Dynamic Signal Compression for Robust Motion Vision in Flies

Highlights

- *Drosophila* motion processing robustly estimates the velocity of moving natural scenes

- Visual interneurons in the fly visual system dynamically adapt to stimulus contrast

- This adaptation relies on fast spatial integration of neural feedback

- Contrast adaptation accounts for robust motion vision in computational circuit models

Authors

Michael S. Drews, Aljoscha Leonhardt, Nadezhda Pirogova, ..., Lukas Braun, Etienne Serbe, Alexander Borst

Correspondence

drews@neuro.mpg.de (M.S.D.), leonhardt@neuro.mpg.de (A.L.)

In Brief

Flies reliably estimate the velocity of moving natural scenes regardless of image statistics. Current models of *Drosophila* motion vision fail to explain this robustness. Drews, Leonhardt, et al. show that flies achieve this performance by rapidly adjusting the sensitivity of visual interneurons in the medulla to surround contrast.

CellPress

CellPress

Dynamic Signal Compression
for Robust Motion Vision in Flies

Michael S. Drews,[1,2,4,*] Aljoscha Leonhardt,[1,4,5,*] Nadezhda Pirogova,[1,2] Florian G. Richter,[1,2] Anna Schuetzenberger,[1,2] Lukas Braun,[3] Etienne Serbe,[1] and Alexander Borst[1]

[1]Department Circuits-Computation-Models, Max-Planck-Institute of Neurobiology, 82152 Martinsried, Germany
[2]Graduate School of Systemic Neurosciences, LMU Munich, 82152 Martinsried, Germany
[3]Bernstein Center for Computational Neuroscience, 10115 Berlin, Germany
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: drews@neuro.mpg.de (M.S.D.), leonhardt@neuro.mpg.de (A.L.)
https://doi.org/10.1016/j.cub.2019.10.035

SUMMARY

Sensory systems need to reliably extract information from highly variable natural signals. Flies, for instance, use optic flow to guide their course and are remarkably adept at estimating image velocity regardless of image statistics. Current circuit models, however, cannot account for this robustness. Here, we demonstrate that the *Drosophila* visual system reduces input variability by rapidly adjusting its sensitivity to local contrast conditions. We exhaustively map functional properties of neurons in the motion detection circuit and find that local responses are compressed by surround contrast. The compressive signal is fast, integrates spatially, and derives from neural feedback. Training convolutional neural networks on estimating the velocity of natural stimuli shows that this dynamic signal compression can close the performance gap between model and organism. Overall, our work represents a comprehensive mechanistic account of how neural systems attain the robustness to carry out survival-critical tasks in challenging real-world environments.

INTRODUCTION

Visual motion represents a critical source of sensory feedback for navigation. Self-motion results in particular patterns of local directional cues across the retina. Detection of these optic flow fields allows animals to estimate and control their current heading [1]. Flies, for instance, react to whole-field retinal motion by turning in the same direction as their surroundings. This optomotor response enables them to maintain a straight path under perturbations as well as over long distances [2, 3].

For the reflex to work effectively, biological motion detectors need to respond reliably and independently of the particular visual statistics of the environment. This poses a challenge given the complexity of natural scenes [4, 5]. Motion vision systems therefore need to employ processing strategies that maintain robust performance despite the variability of natural visual input.

Recent circuit mapping efforts have yielded unprecedented insight into the neural substrate of motion detection in *Drosophila* [6, 7]. The fly optic lobe consists of sequential neuropils (retina, lamina, medulla, lobula, and lobula plate) and is arranged in columns that process visual input retinotopically. In various combinations, lamina cells L1–L5 feed into a light-sensitive ON or a dark-sensitive OFF pathway, each comprising at least four cell types in the medulla [8]. Medulla units fall into two classes characterized either by transient temporal filtering and moderate center-surround antagonism in their spatial receptive field (Mi1 and Tm3 for ON; Tm1, Tm2, and Tm4 for OFF) or by tonic responses and strong antagonistic surround (Mi4 and Mi9 for ON; Tm9 for OFF) [9–13]. Postsynaptic T4 and T5 cells then compute local ON and OFF motion, respectively, by comparing medulla signals with different dynamics across neighboring columns [8, 14–19]. Jointly, they are necessary for the optomotor response [20]. By pooling appropriate T4 and T5 signals, lobula plate tangential cells (LPTCs) detect optic flow fields that correspond to rotations around different body axes and ultimately control turning [3, 21–23].

For artificial stimuli, fly motion processing is well explained by correlation-based detector models that rely on multiplication of spatially adjacent, asymmetrically filtered luminance signals [24]. These elementary motion detectors (EMDs) account for subtle features of behavioral and neural responses such as pattern-induced shifts in velocity tuning [25, 26], intrinsic velocity gain control [27], or reverse-phi sensitivity [28, 29]. However, EMD output strongly depends on contrast as defined by the average difference between light and dark [26]. EMDs thus invariably confound image contrast with velocity. Since local contrast varies substantially within natural images [4], output from individual EMDs is sparse and fluctuates heavily under naturalistic conditions (Figures S1A–S1C). Motion responses in flies, however, have been shown to be highly robust, across both time and different natural scenes [30, 31].

Various general mechanisms for adaptation to naturalistic signals have been described in the fly visual system. These include gain control in photoreceptors or LPTCs [32–34], redundancy reduction through lateral inhibition [35], subtractive enhancement of flow field selectivity [36], and tailoring of processing to fundamental natural scene statistics [31, 37, 38]. However, none effectively address the problem of contrast fluctuations.
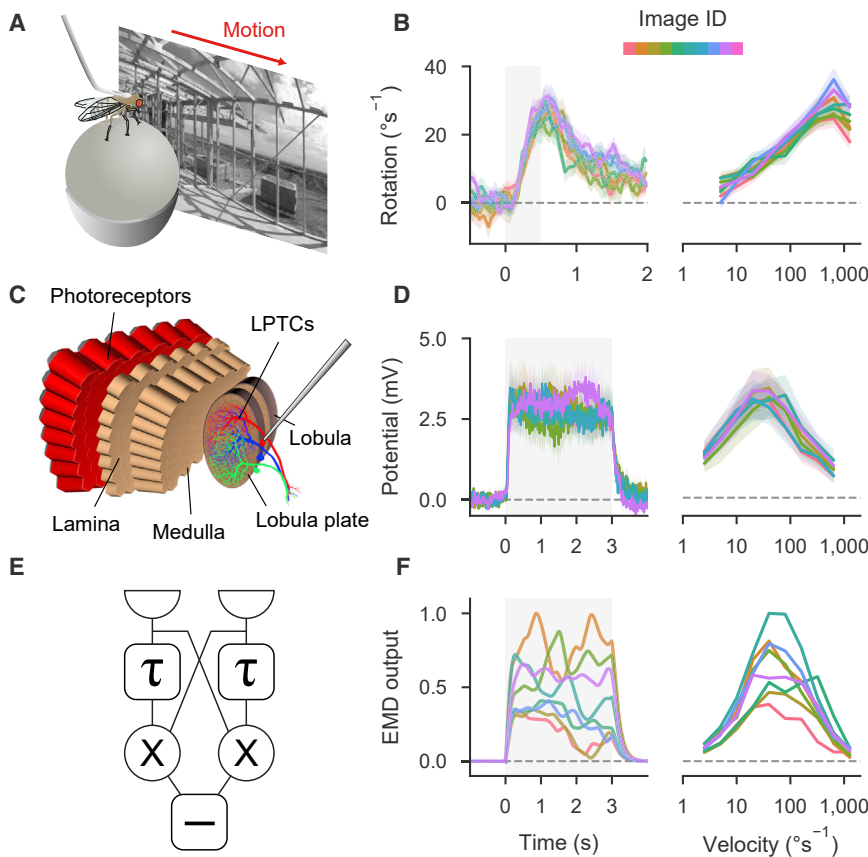
Figure 1. Flies Respond More Robustly to Natural Scene Variability Than Predicted by Correlation-Based Motion Detectors

(A) Illustration of behavioral set-up. Tethered wild-type *Drosophila* were stimulated with translating natural images.

(B) Left: turning responses for images moving at $80°s^{-1}$ (n = 16 flies). Each color indicates a distinct scene. Images moved during gray-shaded period. Right: velocity tuning curves for all measured scenes (averaged between 0 and 1 s after motion onset).

(C) Illustration of fly visual system. Photoreceptor signals are processed in five retinotopically arranged neuropils. Wide-field lobula plate tangential cells (LPTCs) respond to particular optic flow fields.

(D) Left: membrane potential of horizontal system LPTCs in response to images moving at $20°s^{-1}$ (n = 11 cells from 9 flies). Right: velocity tuning curves (averaged between 0 and 3 s after motion onset).

(E) Schematic of an individual correlation-based elementary motion detector (EMD; τ denotes delay line; ×, multiplication; –, subtraction).

(F) Left: responses of an array of EMDs to stimulation with natural images moving at $20°s^{-1}$. Right: velocity tuning curves of EMD array (evaluated like LPTC output). Note that in contrast to experiments, model responses were averaged across many different starting phases. Shaded areas around curves indicate bootstrapped 68% confidence intervals.

See also Figure S1 and Table S2.

In vertebrate visual systems, contrast sensitivity is continuously regulated through the mechanism of divisive normalization [39–41]. Here, the response of a neuron is effectively divided by local contrast, estimated as the average activity within a population of neighboring neurons. The process compresses signals of varying contrast into a fixed range by dynamically adjusting gain to current conditions [5] and renders the neural representation of stimuli largely invariant with respect to contrast. However, so far, no comparable mechanism has been described for the invertebrate visual system.

Here, we investigate how the fly visual system copes with contrast variability and demonstrate that dynamic signal compression based on divisive contrast normalization renders motion processing robust to the challenges imposed by natural visual environments.

RESULTS

Fly Motion Responses Are Robust to Natural Scene Variability

To rigorously assess the robustness of *Drosophila* motion processing, we measured optomotor responses to a diverse set of moving naturalistic panoramas on a walking treadmill setup (Figure 1A). Fly turning was highly consistent across images and velocity tuning curves showed virtually no variation over different scenes, matching previous findings [31] (Figure 1B; Figure S1). To quantify reliability at the neural level, we recorded the membrane potential of horizontal system LPTCs that detect

optic flow fields corresponding to yaw rotation (Figure 1C). Potential was tuned to scene velocity and again exhibited little image-dependent variation (Figure 1D). Additionally, membrane voltage proved highly stable across time. This was consistent with earlier work in hoverflies [30].

To perform a consistent comparison, we tested the robustness of EMDs on the same set of stimuli as in behavior and electrophysiology (Figure 1E). As anticipated from a multitude of similar studies [31, 37, 42, 43], responses were remarkably unreliable across time and images (Figure 1F). For most images, temporally resolved output fluctuated strongly, average amplitudes differed, and tuning curves exhibited peaks at different velocities. Overall, EMDs provided a poor readout of true image velocity. This stands in stark contrast to the experimentally observed robustness of motion responses and leads to the central question: how does the fly visual system compensate for natural contrast variability?

Sensitivity of Optomotor Response Is Modulated by Surround Contrast

We designed an optomotor stimulus to establish whether *Drosophila* dynamically adapt the sensitivity of motion-induced turning to image contrast, which could serve to normalize variation within natural scenes. The stimulus segregated the visual field into a background and a foreground pattern (Figure 2A). The background contained random luminance fluctuations but no net motion. Pattern movement within the foreground window triggered turning. For both, average contrast could be controlled
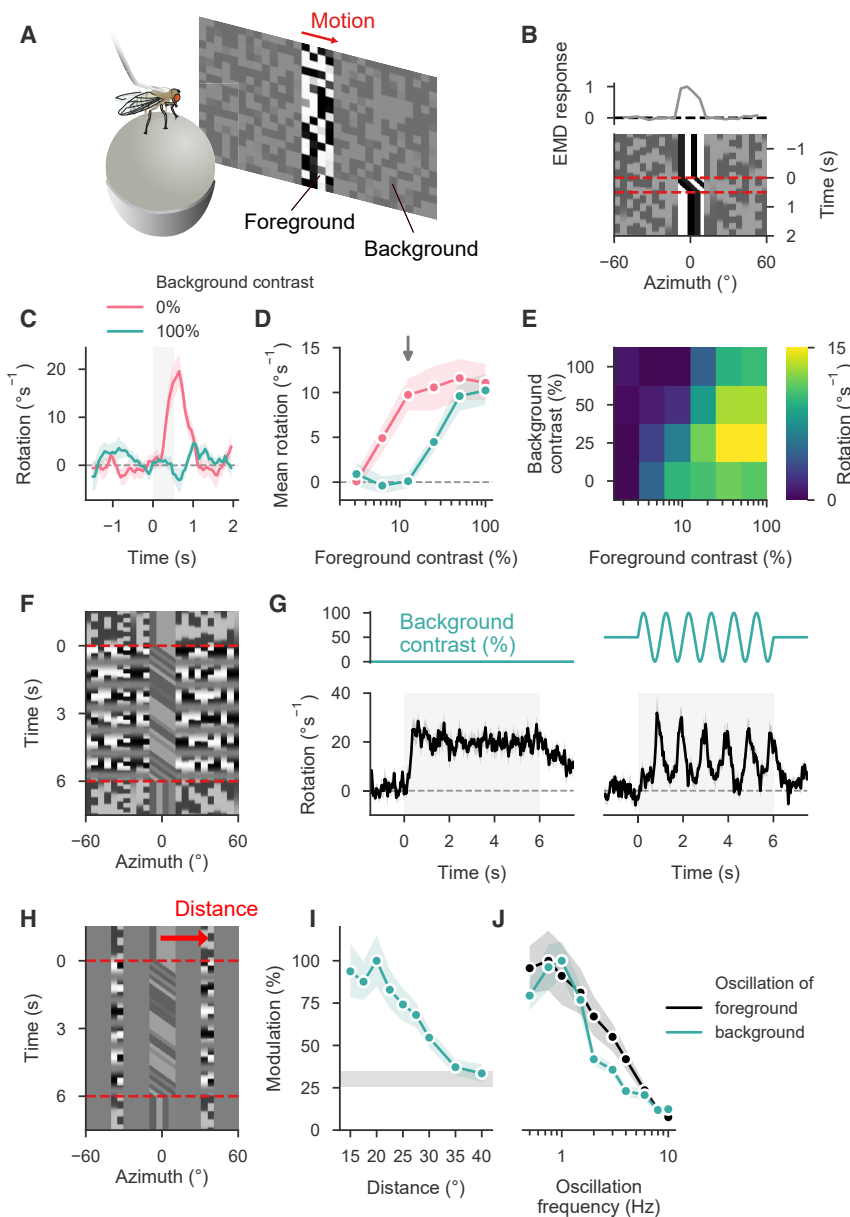
Figure 2. Sensitivity of *Drosophila* Optomotor Response Is Controlled by Surround Contrast

(A) Experimental set-up. Visual display is separated into two areas whose contrast can be set independently.

