INSTITUTE OF PHYSICS PUBLISHING SMART MATERIALS AND STRUCTURES
Smart Mater. Struct. **14** (2005) 413–424 doi:10.1088/0964-1726/14/2/016

Man-made velocity estimators based on insect vision

Sreeja Rajesh[1,2,3,4], **David O' Carroll**[1,3] **and Derek Abbott**[1,2]

[1] Centre for Biomedical Engineering (CBME), The University of Adelaide, SA-5005, Australia
[2] School of Electrical and Electronic Engineering, The University of Adelaide, SA-5005, Australia
[3] School of Molecular and Biomedical Science, The University of Adelaide, SA-5005, Australia

E-mail: srajesh@eleceng.adelaide.edu.au

Received 1 April 2003, in final form 13 September 2004
Published 4 March 2005
Online at stacks.iop.org/SMS/14/413

Abstract
The study of insect vision is of significant interest to engineers for inspiring the design of future motion-sensitive smart sensor devices, for collision avoidance applications. Although insects are relatively simple organisms compared to vertebrates, they are blessed with a very efficient visual system, which enables them to navigate with great ease and accuracy. Biologically inspired motion detection models are bound to replace the conventional machine vision technology because of their simplicity and significant advantages in a number of applications. The dominant model for insect motion detection, first proposed by Hassentein and Reichardt in 1956, has gained widespread acceptance in the invertebrate vision community. The template model is another known model proposed later by Horridge in 1990, which permits simple tracking techniques and lends itself easily to both hardware and software. In this paper, we compare these two different motion detecting strategies. It was found from the data obtained from the intracellular recordings of the steady-state responses of wide-field neurons in the hoverfly *Volucella*, that the shape of the curves obtained agree with the theoretical predictions made by Dror. In order to compare this with the template model, we carried out an experiment to obtain the velocity response curves of the template model to the same image statistics. The results lead us to believe that the fly motion detector emulates a modified Reichardt correlator.

In the second part of the paper, modifications are made to the Reichardt detector that improve its performance in velocity detection by reducing its dependance on contrast and image structure. Our recent neurobiological experiments suggest that adaptive mechanisms decrease the EMD (elementary motion detector) dependence on pattern contrast and improve reliability. So appropriate modelling of an adaptive feedback mechanism is carried out to normalize contrast of input signals in order to improve the reliability and robustness of velocity estimation.

1. Introduction

When we see a group of bees flying together, the first thing that strikes us is their amazing ability to fly with such great speed without colliding into one another. The activities of insects clearly reveal the extraordinary navigational skill they possess despite the fact that they have a very simple visual system. The study of the insect visual system has offered solutions to problems of computational bottleneck and other problems

[4] Author to whom any correspondence should be addressed.

faced by the conventional machine vision algorithms and has led to many elegant strategies that can be profitably applied to motion detection, velocity estimation and has even been used in the design of collision avoidance sensors and autonomous robots [1]. Research over several decades has revealed that the visual system of insects is exquisitely sensitive to motion, inspiring many models of motion detection.

The earliest and probably the most famous model of motion detection inspired by biological systems was developed by Reichardt and Hassentein [2] in 1956 after a series of behavioural experiments examining the optomotor response of insects. The Reichardt or correlation motion detector possess a highly parallel architecture. Each elementary motion detector (EMD) detects motion in a preferred direction by comparing a signal from one receptor with a delayed signal from the other receptor. The comparison is performed using a nonlinear, multiplicative, interaction between the two channels. Two EMDs tuned to opposite directions are combined to form a bidirectional motion detector. The multiplicative interaction employed in this detector is an excitory mechanism.

While the Reichardt correlator model employed multiplicative interaction as an excitory mechanism, Barlow and Levick [3] pointed out that an inhibitory mechanism was also capable of providing directionally selective motion detection. They conducted experiments on rabbit retinal ganglion cells that were stimulated not by a smooth motion but by a sequence of discrete illumination steps in two neighbouring locations, in either the preferred or null direction of the cell. Barlow and Levick found that the response to the null direction sequence was significantly reduced compared to the sum of the individual responses, whereas responses to the preferred direction sequence was roughly equal to the sum of individual responses. The authors proposed a veto mechanism or 'null-direction inhibition' as the basis for direction selectivity.

Alternative mechanisms of inhibition have also been proposed. One of the mechanisms known as lateral inhibition was proposed by Hartline and Ratliff as a result of experiments on the compound eye of *limulus* [4]. However, in order to be able to model directional selectivity for motion detection, nonlinear interactions are required. Hence a nonlinear version of lateral inhibition, known as shunting inhibition, was developed by Pinter [5] after studies on the neurochemistry of visual cells.

The delay and compare models were derived from behavioural and physiological studies of biological systems. A second class of models is derived from spatio-temporal energy characteristics of moving images—these models include those using a Fourier domain description of moving images [6, 7] while others consider time as another spatial dimension [8].

In 1990, after conducting many experimental studies on the behaviour of the insect visual system, Horridge [9, 10] proposed a simplified model of the insect visual system called the template model. This empirical model compares the contrast between two adjacent receptors, at two sampling instances, to form a 2×2 template. These templates can give simple directional information. Operations are performed mainly between adjacent receptors, which make this model easy to implement in a parallel architecture. Several generations of insect vision chips based on the template model have been developed by Moini *et al* [11–13].

Most of these spatio-temporal models, currently the dominant models for motion detection in vertebrates, are mathematically equivalent to correlator models [8]. Correlator models have been applied to explain motion detection in humans, birds and cats [14–16]. In the absence of additional system components or assumptions, the raw output of a basic Reichardt correlator provides an inaccurate, ambiguous indication of image velocity [17]. Several authors have proposed elaborations of the 'basic' Reichardt detector by either adding temporal, spatial or spatio-temporal filters before applying the input to the EMD cell [18, 19]. In one of the most recent elaborations of this model, Dror [20] has used various forms of spatial filtering, temporal filtering, saturation, and integration within the motion detection system to improve the performance of the correlator-based system in response to many stimuli, including complex natural images. Dror used experimental results from the fly visual system to confirm these predictions. The fly is chosen as a model organism here for both experiments, as well as computational simulations, due to the abundance of behavioural, anatomical and electrophysiological data available in relation to motion detection.

In this study, we have conducted an experiment to compare the response of the template model with the elaborated Reichardt correlator using the same image statistics. Since the Horridge model produces templates as its response, we use the velocity versus the number of templates curve to compare it with the experimental results obtained from the fly HS neurons. The results show that Dror's elaborated model has more similarity with the response of the fly neurons. But Dror's model remains contrast dependent. Hence our work will extend Dror's model to include contrast adaptation, with the aim of achieving 'velocity constancy' with respect to contrast.

2. A simple Reichardt correlator

Figure 1 shows a simplified version of the correlator model. Receptors A and B are separated by an angular distance $\Delta\phi$. The signal from A is temporally delayed by the low-pass filter D before multiplication by the signal from B. This multiplication produces a positive output in response to rightward image motion. To achieve similar sensitivity to leftward motion and to cancel excitation by stationary stimuli, a parallel delay-and-multiply operation takes place with a delay on the opposite arm. The outputs of the two multiplications are subtracted to give a single time-dependent correlator output R.

