Local and global responses of insect motion detectors to the spatial structure of natural scenes

David C. O’Carroll  
Adelaide Centre for Neuroscience Research, School of Medical Sciences, The University of Adelaide, SA, Australia

Paul D. Barnett  
Adelaide Centre for Neuroscience Research, School of Medical Sciences, The University of Adelaide, SA, Australia

Karin Nordström  
Adelaide Centre for Neuroscience Research, School of Medical Sciences, The University of Adelaide, SA, Australia

As a consequence of the non-linear correlation mechanism underlying motion detection, the variability in local pattern structure and contrast inherent within natural scenes profoundly influences local motion responses. To accurately interpret optic flow induced by self-motion, neurons in many dipteran flies smooth this “pattern noise” by wide-field spatial integration. We investigated the role that size and shape of the receptive field plays in smoothing out pattern noise in two unusual hoverfly optic flow neurons: one (HSN) with an exceptionally small receptive field and one (HSNE) with a larger receptive field. We compared the local and global responses to a sequence of panoramic natural images in these two neurons with a parsimonious model for elementary motion detection weighted for their spatial receptive fields. Combined with manipulation of size and contrast of the stimulus images, this allowed us to separate spatial integration properties arising from the receptive field, from other local and global non-linearities, such as motion adaptation and dendritic gain control. We show that receptive field properties alone are poor predictors of the response to natural scenes. If anything, additional non-linearity enhances the pattern dependence of HSN’s response, particularly to vertically elongated features, suggesting that it may serve a role in forward fixation during hovering.

Keywords: pattern noise, motion detection, natural images, velocity encoding, EMD, dendritic gain, insect vision


Introduction

Neural pathways for visual motion analysis in a variety of animal species utilize correlation of changes in light intensity across space and time to estimate the motion of local contrast features (for review, see Borst & Euler, 2011; Borst, Haag, & Reiff, 2010; Clifford & Ibbotson, 2002). Consequently, when using simple grating patterns, local motion responses show pronounced dependence on pattern structure and contrast (Egelhaaf, Borst, & Reichardt, 1989; Meyer, Lindemann, & Egelhaaf, 2011; Reichardt, 1987; Reichardt & Egelhaaf, 1988). Natural scenes, however, contain a rich mixture of pattern components that vary enormously in local structure, luminance, and contrast (Girshick, Landy, & Simoncelli, 2011; Ruderman & Bialek, 1994; Simoncelli & Olshausen, 2001), which generate location-dependent response variations or “pattern noise” (Dror, O’Carroll, & Laughlin, 2001). Accurately interpreting motion in the natural world thus presents a major challenge for biological visual systems.

In flies, optic flow is averaged by neurons with large receptive fields, the lobula plate tangential cells (LPTCs). LPTCs achieve large receptive fields by spatially summing large arrays of retinotopically arranged elementary motion detectors (EMDs; see, e.g., Borst & Euler, 2011; Borst et al., 2010; Egelhaaf et al., 1989). While each LPTC collates the output of EMDs viewing spatially circumscribed subregions of their visual world, many LPTCs also share axonal gap junctions with their neighbors (Elyada, Haag, & Borst, 2009; Farrow, Borst, & Haag, 2005; Haag & Borst, 2004, 2005). In the context of natural scenes, large receptive fields are hypothesized to assist in robust motion coding by smoothing out pattern noise (Dror, O’Carroll, & Laughlin, 2000; Haag, Egelhaaf, & Borst, 1992). In particular, spatial averaging by elongation of the receptive fields along the direction of motion provides an optimal solution to smoothing pattern noise, but it also generates a trade-off between response robustness and providing information about the location of motion (Meyer et al., 2011).

Males of the hoverfly *Eristalis* have two fronto-dorsal LPTCs (HSN and HSNE) with unusually narrow receptive...
fields compared with those of most other dipteran flies (Hausen, 1982b; Krapp, Hengstenberg, & Egelhaaf, 2001; Nordström, Barnett, Moyer de Miguel, Brinkworth, & O’Carroll, 2008). The HSN neuron, in particular, has a receptive field largely confined to a specialized dorso-frontal eye region involved in mate pursuit behavior and that is unusually sensitive to motion (Straw, Warrant, & O’Carroll, 2006). Surprisingly, given the findings of recent modeling in other species (Meyer et al., 2011), the hoverfly HSN receptive field is particularly narrow along the preferred direction of motion, being instead elongated orthogonal to this direction (Nordström et al., 2008). While this may suggest a role in enhancing rather than suppressing pattern noise, HSN nevertheless provides a reliable estimate of yaw velocity from a complex mixture of translational and rotational optic flow during prolonged stimulation with a sequence of natural scene images (Nordström et al., 2008). Indeed, our recent recordings from both HSN and HSNE show that these neurons also provide surprisingly robust estimates of image velocity when stimulated with a large set of panoramic natural images, despite these images varying enormously in spatial structure and contrast (Barnett, Nordström, & O’Carroll, 2010; Straw, Rainsford, & O’Carroll, 2008).

How does HSN achieve highly robust responses despite its small receptive field? Recent work using physiological techniques highlights contributions to robust motion coding from additional non-linearities. For example, the postsynaptic membrane voltage of LPTCs saturates with increasing pattern size, leading to a typically sublinear dependence of the overall response on pattern size—a static non-linearity known as “dendritic gain control” (Borst, Egelhaaf, & Haag, 1995). Small signals integrate linearly if the inputs are located on different dendritic branches but sublinearly when the signals are integrated close-by (Elyada et al., 2009; Haag et al., 1992). The specific shape and organization of the major dendritic organization could, thus, strongly influence the responses of LPTCs to natural scenes in which high-contrast features are locally clustered.