(B) Bottom: space-time plot of base stimulus. Foreground pattern moved during time span indicated by dashed lines; background is dynamic but contains no coherent motion. Top: time-averaged response of EMD array along azimuth. Only foreground produced net activity.

(C) Turning responses for extreme background contrast conditions (n = 16 wild-type flies) at foreground contrast 12.5%. Gray-shaded area indicates motion.

(D) Mean rotation (averaged between 0 and 1 s after stimulus onset) as a function of foreground contrast for two background conditions (n = 16; gray arrow indicates foreground contrast depicted in C).

(E) Heatmap of mean rotation for multiple background conditions. With increasing background contrast, optomotor sensitivity shifted rightward (n = 16).

(F) Example stimulus for mapping magnitude of sensitivity shift. Background contrast was modulated at 1 Hz.

(G) Left: baseline turning response in the absence of background contrast (n = 16, foreground contrast 25%). Right: turning response for sinusoidal change in background contrast (data taken from spatial experiment evaluated in I at distance 15°). During high-contrast phase, optomotor response was suppressed; turning modulation allowed readout of background-induced changes in gain.

(H) Illustration of spatial oscillation experiment. Distance indicates separation between centers of foreground motion and flanking background.

(I) Turning response modulation as a function of distance between motion stimulus and background (n = 16). Gray-shaded bar indicates 68% confidence interval around baseline modulation in the absence of background.

(J) Turning response modulation as a function of carrier frequency for either foreground (n = 13) or background (n = 13). Shaded area around curves indicates bootstrapped 68% confidence interval.

See also Figure S2, Table S2, and Videos S1 and S2.

independently. We confirmed that the background by itself produced no net activity in EMDs (Figure 2B).

At zero background contrast, foreground motion induced a reliable optomotor response (Figure 2C). Turning was fully suppressed at maximum background contrast, proving that turning gain is controlled by surround contrast. Average field luminance was constant for all conditions, so linear processing could not account for the phenomenon. A full measurement of contrast tuning curves for foreground motion revealed a smooth shift of the dynamic range of the optomotor response toward the current surround contrast (Figures 2D and 2E).

To efficiently map features of contrast gain control in a single stimulus condition, we sinusoidally modulated background contrast over time, which resulted in oscillations around mean turning (Figures 2F and 2G). Whenever background contrast was high, syndirectional rotation in response to motion was transiently suppressed. Evaluating oscillation amplitude thus allowed a readout of the level of contrast-induced gain adjustment. We determined the spatial scale of suppression by varying the spacing between foreground and a windowed background, separated by uniform gray (Figure 2H). Modulation fell with distance between motion stimulus and background stripe and dropped to baseline at approximately 35°, so contrast estimation was non-local but spatially limited (Figure 2I; Figures S2A and S2B; full width at half maximum of 43.8° for zero-centered Gaussian least-squares fit to mean tuning curve).

When we varied oscillation frequency in the background, suppression followed contrast changes up to fast timescales beyond 3 Hz (Figure 2J; Figures S2C–S2F). However, modulation decreased at lower frequencies than for equivalent foreground oscillations, which is indicative of temporal integration. We additionally evaluated the lag between contrast oscillation and

turning by means of cross-correlation (Figures S2G and S2H). The maximum suppressive effect of background modulation was delayed with respect to the effect of foreground modulation by approximately 70 ms (bootstrapped 95% CI: 33–114 ms). This supported the previous conclusion that the mechanism for background contrast estimation operates on slower timescales than the primary motion pathway. Silencing T4 and T5 cells abolished all contrast-guided oscillatory turning (Figures S2I–S2K), suggesting that contrast adaptation is not mediated by a system parallel to motion detection [44]. Our experiments thus point to a rapid, spatially distributed gain control mechanism that arises in early visual processing.

Signal Compression Emerges in Transient Medulla Neurons

We next used two-photon calcium imaging to locate the neural origin of contrast adaptation. The calcium indicator GCaMP6f was genetically expressed in particular cell types [45]. We targeted visual stimuli to individual neurons by determining receptive field coordinates through a combination of stochastic stimuli and online reverse correlation (Figures 3A and 3B; STAR Methods). This procedure additionally yielded estimated linear receptive fields for L1–L5, analogously to the ones previously described for medulla neurons [9] (Figures S3A–S3T). Consistent with earlier functional work [29, 46], spatiotemporal filters grouped into tonic (L3) or transient units (L1, L2, L4, and L5) like they did in the medulla. In contrast to all other lamina cells, we found that the polarity of the L5 receptive field center is ON.

To precisely map context-dependent changes in contrast sensitivity for a given cell type, we then presented drifting sine gratings with separately controlled contrast in the foreground (as defined by a 25° circular window centered on the receptive field) and the background (Figure 3C). At a fixed foreground contrast, L1 activity followed local grating luminance and was independent of background contrast (Figure 3D). Responses in downstream synaptic partner Tm3, however, showed the signature of gain control as signal amplitude was increasingly suppressed by growing surround contrast (Figure 3E).

We performed these experiments for all major columnar cell types in the circuit as well as T4 and T5 cells (Figure 3F). To obtain contrast tuning curves, we evaluated calcium modulation at the stimulus frequency. Lamina units tracked foreground contrast but were weakly, if at all, modulated by the surround except for a vertical shift at low levels (Figures 3G–3K). This was likely due to background leaking into the receptive fields since antagonistic surrounds extend beyond 25° for some cell types (Figure S3) [9]. In the medulla (Figures 3L–3U), tonic Mi4, Mi9, and Tm9 showed similar tuning as L1–L5 and again little surround dependency. However, for all transient cells (Mi1 and Tm3 for ON; Tm1, Tm2, and Tm4 for OFF), increasing background contrast had a strongly suppressive effect, which is a hallmark of divisive contrast normalization [41].

As with the corresponding behavioral experiments (Figure 2), linear receptive fields could not explain the effect given that the average luminance was constant for all conditions. Curves were shifted rightward on the logarithmic axis, which corresponds to divisive stretching in linear contrast space. Importantly, preferred direction responses in T4 and T5 were also strongly background dependent (Figures 3P and 3U) even

though not all their medulla inputs are subject to gain control. Finally, sensitivity to foreground contrast was generally higher in ON than OFF units.

Several cell types—particularly medulla transient cells—showed a dependency between fluorescence modulation at the target frequency and average response (Figure S4), possibly due to temporal integration by the calcium indicator [47]. Depending on this average activity, a saturating transformation between calcium signal and GCaMP fluorescence could by itself introduce compression of strong signal amplitudes due to ceiling effects at the far end of the sensor's dynamic range. To rule this out, we directly compared mean activity with oscillation amplitude and found no region in which this correlation was negative (Figures S4Q–S4S).

To quantify tuning curves in detail, we fit a closed-form model resembling common models of divisive normalization to the data (Figure 3V; STAR Methods) [41, 48]. Here, response gain is regulated by a divisive term that depends on background contrast while a linear term represents the combined contribution of foreground contrast and background leakage. The model accurately reproduced tuning curves for each cell type (Figure 3W; Table S1). Critically, it accounted for vertical shifts as well as sigmoidal tuning curves and context-dependent changes in contrast sensitivity.

We computed a normalization index from model parameters that estimates the degree of normalization. Given that different cell types had different baseline sensitivities and that horizontal shifts on a logarithmic scale correspond to multiplication, we quantified the relative factor by which tuning curves would shift when background contrast was increased from 0% to 100% (STAR Methods). This index was substantially higher in transient medulla cells (Mi1, Tm3, Tm1, Tm2, and Tm4) and direction-selective T4 and T5 cells than in L1–L5 or tonic medulla units (Mi4, Mi9, and Tm9; Figure 3X). Interestingly, L2 and L5 exhibited mildly elevated normalization indices. For L2, this may be related to previously described non-linearities in its receptive field structure [49].

Normalization Relies on Fast Integration of a Pool of Transient Units

Overall, fly contrast gain control appeared to be based on divisive normalization that predominantly originates in medulla units with transient response dynamics. We focused on these neurons to investigate the mechanism in detail. Responses in Mi1, Tm1, Tm2, and Tm3 were equally suppressed for all background grating directions relative to a reference stimulus with zero background contrast (Figure 4A). Temporal frequency tunings for suppression resembled band-pass filters with a peak at 2 Hz (Figure 4B). Crucially, static backgrounds did not have a suppressive effect. Suppression steadily increased with the outer diameter of an annulus containing the background pattern, which again indicated an extended integration area (Figure 4C). Spatiotemporal features of neural gain control thus matched our findings from behavior (Figure 2).

To determine the temporal scale of normalization, we designed a contrast-step stimulus in which the foreground was replaced by a single light pulse matching each cell type's polarity (Figure 4D). By varying the time interval between motion onset of the background grating and the onset of the pulse, we scanned
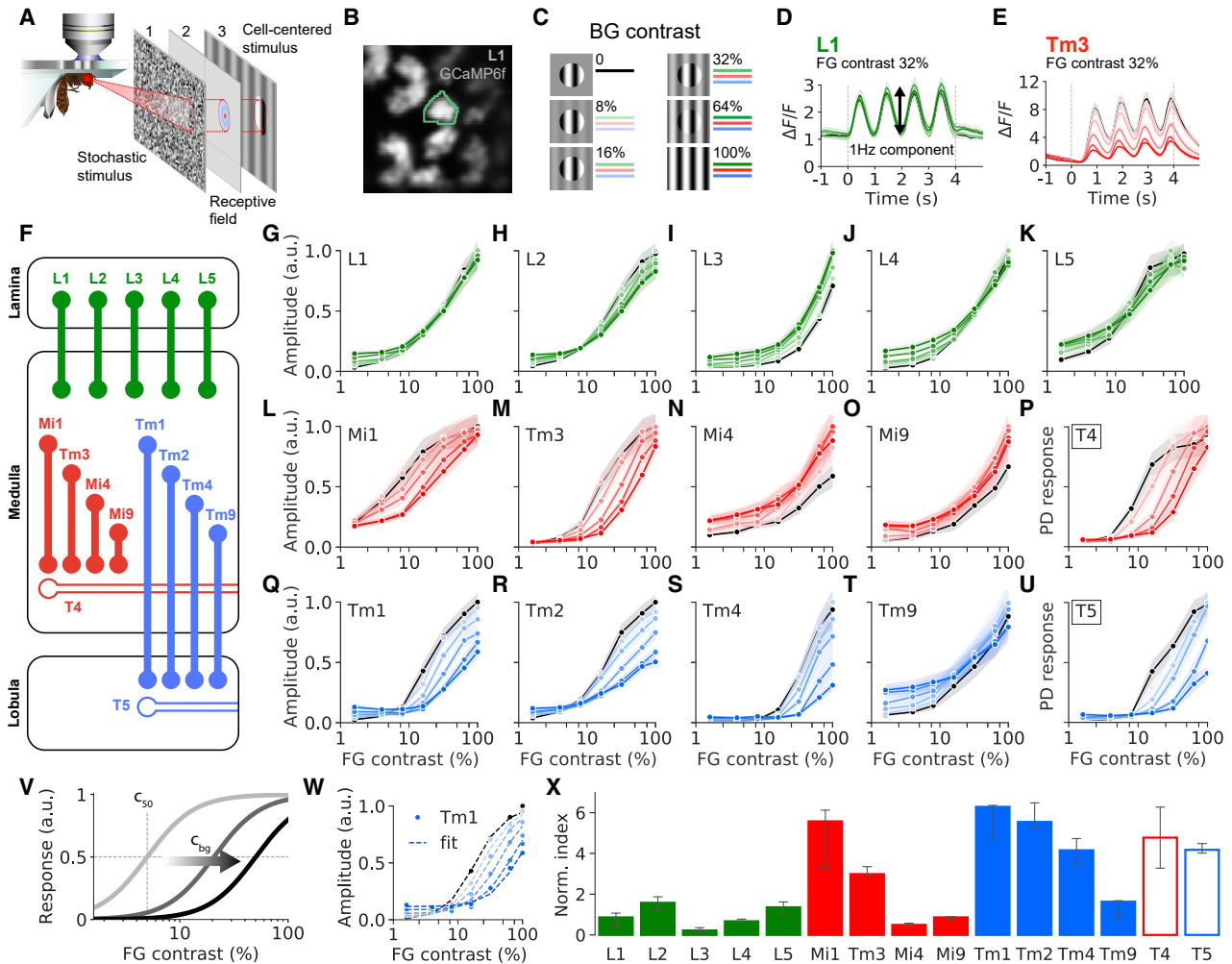
Figure 3. Contrast Normalization Emerges in Transient Medulla Neurons

(A) Schematic of experimental procedure. (1) White noise stimulus. (2) Receptive field reconstruction from single-neuron calcium signals. (3) Drifting grating with different contrasts in foreground and background.

(B) Two-photon image of L1 axon terminals expressing GCaMP6f. Green line indicates example region of interest.

(C) Experimental protocol. Darker color shade corresponds to higher background contrast as used in (G)–(U). Zero background contrast condition is shown in black.