Although the correlator is nonlinear, its response to sinusoidal stimuli is of interest. If the input is a sinusoidal grating that contains only a single frequency component, the oscillations of the two subunits cancel and the correlator produces a constant output. For any linear delay filter, the output level depends separably on spatial and temporal frequency [17]. If the delay filter D is a first-order low pass with time constant T, as in most modelling studies, a sinusoid of amplitude C and spatial frequency f_s travelling to the right at velocity v produces an output

$$R(t) = \frac{C^2 f_t}{2\pi\tau f_t^2 + 1/(2\pi\tau)^2} \sin(2\pi f_s \Delta\phi) \qquad (1)$$

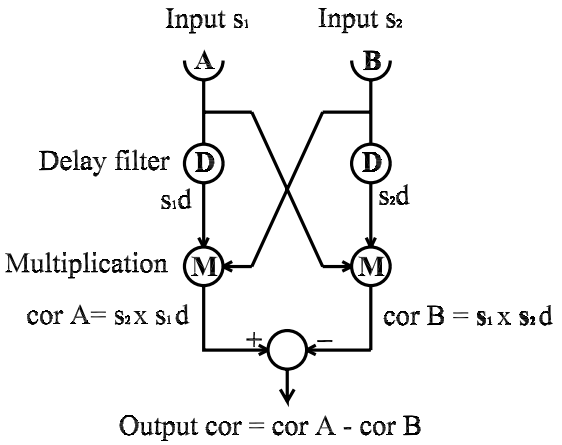

Figure 1. The Reichardt correlator has two arms A and B. Each of the two time-dependent inputs s_1 and s_2 has a fixed angular separation $\Delta\phi$, and passes through a linear delay filter (D) before being multiplied by the other, undelayed signal. The results of the two correlations thus obtained 'cor A' and 'cor B' are subtracted to produce a single output 'cor'. An object moving to the right will produce a positive output; an object moving to the left will produce a negative output.

where $f_t = f_s v$ is the temporal frequency of the input signal [17]. At a given spatial frequency, the magnitude of correlator output increases with temporal frequency up to an optimum $f_{t,opt} = 1/(2\pi\tau)$ and then decreases monotonically as the velocity continues to increase. Output also varies with the square of C, which specifies grating brightness or, in the presence of preprocessing stages, grating contrast. A physical luminance grating must have positive mean luminance, so it will contain a dc component as well as an oscillatory component. In this case, the output will oscillate about the level given by equation (1).

3. Dror's elaborated Reichardt correlator model

Although the simple correlator model produces more meaningful estimates of velocity for natural images than for arbitrary sinusoids, it suffers from two major shortcomings. First, the standard deviation of the correlator output is large relative to its mean, with relative error values ranging from 3.3 to 76 for natural images. Second, the simulated mean correlator response of a simple correlator (with $\Delta\phi = 1.08°$ and with a first-order delay filter of time constant τ = 35 ms) for most natural images peaks at a velocity in the 35°–40° s^{-1} range [20]. Because the velocity range below the peak response corresponds to the most probable range of inputs, we assume that in the absence of contradictory information a correlator response is interpreted as the lower of the two putative velocities. Image velocities above the peak will therefore be misinterpreted. A shorter delay filter time constant would raise the peak response velocity, but experimentally described time constants [21] are not sufficiently low to account for the fact that insects may turn and track targets at velocities up to hundreds of degrees per second [22]. Dror has elaborated the basic Reichardt correlator to include additional physiological components, which helps to overcome these problems, raising the peak response velocity and lowering the relative error of the correlator output.

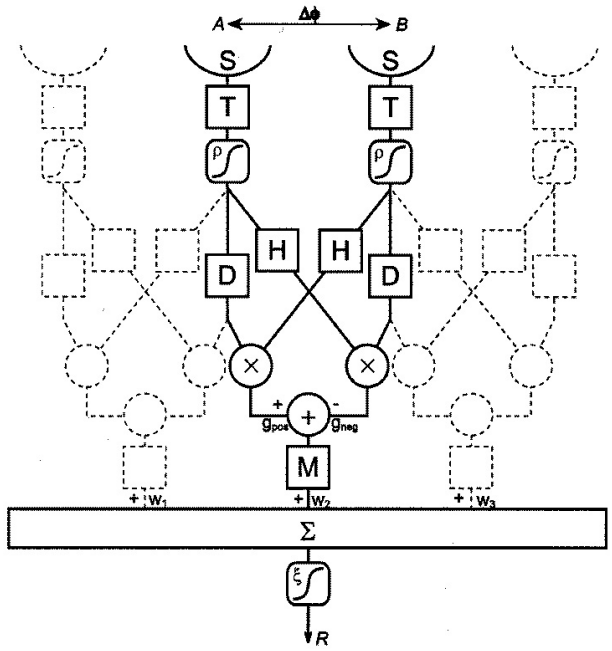

Figure 2. Block diagram of an elaborated correlator model, after [20]. T, D, H and M represent temporal filters; S is a spatial filter; ρ and ξ are saturation functions (compressive nonlinearities). The subunit subtraction may be unbalanced, with weights g_{pos} and g_{neg}. The outputs of the various EMDs undergo two-dimensional spatial integration (Σ), which may be non-uniform, with the weights represented by w_i. For simplicity, this figure omits a number of demonstrated nonlinear and adaptive phenomena.

Figure 2 illustrates an elaborate correlator, including spatial and temporal prefiltering, compressive nonlinearities, and output integration [20].

4. The template model

The template model proposed by Horridge [9, 10] models the function of the small field motion detection neurons in the medulla. In the template model the temporal contrast of the adjacent cells at consecutive instances is used to determine the direction of motion of an object. For simplicity, only temporal contrast at two instances and from two neighbouring cells is considered.

The visual field is sampled spatially. Each sampling channel detects changes in light intensity at two consecutive sampling instances to show either an increase ↑, a decrease ↓ or a no change –. The changed states are then spatially combined between adjacent channels. Hence for a pair of adjacent sampling channels, there are nine possible combinations: (––), (↑ –), (↑↑), (↑↓), (↓ –), (↓↓), (– ↓), (↓↑), (– ↑).

Since in any visual system directional motion is accomplished as a spatio-temporal operation, the temporal domain is included by associating the combinations obtained at two consecutive sampling times t_0 and t_1, thus yielding 81 spatio-temporal combinations or templates. Thus the temporal contrast of two neighbouring cells, at two sampling instances, is combined to give simple motion measures called 'templates'.

Out of these 81 templates, there are only 8 templates which indicate coherent motion. The templates sensitive to coherent motion are the ones in which one of the four entries indicated no

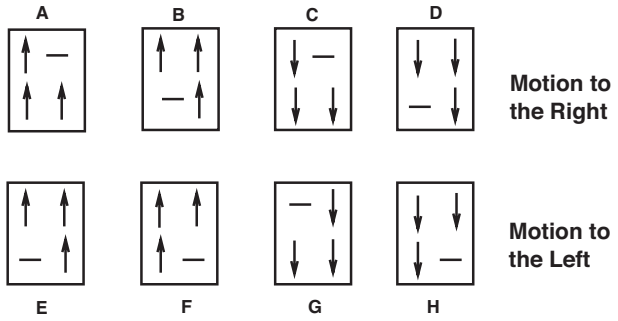

Figure 3. Directionally motion sensitive templates.

change in intensity and the other three must identically either increase or decrease in intensity. These templates are known as directionally motion sensitive templates (DMSTs) [23] as shown in figure 3. These templates possess the 3:1 diagonally symmetric structure.

The template model is easily implemented in VLSI, because the thresholding and template formation operations are simple and can be carried out in parallel. The presence of only three states reduces the bandwidth requirements compared to other vision systems.

A series of 'bugeye' chips [11–13] based on the template model has been implemented using greyscale photodetectors. In these custom insect vision chips, the retina was modelled using an array of on-chip photodetectors with dynamic range control circuitry. This is replaced by a CMOS camera in our prototype. It should be noted that the CMOS camera does not have automatic gain control (AGC) circuitry at the pixel level, but this feature is not essential for indoor controlled lighting conditions.

The filtering operation in the lamina layer was implemented using temporal differentiation in the custom chips, whilst in our CMOS camera prototype this is performed using simple frame differencing. The CMOS camera is interfaced to a personal computer and the captured image is stored in the buffers. The difference between the current image in the current frame buffer and the previous image in the previous frame buffer indicates changes in motion, which are thresholded and stored as templates. Template generation here is performed in software, whereas some custom chips have template generation elements in the hardware. Thus by comparing the templates obtained from frame differencing with directionally sensitive motion templates, the direction of motion of the moving object can be obtained. The time step size between the frames or the frame interval can be adjusted. But it is found that by keeping the frame interval low, more frames can be captured, giving us more template information. When the frame interval is high, a lot of frames and information is lost. In our case, we have chosen a low frame interval of 60 ms as it was found to give the optimum results.

