

SPATIOTEMPORAL INTEGRATION IN THE DETECTION OF COHERENT MOTION

ANDREA J. VAN DOORN and JAN J. KOENDERINK

Department of Medical and Physiological Physics, Physics Laboratory, State University Utrecht,
Princetonplein 5, 3584 CC Utrecht, The Netherlands

Abstract—We present data on the signal to noise threshold (SNR) for the impression of coherent, smooth motion in stroboscopically moving dot patterns (frame repetition rate 100 Hz) degraded with additive noise (uncorrelated pattern each new frame). (Both patterns were 250×255 square arrays in which a pixel was randomly assigned a light or dark value with probability 50%.) It is found that the threshold varies roughly as the inverse square root of the number of available basic dot correlation pairs, subject to certain constraints that can be interpreted as due to the geometry of basic correlator units subserving the detection of motion. This geometry is estimated from the data and compared with values that have been obtained with independent methods. Phenomenologically the results are simple: when the target field is enlarged in the direction perpendicular to the motion the SNR threshold follows the theoretical expectations. Otherwise the variation is stronger, which points to a recruitment of new units. For an impression of movement to occur the smallest targets are short thin strips elongated in the direction of motion (e.g. $2.5' \times 12'$ for $2.08 \text{ deg sec}^{-1}$), but the threshold continues to fall at least to $5' \times 5'$. There is a trade-off between target area and presentation time. For a $5' \times 5'$ target coherent motion can be seen in 3 successive frames, but for a $10' \times 10'$ target ten times as many frames have to be presented, whereas an impression of movement is not obtained for still smaller targets.

INTRODUCTION

Human performance in the detection of coherence in moving random speckle patterns can be conveniently described with correlation models (Lappin and Bell, 1976). In previous papers (van Doorn and Koenderink, 1982a, b) we have shown that the model proposed by Reichardt (1961) as the mechanism behind the optomotor responses of insects is also applicable to human vision. This model consists of an ensemble of structurally identical basic units (Fig. 1), consisting of two sub-receptive fields (diameter δ) at a mutual distance Δ , whose output signals are correlated for some fixed delay T . Such a unit is selectively sensitive to motions with velocity Δ/T . A detection of motion would occur if the pooled response of a number (N say) of basic units exceeds the noise level. Simple reasoning shows that one should expect threshold signal to noise ratios approximately equal to $N^{-1/2}$ (van Doorn and Koenderink, 1982a; Lappin and Bell (1976) have derived a similar equation). We have shown that this model is able to predict many of the peculiarities of human motion detection, and we succeeded in determining the parameters Δ and T as a function of the velocity (van Doorn and Koenderink, 1982a, b).

From these experiments it appears that a great many basic units cooperate for a detection threshold. Threshold signal to noise ratios as low as 10^{-2} were found. In this paper we present results that enable a more precise estimate of the number of basic pair correlations necessary for the impression of coherent movement.

METHODS

Random dot patterns of 250×255 dots were electronically generated and displayed on a HP-

monitor (model 1317-A, P4 phosphor, 470 μsec decay time to 1% intensity). Each 10 msec a new frame was generated. The impression of coherent motion is obtained when the patterns in different frames are mutually shifted in position. The velocity of this stroboscopic movement can be varied by adjustment of the spatial shift and the adjustment of the number of frames between successive shifts. Apart from this stroboscopically moving pattern another electronic generator generated a fresh uncorrelated speckle pattern every new frame. This signal in isolation looks like a "snow storm". It is spatially and temporally completely random. In both patterns the probability for the generation of a dot is 50% for any pixel. In our experiments the correlated and the uncorrelated pattern are electronically mixed in such a way that the root mean square contrast is held invariant whereas the ratio of the variances of the luminance in the two patterns (that is the signal to noise ratio, henceforth abbreviated as SNR) is variable. The subject controls the SNR and seeks the threshold for the impression of coherent motion. When coherent motion is detected the subject is always sure of its direction. The threshold is bracketed at least 16 times.

