PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

On the Elementary Mechanism Underlying Secondary Motion Processing

Johannes M. Zanker

Phil. Trans. R. Soc. Lond. B 1996 **351**, 1725-1736
doi: 10.1098/rstb.1996.0154

On the elementary mechanism underlying secondary motion processing

JOHANNES M. ZANKER*

Max-Planck-Institut für biologische Kybernetik, Spemannstr. 38, D-72076 Tübingen, Germany

SUMMARY

The movement of luminance-defined targets can be easily extracted by elementary motion detectors (EMDs) of the correlation type which often are referred to as Reichardt-detectors. In contrast to such 'primary motion', in 'secondary motion' the moving target is defined by more complex features, like changes in texture, flicker, or local contrast. Such stimulus attributes have to be extracted from the retinal intensity distribution by some nonlinear preprocessing, before they are fed into motion detectors. An intriguing case is the perception of the movement of the motion signal, as is present in theta motion, where an object moves in a different direction than the texture on its surface. A two-layer model of hierarchically organised EMDs has been postulated to account for such motion extraction. Other than for the first layer, the computational nature of the mechanism underlying motion processing in the second layer so far is a matter of speculation, and is therefore characterized here by means of computer simulations and psychophysical experiments. Random dot kinematograms were generated in which sinusoidally modulated vertical dot motion defined gratings, and coherence thresholds were measured for the direction discrimination of a horizontally travelling modulation function. This was done for a variety of spatial frequencies and speeds of the modulation sinusoid. Thresholds turn out to be lowest not for a particular speed, but for a fixed temporal frequency of the modulation function (about 1 cycle per second), when various combinations of fine and coarse, and fast and slow secondary gratings are tested. This result favours a correlation-type mechanism over a gradient-type scheme which should lead to a speed-optimum independent of spatial frequency.

1. INTRODUCTION

Because the visual processing of motion has outstanding relevance for biological and artificial systems, it has been extensively investigated both theoretically and experimentally. There are a number of model proposals for detecting the displacement of a luminance distribution in two-dimensional images, accounting for various aspects of animal behaviour and human perception (Egelhaaf & Borst 1993). One particular scheme, the so-called *elementary motion detector* of the correlation type (EMD) which is often referred to as a *Reichardt-detector*, was originally based on the analysis of insect behaviour (Reichardt 1961), but later gained considerable influence in the psychophysical literature (Sekuler *et al.* 1990), and also has been applied in computer vision (Anandan 1989; Hildreth & Koch 1987). This mechanism, which is based on autocorrelation of the spatiotemporal intensity distribution, can assume various forms. Besides the purely algorithmic formulation it can be interpreted in terms of logical interactions as proposed independently by Barlow & Levick (1965), or it can be implemented by combining spatial and temporal filters in the so-called *energy model* (Adelson & Bergen 1985) which can be formally equivalent to an appropriate version of the Reichardt-detector (van Santen & Sperling 1985). A block diagram of the basic model structure is sketched in the inset left of figure 1*a*: the (possibly spatially filtered) signals from two input elements interact in a nonlinear way, for instance in form of a multiplication, after one of the signals is temporally filtered with respect to the other, for instance by sending the signal through a lowpass filter; finally the multiplication result of this subunit is subtracted from that of an antisymmetrical unit using the same inputs, to increase directional selectivity (for review, see Borst & Egelhaaf 1989). For image processing in artificial systems, on the other hand, often another model is applied, the so-called *gradient scheme*, which formally derives image speed from the ratio of the spatial and temporal derivative of the luminance distribution (Fennema & Thompson 1975; Limb & Murphy 1975; Marr & Ullman 1981). A block diagram of a simple realization of this scheme in terms of discrete operations is sketched in the inset left of figure 1*b*: in the example the spatiotemporal intensity distribution is spatially processed by a Gabor filter, and the temporal derivative of this signal is divided by the spatial derivative, which is increased by a small constant ϵ, in order to prevent zero denominators (adapted after Buchner 1984; Hildreth & Koch 1987).

* Present address: Department of Psychology, University College London, Gower Street, London WC1E 6BT, U.K.

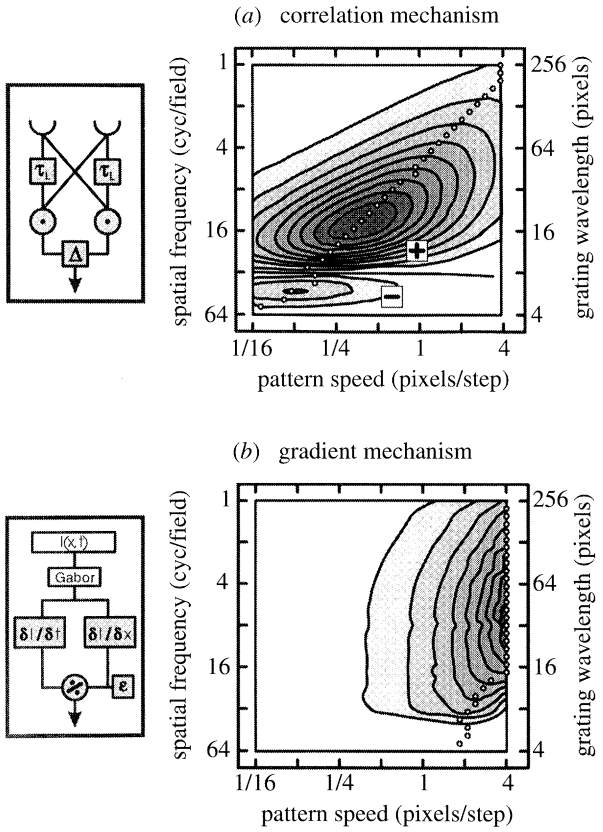

Figure 1. Motion detector model predictions for luminance-defined gratings. The average output (of a single detector) for one stimulus cycle is given in a grey-scale contour plot (lines indicate steps of 10% of maximum response; strong responses are indicated by dark regions) for sinewave gratings of variable speed (plotted horizontally, in pixels displacement per simulation step) and grating wavelength (plotted vertically, in pixels per stimulus cycles); the response maximum for each wavelength is shown by a dot. (*a*) The correlation model consists of 2 input elements (indicated in the inset model sketch by semicircles; Gaussian shaped input sensitivity functions with halfwidth of 2.4 pixels, separated by 4 pixels), 2 temporal filters (first-order lowpass indicated by boxes labelled 'τ_L', time constant 8 simulation steps), two multiplication units (nonlinear interaction circles labelled '●') and a subtraction stage (box 'Δ'). Gratings with larger pattern wavelength elicit maximum response at higher speeds, leading to an oblique optimum response line corresponding to a constant temporal frequency. (*b*) In the gradient model the temporal derivative (box '$\delta I/\delta t$', approximated by differences between successive simulation steps) and spatial derivative (box '$\delta I/\delta t$', approximated by differences between neighbouring points) of the bandpass-filtered (indicated by box 'Gabor', centre-diameter 8 pixels) intensity distribution (indicated by box '$I(x,t)$') are divided (nonlinear interaction circles labelled '÷'), and a small constant ϵ (0.01) is added to prevent zero values in the denominator. This model by definition leads – within certain limits – to the strongest responses at the highest image velocities.

During the last decade evidence has accumulated that humans are not only able to detect the motion of contours which are directly defined by luminance, and which thus is called *primary motion*, or *Fourier motion* because the perceived direction is predicted from the spatiotemporal Fourier transform. Human observers can also perceive *secondary motion*, which is the movement of contours defined by other features like changes in the local contrast, the texture, or flicker (Cavanagh & Mather 1989). Such stimuli are interesting because in the spatiotemporal average the Fourier components of the stimuli do not correspond to the perceived displacement. Some nonlinear preprocessing is required for such patterns before the directional information can be computed, for example by Reichardt detectors. Accordingly, such motion stimuli, to provide a distinction with luminance-defined Fourier motion are referred to as *non-Fourier motion* (Chubb & Sperling 1988). When the displacement of motion-defined contours is considered, as will be explained below, this term can be misleading and will thus be avoided in the present context. In a particular case of secondary motion, which has been termed 'theta motion' (Zanker 1990), the borders of a moving object are defined by discontinuities of the primary motion signal, having an independent direction: when, for instance in a random dot kinematogram, an object is moving to the right which only can be discriminated from the background because the dots on its surface are moving to the left, primary and secondary motion information point in opposite directions; human subjects can nevertheless clearly identify the direction of the object's motion. A two-layer model has been proposed in order to account for the detection of theta motion (Zanker 1993), in which a network of EMDs feeds into a second layer of EMDs and thus extracts the motion of the motion signal. By means of computer simulations it was shown that this model structure is not only sufficient to account qualitatively for the basic phenomenon, but also predicts the psychophysical results in a quantitative way.