5. Comparison of velocity response curves

In order to test the relationships between image power spectra and velocity response curves, a set of experiments was carried out in which the recordings of steady-state responses of wide-field neurons in a hoverfly to motion of broadband images at different velocities were taken and compared with the analytical and computational predictions. Male specimens of the hoverfly *Volucella* were used for these experiments. The HS (horizontal system) cells typically consist of three horizontal cells termed north (HSN), equatorial (HSE), and south (HSS) horizontal cells because their dendritic trees cover the dorsal, medial and ventral regions of the lobula plate, respectively, with corresponding physiological receptive fields. In syrphids such as *volucella*, the HS system consists of four neurons—an HSNE neuron in addition to the three HS neurons mentioned. Here recordings from HSN, HSNE, and HSE neurons are taken, which are tangential cells of the horizontal system. Additional processing may occur subsequent to or in parallel with the wide-field neurons, so we are not necessarily measuring the fly's actual perception of velocity. However, because the wide-field neurons perform extensive spatial integration and because recorded output is averaged over time, the results are effectively velocity response curves for wide-field neurons [24, 20]. It was found that the shapes of these curves and their dependence on image statistics agree with theoretical predictions. The response of the model correlator is then compared to that of the Horridge model using the same stimulus. Since the template model is also developed from an EMD, we expect some similarity in the results. The velocity response curves obtained with a modelled correlator, the velocity response curves obtained from the physiological recordings of the wide-field neurons of the hoverfly and velocity response curve obtained from the software implementation of the Horridge model are explained clearly and compared in the subsections below.

5.1. Model correlator response

Figure 4 shows velocity response curves for a model correlator predicted analytically from the power spectra of random textures of different densities. This correlator model included spatial blurring by the optics and temporal filtering by light-adapted photoreceptors and LMCs. Saturation effects were not included because the texels themselves provide a binary-valued input signal. The exact shape of the curves and their peak response velocities depend on a number of parameters which are difficult to predict on the basis of available data for these cells, such as the extent of temporal high-pass filtering. Moreover, the predicted curves in this figure do not take account of known nonlinear effects such as gain control. However, several important features of the fly's response to the texels can be predicted with confidence. First, they will have the same general shape as the curves predicted for natural images, increasing monotonically up to a peak response velocity and then falling off. Second, the velocity response curves should shift to the left as the texture density increases.

5.2. Response of the wide-field neuron of the hoverfly

Figure 5 shows velocity response curves for one HSNE neuron measured at six texture densities. The velocity response curves show the expected shape, rising to a peak response at some optimum velocity and then falling off again. The tuning curves for broadband images have higher optimal velocities than the corresponding tuning curves for sinusoidal gratings of optimal spatial frequency. As texture density increases, the

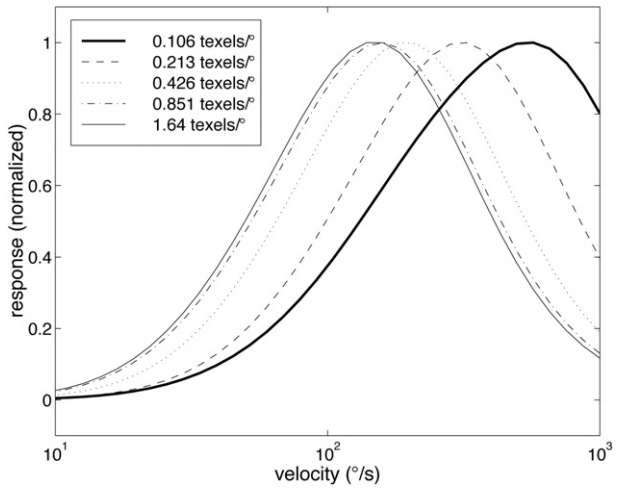

Figure 4. Angular velocity response curves for a model correlator for random texture images of various densities, as predicted from the power spectra. Like models used in previous simulations, the correlator had an inter-receptor angle $\Delta\phi = 1.05°$ and a first-order delay filter with $\tau = 35$ ms. It also included a spatial prefilter corresponding to optic blurring and a temporal prefilter corresponding to light-adapted LMCs, with impulse responses and parameters. Each successive curve corresponds to a random texture with double the density of the preceding one; these correspond to five of the texture densities. These curves were calculated theoretically. While the maximum response levels vary significantly with density, each curve is normalized to a maximum value of 1.0 to facilitate comparison with experimental data. Due to effects such as contrast normalization and output saturation not included in this model, these overall variations in magnitude are not expected to match those recorded experimentally. From [20].

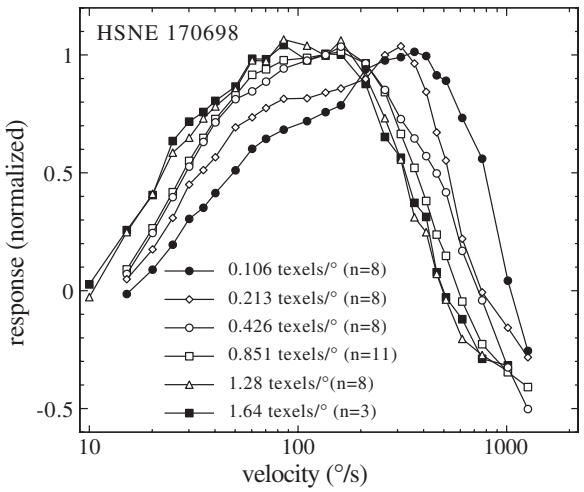

Figure 5. Velocity response curves measured at six different texture densities for a single HSNE neuron. As the texture density increases, the mean horizontal power spectrum of the texture field becomes flatter and the velocity response curves shift to the left. The legend indicates texture density and the number of measurements averaged to obtain each point in the figure. The magnitude of responses of the cell at different texture densities differed significantly, but the curves presented here have been normalized for comparison. The n value gives the number of trials performed for each case. From [20].

curves shift to the left, with the optimal velocity decreasing from over 300°–100° s^{-1} over the range of densities used. As predicted by the model of figure 4, the curves cease to

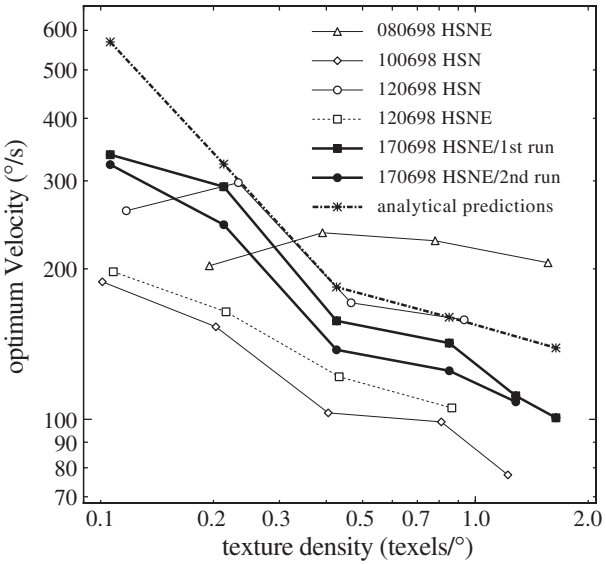

Figure 6. Optimum velocity as a function of texture density for several neurons. The optimal velocity is the velocity of pattern motion for which the neuron gives a maximum steady-state response; this velocity generally decreases as the texture density increases, in agreement with theoretical predictions. Neuron 080698 HSNE exhibited anomalous behaviour, while a noisy velocity response curve led to a single anomalous measurement for neuron 120698 HSN at the second lowest texture density. The legend indicates the type of each neuron. The two thick solid lines indicate two sets of recordings from the neuron whose velocity response curves are shown in this figure. In order to reduce the effects of measurement noise, we estimated the velocity optimum for each cell at each texture density as the velocity optimum of a second-order thin-plate spline fit to the velocity response curve on a logarithmic velocity axis. The amplitude of the curve corresponding to analytical predictions is highly sensitive to the properties of the high-pass temporal prefilter. A slightly weaker high-pass temporal prefilter or the addition of a realistic high-pass spatial prefilter would decrease the predicted response velocities into the range of those observed experimentally. From [20].

shift left at the very highest densities, as the image power spectrum becomes almost completely flat in the relevant frequency range. The model predicts qualitative aspects of the recorded data surprisingly well, given that model parameters were literature values for typical large hoverflies and were not tuned to the cell in question. The data of figure 5 represent a particularly successful recording session lasting over 2 h, during which response curves at a wide range of densities were measured and test protocols were repeated twice at most densities. Neurons in other animals gave similar results, but typically with more noise due to shorter recording sessions.