Importantly, spatial integration of local motion by LPTC dendrites represents a static non-linearity at the end of a long sequence of local motion processing stages, many of which are also known to exhibit pronounced dynamic non-linearity, beginning with luminance adaptation by the photoreceptors themselves. In particular, LPTCs rapidly adapt to prior motion stimuli with strong reduction in contrast gain (e.g., Harris, O’Carroll, & Laughlin, 2000; Kurtz, 2007; Maddess & Laughlin, 1985; Nordström, Moyer de Miguel, & O’Carroll, 2011). These adaptive changes appear to be part of an active normalization strategy that minimizes the global response variability from one image to the next (Barnett et al., 2010). While still speculative in terms of the underlying mechanisms, highly elaborated bio-inspired models have recently been developed for the rapid adaptation operating locally in the photoreceptors and second-order lamina monopolar cells (LMCs), through the EMD itself, and then globally within the modeled LPTC (Brinkworth & O’Carroll, 2009; Shoemaker, O’Carroll, & Straw, 2005; van Hateren & Snippe, 2001). These models show that luminance and motion adaptation can potentially provide robust contrast normalization of natural scenes.

In this paper, we investigate the consequences of receptive field size and shape in the processing of natural image motion by comparing the time-dependent responses recorded from the unique male hoverfly neurons HSN and HSNE (Nordström et al., 2008). We implemented a parsimonious correlation-based model for motion detection, which takes into account the size and shape of the receptive fields of HSN and HSNE in order to predict pattern dependence in response. We then compared the output of this model with the observed response of HS neurons to large-field natural image motion. While our model reflects the local spatiotemporal tuning and receptive field weights of the neurons, it does not incorporate non-linearities in spatial integration such as dendritic gain control or other dynamic non-linearities such as motion adaptation. We instead use experimental manipulation of the stimuli to differentiate between these known non-linear contributions to response summation in space and time and the predictions of our parsimonious model.

Methods

Electrophysiology

Wild-caught male hoverflies (Eristalis tenax) were immobilized with wax. We cut a small hole over the left lobula complex through which we inserted an alumino-silicate electrode pulled on a Sutter Instruments P97 electrode puller. Electrodes were backfilled with 2 M KCl and had tip resistances of 80–250 MΩ. HS neurons were identified based on their physiological response properties and receptive field (Nordström et al., 2008).

Data acquisition and analysis

Data were digitized at 5 kHz using a 16-bit A/D converter (National Instruments) and analyzed offline with Matlab (http://www.mathworks.com). In all experiments, we normalized the response to the resting membrane potential. HS neurons respond with graded membrane potential changes on which activity-induced spikelets ride. As these add additional non-linearity to the membrane potential, we spike filtered our data by removing spike-like events and replacing them with the local mean membrane potential (see Nordström & O’Carroll, 2009).
Statistics

We performed cross-covariance analysis (r) in Matlab (http://www.mathworks.com) to compare pattern dependence between conditions and with model predictions. Quoted cross-covariance coefficients assume a zero lag between the two data sets, i.e., when the data are time aligned, unless otherwise mentioned. Wilcoxon signed rank tests and t-tests were performed using GraphPad Prism software (http://www.graphpad.com), with significance allocated to \( p < 0.05 \). All data are presented as mean ± standard deviation (SD), unless otherwise mentioned, with the \( N \) number for the data displayed in each figure given in the figure legend (minimum of 3 individual flies for each experiment). Capital \( N \) denotes the number of flies, and lower case \( n \) denotes the total number of replicates for each experiment.

Image collection and display

The panoramic natural images represent a subset of the images used by Barnett et al. (2010) and include field sites around South Australia. They range from densely forested areas (e.g., “Hamlin,” “Botanic,” “Creekbed”), through vast open hillsides (e.g., “Field”), to entirely man-made environments (“Carpark”). We displayed 8-bit versions of the panoramic images using Vision Egg software (Straw, 2008) on a linearized RGB CRT monitor at a spatial resolution of 640 × 480 pixels, corresponding to ca. 100° azimuth × 75° elevation of the fly’s visual field. The CRT had a 200-Hz refresh rate and a mean luminance of 100 Cd/m².

We artificially reduced image contrast by scaling the value for each pixel (\( I_{\text{final}} \)) about the mid-gray level of our display, such that

\[
I_{\text{final}} = C (I_{\text{image}} - 0.5) + 0.5,
\]

where \( I_{\text{image}} \) is a floating-point number from 0 to 1 representing luminance intensity in the original image, and \( C \) is the contrast scaling factor (Barnett et al., 2010; Straw et al., 2008).

EMD model

We used a parsimonious Hassenstein–Reichardt correlator model that incorporates many of the spatial and temporal filtering processes on the motion processing pathway (Figure 1). The model includes spatial and temporal filtering matched to Eristalis tenax optics, early visual processing (specified below), and motion correlation to ensure that its output reflects the spatiotemporal passband of the insect motion-sensitive neurons we recorded from. The parsimonious EMD uses an inter-receptor angle of \( \Delta \phi = 1.1^\circ \), a physiologically realistic value for the separation of frontally orientated EMDs in

---

Figure 1. Parsimonious EMD model. (A) Schematic of our parsimonious EMD model. The model uses an inter-receptor angle of \( \Delta \phi = 1.1^\circ \), based on Eristalis (Straw et al., 2006), spatial prefiltering implemented as a 2D Gaussian blur (\( \Delta \rho = 1.4^\circ \)), and temporal pre-filtering matching LMCs modeled as the difference of two lognormals with different time constants (\( t_p = 10.3 \text{ ms} \) and \( \sigma = 0.236 \) for the positive lognormal, and \( t_p = 15.6 \text{ ms} \) and \( \sigma = 0.269 \) for the negative lognormal, where \( t_p \) represents the time to peak of the curve and \( \sigma \) represents the curve’s width). The EMD delay, \( \tau \), is 31 ms. (B) The average receptive field of male Eristalis tenax HSN (Nordström et al., 2008). The local motion sensitivity amplitude is indicated in mV by the color bar. Superimposed is an array of EMDs shown as black circles. Note that for display purposes these are not scaled to their actual size relative to the receptive field.
**Local motion detector analysis**

To investigate local motion responses, we used a window mask corresponding to the size of a few EMDs, by displaying the central rows of the image through a $5^\circ \times 5^\circ$ window located at $10^\circ$ azimuth on the equator. For the slit mask, we extended the height of single image rows and stretched these vertically to $5^\circ$ wide $\times 50^\circ$ high. We used the central row of the original image, corresponding to $1.4^\circ$ of the visual field. This row was vertically blurred to account for the optics of *Eristalis* (Straw et al., 2006), and then stretched across the $50^\circ$ slit.

