nature
INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
Volume 377 No. 6546 21 September 1995, page 237-239

Attention-generated apparent motion

Zhong-Lin Lu & George Sperling

For further information, write:
Zhong-Lin Lu or George Sperling
Human Information Processing Lab
Department of Cognitive Sciences
and Institute for Mathematical Behavioral Science
University of California
Irvine, CA 92717
http://www.socsci.uci.edu/HIPLab

Attention-generated apparent motion

Zhong-Lin Lu & George Sperling

Human Information Processing Laboratory,
Department of Cognitive Sciences and Institute for
Mathematical Behavioral Sciences, University of California,
Irvine, California 92717, USA

MOTION perception mechanisms have recently been divided into three categories[1]. First-order mechanisms[2–4] primarily extract motion from moving objects or features that differ from the background in luminance. Second-order mechanism[5,6] extract motion from moving properties, such as a moving area of flicker in which there is no difference in mean luminance between target and background. These first- and second-order motion mechanisms are primarily monocular. The existence of purely binocular, interocular and various other unusual kinds of apparent motion[7–13] has promoted conjectures of a third-order mechanism[1,14,15], but there has been no clear suggestion as to the actual computations that such a mechanism might perform. Here we demonstrate 'alternating feature' stimuli that produce apparent motion only when the observer selectively attends to one of the embedded features in the display. The latent motion in the alternating feature stimuli is invisible to first- or second-order motion mechanisms, and the direction of apparent motion depends on the particular feature attended. These findings suggest the mechanism of third-order motion: the locations of the most significant features are registered in a salience map, and motion is computed directly from this map.

How might a motion mechanism based on feature salience operate? At each moment in time, the x, y locations of the most significant features are marked (the salience map). For example, when figure–ground distinctions are meaningful, being marked leads to being labelled 'figure'. Motion is computed by standard algorithms[2–4] from the spatio-temporal changes of the map. It is difficult to prove the existence of such a process because the same stimulus components that produce salient features usually also stimulate first- and second-order motion processes. The demonstration of feature-salience motion requires stimuli that are invisible to first- and second-order motion processes. The critical aspect is that only a feature's salience is marked from frame to frame, not the feature itself, so a feature-salience motion system transmits information only about the location, direction and speed of movement. Information about what is moving; that is, the features themselves, is carried by a pattern-processing system[16,17].

The key to demonstrating feature-salience motion is to use voluntary attention to influence which features are marked in ambiguous displays. It should not matter if the actual features being marked were to change from frame to frame, as long as their locations followed a constant motion trajectory, so we used alternating-feature displays. Frames made of one type of stimulus material were interleaved with frames made of another, entirely different, type of material[13]. All displays have