Figure 6 indicates the relationship between texture density and optimum velocity for a number of HS cells from several flies, as well as analytical predictions. With one anomaly, all cells show the expected decrease of optimum velocity with increasing density. The velocity optima do differ systematically from cell to cell. At any given texture density, some neurons have optimal velocities 50% greater than those of others. These systematic variations between cells may reflect real variations in physiology; for example, temporal prefiltering might vary between organisms or regions of the visual field, which could explain the differences shown. Alternatively, the variations may be due to differences in the

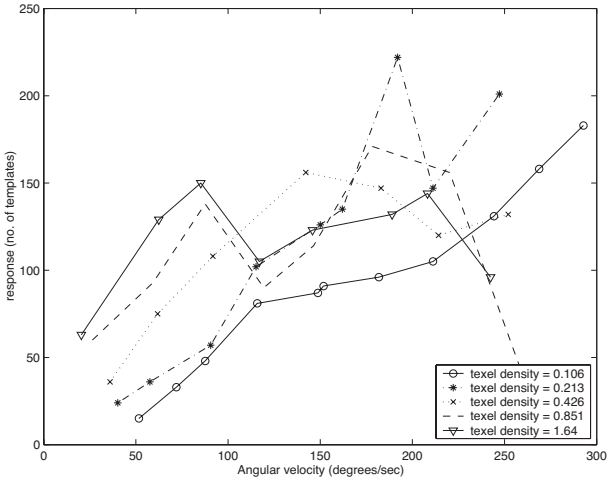

Figure 7. Velocity response curves measured at six different texture densities for the template model. As the density increases, the number of templates produced increases, causing the velocity response curve to shift to the left, emulating the Reichardt model response. The shape of the curve obtained differs to the response obtained from a Reichardt correlator or HSNE neuron of the fly. A number of reasons could be attributed to this. The implementation of the Horridge model here is a discrete-time model, whereas the Reichardt correlator is an analogue model. It could also be due to the noise introduced from the camera and during the analogue to digital conversion, and it could also be due to the motion blur caused by the camera.

recordings, such as the position of the screen relative to the receptive field. Velocity tuning curves as well as spatial and temporal frequency tuning curves were compared for HSE, HSNE, and HSN neurons, but no significant systematic differences were found.

5.3. Horridge template model response

We have set up an experiment to compare the velocity response curve of the Reichardt correlator with that of the Horridge model. In this experiment, the camera was placed in the centre of a hollow cylinder with random horizontal texture elements ('texels') printed on a horizontal roll of paper stuck inside. The cylinder is motor controlled and the angular speed of the cylinder can be adjusted by changing the voltage supply to the motor. Our program then counts the number of templates produced by detecting the motion of the texels. By varying the texture density parameter, texels of different sizes can be produced. The experiment is repeated with each texture density. Since the response of the Horridge model is in the form of templates, the velocity response curve in this case is the velocity versus template count curve, which is plotted for each texel density.

Our program counts the number of templates produced and does it more accurately at lower velocities than at higher velocities. This is expected because here we use the forward tracking method [25] to track the templates and to determine their velocity. Theoretically, this method is more useful in the case of slow motion or at low velocities as the velocity of a slowly moving object is updated at each sampling instant [26]. Hence, in our program, we expect the response to decrease at very high velocities as there are chances of losing or missing the templates, due to the fast motion of the texels.

Figure 7 shows the velocity response curves measured at five texture densities using the template model algorithm. The template model detects the motion of objects as edges. When the texel density is low, the texels are bigger and fewer in number, hence there is a smaller number of edges and thereby fewer templates at low velocities. However, as the texel density increases, the number of edges detected and the number of templates increases, resulting in the curve shifting to the left, showing similarity in response to the response of the HSNE neuron. As the velocity increases, it is seen that the response increases then begins to fall. This is due to the blurring effect caused by the fast motion of the texels. The higher the texel density, the lower the velocity at which the blur occurs. But as we increase the velocity, more texels pass in front of the camera at a shorter rate, resulting in an increase of the response again and then a further increase in velocity causes more motion blur, resulting in decrease of response.

Although the Horridge model is essentially a discrete-time model, the front end of the true Horridge model does a continuous-time differentiation and then the signal is digitized. However, in the set-up with the texels and video camera, that implementation of the Horridge model is fully discrete-time. That could be the reason why the results show variations with the Reichardt model and the experiments done on the fly. But since the template model is basically developed from an EMD, which is the minimum prerequisite for directionally selective motion detection in a visual system, we expect the curves to be similar and we see that the curves do agree to some extent with the Reichardt correlator, in the way that they show increase in response as velocity increases up to an optimum velocity and then the response starts to fall off and in the way the curves shift to the left as the texel density increases.

In the template model, experimental results show that a moving object (or edge) consistently causes the same motion sensitive template to occur at subsequent time steps, and at positions corresponding to the displacement of the edge relative to the detector [25]. The angular velocity may be estimated by evaluating the ratio of the displacement of a motion sensitive template to the time between the template's occurrences. In figure 8, the horizontal axis represents the angular velocity measured using a tachometer and the vertical axis represents the velocity measured, by tracking of the templates, using the template model algorithm. At lower velocities it is seen that the algorithm gives a nearly correct measure of the velocity as the curve coincides with the ideal response, but as the velocity increases the blurring of the texels causes the response to deviate from the ideal line. The slight deviation of the response from the correct value at lower velocities is expected due to the presence of noise in the system. Different noise reduction techniques and filters could be included to reduce its effect [27]. But at higher velocities, because of the blurring of the texels, the edges are not clearly identified, causing the deviation of the response from the ideal line. It is seen that it gets worse at higher texel densities, as there are more edges, and at high velocities they pass the screen at a faster rate, and that, along with the blurring, causes a loss of templates.

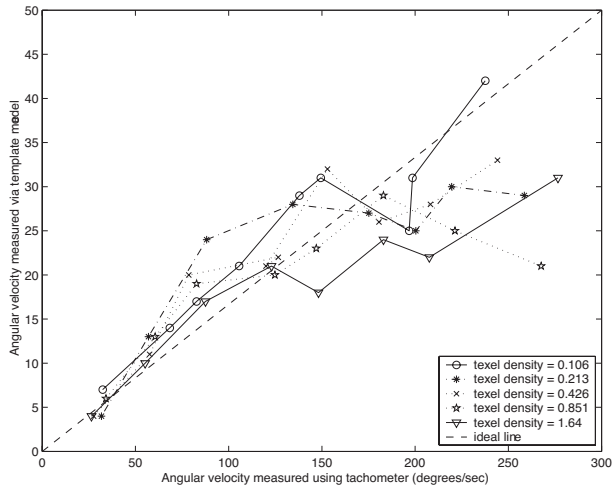

Figure 8. Benchmark angular velocity measured by the tachometer versus velocity measured using the template model. The angular velocity measured by the algorithm deviates from the actual measured velocity or the ideal line for all texel densities, at higher velocities. As the texel density increases it is found that the velocity from which the deviation from the ideal line starts to occur decreases.

