European Journal of Neuroscience, Vol. 18, pp. 2874–2882, 2003

Mechanisms of time-based figure–ground segregation

Farid I. Kandil* and Manfred Fahle
Human Neurobiology, University of Bremen, Argonnenstraße 3, 28211 Bremen, Germany

Keywords: dichoptic stimulation, figure–ground segregation, human, monocular detectors; motion detection, segmentation

Abstract

Figure–ground segregation can rely on purely temporal information, that is, on short temporal delays between positional changes of elements in figure and ground (Kandil, F.I. & Fahle, M. (2001) *Eur. J. Neurosci.*, **13**, 2004–2008). Here, we investigate the underlying mechanisms by measuring temporal segregation thresholds for various kinds of motion cues. Segregation can rely on monocular first-order motion (based on luminance modulation) and second-order motion cues (contrast modulation) with a high temporal resolution of ≈20 ms. The mechanism can also use isoluminant motion with a reduced temporal resolution of 60 ms. Figure–ground segregation can be achieved even at presentation frequencies too high for human subjects to inspect successive frames individually. In contrast, when stimuli are presented dichoptically, i.e. separately to both eyes, subjects are unable to perceive any segregation, irrespective of temporal frequency. We propose that segregation in these displays is detected by a mechanism consisting of at least two stages. On the first level, standard motion or flicker detectors signal local positional changes (flips). On the second level, a segregation mechanism combines the local activities of the low-level detectors with high temporal precision. Our findings suggest that the segregation mechanism can rely on monocular detectors but not on binocular mechanisms. Moreover, the results oppose the idea that segregation in these displays is achieved by motion detectors of a higher order (motion-from-motion), but favour mechanisms sensitive to short temporal delays even without activation of higher-order motion detectors.

Introduction

When dots in two adjacent fields of a stimulus flicker at the same frequency but out of phase, the two fields are perceived as being segregated by an illusory boundary (Ramachandran & Rogers-Ramachandran, 1991; Fahle, 1993; Rogers-Ramachandran & Ramachandran, 1998). In these displays, stimulus elements appear at different times in figure and ground. Subjectively, the display is segregated into two hemifields, or a figure against a ground, even at high frequencies of 20–30 Hz (corresponding to frame durations of 25–17 ms) and phase delays down to 5 ms. These frame rates are too high to allow inspection of individual frames and hence to allow separation on the basis of differences between frames containing either the figure or the ground. Therefore, the underlying segregation mechanism must be more complex than isolating a single frame of the presentation. Ramachandran & Rogers-Ramachandran (1991) and Rogers-Ramachandran & Ramachandran (1998) argued that the contour detection in their displays was achieved in the magnocellular pathway and Fahle (1993) mentioned that direction-selective movement detectors might be able to segregate the fields in his stimuli. First-order motion detectors with an orientation orthogonal to the boundary between figure and ground may detect a shift of luminance across that boundary because these receptors are activated by stimulus appearance at slightly differing times at nearby positions. Hence, at least in principle, motion detectors would be able to detect the boundary between figure and ground. Alternatively, the phase differences between flicker detectors in figure and ground may subserve segregation.

While the temporal segregation demonstrated by these and subsequent studies using similar stimuli may have been detected by first-order motion detectors due to the net luminance shifts between subsequent frames of the stimulus (Fahle & Koch, 1995; Leonards, Singer, & Fahle, 1996; Usher & Donnelly, 1998; Forte, Hogben, & Ross, 1999), there is evidence for temporal segregation which cannot be mediated by these mechanisms. Lee & Blake (1999) presented an array of non-overlapping apertures. In each aperture, contours moved in one of two directions. At random times, directions reversed either in all apertures constituting the figure or else in those of the surround. The advantage of these displays is that all stimuli are permanently displayed and hence global luminance stays constant. Figure and ground cannot be segregated on the basis of a single frame, and first-order motion detectors are permanently stimulated locally and hence will not easily perceive the segregation between figure and ground. However, motion directions reversed randomly and thus independently in figure and ground. As a consequence the temporal delays between reversals in figure and ground varied widely and hence prohibited the quantitative assessment of threshold delays. Adelson & Farid (1999) pointed to another drawback of this procedure. If stimuli in one region undergo a rapid series of reversals while those in the other region continue shifting in their original direction then temporal filtering of the stimulus unveils a contrast difference between these arrays and hence converts the temporal into a contrast discrimination task (Adelson & Farid, 1999; however, cf. Blake & Lee, 1999).

Recently, we introduced displays in which subjects have to detect the temporal asynchrony (phase delay) between local motions (‘flips’) of randomly orientated colons (pairs of dots) in figure and ground in order to perceive the target. All stimulus dots are permanently presented; they undergo small shifts in position at slightly differing times in the figure vs. in the surround (see Fig. 1, Materials and Methods and Kandil & Fahle, 2001). By systematically varying frequency between

---

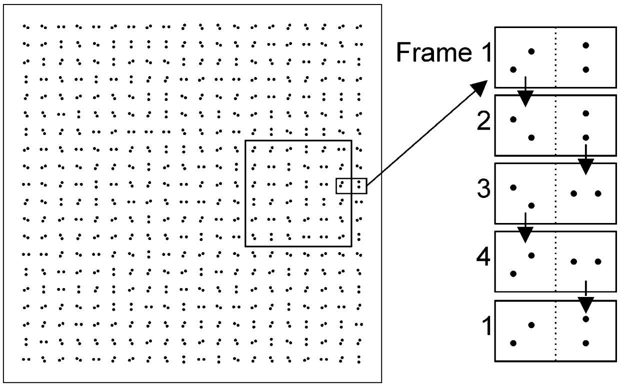
Correspondence: Dr Farid I. Kandil, at current address below.
E-mail: kandil@uni-muenster.de

**Current address*: Department of Psychology II, WW University of Münster, Fliednerstrasse 21, 48149 Münster, Germany.

Received 19 September 2002, revised 9 September 2003, accepted 11 September 2003

doi:10.1046/j.1460-9568.2003.03022.x

(a) Basic Stimulus Field



(b) Temporal Designs

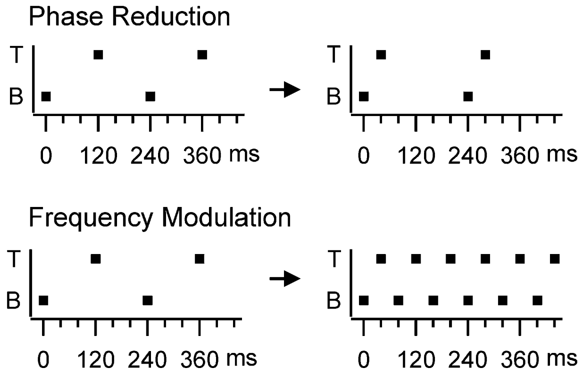


FIG. 1. (a) The basic stimulus display consisted of 20 × 20 randomly orientated colons (dipoles) which flipped (rotated instantaneously by 90°) with a certain frequency. While all colons in the target flipped synchronously between frames 1 and 2 as well as between frames 3 and 4, colons in the background flipped with a phase delay, that is, between frames 2 and 3 as well as between frames 4 and 1. The frame sequence was shown repeatedly for 1 s. (b) In order to measure threshold delays we introduced two temporal designs. In the phase reduction design, the flip frequency was constant at 4.2 Hz (corresponding to 240 ms per period) while the phase (delay) between flip times in figure and surround was reduced from 180° (≈ 120 ms) to 15° (≈ 10 ms). In the frequency modulation design, the frequency was increased from 4.2 to 50 Hz (corresponding to a decrease in period length from 240 to 20 ms) while the phase was constant at 180°, resulting in phase delays diminishing again from 120 ms (at 4.2 Hz) to 10 ms (at 50 Hz).