In all experiments the field subtended $5.21 \times 5.31 \text{ deg}$ of visual angle, mean luminance 180 cd/m^2 . The surround was dark. The subjects viewed the screen binocularly with the natural pupil. They had normal, or corrected to normal vision.

COHERENT MOTION IN LONG STRIPS

The Reichardt correlator mechanism is anisotropic. Thus it is a priori likely that integration properties in the direction of motion and per-

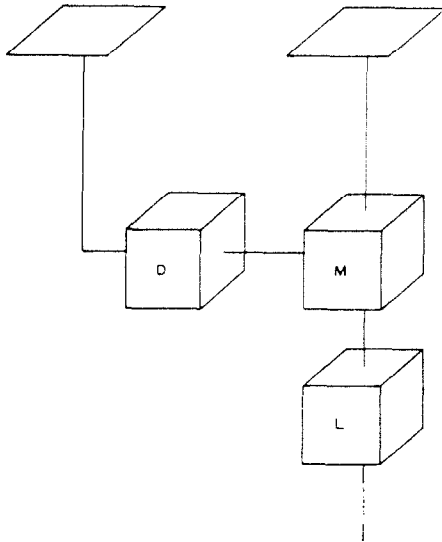


Fig. 1. A simplified scheme of the mechanism proposed by Reichardt (1961, 1976) as the basic building block of the mechanism subserving insect optomotor responses. This simple model can also be used to understand many aspects of human motion detection in a semi-quantitative fashion (van Doorn and Koenderink, 1982a, b). The mechanism consists essentially of two sub-receptive fields (drawn as square photocells here, in the text their sides are denoted δ) at a separation Δ . The signal of one sub-field is delayed (delay box D, the delay is referred to in the text as T msec) and multiplied with the signal of the other sub-field (multiplier box M). The product-signal is fed into a linear filter (box L) which essentially computes a running time average. The output of the mechanism is an estimate of the spatiotemporal correlation of the fluxes collected by the two sub-fields. In general this correlation will be low, except when the input signal is a moving pattern with velocity Δ/T in the direction from left to right. Thus the system is "tuned" to this velocity. Much of the human performance in the detection of coherent motion in the presence of noise can be understood on the assumption that the summed signal of many of these elementary units distributed over large areas in the visual field has to exceed a certain fixed signal to noise ratio near to unity.

pendicular to the direction of motion have to be distinguished. In this experiment we investigate this in the following way: we generate a rectangular window in the target field of size A in the direction of motion and size B perpendicular to the direction of motion centered in the almost square target field of side D (Fig. 2). In the first experiments we took either A or B equal to D and varied the other parameter. The target field outside the rectangle was filled with uncorrelated noise patterns ("snow"). In the rectangle we presented stroboscopically moving noise patterns, degraded to the degree specified by the SNR. For low values of the SNR the rectangle is hardly visible or not visible at all, whereas for high values of the SNR it is clearly perceived with sharp boundaries.

Results

When the assumption of a simple pooling of responses of elementary correlation pairs holds, then the number of stimulated pairs should determine the

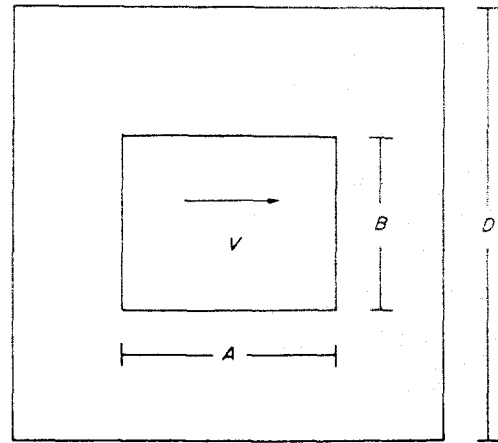


Fig. 2. The target field configuration. The total screen was almost square: $5.21^\circ \times 5.31^\circ$, thus $D = 5.31^\circ$. The sub-field was variable with a length A in the direction of motion and a height B perpendicular to the direction of motion.

threshold and we expect the SNR threshold to be inversely proportional with the square root of the width of the strip. Figures 3 and 4 show the results for velocities ranging from $0.26^\circ \text{ sec}^{-1}$ to $33.3^\circ \text{ sec}^{-1}$. There are indeed conspicuous differences between the cases of movement along the strip or perpendicular to the strip. For a given velocity no coherent motion at all can be seen if the strip is too narrow except for the low velocities along the strip. These limiting strip widths are plotted in Fig. 5 as a function of the velocity.

