Vision Res. Vol. 24, No. 9, pp. 1083–1090, 1984
Printed in Great Britain.

0042-6989/84 $3.00 + 0.00

ILLUSORY MOTION IN VISUAL DISPLAYS

A. M. M. Lelkens and J. J. Koenderink
Department of Medical and Physiological Physics, State University Utrecht, Princetonplein 5, 3584 CC Utrecht, The Netherlands

(*Received* 16 *February* 1983; *in revised form* 2 *December* 1983)

Abstract—The apparent motion of a change in the structure of a random check pattern is studied by spatially masking it with another noise pattern and it is compared with ϕ motion. A fundamental difference with ϕ motion is the insensitivity of second order correlators (Reichardt mechanisms) to this apparent motion. The following experimental characteristics distinguish this motion from ϕ motion: it induces no motion after-effect, it is not transparent to another simultaneous motion, it is strongly influenced by spatial masking and it does not evoke optokinetic nystagmus. A fourth order detector is introduced which is sensitive to this illusory motion as well as to ϕ motion. Simulation experiments with this detector together with the subjective reports of the observers lead us to the conclusion that human subjects inadvertently treat the coarsest spatial structures as signal and the finest as the disturbing noise.

INTRODUCTION

Since the pioneering study of Wertheimer (1912) it has been commonly accepted that motion can be perceived even if there is no real continuous motion: two spatially separate points lit at different times appear as one point moving from one location to the other. The finding of units in the rabbit's retina which respond to both continuous and stepwise motion (Barlow and Levick, 1965) and the occurrence of optokinetic nystagmus in both cases (Morgan and Turnbull, 1978) support the idea that the visual system cannot discriminate between real continuous motion and motion in discrete spatial jumps. For small spatial and temporal separations this is obviously true because of the limited spatio-temporal resolution of the eye. The eye's inability to discriminate between these two is the reason why motion perception is often investigated by means of stimuli moving in jumps. A fruitful stimulus is a random dot pattern which is shifted over one or more dot-size distances in discrete time steps. The stimulus is kept at the same location by omitting a column of dots at the side to which the motion is directed and by adding a new random column at the other side. The advantages of this kind of stimulus are that the moving object has no particular structure and that the sensitivity of the human visual system to this motion can be measured by adding uncorrelated dots (noise) to the stimulus (van Doorn and Koenderink, 1982). Moreover, the target extent is an easily controllable parameter.

In these experiments the detection of motion can be explained in terms of an ensemble of detectors of the Reichardt type (1961) which correlate an illuminance in position x at time t with an illuminance in $x + dx$ at time $t + dt$. The quotient dx/dt defines the speed to which the detector is tuned. It is possible, however, to elicit motion with random dot patterns that are not displaced versions of one another. Two random patterns at different locations replaced by new, uncorrelated patterns at different times give the illusion of motion from the pattern that is changed first to the one changed last, although in fact in this stimulus there is nothing that is now here and then there. This effect was mentioned by Sperling (1976). Because of the lack of correlation between the patterns involved the Reichardt detectors are insensitive to this stimulus and therefore cannot form the underlying mechanism for the perception of motion in this case. The fact that in this stimulus nothing moves at all, in the sense that no spatio-temporal correlation exists, has motivated us to investigate this effect further.

EXPERIMENT

A "check" is formed by a square array of contiguous white pixels. The "density" of the checks is equal to the probability of a check and thus it is a measure of the part of the screen that is filled by the checks. A random pattern of such checks is presented on an HP 1321A monitor screen with P4 phosphor (a square array of 256 × 256 pixels) and viewed from a distance of 4 m. The subjects used their left eye and an artificial pupil with a diameter of 2.9 mm. With a fully white screen the retinal illuminance is 46 td. The subjects are male, experienced observers aged about 25. E.M. has normal vision whereas A.L. uses a correcting lens (−3.5 D).

The pattern is divided into 4 vertical bars each 4° high and 1° wide. The pattern in each bar is replaced by a new uncorrelated one according to the scheme in Fig. 1.

The slope of the curves is 15°/sec. These patterns, referred to as P1, are masked by the superposition of another pattern (P2) which subtends the total field and is replaced every 67 msec, synchronously with the alterations in P1 [masking by visual noise (Kahnemann, 1968)]. This superposition is a logical inclusive "OR": a pixel will be displayed white when

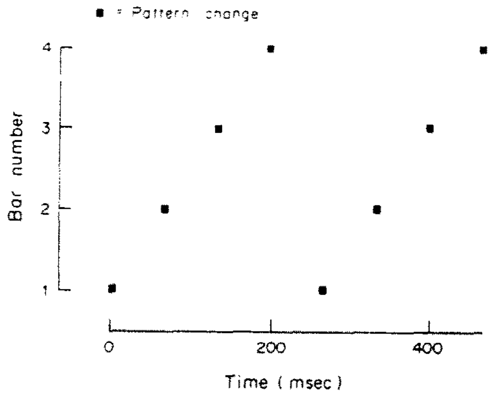

Fig. 1. The sequence of pattern changes in the four bars of the screen. Every 67 msec the pattern in a bar is changed. The width of the bars is 1°. The slope of the line connecting the changes is 15°/sec.

it occurs in P1 or in P2 or in both. An example of a stimulus is shown in Fig. 2. The two different check sizes are easily recognised.

