. Perceived motion magnitudes in surround fixation mode. Comparison of the stereomotion induction in the left and right panel shows that there was no significant change with fixation on the static target versus moving surround (*ns*; $z = -0.135$; $p = 0.823$, > 0.1).

depth motion in all conditions. This result is unexpected because, in the case of frontoparallel motion induction, no target dynamic is required (Duncker, 1929; Nakayama & Tyler, 1978). One way to think about the shift is as a horizontal apparent motion. Another way to think about it is as an interruption and replacement into a new position of each disk. In other words, the horizontal shift interrupted the position signal provided by each disk. Thus, one may ask whether simply interruption of the position of the disks without apparent motion would be sufficient to elicit the stereomotion induction?

Experiment 2. Temporal Interruption of the Target Presentation

One hypothesis for the requirement of a lateral target shift in stereomotion induction could be that the presence of a (lateral) motion component in the target is necessary to allow depth motion to be induced. This requirement was implicitly assumed by Gogel & Griffin (1982) when they studied depth motion induced into a vertically moving target. However, the presence of lateral motion in our paradigm implicitly includes a transient interruption signal. What is the critical variable for stereomotion induction process - the lateral motion component, the position change signal, the transient interruption signal, or some other concomitant factor?

The purpose of Experiment 2 was to test whether interrupting the target presentation in the absence of motion would be sufficient for the stereomotion induction (SMI) to occur.

Stimuli and Procedures

The stimuli were the same dynamic two-frame autostereograms consisting of a target and a surround, both horizontal lines of discs. However, instead of shifting laterally, the target presentation was interrupted in synchrony with the surround transitions, for durations from 0 ms to 575 ms of the 600-ms dwell time of the surround disparities. Thus the experiment included 28 conditions – both the target and surround magnitudes were measured separately for 14 different durations of the temporal interruption (gap) and fixation on the target row, while the surround stereomotion was fixed at 13 arcmin (7 cm in geometric distance). Data were obtained for two of the five observers, with two repetitions of every condition by each observer.

Results and Discussion

In this target interruption paradigm, the target was still presented at only one disparity. However, instead of being shifted laterally in synchrony with the changes in the surround disparity, the target was briefly interrupted at each change of the surround disparity. The data show that all interruption conditions longer than 25 ms elicited stereomotion induction

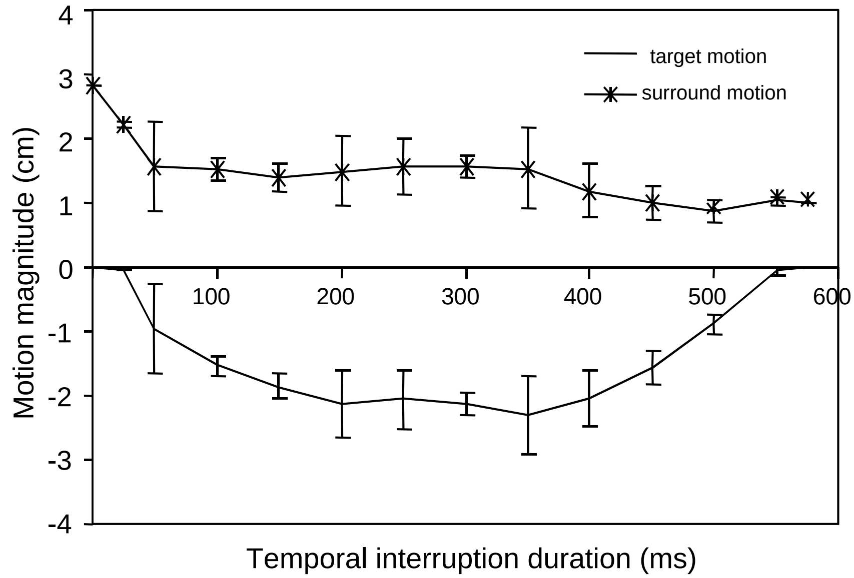

Figure 6. Interruption of the target, ending in synchrony with the surround offset was sufficient to allow full stereomotion induction for a broad range of temporal gap durations. Lower graph (no symbols) plots induced target stereomotion. Upper graph (stars) plots perceived surround stereomotion, from stereomotion relative to the continuous target at the left (gap=0 ms) to absolute stereomotion in the absence of the target at the right. Error bars: ± 1 SEM. Induced motion magnitudes were not significantly different under the two fixation conditions (*ns*; $z = -1.74$, $p = 0.07 > 0.05$).

