Visual and motor signatures of locomotion dynamically shape a population code for visual features in *Drosophila*.

Maxwell H. Turner[1], Avery Krieger[1], Michelle M. Pang[1] and Thomas R. Clandinin[1,⋆]

[1] Department of Neurobiology, Stanford University, Stanford, CA 94103, USA
⋆ Corresponding author

Summary

Natural vision is dynamic: as an animal moves, its visual input changes dramatically. How can the visual system reliably extract local features from an input dominated by self-generated signals? In *Drosophila*, diverse local visual features are represented by a group of projection neurons with distinct tuning properties. Here we describe a connectome-based volumetric imaging strategy to measure visually evoked neural activity across this population. We show that local visual features are jointly represented across the population, and that a shared gain factor improves trial-to-trial coding fidelity. A subset of these neurons, tuned to small objects, is modulated by two independent signals associated with self-movement, a motor-related signal and a visual motion signal. These two inputs adjust the sensitivity of these feature detectors across the locomotor cycle, selectively reducing their gain during saccades and restoring it during intersaccadic intervals. This work reveals a strategy for reliable feature detection during locomotion.

Introduction

Sighted animals frequently move their bodies, heads and eyes to achieve their behavioral goals and to actively sample the environment. As a result, the image on the retina is frequently subject to self-generated motion. This presents a challenge for the visual system, as visual circuitry must extract and represent specific features of the external visual scene in a rapidly changing context where the dominant sources of visual changes on the retina may be self-generated. While this problem has been well studied in the context of motion estimation (Borst et al., 2010; Britten, 2008), the broader question of how visual neurons might extract local features of the scene under naturalistic viewing conditions is relatively poorly understood. How do visual neurons selectively encode local features of interest under these dynamic conditions?

Local feature detection during self motion presents unique challenges. For detecting widefield motion, or large static features of the scene like oriented edges and landmarks, the visual scene is intrinsically redundant, as many neurons distributed across the visual field can encode information that is relevant to the feature of interest even as the scene moves. Conversely, local features like prey, conspecifics, or approaching predators engage only a small part of the visual field, dramatically reducing the redundancy of the visual input. In addition, neurons that selectively respond to small

features could also be activated by high spatial frequency content in the broader scene, potentially corrupting their responses under naturalistic viewing conditions. Neurons that respond selectively to local visual features have been described in many species, including flies, amphibians, rodents, and primates (Keleş and Frye, 2017; Kerschensteiner, 2022; Klapoetke et al., 2022; Lettvin et al., 1959; Pasupathy and Connor, 2001; Piscopo et al., 2013). However, these studies have typically been conducted either in non-behaving animals, or under conditions of visual fixation. Here we explore the neural mechanisms by which local feature detection is made robust to the visual inputs and behavioral signals associated with natural vision.

There are at least four common strategies used to mitigate the impact of self motion on visual feature encoding. First, animals across phyla perform compensatory movements of the eyes, head or body to stabilize the image on the retina during self motion (Angelaki and Hess, 2005; Hardcastle and Krapp, 2016; Land, 1999; Walls, 1962). Second, many animals sample the visual world in a saccadic fashion - that is, the largest changes in gaze (called saccades) occur during compressed periods of time, and between saccades, gaze location is relatively stable (Land, 2019). In primates, saccades are rapid eye movements that occur several times a second and can shift the center of gaze by up to tens of degrees of visual angle (Martinez-Conde et al., 2013; Van Der Linde et al., 2009; Wurtz, 2018). The fruit fly, *Drosophila*, performs "body saccades" which are rapid whole-body rotations during locomotor behavior (Cruz et al., 2021; Geurten et al., 2014). In either case, self-generated movements are temporally discrete, meaning that visual inputs are only transiently corrupted by large self motion signals. The third strategy for reliable vision during self motion uses signals related to the animals' motor commands or behavioral state to modulate neural response gain. For example, the motor commands that initiate primate saccades produce efference copy signals that are associated with neural gain changes and a perceptual decrease in sensitivity called saccadic suppression (Binda and Morrone, 2018; Bremmer et al., 2009; Wurtz, 2018). In flies, efference copy signals can cancel expected motion in widefield motion sensitive neurons during flight (Fenk et al., 2021; Kim et al., 2015), but can also provide independent information about intended movements (Fujiwara et al., 2017; Fujiwara et al., 2022; Cruz et al., 2021). In this way, neural response gain is modulated so that motion sensitive neurons encode unexpected deviations in motion signals after accounting for behavior. Fourth, and finally, some visual neurons incorporate information about global motion signals via a widefield, motion sensitive surround (Aptekar et al., 2015; Baccus et al., 2008; Ölveczky et al., 2003). However, how these self-motion strategies interact, and how they might impact local feature detection is unknown.

We explore this issue using populations of visual projection neurons (VPNs) in *Drosophila*. VPNs are situated at a critical computational and anatomical bottleneck through which highly processed visual information moves from the optic lobes to the central brain. A subset of VPNs, the Lobula Columnar (LC) and Lobula Plate Lobula Columnar (LPLC) cells (Fischbach and Dittrich, 1989; Otsuna and Ito, 2006; Wu et al., 2016) make up a large fraction of all VPN types, thus accounting for a substantial portion of the visual information available to guide behavior. These cell types encode distinct local visual features with behavioral relevance, including looming objects (Ache et al., 2019; Klapoetke et al., 2017) and small moving objects (Keleş and Frye, 2017; Ribeiro et al., 2018) (for a recent survey of VPN visual tuning, see Klapoetke et al., 2022), and project to small, distinct regions in the central brain called optic glomeruli (Wu et al., 2016). Previous work has also implicated some types of LCs in figure-ground discrimination, i.e. the ability to detect an object moving independently of a global background motion signal (Aptekar et al., 2015). Each optic glomerulus receives input from all of the individual cells belonging to a single cell type, resulting

in a functional map in the central brain (Klapoetke et al., 2022). Moreover, both stimulation and silencing experiments argue that at least some VPN classes strongly modulate specific visually-guided behaviors (Hindmarsh Sten et al., 2021; Tanaka and Clark, 2020; Tanaka and Clark, 2022). Finally, the visual tuning of VPN types is heterogeneous across the population, allowing us to explore how strategies for reliable visual encoding during self motion vary across differently tuned populations.

To explore how local visual features are represented across populations of VPNs, we developed a new method to register functional imaging data to the fruit fly connectome, allowing us to measure neural responses across many optic glomeruli simultaneously. We show that this method allowed for reliable and repeatable measurement of VPN responses. This population imaging method allowed us to measure the covariance of optic glomerulus population responses to visual stimuli. This analysis revealed strongly correlated trial-to-trial variability across glomeruli, which improves stimulus encoding fidelity. Importantly, this could not have been inferred from non-simultaneous measurements. We next demonstrate that walking behavior selectively suppressed responses of small object detecting glomeruli, leaving responses to looming objects unchanged. We then show that visual stimuli characteristic of self motion, including those produced by locomotor saccades, also suppressed VPN responses to small objects. Finally, we show that these two forms of gain control - visual and motor-associated - can be independently recruited and reinforce one another when both are active. Together these results reveal that both visual and motor cues associated with self motion can tune local feature detecting VPNs, adjusting their sensitivity to match the dynamics of natural walking behavior. This suggests a strategy for resolving the ambiguities associated with detecting external object motion in a scene dominated by self-generated visual motion.

Results

Global visual motion complicates local feature detection

To build intuition about how self-generated motion might impact local feature selectivity, we designed a task inspired by VPN selectivity to small, moving objects (Keleş and Frye, 2017; Klapoetke et al., 2022). In this detection task, a 15°dark patch moved on top of a grayscale natural image background, through a receptive field whose size was typical of small object detecting LCs (Fig1A). When the natural image background was static, as would be the case if a stationary fly were observing an external moving object in a rich visual environment, detecting the moving patch is trivial given the change in local luminance and/or spatial contrast as the patch traverses the receptive field (Fig1B). How is this detection task impacted by self motion? We simulated self-generated rotational motion by moving the background image at a single, constant velocity (Fig1C). This background motion caused large fluctuations in local luminance and spatial contrast, reflecting the heterogeneous spatial structure of the scene (Fig1D, red traces). These fluctuations were often larger than the changes induced by the moving patch alone (compare Fig1B to Fig1D, for example). Moreover, with an independently-moving patch added to the foreground, the change in local luminance or contrast was negligible for this example image (Fig1D, blue traces), making discrimination between these two conditions very difficult.