The direction of the motion (we shall call it μ motion from now on) can be reversed by reversing the order in which the patterns in the four bars are altered. The direction is selected at random and the subject's task is a forced choice between the two alternatives. The stimulus presentation time is about 1.5 sec. In the inter-stimulus interval of 2 sec the monitor screen is of uniform luminance equal to the mean luminance of the test stimuli. When a correct answer is given the density of the checks in P1 is decreased by 1 step (initially 6%) and for a wrong answer it is increased by 3 steps. This strategy automatically converges to the point where 75% of the answers are correct; this point is defined as the threshold. The density in P2 is adjusted so as to keep the total density of white pixels constant. When the densities in P1 and P2 are denoted by d_1 and d_2

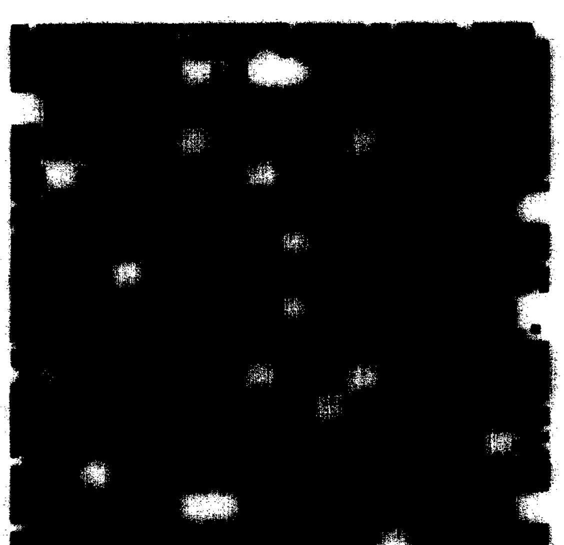
Fig. 2. Example of a stimulus with a signal-check size of 1/4° (with density 1/9) and a mask-check size 1/16° (density 5/32). The total density equals 1/9 + 5/32 − 5/288 = 1/4.

respectively, this means that

$$d_1 + d_2 - d_1 \cdot d_2 = t \text{ (a constant).}$$

When in a bar only P2 is replaced the probability that a given pixel will change from white to black or vice versa is

$$2 \cdot d_2 \cdot (1 - t).$$

When both P1 and P2 change this probability is

$$2 \cdot t \cdot (1 - t).$$

The information possibly contributing to the perception of motion is given by the difference between these two, that is

$$2 \cdot (t - d_2) \cdot (1 - t) = 2 \cdot d_1 \cdot (1 - d_2) \cdot (1 - t)$$

which is referred to as the signal density.

A turning point is defined as a sequence of a right answer followed by a wrong one or the other way around. After 1, 3, 7... turning points the step-size is decreased by a factor 2 (Levitt, 1971). Pilot experiments showed that 10 turning points are sufficient to reach a stable end of a series. Each data point in the figures is the average of 9 such series. These 9 series were taken in 3 runs spread over 2 or 3 days. The experimental parameters were randomly spread over the series.

RESULTS

Figures 3 and 4 show regression lines for the threshold vs the total density. For a mask check size increasing from 1/64° to 1/4° the slope of this line increases from 0.6 to 1.0 when the signal check size is 1/16°. In the case of a signal check size of 1/4° the slope ranges from about 0 for the smallest mask checks to 0.7 for the largest. For a mask much smaller than the signal (1/64° and 1/4° respectively) the threshold is almost independent of t. No matter how many (small) mask checks are present in the stimulus the motion can be detected with the same number of signal checks. Small checks apparently do not mask the motion when the information about the motion is carried by much larger checks. Figure 5 shows this in more detail.

Masking with a pattern of checks larger than the signal checks almost completely destroys the illusion of motion (left half of each figure): the threshold is close to the maximum signal density of 1/2 (for $t = 1/2$). There is a small decrease in threshold for very large mask check sizes. The number of mask checks is small and the motion can be seen in some parts of the screen where, by chance, there happens to be no mask check. When the motion is masked with checks smaller than the signal checks, the threshold decreases with decreasing mask check size. The smaller the signal checks, the lower the threshold.