(Figure 6). Thus, the percept of induced depth motion does not depend on the occurrence of position change of the stimulus (distal or proximal). Retinal displacement or activation of the lateral motion system is irrelevant to SMI. However, no significant stereomotion induction occurred for gaps of 0 and 25 ms, and full induction was only obtained for gap durations of 200-400 ms. It is therefore clear that some temporal interruption of the target is required for stereomotion induction to occur.

The critical variable for stereomotion induction seems to be the occurrence of a transient signal in the target. This transient could operate either by releasing the system from the constraint of a parvocellular position signal in the target, or by providing magnocellular activation in synchrony with the motion signal in the surround.

Without an interruption of the position signal provided by the steady target (interruption duration of zero in Figure 6), any lateral influences in the disparity domain are apparently insufficient to evoke a perceived depth movement in the target (for the long-range disparity conditions evaluated here). On the other hand, if the gap is so long as to reduce the target duration below about 250 ms, the stereomotion induction begins to fall off again. Obviously, the target needs to be visible in order to be perceived as moving. It is somewhat surprising, however, to find that the target needs to be seen for as long as 200 ms to provide the maximum motion induction.

Experiment 3. Cancellation of Stereomotion Induction and Reciprocal Stereomotion Suppression: Data and Modeling

The data in Experiments 1 and 2 were obtained by magnitude estimation of the perceived distance of movement. We performed a cancellation experiment to validate the results and to probe the nature and stability of the stereoinduction process. If it is a weak, imaginary effect, we might expect the presence of a physical disparity change to override the percept, eliminating the stereomotion induction. On the other hand, if the induction derives from a lateral neural pathway, the effect of the compensatory physical stereomotion should add linearly to the induced stereomotion from the surround.

Stimuli and Procedures

Physical disparity changes were added to the target in order to generate depth motion in a direction opposite to its illusory induced motion. The disparity change in the surround was fixed and equal to that used in Experiment 1. Data were obtained for four of the five observers participating in Experiment 1.

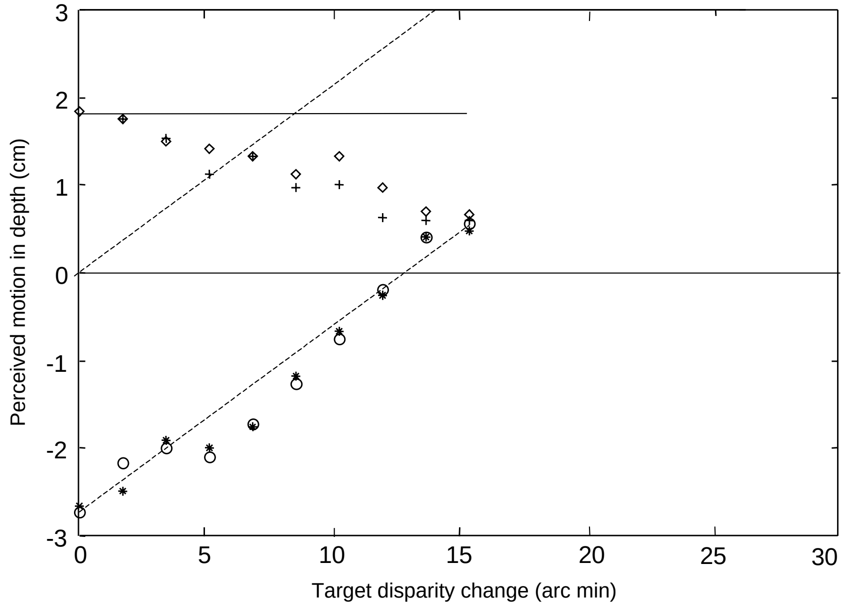

Figure 7. Stereomotion induction cancellation. The target was perceived to have direction opposite to the surround, despite the fact that the directions of both distal stimuli, the target and the surround, were the same. Stars and circles indicate mean perceived target motion for target and surround fixation, respectively. Crosses and diamonds indicate mean perceived surround motion for target and surround fixation, respectively. Motion magnitudes under the two fixation conditions were not significantly different (z = -0.04, p = 0.0964 > 0.05, *ns* for target motion; z = -1.689; p = 0.091 > 0.05, *ns* for surround motion).

Results and Model Fit

The primary result of the previous experiments was that during induction, the surround motion is split between the surround and the target. One interpretation of this behavior is that the induction derives from a frame of reference (FR) against which the motions of the target and surround are judged. If this FR is static, all the motion will be attributed to the surround. If FR itself moves under the influence of the surround, the relative motion between the stationary target and FR will be seen as target motion. The consequence of this model, however, is that the perceived motion of the surround will be proportionately reduced. This model leads to a linear relation between the resultant target and surround stereomotion percepts:

$$PT = -aS \quad (1a)$$

$$PS = (1-a)S \quad (1b)$$

where

PT = perceived target motion magnitude

PS = perceived surround motion magnitude

S = surround motion magnitude defined by disparity change

a = induction coefficient.