In these simulations, a correlation-type EMD was assumed to underlie motion computation in both layers of the model, mainly to keep the model as simple as possible. Is this decision supported by empirical data? One line of evidence will be followed here, namely the characteristic dependence of the EMD output on the speed and spatial structure of the stimulus predicted from the autocorrelation (Varjú 1959; Götz 1964; Buchner 1984), which is illustrated in figure 1 by means of computer simulations (the simulation methods will be fully explained below). As might be expected for a reasonable motion detector, the EMD response depends upon pattern velocity. When sine grating stimuli are tested systematically in the simulations, it becomes obvious that for a given grating the output of the EMD rises to a maximum with increasing speed, and is reduced when the speed is further increased. When the wavelength λ of the grating (the reciprocal of its spatial frequency) is varied in addition, the optimum speed is shifted from low speeds at low λ, i.e. fine gratings, to high speeds for coarse gratings (see figure 1*a*). The oblique line defined by the optimum response in the wavelength-speed plot corresponds to a constant temporal frequency of the stimulus, meaning that the EMD responds best when a fixed number of grating cycles passes the inputs in a given time interval. There is abundant evidence that this feature is characteristic of human perception of moving luminance defined

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

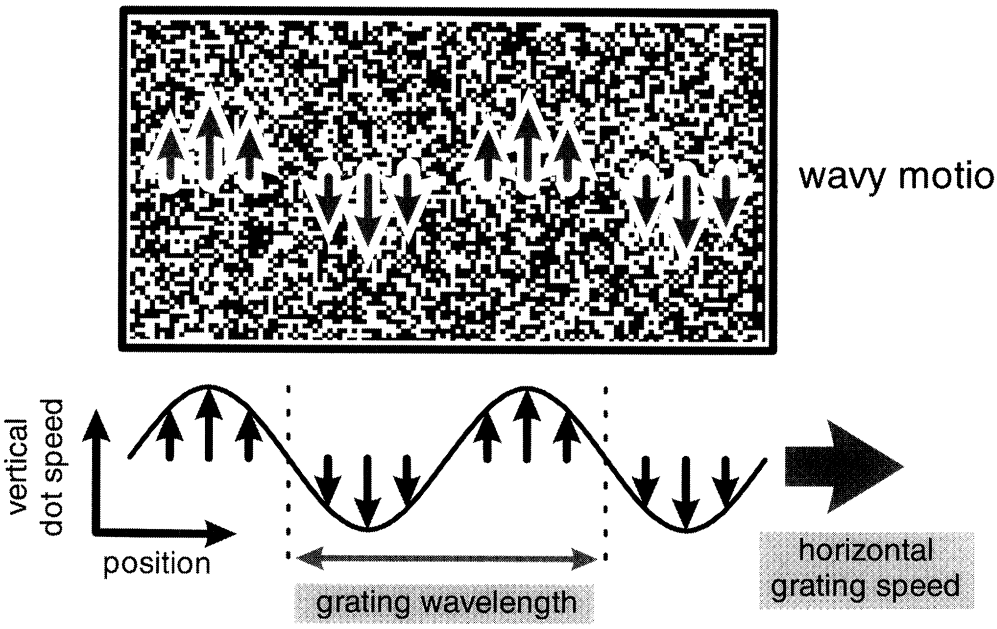

Figure 2. Schematic sketch of the wavy motion stimulus in which a grating defined by vertical dot motion is travelling horizontally. The basic pattern is made of randomly distributed black and white dots (shown in big panel in two space dimensions), which are moving upwards or downwards (indicated by grey vertical arrows). When vertical dot speed is modulated according to a sine function (shown at the bottom) the wavelength λ of this motion-defined grating can be varied. The vertical speed modulation function is displaced horizontally at variable grating speeds v (large grey arrow). Although the dots are moving exclusively up or down, providing a (primary) motion energy only in vertical direction, the horizontal (secondary) motion of the modulation function is clearly perceived by human observers.

gratings (e.g. Watanabe *et al.* 1968; Anderson & Burr 1985), and it is widely agreed that correlation type models can account for the detection of primary motion (for overview, see van Santen & Sperling 1985; Borst & Egelhaaf 1989; Sekuler *et al.* 1990). This particular feature of the correlation type model holds for other model variants, like the motion energy model (Adelson & Bergen 1985), but it clearly differs from the predictions of a gradient type model. The mathematical formulation of the gradient model yields an unambiguous representation of true image velocity in a purely formal sense, dividing the spatial by the temporal gradient of a stimulus. Indeed, this characteristic feature of the gradient scheme is produced even by a rather crude implementation of the model, like that depicted in figure 1*b*, showing that the response increases steadily with speed for almost all grating wavelengths. According to this theoretical prediction, illustrated by the computer simulation, the speed optimum should be largely independent of the spatial frequency of the stimulus grating. One might still argue that introducing more realistic assumptions about temporal properties of biological filters into the model might change the output profile. But it has to be remembered, on the one hand, that using a variety of temporal input filters does not affect the basic results for a wide range of conditions (Srinivasan 1990) and, on the other hand, that the natural dynamical limitations of the human visual system (Kelly 1972) become relevant at much higher temporal frequencies, well above 10–20 Hz, than those considered here (around 1 Hz). If such amendments were to be introduced into the gradient model framework, however, one would have to be aware that the original characteristic feature of the model, and thus the basis of distinguishing separate model proposals, would be compromised. Therefore in the context of the present paper, the somewhat 'idealized' scheme of a gradient detector was used as the alternative to a correlation mechanism.

A correlation mechanism was used previously for both the first and the second layer (Zanker 1993), although this assumption was not based on experimental evidence as far as the second layer is concerned. Knowing that a correlation mechanism can be clearly identified by the prediction of a temporal frequency optimum, this assumption will be tested here. On the one hand it will be shown, by means of computer simulations of a two-layer network, that the predictions formulated for the first layer stimulated with luminance-defined patterns will hold accordingly for the second layer responding to motion-defined targets. On the other hand the task is to provide experimental evidence that human motion perception in fact shows a temporal frequency optimum for secondary motion stimuli. For these two purposes a new variant of the theta-motion stimulus will be used, which may be called 'wavy motion' because the subjects observing the computer-generated kinematograms report the sensation of waves sweeping across the screen. This stimulus exploits the fact that object motion can be detected in random dot kinematograms not only when the dots on the object surface are moving in the same direction as the object (Fourier motion), or in the opposite direction (as is the case in the original version of theta motion), but also when they are moving in an direction orthogonal to the object displacement (Zanker & Hüpgens 1994). The direction of vertical

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

dot motion in a random dot kinematogram can thus be modulated with a sine function which itself is moving horizontally (see figure 2). The essential feature of this wavy motion stimulus is the fact, that the dots are exclusively moving in the vertical direction, and that therefore the primary motion signal is required as the input to a second layer of motion detectors to extract the secondary movement of the motion modulation function.

This wavy motion stimulus makes it possible to vary systematically the spatiotemporal parameters of secondary motion by changing the wavelength and horizontal speed of the modulation function while keeping the primary motion parameters, namely dot size and vertical speed amplitude constant. For this type of stimulus, computer simulations and psychophysical experiments can answer the question of whether a speed or a temporal frequency optimum is found when speed and spatial frequency are varied in such a second order motion stimulus. Part of the results have been published as a conference abstract (Zanker 1994).

2. SIMULATIONS

Computer simulations (IBM workstation Risc 6000, programs in C) were used to check whether the predictions derived from the response of a single EMD to continuously moving luminance gratings will hold for the more complicated situation of a two-layer model where a motion signal originating from a network of EMDs is fed into a second network of motion detectors. This is a non-trivial question because the output of the first-layer EMDs will exactly reflect the profile of the vertical dot displacement only for quasi-linear partitions of the detectors' speed-response characteristic, and random dot kinematograms – being composed of a limited set of discrete displacements – are a highly transient and noisy stimulus. The simulations closely matched the stimuli used in the psychophysical experiments. The basic units of time and space in the simulations were the time step between successive frames of the stimulus sequence, and the size of each picture element in the stimulus arrays.

