Encoding of Yaw in the Presence of Distractor Motion: Studies in a Fly Motion Sensitive Neuron

Suva Roy,¹ Shiva R. Sinha,^{1,2} and Rob de Ruyter van Steveninck¹

¹Department of Physics, Indiana University, Bloomington, Indiana 47405, and ²Medtronic Surgical Technologies Inc., Louisville, Colorado 80027

Motion estimation is crucial for aerial animals such as the fly, which perform fast and complex maneuvers while flying through a 3-D environment. Motion-sensitive neurons in the lobula plate, a part of the visual brain, of the fly have been studied extensively for their specialized role in motion encoding. However, the visual stimuli used in such studies are typically highly simplified, often move in restricted ways, and do not represent the complexities of optic flow generated during actual flight. Here, we use combined rotations about different axes to study how H1, a wide-field motion-sensitive neuron, encodes preferred yaw motion in the presence of stimuli not aligned with its preferred direction. Our approach is an extension of “white noise” methods, providing a framework that is readily adaptable to quantitative studies into the coding of mixed dynamic stimuli in other systems. We find that the presence of a roll or pitch (“distractor”) stimulus reduces information transmitted by H1 about yaw, with the amount of this reduction depending on the variance of the distractor. Spike generation is influenced by features of both yaw and the distractor, where the degree of influence is determined by their relative strengths. Certain distractor features may induce bidirectional responses, which are indicative of an imbalance between global excitation and inhibition resulting from complex optic flow. Further, the response is shaped by the dynamics of the combined stimulus. Our results provide intuition for plausible strategies involved in efficient coding of preferred motion from complex stimuli having multiple motion components.

Key words: adaptation; information theory; insect vision; motion encoding; reverse correlation; spikes

Introduction

Flies often perform complex maneuvers while navigating a 3-D environment (Wagner, 1986; Fry et al., 2003), which results in a wide range of angular velocities. Behavioral studies (Collett and Land, 1975; Budick et al., 2007) have shown that such maneuvers also lead to complex optic flow with rotational components mixed in a way that is hard to disambiguate. Wide-field motion-sensitive neurons in the lobula plate region of the fly’s brain have been known to respond selectively to motion along different directions (Hausen, 1981, 1982a,b; Hengstenberg et al., 1982). Anatomical studies have identified ~60 different motion-sensitive neurons in the lobula plate (Hausen, 1981, 1982a; Hengstenberg et al., 1982; Borst and Haag, 2002), from which only a small subset of neurons, namely H1, V1, HS, and VS, have been extensively studied, owing to their ease of detection and robustness of response.

Much of the classic work on visual motion detection in insects and vertebrates employed simple 1-D spatiotemporal stimuli

such as bar, grating, and sinusoidal patterns moving along a particular direction, usually along the preferred direction of the neuron (Hassenstein and Reichardt, 1956; Hausen, 1982b; van Santen and Sperling, 1984). Despite their practical advantages, these simplified stimuli fail to capture essential elements of a complex image flow generated during flight, such as rotations about the three principal body axes, and translation. These motion elements have been demonstrated in the flights of not only blowflies, but also of bees and wasps (Srinivasan et al., 1991; Hateren and Schilstra, 1999; Schilstra and van Hateren, 1999; Zeil et al., 2008). The receptive field structures of LPTCs further suggest sensitivity to movement along other directions besides the preferred direction (Krapp et al., 2001). It is thus conceivable that motion encoding is affected by the presence of additional rotational degrees of freedom. The geometry of spatial patterns (Borst et al., 1993; Saleem et al., 2012) and velocity components from other rotation types (Kern et al., 2005, 2006) also affect the response of motion-sensitive neurons. However, a quantitative study of preferred motion encoding from a complex multistimulus input in the context where a stimulus is not aligned to the preferred direction of the neuron has so far been lacking.

Here, we investigate the effect of additional wide-field components on the encoding of yaw by H1, a motion-sensitive neuron in the lobula plate of the fly. The stimulus consists of a 2-D pattern subjected to combined rotation about two axes. Combining techniques of information theory and dimensional reduction, we quantify the reduction in H1’s yaw encoding efficiency in the presence of distractor and show that H1 also encodes certain

Received Oct. 7, 2014; revised Feb. 17, 2015; accepted March 12, 2015.

Author contributions: S.R., S.R.S., and R.d.R.v.S. designed research; S.R. performed research; S.R. and S.R.S. contributed unpublished reagents/analytic tools; S.R. analyzed data; S.R. and R.d.R.v.S. wrote the paper.

We thank Anne C. Mennen for participating in experiments [supported by National Science Foundation Grant 1156540 for the Research Experiences for Undergraduates (REU) Program]; and Philip L. Childress for technical support.

The authors declare no competing financial interests.

Correspondence should be addressed to Rob de Ruyter van Steveninck, Department of Physics, Indiana University Bloomington, Bloomington, IN 47405. E-mail: deruyter@indiana.edu.

DOI:10.1523/JNEUROSCI.4256-14.2015

Copyright © 2015 the authors 0270-6474/15/356481-14\$15.00/0

Figure 1. The spatial map and the velocity waveforms for yaw and distractor degrees of freedom. **A**, Spatial map is created by smoothing a matrix of random values with a Gaussian of $\sigma = 6^\circ$. The spatial extension is 360° along the x - and y -axes with a grid resolution of 0.1° . Contrast was set to maximum. **B**, Snapshot of an image seen in the TEK 608 monitor. Pixel intensities are interpolated for successive images displayed on the screen at 500 Hz frame rate. **C, D**, Yaw velocity traces (blue) and distractor (roll or pitch) velocity traces (red) within a 200 ms window for six segments, each lasting 3 s. The velocities of each of the example traces are drawn from Gaussian distributions with $\sigma = \{0, 100\}^\circ/\text{s}$. “Rep” and “Non-rep” stand for velocity traces that are repeated and nonrepeated, respectively, across trials. A total of 200 trials are used for analysis. The solid dark and solid pale lines represent velocity traces in the same segment from two different trials.

distractor features; the latter depending on the global excitation and inhibition from the optic flow. We further show that the neural response matches the collective dynamic range of the input, which indicates that response scaling is preserved for these multiple motion components.

Materials and Methods

Fly preparation. Wild-caught blowflies, *Calliphora vicina*, were housed in an enclosed cabinet that was maintained at a temperature of 21°C and a relative humidity of 55%, and was illuminated with a 12 h light/dark cycle. Experiments were conducted primarily in adult female flies, and consistent results were obtained from six flies. We also checked for consistency of results against male blowflies. The fly was placed in a cylindrical tube, its wings and legs restrained using dental wax. Its head was immobilized by wax bridges from the genae (“cheeks”) to the rim of the tube, leaving the proboscis free so that the fly could be fed in between experiments. The spiracles were kept free to ensure proper respiratory intake during the experiment. After opening the rear of the head capsule, dorsolateral muscles and air sacs covering the lobula plate were removed. The fly holder was then transferred to a goniometer platform near the display screen. Room temperature was maintained between 19°C and 20°C and was monitored throughout the experiment.

Stimuli. A spatial pattern was created generating a 2-D random matrix on a square $0.1^\circ \times 0.1^\circ$ grid, filtering these data with a 2-D spatial Gaussian of SD $\sigma = 6^\circ$ and thresholding the resulting numbers around their mean to produce a set of binary intensity values. Subsequently, the image was linearly filtered to prevent spatial aliasing during stimulus presentation. The temporal stimuli consist of velocity waveforms corresponding to yaw, pitch, and roll, each of which is drawn from independent Gaussian distributions. Because H1 is primarily sensitive to horizontal motion, roll and pitch motion will generically be referred to as “distractor” motion. A wide range of SD values for roll, $\sigma_R = \{100, 300, 1000, 3000, 5000, 7000, 10000\}^\circ/\text{s}$, and pitch, $\sigma_P = \{10, 30, 100, 300, 500, 700, 1000\}^\circ/\text{s}$, were chosen. The SD of yaw motion (σ_Y) was kept unchanged at $100^\circ/\text{s}$ for all yaw–distractor SD pairs. The relative strength of roll and pitch to yaw is defined as the ratio of their SDs (i.e., $\xi_R = \sigma_R/\sigma_Y$ and $\xi_P = \sigma_P/\sigma_Y$, respectively). For each value of ξ_R and ξ_P , six different motion variants were presented sequentially in a trial (Fig. 1C), with each segment lasting 3 s. Segments 1, 2, and 6 consist of repeated yaw with nonrepeated distractor motion, while segments 3, 4, and 5 consist of yaw and distractor motions both of which were nonrepeated across trials. A total of 210 trials were used for each experiment.

We choose different ranges for ξ_R and ξ_P because roll motion, which is generated about a point of origin, introduces local velocity components that are different from those generated by pitch motion, which acts in the vertical direction. This serves two purposes. First, we obtain a reasonably wide range of distractor noise over which we can test the efficiency of H1.

With the given size of the visual field (Fig. 1B), the tangential component induced by a $10,000^\circ/\text{s}$ roll has a velocity of $\sim 3000^\circ/\text{s}$ at the edge of the field. Tangential components at points closer to the center are smaller. Thus, the tangential velocity components of roll can effectively reach up to an order of 10 times the yaw velocity ($\sigma_Y = 100^\circ/\text{s}$). Since the maximum value of σ_P is $1000^\circ/\text{s}$, pitch velocities also can reach up to an order of 10 times the yaw velocity. Similarly, velocity components of roll and pitch at the lower limits of σ_R and σ_P are ~ 10 times lower than that for yaw. Second, the velocities lie in the neighborhood of those generated during actual flight, though one must recognize the tradeoff between choosing a wide range of parameter values and the ecological relevance of the values. Behavioral studies suggest that the angular velocity of the head of the fly reaches up to $\sim 2500^\circ/\text{s}$ for yaw, $\sim 1200^\circ/\text{s}$ for roll, and $< 1000^\circ/\text{s}$ for pitch during saccades. The velocities during nonsaccades are typically $\leq 100^\circ/\text{s}$ (Wagner, 1986; Hateren and Schilstra, 1999; Kern et al., 2005). Therefore, the velocities induced by roll and pitch used in our experiments approximately overlap with the range of angular velocities accessed during saccadic flights. Since the retinal image speed is on the order of $100^\circ/\text{s}$ for nonsaccadic flights, which account for $\sim 60\%$ of the total flight time (Hateren and Schilstra, 1999), we choose a value of σ_Y that is of the same order. This allows us to study yaw encoding in the presence of fluctuations ranging from nonsaccadic to saccadic limits.