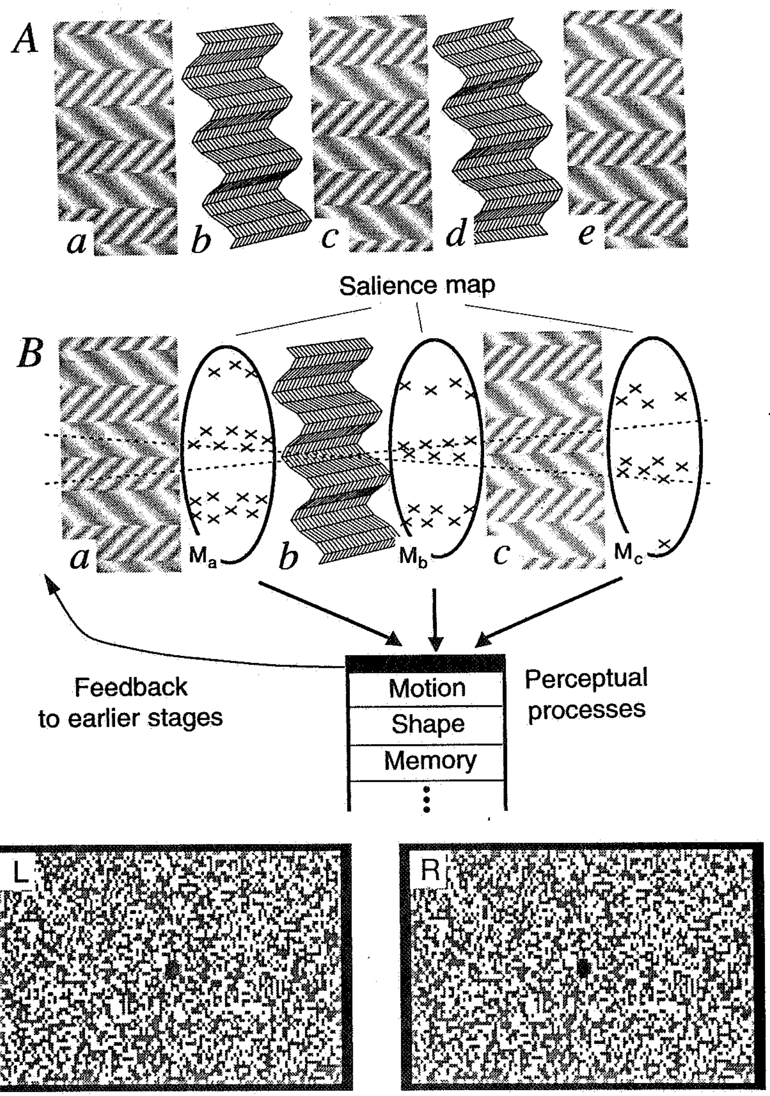

FIG. 1 Stimuli: the depth/texture alternating feature display. The top row (A) shows a sequence of five consecutive frames (a–e); each is displaced by 90 deg from the previous one. The depth stereograms (a, c, e) are indicated schematically; an actual stereogram is shown at the bottom, left (L) and right (R). The second row (B) shows the first three frames (a, b, c) and a feature-salience map (M_a, M_b, M_c) that might be associated with each frame. The most salient features of the depth frames are the near peaks (upper peaks in the panels) marked by crosses in the salience map (M_b). When a subject attends to the coarse grating, the coarse stripes are marked in the salience map (crosses in M_a and M_c). The direction of perceived motion follows the space–time trajectory of the crosses as indicated by the broken line from upper left to lower right. The opposite direction of motion has no support in M_a and M_c; perceiving this upward motion would require attention to the fine stripes. Left (L) and right (R) eye images of a stereogram illustrating the bottom half of one frame of the depth stimulus. Viewing the left panel with the left eye and the right panel with the right eye shows 1.7 (of 3.7) cycles of the grating. The sinusoidal depth variation is quantized to three levels (−1, 0, +1 pixel disparity; each pixel subtends 1.5 min). The overall size of the stimulus as viewed by the subjects was 5.94 × 2.97 deg. The contrast of the texture gratings was 40%. The coarse stripes were sine waves, 2.5 cycles deg^{-1}, and the fine stripes were 5.0 cycles deg^{-1}.

the following properties. (1) Motion is invisible to the first- and second-order systems, and would be visible only to a feature-salience motion system. Perceiving motion would require combining salience information from the two different materials. (2) One stimulus feature perceptually dominates (is more salient) in one of the types of frames. (3) The to-be-attended features are entirely ambiguous (have equal salience) in the other type of frames. (4) Verbal instructions direct the subject to attend only to one of the two kinds of features in the feature-ambiguous frames. (5) Opposite motion directions are perceived depending on which feature is attended (marked as salient) in the feature-ambiguous frames.

Two different kinds of alternating-feature stimuli were created: depth/texture gratings (Fig. 1) and fullwave/halfwave gratings (Fig. 2). In the depth grating, the near-appearing peaks are naturally salient, and the two textures have equal salience. In the fullwave gratings, the high-contrast areas are naturally salient. In the halfwave gratings, after calibration for each subject, the areas of white and of black spots have equal salience.

In both stimulus types, the phase shift between successive frames is 90° (Figs 1 and 2), so successive even-numbered frames in the sequence are separated by 180°, as are successive odd-numbered frames. Motion between frames separated by 180° is totally ambiguous. The perception of coherent motion would require combining information from even and odd frames, which are separated by 90°, the ideal separation for motion detection[18]. For each given alternating-feature motion display, the perceived motion direction depends on which stimulus feature is being attended in the feature-equivalent frames.

Some subjects immediately perceive motion in a direction consistent with the attentional instruction. Others do not. However, after several hundred trials, every subject who participated in the experiment perceived motion in a direction consistent with the attentional instruction. At this point, the instruction was reversed (attend to the previously unattended texture), and trials were continued until performance again stabilized. In no case did this take more than four blocks of 100 trials.