Each stimulus sequence consisted of 8 frames, each of which was a two-dimensional array (256×256 pixels) of grey values, $S(i,j)$. They were generated from a large random dot pattern, an array of dots (each dot being 2×2 pixels square) which were randomly set to the relative intensity value of 0.1 or 0.9, thus leading to an average intensity of 0.5 relative units, and to 80% contrast. Individual stimulus dots were shifted in the vertical direction following a sinusoid modulation function with an amplitude of v (maximum vertical displacement in pixels per frame). Thus the vertical displacement Δy of each column i of the random dot pattern was calculated for a given frame n according to a sinewave travelling horizontally with a speed of h pixels per frame:

$$\Delta y_n(i) = v \sin((i - nh)\, 2\pi/\lambda),$$

with the sine period λ and the horizontal stepsize h being treated as stimulus variables. For each frame k, each column i was cut from the large random dot pattern at a position

$$(x_k, y_k) = \left(i_0 + i, j_0 + \sum_{n=0}^{n=k} \Delta y_n(i)\right)$$

starting at a random zero position (i_0, j_0). Since the vertical positions y_k which assumed non-integer values were rounded to the next integer j, at a displacement step size of 0.5 the dots move by 1 pixel every second frame, for instance.

To illustrate the basic steps of the simulation, the input and output of the two layers of EMDs is shown in figure 3 as two-dimensional representations of the upper half of the stimulus (in this particular case the basic dot size was set to 4×4 pixels, $\lambda = 128$ pixels, $h = 8$ pixels, and $v = 4$ pixels/frame) for two steps of the sequence, namely frame 3 and frame 7. In figure 3*a*, the stimulus is depicted with only 10% of the black dots being plotted in such a way that the two preceding frames are shown as medium grey and faint grey pixels below the current frame, leading to vertical streaks (dot paths fading with age) in the regions where the dots are moving. For more clarity, one such region of upwards dot motion is framed by a thin line which shows how the modulation function is displaced (by 32 pixels) to the right between frame 3 and 7.

The 8 frames of such a stimulus sequence were fed into a two-dimensional array of 256×256 elementary motion detectors of the correlation type (EMDs as characterized in the introduction). To preprocess the signals in a bandpass fashion, DOG filters (Marr & Hildreth 1980; Hawken & Parker 1987) were assumed for the input lines, thus eliminating DC components and high-frequency noise from the input intensity distribution. In the digital simulations two rotational symmetric Gaussians of opposite sign (i.e. one excitatory and one inhibitory) were approximated by weighted sums across 32×32 array elements of the input pattern. Thus a spatially filtered image $F(i,j)$ was calculated for each of the stimulus frames $S(i,j)$ according to the formula

$$F(i,j) = \sum_{x=-16}^{x=16} \sum_{y=-16}^{y=16} S(i+x, j+y) \times G_{ex} \exp(-\tfrac{1}{2}(x^2+y^2)/\sigma_{ex}^2) - G_{in} \exp(-\tfrac{1}{2}(x^2+y^2)/\sigma_{in}^2)]$$

The two standard deviations σ_{ex} and σ_{in} which determine the receptive field size of the input elements, were kept constant at 1 and 2 pixels, leading to a halfwidth of the excitatory centre of the filter of about 2 pixels. Excitation and inhibition were balanced by setting the Gaussian scaling factors G_{ex} and G_{in} to $1/2\pi\sigma_{ex}^2$ and $1/2\pi\sigma_{in}^2$, respectively, thus leading to zero filter output for a homogenous input.

Pairs of inputs at the array positions (i_A, j_A) and (i_B, j_B) experience a local signal with the time course $F_A(t)$ and $F_B(t)$, which is derived from the spatially filtered input sequence with an enhanced temporal resolution, by setting 8 samples per frame to the input value of the given image. The lowpassed version $L_A(t)$ of the signal

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

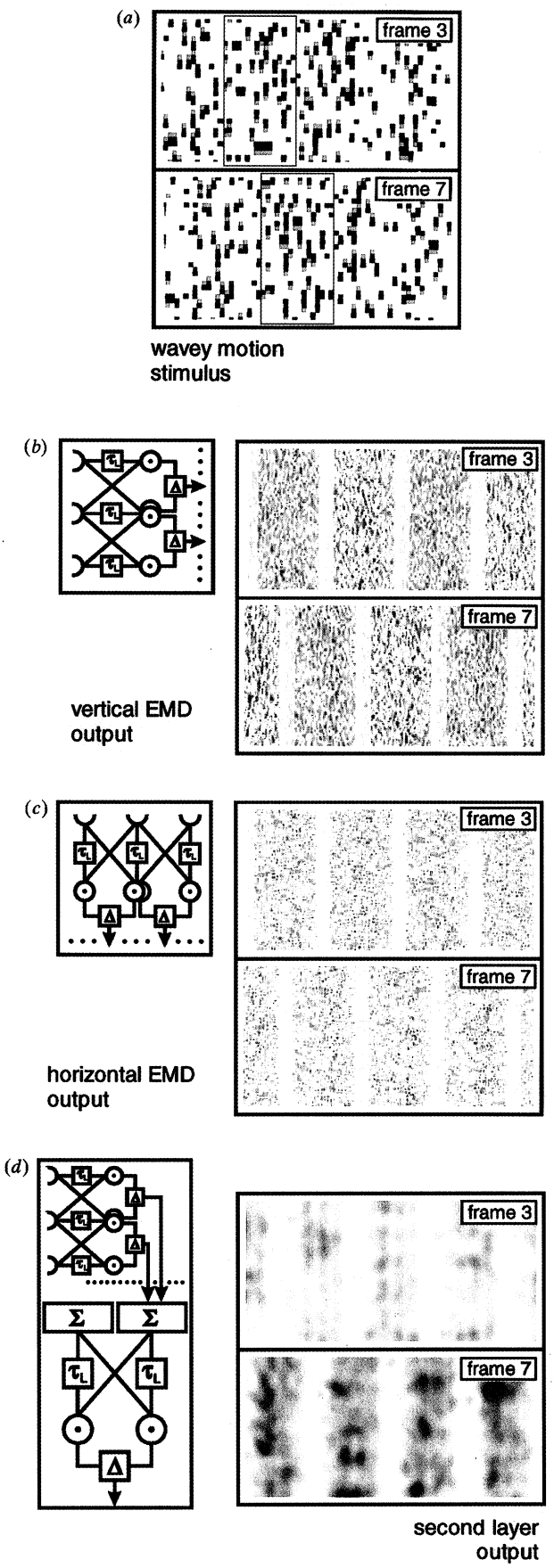

Figure 3. Diagram of the simulation steps of the two-layer model of motion processing. (*a*) 2D sketch of the wavy motion stimulus at two instances of time (frame 3 and frame 7); the position of 10% of the stimulus dots are plotted in black, the positions of the same dots in frame 2 (6) in middle grey, and in frame 1 (5) in light grey; vertical streaks indicate regions of vertical dot motion, the thin line frames a half-cycle (upwards moving dots) of the sine function of motion modulation, which is shifted by 32 pixels between frame 3 and 7. (*b*) and (*c*) 2D representation of motion detector output in pseudo-colour code (red, light grey, and green areas corresponding to regions of negative, zero, and positive response, respectively) for arrays of vertically (*b*) and horizontally (*c*) oriented EMDs (indicated by sketches in boxes, conventions as in figure 1); clear stripes of motion responses can be seen for the vertical EMDs. (*d*) 2D representation of the output of the second layer of EMDs (see two-layer model sketch in box) in pseudo-colour code; the dominating green regions of positive response reflect the detection of the motion signal displacement in the wavy motion stimulus by such a model.

(and $L_B(t)$ correspondingly) was calculated as difference quotien

$$L_A(t) = L_A(t-1) + [F_A(t-1) - L_A(t-1)]/\tau.$$

The time constant τ of the first-order lowpass filter was set to 8.0 units of the enhanced temporal resolution. The time course of the local EMD output after the subtraction stage is given by

$$R(t) = F_A(t)\, L_B(t) - F_B(t)\, L_A(t)$$

assuming a perfect balance between the two mirror-symmetrical subunits. Because the exact time course of the response is not relevant in the present context, the final EMD output was calculated as the temporal average of 8 the consecutive samples which are the result of the temporal resolution enhancement of 8 samples for each input frame. The particular position of this output signal within the array $R(i,j)$ was the centre (again, rounded to the next integer) between the two inputs. The multiplication was done for vertically and for horizontally oriented pairs of inputs which were separated by the sampling base of the detector, $\Delta\phi$, of 2 pixels. A border of 16 pixels of the 256×256 array elements was excluded from the calculations following the spatial filtering, to prevent the input filters from covering areas outside of the stimulus arrays. In the middle panels of figure 3 the output of the vertical and horizontal EMD arrays is shown in pseudo-colour code for the two steps of image motion which are displayed on the left side as stimuli. The horizontal motion signal (figure 3*c*) exhibits a rather irregular pattern of locally positive (green) or negative (red) responses. On the other hand, a clear red-green stripe pattern of predominantly positive or negative response emerges in the vertical motion signal (figure 3*b*); this is shifted to the right in the succession of the two image parts displayed in the figure.

