Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 44 (2004) 3177–3182

Vision Research

www.elsevier.com/locate/visres

Figure–ground segregation can rely on differences in motion direction

Farid I. Kandil *, Manfred Fahle

Human Neurobiology, University of Bremen, Argonnenstraße 3, 28211 Bremen, Germany

Received 9 December 2003; received in revised form 18 May 2004

Abstract

If the elements within a figure move synchronously while those in the surround move at a different time, the figure is easily segregated from the surround and thus perceived. Lee and Blake (1999) [Visual form created solely from temporal structure. *Science, 284*, 1165–1168] demonstrated that this figure–ground separation may be based not only on time differences between motion *onsets*, but also on the differences between *reversals* of motion direction. However, Farid and Adelson (2001) [Synchrony does not promote grouping in temporally structured displays. *Nature Neuroscience, 4,* 875–876] argued that figure–ground segregation in the motion-reversal experiment might have been based on a contrast artefact and concluded that (a)synchrony as such was 'not responsible for the perception of form in these or earlier displays'.

Here, we present experiments that avoid contrast artefacts but still produce figure–ground segregation based on purely temporal cues. Our results show that subjects can segregate figure from ground even though being unable to use motion reversals as such. Subjects detect the figure when either (i) motion stops (leading to contrast artefacts), or (ii) motion directions differ between figure and ground. Segregation requires minimum delays of about 15 ms. We argue that whatever the underlying cues and mechanisms, a second stage beyond motion detection is required to globally compare the outputs of local motion detectors and to segregate figure from ground. Since analogous changes take place in both figure and ground in rapid succession, this second stage has to detect the asynchrony with high temporal precision.

Keywords: Temporal asynchrony; Figure-ground segregation; Segmentation; Motion differences; Spatio-temporal integeration; Psychophysics

1. Introduction

One of the Gestalt laws of visual perception states that a figure stands out from its surround if the elements within the figure change according to a 'common fate'. There has been a heated debate whether figure and ground can be segregated based solely on temporal cues, that is, if changes occur asynchronously between figure and ground (uncommon fate), while synchronously within each of them (common fate) (Adelson & Farid, 1999; Blake & Lee, 1999; Fahle, 1993; Fahle & Koch, 1995; Farid & Adelson, 2001; Kandil & Fahle, 2001, 2003; Kiper, Gegenfurtner, & Movshon, 1996; Lee & Blake, 1999; Leonards, Singer, & Fahle, 1996; Usher & Donnelly, 1998). The underlying question is whether specialised detectors for minute time differences exist in the human visual system or else other types of detectors, such as motion detectors, subserve segregation. It has even been speculated that the fine temporal structure of stimuli may be directly linked to the (a)synchronous firing of neuronal assemblies leading to a binding together of those neurons firing together (see Gray, 1999; Hebb, 1949; Singer, 1999).

Lee and Blake (1999) proposed that an asynchrony in the order of 10 ms between motion reversals in figure and ground could subserve segregation. They presented displays filled with randomly oriented line elements

---

* Corresponding author. Address: Department of Psychology II, WW University of Münster, Fliednerstraße 21, 48149 Münster, Germany. Tel.: +49 251 8334181; fax: +49 251 8334173.

E-mail address: kandil@uni-muenster.de (F.I. Kandil).

doi:10.1016/j.visres.2004.07.027

moving in random directions. [1] All elements in the figure reversed their motion direction after random multiples of 10ms. All background elements also reversed direction simultaneously after random multiples of 10ms, independently from those in the figure. Lee and Blake argued that the target was visible due to the temporal delays between motion reversals in figure and ground. The shortest intervals were 10ms, with increasingly longer delays occurring with decreasing probability. However, due to the random intervals between motion reversals in figure and ground, elements in the ground sometimes move in the same direction for several frames while those in the figure undergo successive reversals and hence 'jitter' in place (or vice versa). Stimuli moving forward produce different contrasts when filtered with a temporal low- or band-pass filter—as they are present in the human visual system—than those jittering in place. Hence, in these displays the visual system may segregate figure from ground based on contrast artefacts as opposed to purely temporal cues (cf. Adelson & Farid, 1999; Blake & Lee, 1999; Farid & Adelson, 2001).