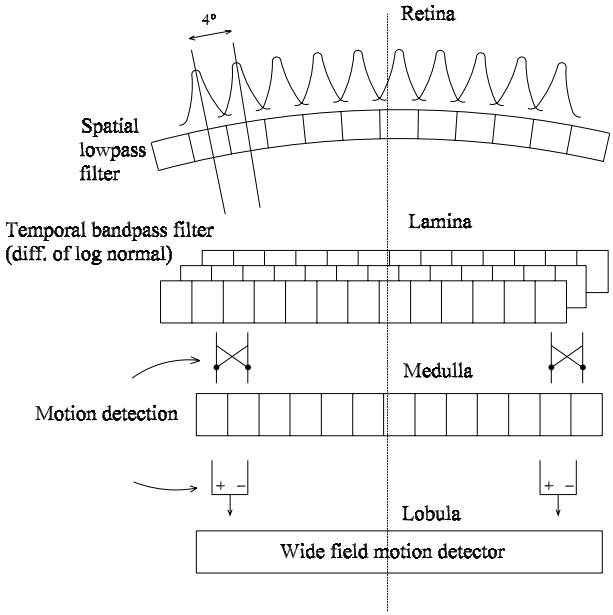

Figure 9. The elaborated version of the fly EMD model with spatial and temporal prefiltering. Spatial filtering is done using a Gaussian filter of half width 2° and temporal filtering is done with the difference of log-normal filters. Motion is detected locally by a correlation method and is pooled so as to enable wide-field motion detection.

5.4. Discussion

This comparative study analyses the important aspects of accurately estimating the local image velocity using a Reichardt correlator and a template model. It is found that both these models give a correct estimate of velocity at low velocities but their efficiency decreases at higher velocities. It also proves that experimentally supported elaborations of the basic Reichardt correlator enhance its reliability as a velocity estimator. Since the basic principle of a template model is the same as the Reichardt correlator, the elaborated Reichardt correlator and the template model produce similar response. This result has been verified by the data obtained from the experiments done on the wide-field neurons of the hoverfly which qualitatively matches the above two results in several important aspects. But it is observed that there is greater similarity with the elaborated Reichardt correlator response than with the template model result, which leads us to believe that the Reichardt correlator model is more biomimetic. However, it may be of interest in the future to elaborate the Horridge model itself, in order to make a fairer comparison.

6. Extension of Dror's model

Dror has demonstrated that the addition of spatial and temporal filtering, saturation, integration and adaptation in a correlator-based system can improve its performance as a reliable velocity estimator. In the second part of this paper, we further investigate and expand his model to improve the correlator performance. The elaborated version of our EMD model with spatial and temporal prefiltering is shown in figure 9. First, the EMD correlator model for the early stage of motion detection in insects is elaborated to mimic the properties of the fly visual system. Then based on motion adaptation studies done on the insect visual system [21, 28], contrast gain reduction is implemented using a feedback mechanism. It is found that the resulting adaptive feedback EMD array model is successful in making the response less sensitive to contrast.

7. Our elaborated motion detector model

In our experiments, we use natural images like the image shown in figure 11 as our stimuli. Recent work has shown that certain image statistics are highly predictable in the natural world [29–31] and that the biological visual system is optimized to take advantage of these statistics [32]. Hence natural images photographed from favoured hovering positions of the hoverfly are used. Then the edges of the images are wrapped to form a panoramic image, which is given as a stimulus to our fly EMD model. A panorama is formed by 'warping' 12 image 'tiles' at 30° intervals to remove lens distribution and then wrapping its ends together. This is done using Apple Quicktime VR software on a Macintosh computer. The resulting image has a width of 8352 pixels and height of 1264 pixels. Spatial prefiltering is implemented by two-dimensional convolution of the image with a Gaussian kernel of half width (standard deviation) 2°, which approximates the acceptance function of a typical fly photoreceptor [33]. Only the luminance (grey scale) information is taken from the image using the green channel since photoreceptors are green sensitive. The spatially low-pass filtered image is illustrated in figure 12. The distance between two ommatidia in an insect eye is between 1° and 1.5°. Since the insect is looking at an image of 360° with 8352 pixels in it, if we consider the inter-ommatidial angle as 1.5°, there will be a total of 240 ommatidia looking at the image. So there will be an array of typically 240 EMDs working together to detect motion as shown in figure 10.

The image in figure 12 is temporally filtered with a difference of log-normal filter to copy the response of the

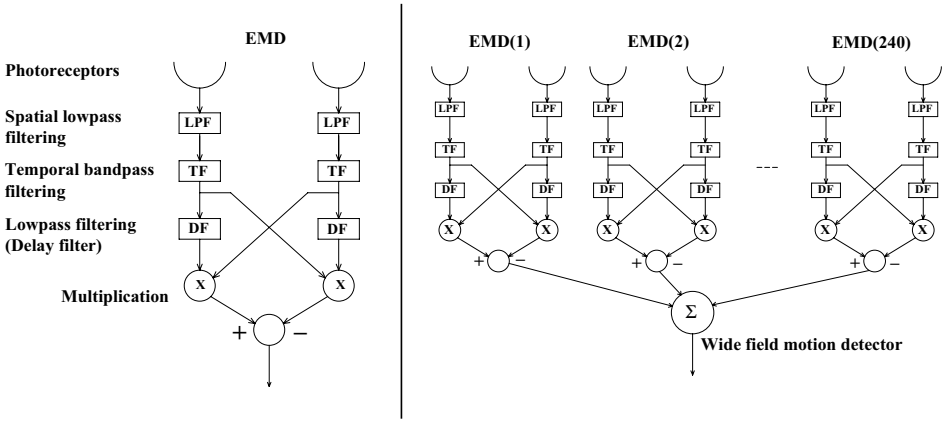

Figure 10. The elaborated EMD array. In our EMD array model, an array of 240 elaborated EMDs is used to detect motion. The input stimulus given is an image of width 8352 pixels, and considering the inter-ommatidial angle as 1.5° there will be typically 240 ommatidia (EMDs) working together to detect motion. The output of these EMDs is pooled to enable wide-field motion detection.

Figure 11. The panoramic natural image given as stimulus to the EMD model. A panorama of the image is formed by 'warping' 12 image tiles at 30° intervals to remove lens distortions and then by wrapping its ends together using Apple Quicktime VR software on a Macintosh computer.

(This figure is in colour only in the electronic version)

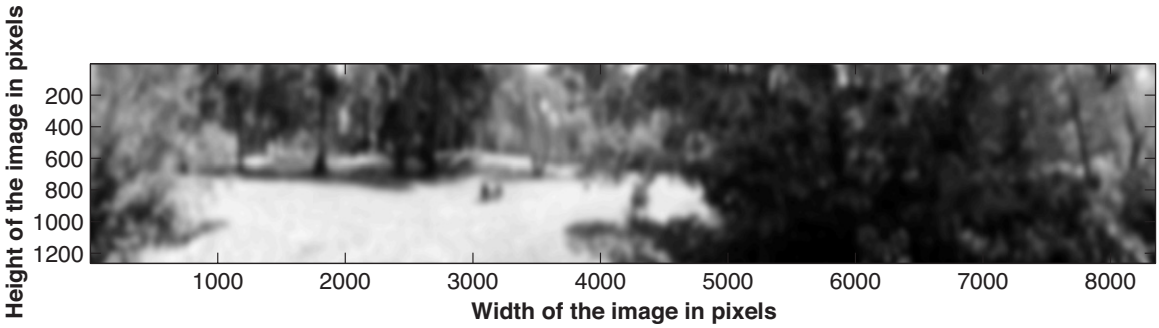

Figure 12. The natural image is prefiltered using a spatial low-pass filter, imitating the characteristics of the photoreceptor. The spatial filter used here is a Gaussian filter of half width (standard deviation) 2°.