Discussion

When the movement is along the length of the strip we find a similar threshold behaviour for the lowest

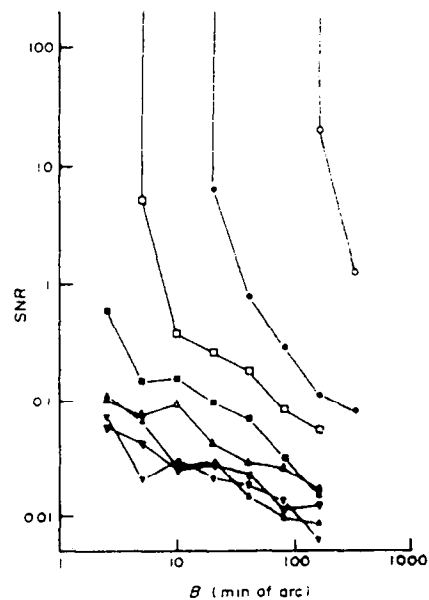


Fig. 3. In this experiment we set $A = D$. The threshold SNR is plotted as a function of B . The velocities were: \blacktriangledown $v = 0.26^\circ \text{ sec}^{-1}$, ∇ $v = 0.52^\circ \text{ sec}^{-1}$, \blacktriangle $v = 1.04^\circ \text{ sec}^{-1}$, \triangle $v = 2.08^\circ \text{ sec}^{-1}$, \blacksquare $v = 4.17^\circ \text{ sec}^{-1}$, \square $v = 8.33^\circ \text{ sec}^{-1}$, \bullet $v = 16.7^\circ \text{ sec}^{-1}$, \circ $v = 33.3^\circ \text{ sec}^{-1}$.

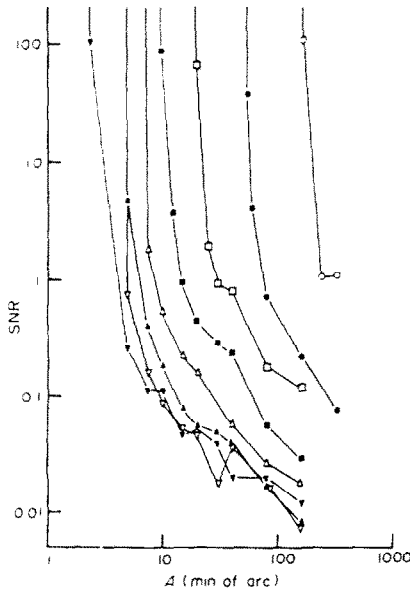


Fig. 4. In this experiment we set $B = D$. The threshold SNR is plotted as a function of A . The symbols are identical to those used in Fig. 3.

velocities. In fact the behaviour is identical within the accuracy of the experiment for velocities of 0.26° to $1.04^\circ \text{ sec}^{-1}$. A regression line in log-log coordinates yields

$$\text{SNR} = \left(\frac{B}{8.3 \times 10^{-3}} \right)^{-0.5}$$

(regression coefficient -0.92). Thus the assumption of simple pooling is well borne out. The square root behaviour can be detected even for velocities up to $16.7^\circ \text{ s}^{-1}$. Thus the same kind of pooling appears to apply to a wide velocity range. However, for the highest velocities a minimum strip width is required. This can be understood if we assume that the sub-fields of the Reichardt correlator have equal height and width (δ). When the strip width B is smaller than δ the noisy surround encroaches on the signal for all