(D and E) Average calcium responses of L1 (D) and Tm3 (E) for fixed foreground and various background contrasts.

(F) Schematic of the motion circuit including all neurons measured.

(G–K) Contrast tuning curves measured as amplitude of calcium signals at stimulus frequency for L1–L5. Shaded areas show bootstrapped 68% confidence intervals around the mean (L1 in G: 21/7 cells/flies, L2 in H: 26/8, L3 in I: 23/6, L4 in J: 19/6, L5 in K: 18/9).

(L–P) Contrast tuning curves for ON pathway neurons (Mi1 in L: 20/5, Tm3 in M: 21/8, Mi4 in N: 20/13, Mi9 in O: 21/9, T4 in P: 23/10).

(Q–U) Contrast tuning curves for OFF pathway neurons (Tm1 in Q: 21/7, Tm2 in R: 20/6, Tm4 in S: 20/13, Tm9 in T: 19/6, T5 in U: 21/9).

(V) Illustration of divisive normalization model for tuning curves. Increasing background contrast c_{bg} shifts the sigmoidal tuning curve from baseline sensitivity c_{50} to higher contrasts.

(W) Example fit of model for Tm1.

(X) Normalization index for all neurons shown as median with 68% bootstrapped confidence intervals. Transient medulla neurons Mi1, Tm3, Tm1, Tm2, and Tm4, as well as T4 and T5, exhibited strongest degree of normalization.

See also Figures S3 and S4, Tables S1 and S2, and Video S3.

the temporal profile of the suppressive signal. For the tested neurons Tm3 and Tm2, we found virtually immediate response reduction within a measurement precision of 50 ms given by the smallest tested onset difference. We observed transient ringing of suppression strength at the background temporal frequency. Ringing was stronger when the grating was present before motion onset compared to when it was masked by uniform gray. A similar effect has been described in LPTCs [26], where it results from neural integration of multiple transient, out-of-phase inputs. In sum, these findings indicated that surround suppression derives from a pool of transient neurons that are not selective for direction. Both isotropy and frequency
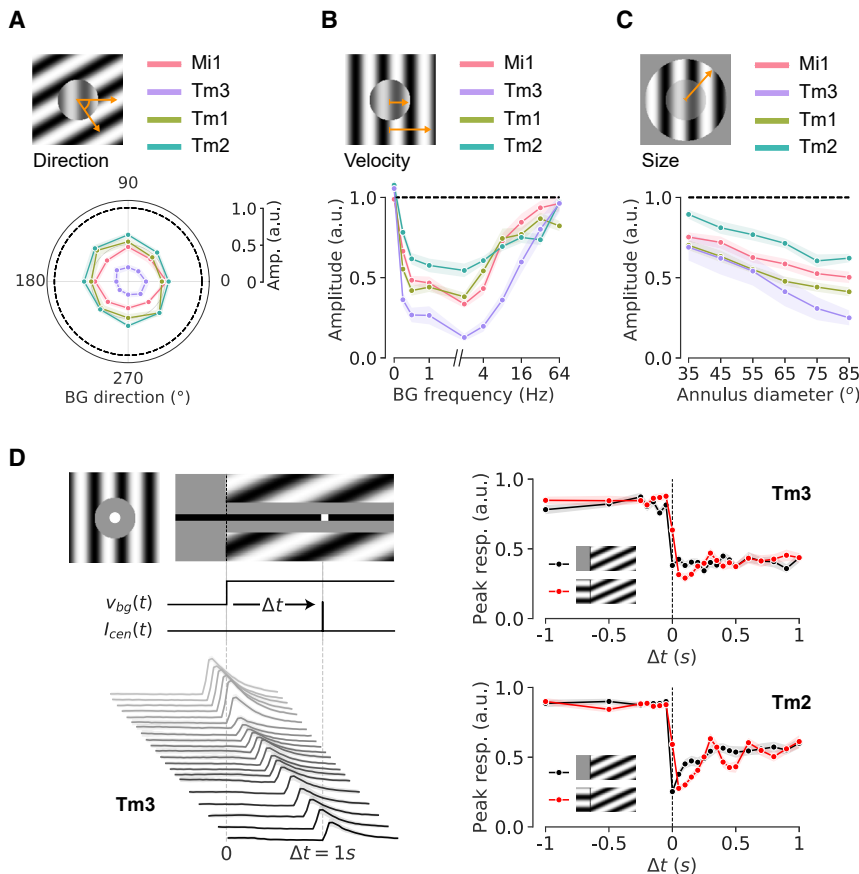
Figure 4. Neural Contrast Normalization Relies on Rapid Integration of a Pool of Transient Units

(A) Polar plot of response amplitude for different directions of background motion. Black dashed line represents response to reference stimulus with background contrast of 0%. For each neuron, foreground contrast was chosen to maximize possible background suppression (Mi1: 16%, Tm3: 32%, Tm1: 64%, Tm2: 100%).

(B) Responses for different background contrast frequencies, revealing band-pass tuning of suppression.

(C) Suppression strength increased with outer diameter of background annulus (Mi1: 21/9 cells/flies, Tm3: 20/6, Tm1: 18/6, Tm2: 21/4 in A–C).

(D) Top left: x-y and x-t plots of contrast-step stimulus for Tm3 (ON center). Background contrast frequency was 3 Hz. Center left: velocity function $v_{bg}(t)$ of background and intensity function $I_{cen}(t)$ of center pulse. Bottom left: mean responses of Tm3 for different time intervals Δt. Right: mean peak amplitude for Tm3 and Tm2 (Tm3: 19/6, Tm2: 20/5). Black line shows condition where the background grating was masked before onset; red where background was visible but static.

Shaded areas around curves indicate bootstrapped 68% confidence intervals. See also modeling in Figure S5, Table S2, and Video S4.

When measuring tuning curves (similar to Figure 3 but only for background contrasts 0% and 100%), baseline contrast sensitivity as measured by the semi-saturation constant of model fits was significantly increased (Figure 5D). This suggests that Tm3 cells were disinhibited due to a reduced pool signal. We observed similar effects for ON-sensitive Mi1 cells (Figures 5E and 5F), but the impact was less pronounced than for Tm3 cells. Absolute signal amplitude was generally not affected by silencing, demonstrating that cells remained visually responsive in the presence of TNT (see Figure 5B).

In the OFF pathway, blocking Tm1 cells did not have any significant effects (Figures 5G and 5H). In contrast, when blocking Tm2, we observed an almost complete loss of background suppression across frequencies (Figure 5I). For this cell type, we did not observe any change in contrast tuning curves for the 0% background condition, and consequently, the fitted semi-saturation constant was not affected (Figure 5J). For full background contrast, however, suppression at high foreground contrasts was strongly reduced. Additionally, background leakage at low foreground contrasts increased substantially compared to control flies. As with Tm3 and Mi1, this is compatible with Tm2 cells being disinhibited due to the silencing of a suppressive signal derived from recurrent output. We therefore conclude that in the fly, contrast normalization is at least partially based on feedback from a combination of medulla neurons.

tunings were strikingly similar to filter properties of the transient lamina and medulla units involved in motion detection (Figures S3U and S3V). This suggested that one or more of these cell types provides input to the suppressive pool.

To determine whether a mechanism that integrates transient units across space to divisively suppress local responses could reproduce our findings, we built a time-resolved, data-driven model. The model faithfully predicted direction, frequency, and size tunings, as well as contrast-step ringing, T4 and T5 responses, and LPTC output for our behavioral stimuli (Figure S5A–K).

Neural Feedback Is Critical for Contrast Normalization

Spatial pooling, however, could occur over either feedforward signals from the lamina or feedback from the medulla (Figure 5A). In vertebrate systems, it has proven difficult to distinguish the two [41, 50, 51]. Fly transient units in the lamina or medulla have similar temporal properties (Figures S3U and S3V), and both implementations produce equivalent steady-state output [48], so we used genetic silencing to pinpoint the source. We co-expressed a calcium indicator and the tetanus toxin light chain (TNT; STAR Methods) [52] in different medulla cell types, blocking chemical synaptic output and thus feedback from the entire neuron array but leaving feedforward input and calcium signals intact.

For the ON pathway unit Tm3, we observed significantly reduced suppression across background frequencies when compared to controls with inactive TNT (Figures 5B and 5C).

Contrast Normalization Improves Robustness to Natural Scene Variability

Could this type of response normalization account for the robustness of fly motion detection? Previous work on EMDs
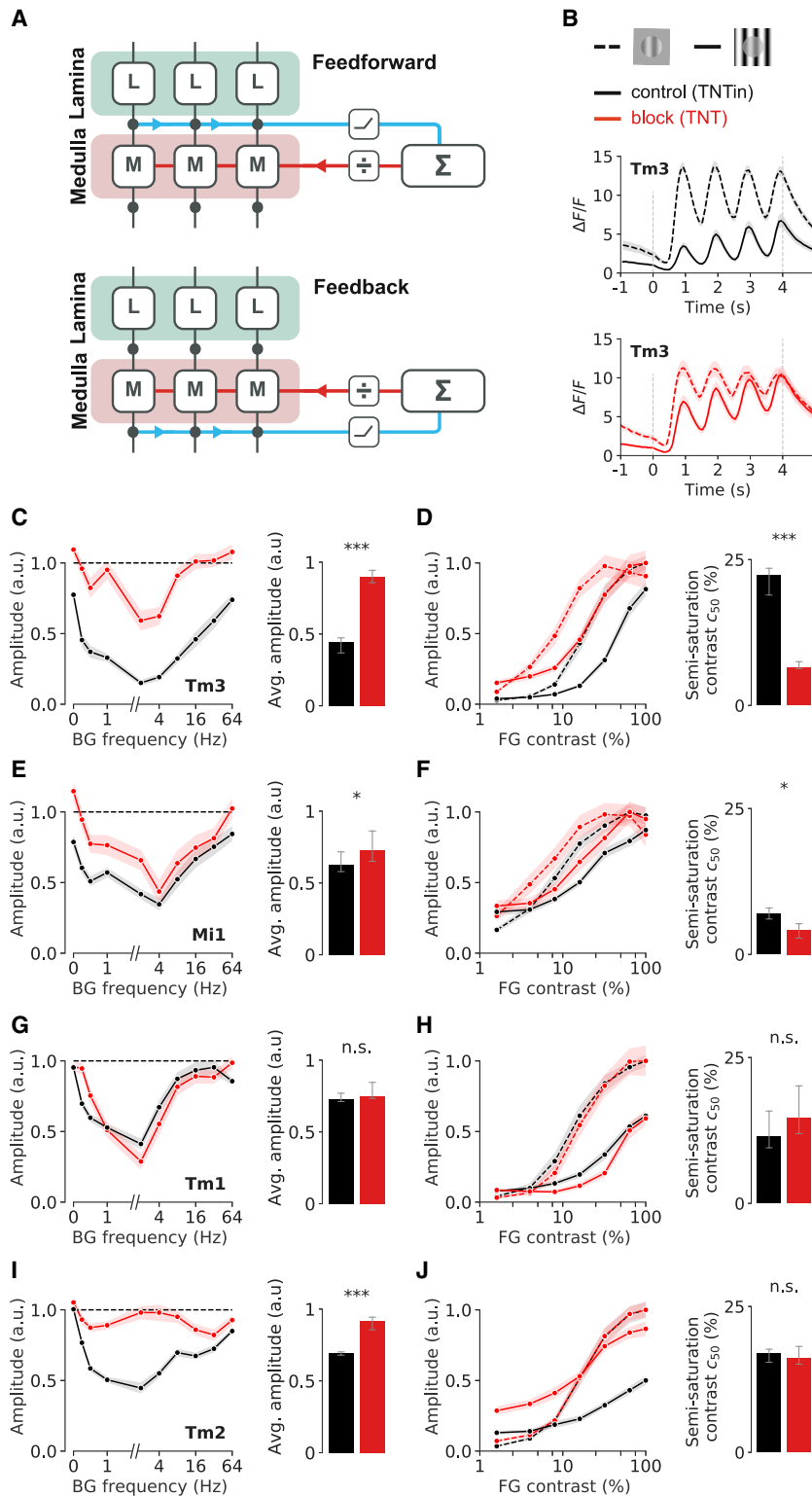
Figure 5. Neural Feedback Underlies Contrast Normalization

(A) Schematic of feedforward and feedback model for surround suppression.

(B) Mean responses of Tm3 for TNT block (red) and TNTin controls (black) at background frequency 16 Hz (dashed line indicates reference response and solid line the response at full background contrast; Tm3 block: 21/5 cells/flies, Tm3 control: 20/5).

(C) Left: frequency tuning for block experiment. Black dashed line represents response to reference stimulus. Right: average amplitude over all frequencies was higher for Tm3 block flies (Mann-Whitney U: 8, ***p < 0.001).

(D) Left: foreground contrast tuning for block experiments at 0% and 100% background contrast. Right: contrast sensitivity was increased for Tm3 block flies as measured by lowered semi-saturation constant c_{50} (Mann-Whitney U: 39, ***p < 0.001).

(E) Blocking results for Mi1 (as in C). Average amplitude over all frequencies was reduced for Mi1 block flies (Mi1 block: 20/5, Mi1 control: 21/6; Mann-Whitney U: 143, *p = 0.04).

(F) Blocking results for Mi1 (as in D). Contrast sensitivity was increased for Mi1 block flies (Mann-Whitney U: 128, *p = 0.02).

(G) Blocking results for Tm1 (as in C). No significant effect was found for Tm1 block flies (Tm1 block: 20/5, Tm1 control: 19/5; Mann-Whitney U: 169, NS p = 0.28).

(H) Blocking results for Tm1 (as in D). Sensitivity was not affected (Mann-Whitney U: 158, NS p = 0.19).

(I) Blocking results for Tm2 (as in C; Tm2 block: 20/5, Tm2 control: 25/6; Mann-Whitney U: 17, ***p < 0.001).