subsequent flips, we obtained threshold delays of 23 ms when flip frequency was increased and colons in figure and ground flipped in counter-phase (i.e. at the largest possible delay for each frequency). When the phase between flips in the figure vs. in the ground was reduced while keeping the frequency constant at 4.2 Hz, threshold delays were ≈13 ms (Kandil & Fahle, 2001).

As in the model of Motoyoshi & Nishida (2001), we argue that two stages are required to segregate figure from ground in these displays: a standard first-order motion detector (e.g. of the Reichardt type) or flicker detector at the first stage which detects the local changes in position (flips), and a second-stage segregation mechanism which reads and globally combines the output of the first stage with a high temporal resolution of ≈20 ms. Moving stimuli will usually stimulate flicker detectors, too, so when we refer to (first-order) motion detectors throughout the text, flicker detectors will be included. Actually, many researchers in motion perception regard flicker detectors as special forms of motion detectors (e.g. Zanker, 1990, 1993, 1996). We examined the temporal precision of the segregation mechanism for various conditions: motion (i) from monocular achromatic, luminance-defined stimuli (experiment 1); (ii) from isoluminant colour-defined stimuli (experiment 1); (iii) from non-Fourier (i.e. contrast-modulated) stimuli (experiment 2); and (iv) from monocular vs. dichoptic (differing between the eyes) presentation of stimulus dots (experiments 3, 4 and 5). Isoluminant stimuli require slightly higher delays than those defined by luminance or contrast. Dichoptic presentation of stimulus elements, however, does not allow time-based segregation between figure and ground. Our results strongly argue against underlying mechanisms such as motion-from-motion detectors.

Materials and methods

Six students, aged between 23 and 30 years, participated in each of the five experiments. All were naive as to the purpose of the study and all had normal or corrected-to-normal visual acuity and normal stereopsis. Prior to the main experiments, subjects were tested with the basic task described in Fig. 1 to ensure that they had a high resolution in temporal figure–ground segregation tasks. This type of experiment was approved by the Bremen ethics review and approval committee.

The basic stimulus was identical to the one used in Kandil & Fahle (2001). It consisted of a regular grid of 20 × 20 randomly orientated colons (dipoles) which flipped around their theoretical midpoints by 90° at a given frequency (Fig. 1a). The colons flipped synchronously within the target as well as within the background, but asynchronously (that is, with a defined phase delay) between regions. The target region was defined by a 6 × 6 square of synchronously flipping colons. For sufficiently large delays of the target flips relative to the background flips, the target region becomes detectable. Subjects had to indicate the position of the target region by pressing one of four cursor keys on a standard PC keyboard in a four-alternative forced-choice task. The distribution of the colons remained constant during each stimulus presentation, all colons had the same contrast and were randomly orientated, and motion directions were ambivalent and did not differ between figure and ground. The segregation of the target region was not signalled by any luminance, orientation or global motion cues.

We used two designs to determine threshold phase delays (Fig. 1b). In the 'phase reduction' (PR) design, the frequency remained fixed at 4.2 Hz (corresponding to an interval of 240 ms between flips) while the delay was reduced stepwise from 120 to 10 ms. In the 'frequency modulation' (FM) task the flip frequency increased from 4.2 to 50 Hz (corresponding to a decrease between flips from 240 to 20 ms) while the phase difference was constant at 180° (i.e. counter-phase), resulting in delays between 120 and 10 ms. Sixteen trials were presented per design and condition. Thresholds were calculated separately for both designs by linearly interpolating the minimum delay required to obtain 62.5% correct responses because this level is midway between chance and perfect performance.

Stimuli were presented on a 21-inch colour CRT monitor (EIZO FlexScan F-784T) with a spatial resolution of 1152 × 864 pixels and a frame rate of 100 Hz, driven by an AMD Duron 800 MHz PC via an Asus V7700 graphics board. Subjects were seated in a dimly lit room and viewed the display from a distance of 40 cm. The stimulus displays subtended 37° of visual angle. The mean distance between colons was 111 arcmin and the distance between the two dots constituting a colon measured 32 arcmin, while each dot had a diameter of 7.3 arcmin.

In experiments 3, 4 and 5, the inputs to the right and the left eye differed from each other; hence the ray paths of the eyes had to be separated. We used either ferro-electric shutter goggles or red–green anaglyphs, which allow the presentation of both eyes' views on the same monitor. In experiments 4 and 5 where maximal temporal resolution was required, we used red–green anaglyphs. In contrast, in experiment 3, where colour differences of the stimuli presented to the two eyes would have diminished any effects of completion (masking), we used shutter goggles which allow the stimuli for both eyes to be maximally similar so subjects cannot distinguish between the inputs from the two eyes. However, as the images for both eyes are presented on successive frames, the monocular frame rate is reduced by a factor of two. The resulting monocular flicker of 50 Hz had no negative impact on the subject's performance (see results of experiment 3).

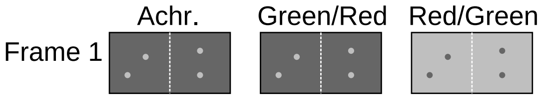
Results

Experiment 1. Motion of luminance- and colour-modulated dots

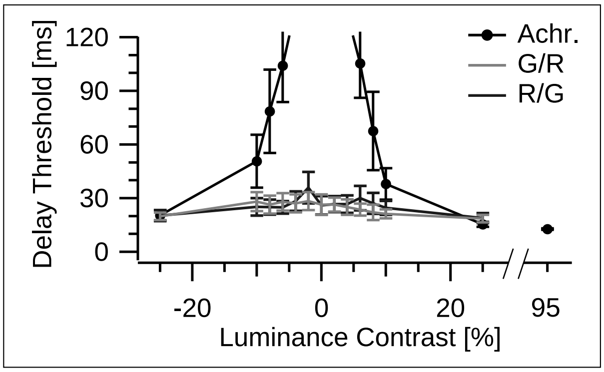
We first tested time-based figure–ground segregation based on presentation of dots defined by achromatic luminance contrast, or else by isoluminant colour contrast. In the luminance modulation condition ('Achr.' in Fig. 2), performance in the basic task was determined as a function of the luminance contrast.