We quantified discriminability, d', between traces where only the background image moved and traces where the small patch moved on top of the moving background, using either local luminance

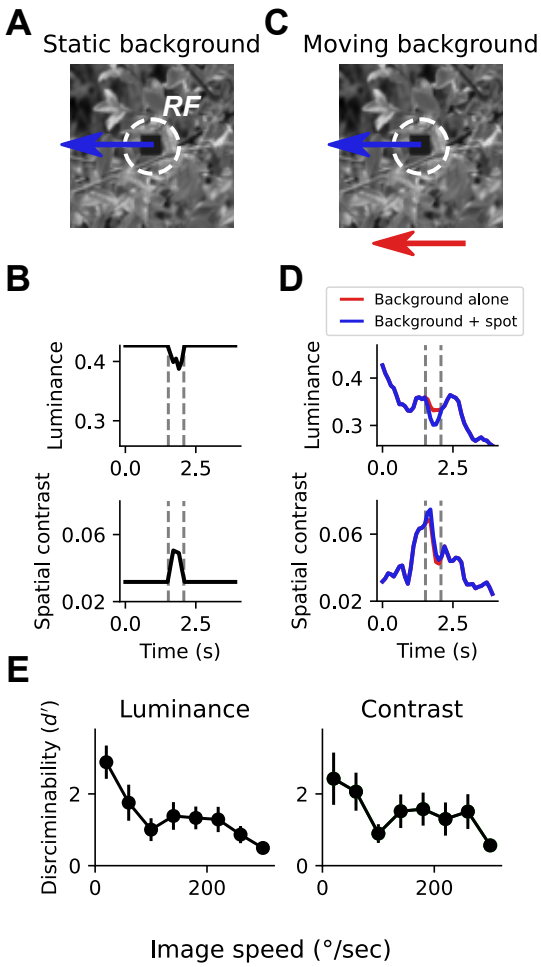

Figure 1: Small spot detection is unreliable during self motion (A) Schematic illustrating the stimulus and detection task. A small dark spot on top of a grayscale natural image moved through a visual receptive field (white dashed circle). (B) When the background image was stationary, small spot detection was trivial using either local luminance (top) or spatial contrast (bottom) cues. Vertical dashed lines indicate the window of time that the spot passes through the receptive field. (C) To mimic an object detection task during self motion, we moved the background image independently of the spot with a variable speed. (D) Movement of the background image alone (red trace) caused dramatic fluctuations in luminance and contrast within the receptive field. The addition of the small moving spot (blue trace) caused relatively small changes in the luminance or contrast signal, which depends on the spatial structure of the image. (E) Discriminability of the spot based on luminance (left) or contrast (right) during the time period when the spot passed through the receptive field, as a function of background image speed. Points indicate mean +/- S.E.M. across a collection of 20 grayscale natural images. Even for slow background speeds, detection was corrupted, and the discriminability of the spot decreased further as the background speed increased.

signals (Fig1E, left) or local contrast signals (Fig1E, right). This metric captures the difference between the mean responses to "spot present" vs. "spot absent" normalized by the standard

deviation of the response traces (see Methods). d' reflects the z-scored difference between the responses to these two conditions, meaning that a d' of 0 corresponds to chance under an ideal observer model. With a static or absent background, the discriminability of the patch is perfect. Across a collection of 20 natural images (van Hateren and van der Schaaf, 1998), moving at velocities between 20°/sec and 320°/sec, small object detection was corrupted even for small amounts of background motion, and discriminability decreased further as background motion increased (Fig1E). These observations suggest that as self-motion signals increase, neurons that respond selectively to local features like small moving objects might increase their response thresholds in order to avoid relaying false positive signals.

A connectome-based alignment method to measure population activity across optic glomeruli

In order to efficiently characterize the responses of individual VPNs to many visual stimuli, and to relate the gain of multiple VPNs with one another and to animal behavior, we needed to measure responses across different VPN types simultaneously. Presently, specific driver lines exist to target single VPN types in a single experiment (Wu et al., 2016), but no approach exists to measure across many VPN types simultaneously. To develop such a population recording approach, we exploited the fact that optic glomeruli are physically non-overlapping (Fig. 2A). Each optic glomerulus receives dominant input from one type of LC or LPLC cell (with one known exception being LPLC4/LC22 (Wu et al., 2016), not included in this study). At the same time, the fly brain is highly stereotyped, meaning that by aligning functional imaging data to the Drosophila connectome (Scheffer et al., 2020) we could use the positions of VPN presynaptic active zones (T-bars) to identify voxels that correspond to specific glomeruli.

We selected the optic glomeruli in the Posterior Ventrolateral Protocerebrum (PVLP) and Posterior Lateral Protocerebrum (PLP) for imaging (Fig. 2A), because this region of the brain contains the majority of known optic glomeruli in a confined volume. We imaged the left PVLP/PLP using a two-photon resonant scanning microscope, which allowed for sampling of the volume of interest at ∼7 Hz (Fig. 2B, see Methods). As previous work had demonstrated that individual VPN cells respond to visual stimuli with monophasic calcium responses that span several hundred milliseconds (as measured using GCaMP6F (Klapoetke et al., 2022)), this volume rate provides dense temporal sampling of each VPN type.

Optic glomeruli contain neurites from many neuron types, including the presynaptic terminals of their dominant VPN input, but also postsynaptic targets of those cells as well as other local interneurons. We used a two pronged approach to bias measured calcium signals towards those selective to presynaptic terminals of VPNs. First, we developed a GCaMP6F variant that preferentially localizes to presynaptic terminals (syt1GCaMP6F). This construct showed much brighter GCaMP6F fluorescence in axon terminals in the optic glomerulus compared to dendrites in the lobula (Fig. 2C). Second, as almost every LC & LPLC neuron is cholinergic, we specifically targeted cholinergic neurons using a ChAT-T2A knock-in Gal4 driver line (Deng et al., 2019). Using this driver line, we expressed both syt1GCaMP6F as well as myr::tdTomato, a plasma-membrane bound red structural indicator that was used for motion correction and alignment (Fig. 2D).

To extract glomerulus responses from our *in vivo* imaging volumes, we used techniques similar to other recent imaging alignment studies in the *Drosophila* brain (Brezovec et al., 2022; Mann et al., 2017; Pacheco et al., 2021; Turner et al., 2021). First, we generated a "mean brain" volume