A small proportion of catch trials with unambiguous motion was used to determine whether subjects ever deliberately misreported the direction of apparent motion. No such reports occurred.

Very brief displays (total duration ⩽333 ms) were constructed to exclude possible eye-movement[19] mediation of perceived motion direction. Because the motion computation is the same with and without eye movements, eye-movement mediation would not logically alter any of the arguments of this paper.

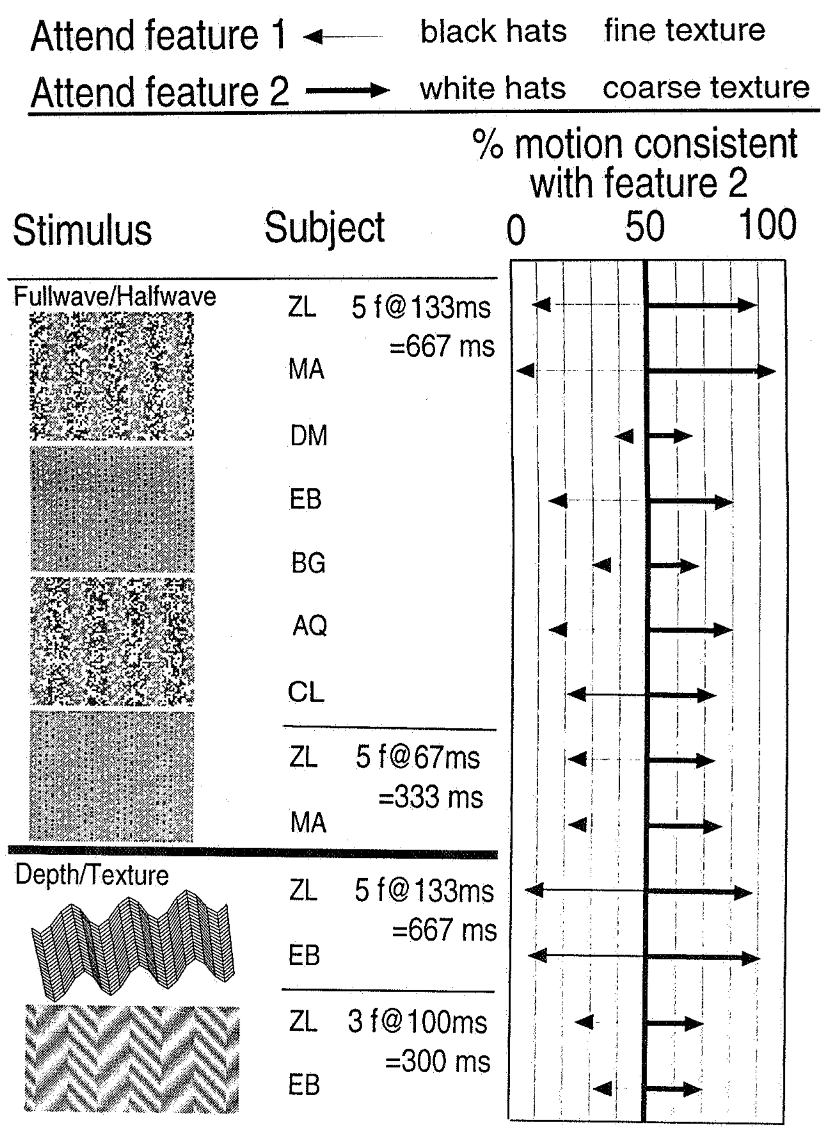

FIG. 2 Procedures and results of the attention experiments. Subjects initiate a trial by pressing a key. A fixation point immediately appears in the centre of the display and remains on throughout the trial; 0.5 s later, a sequence of frames is presented. At the end of each trial, the subject indicates the direction of perceived motion. The notation 5f@133 ms = 667 ms indicates five frames, each of duration 133 ms, with a total display duration of 667 ms. Four frames of the fullwave/halfwave stimuli and two of the depth/texture stimuli are illustrated. Here the term fullwave[28] merely indicates alternating areas of high- and low-contrast random binary noise; halfwave indicates alternating areas of white dots on a dark-grey background and black dots on a light-grey background. Both frames have the same expected luminance everywhere and therefore do not stimulate the first-order (luminance) motion system. Before any attention instructions, the halfwave frames are calibrated individually for each subject to find a contrast ratio of white to black (between 1.00 and 1.18 for our subjects) for which the direction of perceived motion of the displays was completely random. This requires approximately 500 trials, 3 s per trial. This contrast ratio is used for all subsequent observations. Some subjects are then instructed to attend only to the white spots and others to attend only to the black spots. For example, the trajectory from high-contrast areas of fullwave frames (which are automatically salient) to white spots of halfwave frames (made salient by attention) produces perceived motion to the right. Only one type of attentional instruction is given within an experimental session. From trial to trial, the motion direction, the feature type of the starting frame, and the spatial phase of the starting frame are varied randomly. After a subject's performance has stabilized, the attention instructions are reversed, and the procedure is repeated. Two subjects, ZL and EB, were experienced psychophysical observers. The other subjects were naive student observers who were uninformed about the purpose of the experiments. For each subject the left-pointing arrow indicates data obtained from the last 200 trials with the instruction to attend to black spots (in fullwave/halfwave stimuli) and to fine textures (in depth/texture stimuli). A right-pointing arrow indicates data from attention to white spots (top) and to coarse texture (bottom). Initial calibration data, obtained under neutral attention conditions, fall within the black area occupied by the line under 50. Thus the tip-to-tip distance between oppositely pointed arrows indicates the extent to which attention determines the apparent direction of apparent motion in these alternating feature displays.

Indeed, there is very little inclination to make eye movements: the display is 'much too quick'. The best strategy for seeing motion is to maintain very careful fixation and to peer through the display attempting to see the to-be-attended feature.