Stimuli were grey dots either brighter or darker than their grey surround (5 cd/m^2). Luminance contrast was defined as ±95, 25, 10, 8, 6, 4, 2 or 0%. To obtain appropriate starting values for the chromatic green–red and red–green conditions, individual isoluminance points for green dots on red ground and vice versa were established first. This was achieved by requiring subjects to use the method of self-adjustment to modify the dot luminance until the flipping colons evoked a minimal perception of motion at a flip frequency of 12.5 Hz. Thereafter, a 'luminance bracketing' method was used, i.e. the test was conducted 15 times with contrast values of ±25, 12, 10, 8, 6, 4, 2 and 0% around the individual's subjective isoluminant point. In contrast to conducting the test only once at the subjective point of isoluminance, the bracketing procedure allowed a *post hoc* definition of the real point of isoluminance for this very task, as the contrast leading to the lowest temporal resolution. The luminance of the (monitor) background was again constant at 5 cd/m^2.

(a) Achromatic and chromatic stimuli



(b) Results: Phase Reduction



(c) Results: Frequency Modulation

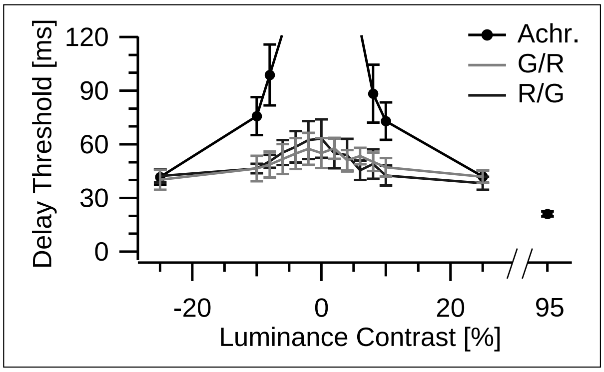


FIG. 2. (a) Performance in the basic task was tested as a function of luminance contrast between dots and monitor background with achromatic displays (Achr.) as well as chromatic displays by presenting either green dots on red ground (G–R) or inversely red dots on green ground (R–G). (b and c) Threshold delays for all three conditions for the two temporal designs, phase reduction (PR) and frequency modulation (FM).

Figure 2b and c shows the results for the PR and the FM designs, respectively. In each graph, thresholds are plotted as a function of the luminance contrast between the dots and their immediate background for all three conditions: achromatic stimuli (dark lines with circular symbols), green dots on red ground (bright lines without symbols), and red dots on green ground (dark lines without symbols).

With high luminance contrast in the achromatic displays, we obtained results similar to those previously reported (Kandil & Fahle, 2001). Specifically, we observed a threshold phase delay ≈10 ms in the PR design and ≈20 ms in the FM design. Reducing the luminance contrast in achromatic displays from 25 to 10% increased thresholds as determined by PR and FM by a factor of 1.6 and 2.0, respectively. Reducing the contrast further increased thresholds monotonically. For luminance contrasts below ≈5% no segregation occurred although the dots and the local motion were still perceived.

Optimal results for the chromatic displays were similar whether they were obtained with presentations of green dots against a red background or with red dots against a green background (Fig. 2b and c). At 25% added luminance contrast, thresholds were identical to the ones obtained under achromatic conditions. However, the thresholds for the red–green stimuli increased only by a factor of 1.5 when the added luminance contrast was reduced from 25 to 0%. This resulted in highest thresholds of 30 ms for the PR design and of 60 ms (corresponding to ≈8.3 Hz) for the FM design.

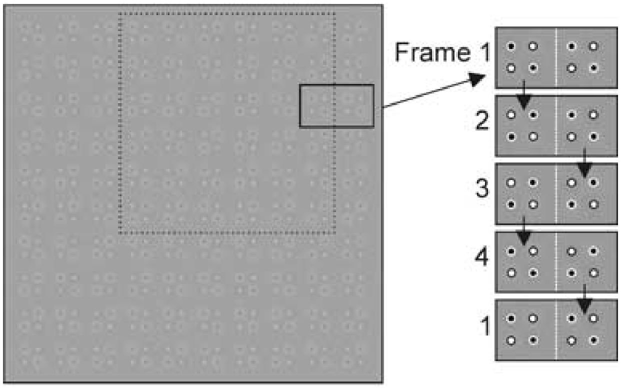
The equivalent luminance contrast describes how much luminance contrast is required in achromatic stimuli for subjects to reach results similar to those obtained under isoluminant chromatic stimulation. In both designs, PR and FM, the equivalent luminance contrast lay between 13 and 20%.

Both the very broad contrast range, from −5 to +5% in which no segregation could be detected in achromatic displays, and the narrow steps (2%), by which the luminance contrast was varied in the red–green conditions, ensured that there were no hidden peaks of deteriorated performance which might have been overlooked by the procedure used here. Obviously, pure colour contrast is sufficient to activate time-based figure–ground segregation.

Experiment 2. Non-Fourier (second-order) motion

The next question was whether, in addition to luminance- and colour-defined inputs, second-order stimuli might also be able to evoke a time-based figure–ground segregation. In this experiment, the elementary stimuli should be invisible to first-order motion detectors and only be detected by second-order motion mechanisms. To prevent first-order motion detectors as far as possible from detecting local motion, colons were replaced by a set of four differences of Gaussians (DOGs). Two of these had a negative polarity and two had a positive polarity. Contrast reversals of these DOGs would not activate first-order motion detectors with sufficiently large receptive fields.

(a) Differences of Gaussians (DOGs)



(b) Results

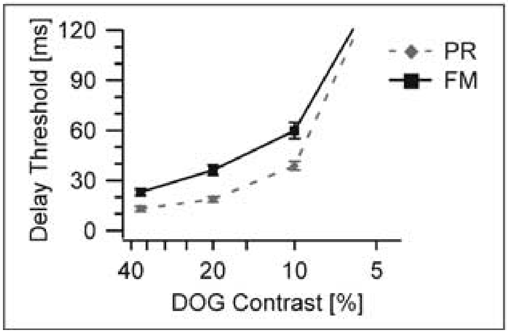


FIG. 3. (a) Colons were replaced by a set of four differences of Gaussians (DOGs), two with negative and two with positive polarity, interchanging their polarity according to the same temporal designs as used in experiment 1. Note that, for clarity of this figure, only a subgroup of the original 20×20 sets of DOGs are shown, the target being the enframed part. Moreover, the contrast of the DOGs in the five clippings on the right side is extremely exaggerated. (b) Threshold delays obtained using the phase reduction (PR, dashed line) and the frequency modulation (FM, solid line) design.

In this experiment, each 'group' of the stimulus consisted of four rather than two elements. The colon flips in experiment 1 corresponded to a reversal of polarity in all four DOGs. This was equivalent to an interchange of the positions of the DOGs with negative and positive polarity (Fig. 3a). Using the additional elements allowed us to keep the local luminance constant; hence the contrast reversals should not activate first-order motion detectors with sufficiently large receptive fields assuming sufficiently linear characteristics of both display and visual system. Under these conditions, the first-order motion detectors would not detect a change in luminance and hence not signal motion. Thresholds were measured for DOGs' modulation contrasts of 5, 10, 20 and 37%.