From Fig. 5 it follows directly that the masking effect is not the result of covering a part of the signal and resultant lowering of the amount of motion information available to the subject. If this were the

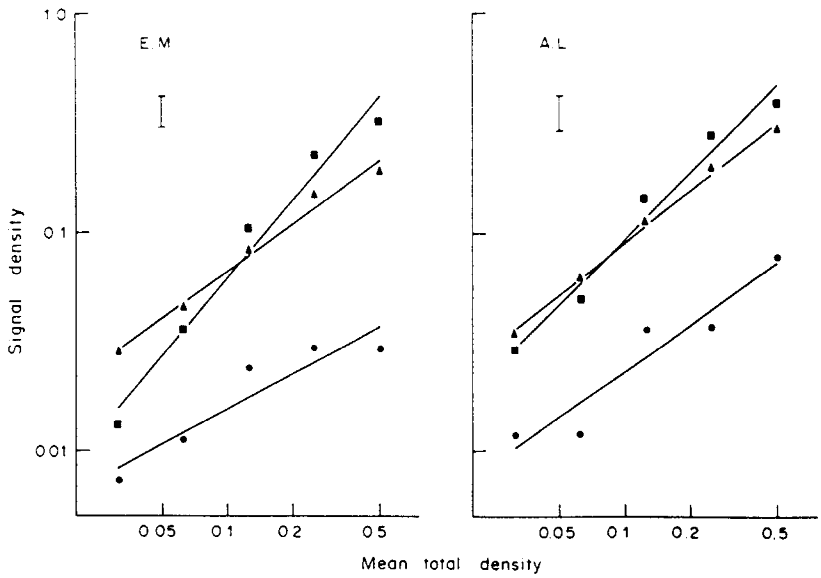

Fig. 3. Thresholds for two subjects E.M. and A.L. as a function of the total mean density for signal check size of 1/16°. Sizes of the mask checks are 1/64° (circles), 1/16° (triangles) and 1/4° (squares). The vertical bar indicates the standard error.

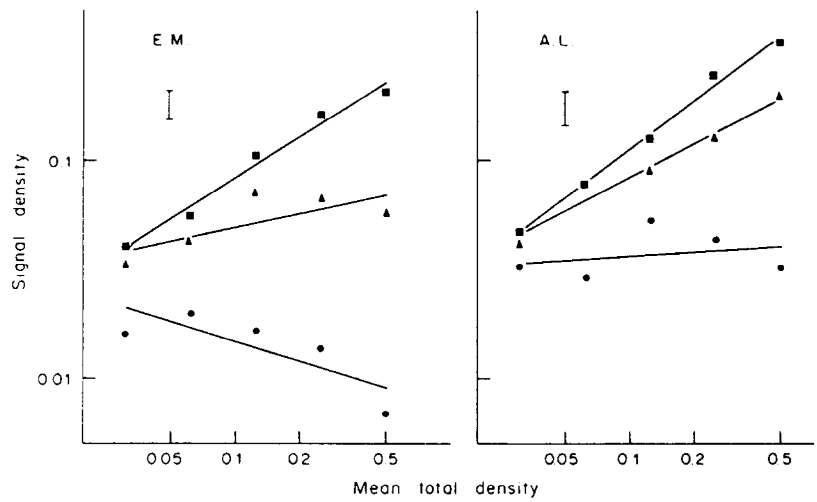

Fig. 4. The same as Fig. 3 except for signal checks of 1/4°.

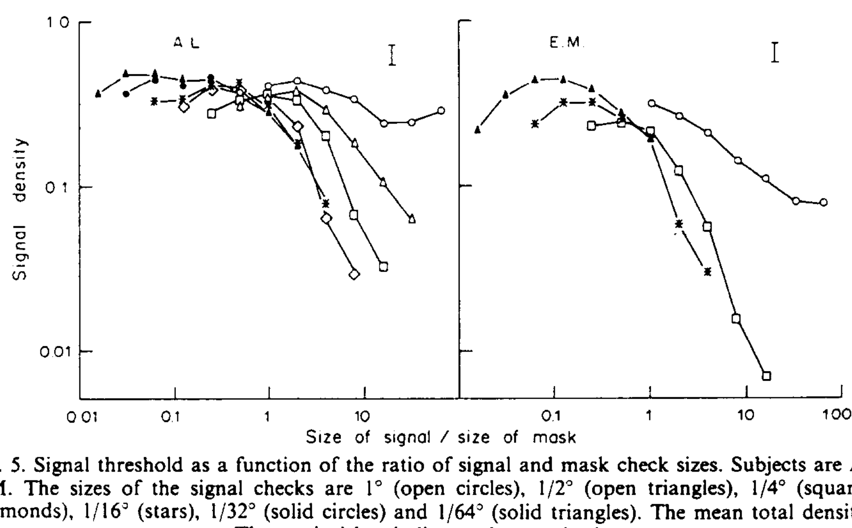

Fig. 5. Signal threshold as a function of the ratio of signal and mask check sizes. Subjects are A.L. and E.M. The sizes of the signal checks are 1° (open circles), 1/2° (open triangles), 1/4° (squares), 1/8° (diamonds), 1/16° (stars), 1/32° (solid circles) and 1/64° (solid triangles). The mean total density is 1/2. The vertical bar indicates the standard error.