Euler rotations in a 3-D coordinate system do not commute. In simple terms, this means that the collective rotation of two successive rotations about two different axes depends on the order in which the rotations are performed. However, in our stimuli, yaw, roll, and pitch rotations are small from one frame to the next, such that the rotations commute to a good approximation. Therefore, the order in which rotations are combined does not matter. Also, the rotations are displayed on a 2-D screen, such that in a flattened approximation, homogeneous movements along the horizontal and vertical axes of the screen are equivalent to yaw and pitch motion, respectively. Roll motion, resulting from rotation about the longitudinal axis, is always produced about the center of the screen.

The stimulus was presented on a Tektronix 608 monitor at a 500 Hz frame rate and a radiance of $150 \text{ mW} \cdot \text{sr}^{-1} \cdot \text{m}^{-2}$ (sr: steradian), which corresponds to an estimated 5×10^4 transduced photons per second per photoreceptor. The spatial pattern was displayed on a hexagonal raster of 833 pixels with 38° horizontal and 44° vertical extents (Fig. 1B), which subtends $\sim 8\%$ of the solid angle accessed by the eye of the fly (Krapp and Hengstenberg, 1997; Lewen et al., 2001). Because the pixels were arranged in a hexagonal array, the intensity of each pixel was calculated as the linear weighted average of intensities of four nearest-neighbor pixels. The fly was positioned at 8.4 cm from the screen, such that the projected interpixel separation approximately matches the interommatidial distance of 1.5° in the blowfly (Land, 1997). The match is, however, approximate because pixels were spatially blurred to prevent aliasing. This makes the experiment insensitive to the misalignment between the raster

of the eye of the fly and the pixel-sampling raster. The fly watched the center of the display screen at an azimuth angle of 30°. The stimulus files were precomputed using MathWorks Matlab version 7.4. A custom-written C program was used to control a National Instruments PCI-6259 data acquisition card that processed the conversion of stimulus data to the analog voltage displayed on the screen.

Electrophysiology. Tungsten microelectrodes (1 MΩ; tip diameter, ~5 μm; FHC) were used to make differential single-unit recordings from H1 in the contralateral lobula plate in the right half of the head. H1 was identified by its characteristic excitatory response to horizontal back-to-front motion and its suppressive response to horizontal front-to-back motion. Control experiments using bar patterns of a 7° wavelength were conducted to adjust elevation and twist, such that the response of H1 is the maximum for horizontal motion and the minimum for vertical motion. Care was taken to achieve distinct isolated spikes with a high signal-to-noise ratio of at least 6. The voltage difference between the recording and reference electrodes was filtered, amplified, and thresholded by a window discriminator to generate discrete pulses. The pulses were time stamped with precision at 10 μs intervals by the National Instruments data acquisition card that generates the analog output. This guaranteed that the displayed stimuli and recorded spikes were synchronized to the same board clock. During the course of the experiment, the drift in spike amplitude was <30% and the amplitude remained well above the spike-discriminating threshold set before the start of the experiment.

Spike train information. Noise, both internal and external to the system (de Ruyter van Steveninck and Bialek, 1995; Manwani and Koch, 1999; White et al., 2000), limits the information transmitted by a spike train. The presence of a distractor effectively adds another “noise” component, which further limits the transmitted information. We quantify this by estimating the information transmitted about a repeated yaw motion in the presence of a nonrepeated distractor motion across multiple trials (Fig. 1C,D, segments 1,2; de Ruyter van Steveninck et al., 1997). To remove the effect of initial transients, we discard data from the first 10 trials of each experiment.

To estimate entropy of the spike train, spikes and spike-absent events within a time bin $\Delta t = 1$ ms were assigned values 1 and 0, respectively. This guaranteed a maximum of one spike in each time bin. A distribution of binary values or “words,” $P(W)$, was obtained from the sequence of 1 and 0 s lying within each time window, $T \geq \Delta t$ (de Ruyter van Steveninck et al., 1997). The entropy rate for infinite word length, or $T \rightarrow \infty$, was obtained from linear extrapolation of the flattened part of the R versus $1/T$ curve. For a time window T , the total entropy rate of the spike train is obtained from $P(W)$ as follows:

$$R_{\text{Tot}}(T, \Delta t) = -\frac{1}{T} \sum_W P(W) \log_2 P(W). \quad (1)$$

The noise entropy rate is obtained by averaging the distribution of words obtained at each time instance $P(W|t)$, over the time instances in a trial, as follows:

$$R_{\text{Noise}}(T, \Delta t) = \left\langle -\frac{1}{T} \sum_W P(W|t) \log_2 P(W|t) \right\rangle_t. \quad (2)$$

The rate of yaw information transmitted by the spike train in the presence of the distractor is given by the difference between the total entropy rate and the average noise entropy rate, as follows:

$$R_{\text{Info}}(T, \Delta t) = R_{\text{Tot}}(T, \Delta t) - R_{\text{Noise}}(T, \Delta t). \quad (3)$$

Information per spike can be obtained either by setting $T = \Delta t$ and dividing by $\langle r \rangle$ in Equation 3, or by using the time-varying firing rate $r(t)$ (Brenner et al., 2000b), as follows:

$$I(\text{spk}) = \frac{1}{T_{\text{trial}}} \sum_{i=1}^{T_{\text{trial}}/\Delta t} \Delta t \left(\frac{r(t_i)}{\bar{r}} \right) \log_2 \left(\frac{r(t_i)}{\bar{r}} \right). \quad (4)$$

Here, T_{trial} is the time period of repeated stimulus. To estimate the information carried by the spike train about repeated distractor motion in the

presence of nonrepeated yaw, we use the same method of analysis as above, but with data from the sixth segment of the trial (Fig. 1C,D).

Error estimation. It is well known that direct estimates of the entropy rate (R_{Tot}) and the noise entropy rate (R_{Noise}) are biased due to under-sampling, and that this bias becomes more prominent at larger values of T . (Treves and Panzeri, 1995; Panzeri et al., 2007). With the assumption that spike correlations are finite, the true entropy rate can be estimated by taking the limiting conditions that data size $\rightarrow \infty$ and $T \rightarrow \infty$ (Treves and Panzeri, 1995; Strong et al., 1998), as follows:

$$\lim_{\substack{T \rightarrow \infty \\ \text{size} \rightarrow \infty}} \left[R_{\text{Tot}}(T, \Delta t) + \frac{R_{\text{Tot}}(T, \Delta t)}{\text{size}} + \dots \right] = R_{\text{Tot}}(T, \Delta t). \quad (5)$$

Linear extrapolation is performed separately for the curves R_{Noise} and R_{Tot} , using only the time windows in which each curve satisfies linearity. The information rate (R_{Info}) is obtained by subtracting the extrapolated R_{Noise} from the extrapolated R_{Tot} .

To obtain statistical error bounds for entropy estimates, we used bootstrap analysis (Efron and Tibshirani, 1993). From the 200 trials, an ensemble of 100 datasets (bootstrap samples), each containing 200 trials, was obtained by resampling. The SD of entropy for different values of T was obtained from the bootstrap samples and was used to minimize χ^2 (Bevington and Robinson, 2003; i.e., the least-square error between estimates from the bootstrap and the parent distribution). The slope of the linear extrapolation is determined from the best-fit curve with a 95% confidence interval.

Reverse correlation methodology. To extract the set of the most significant stimulus features associated with spikes, we use reverse correlation analysis (de Boer and Kuyper, 1968), which was extended to second order by sampling spike-triggered covariances (STCs; de Ruyter van Steveninck and Bialek, 1988). This allows the extraction of a low-dimensional stimulus space, which is spanned by a set of orthogonal eigenvectors associated with leading eigenvalues of the covariance matrix. Since the stimuli consist of combined yaw–distractor motion, both yaw and distractor stimuli are used in the analysis.

The spike time resolution in our analysis is 1 ms, but the stimulus velocity is sampled at 2 ms, corresponding to the frame rate of 500 Hz. In all results presented in this work, the yaw (Y) and distractor (D) velocities are scaled to their respective SDs, σ_Y and σ_D . The spike-triggered average (STA) is obtained by averaging 100 ms time windows (τ) of stimulus fluctuations ($\tilde{s}_0$) preceding each spike, over all the N spikes, as follows:

$$\text{STA}(\tau) = \frac{1}{N} \sum_{n=1}^N \tilde{s}_0(t_n + \tau) / \sigma, \quad (6)$$

$$= \langle \tilde{s}_0(\tau) \rangle / \sigma, \quad (7)$$

where t_n represents the time of n th spike. The STC is calculated by using the concatenated spike-triggered yaw and distractor stimuli with their STAs subtracted (de Ruyter van Steveninck and Bialek, 1988), as follows:

$$[C_{\tau, \tau_j}]_{\text{spk}} = \left\langle \left(\frac{\tilde{s}_{0Y} - \langle \tilde{s}_{0Y} \rangle}{\sigma_Y}, \frac{\tilde{s}_{0D} - \langle \tilde{s}_{0D} \rangle}{\sigma_D} \right)_{\tau_i} \times \dots, \left(\frac{\tilde{s}_{0Y} - \langle \tilde{s}_{0Y} \rangle}{\sigma_Y}, \frac{\tilde{s}_{0D} - \langle \tilde{s}_{0D} \rangle}{\sigma_D} \right)_{\tau_j} \right\rangle_{\text{spk}} \quad (8)$$

$$= \langle (\tilde{s}_Y, \tilde{s}_D)_{\tau_i} (\tilde{s}_Y, \tilde{s}_D)_{\tau_j} \rangle_{\text{spk}} \quad (9)$$

where $i, j \in [1, 100]$ and $\tilde{s}_Y, \tilde{s}_D$ represent stimuli with the mean subtracted normalized to SD. The prior covariance is obtained by averaging over all time bins in the experiment, as follows:

$$[C_{\tau, \tau_j}]_{\text{pri}} = \langle (\tilde{s}_Y, \tilde{s}_D)_{\tau_i} (\tilde{s}_Y, \tilde{s}_D)_{\tau_j} \rangle_{\text{expt}}. \quad (10)$$

Eigenvalues and eigenvectors are then calculated from the difference between the spike-triggered and prior covariance matrices, as follows:

$$\Delta C_{\tau, \tau_j} = [C_{\tau, \tau_j}]_{\text{spk}} - [C_{\tau, \tau_j}]_{\text{pri}} \quad (11)$$

$$\equiv \begin{bmatrix} YY & YD \\ DY & DD \end{bmatrix}. \quad (12)$$

The off-diagonal blocks depicted in Equation 12 represent yaw–distractor cross-covariances YD and DY , while the diagonal blocks represent the yaw and distractor autocovariances YY and DD .