(J) Blocking results for Tm2 (as in D; Mann-Whitney U: 239, NS p = 0.49). Semi-saturation constant at 0% background contrast did not change for Tm2 block flies. Shaded areas show bootstrapped 68% confidence intervals around the mean. Error bars show bootstrapped 68% confidence intervals around the median.

See also Table S2.

data-driven LPTC model and found moderate reduction of cross-image variability compared to a model with bypassed normalization (Figures S5L–S5N). However, post hoc ablation may specifically disadvantage the simpler model. To investigate performance limits in a principled way, we pursued a task-driven approach.

Recent progress in deep artificial networks has made it feasible to use image-processing models of neural systems for rigorously assessing performance on real-world problems [54–56]. EMD-like architectures are concisely expressed as multi-layer convolutional networks [54] and fully differentiable, rendering them amenable to optimization methods like gradient descent.

and natural scenes has exploited compressive transforms but did so heuristically or without surround-dependent gain control [42, 43, 53]. We evaluated natural image responses in the

We designed a fly-like neural network and independently trained possible types of contrast processing such that each model class could optimally adapt to a specific, behaviorally relevant task.
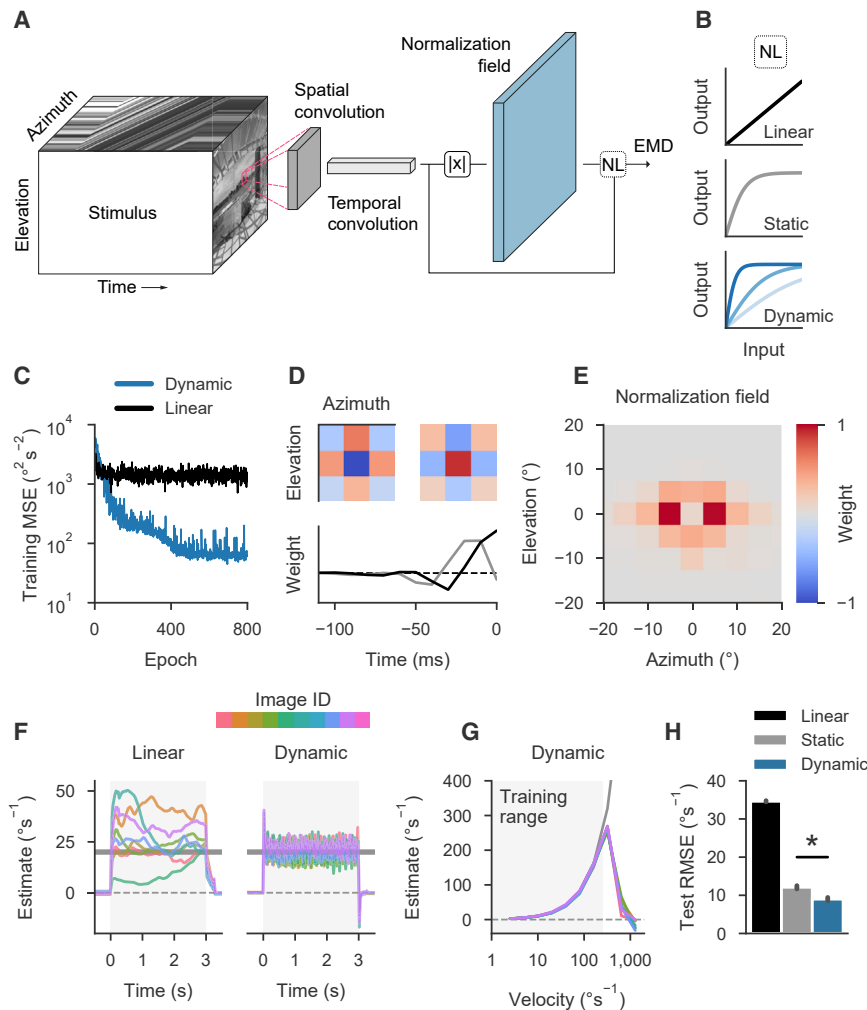
Figure 6. Contrast Normalization Enhances Robustness to Natural Scene Variability

(A) Schematic of single convolutional input filter. Motion stimuli are sequentially processed by a spatial 3 × 3 × 1 (azimuth, elevation, time) and a temporal 1 × 1 × 30 filter. Through a transfer function, the signal is combined with a normalization signal generated by a 11 × 11 × 1 convolution operating on full-wave rectified input signal. The output of two distinct channels is processed analogously to multiplicative EMDs.

(B) Input-output relationships for linear, static, and dynamic models. In the dynamic model, response sensitivity is a function of normalization field activity.

(C) Training mean squared error (MSE) for two example models during stochastic gradient descent.

(D) Spatial and temporal receptive fields for the two channels of a typical dynamic model. Depicted are normalized filter weights.

(E) Spatial receptive field of normalization pool for the model from (D).

(F) Model output for individual images moving at $20°s^{-1}$ during gray-shaded period. Gray line indicates target velocity. Left: example model without non-linearity. Right: example model with dynamic non-linearity.

(G) Velocity tuning curves of example dynamic model for individual images (averaged between 0 and 3 s after motion onset). Gray line indicates true velocity. Gray-shaded area indicates the 99th percentile of absolute velocities in training set.

(H) Mean performance of trained models on held-out test set, estimated as root mean square error (RMSE; n = 22/23/16 for linear/static/dynamic; *p < 0.001, t = 9.01, Student's t test with assumed equal variance; only difference between static and dynamic was tested). Error bars indicate bootstrapped 68% confidence intervals.

See also Figure S6.

All models featured linear, spatiotemporally separable input convolutions (Figure 6A). We evaluated three alternatives for contrast transformation: a linear stage where output was transmitted unchanged, a statically compressive stage that limited signal range independently of context, and a dynamic compression stage with adaptive gain depending on the output of a contrast-sensitive surround filter (Figures 6A and 6B; STAR Methods). Resulting output from two distinct channels was then processed according to a multiplicative EMD scheme. Through backpropagation and stochastic gradient descent, models were trained to estimate the true velocity of natural images translating at random speeds.

All models successfully learned the task on the training set (Figure 6C). We initialized convolutions randomly but after training observed antagonistic spatial filters and transient temporal filters where one channel was phase delayed with respect to the other (Figure 6D; Figures S6A–S6C). Models thus made extensive use of redundancy reduction through center-surround configurations [35] and discovered the EMD strategy of delay and compare [26]. Normalization fields for the dynamic model spanned approximately 30° in azimuth and invariably excluded information from the center of the filter (Figure 6E; Figure S6C). Interestingly, dynamic models exploited normalization in both

channels and switched normalization strategies during training, transitioning from purely static to purely context-dependent compression (Figures S6D and S6E). Overall, normalized networks acquired representations that matched filtering and gain control properties of the fly medulla.

When tested on previous experimental stimuli (Figure 1), linear models exhibited improved velocity tuning curves compared to a standard EMD (Figures 1F and 6F; Figure S6F), but estimates still varied substantially across time. Dynamic models, on the other hand, proved extremely robust at extracting scene motion across time, images, and velocities within the velocity range of the training set (Figures 6F and 6G). Given that all networks were based on a multiplicative EMD scheme, typical phenomena like the velocity optimum were still present. We compared average estimation error on a held-out test set and found both types of non-linear compression to vastly outperform the linear stage (Figure 6H). The performance of static compression indicates that simple response saturation already enhances robustness to contrast fluctuations in natural scenes. However, fly-like context sensitivity consistently decreased test error over the static non-linearity (error reduction 22.0%–29.2%; bootstrapped 95% CI). Finally, we benchmarked generalization on a fully independent image set (Figure S6G), where linear models failed

catastrophically while both compressive stages retained performance. This was particularly pronounced when testing images with high dynamic range (STAR Methods). Critically, on all datasets, dynamic compression resulted in substantial error reduction with respect to both linear transfer and static compression.

DISCUSSION

In summary, our work represents the first demonstration that divisive contrast normalization occurs in the fly visual system and offers a comprehensive look at non-linear response properties in a virtually complete motion vision circuit. We established at multiple levels of motion processing that responses to moving panoramas are substantially more robust than predicted by correlation-based models of the system. Our behavioral experiments indicate that the sensitivity of the optomotor response is regulated by average contrast in a spatially confined part of the visual field. Critically, we traced the emergence of this dynamic signal compression to local elements in the medulla of the fly optic lobe and used targeted circuit manipulation to identify neural feedback as a critical underlying mechanism. Finally, our task-driven approach revealed that the inclusion of spatial contrast normalization drastically improves velocity estimation in correlation-based models of fly motion vision.

Implications for Fly Motion Vision
Previous work on the function of local units in the *Drosophila* optic lobe mostly explored linear properties of light responses, often relying on first-order systems identification techniques like reverse correlation [9–11, 29]. Investigation of non-linear contributions generally focused on computations in direction-selective T4 and T5 cells [13–19, 57, 58].

Here, we describe a powerful non-linearity, adaptive gain control that occurs in a majority of columnar neurons involved in the detection of motion. This casts doubt on the extent to which existing functional descriptions can be generalized. Linear filter estimates are typically based on responses to dynamic noise stimuli of fixed amplitude [9, 10, 29]. Our work suggests that this contrast regime only corresponds to one particular adaptation state for any measured cell type, so filter properties may well differ for stimuli with differing contrast characteristics. Step and edge responses, for instance, are usually measured on backgrounds with uniform luminance [11, 12, 18, 59]. This places cells in a maximally sensitive state due to lack of surround inhibition and is likely to affect both response amplitude and kinetics. Signal compression may reconcile observed discrepancies between studies conducted with different stimuli.

Interestingly, visual interneurons exhibited qualitatively different sensitivity curves even at constant background contrast. In the lamina, for instance, only tonic cell type L3 responded linearly to increasing visual contrast. Sensitivity curves of transiently responding cell types like L1 and L2, on the other hand, proved approximately logarithmic. This is in line with expectations from previous work in other fly species [60] but deviates from predictions based on white noise characterizations [29].

Moreover, we observed a stark discrepancy in baseline sensitivity between ON- and OFF-sensitive neurons, where tuning curves of dark-selective units were shifted toward higher pattern contrast. Notably, due to strong surround suppression, full-field

gratings elicited comparatively weak responses in T5 units whereas T4 cells were driven effectively by the same stimuli. This adds to previous work on ON-OFF asymmetries in the *Drosophila* visual system [31, 38]. We conclude that even at primary processing stages, the fly visual system represents contrast in a multiplexed fashion where individual channels diverge with respect to how they transmit information about luminance differences. The function of these asymmetries remains to be investigated.

The proposed model based on divisive normalization accurately captures most features of the observed contrast tuning curves (see Figure 3; Figure S5; Table S1). Certain discrepancies remain. For instance, the normalization model predicts that responses for different background contrasts eventually plateau at the same level. However, we observed in both behavior (Figure 2E) and T4 responses (Figure 3P) that in the absence of background contrast, saturation occurred at a lower level than for other conditions. To explain such non-monotonic behavior, further investigation of the underlying mechanism is required.

Divisive normalization of local motion signals has previously been suggested to occur at the level of LPTCs, through either isotropic pooling of EMDs in hypothetical secondary cell types [33] or passive membrane properties of LPTCs [61, 62]. Here, we show that gain control already originates upstream of motion-sensitive cell types T4 and T5. However, LPTC-intrinsic gain control mechanisms, including temporal adaptation [32], could well be complementary such that at each processing stage, the fly visual system makes use of compression to optimize the reliability of output signals.

In flies, there is ample evidence for changes in visual coding that depend on the behavioral state of the animal. Various interneurons within the optic lobe, for instance, are affected by the activity of octopaminergic projection units, leading to drastic shifts in response gain or temporal tuning [9, 59, 63–67]. Our calcium imaging experiments were performed in immobilized *Drosophila*. It will be of interest to explore whether the properties of contrast gain control are modulated by locomotion, particularly in highly state-sensitive units like Mi4 [59].

Mechanism of Signal Compression
Our experiments suggest that neural feedback plays a crucial role in gain adjustment. At this point, the cellular origin of feedback is unknown. Present experiments indicate a visual integration field that spans many columns (Figures 2 and 4). Moreover, the observed contrast compression appears to be suppressive. All tested medulla cell types with strong background contrast dependency emit acetylcholine, which, in the *Drosophila* visual system, is generally thought to be excitatory [68, 69]. Inhibitory interneurons could mediate the required synaptic sign reversal. Signal compression could then be implemented through lateral neighbor-to-neighbor interactions between columnar medulla units where suppressive signals spread through a local network. Alternatively, we hypothesize that wide-field interneurons pool local medulla units across multiple columns and provide recurrent inhibitory input to the same cells. In our data-driven model, such a pool cell mechanism accounted for all observed spatio-temporal properties of signal compression including ringing effects (Figure S5). Finally, our TNT-based intervention strategy

should leave coupling via electrical synapses intact [52, 70]. We can therefore not exclude that gap junctions are also involved in shaping contrast response properties.

Silencing feedback from individual medulla cell types had differential effects, ranging from completely abolished suppression in Tm2 to unchanged responses in Tm1 (Figure 5). This suggests either that multiple cell types feed into the pool signal with varying weight or that alternative mechanisms provide the compressive signal, for example, in Tm1. Moreover, it is an open question whether all cell types are suppressed by one or multiple pool cell types. Asymmetries in sensitivity between ON and OFF pathways, for instance, could be an indicator for polarity-specific sources of suppression.

In both distal and proximal layers of the medulla, the class of neuropil-intrinsic Dm and Pm neurons contains approximately 20 cell types and offers a possible substrate for the mechanism [71, 72]. These neurons arborize within the medulla and exhibit diverse stratification and tiling patterns, often spanning dozens of columns and thus approximately matching the observed suppression field of local units. Dm and Pm units release either GABA or glutamate for which receptors in the fly visual system are mostly inhibitory [68], pointing to these cell types as potential candidates for gain control.