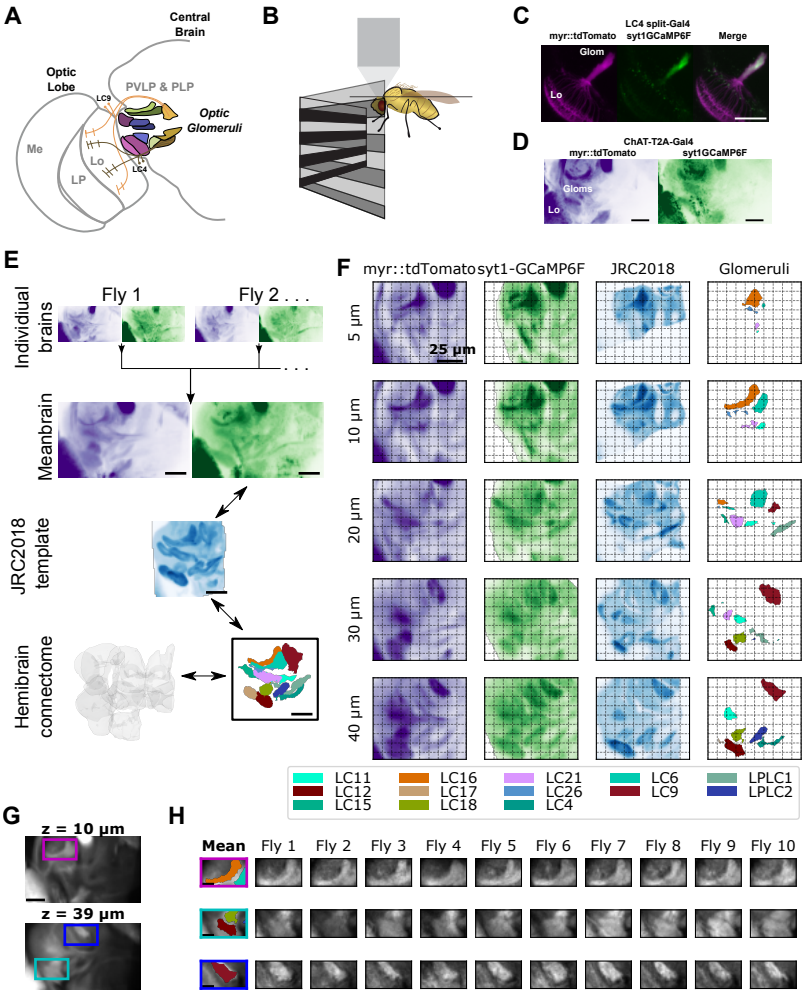

Figure 2: A method to extract optic glomerulus population responses from bulk-labeled neuropil (A) Schematic of the left half of the brain showing optic lobe and the optic glomeruli of the central brain, which receive inputs from distinct visual projection neurons. Me: Medulla, LP: Lobula Plate, Lo: Lobula, PVLP: Posterior Ventrolateral Protocerebrum, PLP: Posterior Lateral Protocerebrum. (B) Schematic of imaging and stimulation setup. (C) LC4 neuron expressing plasma membrane-bound myr::tdTomato (magenta) and presynaptically-localized syt1-GCaMP6F (Green), which is enriched in axons in the optic glomerulus. (D) For pan-glomerulus imaging, cholinergic neurons express myr::tdTomato (purple) and syt1-GCaMP6F (green). The optic glomeruli in the PVLP/PLP can be seen. (E) Pipeline for generating the mean brain from *in vivo*, high-resolution anatomical scans and aligning this mean brain to the JRC2018 template brain. Using this bridging registration, neuron and presynaptic site locations from the hemibrain connectome can be transformed into the mean brain space, allowing *in vivo* voxels to be assigned to distinct, non-overlapping optic glomeruli. (F) Montage showing z planes (rows) of the registered brain space for the mean brain myr::tdTomato (purple) and syt1-GCaMP6F (green) channels (first and second columns, respectively), JRC2018 template brain (third column) and optic glomeruli map (fourth column). (G) Mean brain images at indicated z levels showing distinct glomerulus locations of interest. (H) For the locations of interest in (G), the optic glomerulus map is overlaid on the mean brain (first column) and alignment is shown for each of 10 individual flies (remaining columns). For all images, scale bar is 25 μm.

by iteratively aligning and averaging a collection of high-resolution, *in vivo* anatomical scans of the volume of interest (Fig. 2E, n=11 flies). Next, we used the syt1GCaMP6F channel of the mean brain to align to the JRC2018 template brain (Fig. 2F; (Bogovic et al., 2020)). Finally, we generated a glomerulus map using locations of the presynaptic T-bars belonging to LC & LPLC neurons, which we extracted from the hemibrain connectome (Scheffer et al., 2020), and aligned it to the JRC2018 template brain. Using the mean brain and mean brain-template alignment, we could consistently align individual volumes to the mean brain and to the glomerulus map (Fig. 2G, H). This method, which we refer to as pan-glomerulus imaging, allowed us to assign voxels in a single fly's *in vivo* volume to a specific optic glomerulus. In this paper, we focus on 13 glomeruli (Fig 2F, Fig. S1).

To test whether pan-glomerulus imaging reliably captured visually-driven calcium responses across glomeruli, we presented a suite of synthetic stimuli meant to explore VPN feature detection (Keleş and Frye, 2017; Klapoetke et al., 2017; Klapoetke et al., 2022; Wu et al., 2016). Our stimulus suite therefore consisted of small, moving spots, static flicker, looming spots, moving bars and other stimuli (see Methods). Fig. 3A shows mean glomerulus responses across animals to these stimuli. As expected, the visual tuning measured in one glomerulus in one fly was very similar to tuning seen in corresponding glomeruli measured in other animals (Fig. 3B, C). To determine whether our pan-glomerulus imaging method accurately captured the visual tuning of the VPN that provides the major input to that glomerulus, we used cell-type specific split-Gal4 driver lines for select VPN types (LC18, LC9 and LC4), chosen because together they span the anatomical volume of interest, and presented the same stimulus suite (Wu et al., 2016). We then compared these targeted recordings to those previously measured in the corresponding glomeruli, using our population imaging approach. For each of these VPN/glomerulus pairs, the responses and visual tuning look qualitatively similar (Fig. 3D) and were highly correlated (Fig. 3E). Together, these results show that pan-glomerulus imaging reliably measures visually-driven responses across a population of optic glomeruli, and that these visual responses are dominated by VPN signals.

At a high level, this initial suite of stimuli revealed that optic glomeruli show broad, overlapping tuning (Fig. 3A) in line with previous observations using cell-type specific driver lines (Klapoetke et al., 2022). To conveniently organize the results presented in subsequent analyses, we applied a hierarchical clustering approach to identify functional groupings of VPN types based on their responses to our synthetic stimulus suite. Group 1 was characterized by LCs that responded to moving spots 5°in diameter (the smallest stimuli presented here), and showed relatively weak responses to loom and vertical bars. Group 2 contained glomeruli that were not sensitive to very small objects and showed strong loom responses. Group 3 contained glomeruli that were typically only weakly driven by any of these stimuli but responded to looming stimuli. Finally, group 4 glomeruli had large responses to vertical bars and medium and large moving spots as well as some loom sensitivity.

Population activity is modulated by a dominant gain factor which impacts stimulus coding fidelity

Previous characterization of VPNs relied on targeting each individual cell class using cell-type specific driver lines (Klapoetke et al., 2022; Wu et al., 2016). This allows for the measurement of neural response mean and variance, but not the covariance among different VPNs, which requires simultaneous measurement. Trial-by-trial covariance can have a dramatic impact on stimulus encoding (Averbeck and Lee, 2006; Averbeck et al., 2006; Romo et al., 2003; Zylberberg et al., 2016), and