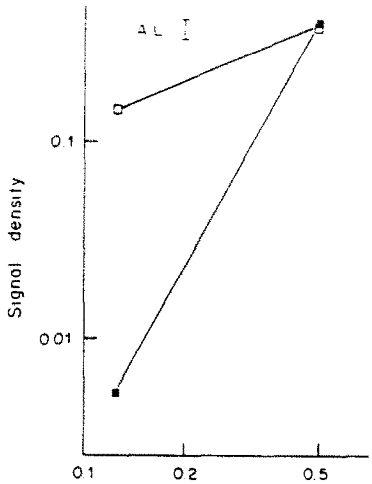

Fig. 6. Results of the experiment in which the mask checks are black (solid squares) compared with the results of the original experiments with white mask checks (open squares). The size of the signal checks is 1/16° and the size of the mask checks is 1/4°.

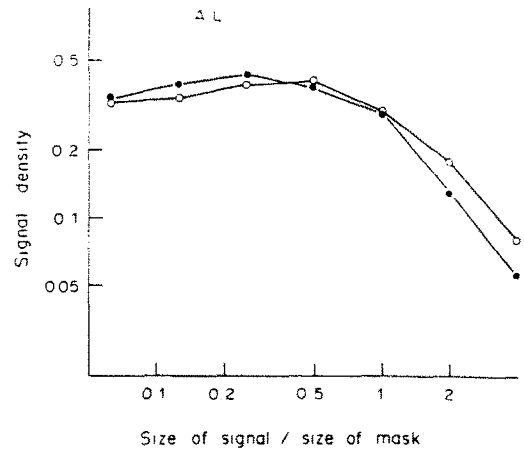

Fig. 7. The results of the experiment in which black and white are exchanged (solid circles) together with some results of the original experiment (open circles). The size of the signal checks is 1/16°. The mean total density is 1/2.

case the threshold would have to be independent of mask check size. To make the above conclusion even more evident we changed the stimulus in such a way that a pixel belonging to a mask check was not displayed and therefore mask checks were black. Because the background is black as well these checks are not visible except perhaps as holes in the signal distribution. In this way the masking erases the same number of pixels as in the former experiments.

In Fig. 6 it is shown that for $t = 1/2$ the resulting masking effect is as strong as it is for the white mask. This may be due to the fact that the screen is filled sufficiently for the subject to discriminate the black mask checks. For $t = 1/8$ however, the mask checks merge with the background and are no longer distinguishable; this strongly reduces the masking effect. It is therefore the visibility of the mask checks which thwarts perception and not the loss of information they cause.

The assumption that the colour of the checks (black or white) is of no importance is confirmed experimentally by making all the white checks black and vice versa (i.e. a logic "NOT" operation performed on the z-axis). The results of such an experiment are shown in Fig. 7. There is no significant difference in threshold in either case.

MODEL SIMULATION

Due to the lack of correlation between the patterns in the stimulus we expect the Reichardt correlator not to be able to detect μ motion. We verified this by means of a computer simulation, which showed that even in the absence of a mask pattern the percentage of correct decisions about whether the motion was to the left or to the right did not exceed the 50% probability of a pure guess. In this way μ is discriminated from ϕ motion to which these correlators are sensitive. In fact the only correlation that is present in the stimulus is the one between the changes in the patterns of each bar. In order to extract this information from the stimulus it is necessary for the subject to detect changes in each bar prior to correlating between bars. A mechanism that can perform such a task is at least of the fourth order and a possible implementation is shown in Fig. 8.