The 8 output images of the vertical EMD arrays were then used as input frames for the second array of horizontally oriented EMDs (see model sketch in figure 3). For the purposes of this demonstration, this second layer was only fed by the output of vertical first-layer EMDs carrying a strong motion signal, because no specific assumptions about the interaction of horizontal and vertical motion information should be made in a first simulation. The processing of the array of EMDs in the second layer basically consisted of the same set of calculations as that in the first layer, with a few notable

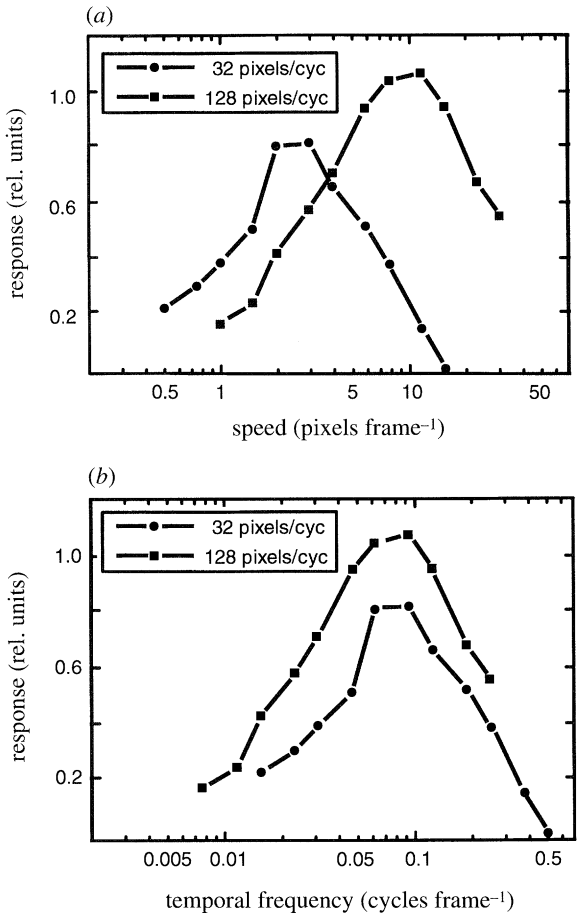

Figure 4. Spatiotemporal averages of the response (ordinates) of the two-layer model (as sketched in figure 3) to gratings of vertical dot motion displaced horizontally with various speeds (*a*) with two grating periods (32 pixels per cycle indicated by dots, 128 pixels per cycle indicated by squares). Coarse patterns elicit optimum response at higher speeds than fine patterns. When the same simulation results are plotted as function of temporal frequency (*b*) the peaks appear coherent.

changes in details introduced to achieve a reliable signal in this operation. (i) The output of a set of 12×12 EMDs of the first layer, centred around a given array element of the second layer was averaged, thus reducing the spatial resolution effectively by a factor of about 6 similar to a spatial lowpass filter, instead of using DOG preprocessing. (ii) The sampling base $\Delta\phi$ was increased to 12 pixels, corresponding to the increased input filter size. (iii) The time constant τ was scaled up correspondingly by a factor of 4, to adjust the sensitivity of the second EMD layer to a similar temporal frequency range as the first layer. In the right part of figure 3, the two-dimensional distribution of the second-layer output is again shown in pseudo-colour code for two time steps, for the upper half of the EMD array. The dominating green regions, in these panels indicate that the second-layer response on average is positive, corresponding to the motion signal travelling from left to right. It generally tends to be stronger and smoother for the later time step (cf. frame 3 output with frame 7 output). Thus, the simulation shows that the two-layer network of EMDs extracts the direction of motion of the wavy motion stimulus.

In a systematic set of simulations, it was investigated, how the output of the two-layer model depends on the wavelength and the speed of the wavy motion stimulus. For this purpose, the output of the second layer of motion detectors was averaged temporally for all 8 frames and spatially across the complete EMD array. This average response was calculated for stimuli with two modulation wavelengths λ (32 and 128 pixels per cycle) and modulation function speeds h ranging between 0.5 and 32 pixels per frame. The dot size was 2×2 pixels, and the amplitude of the sinusoid modulation function ν was 2 pixels per frame. All model parameters were the same as listed in the previous paragraphs. The results of these simulations are shown in figure 4, plotted for immediate comparison as function of speed (*a*) and temporal frequency (*b*). It is immediately clear that the coarse grating of vertical motion (indicated by squares in the plot) leads to a response optimum at higher horizontal speeds of the modulation function than the fine motion grating (dots). When the same data are plotted versus temporal frequency, the two curves are largely covariant, and in particular have their peak at the same temporal frequency. Thus, despite the considerable irregularities of local motion detector responses related to the stochastic nature of random dot kinematograms, a two-layer network of EMDs exhibits a clear temporal frequency optimum instead of a speed optimum, when motion-defined gratings of different spatial frequency are used as inputs to such a model.

3. PSYCHOPHYSICAL EXPERIMENTS

Subjects were seated in a quiet, moderately dark room and were asked to watch a computer monitor (Atari SM 124, driven by Mega ST4) binocularily at a viewing distance of 50 cm. No fixation target was provided. Sets of Random Dot Kinematograms (RDKs) were generated in advance of each experiment and stored in the computer memory, from which later the sequences of 24 images were recalled at a frequency of 50 frames per second (i.e. 20 ms interframe interval, leading to a presentation duration of 480 ms). The average luminance was about 48 cd/m² and the contrast about 98%. The 6 volunteer subjects (3 female, 3 male; aged between 22 and 37 years) who were recruited from the institute members and an undergraduate course, had normal or corrected to normal vision. One of them was the author, the others were mildly experienced psychophysical observers, and all naive to the purposes of the experiment.

Each frame of a given stimulus consisted of an array of 512×128 screen pixels which at the viewing distance of 50 cm subtended 18.1° visual angle in horizontal, and 4.7° in vertical direction. Each screen pixel was set to the bright or dark luminance value at random, leading to a 50:50 random dot pattern. A vertical dot speed was assigned to each vertical strip of this pattern according to a sine function, as described in the simulation section, which varied between 2 dots per frame upwards and 2 dots per frame downwards (corresponding to $\pm 3.7°\ s^{-1}$). The initial phase of this modulation function, i.e. the position of maximum

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

upwards and downwards motion at stimulus onset, was randomised between the stimulus conditions. The speed modulation function was shifted horizontally, either to the left, or to the right, leading to the impression that a smooth wave is sweeping across the screen. The subjects had to detect the horizontal direction of this 'wavy motion', which turned out to be a stable percept despite the fact that the only coherent dot motion, and the corresponding Fourier motion energy admitted to the visual system, was in vertical direction. In the experiments the same two variables were varied as in the simulations (cf. figure 2), namely the wavelength λ of the modulation function, ranging between half a cycle on the screen and 32 cycles on the screen (1024 and 16 pixels per cycle, corresponding to 0.03 cyc/° to 1.7 cyc/°), and the speed h of the modulation function, ranging between one pixel per ten frames and 20 pixels per frame (0.18° to 36° s^{-1}). Before and after the motion stimulus, for an interval of 480 ms dynamic noise (a sequence of totally uncorrelated random dot patterns) was presented. The strength of a given stimulus was varied without changing the basic spatial or temporal parameters, or the luminance contrast, by manipulating the signal-to-noise-ratio of the motion signal, i.e. the number of coherently moving dots of the RDK, thus specifically changing the spatiotemporal correlation which is decisive for motion processing. Noise was superimposed on the previously calculated stimulus sequences by exchanging a certain percentage of screen pixels within each frame in a random fashion, which introduces signal components into the sequence which are totally uncorrelated in time and space (methods described in detail in Zanker 1993).