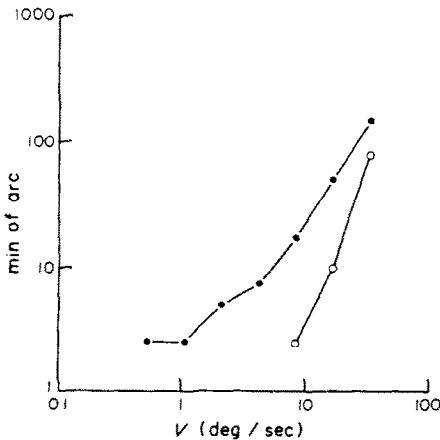


Fig. 5. The critical values of A from Fig. 4 and B from Fig. 3 for which the threshold tends to infinity and here plotted as a function of the velocity.

elementary units. Thus an increased SNR threshold has to be reckoned with whenever $B \leq \delta$.

Simple geometrical considerations show that the number of stimulated pairs equals $N = DB\delta^2$. Thus theory leads us to expect that the SNR threshold should equal $[B/(\delta^2/D)]^{-0.5}$. Comparison with the regression line reveals $\delta = 1.6'$. This value is comparable to the value of Δ found in previous experiments (van Doorn and Koenderink, 1982b). It appears that the sub-fields of the Reichardt correlator must be nearly contiguous or even overlapping.

In the case where the patterns move perpendicular to the direction of the strip we find again that the results for the velocities from 0.26° to $1.04^\circ \text{ s}^{-1}$ merge. A regression line for the data for stripwidths in excess of $10'$ is

$$\text{SNR}_{\text{threshold}} = \left(\frac{A}{0.45} \right)^{-0.8}$$

(regression coefficient -0.92). Thus the simple pooling idea is not borne out here. In fact if we wish to stick to this notion we have to assume that the number of stimulated pairs grows with the 1.6th power of the strip width. This appears to be the case for the whole velocity range. Odd as this may seem at first sight, it is understandable if we assume that the elementary correlators occur in a continuous spectrum of Δ , T values. In that case extra pairs with large separations are progressively recruited as the strip width is increased.

In this case we find a minimum strip width for all velocities. Figure 5 shows the minimum strip widths as a function of the velocity. For velocities up to $8.33^\circ \text{ s}^{-1}$ the regression line yields

$$A = 3.1v^{0.7}$$

(regression coefficient 0.96), (A in minutes of arc, v in degrees per second).

This is in excellent agreement with the dependence of Δ on the velocity found in previous experiments (van Doorn and Koenderink, 1982b): $\Delta = 4.2v^{0.6}$ (Δ in minutes of arc, v in degrees per second). Thus the minimum strip width equals the separation of the Reichardt machine, which seems intuitively reasonable. However, for the highest velocities the minimum strip width rises much faster with the velocity than Δ does.

COHERENT MOTION IN RECTANGULAR AREAS OF VARIOUS SHAPES

In the previous experiment it appeared that the concept of simple pooling works well except for the fact that there seems to be recruitment of units of different size. In this experiment we vary the shape of the rectangular window (A/B) for constant area (AB) in order to see if there is any optimal shape. From the theory we can predict the SNR as a function of the shape as follows. From geometrical considerations we

find that the number of stimulated submechanisms equals approximately

$$\left(\frac{A - \Delta}{\delta}\right) \left(\frac{B}{\delta}\right).$$

Thus

$$\text{SNR}_{\text{threshold}} = \frac{C}{\sqrt{1 - \frac{\Delta}{A}}}$$

with

$$C = \frac{\delta}{\sqrt{AB}}.$$

This experiment was done with the velocity $2.08^\circ \text{sec}^{-1}$, for which the pattern shifts by exactly the width of one pixel spacing every new frame.