Significance testing. Although the symmetry of ΔC guarantees that the eigenvalues (λ) are real, they can be positive or negative, depending on an increase or decrease in variance along the eigenvector directions. To extract the set of most significant eigenvalues, we used a radius of threshold determined from the Wigner distribution (Wigner, 1958). The eigenvalue spectrum is parsed into a relevant set and an irrelevant set that embody the correlations and random noise in ΔC , respectively, such that the eigenvalue distribution of the irrelevant set resembles the distribution of eigenvalues of a symmetric random matrix (C_{Rnd}) with no element-wise correlations. In the limit of infinite data, this distribution converges to the Wigner semicircle, as follows:

$$\rho(\lambda; M, \sigma_M) = \frac{1}{2\pi M \sigma_M^2} \sqrt{4M\sigma_M^2 - \lambda^2}, \quad (13)$$

$$= \frac{2}{\pi R^2} \sqrt{R^2 - \lambda^2}, \quad (14)$$

where M is the dimensionality of the covariance matrix, σ_M^2 is the variance of matrix elements of C_{Rnd} , and $R = 2\sqrt{M}\sigma_M$ is the radius of the Wigner semicircle. The cumulative Wigner distribution is given by the following:

$$\rho_c(\lambda; M, \sigma_M) = \frac{\lambda \sqrt{R^2 - \lambda^2}}{\pi R^2} + \arcsin\left(\frac{\lambda}{R}\right) + \frac{1}{2}. \quad (15)$$

We implement this method by first generating C_{Rnd} from N randomly chosen prior stimulus time windows, where N represents the total number of spikes. We then determine the radius (R) of the semicircle from a nonlinear least-squares fit of the irrelevant eigenvalues of ΔC to the theoretical cumulative Wigner distribution (Eq. 15). We can also formulate R analytically. Each off-diagonal element of C_{Rnd} is the product of two independent random variables with an SD of 1, averaged over N spikes. Therefore, the SD of the elements of C_{Rnd} is $1/\sqrt{N}$. Using this, R can be written as $R = 2\sqrt{M}\sigma_M = 2\sqrt{M/N}$. The eigenvalues lying outside the perimeter of the circle defined by R are considered relevant. A bootstrap resampling procedure accomplished by randomly shifting the spike train relative to the stimulus (Schwartz et al., 2006) was also used to cross-check the set of relevant eigenvalues.

Input–output function. Bayes' theorem allows us to map our prior stimulus knowledge to the spiking response of H1. The functional form of this map is represented by the input–output function (Brenner et al., 2000a), as follows:

$$\frac{P(\text{spk}|\tilde{s}_1, \dots, \tilde{s}_K)}{P(\text{spk})} = \frac{P(\tilde{s}_1, \dots, \tilde{s}_K|\text{spk})}{P(\tilde{s}_1, \dots, \tilde{s}_K)}, \quad (16)$$

$$\frac{r(\tilde{s}_1, \dots, \tilde{s}_K)}{\bar{r}} = \frac{P(\tilde{s}_1, \dots, \tilde{s}_K|\text{spk})}{P(\tilde{s}_1, \dots, \tilde{s}_K)}, \quad (17)$$

where $P(\tilde{s}_1, \dots, \tilde{s}_K)$ and $P(\tilde{s}_1, \dots, \tilde{s}_K|\text{spk})$ represent the prior and the posterior distribution of stimulus projections ($\tilde{s}_k$) on the relevant eigenvectors ($\hat{e}_k$). The normalized firing rate shown in Equation 17 describes the input–output relationship.

Yaw–distractor subspace information. The information per spike associated with a relevant dimension can be found by calculating the entropy difference between distributions of spike conditional stimulus projections and prior stimulus projections. For a K -dimensional subspace, this is given by the following:

$$I(\text{spk}) = H(P(\tilde{s}_1, \dots, \tilde{s}_K)) - H(P(\tilde{s}_1, \dots, \tilde{s}_K|\text{spk})). \quad (18)$$

Note that this formulation (Eq. 18) is the same as that used by DeWeese and Meister (1999), who define it as “stimulus-specific” information.

According to this definition, the information gained from the observation of a symbol can be positive or negative. Recall that positive eigenvalues are associated with the broadening of distribution, which leads to a posterior entropy that is higher than the prior, and consequently negative, symbol information. Similarly for negative eigenvalues, the symbol information is positive. Therefore, choosing this formulation instead of the alternative “surprise” (DeWeese and Meister, 1999) allows us to make a connection between symbol information and the expansion or compression of the stimulus space.

The information associated with nonoccurrence of spikes can be formulated as follows:

$$P(\tilde{s}_1, \dots, \tilde{s}_K) = P(\text{spk}) \cdot P(\tilde{s}_1, \dots, \tilde{s}_K|\text{spk}) + P(\text{no-sp}) \cdot P(\tilde{s}_1, \dots, \tilde{s}_K|\text{no-sp}), \quad (19)$$

$$I(\text{no-sp}) = H(P(\tilde{s}_1, \dots, \tilde{s}_K)) - H(P(\tilde{s}_1, \dots, \tilde{s}_K|\text{no-sp})). \quad (20)$$

Though a symbol can carry negative information, the information averaged over the symbols or, mutual information (MI) is still positive, as follows:

$$MI = \sum_{\text{spk}} I(\text{spk}) \cdot P(\text{spk}) + \sum_{\text{no-sp}} I(\text{no-sp}) \cdot P(\text{no-sp}). \quad (21)$$

We perform this information-theoretic analysis on the yaw subspace, the distractor subspace, and the combined yaw–distractor subspace.

Results

Firing pattern modulation

H1 is responsive to yaw motion, and the peristimulus time histogram (PSTH) for repeated presentations of the same pseudorandom yaw waveform therefore shows strong variations in firing rate, often with sharp onsets and offsets. The addition of random nonrepeated distractor motion to the original yaw stimulus will tend to wash out the structure of such yaw-induced rate fluctuations, lowering the information that H1 carries about yaw. This is illustrated in Figure 2 for both roll and pitch distractors at $\xi_R = \sigma_R/\sigma_Y = 10$ and $\xi_P = \sigma_P/\sigma_Y = 3$, respectively. For the pure yaw case, the spike raster locks on to preferred yaw motion but shows suppression to nonpreferred yaw motion (Fig. 2A). From the modulation in firing rate, we can compute the information per spike according to Equation 4, and for the conditions of the experiment shown in Figure 2, we get $I(\text{spk}) = 1.82$ bits. The spike rasters for yaw–roll and yaw–pitch combinations show similar locking behavior but have increased variability compared with the pure yaw case. This increase in variability results in broadening of the PSTH peaks (Fig. 2B) and a decrease of single-spike information. Conversely, regions with very low activity experience an increase in firing rate (Fig. 2B, insets). As an example of the spread in PSTH peaks, we fit a Gaussian to an isolated peak in the PSTH (Fig. 2C) over the region where its shape is preserved. The SD of the Gaussian increases from 4 ms with pure yaw, to 14 ms with combined yaw and roll, and to 8 ms with combined yaw and pitch. The effect of the distractor on firing rate is relatively small at intermediate values of ξ_R and ξ_P . For example, the mean firing rate ($\bar{r}$) changes from 35 spikes/s without the distractor to 31 spikes/s at $\xi_R = 10$ and 34 spikes/s at $\xi_P = 3$. The effect is much larger at higher values of ξ , especially for roll, when $\bar{r}$ decreases to ~ 25 spikes/s at $\xi_R = 100$.

Effect of distractor on yaw information rate

A more detailed analysis of spike train information takes into account how spike sequences are reproduced across trials. For pure yaw ($\sigma_Y = 100^\circ/\text{s}$), the R_{Tot} is 175 bits/s and the R_{Noise} is 118 bits/s (Fig. 3A), resulting in an R_{Info} of 57 bits/s. With a combined roll ($\xi_R = 10$), R_{Info} drops to 38 bits/s, but R_{Tot} and R_{Noise} both

Figure 2. Impact of distractor on spiking response to repeated yaw stimulus. **A**, A 60-trial spike raster over a 350-ms-long time window, from pure yaw at $\sigma_Y = 100^\circ/\text{s}$, combined yaw-roll at $\xi_R = 10$, and combined yaw-pitch at $\xi_P = 3$. **B**, PSTH (2 ms bins) shows response averaged over 200 trials. **C**, Gaussian fits to isolated peaks in PSTH to determine spiking precision are shown for the three cases. The stimuli corresponding to these responses are shown in segment 1 and segment 2 of Figure 1, **C** and **D**. An example of change in local firing rate from 0.1 spikes/s representing a yaw-suppressed state to 13 and 4 spikes/s, respectively, for roll- and pitch-induced excitatory states is shown by the insets in **B**.

increase. Notably, the increase in R_{Noise} is higher by almost a factor of 2 than the increase in R_{Tot} . This suggests that the interspike correlation decreases and the variability in spike sequences across trials increases. Because limited sampling affects R_{Noise} more than R_{Tot} , the R_{Noise} curve collapses earlier than the R_{Tot} curve (Strong et al., 1998).

Spike train information rates at different distractor strengths parametrized by ξ_R and ξ_P are shown in Figure 4A. In the region where ξ_R increases from 1 to 10, R_{Tot} remains virtually unchanged at 181 bits/s, while R_{Noise} increases from 122 to 143 bits/s. For $\xi_R > 10$, both R_{Tot} and R_{Noise} steadily decrease. The onset of the decrease of R_{Tot} occurs earlier than the decrease of R_{Noise} . This suggests that when the strength of the distractor is low, R_{Tot} increases due to a decrease in interspike correlations. When the strength of the distractor becomes very large, the mean firing rate drops. This leads to a lower number of spikes and therefore a smaller repertoire of binary words (W), which reduces both R_{Noise} and R_{Tot} .

Over the range of ξ_R from 1 to 100, R_{Info} decreases monotonically from 57 to 8 bits/s (i.e., by a factor of 7; Fig. 4A). Over the range of ξ_P from 0.1 to 10, R_{Info} decreases from 58 to 22 bits/s (i.e., by a factor of 2.6; Fig. 4B). The decrease in R_{Info} for the two cases, however, is subtly different. While the decrease in R_{Info} for pitch occurs largely due to an increase in R_{Noise} , for roll it occurs due to a decrease in both R_{Tot} and R_{Noise} . Given the geometry of the display used in our experiments (Fig. 1B), the local tangential velocity components of roll at the edge of display is $\sim 3000^\circ/\text{s}$ for a roll velocity of $10,000^\circ/\text{s}$ (see Materials and Methods). Conceivably, the spatial contrast reduces at these high velocities, which could very well decrease both R_{Tot} and R_{Noise} . The information rate shows approximately a linear dependence on the logarithm of ξ_R and ξ_P . The largest error in entropy estimates stem from the

uncertainty in determining the slope of the extrapolated curves (Fig. 3). Although the absolute value of R_{Info} changes from one fly to the other, we found the characteristic dependence of R_{Info} on ξ to be consistent across all six tested flies.