This mechanism first determines the autocorrelation of the signal in a part of the stimulus. It then correlates this to the autocorrelation in another part of the stimulus at another time. Since displacement of a pattern also evokes local changes in time, this correlator is able to detect ϕ motion as well. In our experiment the time delay T necessary for a

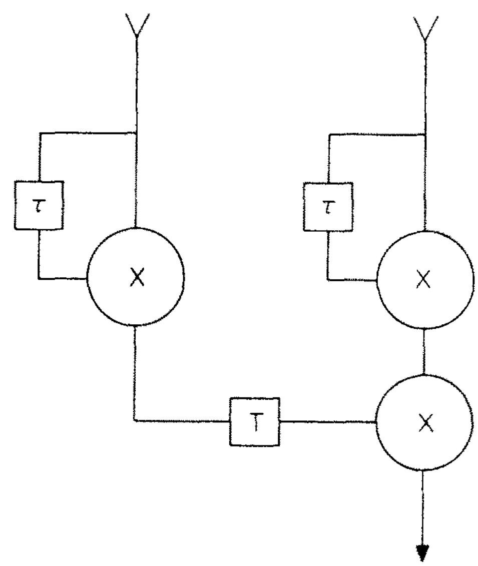

Fig. 8. The structure of a mechanism that is capable of detecting μ motion. The squares denote time delays, the circles correlators. The output of a receptor is autocorrelated. The result is correlated with the retarded autocorrelated output of another receptor that is situated in another part of the stimulus.

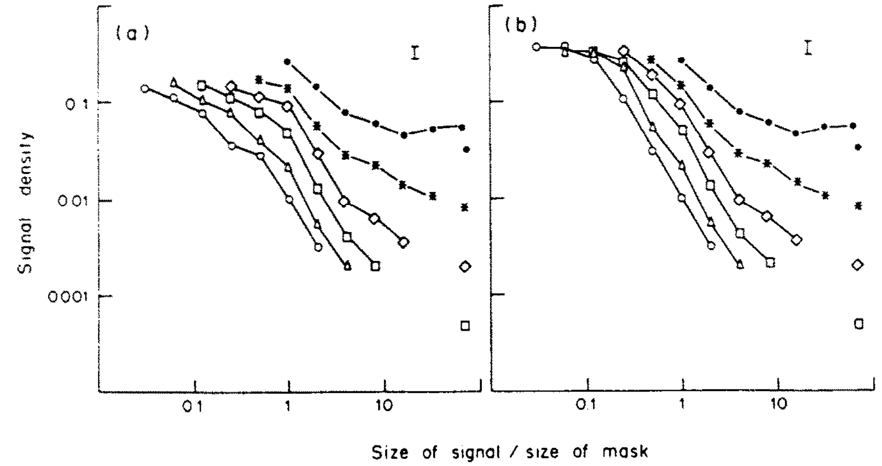

Fig. 9. (a). The thresholds for an ensemble of detectors of the type shown in Fig. 8. The sizes of the signal checks are 1° (solid circles), 1/2° (stars), 1/4° (diamonds), 1/8° (squares), 1/16° (triangles) and 1/32° (open circles). The total density is 1/2. The isolated points indicate the limiting value for very large ratios of signal size versus mask size. The size of the receptor aperture is equal to the size of the signal checks. (b). Same as (a) but with size of the receptors set to the size of the largest checks in the stimulus.

positive correlation is 67 msec for every bar-width separation between the receptors.

We carried out a computer simulation using a contiguous set of these mechanisms which covered the whole stimulus. The stimulation followed the same experimental forced choice procedure as was used for human subjects. The stimulus consisted of one sweep of motion across the four bars. The sampling aperture of the detector unit is set either to the size of the signal [Fig. 9(a)] or to the largest of mask and signal check size [Fig. 9(b)].

The strong decrease in threshold for smaller mask grain is due to the fact that this mask is averaged over the larger detector units. The limit value for very small mask checks can be calculated in the following way. The white-pixel density in this case has zero variance. A detector unit which samples over a certain area cannot detect any changes in the mask pattern. The probability that a signal check will change is

$$2\cdot d_1\cdot(1-d_1).$$

For N checks in a bar the probability that no check will change is

$$[1-2\cdot d_1\cdot(1-d_1)]^N = p.$$

An optimum detector summates all the changes in a bar and correlates these not only over 1 bar width but also over 2 and 3 bar widths. The probability that a stimulus is presented which forces the detector to guess is given by the probability that in no more than 1 bar a change occurs

$$P_g = p^3\cdot(4-3p).$$

The probability that a correct answer will be given is

$$P_c = 1 - P_g/2.$$

The only solution to $P_c = 75\%$ is $p = 0.6143\ldots$ (for $0 < p \leqslant 1$).

The signal density is

$$2\cdot d_1\cdot(1-d_2)\cdot(1-t)$$
$$= -\frac{e_{\log p}}{4N}\cdot\left(1-\frac{e_{\log p}}{2N}\ldots\right).$$

These thresholds are shown in Fig. 9 as isolated points. It is obvious that Fig. 9(b) looks more like the results for human subjects than does Fig. 9(a), since in the latter figure the threshold never reaches 50%. In Fig. 9(b) the detector size is set to the largest checks in the stimulus. This would indicate that the human subjects always take the larger checks as the signal and the smaller ones as the disturbing noise, despite the fact that they know that the reverse is true. This is in agreement with the reports of the subjects who stated that the large checks attracted all their attention and the smaller ones seem stationary.