The present experiment introduces the additional factor of a compensatory (physical) stereomotion into the target row of dots. With the compensatory motion added to the target, the data show that the induced stereomotion was so strong that it could be progressively cancelled with a physical disparity-defined motion in the target. If the cancellation process is both local, i.e., the canceling affects only the target, and linear, then the interaction of the induction and cancellation components should again be additive (Equation 2, Figure 7).

$$PT = -aS + C \quad (2a)$$

$$PS = (1-a)S \quad (2b)$$

where C =is the compensatory motion magnitude.

Figure 7 plots the results for four observers for perceived target and surround motion versus compensatory disparity change in the target, separately for fixation on the target and fixation on the surround (on the row just above the target). The average SEMs for the four conditions were $\sigma_{TT} = 0.211$, $\sigma_{TS} = 0.256$, $\sigma_{ST} = 0.356$, and $\sigma_{SS} = 0.228$ cm, where the first subscript denotes the fixation condition and the second the response variable. The two motions were perceived to have opposite directions, despite the fact that the directions of both distal stimuli, the target and the surround, were same. The target motion should be the result of subtracting the induced motion component in the target from its disparity-defined component. With

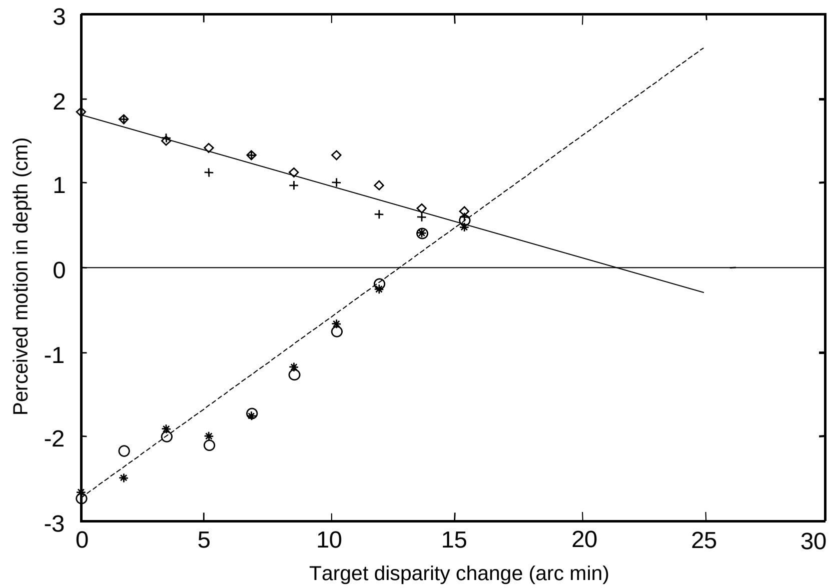

Figure 8. Reciprocal stereomotion suppression. Data as in Figure 7. The upper line denotes the continuation of both the additive and reciprocal models (Equations 2a and 3a) for the target motion beyond the region of cancellation. The lower line shows the prediction of the reciprocal model for the perceived surround motion (Equation 3b).

increasing magnitude of the compensatory stereomotion, the perceived target magnitude decreases in a linear manner, as is predicted by the summation with the compensatory motion C in Equation 2a.

On the other hand, Equation 2b does not fit the behavior of the surround data (Figure 7, full line). The disparity change in the surround was constant, implying that a constant stereomotion magnitude should be expected. However, as is seen from the data, the perceived surround magnitude does not stay constant, but decreases in an almost reciprocal manner relative to the perceived target motion magnitude. We term this novel phenomenon of loss of motion in the stereomoving surround reciprocal stereomotion suppression, because it implies that the target depth motion suppresses the physical stereomotion of the surround, reversing of the process of attributing stereomotion from the surround into the stationary target during direct induction. A reciprocal model, based on the assumption that the disparity changes in both the surround and the target have reciprocal effects on each other but in opposing directions, would give the following predictions:

$$PT = -aS + C \quad (3a)$$

$$PS = (1-a)S - bC \quad (3b)$$

where b is the coefficient of the reciprocal influence from the target to the surround.

Figure 8 shows that this reciprocal model provides a good fit to the data up to the point of motion equality. The surround motion is well fit by the prediction of a linear decline from the influence of the target motion. Because this influence is governed by a free parameter, b, it allows a nonzero motion at the point of equality between perceived target and surround motion. The extrapolation of the reciprocal model beyond the null point would, however, have striking predictions. Beyond the null point, not just the target motion, but both the target and the surround motions, are predicted to switch their directions (Figure 8, lines), so that the surround would be perceived as moving opposite to its physical direction and with an increasing amplitude.