Reverse correlation

The previous subsection quantified the effect of distractor motion on the information encoded about yaw. Here we sketch a more detailed picture of the interactions between yaw and distractor motion associated with spiking response. In principle, this is a problem of describing interactions in a high-dimensional space of stimulus waveforms, but in practice it is often the case that only a few dimensions are relevant. The analysis is based on a generalization of the reverse correlation method (de Boer and Kuypers, 1968), which starts by extracting a few stimulus dimensions that define the relevant subspace. Given that we use a symmetric Gaussian stimulus, this method is guaranteed to provide unbiased estimates of the linear filters, namely, the STA and the eigenvectors of the STC matrix (Chichilnisky, 2001; Paninski, 2003). To estimate the filters, we need many independent samples of the stimulus waveform. For this, we use data from the fifth segment (Fig. 1C,D) of the

trial, where yaw and distractor waveforms are nonrepeated across trials. The results are illustrated using the following subset of parameter values: $\xi_R = \{0, 3, 10, 30\}$ and $\xi_P = \{0, 1, 3, 10\}$.

Spike-triggered average

The first moment of the stimulus leading up to a spike is given by the STA. For calculating the STA, we use the 100 ms history of yaw and distractor motion preceding each spike (de Ruyter van Steveninck and Bialek, 1988).

As expected, the yaw STAs from both yaw-roll and yaw-pitch stimulus combinations are positive and unimodal, with a peak preceding the spike by $\sim 20 \pm 2$ ms. As ξ_R increases from 0 to 30, the yaw STA amplitude decreases to 45% of its amplitude at $\xi_R = 0$ (Fig. 5A). The roll STA (Fig. 5A, dotted line) is positive and relatively small, but shows signs of increase with an increase in ξ_R , which suggests that H1 is excited more by positive roll than by negative roll. Over the range of $\xi_R \in [0, 30]$, the mean firing rate, $\bar{r}$, decreases by $<10\%$. The pitch STA (Fig. 5B, dotted line) has a downward peak, the amplitude of which increases with an increase in ξ_P . This could be due to horizontal motion components arising from the vertical movement of slanted pattern edges (Marr and Ullman, 1981). Additionally, projections from VS1 neuron could play a role in it (Haag and Borst, 2003). To test the dependence on pattern structures, we used a symmetric checkerboard pattern of square size $6^\circ \times 6^\circ$ and contrast 1 (not shown). The resulting reduction in pitch STA (Fig. 5B, gray curve) indicates that the structure of patterns does play a significant role in modulating the response of H1 to vertical motion.

To study the correspondence between the time-lagged mean stimulus and a spike, we fit an exponential to the tail of the yaw STA (Fig. 5A,B, insets). The decay constant of the exponential

Figure 3. Distractor-induced change in spike train entropy. **A–C**, Asymptotic changes in R_{Tot} , R_{Noise} , and R_{Info} with increasing length of time window T are shown for one set of parametric values: repeated yaw at $\sigma_Y = 100^\circ/\text{s}$ or $\xi_{R/P} = 0$ (**A**); repeated yaw with nonrepeated roll at $\xi_R = 10$ (**B**); and repeated yaw with nonrepeated pitch at $\xi_P = 3$ (**C**). Linearly extrapolated values for $T \rightarrow \infty$ are marked by squares on the $1/T = 0$ axes with the respective color codes. The time resolution for spikes was chosen as 1 ms. The difference between the extrapolated R_{Tot} and R_{Noise} gives R_{Info} for $1/T = 0$. Bootstrap errors in entropy estimates were $<5\%$.

Figure 4. The yaw information rate decreases with increasing strength of the distractor. The change in R_{Info} (in bits per second) from spike train to repeated yaw with nonrepeated roll (**A**) and repeated yaw with nonrepeated pitch (**B**; Fig. 1C,D, segment 2). Entire ranges of ξ_R values from 10^0 to 10^2 , and ξ_P values from 10^{-1} to 10^1 are used. The three color-coded curves (brown, blue, and red) stand for R_{Tot} , R_{Noise} , and R_{Info} , respectively. Each point represents extrapolated values for the limiting case $1/T \rightarrow 0$, corresponding to the colored squares in Figure 3. The values of ξ_R and ξ_P highlighted by the gray shaded boxes will be used later as guides for illustrations in the subsection Spike-triggered covariance. The x-axis is shown on a logarithmic scale. The statistical error from bootstrap is $<5\%$.

decreases from 40 to 24 ms as ξ_R increases from 3 to 30, while it decreases from 40 to 28 ms as ξ_P increases from 1 to 10. This suggests that as the distractor variance increases, spikes are more likely to be generated by stimulus features from the immediate past.

Spike-triggered covariance

We compute the joint yaw–distractor spike-triggered covariance around the respective yaw and distractor STAs (de Ruyter van Steveninck and Bialek, 1988). To illustrate the covariance structure, we choose parameter values $\xi_R = 10$ and $\xi_P = 3$, at which R_{Info} for yaw–roll and yaw–pitch stimulus combinations are comparable (Fig. 4, gray shaded boxes).

The structure of the temporal auto-covariance and cross-covariance between stimuli are displayed in the diagonal blocks

Figure 5. Yaw STA change induced by distractors. **A**, Yaw STAs from combined yaw–roll stimulus at four different values of ξ_R . The dotted line shows the roll STA at ξ_R (STA_R) = 10. **B**, Yaw STAs from combined yaw–pitch stimulus at four different values of ξ_P . The dotted line shows the pitch STA (STA_P) with the 2-D random pattern, and the gray line shows the pitch STA with a checkerboard pattern having the same spatial correlation length as the 2-D random pattern. These two illustrative curves are shown for the case $\xi_P = 3$. The insets in **A** and **B** show exponential fits (color coded) to the tail of the yaw STA curves.

(YY, RR, PP) and off-diagonal blocks (YR, YP), respectively, in Figure 6, A and B. The matrix values represent covariance calculated around the mean, or the STA. For example, negative covariance in the YR block indicates that spikes are associated with interactions between an increasing yaw and a decreasing roll around their respective means, or vice versa. The covariance structures indicate that spikes are associated with yaw, distractor, and yaw–distractor correlations. The extension of the structures indicates existence of short- and long-range temporal correlations, with the latter lasting for at least 50 ms.

Using the cutoff set by the radius of the Wigner semicircle, we found one positive and two negative significant eigenvalues for the case $\xi_R = 10$, and one positive and three negative significant eigenvalues for the case $\xi_P = 3$ (Fig. 6C,D, red circles). For the other values of distractor variance, we found at the most one positive and four negative significant eigenvalues. This confirms that, even for a complex stimulus with multiple motion components, the relevant subspace has only a few dimensions. An interesting offshoot of this result, the significance of which will be discussed later in the section Relevant subspace information, is that spikes carry negative information about features defined by $\hat{e}_+$.

The radius of the Wigner semicircle provides a hard cutoff for determining the relevant eigenvalues, unlike the bootstrap bounds, the accuracy of which depends on the number of bootstrap samples. To obtain the radius, we fit the set of irrelevant eigenvalues of ΔC from the central portion of the eigenvalue spectra (Fig. 6E, gray) to the theoretical cumulative Wigner distribution (Eq. 15). For the case $\xi_R = 10$ illustrated in Figure 6, E and F, the radius of the Wigner semicircle is $r = 0.112$. The three eigenvalues (Fig. 6E,F, red markers) lying outside the perimeter of the Wigner semicircle are identified as relevant. With $N = 28,236$ spikes and $M = 100$ dimensions, the analytic estimate of the radius turns out to be $R = 2\sqrt{M/N} = 0.118$ (see Materials and Methods), which matches the numerical estimate within 5%. For all other values of ξ_R and ξ_P , the set of significant eigenvalues determined from the bootstrap analysis matches the corresponding set from the Wigner analysis (data not shown), thus confirming that all of the leading eigenvalues are correctly identified.

Figure 6. Joint yaw–distractor covariance matrices and significant eigenvalues. **A, B**, The heat maps represent temporal covariances for the yaw–roll combinations at $\xi_R = 10$ (**A**) and yaw–pitch combinations at $\xi_P = 3$ (**B**). The covariances are shown for 100 ms before a spike (at $t = 0$ ms), where time is represented along x - and y -axes binned at 2 ms. The diagonal and off-diagonal matrix blocks represent auto-covariances YY , RR , and PP , and cross-covariances YR and YP , respectively (refer to Eq. 12). The color bar shows the range of covariance change from the prior. **C, D**, Eigenvalue spectra of the ΔC matrices shown in **A** and **B**. The black dashed lines represent the upper and lower bounds of eigenvalues based on bootstrap resampling of randomly shifted spike trains. The significant eigenvalues are shown by red circles. **E**, Exemplified for the case $\xi_R = 10$, significant eigenvalues of ΔC (red marker) lie outside the radius (R) of the Wigner semicircle. The grayed out areas show sections of the eigenvalue spectra (solid black circles) used for nonlinear least-squares fit to the theoretical cumulative Wigner distribution (solid green line). The left panel shows the frequency distribution of eigenvalues $\rho(\lambda)$ in black, the Wigner semicircle in green, and the eigenvalue outliers in red. **F**, Spectrum and distribution of eigenvalues of the spike-independent covariance matrix C_{Rnd} obtained from an ensemble of N time windows chosen randomly from the prior stimulus. The radii of Wigner semicircles from the nonlinear least-squares fit for the cases depicted in **E** and **F** match within 1% of each other.

Leading dimensions of yaw–distractor subspace

The eigenvectors corresponding to the leading eigenvalues of ΔC define the stimulus subspace most relevant to H1. To obtain an interpretation of the actual motion features that might trigger spiking, we examine the shapes of the three eigenvectors $\hat{e}_{-1}$, $\hat{e}_{-2}$ and $\hat{e}_{+1}$ corresponding to the two leading negative and the leading positive eigenvalues (Fig. 7A–C). An exception is drawn for the pure yaw stimulus case, for which we did not find any positive eigenvalues.