**Results**

**Local features in natural images induce large response fluctuations**

To what extent does receptive field size and shape contribute to reliable encoding of yaw velocity of natural scenes? To address this question, we analyzed the pattern dependence in the global response of hoverfly HS neurons. Our stimulus display is a viewport onto the panoramic scene, which was animated continuously past this port (Figure 2A). Each scene was perspective-corrected for the flat computer screen on each video frame so that both angular size and velocity ($45^\circ$/s in the preferred direction) of the scene were constant at all points in the image (see Straw et al., 2008). The receptive fields of HSN and HSNE differ, with the receptive field of HSN being particularly narrow in the direction of image motion (Figure 2B).

To avoid confounding pattern dependence with motion adaptation, we pseudorandomly varied the stimulus start position over eight initial start phases each separated by $45^\circ$. Figure 2C shows the response at 4 of these start phases for the image “Field” in Figure 2A. The slope of the diagonal lines in Figure 2C represents an offset of 2 s between each of the 4 initial phase presentations (i.e., $90^\circ$ at a speed of $45^\circ$/s). In each presentation, the image moved for 9 s and thus rotated through $405^\circ$ in total. For subsequent analysis, we excluded the first second of image motion and then averaged the response across each of these presentations starting from the same initial phase (as indicated by the diagonal lines in Figure 2C). The resulting phase-aligned response thus transposes parts of the recorded membrane potential from different absolute times following stimulus onset and spreads any remaining time-dependent changes (i.e., motion adaptation) evenly. A similar analysis of the response to a high-contrast grating pattern produces an essentially flat line (data not shown). The response fluctuations in response to local variations in the specific features within each image are repeatable from one recording to the next with relatively small variance (Figure 2D).

The consistent temporal structure of each response, once we allow for the phase offset, confirms that these modulations reflect the underlying pattern noise generated in response to features at unique locations within the image (Figure 2D). For example, the group of vertically aligned clouds shown on the left-hand edge of the monitor (Figure 2A) produced a large depolarization around 3.5 s (Figure 2D; the response is phase aligned with the panorama in Figure 2A). The region immediately preceding it, on the other hand, has few vertically aligned features and produced a negligible response (obvious at around 2–2.5 s, Figures 2A and 2D). Over the course of a full image rotation, 8 s, average neural response varied by 85% from maximum to minimum for this image, despite image velocity remaining constant at $45^\circ$/s. These large pattern-dependent response fluctuations mean that the neuron’s response is temporally ambiguous with respect to the velocity of image motion.

**Receptive field size and shape influence neuron pattern dependence**

How does receptive field size and shape influence pattern dependence? The male HSN and HSNE neurons have largely overlapping receptive fields with similar centers along the azimuth (Nordström et al., 2008) and similar spatiotemporal frequency tuning if stimulated frontally (Straw et al., 2006). However, the neurons have demonstrable differences in receptive field shape and size (Figure 2B). The male HSN neuron has an extremely narrow fronto-dorsal receptive field (Nordström et al., 2008), whereas HSNE has a much larger receptive field

---

*Eristalis* (Straw, Warrant, & O’Carroll, 2006). Spatial pre-filtering was implemented as a 2D Gaussian blur, with $\Delta p = 1.4^\circ$, which approximates the acceptance function of typical fly photoreceptors (Dror et al., 2001; Hardie, 1985). Temporal pre-filtering was based on the response of *Eristalis* LMCs to continuously varying white noise stimuli (James, 1990). James (1990) showed that LMC responses could be modeled as the difference of two lognormals with different time constants. At high light levels, he found typical values of $t_p = 10.3$ ms and $\sigma = 0.236$ for the positive lognormal and $t_p = 15.6$ ms and $\sigma = 0.269$ for the negative lognormal, where $t_p$ represents the time to peak of the curve and $\sigma$ is a dimensionless parameter that determines the curve’s width (Dror et al., 2001; Payne & Howard, 1981). The EMD delay was implemented as a first-order low-pass filter with a time constant, $\tau$, of 31 ms. The exact value of the time constant does not change the conclusions of this paper. We weighted the outputs of the EMD array with the average HSN and HSNE receptive fields (Nordström et al., 2008).
that extends laterally and ventrally. Figure 3 shows the average responses of HSN (red) and HSNE (blue) neurons to six different natural images (Barnett et al., 2010), displayed as described in Figure 2. Despite producing time-averaged responses of similar magnitude from one image to the next (HSN: 7.8 ± 1.5 mV; HSNE: 6.5 ± 0.9 mV; mean ± SD), structures within some images clearly induce more response fluctuations than other images (e.g., compare Figure 3A with Figure 3F).

Responses of HSN with the smaller receptive field were consistently more variable than those of HSNE. This is confirmed by a larger standard deviation (averaged across the set of 6 images) for HSN (SD = 1.29 ± 0.15, mean ± SEM) than HSNE (SD = 0.98 ± 0.11, p < 0.05, paired, one-tailed t-test). While HSN’s receptive field is largely confined to our display (Figure 2B), the lateral extent of HSNE’s receptive field extends beyond the limit of the stimulus monitor (Nordström et al., 2008), so if anything, our results underestimate the degree to which its larger receptive field averages out pattern noise.