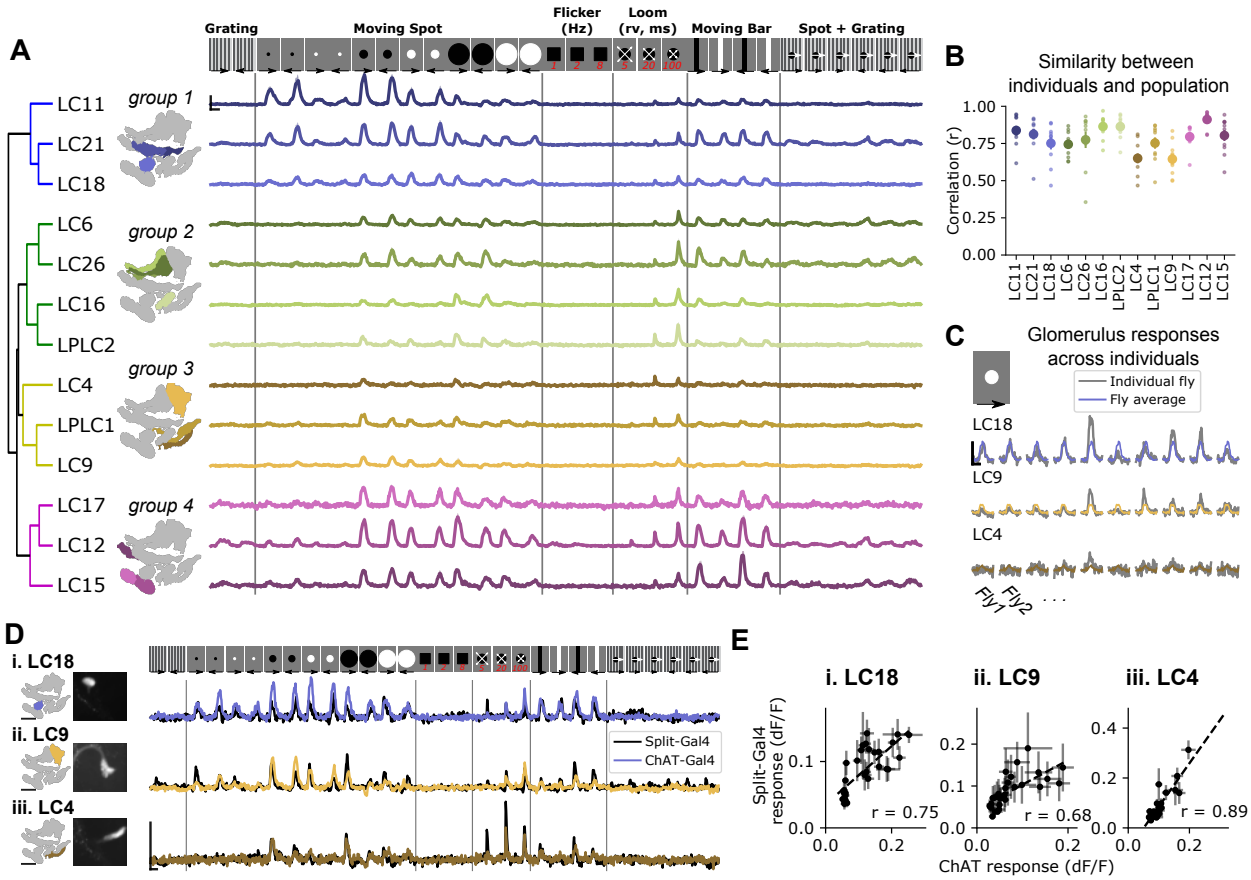

Figure 3: Pan-glomerulus imaging reliably measures optic glomerulus responses dominated by visual projection neurons (A) Responses of thirteen optic glomeruli to a panel of synthetic visual stimuli (top, see methods). Responses are shown according to the indicated stimulus order, but stimuli were presented in randomly interleaved trial order. Shown are mean glomerulus response traces across 10 flies, shading indicates S.E.M. We hierarchically clustered mean glomerulus responses to yield four functional groups of glomeruli. (B) Visual tuning can be reliably estimated in single flies using pan-glomerulus imaging. For each glomerulus, we computed the correlation between each individual fly tuning and the mean tuning (excluding that fly). Large dots and bars indicate mean +/- S.E.M., and small dots correspond to individual flies. (C) Example responses of three glomeruli in individual flies to a 15° bright moving spot. For each panel, the colored trace is the across-fly average response and the gray trace is the individual fly response. (D) Comparison of glomerulus tuning to the LC neurons that dominate their input, for three glomerulus/LC pairs. Left: glomerulus map and example image of the LC axons. Right: syt1GCaMP6F responses to the stimulus panel above. Colored traces show tuning measured using pan-glomerulus imaging procedure and black traces show split-Gal4 LC responses. (E) For the three LC types in (D), the tuning of LC axons is highly correlated with the tuning of corresponding optic glomeruli (pan-glomerulus imaging: n=10 flies; Split-Gal4 imaging: n=5, 6, and 4 flies for LC18, LC9, and LC4, respectively). For all calcium traces, Scale bar is 2 sec and 25% dF/F

can shed light on the circuit mechanisms that govern sensory computation (Ala-Laurila et al., 2011; Rabinowitz et al., 2015). To examine the covariance structure of optic glomerulus responses, we presented a subset of the synthetic stimuli (Fig. 3), and collected 30 trials for each stimulus. We observed significant trial-to-trial variability. Indeed, on some presentations of a stimulus which, on average, drives a strong response, many glomeruli failed to respond at all. Moreover, this large modulation in response gain was shared across many glomeruli on a trial-by-trial basis (Fig. 4A, B). When we averaged the trial-to-trial correlations across flies, we observed strong, positive pairwise correlations across the glomerulus population (Fig. 4C), and across stimuli (Fig. S2).

This large response variance suggests a challenge for downstream circuits integrating information across optic glomeruli: how can a visual feature be reliably decoded when response strength shows such large variability from trial to trial? To explore this issue, we implemented a multinomial logistic regression decoder to predict the identity of a stimulus given single trial population responses. Since the animal does not have *a priori* information about when or where a local visual feature might appear, we didn't want the model to be able to use different stimulus dynamics to trivially learn the decoding task based on response timing. Therefore, we trained the model using only the peak response amplitude from each glomerulus on each trial (Fig. 4D), and tested the ability of the model to predict stimulus identity on held-out trials. This decoding model performed with an overall accuracy rate of around 40%, on average (compared to a chance performance of 7%), and performance for some stimulus classes was considerably higher (Fig. 4E). For example, for dark moving spots with diameter 5°, 15°, and 50°, performance was 55%, 67% and 78%, respectively. For a slowly looming spot, performance was 72%. This high performance was surprising given that the model only had access to scalar response amplitudes on each trial, which themselves displayed high trial-to-trial variability.

We next asked how a model provided with different subsets of optic glomeruli performed on the decoding task by training the model using only responses from a single functional group (identified in Fig. 3). As expected, decoding models with access to responses from only a subset of the population performed more poorly than those with access to the full glomerulus population. Strikingly, however, subpopulations of glomeruli were unable to perform as well as the full population even for correctly classifying the stimuli to which they were most strongly tuned (Fig 3F). For example, group 1 contains the glomeruli that showed strong responses to small, 5°spots. Yet a model trained using the responses from that group alone was unable to encode information about this stimulus nearly as well as the full population model.

In other sensory systems, positive correlations in neural responses can mitigate the effects of trial-to-trial variability in cases of heterogeneous population tuning (Averbeck and Lee, 2006; Franke et al., 2016; Romo et al., 2003; Zylberberg et al., 2016). We therefore hypothesized that the strong trial-to-trial gain correlations (Fig. 4C) were partly responsible for the high decoding performance for some stimuli in spite of the high response variance. To test this, we trained and tested the decoding model using trial-shuffled responses, such that for each glomerulus the mean and variance of each stimulus response was the same, but the trial-to-trial correlations were removed (Fig. 4G, left). With trial-to-trial correlations removed, the decoding model performed about 35% worse than the model trained on correlated single trial responses (Fig. 4G, right). The decrease in performance upon trial shuffling was present across stimuli, indicating that this is a general feature of stimulus encoding for this population, and not specific for select visual features (Fig. 4H). This result highlights the importance of performing simultaneous measurements to characterize population responses: using independent measurements in this case would suggest a significantly worse single