Functional Relevance

Normalization has often been described as a generic mechanism for removing higher-order correlations from natural signals [5, 73–75]. Here, we close the loop between neural mechanism and an ecologically critical behavior, the optomotor response, and demonstrate how contrast gain control can render motion detection resilient to challenges imposed by natural scene statistics. Specifically, normalization serves to distinguish between ecologically relevant parameters like retinal image velocity and nuisance factors like image contrast.

Various biomimetic modeling studies have incorporated compressive transforms along the motion processing cascade to improve robustness under naturalistic visual conditions [43, 53, 76]. In contrast to our work, these normalization stages were not based on experimental evidence, required *ad hoc* parameter tuning, and generally operated in the temporal domain. Interestingly, the fly visual system bases gain control on a temporally immediate, spatially extended estimate of contrast. This represents a trade-off where spatial resolution is sacrificed in favor of temporal resolution, which may be advantageous for global optic flow estimation in rapidly moving animals.

To assess the exact causal contribution of contrast compression to the robustness of velocity estimation in *Drosophila*, one would need to disrupt this mechanism specifically while leaving all other visual processing intact. Silencing the synaptic output of medulla neurons (Figure 5) demonstrates the importance of neural feedback for gain control but should additionally affect feedforward processing in downstream units, particularly T4 and T5 [11, 12, 77, 78]. Future mapping of the circuits underlying contrast compression will provide the tools for establishing causality.

The convolutional network (Figure 6) solves the task of estimating velocity across diverse environments and at little computational cost, particularly compared to standard optic flow algorithms like the Lucas-Kanade method [79]. Present findings may thus aid the design of low-power, low-latency machine vision systems suitable for autonomous vehicles [80, 81].

Comparison with Other Sensory Systems

Gain control in the *Drosophila* optic lobe bears a striking resemblance to normalization in other systems and modalities like fly olfaction [82] or mammalian auditory cortex [83] as well as processing in vertebrate visual areas from retina to V1 [48, 84–86]. Spatial and temporal tuning or isotropy of non-linear surround suppression in the lateral geniculate nucleus, in particular, qualitatively match that of transient units in the fly medulla [40]. The present study suggests differences at the implementation level. For instance, investigations into divisive normalization in mammalian V1 cells point to feedforward mechanisms underlying gain control whereas the fly visual system appears to rely primarily on feedback signals (Figure 5) [50]. Both systems, however, realize a similar algorithm. This provides further proof for evolutionary convergence on canonical solutions in neural sensory processing [41].

Overall, our work establishes the *Drosophila* visual system with its defined cell types, known connectivity patterns, powerful genetic toolkit, and direct correspondence between circuit and task as a novel model for the study of normalization. It thus lays the foundation for future mechanistic inquiries into the functional, cellular, molecular, and biophysical underpinnings of a crucial computation in sensory processing.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Natural image sets
  - Behavioral experiments
  - Electrophysiology
  - Calcium imaging
  - Modeling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data evaluation for behavioral experiments
  - Data evaluation for electrophysiological experiments
  - Data evaluation for calcium imaging experiments
  - Statistical tests
- DATA AND CODE AVAILABILITY

REFERENCES

1. Gibson, J.J. (1950). The Perception of the Visual World (Houghton Mifflin).

2. Dickinson, M.H. (2014). Death Valley, *Drosophila*, and the Devonian toolkit. Annu. Rev. Entomol. *59*, 51–72.

3. Borst, A. (2014). Fly visual course control: behaviour, algorithms and circuits. Nat. Rev. Neurosci. *15*, 590–599.

4. Geisler, W.S. (2008). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. *59*, 167–192.

5. Rieke, F., and Rudd, M.E. (2009). The challenges natural images pose for visual adaptation. Neuron *64*, 605–616.

6. Mauss, A.S., Vlasits, A., Borst, A., and Feller, M. (2017). Visual circuits for direction selectivity. Annu. Rev. Neurosci. *40*, 211–230.

7. Yang, H.H., and Clandinin, T.R. (2018). Elementary motion detection in *Drosophila*: Algorithms and mechanisms. Annu. Rev. Vis. Sci. *4*, 143–163.

8. Shinomiya, K., Huang, G., Lu, Z., Parag, T., Xu, C.S., Aniceto, R., Ansari, N., Cheatham, N., Lauchie, S., Neace, E., et al. (2019). Comparisons between the ON- and OFF-edge motion pathways in the *Drosophila* brain. eLife *8*, e40025.

9. Arenz, A., Drews, M.S., Richter, F.G., Ammer, G., and Borst, A. (2017). The temporal tuning of the *Drosophila* motion detectors is determined by the dynamics of their input elements. Curr. Biol. *27*, 929–944.

10. Behnia, R., Clark, D.A., Carter, A.G., Clandinin, T.R., and Desplan, C. (2014). Processing properties of ON and OFF pathways for *Drosophila* motion detection. Nature *512*, 427–430.

11. Serbe, E., Meier, M., Leonhardt, A., and Borst, A. (2016). Comprehensive characterization of the major presynaptic elements to the *Drosophila* OFF motion detector. Neuron *89*, 829–841.

12. Strother, J.A., Wu, S.-T., Wong, A.M., Nern, A., Rogers, E.M., Le, J.Q., Rubin, G.M., and Reiser, M.B. (2017). The emergence of directional selectivity in the visual motion pathway of *Drosophila*. Neuron *94*, 168–182.e10.

13. Salazar-Gatzimas, E., Agrochao, M., Fitzgerald, J.E., and Clark, D.A. (2018). The neuronal basis of an illusory motion percept is explained by decorrelation of parallel motion pathways. Curr. Biol. *28*, 3748–3762.e8.

14. Maisak, M.S., Haag, J., Ammer, G., Serbe, E., Meier, M., Leonhardt, A., Schilling, T., Bahl, A., Rubin, G.M., Nern, A., et al. (2013). A directional tuning map of *Drosophila* elementary motion detectors. Nature *500*, 212–216.

15. Gruntman, E., Romani, S., and Reiser, M.B. (2018). Simple integration of fast excitation and offset, delayed inhibition computes directional selectivity in *Drosophila*. Nat. Neurosci. *21*, 250–257.

16. Salazar-Gatzimas, E., Chen, J., Creamer, M.S., Mano, O., Mandel, H.B., Matulis, C.A., Pottackal, J., and Clark, D.A. (2016). Direct measurement of correlation responses in *Drosophila* elementary motion detectors reveals fast timescale tuning. Neuron *92*, 227–239.

17. Leong, J.C.S., Esch, J.J., Poole, B., Ganguli, S., and Clandinin, T.R. (2016). Direction selectivity in *Drosophila* emerges from preferred-direction enhancement and null-direction suppression. J. Neurosci. *36*, 8078–8092.

18. Haag, J., Arenz, A., Serbe, E., Gabbiani, F., and Borst, A. (2016). Complementary mechanisms create direction selectivity in the fly. eLife *5*, e17421.

19. Fisher, Y.E., Silies, M., and Clandinin, T.R. (2015). Orientation selectivity sharpens motion detection in *Drosophila*. Neuron *88*, 390–402.

20. Bahl, A., Ammer, G., Schilling, T., and Borst, A. (2013). Object tracking in motion-blind flies. Nat. Neurosci. *16*, 730–738.

21. Busch, C., Borst, A., and Mauss, A.S. (2018). Bi-directional control of walking behavior by horizontal optic flow sensors. Curr. Biol. *28*, 4037–4045.e5.

22. Haikala, V., Joesch, M., Borst, A., and Mauss, A.S. (2013). Optogenetic control of fly optomotor responses. J. Neurosci. *33*, 13927–13934.

23. Krapp, H.G., and Hengstenberg, R. (1996). Estimation of self-motion by optic flow processing in single visual interneurons. Nature *384*, 463–466.

24. Hassenstein, B., and Reichardt, W. (1956). Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*. Z. Naturforsch. B *11*, 513–524.

25. Joesch, M., Plett, J., Borst, A., and Reiff, D.F. (2008). Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila melanogaster*. Curr. Biol. *18*, 368–374.

26. Borst, A., Reisenman, C., and Haag, J. (2003). Adaptation of response transients in fly motion vision. II: Model studies. Vision Res. *43*, 1309–1322.

27. Borst, A., Flanagin, V.L., and Sompolinsky, H. (2005). Adaptation without parameter change: Dynamic gain control in motion detection. Proc. Natl. Acad. Sci. USA *102*, 6172–6176.

28. Eichner, H., Joesch, M., Schnell, B., Reiff, D.F., and Borst, A. (2011). Internal structure of the fly elementary motion detector. Neuron *70*, 1155–1164.

29. Clark, D.A., Bursztyn, L., Horowitz, M.A., Schnitzer, M.J., and Clandinin, T.R. (2011). Defining the computational structure of the motion detector in *Drosophila*. Neuron *70*, 1165–1177.

30. Straw, A.D., Rainsford, T., and O'Carroll, D.C. (2008). Contrast sensitivity of insect motion detectors to natural images. J. Vis. *8*, 1–9.

31. Leonhardt, A., Ammer, G., Meier, M., Serbe, E., Bahl, A., and Borst, A. (2016). Asymmetry of *Drosophila* ON and OFF motion detectors enhances real-world velocity estimation. Nat. Neurosci. *19*, 706–715.

32. Harris, R.A., O'Carroll, D.C., and Laughlin, S.B. (2000). Contrast gain reduction in fly motion adaptation. Neuron *28*, 595–606.

33. Reichardt, W., Poggio, T., and Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual system of the fly. Biol. Cybern. *46*, 1–30.

34. van Hateren, J.H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. Vision Res. *37*, 3407–3416.

35. Srinivasan, M.V., Laughlin, S.B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. Proc. R. Soc. Lond. B Biol. Sci. *216*, 427–459.

36. Mauss, A.S., Pankova, K., Arenz, A., Nern, A., Rubin, G.M., and Borst, A. (2015). Neural circuit to integrate opposing motions in the visual field. Cell *162*, 351–362.

37. Fitzgerald, J.E., and Clark, D.A. (2015). Nonlinear circuits for naturalistic visual motion estimation. eLife *4*, e09123.

38. Clark, D.A., Fitzgerald, J.E., Ales, J.M., Gohl, D.M., Silies, M.A., Norcia, A.M., and Clandinin, T.R. (2014). Flies and humans share a motion estimation strategy that exploits natural scene statistics. Nat. Neurosci. *17*, 296–303.

39. Demb, J.B. (2008). Functional circuitry of visual adaptation in the retina. J. Physiol. *586*, 4377–4384.

40. Bonin, V., Mante, V., and Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. J. Neurosci. *25*, 10844–10856.

41. Carandini, M., and Heeger, D.J. (2011). Normalization as a canonical neural computation. Nat. Rev. Neurosci. *13*, 51–62.

42. Dror, R.O., O'Carroll, D.C., and Laughlin, S.B. (2001). Accuracy of velocity estimation by Reichardt correlators. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *18*, 241–252.

43. Brinkworth, R.S.A., and O'Carroll, D.C. (2009). Robust models for optic flow coding in natural scenes inspired by insect biology. PLoS Comput. Biol. *5*, e1000555.

44. Bahl, A., Serbe, E., Meier, M., Ammer, G., and Borst, A. (2015). Neural mechanisms for *Drosophila* contrast vision. Neuron *88*, 1240–1252.

45. Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

46. Silies, M., Gohl, D.M., Fisher, Y.E., Freifeld, L., Clark, D.A., and Clandinin, T.R. (2013). Modular use of peripheral input channels tunes motion-detecting circuitry. Neuron *79*, 111–127.

47. Strother, J.A., Nern, A., and Reiser, M.B. (2014). Direct observation of ON and OFF pathways in the *Drosophila* visual system. Curr. Biol. *24*, 976–983.

48. Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. Vis. Neurosci. *9*, 181–197.

49. Freifeld, L., Clark, D.A., Schnitzer, M.J., Horowitz, M.A., and Clandinin, T.R. (2013). GABAergic lateral interactions tune the early stages of visual processing in *Drosophila*. Neuron *78*, 1075–1089.

50. Freeman, T.C.B., Durand, S., Kiper, D.C., and Carandini, M. (2002). Suppression without inhibition in visual cortex. Neuron *35*, 759–771.

51. Li, B., Thompson, J.K., Duong, T., Peterson, M.R., and Freeman, R.D. (2006). Origins of cross-orientation suppression in the visual cortex. J. Neurophysiol. *96*, 1755–1764.

52. Sweeney, S.T., Broadie, K., Keane, J., Niemann, H., and O'Kane, C.J. (1995). Targeted expression of tetanus toxin light chain in *Drosophila* specifically eliminates synaptic transmission and causes behavioral defects. Neuron *14*, 341–351.

53. Shoemaker, P.A., O'Carroll, D.C., and Straw, A.D. (2005). Velocity constancy and models for wide-field visual motion detection in insects. Biol. Cybern. *93*, 275–287.

54. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.

55. Yamins, D.L.K., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.

56. Mcintosh, L., Maheswaranathan, N., Nayebi, A., Ganguli, S., and Baccus, S. (2016). Deep learning models of the retinal response to natural scenes. In Advances in Neural Information Processing Systems 29, D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, and R. Garnett, eds., pp. 1369–1377.

57. Haag, J., Mishra, A., and Borst, A. (2017). A common directional tuning mechanism of *Drosophila* motion-sensing neurons in the ON and in the OFF pathway. eLife *6*, e29044.

58. Wienecke, C.F.R., Leong, J.C.S., and Clandinin, T.R. (2018). Linear summation underlies direction selectivity in *Drosophila*. Neuron *99*, 680–688.e4.

59. Strother, J.A., Wu, S.-T., Rogers, E.M., Eliason, J.L.M., Wong, A.M., Nern, A., and Reiser, M.B. (2018). Behavioral state modulates the ON visual motion pathway of *Drosophila*. Proc. Natl. Acad. Sci. USA *115*, E102–E111.

60. Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. Z. Naturforsch., C, Biosci. *36*, 910–912.

61. Borst, A., Egelhaaf, M., and Haag, J. (1995). Mechanisms of dendritic integration underlying gain control in fly motion-sensitive interneurons. J. Comput. Neurosci. *2*, 5–18.