We report here the results of four new experiments testing the contribution of three cues suspected to separate figure from ground based on temporal cues: (1) time differences between *reversals* of motion direction (Lee & Blake, 1999), (2) time differences between *contrast* artefacts emerging from individual reversals (Adelson & Farid, 1999) and (3) differences in motion *direction*—caused by the reversals occurring at different times. The contrast artefact resulting from individual motion reversals did not allow figure–ground segregation in our subjects, and neither did temporal delays between motion reversals. Only differences in motion direction produced a clear discrimination between figure and ground, demonstrating the high temporal accuracy of the neuronal mechanisms separating figure from ground.

2. General methods

Displays presented 20 × 20 bright dots (20 cd/m^2) moving forwards and backwards on horizontal, non-overlapping trajectories on a dark (0.03 cd/m^2) monitor background (Fig. 1a). The time required for motion from the start of the trajectory to its turning point is defined as the period. The dots within the figure precede or lag behind those in the ground by a defined delay. Sufficiently large delays allowed observers to detect the position of the figure in the display.

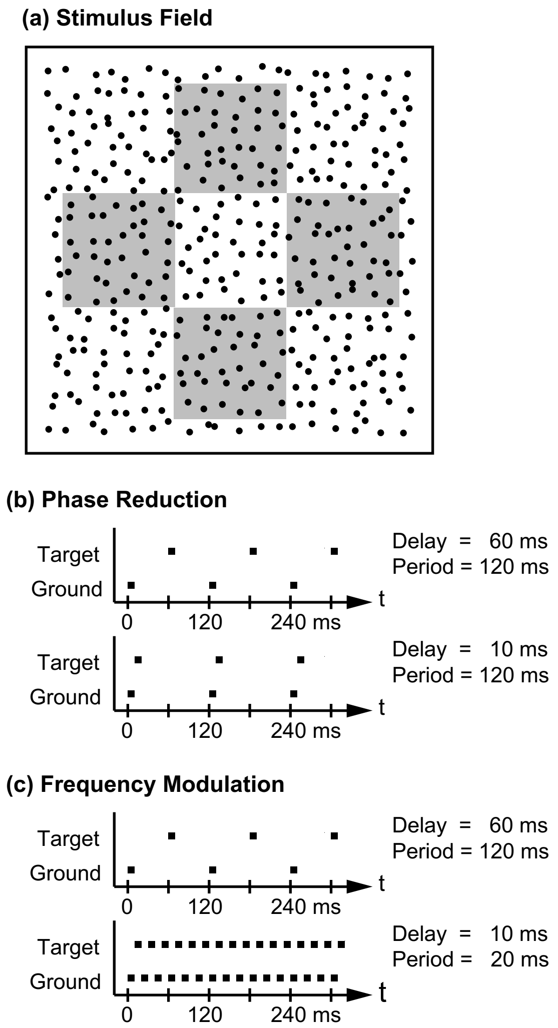

Fig. 1. (a) Contrast-reversed reproduction of one stimulus frame. Displays present 20 × 20 dots moving locally back and forth on parallel horizontal trajectories (see Fig. 2a). The 36 dots that constitute the square target move asynchronously from the rest of the dots. Subjects had to localise the target at one of four fixed positions (four-alternative forced-choice task). The four possible locations are depicted by grey shades here, which were not present in the original display. (b) In the phase reduction design the time a dot moves between the endpoints of its trajectory was constant at 120ms, while the phase delay was reduced from 60ms (counter phase) in steps of 10ms down to 10ms across conditions. The first and the last condition are indicated here. (c) In the frequency modulation design the time a dot moves between the endpoints of its trajectory was reduced from 120ms in the first condition with the phase delay always at counter phase.