A set of stimulus sequences were prepared before each experiment in which the superimposed noise was varied between 0 % and 100 % in steps of 10 %. Linear increments of noise percentage were used in order to provide approximately equal differences in subjective signal strength (see Zanker 1995*b*). Subjects had to decide in a 2AFC situation between grating motion to the left or to the right. Coherence thresholds for direction discrimination were measured in a staircase procedure using sets with variable stimulus strength defined by the signal-to-noise ratio. A simple up-down procedure was used in which three correct decisions were needed to increase the amount of superimposed noise, whereas it was decreased after each false decision (Rose *et al.* 1970; Levitt 1971). The subjects climb up the staircase from the starting level of 0 % noise (i.e. the pure motion signal) from which they needed six correct decisions in a row to proceed, in order to make sure that this was seen reliably and is not overcome by chance. When they reach higher noise levels and make mistakes, in this procedure they will make sequences of downwards and upwards steps several times, oscillating around a equilibrium level. The staircase was exited after 4 tops (reversals in the staircase direction from going up to going down), and the equilibrium level was calculated as the average percentage of noise presented for all decisions between the second and the fourth top (inclusively). This value estimates the amount of noise dots just tolerated by the subjects while they still detect the direction of pattern motion correctly in 79 % of the presentations, which is defined as coherence threshold of direction discrimination. The percentage of noise is used directly in the figures as a measure of the sensitivity of the observers to a given motion stimulus.

In a first set of experiments, the wavelength λ of the motion modulation function was varied between the extreme values of 1024 and 16 pixels (33° and 0.59°), corresponding to spatial frequencies between 0.03 and 1.7 cyc/° (0.5 and 32 cycles displayed on the screen). Such gratings were displaced by 1 or 5 pixels between successive frames, corresponding to speeds $h = 1.8$ or 9.1° s^{-1}. The observers were first allowed one or two practice staircases with a highly visible stimulus parameter combination, to get acquainted with the task. Then for each subject one threshold was measured for each of 12 stimulus conditions which were arranged in random order. The average sensitivity for all subjects ($n = 6$), together with the standard error of the mean, is displayed in figure 5, because there were no obvious systematic differences between the individual data. At least in the speed range tested here, the spatial resolution for motion-defined gratings is far below that of luminance-defined gratings; whereas standard spatial resolution of black and white gratings easily goes up to of 30–50 cyc/° (Kelly 1985), in the present case sensitivity clearly drops for gratings above 1 cyc/° (with a minimum period of 16 pixels this is still well above the pixel resolution). This reduced spatial resolution resembles the capability of human observers to detect and discriminate motion direction of other kinds of non-Fourier motion like contrast-modulated random dot patterns (Smith *et al.* 1994). In the model simulation, this feature is reflected by the spatial integration at the input stage of the second EMD layer which is needed to smooth the irregularities of the primary motion signal in a kind of low-pass filtering before it is used for secondary motion extraction.

When the spatial tuning curves measured at the two speeds are compared, it is clear that they are shifted along the spatial frequency axis (see figure 5*a*). For slow motion-defined gratings, the human visual system is sensitive to a range of higher spatial frequencies as compared to fast gratings where lower spatial frequencies are covered by the secondary motion extracting system. To show that the peak sensitivity is not found at the same spatial frequency, the 12 individual data sets each were interpolated by natural cubic splines (Stoer & Bulirsch 1980) for each subject on a logarithmic frequency scale. A curve passing through the mean data points of each of the data sets in figures 5 and 6 is presented to aid illustration. These curves are natural splines which have the conservative property that while being smooth they must pass through the data points. A spline is useful in the absence of any parametric model for the frequency tuning. For the purpose of comparing the data sets, splines were computed for the data sets of each subject. The averages of peak spatial frequencies (0.119 and 0.377 cyc/°) derived from these 12 functions differ by a factor of 3.2 which is less than a factor of 5 expected from a correlation detector for a speed ratio of 5, but it is highly significant different from a factor of 1 (i.e. the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

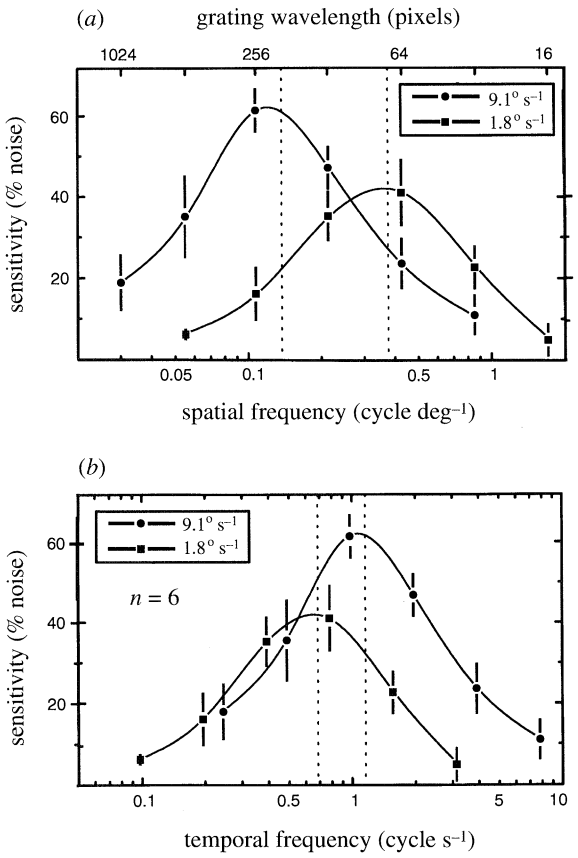

Figure 5. Coherence thresholds in a direction discrimination task (sensitivity given in % noise) to wavy motion stimuli with variable grating spatial frequencies at two grating speeds h (9.1° s^{-1} indicated by dots, 1.8° s^{-1} by squares). Averages of six observers, with spline approximations plotted as thin lines, and SEMs as error bars. When plotted as function of spatial frequency (*a*), highest sensitivity is observed for fast gratings at a lower value (peak average at 0.12 cyc/°, indicated by dashed line) than for slow gratings (at 0.38 cyc/°). The curves have peaks close to each other (peak averages at 1.08 and 0.69 cyc s^{-1}, respectively) when plotted as function of temporal frequency (*b*).

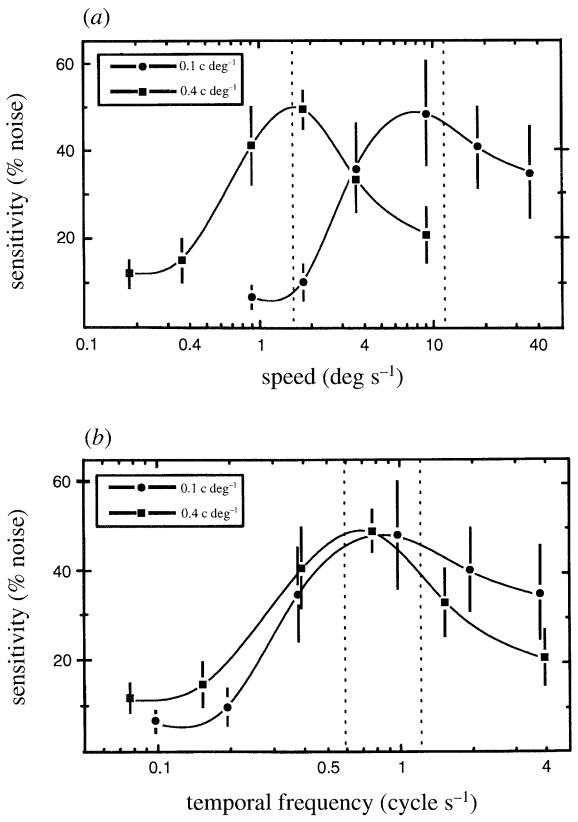

Figure 6. Coherence thresholds in a direction discrimination task to wavy motion stimuli with variable grating speed h at two spatial frequencies (0.4 cyc/° indicated by dots, 0.1 cyc/° by squares). Conventions as in figure 5. When plotted as function of speed (*a*), sensitivity is optimum for fine gratings at a lower value (peak average at 1.37° s^{-1}) than for coarse gratings (peak average at 10.72° s^{-1}). Both curves have maxima at a more similar value (peak averages at 0.58 and 1.14 cyc s^{-1}, respectively) when plotted as function of temporal frequency (*b*).

zero-hypothesis that spatial frequency peaks are the same can be rejected; two-tailed students t-test on peak logarithms, $t_{\text{d.f.}=10} = 6.26$, $p \ll 0.025$) expected from a gradient scheme. However, when the two sensitivity curves are plotted as function of temporal frequency (figure 5*b*), i.e. the number cycles of vertical motion passing by a given position per second, they both have their maximum at similar abscissa values of temporal frequency (1.079 and 0.689 cyc/s^{-1}, not significantly different, $t_{\text{d.f.}=10} = 2.42$, $p > 0.025$). The vertical dashed lines in figures 5 and 6 represent the mean peaks of the individual data sets. As expected these do not exactly line up with the peak of the representative spline on the mean data points shown in these figures, but the fact that the two methods of determining the peak provide similar answers indicates that the spline procedure was reasonable.