In Fig. 10 three different variants on the basic fourth order correlator scheme are given, all based on detection of the largest checks. One variant is identical to the one used in Fig. 9(b), the second one concerns summation over complete bars before correlation between bars and the third uses a threshold which makes the detector insensitive to smaller checks.

Summation over a complete bar before correlation lowers the threshold but keeps the dependence on mask size unchanged, whereas the internal threshold sharpens the curve around the point where signal and mask checks are of equal size.

Figure 11 shows the threshold for the model as a function of total mean density for two signal check sizes. For a signal check size of 1/16° and a mask of 1/4° the slope of the curve is 0.34 when a detector size of 1/16° (equal to the signal check size) is used. But when the size of the detector is chosen to be 1/4° (equal to the mask check size) this slope is 0.81. This latter value is in better agreement with the results of the subjects, for whom a slope of almost 1 was

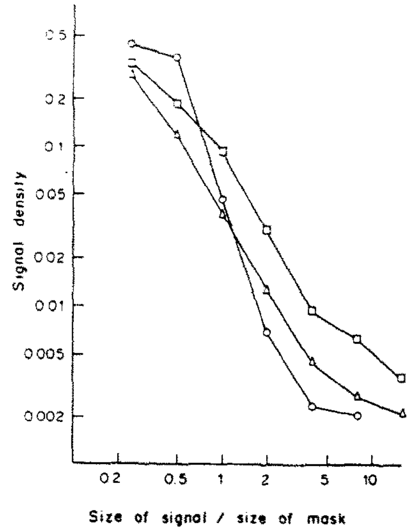

Fig. 10. Several variations of the model detector. The triangles indicate the results when the autocorrelated output of all receptors is summed before the correlation over the bars. Some of the points from Fig. 9(b) are also shown, denoted by squares. The circles are the results when a criterion is implemented in a receptor so that it only gives an output when at least half its area is illuminated. Signal check size is 1/4°.

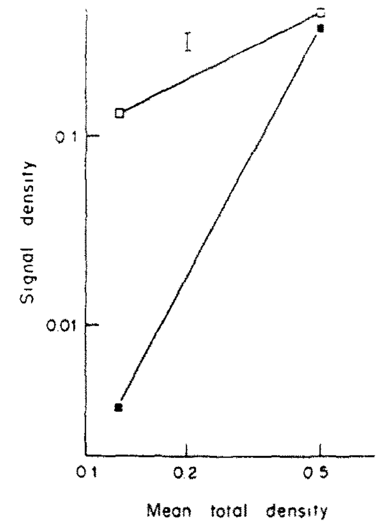

Fig. 12. Simulation thresholds for the experiment in which the mask checks are black (solid squares) compared with the thresholds for the original experiment (open squares). The size of the signal is 1/16° and the size of the mask 1/4°.

measured (Fig. 3). This also indicates that the subjects base their detection on the largest checks. For a signal size of 1/4° it is surprising that the results with a mask of 1/16° in the model are comparable to the results with a mask of 1/64° for the human subjects. The model is clearly better at discriminating between different sizes. This was found earlier when Fig. 5 and Fig. 9(b) were compared: the curves for the model start to decrease for a smaller ratio of check sizes than they do for the subject.

The black-masking experiment of Fig. 6 can now be readily described by the model that uses detectors with aperture sizes equal to the largest checks that are visible in the stimulus (Fig. 12).

When the total density is 1/2 the thresholds coincide, whereas for 1/8 the black masking is much less effective. The model fits the experimental data (Fig. 6) very well.

COMPARISON WITH ϕ MOTION

A stimulus that is frequently used for studying ϕ motion consists of a random check pattern that is displaced at discrete time steps over a small distance.