Results and discussion

We used areas subtending 800, 400, 200 and 100 square minutes of arc. Figure 6 shows the threshold SNR as a function of the shape parameter A/B . The results are unequivocal. The theoretical prediction is well borne out. For the largest area (800 square min of arc) we empirically determine $C \approx 10^{-1}$ from which we find $\delta \approx 2.8'$. This is in excellent agreement with the previous experiment: the curve in Fig. 3 for the same velocity as in the present experiment ($2.08^\circ \text{sec}^{-1}$) can be described with regression line $\text{SNR} = (B/0.22)^{-0.5}$, from which we find $\delta = 2.6'$.

The point where the SNR threshold in Fig. 6 goes to infinity provides a measure of Δ . For the areas of 800 and 400 square min of arc we find $\Delta = 4.9'$, for the two smaller areas (200 and 100 square min of arc) $\Delta = 10'$. These values are somewhat larger than the values determined in previous experiments (van Doorn and Koenderink, 1982b): $\Delta = 2.7'$, although the discrepancy is not large for the largest fields.

From this experiment it is clear that the lowest threshold for constant target area is obtained when

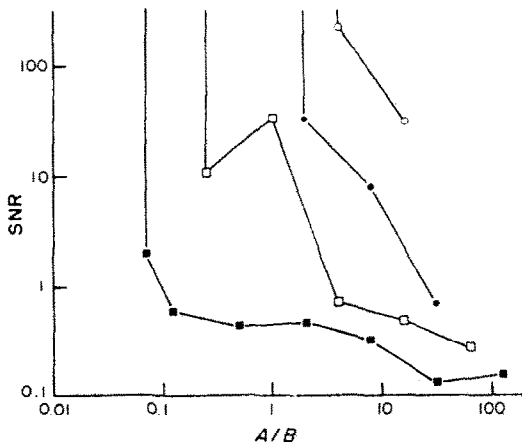


Fig. 6. The SNR thresholds for rectangular targets with fixed area AB arc plotted as a function of the shape parameter A/B . This experiment was done with $v = 2.08^\circ \text{sec}^{-1}$. Four values of the area were used: ■ 800 square min of arc, □ 400 square min of arc, ● 200 square min of arc, ○ 100 square min of arc.

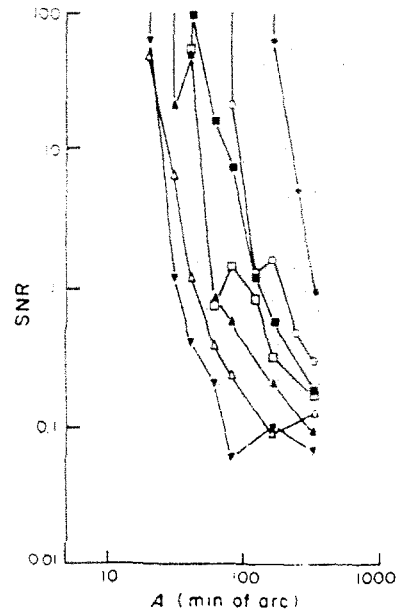


Fig. 7. The target is a thin strip elongated in the direction of movement. The strip width is $2.5'$ for the velocity range $0.26^\circ \text{sec}^{-1} \leq v \leq 2.08^\circ \text{sec}^{-1}$, $5'$ for $v = 4.17^\circ \text{sec}^{-1}$, $10'$ for $v = 8.33^\circ \text{sec}^{-1}$, and $20'$ for $v = 16.7^\circ \text{sec}^{-1}$. These widths were the minimum of B for which the experiment was possible at all (movement invisible in long strips of smaller width). The threshold SNR is plotted as function of the length A of the strip. The symbols denote the various velocities: ▼ $v = 0.26^\circ \text{sec}^{-1}$, △ $v = 0.52^\circ \text{sec}^{-1}$, ▲ $v = 1.04^\circ \text{sec}^{-1}$, □ $v = 2.08^\circ \text{sec}^{-1}$, ■ $v = 4.17^\circ \text{sec}^{-1}$, ○ $v = 8.33^\circ \text{sec}^{-1}$, ● $v = 16.7^\circ \text{sec}^{-1}$.

the target is a narrow strip extended in the direction of motion.