Background luminance, as well as the mean luminance of all the DOGs, was 1 cd/m^2. DOGs with a positive polarity were calculated as the difference between two Gaussians, with respective spatial constants of 6.1 (positive part) and 12.7 arcmin (negative part), and with the positive part having an amplitude 3.57 times than that of the negative part. In DOGs with negative polarity, the constants were reversed. All parameters were computed exactly for the calibrated monitor and the overall luminance as well as that of the brightest and darkest pixels were calibrated using a luminance meter.

Figure 3b shows threshold delays for the segregation task determined with both designs, PR (dotted line) and FM (solid line) for contrast-defined elements. When the contrast was reduced from 37 via 20 to 10%, mean thresholds increased from 13 via 18 to 40 ms for the PR design and from 22 via 38 to 60 ms for the FM design. In the 5% contrast condition, subjects were still able to detect the DOGs and their local motion, yet even at the lowest frequency (4 Hz) they could not segregate the figure from ground. Hence, for sufficient contrasts, even stimulus elements which reverse contrast at constant luminance and which therefore will hardly activate elementary motion detectors can subserve figure–ground segregation.

Experiment 3. Binocular vs. monocular motion I: dichoptic masking

The aim of this experiment was to clarify whether the neuronal mechanisms subserving time-based figure–ground segregation operate on an early, monocular or on a later, binocular stage of processing. To that aim, results for monocular stimuli were compared to those supplying conflicting or masking information to the two eyes in this experiment. Presenting different stimuli to the two eyes is generally called 'dichoptic'.

In the binocular and monocular conditions, the same stimulus sequence was presented either to both eyes ($L = R$; Fig. 4a) or else to the left (L; Fig. 4b) or to the right (R; Fig. 4c) eye only. In the dichoptic masking condition ($L \times R$; Fig. 4d), the left eye sees the stimulus sequence starting with frame 1 whereas the right sees the same sequence starting with frame 3. Here, no motion information can be obtained from the stimulus once the images from both eyes are fused and hence information cannot signal the segregation on a binocular stage.

Dichoptic stimulation was achieved by presenting images for the left and the right eye on alternate video frames and alternately masking the eyes with ferro-electric shutter goggles (FE-1, Cambridge Research Systems). This procedure confines the monocular frame rate to 50 Hz corresponding to a minimal phase delay of 20 ms.

In order to allow spatial summation of the images and to prevent (local) binocular rivalry, the background luminance was set to a relatively low level (0.02 cd/m^2) and the luminance of the dots was 1.5 cd/m^2 when measured through open glasses and 0.1 cd/m^2 when measured through closed glasses. The goggles only diminished luminance by a factor of ≈ 15 in the shut state relative to the open state. The remaining crosstalk ('leakage') between the stimuli to the two eyes increased task difficulty in the critical complementing condition ('$L \times R$'), the task becoming impossible to solve for complete crosstalk.

Results are shown in Fig. 4e. For the PR design (light bars), it was impossible to determine valid thresholds because most subjects yielded between 85 and 100% correct responses, reaching suprathreshold performance even with the shortest phase delay possible with this set-up (20 ms). Performance in the $L \times R$ condition was similar to that in the other conditions. In accordance with the results of the previous experiments, performance in the FM condition was below that in the PR condition, and thresholds could be computed. Due to the lower frame rate causing larger differences between technically possible phase delays, threshold values are less reliable. Again, there were no significant differences between the dichoptic masking condition $L \times R$ and the monocular conditions. However, the threshold was significantly lower (≈ 8 ms) in the binocular condition $L = R$ than under monocular stimulation (repeated-measures ANOVA, $F = 6.16 > F_{(3,18,0.95)}$). Obviously, binocular masking of time-defined figure–ground segregation was not effective, a first hint that the mechanisms subserving this segregation operate on a monocular level.

Experiment 4. Binocular vs. monocular motion II: simple dichoptic motion

In this second experiment investigating the impact of binocular information on the temporal segmentation process, we introduced a

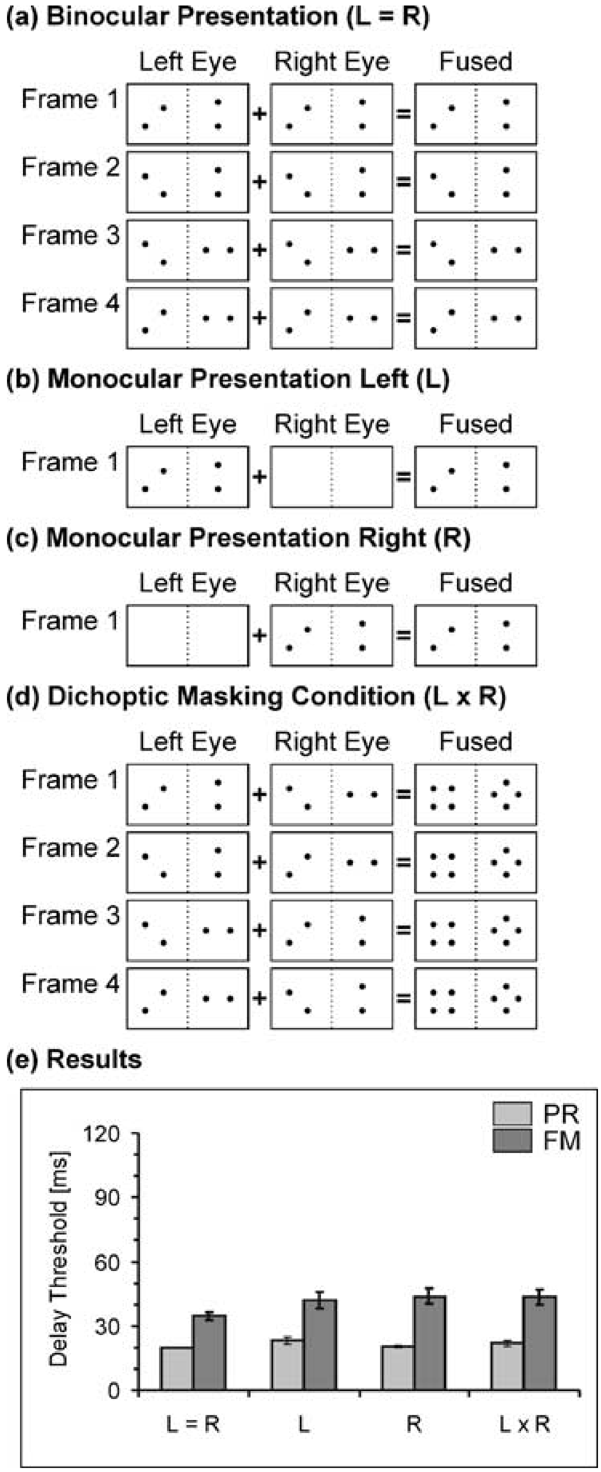


FIG. 4. (a–c) In one binocular dichoptic and two monocular conditions, the basic stimulus was either presented to both eyes (L = R) or restricted to the left (L) or right (R) eye. (d) In the critical complementing condition (L × R) both eyes saw the basic stimulus resulting in the same temporal flip order, yet images to the two eyes complemented each other when binocularly fused, and this masked the segregation information. (e) Results for the four tasks, separately shown for both temporal designs, phase reduction (PR, light bars) and frequency modulation (FM, dark bars). Note that in this first experiment thresholds were not as precisely determined as in the other experiments due to the lower monocular frame rate of 50 Hz.

purely binocular or dichoptic motion signal. This is to say that stimulus elements moved from one position in one eye to another position in the contralateral eye without any monocular cues as to the segregation. Note that, in order to detect the target, the pictures shown to the two eyes have to be compared over time.