Interestingly, however, pattern noise is not simply damped by the larger receptive field. This is evident from pronounced differences in the phase and magnitude of the residual pattern dependencies in the response (Figure 3). Correlation analysis that assumes a 0 time lag between the HSN (red) and HSNE (blue) responses revealed a poor correlation, averaged across all 6 images (r = 0.23 ± 0.13, mean cross-covariance coefficient ± SEM). Given that the maximum sensitivity region of the HSNE receptive field is shifted along the azimuth by 10.5° compared to that of HSN (Figure 2B, and see Nordström et al., 2008), we might expect to see a small temporal delay (0.233 s at 45°/s) in HSNE’s response. When we accounted for such a delay, there was a slight improvement in the correlation in response between the two neurons, although it was still relatively weak (r = 0.47 ± 0.1, r-values for each image in Figure 3). Close inspection of the data reveals substantial variation between images in this respect. For some images (e.g., “Carpark,” Figure 3A and “Hamlin,” Figure 3B), the difference in response fluctuation between HSN and

---

Figure 2. Response to global image motion. The “Field” image (one of six natural panoramic images, see Figure 3) displayed on our monitor, where it covers ca. 100° azimuthal extent, of the 360° panorama, and 72° height. (B) The average receptive field of male HSN (red) and HSNE (blue, Nordström et al., 2008). The panoramic images were perspective distorted (white) on our flat monitor to simulate pure yaw rotation around the fly’s field of view (Straw et al., 2008), moving in the preferred direction (arrows). (C) The images were yaw rotated at 45°/s for 9 s. Because HS neurons adapt to prolonged stimulation, we started each trial at one of eight pseudorandom image start phases (4 shown here, 0°, 90°, 180°, and 270° offset). The diagonal dotted lines connect time points separated by 2 s that phase align the image for each start position. Responses are the average of N = 4 neurons with n = 32 repetitions. (D) The phase-aligned HSN response (mean (black) ± SD (gray)) to “Field” is adjusted for the image start phases. The image start phase of 0° (C) was used for aligning the averaged data. The response of each neuron was normalized by its own mean to reduce the variance of absolute magnitude between recordings. The response is space–time aligned with the image in (A) so that image features correspond to the neural responses below. Note the different time axes in (C) and (D).
HSNE is profound. In these images, some features that produce a strong peak in response in one neuron actually produce a dip in the other (arrows, Figures 3A and 3B). However, for other sparse images (e.g., “Field,” Figure 3E), the responses of the two neurons are similar.

### Computational modeling of receptive field structure

To what extent can receptive field structure itself explain the response differences between HSN and HSNE? To address this, we compared neuron responses with the output of a parsimonious EMD model (Figure 1). The EMD model includes filters matched to several of the motion vision processing stages (Dror et al., 2001; Shoemaker et al., 2005). To account for receptive field size and shape, we weighted the outputs of the EMD array with the average HSN and HSNE receptive field size recorded physiologically (Figure 1B, and see Nordström et al., 2008). While our model should reflect the local spatiotemporal tuning and receptive field weights of the neurons, it makes no attempt to incorporate other local and global non-linearities in spatial integration, such as saturation and dendritic gain control, or other dynamic non-linearities, such as motion adaptation. Differences between our parsimonious model and the global neuronal responses thus reflect the influence of these additional non-linear processing elements.

Figure 4 shows the neural responses (replotted from Figure 3) in pastel colors, with the model prediction for each image overlaid in saturated hues. Since the images differ enormously in their global contrast (see Barnett et al., 2010) and our model does not include any adapting or saturating elements, the mean output varied more than 45-fold between the six images (HSN: 0.29 ± 0.22; HSNE: 0.39 ± 0.35; mean ± SD). To be able to compare the model output and the neuronal responses in the same figure, we therefore normalized the model output to the maximum neuron response, for each image (Figure 4). Non-linear processing by the neurons clearly plays a major role in shaping their global responses to natural scenes: We see a poor correlation between the observed pattern dependence in the response of the HS neurons (pastel colors) and our
parsimonious model for both HSN (red, $r = 0.36 \pm 0.04$) and HSNE (blue, $r = 0.39 \pm 0.11$, Figure 4). In contrast to the neural responses, the model’s HSN and HSNE outputs are well correlated with each other, despite the large difference in receptive field size ($r = 0.78 \pm 0.02$, mean $\pm SEM$). This correlation improves even further when we added the delay that accounts for the lateral shift in receptive field centers (0.23 s at 45°/s; $r = 0.86 \pm 0.02$, mean $\pm SEM$).

We can thus conclude that the observed differences in pattern dependencies in HSN and HSNE (as seen in Figure 3), and the differences between the model and the neurons (Figure 4), are not simply an effect of receptive field structure but are additionally affected by other non-linear processes acting either locally, globally, or both within the receptive field. This is further supported by differences in the model’s predictive power for global responses in the different images. HS responses (replotted from Figure 3) are shown in pastel colors. Cross-covariance coefficients ($r$) are shown at the top of each panel, where $r_{HSN}$ compares the model HSN output (red) with the HSN neuron response (pastel red), and $r_{HSNE}$ compares model HSNE (blue) with HSNE neuron (pastel blue). The gray dashed line indicates zero model output and the resting membrane potential, respectively. For each image and HS type, we normalized maximum model output to the maximum neuronal response (non-normalized model responses show a 45-fold output variation across the six images).

**Artificially lowering image contrast**

Compressive non-linearities, such as response saturation (Dvorak, Srinivasan, & French, 1980) and dendritic gain control (Borst et al., 1995; Single, Haag, & Borst, 1997), influence the responses of LPTCs, and the inclusion of such non-linearities in EMD models reduces inter-scene response variance to natural images (Brinkworth & O’Carroll, 2009; Dror et al., 2001; Shoemaker et al., 2005). As these non-linear influences are largely activity dependent, we should be able reduce their influence by reducing the contrast of the stimulus, thus bringing each stage of motion processing closer to its linear operating range.