The results of the last 200 trials are shown in chronological order from top to bottom in Fig. 2. In the depth/texture condition, each subject consistently perceived motion in the direction that resulted from connecting the dominant depth feature (near-appearing peak) to the dominant texture feature (defined by the attentional instruction) in more than 90% of trials (chance is 50%). The same display was perceived as moving up or down, depending on which component texture the subject attended.

Similar results were obtained in the fullwave/halfwave condition. Subjects consistently perceived motion in the direction that resulted from connecting the dominant fullwave feature (high-contrast peak) to the attended halfwave feature (black or white spot as defined by the attentional instruction). The percentage of attention-conforming motion responses varied between subjects from 65–95%. All effects were highly significant statistically. The very brief displays (viewed only by the most consistently performing subjects) reduced attention-determined motion from over 90% to about 70%, which was still far above chance.

An error in calibration of the displays could have produced a consistent signal that might be detected by the first- or second-order motion system. Such an effect would appear as a bias to perceive motion consistent with one of the attentional instructions and not the other (a shortening or lengthening of the left-pointing arrows relative to the right-pointing arrows in Fig. 2). The near-perfect symmetry of the data verifies the absence of significant first- or second-order motion contamination.

The feature-salience motion mechanism is affected by both top-down attentional processes and by automatic bottom-up processes[20,21]. The large role of bottom-up inputs was demonstrated in some preliminary experiments. In one experiment, five frames were composed of a depth grating (near/far) alternating with the fullwave grating (contrast high/low). When subjects attend to high-contrast areas, they perceived motion in the direction consistent with movement from near depth to high contrast on 87–93% of trials. When they were asked to attend to low-contrast areas, no consistent motion direction was perceived. Attention could not overcome the large inherent asymmetries of feature saliency in the fullwave stimuli.

It is instructive to consider our results in relation to two quite different lines of research.

(1) When a colour and an achromatic grating are optically superimposed and moved in opposite directions, observers perceive only the motion of the achromatic grating. However, when the observers attentionally track a colour stripe, only the motion of the chromatic grating is perceived[22]. Cavanagh[22] argued that this phenomenon is an instance of attentionally based motion perception, rather than merely an instance of attentionally selecting the output of one of two elementary motion computations, achromatic and chromatic. Just because the procedure requires conscious tracking to produce the perception of motion does not mean that tracking is the mechanism of motion perception. In our experiments, observers perceived attentionally determined motion even when stimuli were much too brief for conscious tracking. We assume that 'attentionally tracking' in Cavanagh's experiment[22] involves the same salience marking as does 'selectively attending' in ours, although tracking may also involve other mechanisms (such as efference copy[23]). According to this interpretation, in both paradigms, third-order motion perception depends on an automatic stimulus computation, which in turn depends on the attentional state at the time of stimulus presentation.