Experiment 4. Beyond Stereomotion Induction Cancellation: Dynamic Frame of Reference Factors

To test the predictions of the reciprocal model that both motions should reverse direction in depth beyond their point of intersection, we extended the cancellation motion over a wide range beyond the intersection point.

Stimuli

The only difference from the stimuli in the previous experiment was that the disparity-defined stereomotion in the target was increased beyond the amplitude required to reach the intersection point.

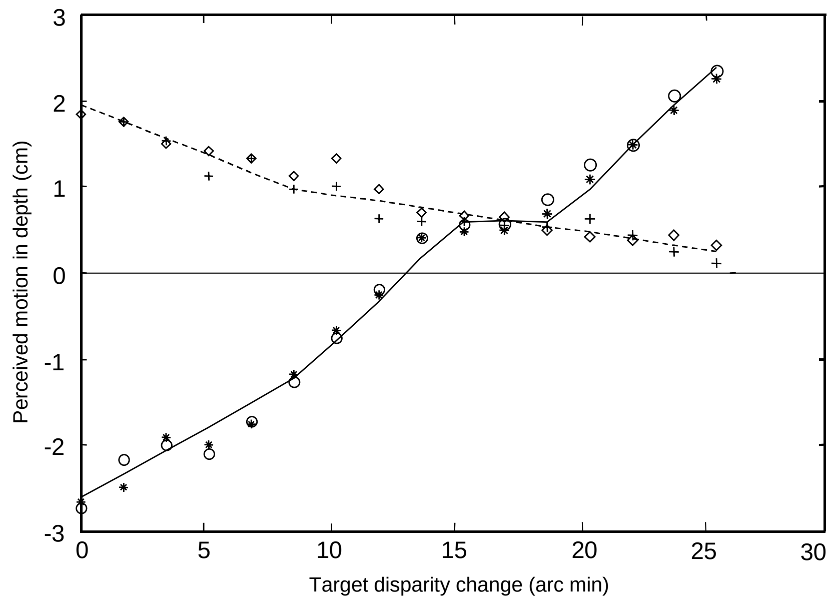

Figure 9. Motion induction cancellation and reciprocal stereomotion suppression. Note that the reciprocal model cannot predict the reciprocal stereomotion suppression characteristics, but the data are well described by the dynamic frame of reference model (see text).

Task and Observers

The task was the same as in the previous experiments: to estimate the relative direction and the magnitudes of the surround and the target motions in centimeters. Three of the observers involved in Experiment 1 participated in Experiment 4, with two repetitions by each observer for each condition.

Results and Discussion

The combined results of Experiments 3 and 4 are shown together in Figure 9. Beyond the null point, the target motion was not perceived in opposition to the surround, but switched direction to move with the surround. In this respect, the depth motion of target followed the prediction of the reciprocal model. Detailed examination of the data reveals that, as the target motion reached the point of physical equality with the surround motion, the two motions both entered a zone of stability, where the perceived motion magnitudes remain almost invariant. Beyond this region, the perceived target motion continued to increase with a linear trend. It should be noted that this linear trend implies that the perceived magnitude of the induced target motion remains translated downward by an amount almost equal to the initial induced magnitude, while combining with added physical disparity change in the target in an almost linear fashion. This is interesting because it implies that significant aspects of the target/surround relationships producing the induction continue to persist implicitly, even though the target undergoes large changes in disparity relative to the surround.

Although the prediction for the target motion in Figure 8 was borne out over a wide range, the perceived surround motion did not follow the reciprocal model prediction of reversing direction in concert with the perceived target motion. Evidently, even the reciprocal model of Figure 8 does not account for the data beyond the null point. Even though the target motion continues upward on an approximately linear trend (dotted line in Figure 9), the surround motion itself declines toward zero across a wide range of disparity changes. The failure of the reciprocal prediction for the surround stereomotion suppression beyond the point of cancellation requires a more elaborated model of the stereomotion induction system.

It is important to note that the fixation mode did not significantly affect the results (compare the data sets for the target and the surround fixation mode, open and cruciform symbols in Figure 9, respectively). The motion magnitudes under both fixation conditions were not significantly different at $p > .05$ ($z = -1.449$, $p = .147$, *ns* for target motion; $z = -1.499$, $p = .134$, *ns* for surround motion). Thus, none of this transfer of perceived motion from one part of the stimulus to the other can be attributed to vergence eye-movement tracking behavior. The twinned data sets serve to emphasize the stability of the results and the requirement of a more accurate model to describe them.