Consistent with a release from the influence of compressive non-linearities, when we reduce the image contrast to 25% the mean response goes down, but relative pattern-dependent modulation increases (Figure 5). The variance-to-mean ratio (VMR, the Fano factor) increases for all images, for both HSN (from 0.27 ± 0.11 to 0.70 ± 0.13, $p < 0.05$, Wilcoxon signed rank test) and HSNE (from 0.17 ± 0.04 to 0.64 ± 0.12, $p < 0.05$). Furthermore, our parsimonious EMD model provides much more
consistent predictions of neural responses (Figure 6). Mean correlation coefficients between model output and neural responses increased significantly (from $0.36 \pm 0.04$ to $0.62 \pm 0.06$ in HSN, $p < 0.05$, and from $0.39 \pm 0.11$ to $0.65 \pm 0.08$ in HSNE, $p < 0.05$). However, the prediction quality varies between images (e.g., compare responses to “Hamlin” with “Field,” Figures 6B and 6E).

The model output in these examples is normalized to the maximum neural response for each neuron. In fact, the absolute model output varies more than 45-fold. Thus, although our parsimonious model, in combination with the receptive field, provides a much better prediction of the pattern dependence for individual images once image contrast is reduced, it fails dismally at capturing the absolute magnitude of the neural response produced from one image to the next. The inclusion of simple compressive non-linearities in a further elaborated version of the model would, without doubt, reduce some of the spread of absolute response magnitudes (as shown by Dror et al., 2001). However, it would also severely “flattens” the local response fluctuations far more than evident in the neural data (Figure 5).

Although there is some evidence of a release from response saturation (Figures 5 and 6), the neural responses did not merely rescale after we lowered image contrast. In fact, for several images, the shape of the response after reducing the image contrast shared little resemblance to the response produced at the full image contrast, despite the structure within the scenes remaining the same. As highlighted by arrows in Figures 5B and 5C, some features that produce local depolarization peaks at normal contrast become troughs in the reduced contrast image, and vice versa. Consequently, on average in both neuron classes, there was only a mediocre mean correlation between the two contrast conditions (HSN, $r = 0.62 \pm 0.06$; HSNE, $r = 0.67 \pm 0.07$). Once again, this varied greatly from image to image. For example, in both neurons, “Carpark” and “Field” showed strong correlations, while “Creekbed” showed a weaker correlation (see Figure 5 for $r$-values). In addition, there were differences between HSN and HSNE. For example, responses to the two contrast versions of “Library” were strongly correlated in HSNE ($r = 0.83$) but more weakly in HSN ($r = 0.54$; Figure 5D).

Local motion responses to natural scenes

The data in Figures 1–5 show the pattern dependence at a global level, with our stimulus largely filling the neurons’ receptive fields. At least at high image contrast, the discrepancy between our parsimonious model and these global responses presumably reflects the influence of both static non-linearities, e.g., dendritic gain control, and dynamic non-linearities, e.g., locally recruited motion.

Figure 5. HS responses to low-contrast natural scenes. We artificially reduced the contrast of the six images [(A) “Carpark,” (B) “Hamlin,” (C) “Botanic,” (D) “Library,” (E) “Field,” (F) “Creekbed”] to 25% and recorded HSN (red) and HSNE (blue) neuronal responses. HS neuron responses to the full contrast images (replotted from Figure 3) are shown in pastel colors. Cross-covariance coefficients ($r$) are shown at the top of each panel, where $r_{\text{HSN}}$ compares the low-contrast HSN response (red) with the full contrast response (pastel red), and $r_{\text{HSNE}}$ compares the low-contrast HSNE response (blue) with the full contrast HSNE response (pastel blue). The gray dashed line indicates the neurons’ resting membrane potential. Arrows in (B) and (C) indicate substantial differences in response to the two contrast images.

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932790/ on 01/01/2019
adaptation. In order to get a better idea of how local features of the images drive the EMDs themselves in the presence of local dynamic non-linearity, but without the confounding influence of spatial gain control, we extended our analysis by using local stimulation of the HS neurons. We used a window to mask the stimulus, limiting its spatial extent to the size of only a few EMDs, using an approach developed for sinusoidal grating stimuli (Figure 7A, and see Egelhaaf et al., 1989; Reichardt & Egelhaaf, 1988).

Figure 7B illustrates our approach using one of our images (“Close”) and the image strip that passed through the 5° window mask (stretched vertically in this figure for display purposes). Figure 7C shows the response of an HS neuron to this same strip. Areas in the image strip that contain few contrast boundaries (e.g., the grassy patch visible at 1–3.5 s or the dark bushy area at 5–7.5 s) generate almost no response. Conversely, obvious contrast boundaries in the image produce sizeable membrane potential depolarizations. Figure 7D shows the output of the parsimonious model for the same image strip. Note that although the model captures the general phase and relative magnitude of the neural pattern dependence, its local response to higher contrast features of the image is quite different from the neuron. Some of these differences clearly arise from dynamic adaptation of the response. For example, the EMD model predicts that three sequential high-contrast features near the boundary of the dark bushy area at t = 4–5 s (dashed lines, Figures 7B–7D) should each produce similar magnitude responses, but the neuronal response to the first of these features is much larger than those that follow. This likely reflects a local reduction in contrast gain induced by the passage of the first feature. The neuronal response to this feature lasts for over 20 ms (see time axis in Figure 7D), which is long enough to induce contrast gain reduction (Nordström et al., 2011).