We used two dichoptic presentation tasks, i.e. two tests with stimulus differences between the two eyes. In the ‘temporal fusion’ (TF-dich) task used in this experiment, the stimulus was alternately presented to the left and the right eye (Fig. 5a and c). The ray paths of the eyes were separated here by presenting red and green dots on the monitor, which subjects examined through red–green goggles. In contrast to the previous experiment, any crosstalk between the stimuli for the two eyes would have transformed the binocular task into a monocular one which can easily be solved, as we saw in the first experiment. Quadruplets rather than colons were used throughout this experiment to assimilate the fused images to those of the fifth experiment. In the second condition (TF-bin), the red–green glasses were removed and subjects repeated the segregation task while watching the same alternating red and green images as before, yet binocularly rather than dichoptically. The control condition (TF-con) tested whether subjects were able to combine the stimuli of both eyes, a combination necessary for the main task of this experiment. Subjects had to put on the red–green glasses again and to observe dichoptic motion stimuli which were obtained by rotation of the quadruplets presented to the right eye. As is shown in the ‘Fused’ column of Fig. 5c, quadruplets in figure and ground rotated in the same direction. Subjects had to report in a two-alternative forced-choice task whether rotation was clockwise or anticlockwise. In this condition, only presentation frequency was modulated because the motion impression in the PR task was too jagged. (In a pilot study we used a segregation task based on differences in rotation direction between figure and ground. However, the figure did not become apparent even under binocular conditions, hence that task was replaced by the present task).

In addition, the binocular (L = R) and the monocular (L or R) conditions of experiment 3 were repeated to obtain thresholds using the higher temporal resolution of 100 Hz. In these last controls all frames of one trial were presented either to both eyes (using yellow dots), to the left (e.g. red dots) or to the right eye (e.g. green dots).

In the critical dichoptic TF condition (TF-dich; Fig. 5f) thresholds could not be determined because, even with a frequency as low as 2 Hz (≈ 240 ms/frame), subjects did not perform any better than chance (24.8% ± 1.7 ≈25%). In contrast, subjects could identify the local rotation sense up to a frequency of 12.5 Hz (≈ 40 ms/frame) in the control condition ‘TF-con’. When glasses were removed and subjects observed the stimuli binocularly but with colour changes from red to green and back between frames (TF-bin), performance was as good as under binocular stimulation without colour changes (L = R) and, similar to the results of the third experiment, somewhat better than under monocular presentation (L or R). Due to the higher frame rate of 100 Hz with a step size of 10 ms between different delays, the threshold values for the L = R and L or R conditions obtained here are more reliable than the ones determined in experiment 3. The results clearly show that observers were unable to use either temporal delays in dichoptic motion information or delays in dichoptic flicker to segregate figure from ground.

Experiment 5. Binocular vs. monocular motion III: complex dichoptic motion

In this last experiment, a more complex stimulus configuration was created which would allow distinction between figure and ground on the basis of temporal information if the underlying neuronal mechanism had access to the fused image. Here, the two images presented to

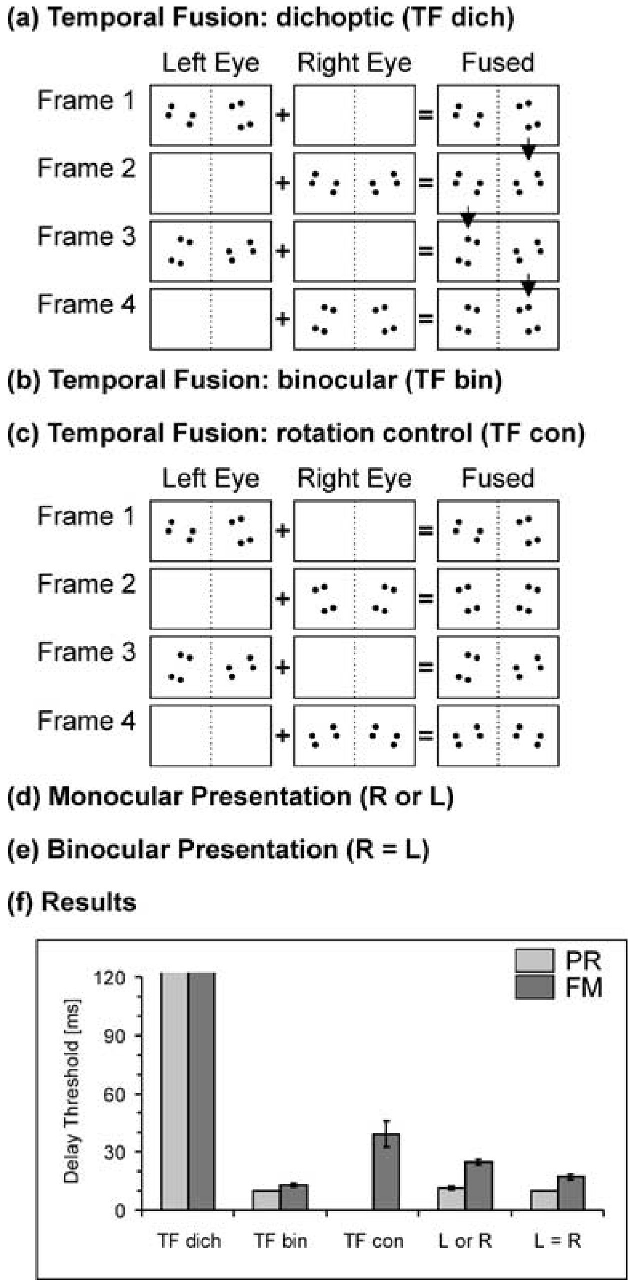


FIG. 5. (a) In this second dichoptic presentation task, the sequence of frames was alternately shown to the two eyes. To obtain the full image sequence needed to perceive the temporal asynchrony between target and surround, the input presented dichoptically to both eyes has to be fused temporally (TF-dich). Displays bound for the left and the right eye were separated by showing red and green dots observed through red–green glasses. (b) In the 'TF-bin' condition subjects observed the same stimulus sequence binocularly, that is, without glasses. (c) For the dichoptic 'TF-con' condition, the sequence shown to the right eye was modified in such a way that the fused sequence presented rotating quadruplets. In this task, subjects had to identify the rotation sense of single quadruplets rather than to identify a target. (d and e) Monocular and binocular controls in which the whole sequence was either presented to the left or right eye (L or R) or, by using yellow dots, to both eyes binocularly (L = R). (f) Threshold delays for all five conditions obtained using both temporal designs, phase reduction and frequency modulation.

the two eyes need to be fused spatially in order to obtain the full sequence of motion.