Analysis and Modeling

One theoretical concept that can be applied to induced motion data is that of a frame of reference, which is a single- or multi-dimensional coordinate system with respect to which the properties of objects are perceived (Palmer, 1999; Mozer, 2002). When the target and the surround move through space, they not only change their location with respect to the observer (observer-specific FR), but they also change their positions relative to each other (environment-specific FR). The 2D FR could be assumed to derive from the retinal anatomy. However, the visual system is more sensitive to relative than to absolute motion at low temporal frequencies (Tyler & Torres, 1972). If the reference stimulus is set at different distances from the target, McKee, Welch, Taylor, & Bowne (1990) showed that its effect on motion estimation is independent of distance (up to about 1°). Thus, motion thresholds do not follow the Weber law for distance, i.e., threshold proportionality is not a property of the receptive fields. It is as though the motion is coded into a map whose absolute position is unknown. As soon as the location of the FR is specified, it pins the position of all objects in the map, giving a high acuity for motion wherever it is over the map.

Among the most important aspects of frame of reference in a 3D dynamic scene is its depth position. When one is in a normal environment, the "stable" FR position should coincide with the background because the dominant natural situation is for the background to be stationary or at least the most stable part of the scene. It has been shown that the intrinsic image structure might constitute information about the 3D structure of environmental objects (Perroti, Todd, Lappin, & Philips, 1998; Koenderink & van Doorn, 1992; Lappin & Kraft, 2000). The finding of superior acuity for relative motion (Tyler & Torres, 1972; McKee, et al., 1990) also implies that the FR is not based on retinal anatomy, leading to an alternative hypothesis - that the FR for vision is derived from the image structure in the neighborhood a given point (Lappin, 2001).

The spectrum of dynamic stereomotion phenomena that we report is in a general agreement with the concept that the image structure provides an intrinsic visual reference frame (Rock 1973, 1997; Palmer, 1999; Lappin, 2001), but our data show that this concept needs to be expanded to reflect the global dynamics of the 3D visual scenes. Not only the spatial but also the spatiotemporal structure of object relationships in a scene form the basis of the intrinsic FR. We conceptualize the FR as a dynamic coordinate system operating in a non-homogenous perceptual space. Specifics of the FR model are developed here to explain the results in Experiment 3, implementing three categories of organizational principles for the FR: (i) spatial (or time-independent) weighting, (ii) spatiotemporal-dependent weighting, and (iii) a dynamic variant of the rigidity principle.

FR Weighting

It has been postulated that the frame of reference tends to be attributed to the object that *dominates* in some perceptual respect (Duncker, 1929). In the paradigm case of just a stationary object and a moving surround (induced motion), such a Bayesian heuristic implies that the motion should be attributed to the object most likely to be stationary. According to this principle, the surround in our stimuli should be perceived as a stationary. How then could we explain the fact that the surround is perceived as moving to some extent? First, one should consider that the display is actually a multi-frame system due to the visibility of the outer frame of the monitor (at a 70-cm distance). Obviously, the perceived motions are a net result of multi-factor weightings from each of the frames. As a result, the outer frame should shift the FR away from the surround location with a partial weighting toward the location of the stationary outer frame, implying that the FR should move with the surround, but to less than its full extent.

Other heuristics that might govern the perceived motion of the surround include the following:

Size Dominance

FR tends to be located nearer to the object that is dominating in size.

Surround Dominance

FR tends to be located nearer to an object that surrounds other objects. The net result of these factors is a weighting giving the empirical ratio of the displacement of the target and its surround relative to the FR. The motions of each component of the system are derived from the corresponding absolute disparities: C_1, C_2 for the compensatory motion C; S_1, S_2, for the surround motion S; and F_1, F_2 for the frame of reference F. The disparity changes may be computed from the absolute disparities of the stimuli in the two alternating frames:

$$dC = C_2 - C_1, \quad (4a)$$

$$dS = S_2 - S_1, \quad (4b)$$

$$dF = F_2 - F_1, \quad (4c)$$

where the weightings for F are partitioned according to:

$$F_1 = a*S_1 + (1 - a)*C_1 \quad (5a)$$

$$F_2 = a*S_2 + (1 - a)*C_2 \quad (5b)$$

where a is the weighting factor for the stationary frame of reference.

To obtain the perceived target (PT) and perceived surround (PS) motions

$$PT = k*(dC - dF) \quad (6a)$$

$$PS = k*(dS - dF) \quad (6b)$$

where PT and PS are the perceived target and surround motions relative to the FR in the first and the second frame, and k is the scaling factor from disparity to perceived depth.