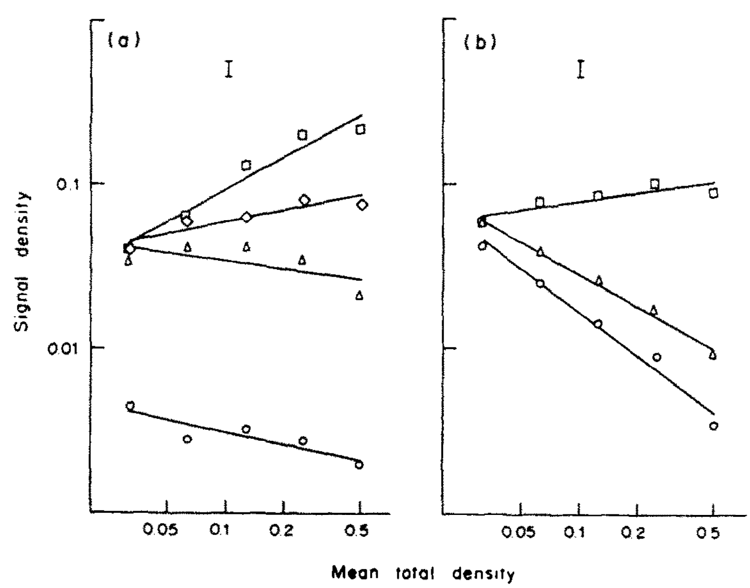

Fig. 11. (a). The results of a model simulation for a signal check size of 1/16°. The sizes of the mask checks are 1/64° (circles), 1/16° (triangles) and 1/4° (squares and diamonds). The size of the receptors is 1/16°, except in the case of the curve with the squares (there it is 1/4°). (b). Simulation results with a signal check size and a detector size of 1/4°.

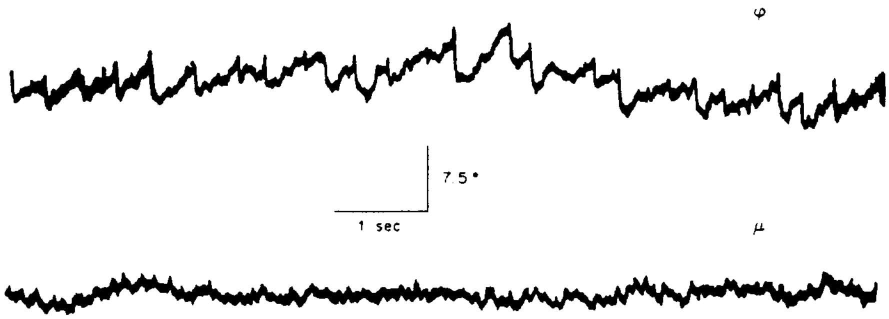

Fig. 13. Eye movement recordings for ϕ and μ motion for subject E.M. The ϕ motion was formed by displacement of a random bar pattern. In both cases, the density was 1/2 and the width of the bars and checks respectively was 1/4°.

At one end of the stimulus a column disappears and at the other end a new uncorrelated column is added. For a logical "OR" of such a stimulus with another, non-moving pattern we did not observe a masking of the motion (apart from the trivial case of a completely filled screen). The second pattern is completely transparent to ϕ motion.

Another difference is the occurrence of a motion after-effect. If, after prolonged viewing of ϕ motion this motion is stopped, the observer gets the impression of a motion in the reverse direction. With a similar procedure for μ motion no such effect was observed.

Figure 13 shows recordings of eye movements for both ϕ and μ motion. The ϕ motion consisted of a random bar pattern shifted 1/8° 60 times per second.

The subject was instructed to fixate in the centre of the stimulus. However, in order to prevent the subject from staring no fixation point was supplied. For the sake of resolution the viewing distance was 1 m so the target subtended 16°. Furthermore the speed of both ϕ and μ motion was chosen to be 7.5°/sec. In the case of ϕ motion nystagmus is visible in the recordings. With μ motion it is not.

For random check patterns with density 1/2 the maximum distance over which the stimulus may be translated without the subject losing the perception of ϕ is about 20′ irrespective of the check size (Braddick, 1974; Morgan and Ward, 1980). Larger displacements of up to 18° (Zeeman and Roelofs, 1954) can be used only for stimuli with recognizable structure. In the case of random check patterns, however, each check has enough neighbours with which it can be correlated for the real correlation over larger distances not to be detected. This is the so-called correspondence problem (Braddick, 1974). Since the neighbours change at random, the orientation of short range correlation is completely random and no resultant motion is seen. For low densities of a random check pattern simple structures occur, which can be tracked easily and thus facilitate motion perception. By experiment we find that this motion with low densities and large displacements is easily masked by another pattern. It does not build up an after-effect. This stimulus can even be combined with μ in a way that is illustrated in Fig. 14.

This sequence of patterns results in a μ motion to the left at four times the speed of the ϕ motion to the right. It is not possible to see μ and ϕ simultaneously. Whether μ or ϕ is visible depends on the check density: for high densities μ is seen and for low densities it is ϕ that is seen. For intermediate values (5–10 checks in each bar) the motion is ambiguous: either μ or ϕ is seen, depending on the actual number of checks present and the pattern they form.