The eigenvector $\hat{e}_{+1}$ of the yaw–distractor space represents the direction of largest expansion of the stimulus space. For both yaw–roll and yaw–pitch combinations, we obtain one such direction. Since positive yaw velocity excites H1, the sign of each eigenvector is set by imposing the constraint that the yaw part of the curve has a positive rising peak (Fig. 7). The profile of distractor and yaw halves of $\hat{e}_{+1}$ clearly shows that the effect of the distractor is much larger than the effect of yaw. The presence of a heavy tail

Figure 7. Eigenvectors associated with the leading eigenvalues (λ) show increased distractor representation. **A–C**, The leading dimensions for the three cases with pure yaw: no positive λ value (**A**); yaw–roll combinations at $\xi_R = \{3, 10, 30\}$ (**B**); and yaw–pitch combinations at $\xi_P = \{1, 3, 10\}$ (**C**). The color codes red, green, and blue represent eigenvectors $\{\hat{e}_{+1}\}$ and $\{\hat{e}_{-1}, \hat{e}_{-2}\}$, corresponding to the largest positive and the two largest negative eigenvalues, respectively. Each eigenvector has a left yaw half (100 ms) and a right distractor half (100 ms), which are separated by the dashed vertical line. The time of spike occurrence is denoted by $t = 0$ ms.

suggests relatively long-lasting temporal effects. In addition, the yaw and distractor halves are monophasic, with their relative phases opposite to each other. This could point to a competition for generating spikes by preferential (or positive) roll, and downward pitch motion. We will discuss this in further detail in the next subsection.

The temporal structure of the eigenvector $\hat{e}_{-1}$ suggests that it is dominated by yaw when the distractor variance is low, and by distractor when the distractor variance is high. The negative eigenvalue associated with $\hat{e}_{-1}$ indicates a compression of stimulus space along the yaw STA and along a distractor direction that varies with ξ . This can be understood from the change in the temporal profile of the distractor half of $\hat{e}_{-1}$, from a monophasic one to a triphasic one (Fig. 7B,C, blue curves). The triphasic profile demonstrates the presence of positive and negative distractor correlations with different time lags, thus representing the feature “jerk.” Since the eigenvector $\hat{e}_{-2}$ resembles the derivative of the STA, it represents the feature “acceleration.”

To estimate the change in yaw variance induced solely by distractor, eigenvalue decomposition was performed on the YY block (Eq. 12) of the ΔC matrix, which contains only yaw correlations. Table 1 shows that the magnitude of the largest negative eigenvalue of yaw stimulus space decreases rapidly with an increase in distractor variance. Over the full range of ξ_R and ξ_P , the

Table 1. Leading eigenvalues of the yaw subspace

ξ_R	1	3	10	30	50	70	100
λ_Y	−0.59	−0.28	−0.11	−0.08	−0.09	−0.07	−0.08
ξ_P	0.1	0.3	1	3	5	7	10
λ_Y	−0.42	−0.45	−0.37	−0.22	−0.15	−0.11	−0.09

corresponding eigenvalues decrease by 86% and 80%, respectively. Thus, spike coupling to yaw stimulus is markedly weakened by an increase in distractor fluctuations.

Response as a function of relevant stimulus

The input–output relationship (Eq. 17) helps bypass the complexities of sensory processing in the intermediate stages and provides the instantaneous firing rate based on the stimulus distribution. To obtain the input–output mapping, we project the spike-triggered and prior stimuli on each of the yaw and distractor halves of the eigenvectors. The 1-D and 2-D distribution of projections are then used to compute the corresponding response curves (Brenner et al., 2000a).

First, we focus on how the stimulus feature associated with $\hat{e}_{-1}$ impacts the firing rate. The response curve obtained using the yaw half of $\hat{e}_{-1}$ shows a characteristic sigmoidal increase, which eventually plateaus at approximately three times the mean firing rate (Fig. 8A,C, top). This is a typical rectifier-type response indicating the direction selectivity of H1. Unlike negative yaw, negative distractor does not completely suppress the response of H1, as indicated by the firing rate, which remains at the mean firing level (Fig. 8A,C, left). This suggests excitation arising from a relatively weak distractor in an otherwise yaw-dominated direction. For positive distractor projections, the firing rate initially increases but declines when ξ becomes large, with the latter clearly evident at $\xi_R = 30$ (Fig. 8A,C, left, blue curve). Considering that the actual SD of roll in this case is 30 times that of yaw, it is reasonable to expect that such velocities would reduce image contrast and wash out otherwise detectable features, which in turn would diminish the response. An overall assessment of the response curves indicates a broadening of yaw posterior and a narrowing of distractor posterior as ξ_R or ξ_P increases. If we define stimulus sensitivity as the maximum slope of the stimulus–response curve, then this indicates an increase in yaw sensitivity and a decrease in distractor sensitivity.

It is essential to set a reference for deriving the sign of the physical velocity from the sign of the velocity projection. As the STA shape is a good indicator of the direction of the mean velocity that is preferred by the neuron, we shall use it as our reference. It is clear that the spike-triggered roll and yaw distributions peak respectively at negative and positive projections of $\hat{e}_{+1}$ (Fig. 8B). This, however, fails to indicate the direction of the physical roll or yaw velocity that excites or suppresses the neuron. Therefore, we compared the roll posterior distribution obtained from the eigenvector $\hat{e}_{+1R}$ with that obtained from the STA of a pure roll stimulus, at the same roll variance (Fig. 8B, inset). As shown, the two curves are asymmetric with a large peak and a prominent shoulder. Because the sign of the roll STA is positive and the peak of the roll posterior obtained from projecting the spike-triggered roll on the positive designated roll half of $\hat{e}_{+1}$ lies on the negative side of the projection axis, the negative roll projection value for $\hat{e}_{+1}$ corresponds to physical positive roll velocities. The peak indicates strong excitation by positive [counterclockwise (CCW)] roll, and the shoulder indicates weak excitation from negative [clockwise (CW)] roll. The asymmetric U-shaped nonlinearity (Fig. 8B, left) demonstrates the bidirectional response to roll.

For the yaw–pitch case (Fig. 8D), the firing rate mainly increases with an increase in negative pitch projection values. Since the pitch STA is negative, using the same line of argument as above, we find that downward-directed pitch motion corresponds to negative projection values for $\hat{e}_{+1}$. Thus, the feature represented by $\hat{e}_{+1}$, associated with an excitatory response, is largely a reflection of the pitch STA.

A topographical map of the 2-D input–output relationship allows us to examine the firing rate dependence on the interaction of the relevant features. The response profile obtained from projections on $\hat{e}_{+1Y}$, $\hat{e}_{+1R}$ and $\hat{e}_{-1Y}$, $\hat{e}_{-1P}$ display a bean-like shape, whose center is located in the positive half of the yaw projection space, with the two ends extending out into the positive and negative halves of the distractor projection space (Fig. 8A,C). As the distractor variance increases, the two ends gradually fade away, leaving only the center. This indicates the dominance of yaw stimulus along $\hat{e}_{-1}$, the effect of which lingers even at high distractor variance, unlike the distractor response, which is weak and ultimately vanishes. Interestingly for $\hat{e}_{+1}$, the bean extends farther into the two diagonally opposite quadrants (Fig. 8B,D). This implies that a coupling between strong preferred distractor and weak positive or negative yaw, and between weak nonpreferred distractor and strong positive yaw, increases the firing rate. Overall, these results demonstrate the presence of excitatory components in distractor motion and a competitive inhibition of the spiking response to yaw caused by strong preferred distractor motion components.

Stimulus energy shapes input–output response

Direct visual inspection of the 2-D input–output relations of Figure 8 shows that they are not separable as the product of two marginal functions. Moreover, their shape changes with the value of ξ . Here we examine those relations in more detail by drawing various 1-D sections. Sections parallel to the horizontal axis describe how yaw encoding depends on the amount of distractor motion, as measured by its projection on the eigenvector. Conversely, vertical sections describe the encoding of the distractor at different values of yaw projection.

First, we focus on the yaw response when the distractor projection is zero (i.e., $s_{R,P} = 0$). This means that the instantaneous distractor energy is zero, but the global distractor energy is not. The input–output curves exhibit two important features. First, with increasing values of $\xi_{R,P}$ the yaw response curves flatten (Fig. 9A,D, left). Second, to a good approximation the input–output curves all cross at a common point, $s_Y = s_{\text{cross}} = 0.5$, where the slope of the curves is maximal (Fig. 9A,D, right). Based on previous findings that the gain of H1 scales with the SD of the yaw stimulus (Brenner et al., 2000a), one might expect similar scaling to exist for a combined stimulus. To test this hypothesis, we scaled the input–output curves with the “total SD,” but allowing different weights for each stimulus. Analytically, the scale factor is written as $SF = \sqrt{1 + \alpha\xi^2}$, where α may differ for roll and pitch stimuli. We find that the curves for different ξ_P values collapse onto a single function for $\alpha_P = 0.1$ (Fig. 9D, right). However, the curves for different ξ_R values collapse for different values of α_R ranging from 0.03 to 0.3 (Fig. 9A, right). This suggests that a universal scaling exists for pitch but not for roll. Since the fitting parameter effectively weights the contribution of distractor variance to yaw variance, this discrepancy might be a consequence of the inhomogeneity of local motion components of roll, something that is absent in pitch.

We can also look at the yaw input–output curves that are conditional on nonzero values of distractor projection. Repre-

Figure 8. Normalized response as a function of stimulus projection on the leading stimulus dimensions $\hat{e}_{+1}$ and $\hat{e}_{-1}$. **A–D**, Response curves obtained from $\hat{e}_{-1}$ corresponding to $\lambda < 0$ are color coded blue (**A, C**), and those obtained from $\hat{e}_{+1}$ corresponding to $\lambda > 0$ are color coded red (**B, D**). The 2-D input–output (I/O) heat map, marginal prior (dashed black) and posterior (solid black) distributions (pdfs), and response curves (solid red, blue) derived from them are shown for each of these two cases. The prior and posterior pdfs are generated by projecting yaw and distractor modes of each eigenvector on to the respective spike-triggered yaw and distractor stimuli. The projection values are denoted by s_Y , s_R , and s_P for yaw, roll, and pitch. The horizontal and vertical axes of the 2-D heat map have units of yaw and distractor projections. The y-axis of the response curve shows the normalized firing rate, $\frac{r(s)}{\bar{r}}$. The parameter values $\xi_R = \{3, 10, 30\}$ and $\xi_p = \{1, 3, 10\}$ are chosen to highlight the change in nonlinearity with increasing distractor variance. The inset in **B** ($\xi_R = 10$) shows a peak as well as a shoulder (gray shaded regions) in the spike-triggered distributions obtained from pure roll STA at $\sigma_R = 1000^\circ/\text{s}$ and from $\hat{e}_{+1R}$. The error bands (color coded light red and blue) represent the SEM from 100 bootstrap samples.