In this third experiment with dichoptic presentation, each image of the sequence was split into two halves, shown to the right or the left

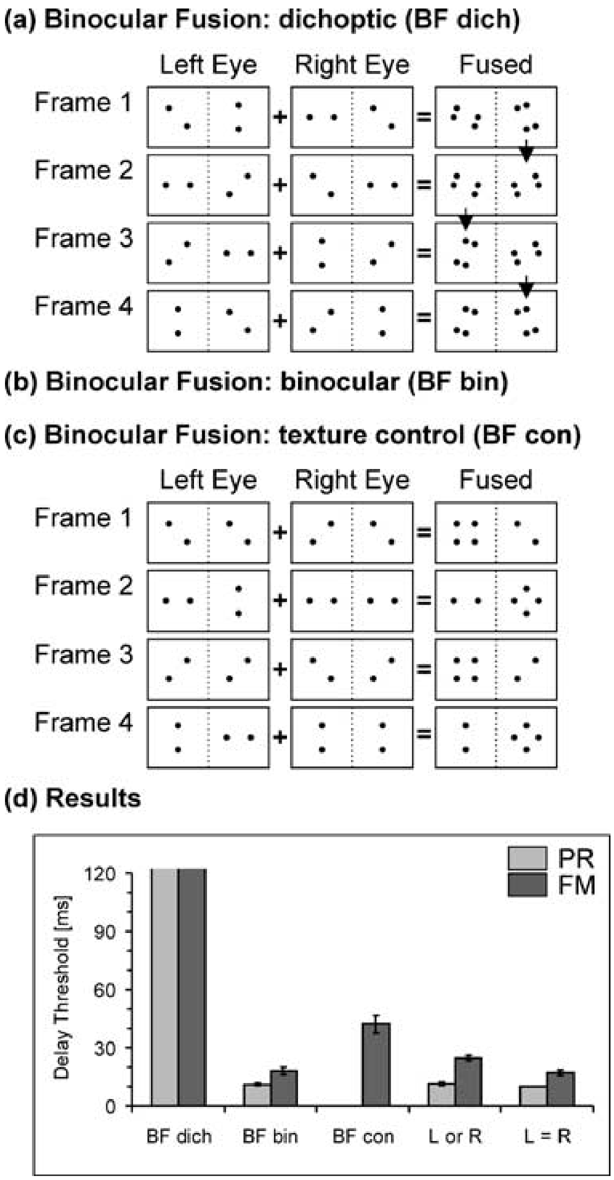


FIG. 6. (a) In this third dichoptic presentation task, the sequence of frames was split in two halves in such a way that each eye saw a continuously rotating colon comprising no hint as to the segregation. Hence, subjects had to binocularly fuse the dichoptic images in order to detect the asymmetry (BF-dich). Again displays bound for the left and the right eye were separated by showing red and green dots observed through red–green glasses. (Note that rotation directions of the colons were randomized across the stimulation field and hence rotation direction could not subserve segregation.) (b) In the 'BF-bin' control, subjects observed the same stimulus sequence binocularly, that is, without glasses. (c) In the dichoptic texture control task (BF-con), the sequence for the right eye was shifted by one frame. In each of the fused images there was a texture difference between the elements in target and ground: in one area monocular colons fell on corresponding positions, hence fusing to a colon, whereas in the other area they added up to a form a quadruplet. Textures in target and ground interchanged between frames. Subjects had to identify the target position. Note that the segregation in this task was cued by texture differences to be seen in each frame, not by motion comprised in between frames. In the phase reduction task the target would be visible in both the short and the long intervals, hence reducing the phase would not make the task more difficult. Therefore, only the frequency modulation task was used here. (d) Thresholds obtained in these three tasks for both phase reduction (PR, light bars) and frequency modulation (FM, dark bars), along with results for the monocular and binocular control conditions obtained in experiment 3.

eye, respectively (Fig. 6a). Both the left and the right eye's image consisted of colons rotating by 45° from frame to frame. Sense of rotation was randomised for each colon in the right eye while the partner colon in the left eye was rotated in the opposite direction. Hence only the binocularly fused (BF) images revealed cues as to the segregation of the display ('BF-dich', Fig. 6a) while both monocular images presented motion in both figure and surround at each transition between frames. As in experiment 4, separation of the images was achieved by means of presenting red and green dots on the monitor and having subjects wear red–green goggles. Also as before, subjects repeated the critical condition under binocular stimulation (BF-bin; Fig. 6b) and underwent a control task, in which the left eye's sequence was kept as before while the right eye's sequence was shifted forward by one frame, resulting in a texture difference between figure and ground which reversed its sign between frames (BF con; cf. Fig. 6c). In this last stimulus, in contrast to all other stimuli presented in this study, the segregation cue was visible within every single frame. Here, the PR design did not lead to valid thresholds because not only the shortened but also the prolonged images could be used to detect the target. Hence, as in the control condition of experiment 4, only the FM design was used in this experiment.

Results (Fig. 6d) were similar to the ones obtained in experiment 4. Performance was at chance level ($25.2\% \pm 1.6 \approx 25\%$) in the dichoptic task (BF-dich) even when the presentation frequency was only 2 Hz (≈ 240 ms/frame). In contrast, in the dichoptic control task (BF con) the display could be segregated up to a frequency of 12.5 Hz (≈ 40 ms/frame). Again, thresholds obtained under binocular stimulation (BF-bin) but with alternating red and green images were as low as those resulting from binocular images without colour changes ($L = R$) and slightly better than those found with monocular stimulation (L or R). Again, a segregation stimulus created on a binocular level, a so-called cyclopean stimulus, failed to allow time-based figure–ground segregation, as was the case with the more straightforward stimulus used in the preceding experiment.

Discussion

Segregation in motion stimuli defined by luminance, colour or contrast modulations

In experiment 1, subjects segregated the time-based figure from its ground for delays ≈ 10 and 20 ms depending on the type of task. These thresholds are slightly better than those reported earlier (Kandil & Fahle, 2001), and are similar to the temporal resolution for texture segregation tasks (Motoyoshi & Nishida, 2001). For these low thresholds, stimulation has to be binocular and luminance contrast of the stimuli must be high (cf. 'Achr.' in Fig. 2b and c and '$L = R$' in Figs 5f and 6d). With decreasing luminance contrast, performance decreased monotonically, that is, temporal threshold increased and segregation failed completely below 6% contrast (Fig. 2b and c).

The mechanism was able to segregate purely colour-defined stimulus elements on the basis of temporal cues, as became obvious under isoluminant conditions. The luminance and the colour signal could be combined: adding luminance contrast to isoluminant stimuli improved thresholds (Fig. 2c). However, the layout of this experiment did not allow conclusions about whether this combination of signals took place (i) before or within the detectors themselves, thus improving primarily the detection of motion, or (ii) on the level of the segregation mechanism, hence improving the segregation. In any case, segmentation is possible for isoluminant stimuli which poorly stimulate first-order motion detectors (see Dobkins & Albright, 1998, for a review).