We used two designs to determine threshold delays. In the phase reduction (PR) design (Fig. 1b), the period had a constant length of 120ms, while the phase delay between movement in figure versus ground was reduced stepwise from 60 to 10ms. In the frequency modulation (FM) design (Fig. 1c), the period length was reduced from 120 to 20ms (in steps of 20ms) while the delay was constant at anti-phase (180°), corresponding to phase delays between 60 (for 120ms periods) and 10ms (for 20ms periods). Each step of each design

---

[1] The stimuli consisted of Gabor patches with carriers moving perpendicular to the long axis of the Gabor elements.

contained 16 stimulus presentations of 1.5 s duration, using the method of constant stimuli. Subjects had to indicate, in a four-alternative forced-choice task, at which of four positions the figure was displayed based on the temporal delays between dot movements in figure versus surround (cf. the grey squares in Fig. 1a). Thresholds were calculated separately for both designs, PR and FM, by determining (through interpolation of the probit function) the minimum delay required to obtain 62.5% correct answers, as this level is midway between chance and perfect performance.

Stimuli were presented on a 21″ 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. From a viewing distance of 40 cm, the stimulus displays measured 40 × 40°. The mean distance between two dots was 40 pixels (120′) and the length of the trajectories was 24 pixels (72′) in all conditions. A single dot was 2 × 2 pixels large, corresponding to 6′. Since the minimum position change was also 2 pixels and anti-aliasing was deactivated no pixel was intensified on two successive frames in order to avoid contrast artefacts. Starting points of dot-trajectories were randomly jittered horizontally and vertically by up to 12 pixels relative to a regular grid to diminish positional cues in single frames.

Six subjects participated in this study. All had normal or corrected to normal visual acuity and five of them were naive as to the purpose of the study. They were seated in a dimly lit room.

3. Experiment 1

The first experiment serves as a baseline task. Dots in figure and ground moved back and forth on parallel straight trajectories. All started with the same initial horizontal motion direction (Fig. 2a). The dots in the target differed from those in the ground only in a phase delay, that is, they both started and returned earlier than their counterparts in the ground. Here, figure–ground segregation can—at least in principle—rely on all three features present: (1) motion directions *reverse* at different points in time, (2) thus producing *contrast* artefacts at different times and (3) elements move in different *directions*, at least during part of each period of the presentation (see Fig. 2a). Due to the delay between movements in figure and ground, motion directions in the figure reverse before the ground, yielding different motion directions until the elements in the ground reverse direction, too.

The results are shown in Fig. 2b. Dots represent the percentage of correct answers obtained with phase reduction (PR, left half) and frequency modulation (FM, right half). In both conditions, subjects reliably perceived the targets for phase delays longer than 20 ms, and threshold delays of about 15 ms can be calculated.