Figure 9. Response curves for yaw and distractor at different yaw–distractor energies. **A, D**, The plots on the left show yaw input–output curves for pure yaw stimulus (brown), and for combined stimulus at zero distractor projection value (green). For illustration, four different variances $\xi_R = 1, 3, 10, 30$ and $\xi_p = 1, 3, 5, 10$ are chosen for roll and pitch, respectively. The plots on the right show the scaled version of the input–output curves, where the projection axes are scaled by $\sqrt{1 + \alpha_R \xi_R^2}$ and $\sqrt{1 + \alpha_P \xi_P^2}$ around the coordinates $r_N = 1.2$, $s_Y = 0.5$, shown by the yellow circle. Here $s_{\text{cross}} = 0.5$, and SF_R , SF_P represent scale factors for roll and pitch. The fitting parameter is unchanged for pitch ($\alpha_P = 0.1$), but for roll (α_R) it varies from 0.03 to 0.3. **B, E**, Yaw response conditional on three different distractor projection values ($-2, 0, 2$) corresponding to the three horizontal cuts on the 2-D I/O map from $\hat{e}_{-1}$. **C, F**, Distractor response conditional on three different yaw projection values ($-2, 0, 2$) corresponding to the three vertical cuts on the 2-D I/O map from $\hat{e}_{+1}$. The color codes blue, green, and red stand for the three projection values, and brown stands for the response to pure stimulus. The yaw, roll, and pitch projection values are denoted respectively by s_Y , s_R , and s_P . The y-axis represents the normalized firing rate (r_N). The light-colored bands around the solid mean lines represent the SD calculated over a range of 0.2 around the conditional values $-2, 0, 2$.

sentative examples are shown in Figure 9, **B** and **E** (dominant-negative eigenvalue). Here the red, green, and blue curves are horizontal slices through the 2-D input–output maps at projection values of $-2, 0$, and 2 SDs of the distractor. These conditions

measure instantaneous projections on the distractor modes. Compared with the yaw response curves that are conditional on zero distractor projection (Fig. 9A,D), the response curves for nonzero distractor projections become more flattened and shift

along the horizontal axis (Fig. 9B,E). The direction of shift is opposite for the opposite sign of the projections. An earlier result that shows that H1 is weakly sensitive to mean distractor velocity (Fig. 5A,B) provides a clue to this behavior. A rightward or leftward shift in the threshold indicates the presence of large instantaneous distractor energy associated with motion components that primarily compete or act in unison with excitatory yaw components, such that in one case higher than normal yaw excitation is required to elicit a response, and vice versa for the other case. The flattening occurs due to an overall reduction in the sensitivity to yaw in the presence of strong instantaneous distractor energy, thus diminishing the gain to yaw.

Finally, Figure 9, C and F, shows distractor input–output curves conditional on different yaw projections for the eigenvector $\hat{e}_{+1}$, as well as the pure distractor input–output behavior. One interesting feature for the roll input–output curve is that it has an asymmetric “U” shape, which means that spikes are generated preferentially for large negative or large positive projections on the distractor half of $\hat{e}_{+1}$. This behavior is related to a broadening of the posterior distribution compared with the prior, resulting in an increased variance, which is also related to the positive sign of the eigenvalue. The asymmetry in the U shape of the roll response curves suggests unequal excitation from CCW and CW roll. A small signature of the U shape is also found in the pitch response curves, and we shall return to a possible explanation for this behavior in the Discussion. The vertical shift of the input–output curves here may be expected simply from the overall effect of yaw on spiking probability.

Relevant subspace information

We begin with the caveat that eigenvalues provide an incomplete description of the posterior distribution because it fails to capture statistical features beyond the second order. Stimulus-specific information per spike (Eq. 18) is a more appropriate choice because it is independent of assumptions about the statistics of distribution and allows us to probe the existence of a synergistic relationship (Brenner et al., 2000b; Schneidman et al., 2003) between yaw and distractor features. Data limitations make it difficult to sample a high-dimensional stimulus space (Rieke et al., 1997; Dayan and Abbott, 2001), so we limit the analysis to a 2-D subspace. The leading stimulus dimensions corresponding to $\hat{e}_{-1}$ and $\hat{e}_{+1}$ are used for the analysis.

Figure 10A shows that spike information about yaw decreases asymptotically from 0.53 bits to zero as ξ_R increases, while spike information about roll and pitch increases from zero to 0.21 and 0.18 bits, respectively (Fig. 10B). Thus, maximum roll and pitch information are respectively 40% and 35% of the maximum yaw information. Since the curves do not plateau, it is conceivable that saturation occurs at yet larger distractor variance. Overall, this characterizes a transition from a yaw-encoding regime to a distractor-encoding regime. Although yaw dominates $\hat{e}_{-1}$, non-zero distractor information indicates that when distractor variance is high, distractor motion overrides yaw motion in driving spike generation. In contrast, spikes carry negative information, up to -0.22 bits about roll and up to -0.14 bits about pitch, corresponding to the $\hat{e}_{+1}$ dimension (Fig. 10C,D). Note that these measures correspond to the difference between posterior and prior entropies, also termed stimulus-specific information (see DeWeese and Meister, 1999; see Materials and Methods). Recalling the shape of nonlinearity (Fig. 8B,D, left), the negative sign results from higher entropy of a widened posterior distribution. We found no evidence of negative yaw information carried by spikes. The mutual information (Eq. 21) is necessarily positive,

Figure 10. Spike information from the leading dimensions of the yaw–distractor subspace. The two leading eigenvectors of ΔC matrix, $\hat{e}_{+1}$ and $\hat{e}_{-1}$, are split into respective yaw and distractor halves. Each of the halves is used to construct the marginal and joint probability distributions for estimating information (Eq. 18). **A–D**, Information (in bits per spike) as a function of $\xi_{R/P}$ corresponding to the eigenvectors $\hat{e}_{+1}$ and $\hat{e}_{-1}$ is shown for the yaw–roll combined stimulus (**A**, **C**) and the yaw–pitch combined stimulus (**B**, **D**). The three curves distinguished by the three marker types correspond to the 1-D yaw subspace (filled triangle), the 1-D distractor subspace (filled square), and the 2-D yaw–distractor subspace (empty circle). Missing values in **C** and **D** are due to the absence of $\lambda > 0$ at the corresponding values of ξ_R and ξ_P . Error bars represent the SEM (100 bootstrap samples).

and is respectively 0.034 and 0.017 bits for $\hat{e}_{-1Y}$ and $\hat{e}_{-1R}$. We also found that the information estimate from the 2-D stimulus subspace spanned by $\hat{e}_{-1Y}$ and $\hat{e}_{-1R}$, or $\hat{e}_{-1Y}$ and $\hat{e}_{-1P}$, was equal to the sum of information estimates from the individual subspaces, within statistical error. This rules out the synergistic effect between yaw and distractor features.

A comparison of information estimates

Stimulus information is conveyed both by individual spikes and by spike patterns (de Ruyter van Steveninck and Bialek, 1988; Schneidman et al., 2011). The excess information that spike patterns carry over single spikes (Brenner et al., 2000b) can be obtained by comparing the estimates from Equations 3 and 4 (see Materials and Methods). For example, the spike train conveys 1.8 bits per spike, while single spikes convey 1.25 bits about the repeated yaw stimulus at $\xi_R = 1$. Thus, single spikes account for 70% of the spike train information at $\xi_R = 1$, which eventually decreases to 47% at large roll variance (Fig. 11A). For pitch, the decrease is from 79% to 45% (Fig. 11B). A decrease in spike-timing precision to repeated yaw, caused by a distractor competing with yaw for available spikes, may play a dominant role here. The overall dependence of information rate on ξ is approximately logarithmic (Fig. 11A,B).

We can assess the completeness of the feature description by comparing the information associated with features (Eq. 18), with single-spike information (Eq. 4). However, such a comparison requires caution because specific distractor features contribute negative symbol information, unlike the estimate from Equation 4, which is always positive. Therefore, for comparison, we choose only those features that contribute positive information. The leading yaw dimension ($\hat{e}_{-1Y}$) of the yaw–roll and yaw–pitch subspaces contributes a maximum of 42% (0.52 bits) and 36% (0.45 bits), respectively, of the single-spike yaw information (Fig. 11A,B). The joint yaw space constructed from the leading yaw dimensions $\hat{e}_{-1Y}$ and $\hat{e}_{-2Y}$ contributes up to 53% of single-spike yaw information. The information we obtained from other relevant dimensions, such as $\hat{e}_{-2}$, $\hat{e}_{-3}$, were lower and decreased with decreasing significance of the eigenvalues.

Figure 11. Spike train information compared with stimulus subspace information, at different distractor strengths. **A, B**, Spike train information ($R_{\text{Info}, \text{yaw}}$) color coded red and single-spike information (I_{Spk} ; Eq. 4) color coded brown shown for the combined stimulus of repeated yaw and nonrepeated distractor. Information per spike (Eq. 18) from the yaw and roll modes of $\hat{e}_{-1}$ are shown color coded green and blue respectively. **C, D**, Spike train information ($R_{\text{Info}, \text{roll}}$, $R_{\text{Info}, \text{pitch}}$) color coded red and single-spike information (I_{Spk} ; Eq. 4) color coded brown are shown for the combined stimulus of repeated roll or pitch with nonrepeated yaw. The x-axis shows the range of ξ_R and ξ_P on a logarithmic scale.

Since distractor features contribute to spike generation (Fig. 10), it is conceivable that spike patterns carry significant distractor information. We test this by using a presentation of repeated distractor with nonrepeated yaw across trials (Fig. 1C,D, segment 6). Figure 11C shows that spike train information rises to 1 bit/spike, while single-spike information rises to 0.8 bits for roll. For pitch, the spike train information rises to 0.63 bits/spike, and single-spike information rises to 0.60 bits/spike (Fig. 11D). As a comparison, the joint feature space of $\hat{e}_{-1R}$ and $\hat{e}_{-2R}$ contributes 0.35 bits, and the joint feature space of $\hat{e}_{-1P}$ and $\hat{e}_{-2P}$ contributes 0.37 bits (Fig. 11A,B). Even if we add the information associated with each of the relevant dimensions, the sum is still less than what we obtain for single spikes from the spike train (data not shown). This indicates that the spike train carries information about stimulus features beyond those identified from the relevant subspace, and it is possible that those features are associated with particular spike combinations.