In experiment 2, the luminance-defined dots were replaced by contrast-defined luminance-balanced differences of Gaussians (DOGs) and hence first-order motion mechanisms with reasonably coarse receptive fields were excluded from detecting the local motion, reserving this task for second-order mechanisms (see for review Lu & Sperling, 1995; cf., however, Cavanagh & Mather, 1989). If segregation is possible under these conditions, indicating that the second stage postulated in the Introduction can segregate figure from ground, this stage must have access to the output of second-order motion detectors which, unlike first-order mechanisms, detect contrast-defined stimuli. Our results show that segregation is indeed successful for delays >13 ms for the PR task and >22 ms for the FM tasks, respectively. As with the achromatic stimuli in experiment 1, the segregation success for this task depended on the modulation contrast of the DOGs. Both PR and FM thresholds for these contrast- and luminance-defined stimuli were highly similar across the various contrast values (Figs 2b and c, and 3b). These findings are consistent with the notion that first- and second-order motion detectors have the same contrast-dependent temporal confinements (Lu & Sperling, 1995). Obviously, the neuronal mechanisms subserving time-based figure–ground segregation have access not only to luminance-defined stimuli but also to purely colour- (wavelength-) and contrast-defined stimuli.

Segregation in monocular vs. binocular motion stimuli

The last three experiments tested whether the segregation mechanism has access to the output of either monocular or binocular mechanisms, or to both. Segregation thresholds obtained under binocular viewing conditions are better than monocular ones, indicating higher temporal resolution for the binocular case. [Compare results for the binocular ($L = R$) vs. monocular (L or R) conditions in Figs 4e and 5f].

Assuming a simple three-level model consisting of a motion or flicker detector at the front end, a segregation mechanism in the middle and a decision stage at the upper end, better performance following binocular stimulation could result from at least one of three reasons. (i) The front-end mechanism detecting the flips of the colons is binocular and benefits directly from a higher signal-to-noise ratio in the motion signal under binocular conditions. (ii) The front-end detectors are monocular but the segregation mechanism is binocular and benefits from a higher signal-to-noise ratio in the signals provided by the monocular detectors. (iii) Both detectors and segregation mechanisms are monocular, find a putative figure at one location and report it to the decision stage. There, the proposals obtained independently reassure each other, thus increasing the probability for the correct solution.

The aim of experiments 3–5 was to distinguish between possibility (i) on one hand and (ii) and (iii) on the other hand, with experiment 3 testing whether or not the segregation mechanism can use information from monocular detectors at all.

In the critical condition ($L \times R$; cf. Figure 4d), monocular detectors can register the apparent local motion while mechanisms located after the binocular fusion see static stimuli and are hence deprived of any significant information for time-based figure–ground segregation. In this condition, subjects reported seeing a 'somehow constant' (fused images) and 'somehow slowly flickering' (monocular images) pattern of quadruplets but were unable to be more precise. Within this pattern they perceived a target as clearly standing out from the ground. Psychophysical performance was as good as under monocular presentation, demonstrating that the segregation mechanism has access to the output of monocular luminance (first-order)-based detectors and does not operate based primarily on the combined (cyclopean) input of both eyes. These findings are consistent with the conclusion of Solomon & Morgan (1999) that mechanisms computing motion from flicker have access to monocular flicker information, although attention cannot be directed to monocular levels of visual processing.

The following two experiments tested whether the segregation stage can use binocular motion mechanisms, similar to the 'third-order' motion detector proposed by Lu & Sperling (1995). In the fourth experiment, we presented successive images alternatingly to the right and left eye (Fig. 5a). Under these conditions quadruplets in both figure and ground flip simultaneously for monocular detectors, thereby eliminating monocular segregation cues. Only after binocular fusion did the temporal delay between motion in figure and ground become detectable. In this task no subject performed better than chance (Fig. 5f). This result is taken as evidence that the segregation mechanism has no access to the output of binocular detectors. The result of the control condition TF-con (Fig. 5c and f) ensures that subjects can temporally fuse the images and detect local motion in a spatially identical stimulus up to a flip frequency of 12.5 Hz.

In this experiment red–green anaglyphs were used to separate the inputs to the two eyes. This implies changes of stimulus colour between stimulus frames from red to green or vice versa, an artefact which might prevent the segregation mechanism from detecting the positional changes in figure and ground although it does not prevent the binocular motion mechanism itself from detecting the rotation sense in the control task TF-con. However, binocularly presenting the sequences with alternating colours (TF-bin in Fig. 5b and f) had no negative effect on threshold delays as compared to binocularly presenting images with constant colours (L = R in Fig. 5e and f), indicating that at least the segregation mechanism operating on the output of the first-order detector can generalize across colour changes.

There is another possible explanation for the lack of segregation with these binocular signals: the segregation mechanism reading the output of the first-order motion detectors receives a strong on- and off-flicker signal in the entire stimulus as a result of the stimulation procedure. This monocular signal might override or attenuate the signal stemming from the binocular motion detectors which perceive the flips. However, the results of the next experiment render this explanation improbable.

In the fifth experiment (cf. Fig. 6), half of the dots are presented to either eye. Similarly to the fourth experiment, the visual system has to fuse the right- and the left-eye image first, before appropriate detectors can perceive the difference in motion between figure and ground. In contrast to the last experiment, however, monocular motion mechanisms see continuously rotating colons rather than luminance flicker between frames. Hence, the segregation mechanism should not be strongly disturbed by the monocular signal in this task. However, much like in the dichoptic condition of the previous experiment, subjects were unable to perceive the segregation (BF-dich in Fig. 6d). Taken together, these results render highly improbable the above hypothesis of monocular flicker signals masking the binocular motion signals at the stage of the segregation mechanism.

A binocular control task ensured that chromatically heterogeneous images activate the segregation mechanism itself if only the first-stage detector is monocular (BF-bin in Fig. 6d). Performance is similar to that found under binocular stimulation with achromatic images. Moreover, in the dichoptic texture-differentiation task (BF con in Fig. 6d) which served as a control, subjects readily identified a figure defined by texture differences of the fused image up to an alternation frequency of 12.5 Hz.

The results of experiments 5 and 6 leave two possible interpretations. Either there is no temporal segregation mechanism reading the output of the binocular detector at all or, if there is one, it is extremely sensitive to colour changes. In pilot studies for the fourth experiment, we used LCD-shutter goggles instead of the red–green anaglyphs. LCD-shutter goggles allow the presentation of dots in the same colour but they suppress the blocked stimulus even less perfectly than the ferro-electric shutter goggles eventually used in experiment 3, allowing a certain amount of crosstalk between the eyes which is precarious for the interpretation of dichoptic experiments. However, even with these goggles we obtained the same result, namely that segregation of the dichoptically presented stimuli was impossible, disproving the interpretation based on a masking by colour effects. Hence, we conclude that the segregation mechanism has no access to the output of dichoptic or binocular motion detectors.

Segregation by a (higher-order) motion detector?