Experiments using the small window mask evoke only weak responses and thus require averaging over many trials to isolate neural response from noise (in the example in Figure 7C, we show the average of 240 repetitions). To allow us to collect the equivalent of local responses for a whole image, but in fewer trials, we adapted a method originally developed to study local motion detector responses to grating patterns (Egelhaaf et al., 1989; Reichardt & Egelhaaf, 1988). We extended the height of single image rows to stretch across the entire vertical extent of our display. Before sampling each row, the image was vertically blurred to account for the optics of Eristalis (Straw et al., 2006). This stimulus gives a stronger response because it stimulates many (vertically aligned) EMDs simultaneously. However, because the stimulus is in phase across the slit, the response is still equivalent to that of a single EMD (Egelhaaf et al., 1989; Reichardt & Egelhaaf, 1988). This allowed us to obtain full 2-dimensional maps of neuronal responses to the
central strips from the “Carpark” and “Botanic” images (Figure 8).

Large transient depolarizations associated with obvious image features (e.g., the edge of the car or the columns in “Carpark”) dominate the neural response for both images (Figures 8B and 8E). Between these transients, the membrane potential tends to remain slightly hyperpolarized, despite the image moving continuously in the preferred direction. The small, prolonged hyperpolarization of the membrane potential between depolarizing events likely reflects the antagonistic motion after-potential (“waterfall effect”), a component of motion adaptation that can last for several hundred milliseconds depending on the strength of the stimulation (Harris et al., 2000; Nordström & O’Carroll, 2009; Srinivasan & Dvorak, 1979).

While the local response of the parsimonious model to the same scenes predicts the location of the gross features of the image, the output only results in a mediocre correlation with the neuronal response (average row coefficients: $r = 0.58 \pm 0.05$ for “Botanic,” and $r = 0.53 \pm 0.15$ for “Carpark”). In the model output, many features that produce strong responses are associated with subsequent transient hyperpolarizations (Figures 8C and 8F), a prediction of basic Hassenstein–Reichardt correlator response to discrete contrasting objects (Geurten, Nordström, Sprayberry, Bolzon, & O’Carroll, 2007). However, such transient hyperpolarizations are less apparent in the neuronal response (Figures 8B and 8E). This difference was also evident using the window mask (arrows, Figures 7C and 7D).

**Spatial summation**

The results in Figures 7 and 8 provide the first experimentally measured responses of individual EMDs that contribute to the response of HS neurons to local components of natural scenes. Our stimulus allows these local input elements to express all of the normal dynamic non-linearity associated with local motion processing, including luminance and motion adaptation along the entire input pathway. This data set thus has a substantial advantage over our recent attempts to account for such adaptation in elaborated adaptive EMD models (e.g., Brinkworth & O’Carroll, 2009; Shoemaker et al., 2005), which infer or guess the nature and gain of many components.
within the EMD that have yet to be studied experimentally, because the neurons involved have thus far proved to be too small to record from in vivo (see, e.g., Borst et al., 2010).

We can potentially treat our data set as an “ideal” model input to further models for the wide-field integration of the local motion signals by the receptive field of HSN.

Before we do so, however, we need to consider the degree to which the global response of an HS neuron should reflect the linear sum of the local responses (at least weighted by the locally measured receptive field strength). When using sinusoidal grating patterns, LPTC response saturates as stimulus size increases, indicating the presence

Figure 8. Response to local image motion. (A) The “Botanic” panoramic image. (B) Thirty individual rows were stretched vertically and displayed to the animal at 45°/s using the slit window (as illustrated with dashed white lines, Figure 7B). The neural response (color coded relative to resting membrane potential) is spatially mapped to the equivalent vertical location of the image rows, with the response time course matched to the horizontal location within the image (N = 3, n = 14). (C) Model output to the same stimulus. The model output shows the average of three neighboring EMDs, to match the width of the stimulus window. (D–F) The same configuration as (A)–(C) but for “Carpark.” In this case, N = 3, n = 12.
of highly non-linear spatial integration (Borst et al., 1995; Gauck & Borst, 1999; Haag et al., 1992). The non-linear spatial integration can be predicted by a compartment model of the LPTC if the subtraction operation of an EMD-type input (Figure 1A) occurs on its dendrites, because motion in one direction jointly activates both excitatory and inhibitory conductances with a ratio that depends on pattern velocity, a phenomenon known as dendritic gain control (Borst et al., 1995; Single et al., 1997).

These studies, however, did not investigate the spatial integration in response to natural images. The dendritic gain control mechanism that leads to sublinear spatial summation requires inputs to be activated simultaneously on nearby dendrites, so small signals integrate linearly if the inputs are located on different dendritic branches (Haag et al., 1992). While the image features of a sinusoid are distributed widely and evenly across the pattern and thus are likely to activate numerous adjacent dendrites, at least at high spatial frequencies, those within natural scenes are less regularly spaced and may thus sum more linearly, at least within local patches of the scene. To test this hypothesis, we used a variant of the slit experiment in Figure 7B, this time recording the response to horizontally extended images with varying vertical extent (Figure 9).

We compared the neuronal response with that of the parsimonious EMD model, where integration across the receptive field was simulated as a weighted linear sum based on the same set of weights used earlier (Figure 9A). Because the HSN receptive field becomes narrower and less sensitive at the upper and lower margins of the screen, the integrated model predicts a weakly sublinear relationship with increasing image height despite the absence of any compressive non-linearity within the model (red dashed line, Figure 9A). The HSN response to the “Botanic” image appears to closely track the model predictions up to 40 degrees in image height but then falls off slightly for full screen images (green, Figure 9A), indicating recruitment of mild additional response compression compared with the linear model. Responses to several other images show a similar pattern, if anything rather more linear, despite absolute differences in the magnitude of the response to images such as the low-contrast “Field” image (Figure 9A). A similar, nearly linear, response was observed up to 60 degrees of image height at several different velocities for the “Library” image (Figure 9B). Despite the literature showing a mechanism for strongly sublinear spatial integration when experimenter-defined stimuli are used (Borst et al., 1995; Gauck & Borst, 1999; Haag et al., 1992; Hausen, 1982a; Single et al., 1997), our results thus suggest that the response to natural scenes sums largely linearly with increasing stimulus size.