THE MINIMUM AREA IN WHICH COHERENT MOTION CAN BE DETECTED

The previous experiment showed that the lowest thresholds are obtained for narrow targets elongated in the direction of motion. In the present experiment we choose targets which are as narrow as possible (while still leading to the percept of coherent motion) and we vary the length of the target in order to determine the minimum area for which the detection of coherent motion can be obtained. For the lowest velocities (the range from 0.26 to $2.08^\circ \text{sec}^{-1}$) we choose $B = 2.5'$; for $v = 4.17^\circ \text{sec}^{-1}$, $B = 5'$; for $v = 8.33^\circ \text{sec}^{-1}$, $B = 10'$ and for $v = 16.7^\circ \text{sec}^{-1}$, $B = 20'$.

Theoretically we would expect that the shortest length for which an impression of coherent motion is possible should equal Δ , as in the first experiment (Fig. 4).

Results

Figure 7 shows the results. The threshold SNR for the total range of velocities is plotted as a function of the length of the strip. Only for the lowest velocities do we find a limiting length beyond which further elongation does not lead to a lower threshold. For all

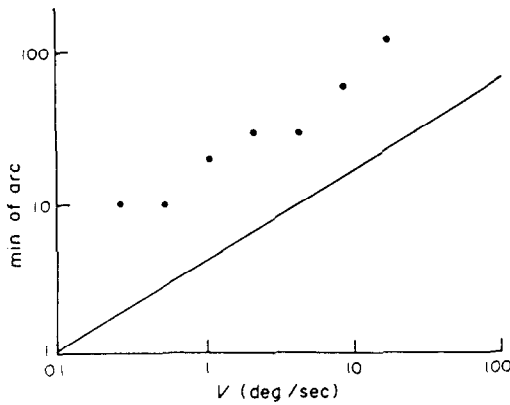


Fig. 8. The critical length of the narrow strip for which the threshold SNR in the experiment shown in Fig. 7 tends to infinity plotted as a function of the velocity (points). A regression line through the points is $A = 18v^{0.6}$ (regression coefficient 0.97, regression line not drawn). The drawn line is the dependence of Δ on v as determined in an earlier (independent) experiment: $\Delta = 4.2v^{0.6}$. (A and Δ in min of arc, v in deg sec^{-1}). Clearly the critical length is proportional to Δ .

curves the threshold SNR depends strongly on the length of the strip. Figure 8 shows the shortest lengths for which an impression of coherent motion is possible at all.

Discussion

The regression line through the points plotted in Fig. 8 is

$$A_{\min} = 18v^{0.6}$$

(regression coefficient 0.97), (A_{\min} in minutes of arc, v in degrees per second).

The straight line plotted in the figure shows the dependence of Δ on the magnitude of the velocity determined in a previous experiment (van Doorn and Koenderink, 1982b): $\Delta = 4.2v^{0.6}$ (Δ in minutes of arc, v in degrees per second).

Obviously the dependence of the minimum length and of Δ on the velocity are identical: a 0.6 power-law. But it is found that the minimum strip length must be 4.3 times the (average) value of Δ for an impression of smooth movement to be obtained. In this respect the result of this experiment differs from the result of the first experiment. Obviously the width of the strip is of importance here. This was also found in the previous experiment where the minimum length of the strip was found to increase with diminishing surface area.

DETECTION OF COHERENT MOTION IN SQUARE FIELDS WITH RESTRICTED PRESENTATION TIME

In the previous experiments the subjects could take as long a time as they wished to set the threshold. But obviously the presentation time must be an important variable too because the theoretical number of available pair correlations is proportional to it. In fact, asymptotically we expect the SNR to vary inversely proportionally with the square root of the

product of the target area and the presentation time. It is to be expected that effects of recruitment would confound this simple relation. However, it is of interest to check whether a trade-off between target area and presentation time is obtained. This experiment was again performed with the velocity of $2.08^\circ \text{sec}^{-1}$ (a single pixel shift per frame).