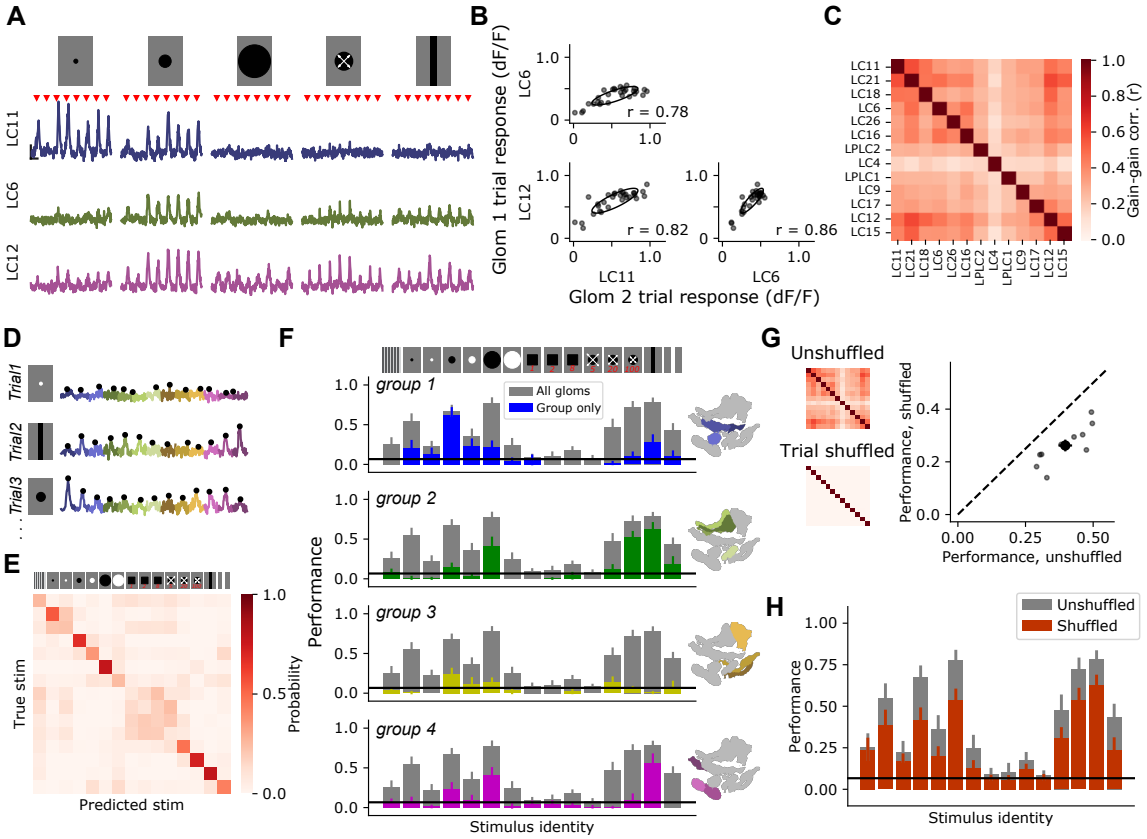

Figure 4: Glomerulus responses are modulated by a shared gain factor that impacts stimulus encoding fidelity (A) For a reduced stimulus set (images above), we presented many trials in a randomly interleaved order. For display, we have grouped responses by stimulus identity. Red marks indicate stimulus presentation times. Example single trial responses for representative glomeruli show large trial-to-trial variability. Scale bar is 4 sec and 25% dF/F. (B) For the example glomeruli in A, plotting one glomerulus response amplitude against another for a given stimulus (here, a 15° moving spot), reveals high correlated variability. Each point is a trial. Ellipses show 2D gaussian fit derived from the trial covariance matrix. (C) Trial-to-trial variability is strongly correlated across different glomeruli. Heatmap shows the average correlation matrix across all stimuli and flies (n=17 flies). (D) Single trial responses of 13 optic glomeruli, concatenated together for each trial. Stimulus identity is indicated to the left. Black dots indicate the peak responses of each glomerulus on each trial, which will be used for stimulus decoding. The response amplitudes for all 13 glomeruli were used to train a multinomial logistic regression model to predict stimulus identity (see methods). (E) Confusion matrix, with rows and columns corresponding to stimulus identity above. (F) We used held-out data to test the ability of the decoding model to predict the stimulus identity given the single trial response amplitudes of different groupings of glomeruli. In each barchart, gray bars correspond to a model with access to all 13 glomerulus responses for each trial. Colored bars show mean + S.E.M. performance of a model with access to only the indicated functional group of glomeruli. (G) Trial shuffling population responses removes pairwise correlations among glomeruli (insets to the left show correlation matrices before and after trial shuffling). Right: Decoding model performance, averaged across all stimuli, for each fly (n=11 flies). (H) Performance of the decoding model suffers across all stimulus classes when trial-to-trial correlations are removed.

trial decoding ability than is present in the full population. Taken together, these results show that, rather than a single visual feature being encoded by one or a few VPNs, all visual features are likely represented jointly across the population. Moreover, positive correlations across this population enhance stimulus decoding.

Walking behavior selectively suppresses responses of small-object detecting glomeruli

Because sensory neural activity has been shown to be modulated by behavior in flies (Chiappe et al., 2010; Fenk et al., 2021; Kim et al., 2015; Strother et al., 2018) and other animals (Maimon, 2011; Niell and Stryker, 2010), we wondered whether the trial-to-trial gain changes shown above were related to the behavioral state of the animal. To test this, we measured glomerulus population responses while the animal walked on an air-suspended ball (Fig. 5A-B, see Methods). To simplify the gain characterization, we showed a repeated probe stimulus on every trial, for 100 trials. First, we showed a 15°dark moving spot, since this stimulus drives strong responses in many glomeruli, including LC11, LC21, LC18, LC6, LC26, LC17, LC12, & LC15. We will refer to these glomeruli as "small object detecting glomeruli", recognizing that they also respond to other stimuli (Fig. 3). Examining the single trial responses to the probe alongside fictive walking behavior revealed a striking relationship: probe stimuli that appeared when the fly was walking drove much weaker responses in some glomeruli than stimuli that appeared while the fly was stationary (Fig. 5C, D). On average, responses of the LC11, LC21, L18, LC12 & LC15 glomeruli showed significant negative correlation with behavior. Conversely, responses of the LC6, LC26 and LC17 glomeruli did not show significant negative correlation with behavior.

We next tested whether a similar behavioral modulation exists for those glomeruli which respond more strongly to loom, namely LC6, LC26, LC16, LPLC2, LC4, LPLC1, LC9, LC17 and LC12, using a dark looming spot as a probe (Fig. 5E). Across animals we saw no significant modulation of loom responses by walking (Fig. 5F). Thus, walking behavior selectively suppressed the visually evoked responses of specific optic glomeruli, with the strongest effects on a subset of small object detecting glomeruli, while having no significant effect on glomeruli that respond most strongly to loom.

The gain changes associated with walking strongly resemble the correlated gain changes we saw in earlier experiments with the broader stimulus suite (Fig. 4). This suggests that the trial-to-trial shared gain was associated with the behavioral state of the animal. To test this idea, we examined the subset of flies from the experiments in Fig. 4 where we also collected walking behavior. We found that for each fly, the first principal component of the population response, corresponding to the large shared gain factor, was negatively correlated with walking (Fig. S3), with an average rank correlation coefficient of ρ = -0.23. Thus, the shared gain modulation is associated with walking, but importantly, this relationship is incomplete. This means that one could not infer the population correlation structure seen in Fig. 4 by leveraging information about walking behavior.

Visual inputs associated with self-generated motion modulate glomerulus sensitivity

Self-generated motion is associated with characteristic visual cues, including wide-field, coherent visual motion on the retina. In the next series of experiments, we set out to test the hypothesis that optic glomerulus gain might be modulated by these visual signatures of self-generated motion. To test whether glomeruli respond to visual cues characteristic of walking, we first created a complex