Using local responses as an elaborated model for the global response

Given that spatial summation for natural scenes is approximately linear, we can now use the receptive field weights derived from our earlier experiments on hoverfly HS neurons, combined with the locally measured neural response to produce a “hybrid” model for the global response that takes full account of any dynamic non-linearity, such as local motion adaptation. We animated the neural image of the locally measured neuron response
through the same linear summation mask as used for the parsimonious EMD model to simulate the weights of the receptive field (Figure 1B) to generate a model for the pattern dependence using the whole stimulus display (Figure 10). Interestingly, once again we see poor correlation between recorded pattern noise and that predicted by the local motion detector responses (Figure 10A). Moreover, if we compare this same prediction with the response to pattern dependence recorded under reduced image contrast conditions (Figure 5), the correlations improve but remain relatively weak (Figure 10B). Because this hybrid model incorporates realistic dynamic non-linearity (i.e., motion adaptation), discrepancies between the predicted and measured responses presumably indicate the strong influence of additional spatial integration non-linearities. In particular, large depolarizing spikes in the experimentally measured global response for some images are not evident in the summed local responses.

Surprisingly, predictions of this hybrid neuronal response model for spatial integration are, if anything, even worse than those of our EMD model (Figures 4 and 6), particularly for the “Carpark” image, which contains a predominance of vertical and diagonal edge features. Unfortunately, the technically demanding nature of this technique means that it is very time consuming to collect repetitions of locally generated neuronal responses or, for that matter, explore this effect across a wider range of images (each such “neural image” takes several hours of intracellular recording to generate reliably). Part of this discrepancy may, thus, be due to recording noise that persists in our locally measured responses. However, an intriguing additional possibility is that the appearance of linear summation evident in Figure 9 results from two counteracting influences within the HSN receptive field: Dendritic gain control generated by local “clusters” of features within the scenes that stimulate adjacent dendrites may lead to locally sublinear summation, while responses to larger features could generate expansive non-linearity for depolarizing transients via voltage-gated conductances within the HSN axon itself. Axonal sodium conductances have previously been shown to enhance positive (depolarizing) transient response in blowfly HS neurons (Haag & Borst, 1998). Axonal conductances would certainly

**Figure 10.** Local responses cannot predict the global response. We used local motion responses (shown in Figure 8) to generate a prediction of pattern dependence based on the receptive fields of the HSN (red) and HSNE (blue) neurons. (A) Predicted pattern dependence (hybrid model, data from Figure 8, N = 3) overlaid on the observed pattern dependence recorded with the whole screen stimulus (pastel colors, replotted from Figure 3) for “Botanic” and “Carpark.” (B) The same as in (A) but with the pattern dependence recorded at 25% image contrast (pastel colors, replotted from Figure 5). The covariance coefficients (r) were computed at zero time lag for the observed versus the predicted pattern dependence. The local response predicted pattern dependence is normalized to the maximum of the observed global pattern dependence for each image.
explain the failure of the hybrid model to capture the sharp depolarizing transients in response to the high-contrast, vertically aligned features within the “Carpark” image (Figure 10).

Discussion

Neuronal responses to natural scenes are influenced by a range of local and global non-linearities operating at different stages of the motion vision pathway. Our experimental approach provided a means for partially segregating the influence of the different effects on HS responses to natural images: (i) The global neuronal responses to the full contrast images include the effect of all saturating and dynamic non-linearities that shape responses to natural scenes. (ii) The reduced contrast images have the same spatial structure and thus stimulate spatial integration similarly to the original images but recruit less dynamic non-linearity, such as motion adaptation. (iii) The parsimonious EMD model reflects the local spatiotemporal tuning and receptive field weights of HSN and HSNE but does not incorporate other local and global non-linearities in spatial integration, such as saturation and dendritic gain control, or other dynamic non-linearities, such as motion adaptation. (iv) The local neuronal responses and, thus, our “hybrid” model include the additional influence of dynamic local motion adaptation but recruit less non-linearity in spatial integration. Our results confirm that both spatial and temporal non-linearities strongly shape the global response.

Influence of receptive field size and shape

Large LPTC receptive fields have been described as a prerequisite for reliable estimation of self-generated optic flow, since local optic flow fields may give ambiguous cues (Krapp, Hengstenberg, & Hengstenberg, 1998; Meyer et al., 2011). Calcium imaging of the input dendrites of LPTCs confirms the smoothing effect of the pattern noise generated locally by grating patterns (Borst & Egelhaaf, 1992). Unusually, however, HSN in male hoverflies has an extremely narrow receptive field, yet it still responds reliably to yaw stimuli when stimulated with complex optic flow (Nordström et al., 2008). Its narrow receptive field, however, is poorly optimized for averaging out residual pattern-dependent response fluctuations: These are more prominent in HSN than in HSNE (Figures 2 and 4) where local irregularities in the scene appear smoothed by the larger receptive field. Even more striking, HSN shows global responses that are far more pattern dependent than predicted from integrating its local motion responses with its receptive field (red, Figure 10). It appears as if integration across the receptive field actually sharpens the response to local pattern features within the image and that the structure of the receptive field of hoverfly HSN seems designed to enhance rather than suppress pattern noise. This is most apparent for responses to the image “Carpark”: Although the prediction captures the two main depolarizing “bumps,” the global response includes many sharp de- and hyperpolarizations within these (red, Figure 10). HSNE responses are better matched with the prediction (blue, Figure 10), largely because the responses are smoothed by the larger receptive field. HSN should, thus, give more reliable optical flow estimates in temporal snapshots from texture-rich environments.