(2) In perceptual search tasks, selective attention to features (such as red items among green items) seems to be mediated by directing attention to the location of the attended feature, and then processing the information at the location. The critical aspect of these results is that, in rapid sequences of dynamic search displays, as in motion, selective attention is mediated via the locations of the attended features; attention does not directly exclude or admit items containing the feature[24,25].

All these phenomena are encompassed within the salience map theory. A dynamic salience map of the locations of the most salient stimulus features is determined jointly by stimulus strength (bottom-up) and by selective attention (top-down). Motion is computed directly and automatically from the salience map, but the map can also be used to guide other processes, such as object recognition in attention-guided search and memory storage in location-cued recall tasks[26,27]. □

Received 24 March; accepted 28 July 1995.

1. Lu, Z.-L. & Sperling, G., *Vision Res.* **35,** 2697–2722 (1995).
2. Reichardt, W. *Z. Naturf.* **12b,** 447–457 (1957).
3. van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am.* **A1,** 451–473 (1984).
4. Adelson, E. H. & Bergen, J. R. *J. opt. Soc. Am.* **A2,** 284–299 (1985).
5. Cavanagh, P. & Mather, G. *Spatial Vision,* **4,** 103–129 (1989).
6. Chubb, C. & Sperling, G. *Proc. natn. Acad. Sci. U.S.A.* **86,** 2985 (1989).
7. Shipley, W. G., Kennedy, F. A. & King, M. E. *Am. J. Psychol.* **58,** 545–549 (1945).
8. Braddick, O. J. & Adlard, A. J. in *Visual psychophysics: Its psychological basis* (eds Armington, J. Krauskopf, J. & Wooten, B. R.) 417–426 (Academic, New York, 1978).
9. Shadlen, M. & Carney, T. *Science* **232,** 95–97 (1986).
10. Julesz, B. & Payne, R. A. *Vision Res.* **8,** 433–444 (1968).
11. Petersik, J. T., Hicks, K. I. & Pantle, A. J. *Perception* **7,** 371–383 (1978).
12. Zanker, J. M. *Vision Res.* **33,** 553–569 (1993).
13. Cavanagh, P., Arguin, M. & von Grunau, M. *Vision Res.* **29,** 1197–1204 (1989).
14. Pantle, A. & Picciano, L. *Science* **193,** 500–502 (1976).
15. Georgeson, M. A. & Shackleton, T. M. *Vision Res.* **29,** 1511–1523 (1989).
16. Victor, J. D. & Conte, M. M. *Vision Res.* **30,** 289–301 (1990).
17. Werkhoven, P., Sperling, G. & Chubb, C. *Vision Res.* **33,** 463–485 (1993).
18. van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am.* **A2,** 300–321 (1985).
19. Hallett, P. E. in *Handbook of Perception and Human Performance* Ch. 10 (eds Boff, K. R., Kaufman, L. & Thomas, J. P.) (John Wiley, New York, 1986).
20. Treisman, A. & Gormican, S. *Psychol. Rev.* **95,** 15–48 (1988).
21. Kahneman, D., Treisman, A. & Gibbs, B. J. *Cognitive Psychol.* **24,** 175–219 (1992).
22. Cavanagh, P. *Science* **257,** 1563–1565 (1992).
23. Helmholtz, H. *Helmholtz's Treatise on Physiological Optics* 3rd edn (ed. Southall, J. P. C.) Menasha, Wisc. (Optical Society of America, Menasha, Wisconsin 1925). (Originally published in 1867).
24. Cave, K. R. & Wolfe, J. M. *Cognitive Psychol.* **22,** 225–271 (1990).
25. Shih, S. & Sperling, G. *J. exp. Psychol., hum. Percept. Perform.* (in the press).
26. Sperling, G. *Psychol. Monogr. 74* (1960).
27. Sperling, G. & Weichselgartner, E. *Psychol. Rev.* **102,** 502–532 (1995).
28. Solomon, J. A. & Sperling, G. *Vision Res.* **34,** 2239–2257 (1994).

ACKNOWLEDGEMENTS. This work was supported by the U.S. Air Force Office of Scientific Research, Life Sciences, Visual Information Processing Program.