62. Weber, F., Machens, C.K., and Borst, A. (2010). Spatiotemporal response properties of optic-flow processing neurons. Neuron *67*, 629–642.

63. Suver, M.P., Mamiya, A., and Dickinson, M.H. (2012). Octopamine neurons mediate flight-induced modulation of visual processing in *Drosophila*. Curr. Biol. *22*, 2294–2302.

64. Tuthill, J.C., Nern, A., Rubin, G.M., and Reiser, M.B. (2014). Wide-field feedback neurons dynamically tune early visual processing. Neuron *82*, 887–895.

65. Jung, S.N., Borst, A., and Haag, J. (2011). Flight activity alters velocity tuning of fly motion-sensitive neurons. J. Neurosci. *31*, 9231–9237.

66. Maimon, G., Straw, A.D., and Dickinson, M.H. (2010). Active flight increases the gain of visual motion processing in *Drosophila*. Nat. Neurosci. *13*, 393–399.

67. Chiappe, M.E., Seelig, J.D., Reiser, M.B., and Jayaraman, V. (2010). Walking modulates speed sensitivity in *Drosophila* motion vision. Curr. Biol. *20*, 1470–1475.

68. Davis, F.P., Nern, A., Picard, S., Reiser, M.B., Rubin, G.M., Eddy, S.R., and Henry, G.L. (2018). A genetic, genomic, and computational resource for exploring neural circuit function. bioRxiv. https://doi.org/10.1101/385476v1.

69. Mauss, A.S., Meier, M., Serbe, E., and Borst, A. (2014). Optogenetic and pharmacologic dissection of feedforward inhibition in *Drosophila* motion vision. J. Neurosci. *34*, 2254–2263.

70. Joesch, M., Schnell, B., Raghu, S.V., Reiff, D.F., and Borst, A. (2010). ON and OFF pathways in *Drosophila* motion vision. Nature *468*, 300–304.

71. Nern, A., Pfeiffer, B.D., and Rubin, G.M. (2015). Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. Proc. Natl. Acad. Sci. USA *112*, E2967–E2976.

72. Fischbach, K.F., and Dittrich, A.P. (1989). The optic lobe of *Drosophila melanogaster*. I: A Golgi analysis of wild-type structure. Cell Tissue Res. *258*, 441–475.

73. Schwartz, O., and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. Nat. Neurosci. *4*, 819–825.

74. Mante, V., Frazor, R.A., Bonin, V., Geisler, W.S., and Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. Nat. Neurosci. *8*, 1690–1697.

75. Mante, V., Bonin, V., and Carandini, M. (2008). Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. Neuron *58*, 625–638.

76. Shoemaker, P.A., Hyslop, A.M., and Humbert, J.S. (2011). Optic flow estimation on trajectories generated by bio-inspired closed-loop flight. Biol. Cybern. *104*, 339–350.

77. Ammer, G., Leonhardt, A., Bahl, A., Dickson, B.J., and Borst, A. (2015). Functional specialization of neural input elements to the *Drosophila* ON motion detector. Curr. Biol. *25*, 2247–2253.

78. Meier, M., Serbe, E., Maisak, M.S., Haag, J., Dickson, B.J., and Borst, A. (2014). Neural circuit components of the *Drosophila* OFF motion vision pathway. Curr. Biol. *24*, 385–392.

79. Lucas, B.D., and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. Proceedings of the International Joint Conference on Artificial Intelligence *2*, 674–679.

80. Plett, J., Bahl, A., Buss, M., Kühnlenz, K., and Borst, A. (2012). Bio-inspired visual ego-rotation sensor for MAVs. Biol. Cybern. *106*, 51–63.

81. Bagheri, Z.M., Cazzolato, B.S., Grainger, S., O'Carroll, D.C., and Wiederman, S.D. (2017). An autonomous robot inspired by insect neurophysiology pursues moving features in natural environments. J. Neural Eng. *14*, 046030.

82. Olsen, S.R., Bhandawat, V., and Wilson, R.I. (2010). Divisive normalization in olfactory population codes. Neuron *66*, 287–299.

83. Rabinowitz, N.C., Willmore, B.D.B., Schnupp, J.W.H., and King, A.J. (2011). Contrast gain control in auditory cortex. Neuron *70*, 1178–1191.

84. Baccus, S.A., and Meister, M. (2002). Fast and slow contrast adaptation in retinal circuitry. Neuron *36*, 909–919.

85. Busse, L., Wade, A.R., and Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. Neuron *64*, 931–942.

86. Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. J. Neurosci. *17*, 8621–8644.

87. Tuthill, J.C., Nern, A., Holtz, S.L., Rubin, G.M., and Reiser, M.B. (2013). Contributions of the 12 neuron classes in the fly lamina to motion vision. Neuron *79*, 128–140.

88. Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online *2*, 13. https://doi.org/10.1186/1475-925X-2-13.

89. Meyer, H.G., Schwegmann, A., Lindemann, J.P., and Egelhaaf, M. (2014). Panoramic high dynamic range images in diverse environments. https://doi.org/10.4119/unibi/2689637.

90. Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401–415.

91. Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv, arXiv:1412.6980. https://arxiv.org/abs/1412.6980.

92. Richter, F.G., Fendl, S., Haag, J., Drews, M.S., and Borst, A. (2018). Glutamate signaling in the fly visual system. iScience *7*, 85–95.

93. Jia, H., Rochefort, N.L., Chen, X., and Konnerth, A. (2011). In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons. Nat. Protoc. *6*, 28–35.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Experimental data (behavior, electrophysiology, and calcium imaging)	This study	https://github.com/borstlab/normalization_paper
Experimental Models: Organisms/Strains		
D. melanogaster: WT: Canton S	N/A	N/A
D. melanogaster: L1-AD: w^{1118}; VT027316-AD; +	Courtesy of A. Nern / Janelia Research Campus	N/A
D. melanogaster: L1-DBD: w^{1118}; +; R40F12-DBD	Courtesy of A. Nern / Janelia Research Campus	RRID: BDSC_69935
D. melanogaster: L2-AD: w^{1118}; R53G02-AD; +	[87]	RRID: BDSC_68990
D. melanogaster: L2-DBD: w^{1118}; +; R29G11-DBD	[87]	RRID: BDSC_70173
D. melanogaster: L3-AD: w^{1118}; R59A05-AD; +	[87]	RRID: BDSC_70751
D. melanogaster: L3-DBD: w^{1118}; +; R75H07-DBD	[87]	RRID: BDSC_69459
D. melanogaster: L4-AD: w^{1118}; R20A03-AD; +	[87]	RRID: BDSC_68957
D. melanogaster: L4-DBD: w^{1118}; +; R31C06-DBD	[87]	RRID: BDSC_68978
D. melanogaster: L5-AD: w^{1118}; R21A05-AD; +	[87]	RRID: BDSC_70588
D. melanogaster: L5-DBD: w^{1118}; +; R31H09-DBD	[87]	RRID: BDSC_68980
D. melanogaster: Mi1-AD: w^{1118}; R19F01-AD; +	[12]	RRID: BDSC_68955
D. melanogaster: Mi1-DBD: w^{1118}; +; R71D01-DBD	[12]	RRID: BDSC_69066
D. melanogaster: Tm3-AD: w^{1118}; R13E12-AD; +	[12]	RRID: BDSC_68830
D. melanogaster: Tm3-DBD: w^{1118}; +; R59C10-DBD	[12]	RRID: BDSC_69153
D. melanogaster: Mi4-AD: w^{1118}; R48A07-AD; +	[12]	RRID: BDSC_71070
D. melanogaster: Mi4-DBD: w^{1118}; +; R13F11-DBD	[12]	RRID: BDSC_69722
D. melanogaster: Mi9-AD: w^{1118}; R48A07-AD; +	[12]	RRID: BDSC_71070
D. melanogaster: Mi9-DBD: w^{1118}; +; VT046779-DBD	[12]	RRID: BDSC_74714
D. melanogaster: Tm1-AD: w^{1118}; R41G07-AD; +	[68]	RRID: BDSC_71049
D. melanogaster: Tm1-DBD: w^{1118}; +; R74G01-DBD	[68]	RRID: BDSC_69767
D. melanogaster: Tm2: w^{1118}; +; VT012282	[11]	N/A
D. melanogaster: Tm2split-AD: w^{1118}; R28D05-AD; +	[68]	RRID: BDSC_68974
D. melanogaster: Tm2split-DBD: w^{1118}; +; R82F12-DBD	[68]	RRID: BDSC_69250
D. melanogaster: Tm4: w^{1118}; +; R35H01	[11]	RRID: BDSC_49922
D. melanogaster: Tm9: w^{1118}; +; VT065303	[11]	N/A
D. melanogaster: T4-AD: w^{1118}; VT016255-AD; +	Vienna Drosophila Resource Center	N/A
D. melanogaster: T4-DBD: w^{1118}; +; VT012314-DBD	Vienna Drosophila Resource Center	N/A
D. melanogaster: T5-AD: w^{1118}; VT013975-AD; +	Vienna Drosophila Resource Center	N/A
D. melanogaster: T5-DBD: w^{1118}; +; R42H07-DBD	Bloomington Drosophila Stock Center	RRID: BDSC_69609
D. melanogaster: T4/T5-AD: w^{1118}; R59E08-AD; +	[44]	RRID: BDSC_71101
D. melanogaster: T4/T5-DBD: w^{1118}; +; R42F06-DBD	[44]	RRID: BDSC_69285
D. melanogaster: w+; P{20XUAS-IVS-GCaMP6f}attP40; +	Bloomington Drosophila Stock Center	RRID: BDSC_42747
D. melanogaster: w+; +; PBac{20XUAS-IVS-GCaMP6f}VK00005	Bloomington Drosophila Stock Center	RRID: BDSC_52869
D. melanogaster: UAS-TNT: +; UAS-TNT; +	[52]	N/A
D. melanogaster: UAS-TNTin: +; UAS-IMPTNT-Q; +	[52]	N/A

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Custom-written software in Python	This study	https://github.com/borstlab/normalization_paper
ScanImage 3.8	[88]	http://scanimage.vidriotechnologies.com/display/SIH/ScanImage+Home
Other		
Natural images for experiments and modeling	[43]	N/A
Natural images for modeling	[89]	https://doi.org/10.4119/unibi/2689637

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Aljoscha Leonhardt (leonhardt@neuro.mpg.de). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Drosophila melanogaster were kept on a 12 h light/12 h dark cycle at 25°C and 60% humidity on standard cornmeal-agar medium. Genetic expression of effectors was targeted through the Gal4-UAS system [90]. Resulting genotypes and their abbreviations are listed in Table S2.

Unless stated otherwise, locomotion and tangential cell responses were recorded in wild-type Canton S flies 1 to 5 days after eclosion (Figures 1 and 2). We used the genetically encoded calcium indicator GCaMP6f [45] to determine the functional properties of individual cell types (Figures 3, 4, and 5). Throughout silencing experiments (Figure 5; Figure S2), we expressed tetanus toxin light chain (TNT) or an inactive version (TNTin) in the cell type of interest [52]. For calcium imaging experiments involving silencing (Figure 5), one day old flies were collected and put on 29°C for 3 days to boost expression of TNT or TNTin.

METHOD DETAILS

Natural image sets

For electrophysiology, behavioral, and modeling experiments, we used images from a published set of 20 natural panoramic scenes [43] termed dataset A. All images were independently processed as follows: We averaged across color channels and downsampled the scene to a resolution of 1,600 × 320 pixels (covering 360° sampled at 0.225 pixels per degree along the azimuth) using linear interpolation. To be able to render 12 bit images on conventional screens with 8 bits of dynamic range, we first performed standard gamma correction by raising raw pixel values to a power of 0.45 and then clipped the top percent of pixel intensities. The resulting image was scaled to fill the range between 0 and 255.

For optomotor experiments (Figure 1), we selected a subset of 8 images that covered different types of terrain. From this set, we again selected a subset of 6 images to determine tangential cell responses. We used all 20 images to build the convolutional network (Figure 6), randomly assigning 15 scenes to the training and 5 to the test set. Finally, we validated the trained convolutional model with images from an independent panoramic scene collection [89] consisting of 421 images (Figure S6G). These scenes were kept at their native resolution of 927 × 251 pixels (corresponding to an azimuthal sampling rate of 0.39 pixels per degree) and processed as above, yielding dataset B. We then generated two test sets: One had gamma correction applied to limit the images' bit depth ("low dynamic range" or LDR) and the other one was left at 12 bit depth to produce a dataset with high dynamic range (HDR).

Behavioral experiments

Experiments on the treadmill setup were conducted as described before [20, 31, 44]. Briefly, we tethered flies to a thin metal rod and placed them on air-cushioned polyurethane balls whose movement was tracked at 4 kHz, allowing for direct readout of rotational motion along all three axes. Temperature within the vicinity of the fly was 25°C at the start of each experiment. Using a closed-loop thermoregulation system, we linearly increased it to 34°C within 15 min to encourage locomotion.

For visual stimulation, we used three identically calibrated computer screens that were placed in a rectangle surrounding the fly. To simulate a cylindrical display, all stimuli were rendered onto a virtual cylinder and distorted accordingly before projection onto screens. Our setup covered approximately 270° in azimuth and 120° in elevation of the visual field. All stimuli were displayed at 144 Hz and at a spatial resolution greatly exceeding that of the fly eye. Screens had a maximum luminance of approximately 100 cd m^{-2} and a luminance depth of 8 bit; for all descriptions below, we assume pixel brightness to range from 0 to a maximum of 1. Patterns were generated in real-time and programmed in Python 2.7 using the game engine Panda3D.

We measured velocity tuning curves (Figure 1) for 8 distinct natural images at 6 logarithmically spaced velocities ranging from 5 to 1,280°s^{-1}. Initial image phase was randomized on each trial. Scenes were displayed at their native gamma-corrected mean luminance and contrast (see above). On each trial, images stood still for 1.5 s, then were rotated at the chosen velocity for 0.5 s, and remained fixed for another 1.5 s.