Discussion

In this article, we provide a quantitative analysis of preferred motion coding from stimuli not aligned with the preferred direction of the neuron. The presence of distractor reduces the efficiency of coding the yaw of H1, but increases the efficiency of coding the distractor or, specifically, the features of the distractor that excite H1. Geometry of the spatial pattern, disparity in the strength of local excitation and suppression, and blurring are some of the factors that may impact the efficiency of coding preferred motion. Our results provide insight into how the response is shaped, not only by the stimuli statistics, but also by the complexity of the optic flow. The method used here can be generalized to studies in other systems, including, but not limited to, the coding of complex, multidimensional stimulus.

It has been shown that H1 can operate with a reliability close to the statistical limits set by noise in the visual input (de Ruyter van Steveninck and Bialek, 1995). Other, internal, noise sources (Schneidman et al., 1998; Manwani and Koch, 1999; White et al.,

2000) may also limit the information communicated about the stimulus. How large is the distractor noise in comparison with noise from these other sources? We found that distractor motion acts as the primary source of noise for yaw (“signal”) encoding, dominating noise from other sources (Fig. 3). If this were not the case, then we would have observed a minimal decrease in spike train information with an increase in distractor variance. The most direct effect of this distractor noise is a decrease in information rate, which is caused by a reduction in firing precision to repeated yaw (Bialek et al., 1991) at moderate distractor variances, and a reduction in both firing precision and firing rate at very large distractor variances (Optican and Richmond, 1987; Bialek et al., 1991; Victor and Purpura, 1996; Warland et al., 1997).

There are two possible explanations for the above findings. First, at low-to-moderate distractor variance, local variation of yaw components is induced by the distractor due to a combination of the aperture effect (Marr and Ullman, 1981) and varying local flow directions, the latter in the case of roll. The resulting ambiguity between components of pure yaw and those induced by distractor cannot be resolved by a local measurement. Rather, the percept of a wide-field motion is built out of many such local ambiguous measurements. This effectively induces noise, which becomes stronger as distractor variance increases. Second, at very high distractor variance, rapid fluctuations from the distractor cause blurring, leading to diminished image contrast. Since the incoming visual signal is low-pass filtered by photoreceptors (Howard et al., 1987), only low spatial frequencies survive at those high-velocity amplitudes. With a typical spatial correlation length of 6° (Fig. 1A) and a pitch SD of $1000^\circ/\text{s}$, the correlation time is $6/1000 = 6$ ms, which is of the same order as the photoreceptor integration time (Laughlin and Weckstrom, 1993). The decrease in R_{Tot} may be related to these effects.

Previous work suggests that optic flow from natural flight (Kern et al., 2001; Boeddeker et al., 2005; van Hateren et al., 2005; Karameier et al., 2006) and higher-order motion stimulus (Quenzer and Zanker, 1991; Lee and Nordström, 2012) contain motion components that can evoke neural response. A strong enough motion along a nonoptimal axis can also induce neural excitation (Karameier et al., 2003). Our results provide evidence that the structure of the optic flow field shaped by the combined motion plays a crucial role in preferred motion encoding. Since roll generates local motion, with yaw and pitch components speeding up as the distance increases from the center of rotation, it induces both excitation and suppression in H1. But a net positive roll STA (Fig. 5A) indicates that excitation is stronger in the ventral half than in the dorsal half, corroborating earlier observations (Eckert, 1980; Hausen, 1981). This means that effective excitation from counterclockwise roll (ventral excitation with dorsal inhibition) overrides effective inhibition from clockwise roll (dorsal excitation with ventral inhibition). Excitation and inhibition could also have different levels of saturation, depending on the stimulus strength (Borst and Egelhaaf, 1990), such that with a strong enough clockwise roll, inhibition in the ventral half cannot overcome excitation in the dorsal half. This may qualitatively describe the asymmetric U shape of the roll response curves (Fig. 9C).

The geometry and size of the spatial pattern have also been reported to impact motion response, both in insects (Borst et al., 1993; Meyer et al., 2011; O’Carroll et al., 2011; Saleem et al., 2012) and in vertebrates (Rodman and Albright, 1989). The response of H1 to pitch motion in our case can perhaps be attributed to two factors. First, the apparent horizontal motion induced by vertical

movement of slanted pattern edges (i.e., the aperture effect; Marr and Ullman, 1981) can induce excitation of H1, and an excitation/inhibition asymmetry, as noted above, may result in residual net effects. Second, excitatory input from ipsilateral VS1 to H1 also can increase the response sensitivity to downward pitch (Haag and Borst, 2003).

What motion features does H1 encode? The eigenvector profiles demonstrate that H1 preferentially encodes stimulus features such as velocity, acceleration, and jerk, among others (Fig. 7), which was also reported in the study by Brenner et al. (2000a). These types of features are also coded by motion-sensitive neurons in pigeons (Cao et al., 2004) and parieto-insular vestibular cortex neurons in macaques (Chen et al., 2010). Our results further establish that these features carry a large portion of stimulus information. For example, the 2-D subspace spanned by $\dot{e}_{-1Y}$ and $\dot{e}_{-2Y}$ contributes 40% of the spike train yaw information. This raises the question of whether natural stimuli are better alternatives to Gaussian stimuli. Studies on fly motion-sensitive neurons under natural conditions (Lewen et al., 2001) and using natural stimulus (Karmeier et al., 2006) suggest that neurons encode features beyond velocity. Further, natural stimuli have been linked to a higher coding efficiency in auditory systems (Rieke et al., 1995; Escabí et al., 2003) and visual systems (Vinje and Gallant, 2002; Felsen et al., 2005), indicating the presence of a richer repertoire of features in such stimuli (Dong and Atick, 1995). However, the statistical complexity of natural stimuli and the strong response nonlinearities make the analyses nontrivial (Egelhaaf and Borst, 1989; van Hateren, 1997; O'Carroll et al., 2011). Nonparametric methods such as Maximally Informative Dimensions (Sharpee et al., 2004) and generalizations of it (Rajan and Bialek, 2013) may help identify which features are relevant and how informative those features are.

It is well known that the visual system adapts to stimulus properties, such as contrast and stimulus variance (Maddess and Laughlin, 1985; Smirnakis et al., 1997; Brenner et al., 2000a; Harris et al., 2000; Fairhall et al., 2001; Baccus and Meister, 2002). This has been linked to efficient signal coding (Barlow, 1961) and an increase in neural information throughput (Laughlin, 1981; Brenner et al., 2000a; Safran et al., 2007). We find that H1 adapts its response to yaw in a way that depends on the global distractor variance even when the instantaneous distractor energy is zero. The collapse of scaled yaw response curves onto a single curve (Fig. 9A,D) may indicate a universal strategy for scaling. In contrast, a change in the fitting parameter for roll (α_R) suggests that scaling depends crucially on the structure of the optic flow, which determines the effective wide-field excitation, and not on the stimulus variance alone. Overall, these empirical findings provide insight into how a system might adapt to the combined dynamics of different stimulus components. Whether, and to what extent, this can be understood in a framework of optimal coding (Laughlin, 1981) remains an open question. Since different directions of motion are relevant here, a complete answer would require an analysis of coding across multiple neurons, which would necessitate experiments that simultaneously record from motion-sensitive cells with different directional selectivities (see also van Hateren, 1990).

In summary, we have characterized the features encoded by H1 from a stimulus with competing motion components and have shown how such features shape the neural response. Although scaling of the output to the collective dynamic range of the input might be viewed as a strategy for optimizing information transmission about the preferred motion buried in a complex stimulus, further studies are needed, with multiple neurons

with different directional selectivity, to validate this argument. The Gaussian white noise stimulus imposes an obvious limitation in our study, because it lacks the statistical variability and functional relevance of a natural stimulus. But its merit lies in the mathematical tractability (Marmarelis and Marmarelis, 1978), which allows complete characterization of the stimulus feature space. A natural next step is to incorporate stimuli with higher-order correlations and, ultimately, natural stimuli into this information-theoretic framework. Modeling (Prenger et al., 2004; Butts et al., 2011) and nonparametric approaches (Sharpee et al., 2004) will prove particularly useful in such studies for characterizing the physiologically relevant features and their neural representation.

References

- Baccus SA, Meister M (2002) Fast and slow contrast adaptation in retinal circuitry. *Neuron* 36:909–919. [CrossRef Medline](#)
- Barlow HB (1961) Possible principles underlying the transformations of sensory messages, pp 217–234. Cambridge, MA: MIT.
- Bevington P, Robinson D (2003) Data reduction and error analysis for the physical sciences. New York: McGraw-Hill.
- Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252:1854–1857. [CrossRef Medline](#)
- Boeddeker N, Lindemann JP, Egelhaaf M, Zeil J (2005) Responses of blowfly motion-sensitive neurons to reconstructed optic flow along outdoor flight paths. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 191:1143–1155. [CrossRef Medline](#)
- Borst A, Egelhaaf M (1990) Direction selectivity of blowfly motion-sensitive neurons is computed in a two-stage process. *Proc Natl Acad Sci U S A* 87:9363–9367. [CrossRef Medline](#)
- Borst A, Haag J (2002) Neural networks in the cockpit of the fly. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 188:419–437. [CrossRef Medline](#)
- Borst A, Egelhaaf M, Seung HS (1993) Two-dimensional motion perception in flies. *Neural Comput* 5:856–868. [CrossRef](#)
- Brenner N, Bialek W, de Ruyter van Steveninck R (2000a) Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702. [CrossRef Medline](#)
- Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000b) Synergy in a neural code. *Neural Comput* 12:1531–1552. [CrossRef Medline](#)
- Budick SA, Reiser MB, Dickinson MH (2007) The role of visual and mechanosensory cues in structuring forward flight in *Drosophila melanogaster*. *J Exp Biol* 210:4092–4103. [CrossRef Medline](#)
- Butts DA, Weng C, Jin J, Alonso JM, Paninski L (2011) Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. *J Neurosci* 31:11313–11327. [CrossRef Medline](#)
- Cao P, Gu Y, Wang SR (2004) Visual neurons in the pigeon brain encode the acceleration of stimulus motion. *J Neurosci* 24:7690–7698. [CrossRef Medline](#)
- Chen A, DeAngelis GC, Angelaki DE (2010) Macaque parieto-insular vestibular cortex: responses to self-motion and optic flow. *J Neurosci* 30:3022–3042. [CrossRef Medline](#)
- Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network* 12:199–213. [CrossRef Medline](#)
- Collett T, Land M (1975) Visual spatial memory in a hoverfly. *J Comp Physiol* 100:59–84. [CrossRef](#)
- Dayan P, Abbott L (2001) Theoretical neuroscience. Cambridge, MA: MIT.
- de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc R Soc Lond B Biol Sci* 234:379–414. [CrossRef](#)
- de Ruyter van Steveninck R, Bialek W (1995) Reliability and statistical efficiency of a blowfly movement-sensitive neuron. *Philos Trans R Soc Lond B Biol Sci* 348:321–340. [CrossRef](#)
- de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. *Science* 275:1805–1808. [CrossRef Medline](#)
- de Boer R, Kuypers P (1968) Triggered correlation. *IEEE Trans Biomed Eng* 15:169–179. [Medline](#)