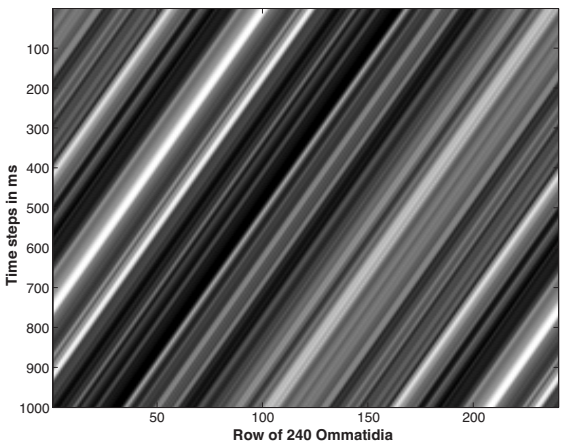

Figure 13. A single row of the image animated at a constant velocity of 200° s^{-1} and sampled onto an array of 240 ommatidia.

lamina monopolar cells [34–36]. The temporally prefiltered image is converted to a space–time matrix as shown in figure 13 based on desired velocity, which is then simulated along the height of the image in pixels. The height of the image is also divided into 36 ommatidia by calculating the pixels per degree and keeping the inter-ommatidial angle as 1.5°.

Then this spatio-temporally prefiltered image is given to the EMD array, which correlates the inputs to give an array of outputs as done in the insect eye [37, 38]. Then the EMD array model copies the lobula by averaging the outputs to produce an average EMD response. The simulations are first performed with a constant velocity of 200° s^{-1} per second and the input–output responses of a single EMD row from the total array is shown in figure 14. Then the simulations are repeated by increasing the velocity in steps as shown in figure 15(a) and the spatial average or the average response of all the EMD rows is shown in figure 15(b).

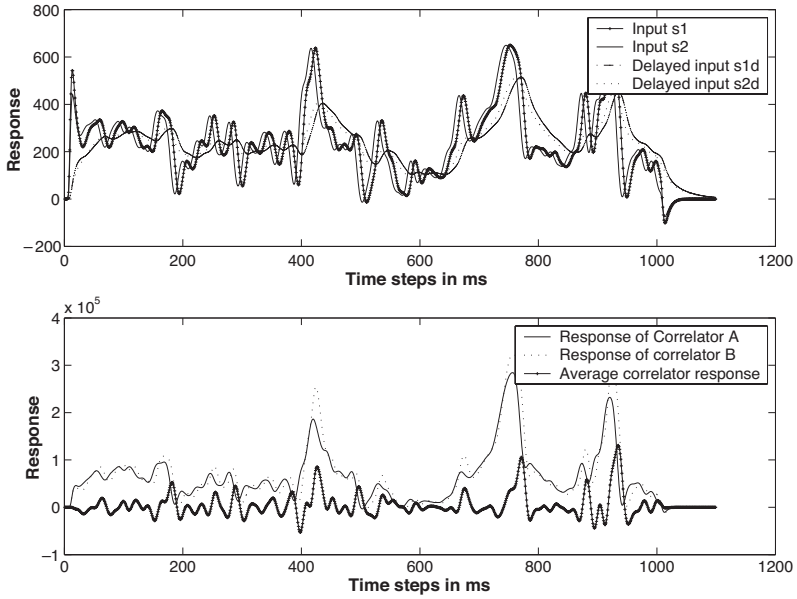

Figure 14. The input and the output simulated responses of a single EMD row from the modelled array moving with constant velocity of 200° s^{-1}. Refer to figure 2 for explanation of each component. Note that the EMD output as a function of time is highly variable, and is maximal whenever high contrast features move across the EMD inputs (s_1, s_2).

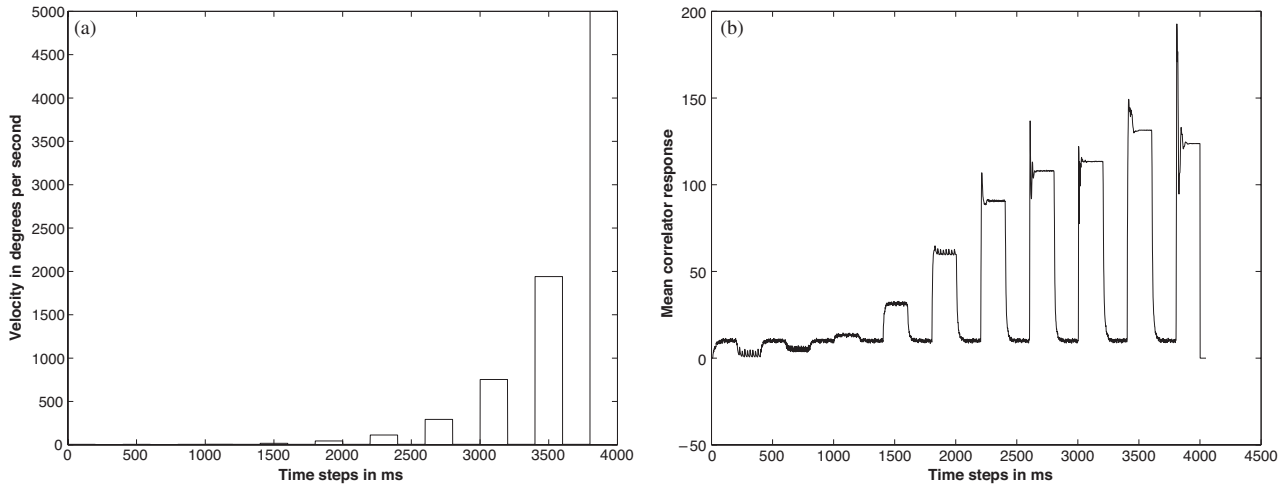

Figure 15. (a) To test motion adapted responses, we increased the velocity stepwise, with interleaved bursts of adapting motion (constant speed). The velocity is increased in steps with time. (b) The mean correlator simulated response of the EMD array model to a natural image moving with velocity increasing in steps across the EMD array.

8. Contrast adaptation

Recent studies by Harris *et al* (1999) [21] on motion adaptation in the fly reveal that motion adaptation induces a profound decrease in contrast sensitivity of fly motion sensitive cells via two proposed mechanisms, a local direction sensitive after potential and a local direction insensitive contrast gain reduction. We propose that this gain reduction may serve to reduce sensitivity to image contrast, as well as reducing the tendency of the motion detectors to saturate.

In the correlation model of motion detection, each correlator contains an expansive nonlinearity (multiplication). This would make the correlator output particularly sensitive to the magnitude of input signals and so potentially vulnerable to saturation. The after-potential and the gain reduction serve to release the motion pathway from this saturation, allowing us to maintain a wide sensitivity across a wide range of stimulus conditions. The after-potential acts antagonistically to recent activity in the cell, repositioning the cell's responses within the available signalling range. By analogy with the retina, this type of subtractive mechanism may exploit correlations in continuous signals, reduce redundancy and maintain the operations of synapses in favourable regions of their input–output functions.

Similarly the gain reduction component of adaptation scales down the magnitude of signals in the motion pathway. If the correlator nonlinearity is to be protected from saturation the gain control should act on the inputs. Furthermore, since the output of the correlators depends on the spatio-temporal correlation between the input signals as well as their

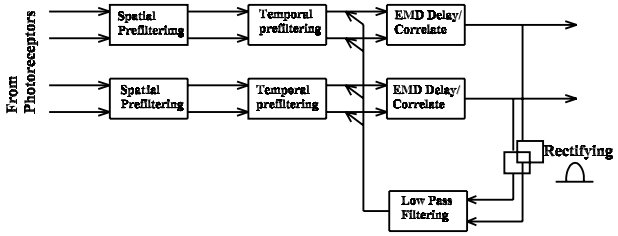

Figure 16. Block diagram of a feedback adaptive EMD model. In this model, the output of the correlator is rectified and low-pass filtered and is fed back to control the gain of the EMD inputs.

magnitude, the gain control would be best regulated by the magnitude of the correlator output, not the magnitude of the inputs. This is consistent with the observation that adaptation is recruited by motion but not flicker. The gain reduction mechanism will be recruited strongly when either the mean or the variance of the stimulus velocity distribution becomes large.