Results

Figures 9 and 10 show the threshold SNR as a function of the side A of the square target fields and the presentation time T for the velocity $v = 2.08^\circ \text{s}^{-1}$. In Fig. 11 contours of equal threshold SNR are drawn (by linear interpolation) in the (A , T) plane. For the presentation times used (up to 1280 msec) we do not find a saturation value for the threshold SNR. Instead there appear to be minimum values for the size of the target and the presentation time below which no perception of coherent motion is possible, whereas in the other regions we find a dependence somewhat stronger than inversely proportional with either size or presentation time.

Discussion

Clearly there does exist a trade-off between target area and presentation time. In fact, for large target areas presentation times down to 30 msec can lead to a perception of movement. This is somewhat amazing in view of the fact that the (average) value of the delay of the Reichardt correlators as measured in previous experiments (van Doorn and Koenderink, 1982a) was 66 msec, for theory obviously predicts

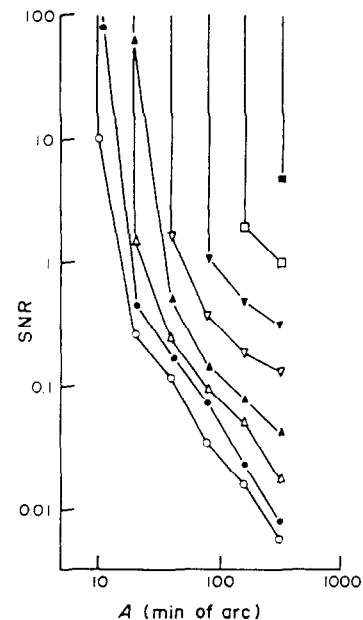


Fig. 9. Square target fields of side A ($= B$) were presented for T msec in a noisy background. SNR thresholds for the impression of smooth movement are plotted as a function of A for the velocity $v = 2.08^\circ \text{sec}^{-1}$. Parameter is the presentation time: \circ $T = 1280$ msec, \bullet $T = 640$ msec, \triangle $T = 320$ msec, \blacktriangle $T = 160$ msec, ∇ $T = 80$ msec, \blacktriangledown $T = 60$ msec, \square $T = 40$ msec, \blacksquare $T = 30$ msec.

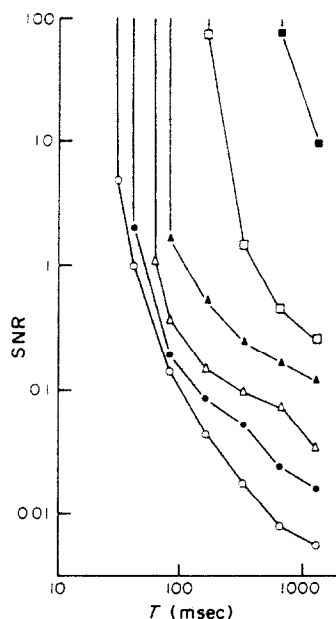


Fig. 10. Square target fields of side $A (=B)$ were presented for T msec in a noisy background. SNR thresholds for the impression of smooth movement are plotted as a function of T for the velocity $v = 2.08^\circ \text{sec}^{-1}$. Parameter is the target width: \circ $A = 320'$, \bullet $A = 160'$, \triangle $A = 80'$, \blacktriangle $A = 40'$, \square $A = 20'$, \blacksquare $A = 10'$.

that detection of motion should be impossible for phenomena shorter than the delay time. This may be taken as an indication that there exists a continuous spectrum of (Δ, T) values, a notion that was also suggested by the recruitment effect described in the first experiment. Moreover, the pure delay in the Reichardt machine is of course an idealization. In practice, a leaky integrator appears physiologically more reasonable and this would automatically entail a finite response of the mechanism for a range of velocities. In previous experiments we have indeed obtained abundant evidence for an appreciable tuning width (van Doorn and Koenderink, 1982c, 1983).