However, it is obvious that the perceived motion even of this simple system of two moving objects could not be explained with these principles alone, because Equation 6 predicts a linear decrease in surround motion (gray line in Figure 9). The change in slope of the surround data at a target-motion amplitude of about 8.5 arcmin implies the involvement of further principles. One problem with the above principles is that they implicitly presume time-independence of the FR weighting. When the scene is dynamic, the variations in input locations will imply a particular dynamic of the FR, given the constant weightings in Equation 4c. The deviations of the data from these predictions imply that further spatiotemporal dependencies need to be incorporated beyond the linear FR hypothesis. Each of these factors was incorporated into the model to explain the full data set, and removal of any of them resulted in a qualitatively poorer fit to the data.

Spatial Blockage

According to this concept, the FR can move freely in empty space according to the weights of Equation 4, but cannot pass through the interpolated "wall" defined by the surround stimuli with physical disparities. The movement of the FR is therefore blocked as it reaches the location of the back disparity plane, and is effectively contained within the space defined by the two walls of the two surround disparities, as shown in Equation 7:

$$F_2' = S_2, \text{ when } F_2 \geq S_2 \qquad (7)$$

(Under this concept of spatial blockage, note that the outer frame of the monitor is defined only by the edges of the display, which are not strong enough to evoke an effective wall in the way that the extended array of surround dots could.)

Elastic Rebound

If the FR positions are interdependent over time, the spatial blockage heuristic may be postulated to affect the FR in the following frame by exerting pressure on it in the opposite direction, i.e., toward the opposite "surround wall," in proportion to a "rebound force." This effect is conceived to operate as though the energy for displacement of the FR, which was lost on being blocked by the "back wall," actually rebounds back in the opposite direction. Thus, the change of the FR position was estimated by the following equation:

$$F_1 = a*S_1 + (1 - a)*C_1 - r*(F_2 - F_3) \qquad (8)$$

where r is an empirical parameter for the strength of the "rebound force."

Dynamic Figure/Ground Rigidity

Even with the static modifications of Equations 4-8, the weighted FR could not account for the full dynamics of the system. The predictions of this model for perceived target motion are shown as the straight dotted line in Figure 9. The data conform well to this model away from the region of intersection of the trajectories, but near the intersection there is a region of stability where the perceived motions are almost invariant. This behavior can be explained by supposing the existence of an entirely dynamic linkage between the two motion domains. This linkage imposes the constraint that systems of 3D motions that are close to rigid are seen as completely rigid. Thus, when the difference between two motions is small enough ($dP \approx dC - dS$), both tend to be perceived moving at the same rate. In effect, the target motion is "locked" to the surround motion, thus satisfying the dynamic rigidity principle. This dynamic rigidity is implemented by a hyperbolic threshold function on the difference dP between the target and surround motions:

$$PT_{\pm} = PS \pm \frac{|dP|^{(p+1)}}{\theta + |dP|^{(p)}} \qquad (9)$$

where p and q are the parameters of the hyperbolic threshold and the ± suffix means that the equation is computed separately for +ve and -ve values of dP.

The rigidity constraint of Equation 9 may sound similar to the classic rigidity heuristic of projection from two dimensional to three dimensional, as defined for example by Palmer (1999): "The rigidity heuristic is a bias toward perceiving rigid motions in 3D space rather then plastic deformations, provided that the sensory stimulation is consistent with such an interpretation." The essence of that heuristic is that a rigid solution is likely to be found if it exists. On the other hand, the rigidity constraint in our model enforces perceptual rigidity even when it is *not* precisely consistent with the explicit physical information in the dynamic 3D interpretation (because both stereomotions are explicitly defined by their disparity changes).

The full model (Equations 4-9) of the dynamic target/surround interactions incorporates all of the organizational heuristics formulated above. Figure 9 shows the model fitting of the results under both fixation modes for the whole range - from stereomotion induction to reciprocal stereomotion suppression. The model was fitted in Matlab using the Fmins least-squares fitting algorithm. The model's complexity lies in its derivation and equation structure. It fits the data with only four free parameters beyond the overall scaling factor k, with best-fitting values in Equations 4-9 of k = 0.37, a = 0.57, r = 0.38, q = 0.53, and p = 2.94. The fact that the target motion is attributed to the common "rigid" motion defined by the surround shows that the target/surround classification is taken into account in applying the rigidity constraint.

Discussion

The visual mechanisms for detecting spatiotemporal relationships between objects in a dynamic scene are not well understood. In this study, we first establish purely stereoscopic motion induction with direct estimation and validate it by a cancellation technique over a range of principal spatial target/surround configurations. Furthermore, by conceptualizing the motion induction phenomenon as a particular case of spatiotemporal relationships between two objects in a dynamic scene, we expand the study to cover several basic dynamic interactions between a stereodefined target and its moving surround: (i) stationary target; (ii) the target moving less than the surround; (iii) target moving with the same magnitude as the surround; and (iv) target moving more than the surround.