A recent modeling study emphasizes the importance of elongating the receptive field of LPTCs in the direction of motion, thereby providing more reliable estimates of image velocity (Meyer et al., 2011), at the cost of being able to pinpoint the location of motion. This trade-off leads to different LPTCs having different receptive fields, where their size and shape should reflect their underlying computational task (Meyer et al., 2011). Even though HSN undoubtedly responds strongly to optic flow, and its time-averaged responses are on par with those of HSNE (Nordström et al., 2008), it is thus possible that it evolved to serve a species-specific role in visual processing. As the name implies, male hoverflies have a distinctive behavior where they hover stationary in their territory before switching to high-speed pursuits of conspecifs (e.g., Collett & Land, 1978). Hovering is a challenging task and hoverflies will often place themselves in small openings in the foliage or aligned with the edges of bushes where they can use visual cues, such as vertical edges, to maintain their stationary stance. In windy conditions, they continuously adjust their hovering position, staying synchronized with the swaying foliage. The vertically elongated, narrow receptive field of HSN, along with its increased sensitivity to structural features within natural scenes (red, Figure 3), would be well adapted to respond with high gain to prominent vertically oriented edge features and, thus, may be an important adaptation to aid in such forward fixation during hovering. The same velocity-servo mechanism proposed to underlie the syn-directional response to rotational motion in insects would, thus, allow this neuron to fixate frontal features and correct for both yaw and sideslip (Hausen & Egelhaaf, 1989; Reichardt, Poggio, & Hausen, 1983), a strategy observed also in hovering hawk moths (Farina, Kramer, & Varjú, 1995).

Our stimuli provide pure, continuous preferred-direction yaw rotation at near-optimal but constant velocity, which is an unrealistic optic flow input for a hoverfly during normal flight. Due to the technical difficulties associated with intracellular electrophysiology, closed-loop recordings of LPTCs are still impossible in freely behaving flies. However, within the narrow ipsilaterally biased receptive fields of HSN and HSNE, prolonged progressive motion would be generated locally during forward flight, albeit not at constant speed. Importantly, even if the stimuli as presented here may not be a realistic representation of the
type of optic flow experienced by a behaving insect in the field, we see similar spatial integration properties irrespective of the velocity used (Figure 9A, and Barnett et al., 2010), suggesting that our conclusions are not critically dependent on the velocity itself.

Motion adaptation

During prolonged motion stimulation, as in the experiments described here, strong motion adaptation is recruited. Dynamic adaptation to the natural scenes used here occurs very rapidly, within the first few hundred milliseconds, with no further significant adaptation for periods of up to 4 s after the first 300 ms (Barnett et al., 2010). The rapid time course for adaptive effects is confirmed using sinusoidal grating stimuli (Nordström et al., 2011), where the contrast gain reduction is evident after as little as 20-ms adaptation, while the antagonistic motion aftereffect builds up during the first few hundred milliseconds of stimulation. It is, thus, unlikely that the response to the natural scenes that we see here, when averaged across different initial display phases, recruits strong differences in adaptation state, compared with changes that would occur during the first few hundred milliseconds.

The contrast gain reduction associated with motion adaptation is predominantly locally recruited, in those input dendrites directly stimulated with visual motion (Nordström & O’Carroll, 2009), and is associated with increased sensitivity to changes in stimulus velocity and direction (Kurtz, Egelhaaf, Meyer, & Kern, 2009; Maddess & Laughlin, 1985). The consequences of such adaptation in natural scenes is difficult to predict since the gain of local motion detectors viewing a moving natural scene at any instant will be strongly influenced by features of the scene that passed by the same point in recent history. Recent work using sinusoidal gratings show that while contrast gain is locally recruited in those input dendrites directly stimulated by motion, the antagonistic after-potential operates at a global level by affecting subsequent responses in the neuron’s entire receptive field, whether directly stimulated or not (Nordström & O’Carroll, 2009). This implies that if the neuron responds strongly to a feature in one part of the image, the subsequent antagonistic after-potential would suppress responses to features anywhere in the scene, even if these pass over a completely different part of the dendritic tree.

Contrast gain reduction will be strongly influenced by higher order scene statistics. For example, an EMD viewing a heavily cluttered part of a scene could be much more strongly motion adapted than one viewing a sparse part of the same image. Depending on the subsequent features that pass the strongly adapted EMD, gain reduction may either contribute to a weaker response than a weakly adapted EMD (for subsequent low or moderate contrast features) or an increased response modulation by higher contrast features (due to relief from saturation). Interestingly, since dynamic adaptation is so rapid (Nordström et al., 2011), and spatial gain control is effectively instantaneous, the adaptation state in free-flight might change dramatically under diverse behaviors, such as during conspecific pursuit versus hovering flight.

Spatial gain control

LPTC responses spatially saturate when the size of a sinusoidal grating stimulus increases (Haag et al., 1992; Hausen, 1982b). This response saturation can be accurately recaptured by an LPTC compartment model (Single et al., 1997). No such compartmental model currently exists for hoverfly HS neurons, since we do not have the required detailed 3-dimensional morphometric data for the dendritic structure nor the required detailed biophysical measurements. While compartmental models are available for blowfly tangential neurons (e.g., Weber, Machens, & Borst, 2009), these lack the narrow, strongly directed receptive fields of male hoverflies. If such data become available in the future, it would be interesting to combine realistic models for dendritic integration in hoverfly HS neurons with either elaborated EMD models or our locally measured responses (Figure 8) to see if these resolve the remaining discrepancies between our models and globally measured responses. Similarly, it will be interesting for further experiments to test the hypothesis that the HSN dendritic structure and receptive fields may be optimized to enhance responses to vertically aligned features as an aid for stabilizing sideslip and yaw during hovering, in addition to a role in optic flow analysis during other behavior.

Acknowledgments

We thank the Managers of the Adelaide Botanical Gardens for allowing insect collection. The work was funded by ARC (LP 0667744), AFOSR (FA2386-10-1-4114), and VR (2008-2933).

Author contributions: David C. O’Carroll and Paul D. Barnett contributed equally to the work.

Commercial relationships: none.

Corresponding author: Karin Nordström.

Email: karin.nordstrom@neuro.uu.se.

Address: Department of Neuroscience, Uppsala University, Box 593, 751 24 Uppsala, Sweden.

References