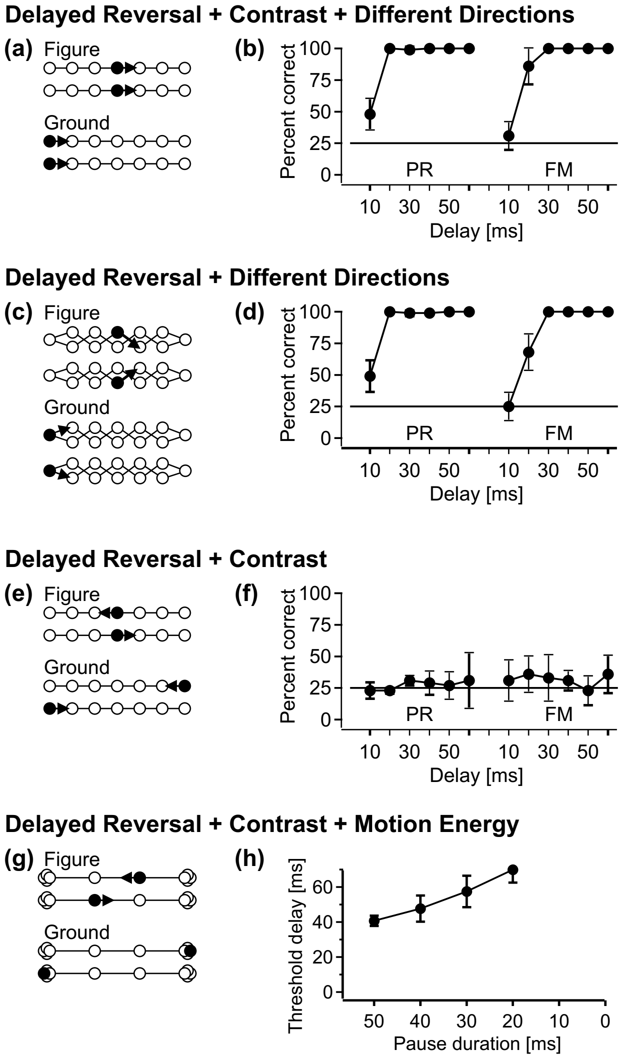

Fig. 2. (a, c, e, g) Stations of the trajectories (○), actual positions (●) and motion directions (→) of two dots from within the figure and two from the ground. In the frequency modulation (FM) condition shown here, dots need 6 frames with 10 ms each from one end of the trajectory to the other, an interval of 60 ms ('the period'). Dots in the figure lag behind those in the ground by 3 frames (phase delay of 30 ms). (c) The dots move on zigzag rather than on straight trajectories in the second experiment. (e) Initial directions of motion are randomised in the third experiment that otherwise corresponds to the first experiment. (g) In the fourth experiment, period length and phase delay stay constant. The number of steps (circles) on the trajectory is reduced, while dots pause at the ends of the trajectories, as indicated by the multiple circles at both ends of the route. Here, the pauses each last for 3 frames (30 ms). (b, d, f) Means and standard errors in percent correct answers across six subjects for a phase reduction task (PR, left side) as well as a frequency modulation task (FM, right half). Grey lines indicate the chance level of 25%. Observers are unable to detect the figure on the basis of contrast plus delayed reversal, but require differences in motion direction. (h) Means and standard errors for the fourth experiment. Threshold delays for the frequency modulation task (FM) are plotted as a function of pause duration.

These very low threshold delays correspond well with data reported for other time-based figure–ground segregation tasks (Fahle, 1993; Kandil & Fahle, 2001, 2003; Usher & Donnelly, 1998). However, since all three features are present, one cannot differentiate between their individual impacts on performance.

4. Experiment 2

In the second experiment, the dots use different (zigzag) trajectories for their way forwards versus backwards (Fig. 2c), that is dots were displaced by 2 pixels alternatively down- or upwards relative to a straight line through the endpoints. Due to the zigzag trajectories, the dots do not return, at the time of reversal, to points they had just occupied before, as was the case in the first experiment. In any system that integrates over time—as does the visual system—it is intuitively clear that moving objects are smeared due to motion smear and hence have lower contrast than stationary ones. Consequently, if one considers two objects over a short time, an object reversing its direction and returning in the opposite direction on exactly the same path covers only half the number of different spatial positions (pixels) compared to an object that does not reverse but moves on. Therefore, the reversing object is less smeared and appears to have a higher contrast than its linearly moving counterpart (see Adelson & Farid, 1999, for a formal argument). The zigzag trajectory should eliminate a potential contrast artefact for small receptive fields, while neurons with larger receptive fields would be minimally sensitive to contrast changes occurring in a small region of their field. Phase delay and identical initial motion direction remain as in the first experiment. Hence, temporal differences between the reversals (feature 1) as well as the opposing motion directions (feature 3) can serve as segregation cues in this second experiment.

As shown in Fig. 2d, overall performance and thresholds (15 ms) for both designs (PR and FM) are comparable to the ones obtained in the first experiment.

Hence, if other cues are present, the influence of contrast artefacts induced by individual motion reversals seems to be negligible. Eliminating, or at least decreasing the size of the contrast artefacts by zigzag trajectories, did not influence performance in the presence of the other cues. The next experiment further clarifies the (non-existent) role of contrast artefacts based on individual reversals.

5. Experiment 3

The third experiment differs from the *first* in that initial motion direction is randomised for all individual dots of the display (Fig. 2e), thus eliminating information for all mechanisms based on motion direction (feature 3) while providing the same amount of information for detectors sensitive to delays between motion reversals (feature 1) and for contrast differences (feature 2) as in Experiment 1.