8.1. Experimental verification

As described in section 5, a set of experiments was carried out in which the recordings of steady-state responses of wide-field neurons in a hoverfly to motion of broadband images at different velocities were taken and compared with the analytical and computational predictions. It was found that the shapes of these curves and their dependence on image statistics agree with theoretical predictions [39].

The results of the experiments suggested that while the system behaves like a simple Reichardt correlator at low contrasts, these curves support the presence of some form of contrast gain control at higher contrasts. This invariance with contrast is characteristic of the motion-adapted system; the unadapted system exhibits larger variations of the response level with stimulus contrast. From a practical point of view, the invariance of the motion-adapted velocity response curve with overall image contrast implies that mean correlator output may indeed provide an accurate estimate of velocity for a wide range of natural images [20].

9. Contrast gain reduction—feedback adaptive EMD model

In our elaborated Reichardt correlator array model, in order to reduce the dependancy of the response to changes in contrast and spatial frequency and to get a more accurate estimate of velocity, contrast gain reduction is implemented by a feedback adaptive process as in figure 16. The gain of the EMD inputs is reduced by a signal derived from the rectified and low-pass filtered outputs of a local EMD pool with different local preferred direction, fed back to control the gain of the EMD inputs. The model was inspired by recent observation of contrast-dependent gain reduction in the responses of HS neurons following motion stimulation. The model captures several aspects of the adaptive phenomena observed in the biological system. In particular, the adaptation is strongest when the local motion detector output is the largest, conferring a robustness in adaptation to motion signals as opposed to static flicker or noise applied to the inputs. This matches data obtained from the electrophysiological experiments, which show that motion stimuli are much more effective at recruiting adaptive gain than other stimuli. Secondly, the adaptive mechanisms remain independent of the direction of local motion, despite the selectivity for motion as the source of adaptation. Figure 17 shows the input and the output simulated responses of one row of EMDs from the adaptive EMD array in response to the test sequence illustrated in figure 15(a).

In order to test the performance of this adaptive EMD array, we have compared the responses of the non-adapted elaborated EMD array with the adaptive elaborated EMD array at three different contrasts. From the graphs (figures 18 and 19)

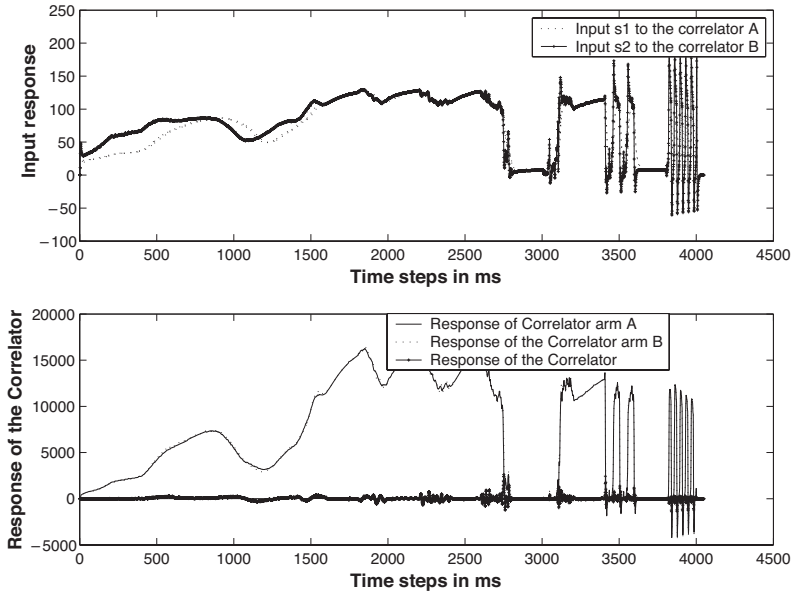

Figure 17. The input and the output simulated responses of one row of EMDs from the adaptive EMD array in response to the test sequence illustrated in figure 15(a).

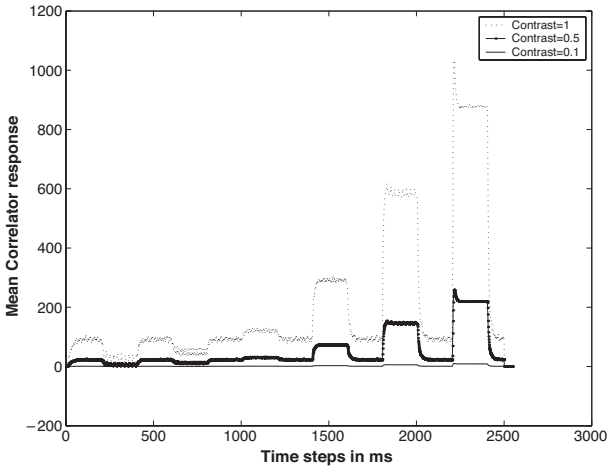

Figure 18. The simulated mean correlator response of an unadaptive EMD array at three different contrasts. The response of the elaborated EMD array without including the adaptive feedback loop is noted at three different contrasts (1, 0.5 and 0.1). It is clearly seen from the graph that there is huge variation of the response with contrast.

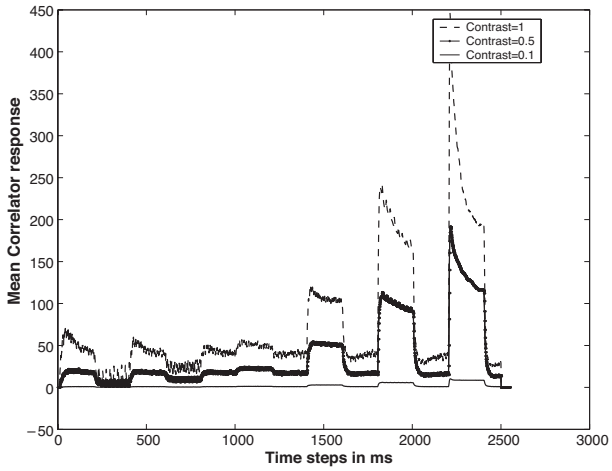

Figure 19. The simulated mean correlator response of an adaptive EMD array at three different contrasts. The response of the elaborated EMD array including the adaptive feedback loop is noted at three different contrasts (1, 0.5 and 0.1). The adaptive feedback mechanism helps in decreasing the dependence of the response to contrast.

it is very clear that the adaptive feedback mechanism has resulted in decreasing the contrast sensitivity of the model, compared with an otherwise similar model that lacks adaptive feedback. It is also found that the adaptation is stronger at higher velocities and at higher contrasts.

10. Conclusion and future work

In this paper, a comparative study of the velocity estimation by the template model with the Reichardt correlator model is performed. The experimental results reveal that Dror's elaborated correlator model has more similarity with the curves obtained from the fly HS neurons. In order to improve its performance, Dror's model is then extended and we present a preliminary implementation of an adaptive EMD array model to take account of recent neurobiological research on the fly visual system. This model leads to contrast gain reduction at local EMDs as observed in fly HS neurons [28]. In our model, this gain reduction is achieved by direct feedback of local EMD outputs. An alternative mechanism in which the feedback signal gates a feedforward contrast normalization [40] provides an alternative explanation for the physiological data. Ongoing physiological experiments in our laboratory aim to test predictions of both models and thus to establish whether either of these models is fully consistent with motion adaptation mechanisms in the fly visual systems.