In a second set of experiments, the wavelength λ of the motion modulation function was set to two fixed values (64 pixels and 256 pixels). At the viewing distance of 50 cm thus the grating had a spatial frequency of 0.43 cyc/° and 0.11 cyc/°, in the optimum range of the prior experiment. These gratings were displaced at step sizes between 0.1 and 20 pixels per frame, leading to speeds h between 0.18 and 36° s^{-1}. For each of 6 subjects (the same observers as in the first set of experiments) one threshold was measured for each of 12 stimulus conditions, again arranged in random order. Because all observers behaved qualitatively the same, in figure 6 again the average results of all subjects is shown, together with the standard error of the mean, and the natural cubic spline interpolations to the averaged data. The sensitivity to discriminate the direction of a travelling motion signal clearly depends on the speed of the modulation function; motion-defined gratings are difficult to be seen at very high and low speeds, and direction discrimination reaches an optimum performance at intermediate speeds. This speed optimum assumes different values when the spatial frequency of the modulation function is altered; coarse gratings peak at higher speeds (10.72° s^{-1}) than fine gratings (1.37° s^{-1}). When the average of peak velocities was determined from spline interpolation, one subject was excluded from the statistical analysis, because the interpolation procedure did not indicate a peak within the range of stimulus speeds. Therefore a slight mismatch between the peaks

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

averages indicated by dashed lines and the peaks of the splines can be observed in figure 6. The difference in optimum speed is highly significant for the two spatial frequencies ($t_{\text{d.f.}=8} = 7.64, p \ll 0.025$), and the factor of 7.8 is even larger than that that expected for temporal frequency tuning from a spatial frequency factor of 4. When the sensitivity is plotted as function of the temporal frequency of the travelling modulation function, however, the maxima of the two tuning functions for secondary motion sensitivity get close, and the difference between peak frequencies is not significant ($t_{\text{d.f.}=8} = 2.47, p > 0.025$).

4. DISCUSSION

In both sets of experiments, the data do not perfectly meet the expectations from either model in a quantitative sense, probably due to large interindividual variability in the difficult psychophysical task. But the deviations from the values predicted by temporal frequency tuning show no trend in a particular direction, and the statistical test does not allow one to reject this hypothesis in either experiment, whereas the alternative hypothesis, of velocity tuning, can be rejected with high significance in both experiments. Thus the critical experimental variable determining performance for direction discrimination of motion-defined gratings is temporal frequency. The optimum temporal frequency, about 1 cycle per second, is close to the values observed for motion of luminance-defined sine-wave gratings (see Anderson & Burr 1985; Borst & Egelhaaf 1989). In combination with the simulation results and the general predictions for luminance-defined gratings, the temporal frequency tuning found here indicates that a correlation type mechanism should be favoured over a gradient scheme as being the basis for the extraction of secondary motion from a spatiotemporal distribution of primary motion signals, because the latter – at least in its original formulation – would predict a speed optimum largely independent of the spatial structure of the function modulating the motion signal. This suggests that the same elementary process operates on the first and the second stage of the two-layer model which was proposed to account for the extraction of primary and secondary motion (Zanker 1993). What are the limitations for this conclusion?

Only two alternative motion detection mechanisms were compared here, and other possible schemes were not considered so far. For instance, a feature tracking mechanism was proposed for motion detection which identifies discrete features in the images, and matches the positions of these features in successive presentations of a kinematogram (Ullman 1979; Georgeson & Harris 1990). In the context of motion processing beyond luminance-defined Fourier motion, the nonlinear pre-processing postulated to underlie secondary motion extraction can be regarded as some kind of feature extraction preceding the operations to detect the displacement of such features. Indeed, a feature tracking mechanism was at least implicitly suggested to account for various kinds of complex object displacements (Cavanagh *et al.* 1989; Lu & Sperling 1995) and to explain the perceived displacement in beat patterns under certain conditions (Hammett *et al.* 1993). Such a high-level mechanism in particular would be useful to detect the position and displacement of discrete objects. But on the other hand, an operator requiring explicit object definition would be inferior to more automatic processing when stimuli are not discrete objects, but periodic or random patterns. In this case matching the features can become computationally extremely expensive. Data provided in the present paper match the specific predictions of the correlation-type model, but would not be expected from an feature tracking mechanism because there is no reason why the ideal token matching, or an implementation of this mechanism, should change its speed characteristic with the spatial structure of the motion stimulus.

As far as the specificity of the correlation model predictions are concerned, it has to be asked how explicit the speed dependence and pattern independence of the gradient model would be, when possible amendments of the actual implementation are taken into account. In an extension of the original proposition (Limb & Murphy 1975; Fennema & Thompson 1975), a small constant was added to the denominator in the demonstration simulation shown in the introduction, in order to prevent the division by zero (Wang *et al.* 1989). Still, such a 'realistic' gradient model produces the characteristic (and desired) proportionality of the model output with image speed. Elaborated gradient schemes get around the zero division problems by combining sets of units with different spatial and temporal parameters, and/or using quadrature pairs of Gabor filters in the input lines, which are integrated before the division (Johnston *et al.* 1992; Johnston & Clifford 1995). It is not clear, whether this and other elaborations (Srinivasan 1990; Uras *et al.* 1988) respond to variations of speed and spatial frequency as to be expected from the original formulation. Under certain conditions gradient schemes are known to behave very similarly to correlation models, and even to can become formally equivalent (Wang *et al.* 1989). Because in this case the underlying mechanism no longer offers alternative solutions, in the present context the term 'gradient scheme' refers to the original model design, attempting to represent velocity independent of the spatial stimulus structure.

Will the conclusion drawn here for one specific type of secondary motion stimulus hold for other types of motion? Besides moving flicker-defined gratings, similar to drift-balanced (Chubb & Sperling 1988) or μ-motion (Lelkens & Koenderink 1984), other motion-defined gratings have to be tested. Modulating vertical speed around various average speeds and with various amplitudes, for instance, gives access to the effects of changing mean intensity and contrast of the primary motion signal. Modulations of vertical motion without zero crossings, for instance when dots are only moving upwards but with variable speed, will avoid a minor problem in the present experiments. In the wavy motion stimulus applied here, the local temporal frequency is not the same everywhere in the grating, in particular there are thin strips at the zero-crossings of the modulation function in which the dots remain