- DeWeese MR, Meister M (1999) How to measure the information gained from one symbol. *Network* 10:325–340. [CrossRef Medline](#)
- Dong DW, Atick JJ (1995) Statistics of natural time-varying images. *Network* 6:345–358. [CrossRef](#)
- Eckert H (1980) Functional properties of the H1-neurone in the third optic ganglion of the blowfly, *Phaenicia*. *J Comp Physiol* 135:29–39. [CrossRef](#)
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. New York: Chapman and Hall.
- Egelhaaf M, Borst A (1989) Transient and steady-state response properties of movement detectors. *J Opt Soc Am A* 6:116–127. [CrossRef Medline](#)
- Escabi MA, Miller LM, Read HL, Schreiner CE (2003) Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J Neurosci* 23:11489–11504. [Medline](#)
- Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412:787–792. [CrossRef Medline](#)
- Felsen G, Touryan J, Han F, Dan Y (2005) Cortical sensitivity to visual features in natural scenes. *PLoS Biol* 3:e342. [CrossRef Medline](#)
- Fry SN, Sayaman R, Dickinson MH (2003) The aerodynamics of free-flight maneuvers in drosophila. *Science* 300:495–498. [CrossRef Medline](#)
- Haag J, Borst A (2003) Orientation tuning of motion-sensitive neurons shaped by vertical-horizontal network interactions. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 189:363–370. [Medline](#)
- Harris RA, O'Carroll DC, Laughlin SB (2000) Contrast gain reduction in fly motion adaptation. *Neuron* 28:595–606. [CrossRef Medline](#)
- Hassenstein B, Reichardt W (1956) Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenbewertung bei der bewegungsperzeption des rüsselkäfers *Chlorophanus*. *Z Naturforsch B* 11:513–524.
- Hateren JH, Schilstra C (1999) Blowfly flight and optic flow. II. Head movements during flight. *J Exp Biol* 202:1491–1500. [Medline](#)
- Hausen K (1981) Monocular and binocular computation of motion in the lobula plate of the fly. *Verh Dtsch Zool Ges* 74:49–70.
- Hausen K (1982a) Motion sensitive interneurons in the optomotor system of the fly. *Biol Cybern* 45:143–156. [CrossRef](#)
- Hausen K (1982b) Motion sensitive interneurons in the optomotor system of the fly. *Biol Cybern* 46:67–79. [CrossRef](#)
- Hengstenberg R, Hausen K, Hengstenberg B (1982) The number and structure of giant vertical cells (VS) in the lobula plate of the blowfly *Calliphora erythrocephala*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 149:163–177. [CrossRef](#)
- Howard J, Blakeslee B, Laughlin SB (1987) The intracellular pupil mechanism and photoreceptor signal:noise ratios in the fly *Lucilia cuprina*. *Proc R Soc Lond B Biol Sci* 231:415–435. [CrossRef Medline](#)
- Karmeier K, Krapp HG, Egelhaaf M (2003) Robustness of the tuning of fly visual interneurons to rotatory optic flow. *J Neurophysiol* 90:1626–1634. [CrossRef Medline](#)
- Karmeier K, van Hateren JH, Kern R, Egelhaaf M (2006) Encoding of naturalistic optic flow by a population of blowfly motion-sensitive neurons. *J Neurophysiol* 96:1602–1614. [CrossRef Medline](#)
- Kern R, Peterreit C, Egelhaaf M (2001) Neural processing of naturalistic optic flow. *J Neurosci* 21:RC139. [Medline](#)
- Kern R, van Hateren JH, Michaelis C, Lindemann JP, Egelhaaf M (2005) Function of a fly motion-sensitive neuron matches eye movements during free flight. *PLoS Biol* 3:e171. [CrossRef Medline](#)
- Kern R, van Hateren JH, Egelhaaf M (2006) Representation of behaviourally relevant information by blowfly motion-sensitive visual interneurons requires precise compensatory head movements. *J Exp Biol* 209:1251–1260. [CrossRef Medline](#)
- Krapp HG, Hengstenberg R (1997) A fast stimulus procedure to determine local receptive field properties of motion-sensitive visual interneurons. *Vision Res* 37:225–234. [CrossRef Medline](#)
- Krapp HG, Hengstenberg R, Egelhaaf M (2001) Binocular contributions to optic flow processing in the fly visual system. *J Neurophysiol* 85:724–734. [Medline](#)
- Land MF (1997) Visual acuity in insects. *Annu Rev Entomol* 42:147–177. [CrossRef Medline](#)
- Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. *Z Naturforsch C* 36:910–912. [Medline](#)
- Laughlin S, Weckström M (1993) Fast and slow photoreceptors: a comparative study of the functional diversity of coding and conductances in the Diptera. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 172:593–609. [CrossRef](#)
- Lee YJ, Nordström K (2012) Higher-order motion sensitivity in fly visual circuits. *Proc Natl Acad Sci U S A* 109:8758–8763. [CrossRef Medline](#)
- Lewen GD, Bialek W, de Ruyter van Steveninck RR (2001) Neural coding of naturalistic motion stimuli. *Network* 12:317–329. [CrossRef Medline](#)
- Maddess T, Laughlin SB (1985) Adaptation of the motion-sensitive neuron h1 is generated locally and governed by contrast frequency. *Proc R Soc Lond B Biol Sci* 225:251–275. [CrossRef](#)
- Manwani A, Koch C (1999) Detecting and estimating signals in noisy cable structures. I: neuronal noise sources. *Neural Comput* 11:1797–1829. [CrossRef Medline](#)
- Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems: the white-noise approach. New York: Plenum.
- Marr D, Ullman S (1981) Directional selectivity and its use in early visual processing. *Proc R Soc Lond B Biol Sci* 211:151–180. [CrossRef Medline](#)
- Meyer HG, Lindemann JP, Egelhaaf M (2011) Pattern-dependent response modulations in motion-sensitive visual interneurons—a model study. *PLoS One* 6:e21488. [CrossRef Medline](#)
- O'Carroll DC, Barnett PD, Nordström K (2011) Local and global responses of insect motion detectors to the spatial structure of natural scenes. *J Vis* 11(14):20. [CrossRef](#)
- Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57:162–178. [Medline](#)
- Paninski L (2003) Convergence properties of three spike-triggered analysis techniques. *Network* 14:437–464. [CrossRef Medline](#)
- Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98:1064–1072. [CrossRef Medline](#)
- Prenger R, Wu MC, David SV, Gallant JL (2004) Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Netw* 17:663–679. [CrossRef Medline](#)
- Quenzer T, Zanker J (1991) Visual detection of paradoxical motion in flies. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 169:331–340.
- Rajan K, Bialek W (2013) Maximally informative “stimulus energies” in the analysis of neural responses to natural signals. *PLoS One* 8:e71959. [CrossRef Medline](#)
- Rieke F, Bodnar D, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc R Soc Lond B Biol Sci* 262:259–265. [CrossRef](#)
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) Spikes: exploring the neural code. Cambridge, MA: MIT.
- Rodman HR, Albright TD (1989) Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp Brain Res* 75:53–64. [Medline](#)
- Safran MN, Flanagan VL, Borst A, Sompolinsky H (2007) Adaptation and information transmission in fly motion detection. *J Neurophysiol* 98:3309–3320. [CrossRef Medline](#)
- Saleem AB, Longden KD, Schwyn DA, Krapp HG, Schultz SR (2012) Bimodal optomotor response to plaids in blowflies: mechanisms of component selectivity and evidence for pattern selectivity. *J Neurosci* 32:1634–1642. [CrossRef Medline](#)
- Schilstra C, Hateren JH (1999) Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics. *J Exp Biol* 202:1481–1490. [Medline](#)
- Schneidman E, Freedman B, Segev I (1998) Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Comput* 10:1679–1703. [CrossRef Medline](#)
- Schneidman E, Bialek W, Berry MJ 2nd (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23:11539–11553. [Medline](#)
- Schneidman E, Puchalla JL, Segev R, Harris RA, Bialek W, Berry MJ 2nd (2011) Synergy from silence in a combinatorial neural code. *J Neurosci* 31:15732–15741. [CrossRef Medline](#)
- Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. *J Vis* 6(4):13 484–507. [CrossRef Medline](#)
- Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16:223–250. [CrossRef Medline](#)
- Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386:69–73. [CrossRef Medline](#)
- Srinivasan MV, Lehrer M, Kirchner WH, Zhang SW (1991) Range perception through apparent image speed in freely flying honeybees. *Vis Neurosci* 6:519–535. [CrossRef Medline](#)

- Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200. [CrossRef](#)
- Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Comput* 7:399–407. [CrossRef](#)
- van Hateren JH (1990) Directional tuning curves, elementary movement detectors, and the estimation of the direction of visual movement. *Vision Res* 30:603–614. [CrossRef](#) [Medline](#)
- van Hateren JH (1997) Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res* 37:3407–3416. [CrossRef](#) [Medline](#)
- van Hateren JH, Kern R, Schwerdtfeger G, Egelhaaf M (2005) Function and coding in the blowfly H1 neuron during naturalistic optic flow. *J Neurosci* 25:4343–4352. [CrossRef](#) [Medline](#)
- van Santen JP, Sperling G (1984) Temporal covariance model of human motion perception. *J Opt Soc Am A* 1:451–473. [CrossRef](#) [Medline](#)
- Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 76:1310–1326. [Medline](#)
- Vinje WE, Gallant JL (2002) Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J Neurosci* 22:2904–2915. [Medline](#)
- Wagner H (1986) Flight performance and visual control of flight of the free-flying housefly (*Musca domestica* L.) II. Pursuit of targets. *Philos Trans R Soc Lond B Biol Sci* 312:553–579. [CrossRef](#)
- Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol* 78:2336–2350. [Medline](#)
- White JA, Rubinstein JT, Kay AR (2000) Channel noise in neurons. *Trends Neurosci* 23:131–137. [CrossRef](#) [Medline](#)
- Wigner EP (1958) On the distribution of the roots of certain symmetric matrices. *Ann Math* 67:325–327. [CrossRef](#)
- Zeil J, Boeddeker N, Hemmi JM (2008) Vision and the organization of behaviour. *Curr Biol* 18:R320–R323. [CrossRef](#) [Medline](#)