As mentioned in the introduction, Fahle (1993) and Ramachandran & Rogers-Ramachandran (1991) and Rogers-Ramachandran & Ramachandran (1998) discussed the possibility that motion detectors achieved the segregation in their displays. Because luminance changes at different times in the two areas (hemifields or figure and ground), a standard first-order motion detector detecting delays between luminance changes might be activated at the border between the two fields. In the case of asynchronously flipping colons, there is no global luminance change between figure and ground between frames and, besides, local motion detectors will be strongly activated through the rotation of the colons. Hence, first-order motion detectors would not be expected to produce pronounced signals indicating the boundary between the areas in our experiments. Instead, we suppose that first-order detectors perceive the local changes and report these to the segregation mechanism, which may be a higher-order motion detector, i.e. a 'motion-from-motion' detector (Zanker, 1990, 1993) or a purely time-based mechanism. However, motion-from-motion detectors have generally been described as having a temporal resolution of 4–8 Hz (Lu & Sperling, 1995) or only 1 Hz (Zanker, 1996), whereas the temporal resolution of the segregation mechanism tested here is ≈20–25 Hz. Besides, no subject ever mentioned having solved the task on the basis of a subjective motion at the borders between the areas, favouring instead a purely time-based mechanism.

'Third-order' motion

Sperling and coworkers (Lu, Lesmes, & Sperling, 1999; Lu & Sperling, 2001) argue that chromatic isoluminant motion is detected by a 'third-order' motion mechanism, the same mechanism responsible for the detection of dichoptic motion. This mechanism should produce similar segregation thresholds for isoluminant and for dichoptically presented stimuli. However, results differ greatly between the isoluminant and the dichoptic tasks. Segregation of isoluminant stimuli was possible up to threshold frequencies of 8.5 Hz (corresponding to threshold phase delays of ≈60 ms in Fig. 2c). This result is consistent with the thresholds of 8–12 Hz for (isoluminant) 'third-order' motion (Lu & Sperling, 2001). However, experiments 4 and 5 suggest that the segregation mechanism cannot read out motion detectors sensitive to dichoptically presented motion. Hence, our results cannot be explained in a straightforward way by a mechanism based on third order motion detectors of the Lu and Sperling type.

Conclusions

A figure can be segregated from its surround on the basis of temporal delays between positional changes (flips) in figure and ground. We hypothesize that these flips are detected by a set of local detectors and that the segregation is accomplished by a second-stage reading out and comparing globally, with high temporal resolution, the transient activation of the local detectors. In this study we found that the segregation mechanism has access to monocular detectors sensitive to (i) luminance-defined, (ii) contrast-defined and (iii) isoluminant colour-defined stimuli, albeit the latter yield a lower temporal resolution. In contrast, the segregation mechanism does not seem to have access to binocular motion detectors as tested with dichoptically presented stimuli. Hence, to explain the small but significant benefit from binocular over monocular stimulation, we assume that at least one of the remaining stages (segregation mechanism and decision-taking level) has to be binocularly activated.

On the basis of the present results, it is highly unlikely that the time-based segregation is achieved by a higher-order (motion-from-motion) mechanism; the results favour a purely temporal mechanism.

Acknowledgements

We would like to thank Michael Morgan for lending us the ferro-electric goggles, Marc Repnow for enabling our set-up to run the goggles and for additional help with the program code, and Derek Arnold for his comments on an earlier draft of the manuscript. The anonymous referees greatly contributed to clarifying our message. Supported by a grant from the German Research Council (Sonderforschungsbereich SFB 517). QuickTimeMovies can be requested from the Corresponding author.

Abbreviations

Achr, achromatic; BF, binocular fusion; -bin, binocular; -con, control; -dich, dichoptic; DOGs, differences of Gaussians; FM, frequency modulation; PR, phase reduction; TF, temporal fusion.

References

Adelson, E.H. & Farid, H. (1999) Filtering reveals form in temporally structured displays. *Science*, **286**, 2231a.

Blake, R. & Lee, S.H. (1999) Response to: Filtering reveals form in temporally structured displays. *Science*, **286**, 2231a–2232a.

Cavanagh, P. & Mather, G. (1989) Motion: the long and short of it. *Spat. Vis.*, **4**, 103–129.

Dobkins, K.R. & Albright, T.D. (1998) The influence of chromatic information on visual motion processing in the primate visual system. In Watanabe, T. (ed.), *High-Level Motion Processing*. MIT Press, Cambridge. MA, pp. 53–94.

Fahle, M. (1993) Figure–ground discrimination from temporal information. *Proc. R. Soc. Lond. B Biol. Sci.*, **254**, 199–203.

Fahle, M. & Koch, C. (1995) Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Res.*, **35**, 491–494.

Forte, J., Hogben, J.H. & Ross, J. (1999) Spatial limitations of temporal segmentation. *Vision Res.*, **39**, 4052–4061.

Kandil, F.I. & Fahle, M. (2001) Purely temporal figure-ground segregation. *Eur. J. Neurosci.*, **13**, 2004–2008.

Lee, S.H. & Blake, R. (1999) Visual form created solely from temporal structure. *Science*, **284**, 1165–1168.

Leonards, U., Singer, W. & Fahle, M. (1996) The influence of temporal phase differences on texture segmentation. *Vision Res.*, **36**, 2689–2697.

Lu, Z.L., Lesmes, L.A. & Sperling, G. (1999) The mechanism of isoluminant chromatic motion perception. *Proc. Natl Acad. Sci. USA*, **96**, 8289–8294.

Lu, Z.L. & Sperling, G. (1995) The functional architecture of human visual motion perception. *Vision Res.*, **35**, 2697–2722.

Lu, Z.L. & Sperling, G. (2001) Three-systems theory of human visual motion perception: review and update. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, **18**, 2331–2370.

Motoyoshi, I. & Nishida, S. (2001) Temporal resolution of orientation-based texture segregation. *Vision Res.*, **41**, 2089–2105.

Ramachandran, V.S. & Rogers-Ramachandran, D.C. (1991) Phantom contours: a new class of visual patterns that selectively activates the magnocellular pathway in man. *Bull. Psychonomic Soc.*, **29**, 391–394.

Rogers-Ramachandran, D.C. & Ramachandran, V.S. (1998) Psychophysical evidence for boundary and surface systems in human vision. *Vision Res.*, **38**, 71–77.

Solomon, J.A. & Morgan, M.J. (1999) Dichoptically cancelled motion. *Vision Res.*, **39**, 2293–2297.

Usher, M. & Donnelly, N. (1998) Visual synchrony affects binding and segmentation in perception. *Nature*, **394**, 179–182.

Zanker, J.M. (1990) Theta motion: a new psychophysical paradigm indicating two levels of visual motion perception. *Naturwissenschaften*, **77**, 243–246.

Zanker, J.M. (1993) Theta motion: a paradoxical stimulus to explore higher order motion extraction. *Vision Res.*, **33**, 553–569.

Zanker, J.M. (1996) On the elementary mechanism underlying secondary motion processing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **351**, 1725–1736.