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

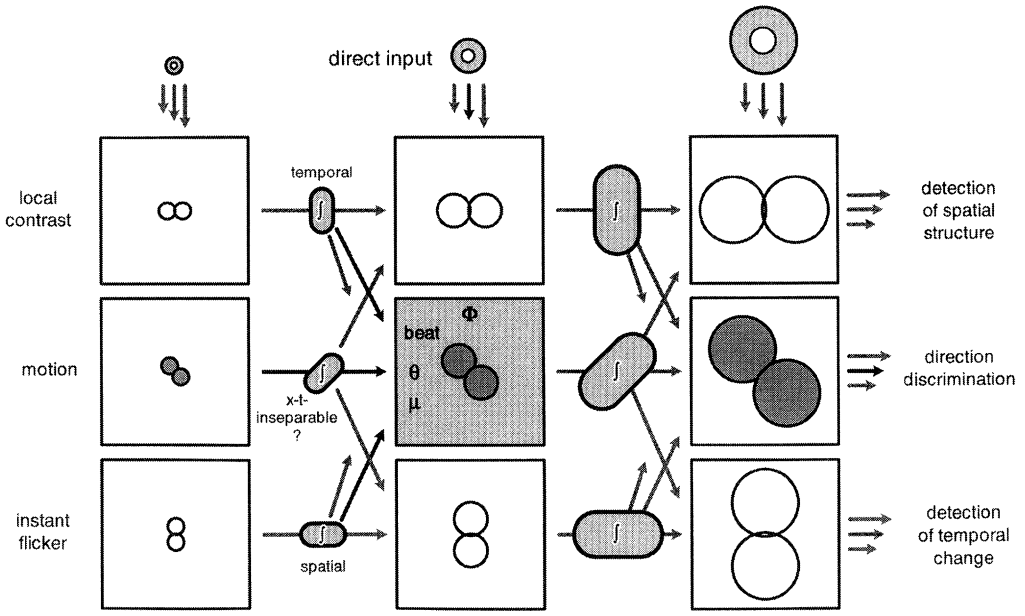

Figure 7. Cascade structure of multiple interactions (arrows) between modules to extract local contrast (first row of boxes), motion (second row), and flicker (third row), which are involved in the visual tasks listed on the right side, respectively. In these space-time boxes the nonlinear interaction between two neighbouring regions in space, space and time, or time is indicated by the pairs of circles which are oriented horizontally, oblique, or vertically; the different circle size indicates the different spatiotemporal receptive field size of the input elements, corresponding to variations in the resolution of the respective modules. A motion detector of given size (e.g. the one indicated by the light grey box) can receive direct input from linear spatial filters of appropriate size (symbolized above the boxes as concentric circles), or after some spatial or temporal integration (ovals labelled by '$\int$'), from higher resolution modules extracting contrast, motion or flicker. These connections account for the detection of Fourier ('Φ'), beat pattern ('beat'), theta ('θ'), or drift-balanced ('μ') motion, respectively (for details, see text).

static. These strips, in principle, could be used to feed motion detectors with a flicker signal instead of a motion signal, and thus would not require two layers of motion processing (cf. Lelkens & Koenderink 1984; Chubb & Sperling 1988). However, the strips of static dots are very difficult to detect in the actual stimuli, and rapidly disappear when noise is superimposed, and subjects consistently reported that they spontaneously tried to concentrate on the peaks and troughs of the vertical motion waves, and not on the borders between the regions of upward and downward motion. When the vertical speed is kept constant, and the signal-to-noise-ratio of vertical motion is modulated by a travelling sine function, it can be furthermore tested whether all directions of primary motion can be utilized for extracting secondary motion (similar to the experiments of Zanker & Hüpgens (1994) for secondary object motion), and how the effective speed ranges in the first and second processing layer are matched to each other. The present study left all these questions open and rather concentrated on one stimulus example to demonstrate the quantitative match between correlation model prediction and experimental results.

Since highly artificial stimuli like theta motion are rather unlikely to be experienced outside the laboratory, one may wonder about the biological significance of extracting the movement of the motion signal. Complex mechanisms of motion processing may be useful when *conventional* stimulus attributes, like brightness or colour, do not provide sufficient contrast to detect objects, for instance in a noisy environment (Braddick 1993). Motion is the decisive visual cue to break texture camouflage in solving the so-called 'figure-ground problem' (Reichardt & Poggio 1979), and tracking of an animal in the shaking and glittering foliage of a tree in the wind is a typical example for spatiotemporal changes of the motion signal distribution conveying the relevant information. The two-layer network of EMDs from a formal point of view extends the model proposed for figure-ground detection, in that it not only detects a motion-defined target, but at the same time extracts its movement (Zanker 1994).

To prevent further complication of the picture, the schematic network in figure 7 is reduced to a skeleton of all the different modules and all the possible interactions the visual system is composed of. Understanding the various links between the modules as fixed connections is another simplification. Adaptation, attentional mechanisms, or experience may be responsible for variations in particular connectivity strengths, thus matching the image analysis to the actual stimulus conditions. Changes *via* top-down processing would explain the results of perceptual learning experiments, in which the sensitivity for a specific motion discrimination task improves with practise (Ball & Sekuler 1982; Zanker 1995*c*). Such plasticity would finally lead to a highly versatile image processing system which can adapt to the momentary task utilizing whatever information source would be available. The crucial point supported by the present findings is that various kinds of input information may be used by the same elementary motion detector. The processing scheme put forward here in some respects

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

resembles an earlier proposal by Wilson *et al.* (1992) who suggest an identical mechanism underlying primary and secondary motion processing, which operates on different inputs and at different spatial scales. After their view received experimental support from primate electrophysiology (Zhou & Baker 1993), it will be interesting to learn how our present and future knowledge on the perception of various kinds of secondary motion may be accommodated in a processing scheme like that sketched above.

I thank A. Borst, O. Braddick, M. Egelhaaf, and A. Johnston for many fruitful discussions on my work and this manuscript, and the anonymous referees for their helpful suggestions. This project was supported by the Max-Planck-Gesellschaft and the EC Human Capital and Mobility Programme (CK 4B4).

REFERENCES

Adelson, E. H. & Bergen, J. R. 1985 Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* A **2**, 284–299.

Anandan, P. 1989 A computational framework and an algorithm for the measurement of visual motion. *Int. J. Computer Vision* **2**, 283–310.

Anderson, S. J. & Burr, D. C. 1985 Spatial and temporal selectivity of the human motion detection system. *Vision Res.* **25**, 1147–1154.

Badcock, D. R. & Derrington, A. M. 1989 Detecting the displacements of spatial beats: no role for distortion products. *Vision Res.* **29**, 731–739.

Ball, K. & Sekuler, R. 1982 A specific and enduring improvement in visual motion discrimination. *Science* **218**, 697–698.

Barlow, H. B. & Levick, W. R. 1965 The mechanism of directionally selective units in rabbit's retina. *J. Physiology* **178**, 477–504.

Borst, A. & Egelhaaf, M. 1989 Principles of visual motion detection. *Trends Neurosci.* **12**, 297–306.

Braddick, O. J. 1993 Segmentation versus integration in visual motion processing. *Trends Neurosci.***16**, 263–268.

Buchner, E. 1984 Behavioural analysis of spatial vision in insects. In *Photoreception and vision in invertebrates* (ed. M. A. Ali), pp. 561–621. New York: Plenum.

Cavanagh, P., Arguin, M. & Grünau, M. W. v. 1989 Interattribute apparent motion. *Vision Res.* **29**, 1197–1204.

Cavanagh, P. & Mather, G. 1989 Motion: the long and short of it. *Spatial Vision* **4**, 103–129.

Chubb, C. & Sperling, G. 1988 Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am.* A **5**, 1986–2006.

Egelhaaf, M. & Borst, A. 1993 Movement detection in arthropods. In *Visual motion and its role in the stabilization of gaze* (ed. F. A. Miles & J. Wallman), pp. 53–77. Elsevier.

Fennema, C. L. & Thompson, W. B. 1975 Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing* **9**, 301–315.

Georgeson, M. A. & Harris, M. G. 1990 The temporal range of motion sensing and motion perception. *Vision Res.* **30**, 615–619.

Götz, K. G. 1964 Optomotorische Untersuchung des visuellen Systems einiger Augenmutanten der Fruchtfliege Drosophila. *Kybernetik* **2**, 77–92.

Hammett, S. T., Ledgeway, T. & Smith, A. T. 1993 Transparent motion from feature- and luminance-based processes. *Vision Res.* **33**, 1119–1122.

Hawken, M. J. & Parker, A. J. 1987 Spatial properties of neurons in the monkey striate cortex. *Proc. R. Soc. Lond.* B **231**, 251–288.

Hildreth, E.-C. & Koch, C. 1987 The analysis of visual motion: From computational theory to neuronal mechanisms. *An. Rev. Neurosci.* **10**, 477–533.

Johnston, A. & Clifford, C. W. G. 1995 Perceived motion of contrast-modulated gratings: predictions of the multchannel gradient model and the role of full-wave rectification. *Vision Res.* **35**, 1771–1783.

Johnston, A., McOwan, P. W. & Buxton, H. 1992 A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond.* B **250**, 297–306.

Kelly, D. H. 1972 Flicker. In *Handbook of Sensory Physiology VII/4 Visual Psychophysics* (ed. D. Jameson & L. M. Hurvich), pp. 273–302. Berlin, Heidelberg and New York: Springer.

Kelly, D. H. 1985 Visual processing of moving stimuli. *J. Opt. Soc. Am.* **A2**, 216–225.

Lelkens, A. M. M. & Koenderink, J. J. 1984 Illusory motion in visual displays. *Vision Res.* **24**, 1083–1090.

Levitt, H. 1971 Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* **2**, 467–477.

Limb, J. O. & Murphy, J. A. 1975 Estimating the velocity of moving images in television signals. *Computer Graphics and Image Processing* **4**, 311–327.

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

Marr, D. & Hildreth, E.-C. 1980 Theory of edge detection. *Proc. R. Soc. Lond.* B **207**, 187–217.