Our data show that perceived motion depends strongly on the global spatiotemporal structure. Under some conditions, this interaction may result in a range of possible misperceptions, from stereomotion induction with inverted direction of the compensatory motion up to the cancellation point, to another novel phenomenon - reciprocal stereomotion suppression. While stereomotion is induced into a static target by a moving surround, the reciprocal effect is a reduction of perceived motion in the surround evoked by motion added to the smaller target.

What basic principles govern motion induction? Duncker (1929) attributed motion induction to what he termed "object-relative displacement." In the case, where one object may be said to surround another or to act as its frame of reference, setting the frame of reference in motion characteristically causes the surrounded object to be perceived as moving in a direction opposite to its frame (Rock, 1997). The heuristic of attributing motion to the smaller object makes ecological sense because moving objects are generally smaller, while their surrounding environment is usually more stable and immobile.

The reference frame plays a basic role in perceiving relative motion. It is well established that the visual system is more sensitive to relative than to absolute motion (Tyler & Torres, 1972; Nakayama, 1985; Lappin, 1995; McKee et al., 1990; Jansson, Bergstrom, & Epstein, 1994; Papathomas, 1995; Watanabe, 1998; Lappin & van de Grind, 2000; Lappin, 2001). Our conditions differ from those of Duncker (1929) because we are investigating the stereodomain where the segmentation of the target from the surround was based only on differences in their relative stereodynamics. All else being equal between a stereodefined target and its surround (the elements' shape, size, color, luminance, and even their disparities in one of the two frames), differences in their stereodynamic will induce stereomotion to be perceived. If the target remains fixed while the surround varies in disparity, the stereomotion in the target will be in a direction opposite to the perceived stereomotion in the surround.

An important observation was that none of the processes we investigated were affected by the fixation mode. Convergence on the target maximizes the stability of the convergence position. Convergence on the surround maximizes the opportunity for vergence-tracking eye movements, especially considering the long (600 ms) duration for which the stimulus was present at each disparity. The similarity of all results between these two fixation conditions thus implies that perceived stereomotion in the target is not governed by retinal motion in the two eyes. It must therefore derive from lateral induction signals of some kind arising from the surround stereomotion. This conclusion was validated by including nonius lines to reveal the degree of vergence tracking. Most observers saw no movement of the nonius lines during target fixation, implying that the entire stereomotion induction effect was perceptual. The model was designed to account for this perceptual interaction, although under conditions of vergence tracking, there would also be a vergence-based component of the effect.

Surprisingly, we found that, in addition to the known necessary and sufficient conditions for generic apparent motion, a lateral shift in the target position was required for the stereomotion to be induced. We further explored this puzzling result with a target interruption paradigm. Instead of shifting laterally, the target offset was interrupted in synchrony with both surround transitions. The fact that all interruption conditions longer than 25 ms elicited stereomotion induction tells us that the percept of induced depth motion does not depend on lateral motion or position change of the stimulus (distal or proximal). On the other hand, no significant stereomotion induction occurred for interruptions of 0 and 25 ms, and full induction was not obtained until the gap duration reached about 200 ms. It is clear that a temporal interruption of the target is the critical feature, rather than retinal displacement or activation of the lateral motion system, and that the mechanism of activation had a much longer time constant than that for simple luminance transients. Our results indicate the importance of the co-occurrence of transient target and surround signals. It seems that the occurrence of transients in the target frees the interpretative perceptual system to provide a heuristic interpretation appropriate to this dynamic stereoscopic context.

This synchrony may imply a role of cortical synchronous oscillations as a possible neural substrate in interpreting the temporal schemas underlying different motion percepts. We developed a model for the dynamics of the perceived behavior of target/surround spatiotemporal configurations. The model assumes that perceived target motion, as well as the perceived surround motion, depends on the relative displacement between the target and the surround over time. The model incorporates the frame of reference idea and a rationale for involving several organizational heuristics. The core of these heuristics govern the weighting of the FR - a spatial blockage heuristic, an elastic rebound heuristic, and a dynamic rigidity heuristic, the last two of which are new conceptualizations in the study of perception. However, when the spatiotemporal configurations require the dynamic rigidity heuristic to be applied, then it overrides the FR heuristic.