The optomotor contrast stimulus separated the visual field into two areas (see Figure 2A; Figure S2). For the so-called background, we tiled the visual field with pixels of size 5° x 5°. At each pixel location we drew a temporal frequency f from a normal distribution ($\mu = 0$ Hz, $\sigma = 1$ Hz) and a starting phase λ from a uniform distribution covering 0 to 360°. Instantaneous luminance of each pixel i was then determined by a random sinusoid of the form

$$I_i(t) = 0.5 + 0.5\, c_{bg} g(sin(2\pi\, f_i\, t + \lambda_i))$$

where the experimental parameter c_{bg} runs from 0 to 100% and controls the effective contrast of the background. To increase average contrast in the visual field, we applied the compressive transform

$$g(x) = x \sqrt{\frac{1 + \alpha^2}{1 + \alpha^2 x^2}}$$

where $\alpha = 5$ determined the degree of curve flattening. Using this method, we generated stochastic and dynamic visual input at a controllable contrast level without introducing coherent motion (see Figure 2B).

The so-called foreground delivered a coherent motion stimulus driving the optomotor response. It consisted of two vertical stripes that were placed at plus and minus 90° from the frontal axis of the fly, each spanning 20° in azimuth and the full screen elevation. We again tiled each stripe with pixels covering an area of approximately 5° x 5°. For each pixel i, luminance was fixed over time and determined by

$$I_i(t) = 0.5 + 0.5 c_{fg} g(sin(\lambda_i))$$

where the experimental parameter c_{fg} controls the effective motion contrast and λ was independently drawn from a uniform distribution covering 0 to 360°. The pixelated noise pattern smoothly wrapped around the azimuthal borders when moving. Note that for all instantiations of the stimulus, mean luminance across the visual field was 0.5. We verified that at typical scales of visual processing in *Drosophila* (approximated as a Gaussian filter with FWHM = 25° that covers a majority of the receptive fields of visual neurons; see [9]), variation in average luminance around this mean was small (Figure S2L).

For the basic contrast tuning experiment (Figures 2A–2E; see Video S1), we exhaustively measured combinations of logarithmically spaced values for c_{fg} (1.6, 3.1, 6.3, 12.5, 25, 50, and 100%) and c_{bg} (0, 25, 50, and 100%). At the beginning of each trial we simultaneously presented the dynamic background and the static foreground pattern. Between 1.5 and 2.0 s following stimulus onset, the foreground pattern moved at a fixed velocity of 50°s^{-1}. For oscillation experiments (Figures 2F–2J), the motion period was extended to 6 s. While the foreground pattern was moving, we sinusoidally modulated the contrast of either fore- or background between 0 and 100% around a mean value of 50% and at the specified temporal frequency (see Figure 2F; Figures S2A, S2C, and S2E; Video S2). When mapping the spatial extent of the contrast-induced modulation, we set the modulation frequency to 1 Hz and restricted the background pattern to two stripes of 10° width flanking each foreground pattern (see Figure S2A). The distance parameter (15, 17.5, 20, 22.5, 25, 27.5, 30, 35, or 40°) determined the separation between centers of foreground and background. In this experiment, we additionally measured a zero-contrast background condition to obtain an appropriate modulation baseline. Here, the motion stimulus had a contrast of 25% and luminance in the rest of the field was set to a uniform 0.5. Example traces in Figure 2G are taken from this spatial experiment (for distance 15° or no background). For the temporal experiments, we measured oscillation frequencies of 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, and 10 Hz (Figure 2J). Background contrast was zero when measuring foreground tuning; for background tuning, foreground contrast was set to 25%.

All stimulus patterns were displayed twice throughout optomotor experiments, once in clockwise and once in counterclockwise direction of motion. We recorded multiple trials to obtain robust turning responses for each fly (15 trials for natural image stimuli, 20 for contrast tuning, 25 for oscillation stimuli). Presentation order was shuffled across conditions within any trial to mitigate adaptation effects. Individual experiments lasted between 60 and 120 min.

Electrophysiology

Our patch-clamp recordings from tangential cells followed established protocols [11]. Cell bodies of horizontal system (HS) units were targeted visually through a microscope. We confirmed their preferred direction by stimulation with oriented moving sine wave gratings before each experiment.

Visual stimulation was delivered using a cylindrical projector-based arena as previously described [9]. Briefly, the screen of the arena covered a viewing angle of the fly of 180° in azimuth and 105° in elevation. Stimuli were generated at a framerate of 180 Hz using green light spanning approximately 500 nm to 600 nm in wavelength. The maximum luminance this arena achieved was 276 ± 48 cd m^{-2} (mean ± SD across devices). All visual stimuli were rendered using custom software written in Python 2.7 and the Panda3D framework. Membrane potential was recorded using custom software written in MATLAB (MathWorks, MA).

We measured tuning curves for 6 distinct natural image panoramas at 9 logarithmically spaced velocities ranging from 2.5 to 640°s^{-1} (Figure 1). On each presentation, the scene was displayed at a fixed phase, stayed still for 1 s, and then rotated horizontally for 3 s at the chosen constant velocity. Image movement was always in the preferred direction of the HS unit. We showed images at

their native gamma-corrected mean luminance and contrast (see above). Each condition was repeated 5 times. Conditions and trials were randomly interleaved to exclude adaptation effects along any stimulus dimension.

Calcium imaging

Calcium imaging experiments were performed using custom-built two-photon microscopes as described before [9]. The imaging acquisition rate was 11.8 Hz for all experiments, or 23.7 Hz for the experiment in Figure 4D, with imaging resolutions ranging from 32 × 32 to 64 × 128 pixels. Image acquisition was controlled using the ScanImage software (version 3.8) [88]. We prepared flies as previously described [9, 14]. Briefly, *Drosophila* were anesthetized on ice and glued onto an acrylic glass holder with the back of their head exposed to a perfusion chamber filled with Ringer's solution. Then the cuticula was surgically opened to allow optical access.

Stimuli were presented using the same projector system as in electrophysiological experiments, with additional long-pass filters (cut-off wavelength of 550 nm) in front of the projectors to spectrally separate visual stimulation from GCaMP fluorescence signals.

To identify receptive field (RF) positions of individual neurons, white noise stimuli of 3 min length were used (except for T4 and T5 cells, see below). The stimuli were pre-rendered at 60 Hz and generated as previously described [9]. Briefly, the spatial resolution of all white noise stimuli was 2.8° of visual angle corresponding to 64 pixels across the 180° screen. For all lamina cells, the same stimulus was used in order to provide a systematic description of their spatiotemporal filtering properties (Figure S3). This stimulus had a Gaussian autocorrelation with a standard deviation of approximately 45 ms in time and a contrast of 25% around a mean intensity value of 50 on an 8 bit grayscale. For some medulla cell types, variants of this stimulus with higher contrast or longer time constants were used if necessary to reliably locate their RFs on the arena. Specifically, we mapped RFs for Tm4, Mi4, Mi9 and Tm9 with a binary stimulus at 100% contrast and a temporal cut-off frequency of 1 Hz. For Mi9, we chose a 1D version of this stimulus, consisting of horizontal (1.5 min) and vertical bars (1.5 min) instead of pixels.

For T4 and T5, we relied on a novel stochastic motion noise stimulus to determine RF coordinates. First, we determined the preferred direction of an ROI using drifting gratings. Then we displayed a stimulus consisting of 20 randomly distributed 15° wide circular windows. Inside of each window, a 30° wavelength sine grating drifted at 30°s⁻¹ in the preferred direction (Figure S3X). The positions of these 20 windows were changed and randomly chosen every second over 4 min. Reverse correlation of T4 and T5 responses with the area covered by those windows at a given time point yielded motion-sensitive RFs which were fit with a Gaussian to determine center coordinates (Figure S3Y). These were verified by presenting 25° windows containing full contrast drifting gratings at the estimated RF center and 6 hexagonally distributed positions around the center. Cells responded only to the grating in the RF center (Figure S3Z).

For the experiments shown in Figure 3, a 25° circular window around the RF center of a cell defined the foreground whereas the rest of the screen was defined as background. Before stimulus presentation, we verified that RF centers were sufficiently distant from the border of the screen to allow full display of the foreground. A drifting sine grating with 30° wavelength and a velocity of 30°s⁻¹ was shown, starting with medium gray at the center of the RF and moving for 4 s after stimulus onset (see Video S3). The contrast of the grating was varied independently between background and foreground. A stimulus matrix of 7 foreground contrasts (1.6, 4, 8, 16, 32, 64 and 100%) and 6 background contrasts (0, 8, 16, 32, 64 and 100%) at a constant mean luminance of 0.5 was presented.

For the experiments shown in Figures 4A–4C, the foreground contrast was chosen depending on the cell type as the point where the suppression elicited by 100% background contrast (as measured in Figure 3) would be greatest. This was 16% for Mi1, 32% for Tm1, 100% for Tm2 and 64% for Tm3. The background had 100% contrast and 30° wavelength. We varied either its direction, its velocity (0, 0.25, 0.5, 1, 2, 4, 8, 16, 32 or 64°s⁻¹), or restricted its presentation to an annulus with changing outer diameter. A reference condition with 0% background contrast was added to the stimulus protocol.

For the contrast-step stimulus experiments shown in Figure 4D (see Video S4), the background grating had 30° spatial wavelength, drifted with 90°s⁻¹ after motion onset and its initial phase was randomized. For Tm2 it had full contrast, for Tm3 44% contrast. The 25° foreground window was 50% gray and we placed a 5° wide dot in the center. For Tm3, the dot was initially black and set to white for a duration of 50 ms at a given time interval after motion onset of the background grating. For Tm2, the dot was initially white and then set to black. The time interval was varied in steps of 50 ms from –250 ms to 500 ms and then in steps of 100 ms. Negative values indicate that the surround grating started to move after the dot changed its intensity. Additional time intervals were –500 ms and –1 s. The block experiments in Figure 5 were performed with the same frequency tuning stimuli as before (Figure 4B). For the contrast tunings, the same stimuli as in Figure 3 were used but with background contrast of either 0 or 100% only.

All stimuli were repeated three times in randomized condition order to prevent adaptation to any stimulus features.

Modeling

Natural motion stimuli

To evaluate the performance of our models under naturalistic conditions, we generated a synthetic set of motion sequences that closely mimicked the experimental stimuli described above. For each sequence we translated 360° images at a fixed horizontal velocity through a virtual window spanning 100° in azimuth. Given their panoramic nature, scenes wrapped around seamlessly at each border. Movies were generated at a time resolution of 100 Hz. To reduce jitter for small velocities, we linearly interpolated non-integer pixel shifts. Fly eye optics were simulated ahead of time. We blurred each frame with a Gaussian filter (full width at half-maximum of 4°) to approximate the acceptance angle of each photoreceptor [26] and then sampled individual signals from a rectangular grid with isotropic spacing of 4° (yielding 23 × 17 receptor signals per frame for dataset A and 23 × 23 for dataset B, as described above).

For the comparison in Figure 1, we modeled the exact stimulus parameters of the electrophysiological experiment including an approximation of the image's starting phase on the arena. We generated sequences for our convolutional detector models (Figure 6) as follows: The set of 20 panoramic images was randomly split into a training group consisting 15 scenes and a test group consisting of 5 scenes. For each sequence, a random image was drawn from the appropriate set. The stimulus lasted 5 s. Between 1 and 4 s, scene velocity stepped from zero to a fixed value drawn from a Gaussian distribution with SD = 100°s^{-1}. The initial window phase followed a uniform distribution spanning 360°. To further augment the dataset, we flipped the underlying image along the horizontal and vertical axes with a probability of 50%. We generated 8,192 such sequences for the training set and 512 for the test set.

Experimental stimuli

For all modeling experiments in Figure S5, we replicated the experimental protocols described above as precisely as feasible. All stimuli were projected onto a field of view that spanned 120° in azimuth and 90° in elevation at a spatial resolution of 1° for calcium imaging experiments and 0.5° for behavioral experiments. Frames were then blurred and sampled as described for natural image stimuli. Brightness values for all stimuli ran from 0 to 1 and we fixed the mean level for contrast stimuli at 0.5. For calcium imaging stimuli, we always placed the foreground disk at the center of the field of view. Patterns were rendered and processed at 100 Hz.

Tuning curves for the basic contrast experiment (Figures S5B–S5D), the frequency experiment (Figure S5F), and the background diameter experiment (Figure S5G) were estimated from a single trial per parameter setting. For the background orientation experiment (Figure S5E) and the step interval experiment (Figure S5H) we averaged 100 trials with randomized background pattern phases to approximate the experimental phase stochasticity that results from individual cell receptive fields being located in different parts of the visual field. We averaged 200 trials for the behavioral stimuli (Figure S5K) to account for the intrinsic stochasticity of the stimulus and to generate reliable model responses. Throughout Figure S5, we calculated point estimates for all tuning curves exactly as described for the behavioral and calcium data.

Tuning curve normalization model

The analytical model for divisive normalization (Figures 3V–3X) resembles previous formulations in the literature [48, 50, 86]. The steady-state response R of a neuron is given by

$$R(c_{fg}, c_{bg}) = \frac{L_{fg}c_{fg}^p + L_{bg}c_{bg}^p}{c_{50}^p + c_{fg}^p + S^p}$$

where c_{fg} and c_{bg} are foreground and background contrast and L_{fg} and L_{bg} are weight factors defining the respective amount of linear contribution of foreground and background to the response. The semi-saturation constant c_{50} determines the contrast at which the cell responds with 50% strength and the parameter p defines the steepness of the saturation curve.

The normalization term

$$S = w_{pool} \cdot c_{bg}^q$$

gives the amount of divisive surround suppression which is proportional to background contrast to a power of q, which accounts for possible non-linear scaling behavior, with a proportionality weight constant w_{pool}. In this model, the normalization index w_{pool}/c_{50} quantifies how much the sigmoidal tuning curve shifts to the right when c_{bg} is increased from 0 to 1 (full contrast), in relation to the semi-saturation constant. It thus describes the fold decrease in contrast sensitivity between no background contrast and full background contrast.