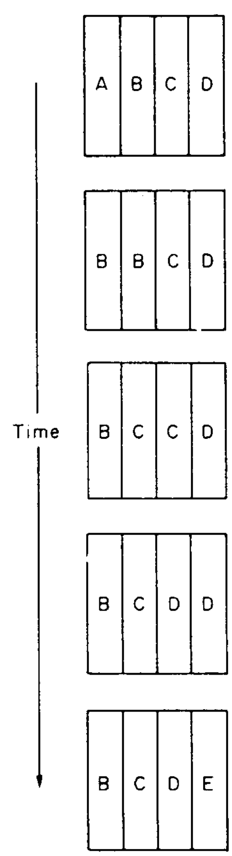

Fig. 14. The arrangement of a stimulus to provide ϕ and μ simultaneously. A, B, C, D and E denote uncorrelated random patterns. At discrete time steps each pattern is replaced by its neighbour, thus eliciting μ directed to the right. After 4 time steps the whole pattern has been shifted over one bar-width to the left and a new uncorrelated pattern was added. This is a ϕ motion to the left.

DISCUSSION

The function of a fourth order detector has been discussed by Reichardt and Poggio (1976). They showed that it is the smallest configuration that can cope with the task of figure-ground discrimination. It can be used to separate two patterns moving at different speeds or in different directions. The autocorrelation of a receptor signal can perhaps be formed by the response of both the "on" and "off" ganglion cells, since the sign of the local change in illuminance is irrelevant.

The fact that two ϕ motions over small displacements with different speeds or directions can be seen transparent (Clarke, 1977) means that correlator mechanisms tuned to different directions and/or speeds work simultaneously and independently. The absence of such transparency for larger displacements and for μ points to only a single mechanism.

The stimulation results, obtained with a detector size equal to the largest of the checks in the stimulus describe the results for human subjects better than the results with a detector size equal to the size of the checks that carry information on the motion. It has therefore been argued that the mechanism used by humans invariably employs the largest spatial structures as the relevant signal. A better strategy would have been to use several sample sizes in parallel and to let the one with the highest signal-to-noise ratio determine the perception. It seems, however, that the subject is forced to use a specific detector size, although another detector may have a lower threshold. It is concluded that the choice of a detector is based on the spatial composition of the stimulus, which means that form perception plays an important role in the seeing of μ. This is another feature that distinguishes μ motion from ϕ motion: according to the experiments of Anstis (1970) luminance rather than form is important for seeing ϕ. The ϕ motion between two granular patterns in the case where one pattern is the photographic negative of the other is reversed. The motion is seen in the direction of the largest correlation of luminance rather than form.

The visibility of μ can be troublesome in the case of a stepwise updated display of moving scenes. When a new scene cannot be updated quickly enough (e.g. real time computer calculations are needed) the update itself becomes visible and interferes with the ϕ motion in the scene.

Acknowledgements—This study was financially supported by the Dutch Organization for the Advancement of Pure Research (ZWO). The authors are indebted to S. M. McNab for her linguistic advice and to E. van Maanen for acting as a test subject.

REFERENCES

Anstis S. M. (1970) Phi movement as a subtraction process. *Vision Res.* **10**, 1411–1430.

Barlow H. B. and Levick W. R. (1965) The mechanism of directional selective units in rabbit's retina. *J. Physiol., Lond.* **178**, 477–504.

Braddick O. (1974) Short-range process in apparent motion. *Vision Res.* **14**, 519–527.

Clarke P. G. H. (1977) Subjective standstill caused by the interaction of moving patterns. *Vision Res.* **17**, 1243.

Doorn A. J. van and Koenderink J. J. (1982) Temporal properties of the visual detectability of moving spatial white noise. *Expl Brain Res.* **45**, 179–188.

Kahneman D. (1968) Method, findings, and theory in studies of visual masking. *Psychol. Bull.* **70**, 404–425.

Levitt H. (1971) Transformed up–down methods in psychoacoustics. *J. acoust. Soc. Am.* **49**, 467–477.

Morgan M. J. and Turnbull D. F. (1978) Smooth eye tracking and the perception of motion in the absence of real movement. *Vision Res.* **18**, 1053–1059.

Morgan M. J. and Ward R. (1980) Conditions for motion flow in dynamic visual noise. *Vision Res.* **20**, 431–435.

Reichardt W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* (Edited by Rosenblith W. A.), pp. 303–317. MIT Press, Cambridge, MA.

Reichardt W. and Poggio T. (1976) Visual control of orientation behaviour in the fly. *Q. rev. Biophys.* **9**, 311–375.

Sperling G. (1976) Movement perception in computer-driven visual displays. *Behav. Res. meth. Instrum.* **8**, 144–151.

Wertheimer M. (1912) Experimentelle Studien ueber das Sehen von Bewegung. *Z. Psychol.* **61**, 161–265.

Zeeman W. P. C. and Roelofs C. O. (1953) Some aspects of apparent motion. *Acta psychol.* **9**, 159–181.