The methods differed from those used in Experiment 1 only in that here the initial motion direction of the dots was randomised. Hence, at each point in time about the same number of dots moved left and right in both figure and ground.

The results (Fig. 2f) confirm that performance was no better than chance for all phase delays and both temporal designs, phase reduction (PR) and frequency modulation (FM)—observers were unable to detect the figure even at the largest delays.

Segregation in the stimuli used in Experiments 1 and 2 cannot rely on motion reversals as such—a result in direct contradiction to the interpretation of Lee and Blake's experiments (Lee & Blake, 1999). An important difference between our experiment and that of Blake and Lee certainly is the constant order of stimuli here with clearly segregated motion reversals while the more complex ones used by these authors contain some clustered reversals. Moreover, our subjects probably were less well trained with this specific task than Lee and Blake's. Finally, the result of this experiment demonstrates that a single motion reversal does not produce contrast artefacts sufficiently salient for segregation in untrained observers. This finding strongly restricts the generality of the explanation for temporal segregation proposed by Adelson and Farid (1999; Farid & Adelson, 2001): single motion reversals do not generally produce a contrast artefact sufficient for figure ground segregation.

6. Experiment 4

Adelson and Farid's (1999) critique on Lee and Blake's experiments was based mainly on a possible contrast artefact caused by motion reversals. This contrast artefact would render the corresponding part of the stimulus, be it figure or ground, appear in higher contrast and hence would transform the task from the detection of time differences to the detection of differences in luminance or contrast. To further investigate the influence of such a contrast cue, we conducted a fourth experiment to ascertain the degree of contrast differences necessary to support figure–ground segregation. Introducing pauses at the reversal points of the dots in the figure while dots in the background move and vice versa renders these dots stationary, thus removing motion smear and further increasing apparent contrast for any system integrating over time. Hence, linearly moving dots produce lowest contrast, reversing dots yield higher contrasts, and stationary (pausing) dots lead to highest contrasts. Of course, during these pauses another cue allows figure–ground segregation, namely discrimination between stationary versus moving dots, i.e. a separation based on motion energy (Adelson & Bergen, 1985; Burr, Ross, & Morrone, 1986; Fahle & Poggio, 1981; Georgeson & Scott-Samuel, 1999; Watson & Ahumada, 1985).

The fourth experiment was a modification of the third experiment. Motion paused shortly for various lengths (10–50 ms) at the end of each trajectory (Fig. 2g). Only the frequency modulation design was used and only pause durations shorter than the period length could be tested. [2] Period length was kept constant by removing one step of the motion trajectory for each additional pause frame added at the end of the trajectory. Thresholds were determined separately for all pause durations $(0, 10, \ldots, 50\,\text{ms})$.

Fig. 2h shows thresholds as a function of pause duration. Only one of the six subjects was able to perceive the segregation with a pause duration of 10 ms, three subjects at pauses of 20 ms and almost all at pause durations of 30 ms. With increasing pause durations, threshold delays of pause onsets required for figure ground segregation and hence target detection decreased from near 60 ms to around 40 ms.

Contrast artefacts caused by a single or double reversal (corresponding to pause durations and hence delays of 10 and 20 ms) do not suffice for most subjects to reliably segregate figure from ground, but 40 ms or more are required. Obviously, the contrast artefact proposed by Adelson and Farid cannot play a crucial role in our experiments using regular intervals as many previously reported segregation tasks where much finer temporal resolutions were obtained (Fahle, 1993; Kandil & Fahle, 2001, 2003; Leonards et al., 1996; Usher & Donnelly, 1998).

7. General discussion

The results obtained here confirm that the mechanism segregating figure from ground can rely on a number of low-level cues including time differences in both contrast artefacts and especially differing motion direction, but not on individual motion reversals per se, at least under the conditions employed here.