The present model demonstrates that a feedback gain control is capable of reducing the dependence of EMD output on contrast in natural images. Although the model does not, in its present form, achieve 'velocity constancy' with respect to contrast, as implied by pilot experiments on the fly HS neurons [28], we have deliberately simplified this model to exclude known additional nonlinearities of the fly vision, including logarithmic encoding of luminance and response saturation. Since earlier work already showed that these additional elaborations improve velocity coding by EMDs [20, 39], the inclusion of such components in future models may serve to achieve substantial improvements in performance of an adaptive EMD over earlier versions of the Reichardt correlator.

Acknowledgments

This research is supported by funding from the US Air Force Research Laboratory/Asian Office for Aerospace Research and Development (contract no. F62562-01-P-0158), the Sir Ross and Sir Keith Smith Fund and the Australian Research Council.

References

[1] Srinivasan M V, Chahl J S, Weber K, Venkatesh S, Nagle M G and Zhang S W 1998 Robot navigation inspired by principles of insect vision *Robot. Auton. Syst.* **26** 203–16

[2] Hassenstein B and Reichardt W 1956 Structure of a mechanism of perception of optical movement *Proc. 1st Int. Conf. on Cybernetics* pp 797–801

[3] Barlow H B and Lewick W R 1965 The mechanism of directionally selective units in the rabbit's retina *J. Physiol. Lond.* **178** 477–504

[4] Ratliff F and Hartline H K 1974 *Studies on Excitation and Inhibition in Retina* (London: Chapman and Hall)

[5] Pinter R B 1984 Adaptation of receptive field spatial organization via multiplicative lateral inhibition *Proc. IEEE Conf. on Systems, Man and Cybernetics* (Piscataway, NJ: IEEE) pp 328–31

[6] Watson A B and Ahumada A J 1985 Model of human visual-motion sensing *J. Opt. Soc. Am.* **2** 322–42

[7] Fleet D J and Jepson A D 1989 Hierarchical construction of orientation and velocity sensitive filters *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 315–25

[8] Adelson E H and Bergen J 1985 Spatiotemporal energy models for the perception of motion *J. Opt. Soc. Am.* A **2** 284–99

[9] Horridge G A 1990 A template theory to relate visual processing *Proc. R. Soc.* B **239** 17–33

[10] Sobey P 1990 Determining range information from self motion—the template model *Proc. SPIE Intell. Robot. Comput. Vis.* **1382** 123–31

[11] Moini A, Bouzerdoum A, Yakovleff A and Eshraghian K 1996 A two dimensional motion detector based on insect vision *Proc. SPIE Int. Symp. Advanced Focal Plane Arrays and Electronic Cameras (Berlin)* ed T Bernard (Bellingham, WA: SPIE Optical Engineering Press) pp 146–157

[12] Moini A and Bouzerdoum A 1997 A biologically motivated imager and motion detector with pixel level image processing *Australian Microelectronics Conf. (Melbourne, Australia)* pp 180–5
[13] Moini A, Bouzerdoum A, Eshraghian K, Yakovleff A, Nguyen X T, Blanksby A, Beare R, Abbott D and Bogner R E 1997 An insect vision based motion detection chip *IEEE J. Solid State Circuits* **32** 279–84
[14] van Santen J P and Sperling G 1985 Elaborated Reichardt detectors *J. Opt. Soc. Am.* A **2** 300–321
[15] Wolf-Oberhollenzer F and Kirschfeld K 1994 Motion sensitivity in the nucleus of the basal optic root of the pigeon *J. Neurophysiol.* **71** 1559–73
[16] Emerson R C, Citron M C, Vaughn W J and Klein S A 1987 Nonlinear directionally selective subunits in complex cells of cat striate cortex *J. Neurophysiol.* **58** 33–65
[17] Egelhaaf M, Borst A and Reichardt W 1989 Computational structure of a biological motion detection system as revealed by local detector analysis in the fly's nervous system *J. Opt. Soc. Am.* A **6** 1070–87
[18] van Santen J P H and Sperling G 1985 Elaborated Reichardt detectors *J. Opt. Soc. Am.* A **2** 300–21
[19] Courellis S H and Marmarelis V Z 1990 An artificial neural network for motion detection and speed estimation *Proc. Int. Joint Conf. on Neural Networks* pp 407–21
[20] Dror R O 1998 Accuracy of visual velocity estimation by Reichardt correlators *Master's Thesis* University of Cambridge, Cambridge, UK
[21] Harris R A, O'Carroll D C and Laughlin S B 1999 Adaptation and the temporal delay filter of fly motion detectors *Vis. Res.* **39** 2603–13
[22] Land M F and Collett T S 1974 Chasing behaviour of houseflies (*Fannia canicularis*): a description and analysis *J. Comp. Physiol.* **89** 331–357
[23] Yakovleff A and Moini A 1997 Motion perception using analog VLSI *J. Analog Integr. Circuits Signal Process.* **2** 1–22
[24] O'Carroll D C, Bidwell N J, Laughlin S B and Warrant E J 1996 Insect motion detectors matched to visual ecology *Nature* **382** 63–6
[25] Yakovleff A, Nguyen X T, Bouzerdoum A, Moini A, Bogner R E and Eshraghian K 1994 Dual-purpose interpretation of sensory information *Proc. IEEE Int. Conf. on Robotics and Automation (San Diego, CA, May 1994)* vol 2 (Piscataway, NJ: IEEE) pp 1635–40
[26] Nguyen X T, Bouzerdoum A and Moini A 1996 Velocity measurement using a smart micro-sensor *Int. Symp. on Robotics and Cybernetics (Lille, France)* vol 943, pp 937–42
[27] Nguyen H X, Rajesh S and Abbott D 2001 Motion detection algorithms using template model *Proc. SPIE Conf. on Electronics and Structures for MEMS II* vol 4591 (Bellingham, WA: SPIE Optical Engineering Press) pp 78–90
[28] Harris R A, O'Carroll D C and Laughlin S B 2000 Contrast gain reduction in fly motion adaptation *Neuron* **28** 595–606
[29] Burton G J and Moorhead I R 1987 Color and spatial structure in natural scenes *Appl. Opt.* **26** 157–70
[30] Tolhurst D J, Tadmor Y and Chao T 1992 Amplitude spectra of natural images *Ophthalmol. Physiol. Opt.* **12** 229–32
[31] Ruderman D L 1994 The statistics of natural images *Network: Comput. Neural Syst.* **5** 517–48
[32] Laughlin S B 1994 Matching coding, circuits, cells and molecules to signals: general principles of retinal design in the fly's eye *Prog. Retinal Res.* **13** 165–95
[33] Hardie R C 1985 Functional organisation of the fly retina *Prog. Sensory Physiol.* **5** 1–80
[34] James A C 1990 White-noise studies in the fly lamina *PhD Thesis* Australian National University
[35] Payne R and Howard J 1981 Response of an insect photoreceptor: a simple log-normal model *Nature* **290** 415–6
[36] Howard J H, Dubs A and Payne R 1984 The dynamics of phototransduction in insects *J. Comp. Physiol.* **154** 707–18
[37] Hausen K and Egelhaaf M 1989 Neural mechanisms of visual course control in insects *Facets in Vision* ed R Hardie and D Stavenga (Berlin: Springer)
[38] Douglass J K and Strausfeld N J 1995 Visual motion detection circuits in flies: peripheral motion computation by identified small-field retinotopic neurons *J. Neurosci.* **15** 5596–611
[39] Dror R O, O'Carroll D C and Laughlin S B 2001 Accuracy of velocity estimation by Reichardt correlators *J. Opt. Soc. Am.* A **18** 241–52
[40] Shoemaker P A, O'Carroll D C and Straw A 2001 Implementation of visual motion detection with contrast adaptation *Proc. SPIE Conf. on Microelectronics and Micro-Electro-Mechanical Systems (Adelaide, Australia)* vol 4591 (Bellingham, WA: SPIE Optical Engineering Press) pp 316–27