For evaluation of the normalization index (Figure 3X), this model was fit individually for each cell. Parameter fits to the average tuning curve per cell type are listed in Table S1. Since tuning curves from individual cells are subject to measuring inaccuracies, we cross-validated fit quality. We optimized model parameters for the average tuning curve of 50% of all measured cells per type and evaluated variance explained for the other 50%. This was repeated 100 times with shuffled training and validation sets. For all cell types, cross-validated variance explained was more than 90% (see $R^2_{DivisiveNorm}$ in Table S1). When we repeated this procedure with a fully linear model

$$R(c_{fg}, c_{bg}) = L_{fg}c_{fg} + L_{bg}c_{bg}$$

variance explained dropped substantially for all units except L3 (see R^2_{linear} in Table S1).

This analysis was implemented using Python 2.7 and NumPy 1.11.3. Optimization of model parameters was performed using the L-BFGS-B algorithm in SciPy 0.19.0.

Data-driven detector model

The reference model in Figure 1 was based on a standard implementation of the Reichardt-type correlational motion detector [26]. Briefly, all receptor signals of the two-dimensional input grid (see above) were filtered with a first-order high-pass (τ = 150 ms). We then multiplied each local signal with the delayed horizontal neighbor (first-order low-pass, τ = 50 ms). This was done twice in a mirror-symmetrical fashion and resulting output was subtracted. Finally, we summed across all local detectors to derive a model of tangential cell output. For the illustration in Figure S1C, we simulated the receptor array at the full image resolution without blurring. These models were implemented in Python 3.6 using PyTorch 0.4.1.

We simulated time-resolved cell models for three basic response types: a purely linear low-pass unit (modeled after L3; Figure S5B), a strongly normalized band-pass unit (modeled after Mi1; Figure S5C), and a weakly normalized low-pass unit (modeled after Mi9; Figure S5D). We hand-tuned parameters based on our and previous work [9] to qualitatively match response properties

of the corresponding cell. Models were implemented as signal processing cascades (see Figure S5A). First, signals at each location in the field of view were filtered with a spatial difference of Gaussians kernel that had a central full-width at half-maximum (FWHM) of $6°$ and a FWHM of $20°$ in the surround. In accordance with results from receptive field mapping (Figure S3), the weight ratio between surround and center was 100% for low-pass units and 50% for the band-pass model. Full-field flashes would thus produce no activation in low-pass units. This was followed by first-order temporal filters: a single low-pass filter for low-pass units ($\tau = 80$ ms) or serial low- ($\tau = 50$ ms) and high-pass filters ($\tau = 150$ ms) for band-pass units. We then left the signal as is for ON cells or sign-inverted it for OFF cells and half-wave rectified the output by setting all negative values to zero.

For normalized cell models, we calculated local input P_i from the normalization field by pooling across rectified signals x_i with a Gaussian kernel (FWHM = $30°$). Final output was then calculated using the divisive normalization equation

$$f(x_i) = \frac{x_i^p}{c_{50}^p + x_i^p + (w_{pool}P_i)^p}$$

where i indexes across points in space and time, c_{50} determines baseline sensitivity, exponent p regulates the static response non-linearity, and w_{pool} adjusts sensitivity to the normalization field signal. We manually tuned normalization parameters for the band-pass ($c_{50} = 0.012$, $p = 1.3$, $w_{pool} = 1.5$) and the low-pass cell ($c_{50} = 0.12$, $p = 1.1$, $w_{pool} = 3.0$) to match critical features of the empirical contrast tuning curves (Figures S5C and S5D).

To generate simulated T4 responses (Figures S5I and S5J), we multiplied the output of spatially adjacent low- and band-pass units. For the linear reference model we bypassed the final normalization step in both arms of the detector. We built the LPTC model (Figure S5K) as a spatial array of T4 and T5 cells covering the full field of view, analogously to the previously described two-quadrant detector [28]. For the T5 model, we used two OFF-sensitive input units with identical parameters as for ON cells. Output from syndirectionally tuned T4 and T5 motion detectors was summed and subtracted from a mirror-symmetric, oppositely tuned array to produce LPTC model output. The same model was used to simulate natural scene responses (Figures S5L–S5N). All models in Figure S5 were implemented using Python 3.6 and NumPy 1.15.

To quantify the robustness of velocity tuning for models and LPTCs (Figure S5N), we calculated per-velocity coefficients of variation as the ratio between response standard deviation across images and response mean across images. For neural data, we used cell-averaged mean potential to estimate these parameters.

Task-driven detector model

We implemented the trained detector model as a four-layer convolutional neural network consisting of linear input filters, a normalization stage, local multiplication, and linear spatial summation. In contrast to typical deep architectures used for object recognition, this network processed three-dimensional inputs spanning two dimensions of space as well as time.

First, receptor signals of shape $23 \times 17 \times 500$ or $23 \times 23 \times 500$ (azimuth, elevation, time), depending on the dataset, were processed in two independent convolutional channels. The convolutions were temporally causal and spatiotemporally separable. Each of the channels was composed of a $3 \times 3 \times 1$ spatial filter (covering 3 simulated receptors in azimuth and elevation) followed by a temporal filter of shape $1 \times 1 \times 30$ (corresponding to 300 ms at the chosen time resolution of 100 Hz). Convolutions had no bias parameter. In contrast to standard Reichardt detectors, each filter weight was allowed to vary freely during optimization.

Second, we passed local output signals x_i (where i indexes points in space and time) through one of three types of local normalization: a simple pass-through (termed "linear")

$$f(x_i) = x_i$$

a static and contrast-independent compression stage (termed "static")

$$f(x_i) = tanh\left(\frac{x_i}{c}\right)$$

where the trained parameter c determines the sensitivity of the saturating function, or an adaptive saturation stage (termed "dynamic")

$$f(x_i) = tanh\left(\frac{x_i}{c + P_i}\right)$$

where c again determines the baseline sensitivity and P_i is the instantaneous output of a $11 \times 11 \times 1$ spatial filter (centered on the location of x_i and operating on full-wave rectified output signals $|x_i|$; see Figure 6A). This models the fast and spatially distributed normalization we observed during experiments. We chose the hyperbolic tangent because it generalizes to positive and negative input values, the transformation closely resembles the normalization model described above, and it is more commonly used in the field of deep learning. Spatiotemporal filters were optimized independently for each of the two channels while the sensitivity parameter c was shared.

Third, we then combined signals from both channels in a EMD-type scheme where adjacent signals were multiplied and output from two mirror-symmetric pairs was subtracted. This stage was parameter-free. Finally, resulting signals were summed across space and multiplied by a trained scalar amplification factor to generate the final time-resolved output of the model. The base model without normalization had 79 trainable parameters; static normalization added one parameter and dynamic normalization another 242.

We trained each model architecture to estimate the true velocity of translation stimuli using automatic differentiation, backpropagation, and stochastic gradient descent. The loss function we applied was the mean squared error (MSE) between model output and current velocity of the scene. Weights were updated using the Adam optimizer [91], with parameters set to standard values ($\beta_1 = 0.9$, $\beta_1 = 0.999$, $\varepsilon = 10^{-8}$). Models were trained over 800 epochs with a batch size of 128; no early stopping was used. We set the initial learning rate to 0.025 and divided it by a factor of 4 after 400, 500, and 600 steps. Input convolutional layers were initialized to random values drawn from a uniform distribution. For the pooling receptive field, we initialized each weight with 0.0001 and the sensitivity factor c with 1.0. Static sensitivity as well as pooling weights were constrained to be positive. In the dynamic normalization model, we applied a L_2 penalty of 400.0 to the spatial weights of the pooling stage. Hyperparameters were determined in preliminary experiments with an independent image set. We optimized each architecture 16 to 23 times with different random number generator seeds to assess reliability and did not select models post hoc.

We implemented all architectures in Python 3.6 using PyTorch 0.4.1 for automatic differentiation. Depending on model type, a single optimization run took between 6 and 14 hs on an NVIDIA Titan Xp GPU.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data evaluation for behavioral experiments

To ensure data quality, we excluded all flies whose average forward velocity during the experiment was below 0.25 cm s^{-1} and whose average turning tendency was either slowly drifting or far from 0°s^{-1}. Fewer than 20% of all experiments failed these criteria. Measurements of ball movement were downsampled via linear interpolation for further processing (to 50 Hz for natural image stimuli, Figure 1; 20 Hz for contrast tuning, Figure 2; 100 Hz for oscillation stimuli, Figure 2). Trials were averaged.

Responses for clockwise and counterclockwise motion were subtracted and divided by two to minimize residual deviations from straight forward walking. Traces for natural image and contrast tuning stimuli were filtered using a first-order low-pass with a time constant of 100 ms. For the contrast oscillation experiments, we evaluated modulation at the relevant carrier frequency by calculating the zero-padded Fourier Transform of the turning trace and averaging the amplitude spectrum in a window of width 0.2 Hz centered on the target frequency. These values were normalized per experiment such that the modulation peak after averaging was 100%. We applied a Savitzky-Golay filter (window length 11 samples, 5th order polynomial) before plotting traces from oscillation experiments; this did not affect the analysis.

All analysis for behavioral experiments was performed in custom-written software using Python 3.6, NumPy 1.15, and SciPy 1.1.

Data evaluation for electrophysiological experiments

Voltage data were digitized at 1,000 Hz. To account for slow drift in potential, we subtracted the average voltage in a 1 s window before stimulus onset from each trace per stimulus condition and trial. Signals were then low-pass filtered (8th order Chebyshev Type 1) and resampled at 100 Hz. Finally, we averaged cell responses across trials. Cells whose mean depolarization during full-contrast sine grating presentation in preferred direction remained below 5 mV were discarded before further analysis. All analysis for electrophysiological experiments was performed in custom-written software using Python 3.6, NumPy 1.15, and SciPy 1.1.

Data evaluation for calcium imaging experiments

Calcium imaging stacks were registered in order to correct for translational movement artifacts of brain tissue using custom-written software. Responses of individual neurons were extracted by manually selecting small regions of interest (ROI) encompassing individual anatomical structures. For T4 and T5 these corresponded to single or few axon terminals; for Mi and Tm cells, individual axon terminals could be identified clearly through visual inspection. For ON pathway medulla cells, signals were measured in layer 10 of the medulla, for OFF pathway medulla cells in layer 1 of the lobula. For lamina cells L1–5, signals were measured at axon terminals in corresponding layers 1–5 in the medulla. For T4 and T5, signals were recorded in the lobula plate.

To reconstruct RFs, calcium signals were mean subtracted and reverse-correlated with the stimulus as previously described [9]. 1D Gaussians were fit to horizontal and vertical cross-sections of spatial receptive fields to obtain precise RF coordinates. For lamina cells (Figure S3), all reconstructed RFs were peak-aligned and analyzed as previously [9]. For 1D projections of spatial RFs (Figures S3F–S3J), an average of 1D projections of 2D RFs along 3600 evenly distributed projection angles between 0° and 360° was calculated. This enhanced the visibility of the center-surround structure but neglected possible anisotropies in the spatial structure of RFs [49]. For impulse responses (Figures S3K–S3O) the temporal receptive field of the 9 center pixels was averaged; frequency responses (Figures S3P–S3T) are the Fourier-transformed impulse responses. Deconvolution (Figures S3U and S3V) was performed by dividing the frequency spectra with the frequency response of a 1st order low-pass filter with time-constant 350 ms as a proxy for calcium indicator dynamics [9, 92].

Relative fluorescence changes ($\Delta F/F$) from raw calcium traces were obtained by adapting an automatic baseline detection algorithm [93]. Briefly, raw data were first smoothed with a Gaussian window (full-width at half maximum, FWHM = 1 s). Then, minima within a 90 s long sliding window were extracted and the resulting trace smoothed with a Gaussian window (FWHM = 4 min). The result was used as a dynamic baseline F_0 and $\Delta F/F$ values were computed as $\Delta F/F = (F–F_0)/F_0$.

For further evaluation, only recordings with good signal-to-noise ratio (SNR) were taken. The criterion was that the standard deviation of the mean signal averaged over trials had to be at least 120% of the mean standard deviation over trials. This criterion filtered out cells with an inter-trial variance larger than the typical cell response (caused by movement artifacts or photobleaching).

In addition, the standard deviation of the mean signal had to be larger than 25% ΔF/F. On average, 90% of all cells measured passed these criteria with slight variations due to different levels of GCaMP expression depending on the genotype.

For experiments with drifting gratings, the driving foreground contrast frequency was 1 Hz. For these experiments, we evaluated the amplitude of the 1 Hz component of the signal. This was achieved by computing the Fourier coefficient at that frequency, using the equation

$$F = \left| \frac{1}{T} \int_0^T dt \, s(t) \, e^{-2\pi i \cdot 1Hz \cdot t} \right|$$

where $s(t)$ denotes the signal and T the stimulation time. For experiments in Figure 4D, we evaluated the peak response of the calcium signal. For Figure S4, we additionally evaluated the average calcium signal (F0) during stimulus presentation and normalized it to the maximum amplitude of the 1 Hz component (F1).

Amplitudes were averaged over trials and normalized to the maximum, then averaged over cells and normalized to the maximum. For Figures 4 and 5, amplitudes were normalized to the response amplitude for the reference stimulus.

Statistical tests

Unless indicated otherwise, error bars show bootstrapped 68% confidence intervals around the mean (estimated as corresponding distribution percentiles after resampling the data 1,000 times). All statistical tests were two-tailed and performed at a 5% significance level. Normality of data distributions was assessed visually but not tested formally. Sample sizes are given in each figure legend and were not based on power analysis but predetermined in line with standards in the field. We did not blind experimenters to genotypes or conditions during data gathering and analysis.

DATA AND CODE AVAILABILITY

Code and experimental data are available on GitHub (https://github.com/borstlab/normalization_paper).