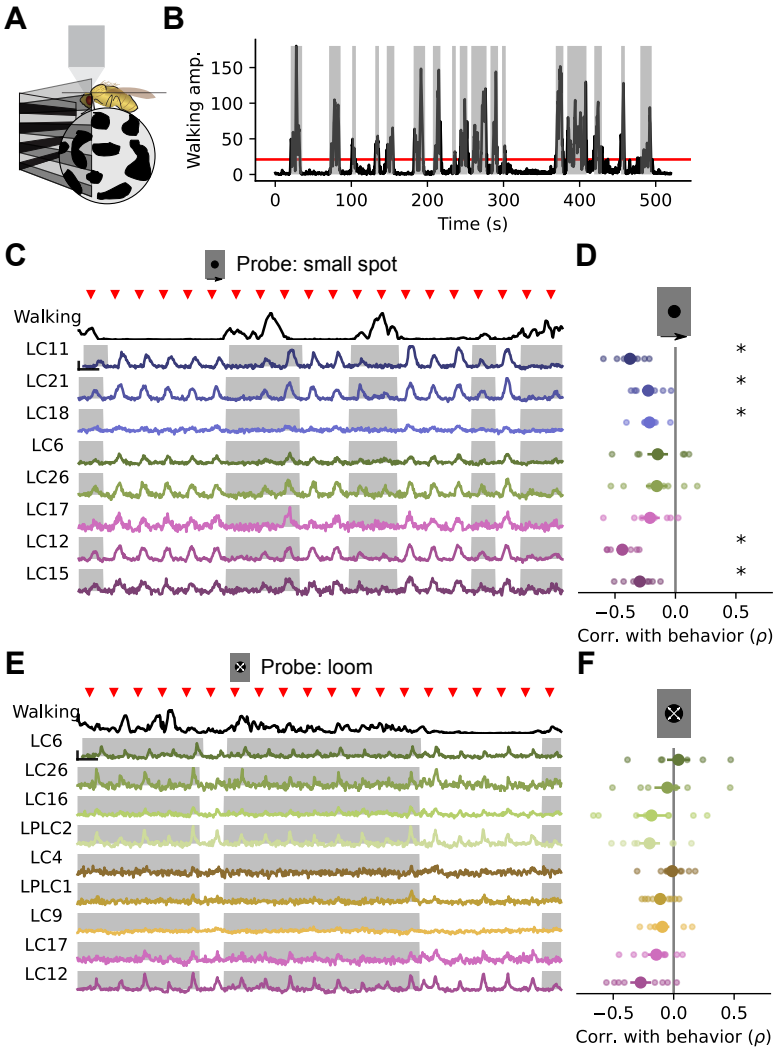

Figure 5: Walking behavior suppresses glomerulus responses to small object stimuli (A) Schematic showing fly on air-suspended ball for tracking behavior. (B) We used the change in ball rotation to measure overall walking behavior. Red line indicates threshold for binary classification of behaving vs. nonbehaving, and gray shading indicates trials that were classified as behaving. (C) We presented a 15° moving spot repeatedly to probe glomerulus gain throughout the experiment. Example responses and quantification are shown only for glomeruli which respond reliably to this probe stimulus. Red triangles above show stimulus presentation times. Black traces (top) show walking amplitude, and gray shading indicates trials classified as behaving. Scale bar is 4 sec and 25% dF/F. (D) Correlation (Spearman's rho) between behavior and response amplitude. Each small point is a fly, large point is the across-fly mean (n=8 flies), and asterisks indicate glomeruli with a significant negative correlation between response gain and behavior. (E-F) Same as C-D for a looming probe stimulus, with responses from the subset of glomeruli that respond strongly to looming stimuli. On average, there is no correlation between behavior and loom response strength (n=8 flies).

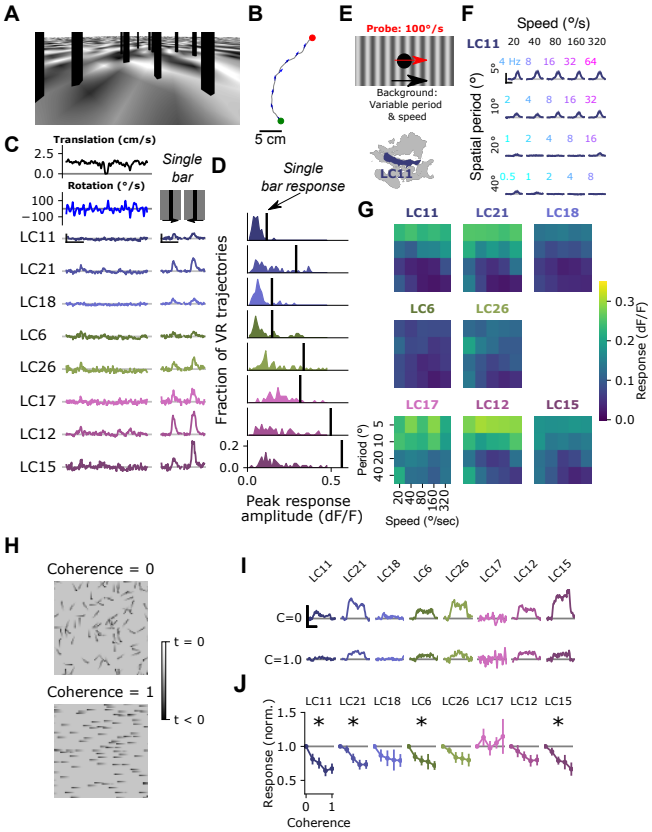

Figure 6: Visual input associated with rotational motion suppresses optic glomerulus responses (A) Still image from a VR stimulus. (B) 20 second snippet of a fly walking trajectory. Green and red points indicate start and end of the trajectory, respectively. Arrows show the fly's heading. (C) For an example fly, responses of small object detecting glomeruli to an example virtual reality trajectory. These glomeruli respond very weakly to the virtual reality stimulus, despite showing strong, reliable responses to solitary vertical bars (right) similar to those present in the virtual reality scene. Scale bars are 5 sec and 25% dF/F. (D) Histograms showing, for each glomerulus, the distribution of peak responses to each VR trajectory. Vertical line indicates mean response to a single dark, vertical bar stimulus (five VR trajectories were presented to each fly, n=10 flies). (E) Schematic showing the surround suppression tuning stimulus. A small dark probe stimulus moves through the center of the screen while a grating with varying spatial frequency and speed moves in the background. (F) For the LC11 glomerulus, probe responses are suppressed by low spatial frequency gratings across a range of speeds consistent with locomotor turns. Small, color-coded numbers indicate the temporal frequency associated with each grating speed and spatial period. Scale bar is 2 sec and 25% dF/F. (G) Heatmaps showing probe responses as a joint function of background spatial period and speed for each of these 8 glomeruli (n=10 flies). (H) Schematic of random dot coherence stimulus. At zero coherence (top image), each dot moves independently of all the other dots. At a coherence of 1.0 (bottom image), all dots move in the same direction. (I) For an example fly, responses of the small object detecting glomeruli are shown to coherence values of 0 (top row) and 1.0 (bottom row). Scale bar is 2 sec and 25% dF/F. (J) Summary data showing response amplitude (normalized to the 0 coherence condition within each fly) of each glomerulus to varying degrees of motion coherence. Asterisk at the top indicates a significant difference between the response to 0 and 1 coherence (n=11 flies).

visual stimulus designed to include several features thought to be components of natural visual inputs to walking flies, including objects at different depths (vertically oriented, dark bars), as well as images dominated by low spatial frequencies (Fig. 6A). To move this scene, we measured fly walking trajectories using a $1m^2$ arena with automated tracking, as described previously (York et al., 2022), and applied short segments of these walking trajectories to the camera location and heading in our visual environment, creating an open loop "play-back" stimulus (Fig. 6B). These VR stimuli drove very weak responses across all glomeruli, including the small object detecting glomeruli (Fig. 6C), despite these glomeruli in the same flies responding very robustly to isolated vertical bars similar to those in the scene (Fig. 6D). The relatively weak responses of most glomeruli to these play-back stimuli suggested that some features characteristic of visual inputs during walking suppress glomerulus responses via the visual surround of each VPN. To test this idea, and to explore the properties of this visual surround, we presented a 15°dark spot, a probe stimulus that many glomeruli respond to (Fig. 3), while drifting a sine wave grating in the background with variable spatial period and speed (Fig. 6E). The LC11 glomerulus, which responds strongly to small moving objects on uniform backgrounds, showed strongly suppressed probe responses to gratings with low spatial frequencies, and across speeds chosen to span the typical range of angular velocities experienced during fly locomotor turning (Fig. 6F) This suppression by low spatial frequency gratings across a range of rotational speeds was seen for all small object detecting glomeruli (Fig. 6G). Thus, these glomerulus responses are subject to a suppressive surround that is sensitive to low spatial frequencies and to a broad range of retinal speeds.

A prominent feature of self-generated visual motion, especially rotational turns, is widefield motion coherence. That is, when an animal turns, all local motion signals across the visual field are aligned along an axis defined by the axis of rotation. To test whether motion coherence impacted surround suppression of optic glomeruli, we designed a stimulus inspired by random dot kinematograms (Britten et al., 1992). This stimulus was composed of a field of small dots, roughly 15°in size, that moved around the fly at constant speed. This moving dot field had a tunable degree of coherence, such that at a coherence level of 0, each dot moved at the defined speed, but in a random direction. At a coherence level of 1, every dot moved in the same direction (Fig. 6H). Intermediate coherence values correspond to the fraction of dots moving along the pre-defined "signal" direction. Importantly, this stimulus has the same overall mean intensity, contrast and motion energy for every coherence level. As expected, at 0 coherence, small object detecting glomeruli responded strongly. However, as the motion coherence was increased, responses of many small object detecting glomeruli decreased (Fig. 6H, I). Taken together, these results are strong evidence that the suppressive surround of these glomeruli is sensitive to widefield motion cues that are characteristic of self motion.