The fixed rather than random delays between subsequent motion reversals as employed in this study allow disentangling the different possible cues for figure–ground segregation present in earlier studies (Lee & Blake, 1999) and allow precise temporal threshold estimations. Since the stimuli are presented repeatedly with a fixed temporal scheme, temporally filtering the stimuli may alter signal amplitudes but cannot prolong the time allowed for the second stage mechanisms to segregate the displays, hence the results should be largely independent of the type of physiologically feasible temporal filter employed, unlike in the earlier studies.

The first cue, time differences between motion reversals, turned out to be unable to subserve figure–ground segregation by itself in our observers (see Fig. 2f). The same is due for possible contrast artefacts caused by motion reversals (again Fig. 2f). Only introducing pauses of 30 ms or longer at the reversals of motion directions allowed subjects to segregate figure from ground. The third cue, differences in motion directions only required phase delays of about 15 ms—thresholds similar to those found in earlier studies (Fahle, 1993; Kandil & Fahle, 2001, 2003; Leonards et al., 1996; Usher & Donnelly, 1998). In the displays used, thresholds for segregation based on differing motion directions are much shorter (15–20 ms) than those for segmentation based on contrast artefacts (35–40 ms) suggesting that the contrast artefact of a single reversal produces a lower signal-to-noise ratio than different motion directions do.

While Adelson and Farid (1999) may be correct in that contrast artefacts contributed to figure–ground segregation in some earlier investigations, their conclusion that (a)synchrony as such was ‘not responsible for the perception of form in these or earlier displays’ (Farid & Adelson, 2001) is clearly unwarranted. Our results demonstrate that even in the absence of detectible contrast cues, segregation is possible on the basis of (a)synchronous signals for temporal delays above 15 ms. The contour between figure and ground is invisible for all generally accepted first-order detectors such as those detecting motion or differences in luminance or colour. Irrespective of the underlying cue, defining local transitions in these displays, a subsequent second stage mechanism is required for figure–ground segregation (e.g. Kandil & Fahle, 2003; Motoyoshi & Nishida, 2001). This mechanism has to read and globally group the incoming low-level signals with high temporal accuracy. In conclusion, short temporal delays can obviously lead to figure–ground segregation in the absence of any contrast cues while individual motion reversals can not.

Acknowledgments

QuickTime demonstrations can be found on the corresponding author’s homepage: www.psy.uni-muenster.de/inst2/lappe/Farid/Farid.html. Supported by a grant from the German Research Council (SFB 517, TP C 10).

---

[2] Hence, for pause durations of 20 and 30 ms, only period lengths between 120 and 40 ms were tested (shortest phase delay was thus 20 ms), and for pause durations of 40 and 50 ms, only periods down to 60 ms were used (with the shortest delay tested being 30 ms). Note that even in the latter case, the threshold of 40 ms is well within the tested range.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A—Optics Image Science and Vision, 2*, 284–299.

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

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

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London Series B—Biological Sciences, 227*, 249–265.

Fahle, M. (1993). Figure–ground discrimination from temporal information. *Proceedings of the Royal Society of London Series B—Biological Sciences, 254*, 199–203.

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

Fahle, M., & Poggio, T. (1981). Visual hyperacuity:spatiotemporal interpolation in human vision. *Proceedings of the Royal Society of London Series B—Biological Sciences, 213*, 451–477.

Farid, H., & Adelson, E. H. (2001). Synchrony does not promote grouping in temporally structured displays. *Nature Neuroscience, 4*, 875–876.

Georgeson, M. A., & Scott-Samuel, N. E. (1999). Motion contrast: a new metric for direction discrimination. *Vision Research, 39*, 4393–4402.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron, 24*, 31–47.

Hebb, D. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.

Kandil, F. I., & Fahle, M. (2001). Purely temporal figure–ground segregation. *European Journal of Neuroscience, 13*, 2004–2008.

Kandil, F. I., & Fahle, M. (2003). Mechanisms of time-based figure–ground segregation. *European Journal of Neuroscience, 18*, 2874–2882.

Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research, 36*, 539–544.

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 Research, 36*, 2689–2697.

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

Singer, W. (1999). Neural synchrony: a versatile code for the definition of relations? *Neuron, 24*, 49–65.

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

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A—Optics Image Science and Vision, 2*, 322–341.