The theoretical prediction that the SNR would depend primarily on the product of target area and presentation time is not borne out. In fact, in the low threshold region we have (very roughly) a dependence

$$\text{SNR}_{\text{threshold}} = \frac{1}{\frac{AT}{3000} - 1}.$$

This may be due to recruitment, but to ascertain this a further study is necessary. For the size of the target field we find a hard limit of $10'$, which is four times larger than the theoretical limit (Δ). This is in agreement with the previous experiment. Even for $T = 30$ msec (3 successive frames) an impression of coherent motion is possible.

This result can be partly compared with that of a study by Morgan and Ward (1980) who determined the parameter region for which an impression of coherent motion is obtained when successive pairs of

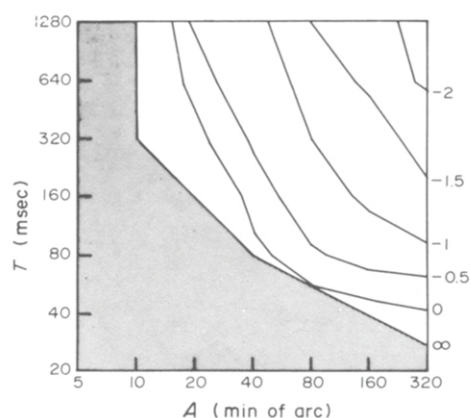


Fig. 11. Contours of equal SNR threshold for the data presented in Figs 9 and 10 are here plotted in the parameter (A, T) plane. The contours are obtained through linear interpolation. The contour $\text{SNR} = \infty$ is the boundary of the shaded area that contains the parameter values for which no impression of coherent motion is obtainable at all.

random dot patterns are presented with the spatial shift and the interstimulus times as variables. For the shift used in our experiment ($1.25'$) their "motion identified" region extends from interstimulus times of 5–60 msec. In our case the frame repetition period was 10 msec, and thus well within their "motion identified" region.

CONCLUSIONS

This study has yielded several new results. We have shown that the detection of coherent motion is possible when only a very few basic correlator units are stimulated, whereas the performance improves when the number of stimulated units goes up to 10^4 or more. We have identified cases in which the threshold SNR varies in accordance with a simple pooling model. However, we also have found evidence of the recruitment of units of different size. In general, the minimum target area in which coherent motion can be detected is obtained when the stimulus is presented in the form of a narrow strip elongated in the direction of motion. The minimum width and length of the strip are connected with the geometrical properties of the Reichardt correlation units as obtained from independent experiments.

These experiments have yielded some results which we have tentatively explained by means of the hypothesis of a continuous spectrum of correlation units, from which units can be progressively recruited when presentation time or target size are varied. This notion certainly merits further study. The problem is a complicated one because the individual units are certainly not sharply tuned, but respond to a certain restricted range of velocities. In order to be able to predict the course of SNR thresholds as a function of spatiotemporal variables it would be necessary to know the tuning properties and the density of correlators with different parameters that are stimulated on a supra-threshold level.

Acknowledgement—The authors are indebted to the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) for financial support of a part of this investigation.

REFERENCES

- Doorn A. J. van and Koenderink J. J. (1982a) Temporal properties of the visual detectability of moving spatial white noise. *Expl Brain Res.* **45**, 179–188.
- Doorn A. J. van and Koenderink J. J. (1982b) Spatial properties of the visual detectability of moving spatial white noise. *Expl Brain Res.* **45**, 189–195.
- Doorn A. J. van and Koenderink J. J. (1982c) Visibility of movement gradients. *Biol. Cybernet* **44**, 167–175.
- Doorn A. J. van and Koenderink J. J. (1983) Detectability of velocity-gradients in moving random-dot patterns. *Vision Res.* **23**, 799–804.
- Lappin J. S. and Bell H. H. (1976) The detection of coherence in moving random-dot patterns. *Vision Res.* **16**, 161–168.
- 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**, 311–438.