Natural images recruit surround suppression

To test whether self motion cues derived from natural scenes can drive surround suppression in small object detecting glomeruli, we used a moving 15°spot to probe response gain while presenting moving natural images (van Hateren and van der Schaaf, 1998) (Fig. 7A). When we presented a rotating natural image behind the probe, LC11 glomerulus responses were strongly suppressed for rotational speeds spanning the range of locomotor turns (Fig. 7B). We next explored surround speed tuning across all eight small object detecting glomeruli (Fig. 7C). The LC11, LC21 and LC18 glomeruli showed strong suppression at all non-zero image speeds tested (Fig. 7C, left). The LC6

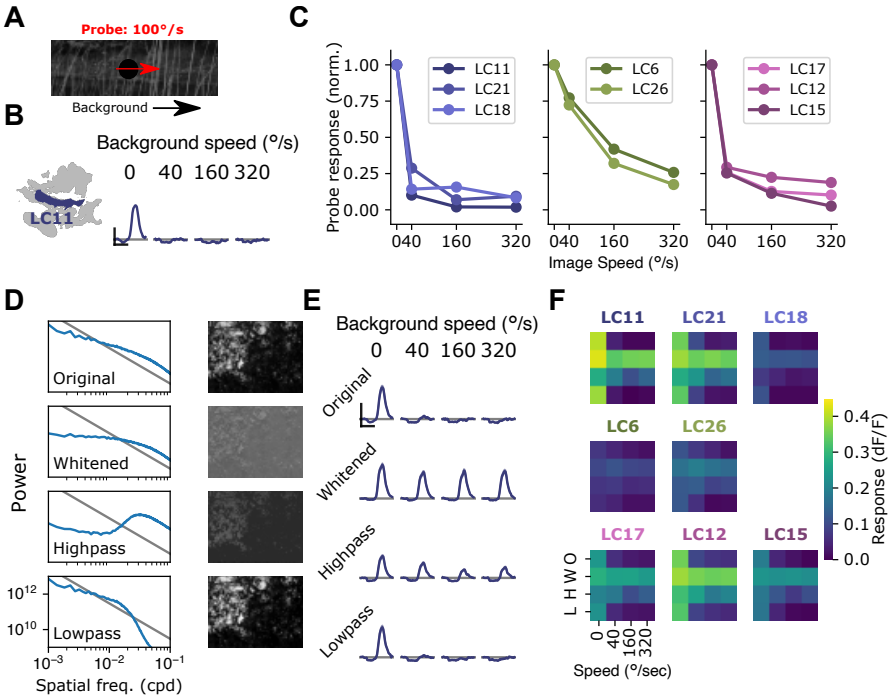

Figure 7: Visual suppression is tuned to natural image statistics (A) Stimulus schematic: a small spot probe is swept across the visual field while a grayscale natural image moves in the background at a variable speed. (B) For the LC11 glomerulus, natural image movement strongly suppresses the probe response across a range of speeds (average across three images, n=8 flies). (C) Mean probe response as a function of background image speed, normalized by probe response with static background. Surround speed tuning curves are grouped by functional glomerulus groupings for the small object detecting glomeruli. (D) Average power spectra (left) and example image (right) for the original natural images (top), whitened images (second row), high-pass filtered images (third row) and low-pass filtered images (bottom row). Gray line shows p $\propto 1/f^2$. (E) For LC11, image suppression of probe responses is attenuated by whitening the image or high-pass filtering it, but not by low-pass filtering the image. (F) Dependence of probe suppression on image speed and filtering for each of the 8 small object detecting glomeruli (n=9 flies). Scale bars are 2 sec and 25% dF/F.

and LC26 glomeruli showed a shallower dependence of surround suppression on image speed (Fig. 7C, center), while the LC17, LC12 and LC15 glomeruli showed intermediate speed dependence (Fig. 7C, right). In summary, natural images suppress small object responses in these glomeruli and surround suppression generally increases with increasing background speed.

We hypothesized that the low spatial frequency content of natural images was critical for these effects, since the grating results (Fig. 6) showed the strongest suppression for low spatial frequency gratings, and because natural images are characterized by long-range intensity correlations and low spatial frequencies (Fig. 7D). To test the effect of spatial frequency content of images on surround suppression, we repeated this experiment with filtered versions of the natural images. For each of three natural images, we presented the original (unfiltered) image, a whitened natural image, which has a roughly flat power spectrum at low spatial frequencies, a high pass filtered image, and a low pass filtered image (Fig. 7D). For LC11, and all other small object detecting glomeruli, the natural image and its low-pass filtered version strongly suppressed responses to the probe, whereas the whitened and high-pass filtered images recruited much weaker suppression (Fig. 7E,F). We note that this spatial and temporal frequency tuning of these suppressive surrounds is broadly consistent with the tuning properties of elementary motion detecting neurons T4 & T5 (Leong et al., 2016; Maisak et al., 2013). These observations raise the possibility that local motion detectors provide critical input to the visual surrounds of small object detecting glomeruli. Moreover, the observation that surround speed tuning was similar for glomeruli that clustered together using the synthetic stimulus suite (Fig. 3) suggests that this functional clustering may reflect properties of the surround. More broadly, the differential speed sensitivities of these surrounds may further diversify feature selectivity across these groups of glomeruli in the context of natural visual inputs.

Behaviorally and visually-driven suppression independently modulate small object detectors

The results presented thus far show that the gain of small object detecting glomeruli was tuned by both locomotor behavior and widefield visual motion. Both of these cues are associated with self-generated movements of the animal. How can the fly reliably track external objects during self motion if small-object detecting glomeruli are suppressed by visual and behavioral cues? We hypothesized that the answer might lie in the temporal dynamics of locomotor behavior. Fly walking behavior is saccadic, interspersing fast turns with periods of relatively straight walking bouts (Cruz et al., 2021; Geurten et al., 2014; Juusola et al., 2017; Reynolds and Frye, 2007). We hypothesized that the saccadic structure of walking ensures that glomerulus gain is suppressed only transiently during a saccade, and once the saccade is over, visual response gain is restored to sample external objects.

To test this idea, we first examined the temporal dynamics of locomotor turns under conditions where animal movement is unconstrained. To do this, we examined walking trajectories from our open behavioral arena (See Fig. 6 and York et al., 2022) (Fig. S4A). Examining the angular velocity of a single walking trajectory revealed saccadic temporal dynamics (Fig. S4B). Across all flies, there were few inter-turn intervals less than ∼0.5 sec, a peak near 1 sec, and a long tail (Fig. S4C, n=81 flies). We chose a threshold angular velocity to classify saccadic turns, here 100°/sec, but across a range of threshold values that include the vast majority of turns, the median inter-turn interval ranged from 0.6-1.7 sec. For all saccade thresholds, there was a low probability of a saccade within ∼0.5 sec of the previous saccade. This means that a typical locomotor saccade is followed by at

least a 500ms, and often a ∼1 second period of relative heading stability.

We next asked whether saccadic visual inputs recruit surround suppression, and whether the timescale of this suppression could support such a visual sampling strategy. We designed a stimulus meant to mimic the retinal input during a locomotor saccade. As before, we presented a probe stimulus on every trial to measure the response gain of small object detecting glomeruli. In the background was a grayscale natural image (Fig. 8A), which underwent a lateral rotation of 70°in 200 msec at a variable time relative to the glomerular response to the probe (Fig. 8B). As a result, the saccade signal could precede, co-occur with, or lag the glomerulus response to the probe. When the saccade occurred within ∼500 milliseconds of the probe response, the probe response was attenuated, suggesting that this saccade stimulus recruits the motion-sensitive suppressive surround. Across many small object detecting glomeruli, including LC11, LC21, LC17, LC12, and LC15, we saw strong gain suppression when the saccade occurred around the time of the probe response (Fig. 8C, D). Interestingly, the other glomeruli, LC18, LC6, and LC26 showed much weaker and more variable saccade suppression, maintaining their response gain regardless of saccade timing. Note that because we presented only one saccade on each probe trial, and because we are quantifying gain using the response amplitude, the timing dependence measured here is independent of calcium indicator dynamics and therefore reflects dynamics associated with the glomerulus response. The timescale of this surround suppression, combined with the temporal dynamics of fly turning (Fig. S4) suggests that visually-driven saccade suppression transiently reduces glomerulus response gain around the time of a locomotor saccade, but gain recovers while the fly's heading is stable and before the next saccade occurs. We infer that this dynamic gain adjustment allows the fly to sample the scene during the inter-saccadic periods of heading stability.