The dissociations established in this study between real motion and perceived motion within a dynamic stereoscene reflect principles of dynamic perceptual organization (beyond basic lateral induction) implemented in global long-range interactions. Overall, on the basis of our results, it seems reasonable to conclude that the spatiotemporal organization of a physical scene has a critical effect on stereomotion perception. Whether and how objects would be perceived to move with respect to each other, with what speed, direction, and trajectory, can be determined to high degree by the global spatiotemporal context.

Conclusions

The principal findings are (i) that purely stereoscopic motion induction exists; (ii) that this phenomenon might be considered to be a specific case of the dynamic spatiotemporal relationships between two objects, when one object could be classified as a surround relative to the other; and (iii) that the whole range of spatiotemporal relationships we studied, from induced stereomotion through the process of its cancellation to reciprocal stereomotion suppression, could be incorporated into a unified model, which combines the concept of a reference frame with other hidden heuristics in producing relative motion percepts. The visual mechanisms for "recognizing" the spatiotemporal relationships among objects in a dynamic scene are not well understood. The observed stereomotion phenomena are a demonstration of general organizational principles that operate in the dynamic visual perception and misperception.

Acknowledgements

Supported by NIH EY 7890.
Commercial relationships: none.

References

Attneave, F., & Block, G. (1973). Apparent movement in tridimensional space. *Perception and Psychophysics 13*, 301-307.

Corbin, H. H. (1942). The perception of moving and apparent movement in visual depth. *Archives of Psychology*, *273*, 1-50.

Duncker, K. (1937). Induced motion. In W.E. Ellis (Ed.), *A Sourcebook of Gestalt Psychology*. London: Routledge & Kegan Paul. (Original work published 1929)

Farné, M. (1972). Studies on induced motion in the third dimension. *Perception*, *1*, 351-357. [PubMed]

Gogel, W. C., & Griffin, B. W. (1982). Spatial induction of illusory motion. *Perception, 11*, 187-99. [PubMed]

Jansson, G., Bergström, S. S., & Epstein, W. (1994). *Perceiving Objects and Events*. Hillsdale, NJ: Lawrence Erlbaum.

Koenderink, J. J., & van Doorn, A. J. (1992). Second-order optic flow. *Journal of the Optical Society of America A, 9*, 530-538.

Lappin J. S. (1995). Visible information about structure from motion. In W. Epstein & B. Rogers (Eds.), *Handbook of Perception and Cognition: Volume 5, Perception of Space and Motion* (pp. 165-199). New York: Academic Press.

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived shapes. *Psychological Review,* 107, 6-38.[PubMed]

Lappin, J. S., & van de Grind, W A. (2000). Visual forms in space-time. In L. Albertazzi (Ed.), *Unfolding Perceptual Continua*. Amsterdam: John Benjamins.

Lappin, J. S. (2001). Coherence of early motion signals. *Vision Research 41*, 1631–1644. [PubMed]

McKee, S., Welch, L., Taylor, D., & Bowne, S. (1990). Finding the common bond: Stereoacuity and the other hyperacuities. *Vision Research 30*, 879-891. [PubMed]

Minev, K., & Likova, L. T. (1999). Autostereograms as a research tool in stereoscopic vision: Interactions between some cues in perception of motion-in-depth [Abstract]. *Perception 28*, 135.

Mozer, M.C. (2002). Frames of reference in unilateral neglect and visual perception: A computational perspective. *Psychological Review 109*, 156-185. [PubMed]

Nakayama, K., & Tyler, C. W. (1978). Relative motion induced in stationary lines. *Vision Research*, *18*, 1663-1668. [PubMed]

Nakayama, K. (1985). Biological image processing: A review. *Vision Research, 25*, 625-660. [PubMed]

Palmer, S. (1999). *Vision Science – Photons to Phenomenology*. Cambridge, MA: MIT Press.

Perroti, V. J., Todd, J. T., Lappin, J. S., & Philips, F. (1998). The perception of surface curvature from optical motion. *Perception and Psychophysics, 60*, 377-388. [PubMed]

Papathomas, T. (1995). Coherence of early motion signals. In T. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early Vision and Beyond*. Cambridge, MA: MIT Press.

Rock, I. (1973). *Orientation and Form.* New York: Academic Press.

Rock, I. (1997). The concept of indirect perception. In I. Rock (Ed.), *Indirect Perception*. Cambridge, MA: MIT Press.

Tyler, C. W., & Clarke, M. B. (1990). The autostereogram. *Society of Photo-optical Instrumentation and Engineers, 1256*, 182-197.

Tyler, C. W., & Torres, J. (1972). Frequency response characteristics for sinusoidal movement in the fovea and periphery. *Perception and Psychophysics 12*, 232-236.

Watanabe, T. (Ed.) (1998). *High-level Motion Processing*. Cambridge, MA: MIT Press.