Marr, D. & Ullman, S. 1981 Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond.* B **211**, 151–180.

McCarthy, J., Pantle, A. & Pinkus, A. 1994 Detection and direction discrimination performance with flicker gratings in peripheral vision. *Vision Res.* **34**, 763–773.

Pantle, A. 1992 Immobility of some second-order stimuli in human peripheral vision. *J. Opt. Soc. Am.* **9**, 863–867.

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

Reichardt, W. & Poggio, T. 1979 Figure-ground discrimination by relative movement in the visual system of the fly. *Bio. Cybernetics* **35**, 81–100.

Rose, R. M., Teller, D. Y. & Rendleman, P. 1970 Statistical properties of staircase estimates. *Perception Psychophysics* **8**, 199–204.

van Santen, J. P. H. & Sperling, G. 1985 Elaborated Reichardt detectors. *J. Opt. Soc. Am.* A **2**, 300–321.

Sekuler, R., Anstis, S. M., Braddick, O. J., Brandt, T., Movshon, J. A. & Orban, G. A. 1990 The perception of motion. In *Visual perception. The neurophysiological foundations* (ed. L. Spillman & J. S. Werner), pp. 205–230. San Diego: Academic Press.

Smith, A. T., Hess, R. F. & Baker Jr, C. L. 1994 Direction identification thresholds for second-order motion in central and peripheral vision. *J. Opt. Soc. Am.*, 506–514.

Srinivasan, M. V. 1990 Generalized gradient schemes for the measurement of two-dimensional image motion. *Bio. Cybernetics* **63**, 421–431.

Stoer, J. & Bulirsch, R. 1980 *Introduction to numerical analysis.* New York, Heidelberg and Berlin: Springer.

Ullman, S. 1979 *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Uras, S., Girosi, F., Verri, A. & Torre, V. 1988 A computational approach to motion perception. *Bio. Cybernetics* **60**, 79–87.

Varjú, D. 1959 Optomotorische Reaktionen auf die Bewegung periodischer Helligkeitsmuster (Anwendung der Systemtheorie auf Experimente am Rüsselkäfer Chlorophanus viridis). *Z. Natur.* **14b**, 724–735.

Wang, H. T., Mathur, B. & Koch, C. 1989 Computing optical flow in the primate visual system. *Neural Computation* **1**, 92–103.

Watanabe, A., Mori, T., Nagata, S. & Hiwatashi, K. 1968 Spatial sine-wave responses of the human visual system. *Vision Res.* **8**, 1245–1263.

Wilson, H. R., Ferrera, V. P. & Yo, C. 1992 A psychophysically motivated model for two-dimensional motion perception. *Visual Neurosci.* **9**, 79–97.

Zanker, J. M. 1990 On the detectability of phi- and theta-motion. *Investigative Ophthalmology Visual Sci.* **31**, 519

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

Zanker, J. M. 1994 What is the elementary mechanism underlying secondary motion processing? *Investigative Ophthalmology Visual Sci.* **35**, 1405.

Zanker, J. M. 1995*a* Is theta-motion detected in the peripheral visual field? *Investigative Ophthalmology Visual Sci.* **36**, S52.

Zanker, J. M. 1995*b* Does motion perception follow Weber's Law? *Perception* **24**, 363–372.

Zanker, J.M. 1995*c* On perceptual learning in human motion processing. In *Goettingen Neurobiology Report* 1995 (ed. N. Elsner & R. Menzel), Stuttgart: Georg Thieme.

Zanker, J. M. & Hüpgens, I. S. 1994 Interaction between primary and secondary mechanisms in human motion perception. *Vision Res.* **34**, 1255–1266.

Zhou, Y.-X. & Baker Jr, C. L. 1993 A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science* **261**, 98–101.

Received 12 March 1996; accepted 25 June 1996

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

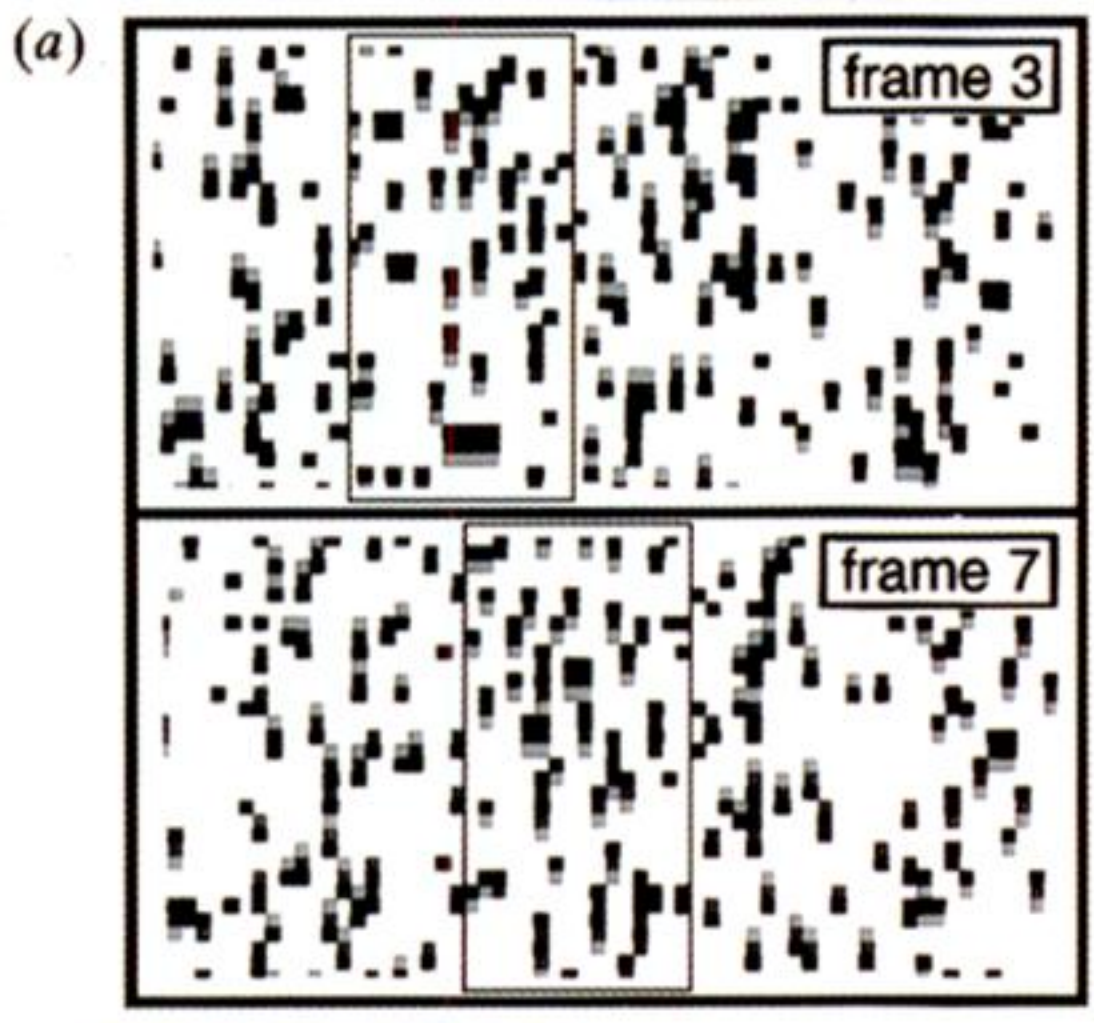

Figure 3. Diagram of the simulation steps of the two-layer model of motion processing. (*a*) 2D sketch of the wavy motion stimulus at two instances of time (frame 3 and frame 7); the position of 10% of the stimulus dots are plotted in black, the positions of the same dots in frame 2 (6) in middle grey, and in frame 1 (5) in light grey; vertical streaks indicate regions of vertical dot motion, the thin line frames a half-cycle (upwards moving dots) of the sine function of motion modulation, which is shifted by 32 pixels between frame 3 and 7. (*b*) and (*c*) 2D representation of motion detector output in pseudo-colour code (red, light grey, and green areas corresponding to regions of negative, zero, and positive response, respectively) for arrays of vertically (*b*) and horizontally (*c*) oriented EMDs (indicated by sketches in boxes, conventions as in figure 1); clear stripes of motion responses can be seen for the vertical EMDs. (*d*) 2D representation of the output of the second layer of EMDs (see two-layer model sketch in box) in pseudo-colour code; the dominating green regions of positive response reflect the detection of the motion signal displacement in the wavy motion stimulus by such a model.