Because the brief saccade stimulus did not completely suppress probe responses, we could examine the relationship between visual- and motor-related gain control mechanisms. One way to interpret these data is that motor signals suppress small object detecting glomeruli, and that widefield, coherent visual motion induces a turning response, recruiting the same motor-command derived suppression. Is the apparent visual suppression a result of the motor feedback, or are the visual- and motor- based suppression mechanisms independent? To test this idea, we monitored walking behavior while presenting saccadic visual stimuli (Fig 8E). We first examined the LC11 glomerulus response to the probe under both behavioral conditions (walking versus stationary), and under both visual conditions (saccade coincident with the probe response, versus no saccade coincident with the probe response). Strikingly, when the fly was stationary, visual saccades still reduced the gain of the response (Fig. 8F). Interestingly, the glomeruli that were subject to stronger gain reductions by the visual saccade also showed stronger gain reductions by walking (Fig. 8G, $r = 0.80$). Finally, to test whether these two gain modulation mechanisms were independent, we compared the measured probe responses when the fly was receiving both saccadic visual input and was walking to the product of each gain change measured independently (i.e. when the fly was receiving either saccadic visual input or was walking, but not both) (Fig. 8H). Across the small object detecting glomeruli, the prediction that these two gain mechanisms were independent accurately captured the jointly measured gain modulation. Taken together, these data indicate that visual suppression is not the indirect effect of an induced turning response and that saccadic visual modulation and behavior-related modulation are both balanced in magnitude and independent.

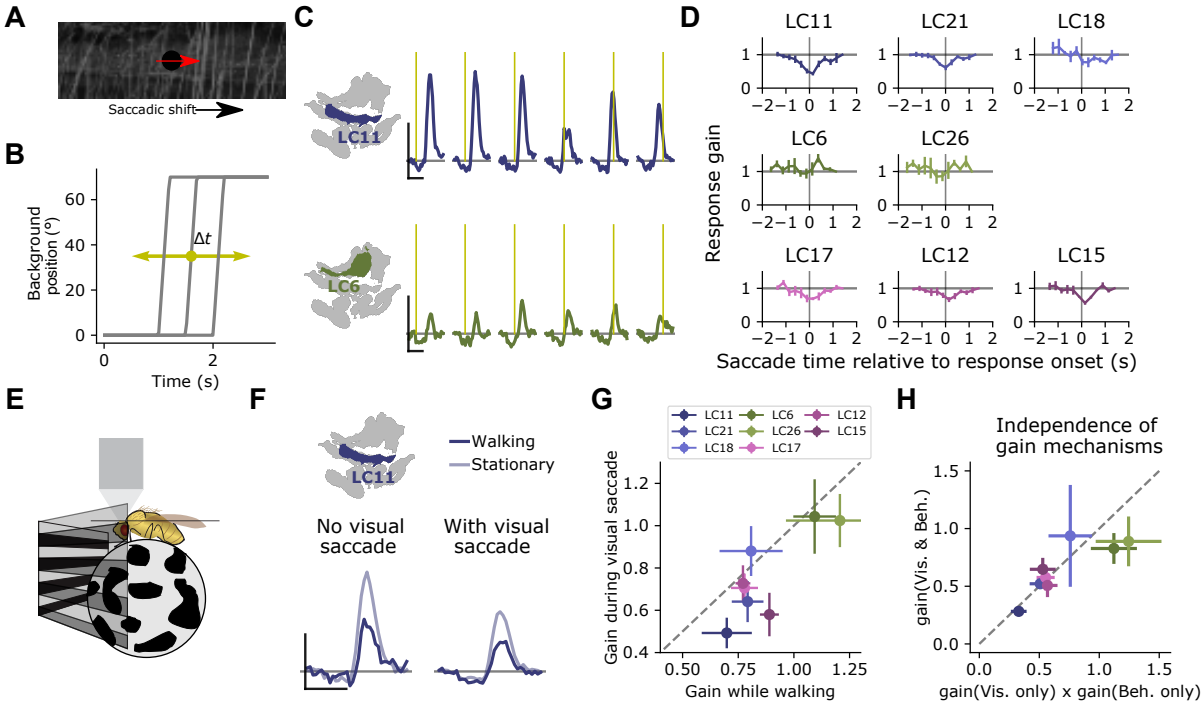

Figure 8: Visually-driven saccade suppression works in concert with behavioral suppression (A) Stimulus schematic: A small moving probe is used to measure glomerulus response gain while a full-field natural image in the background undergoes a fast, ballistic lateral displacement meant to mimic fly walking saccades. (B) The time of the saccade is varied throughout the trial, such that it occurs at different phases of the probe response. Each saccade lasted 200 ms and translated the image by 70°. (C) Trial-average probe responses in an example LC11 glomerulus (top) and an example LC6 glomerulus (bottom). Saccade times are indicated by the yellow vertical line in each panel. As the saccade approaches the probe in time, the probe response is suppressed in LC11 but not LC6. (D) Summary data showing, for each small object detecting glomerulus, the response gain as a function of saccade time relative to the response onset. t=0 corresponds to coincident probe response and saccade onset. Suppression is strongest when the saccade occurs near probe response onset. (n=5 flies) (E) We monitored fly walking behavior while presenting the saccadic visual stimuli above. (F) For an example LC11 glomerulus, both the saccadic stimulus and walking behavior reduced probe responses, and these gain control mechanisms could be recruited separately. (G) Population data showing average response gain for visual saccades versus response gain during walking behavior, for each small object detecting glomerulus. Most glomeruli lie below unity (dashed line), indicating that visual saccade suppression is typically stronger than behavior-linked suppression. These two sources of suppression are also correlated across glomeruli (r = 0.80). (H) For each small object detecting glomerulus, we compared the response gain when both visual and motor related gain modulation were recruited (vertical axis) to the product of both response gains in isolation (horizontal axis). Dashed line is unity, and points falling along that line indicate a linear interaction between those forms of gain control.

Discussion

In this study, we show that local feature detection is challenged by self motion signals in rich visual environments (Fig. 1). To determine how feature detecting neurons might maintain selectivity under natural viewing conditions, we first developed a new connectome-based method to segment functional imaging signals that allowed us to measure neural responses across a heterogeneous population of VPNs (Fig. 2-3). Using this method, we found that VPNs represent visual features jointly across the population, meaning that stimulus identity cannot be decoded from a single glomerulus alone (Fig. 4). Further, we found that strong trial-to-trial response correlations improve stimulus encoding fidelity (Fig. 4). Strikingly, the locomotor behavior of the fly selectively modulated responses of small object detecting, but not loom detecting, glomeruli (Fig. 5). We then showed that visual motion signals characteristic of walking also modulated the responses of glomeruli tuned to small objects (Fig. 6-7). Finally, we demonstrated that visual suppression occurs during naturalistic body saccades made by walking flies, and that behavioral and visual gain modulation are both balanced in magnitude and independent, such that these two cues combine linearly (Fig. 8). Taken together, these two forms of gain control reduce the sensitivity of small object detectors to inputs that can diminish the discriminability of local features, thereby allowing for reliable feature detection during saccadic vision.

Population coding of local visual features

Our characterization of the optic glomeruli using solitary visual features (Fig. 3) largely agrees with what has been described previously, in that many glomeruli respond strongly to small moving objects, others respond to visual loom, and responses to stationary flicker or widefield motion are weak or nonexistent (Hindmarsh Sten et al., 2021; Keleş and Frye, 2017; Keleş et al., 2020; Klapoetke et al., 2022; Städele et al., 2020). These data have been used as evidence that particular VPNs are linked to specific visual features and corresponding visually-guided behaviors (Hindmarsh Sten et al., 2021; Ribeiro et al., 2018). At the same time, the responses of individual VPN classes overlap, in the sense that an individual visual stimulus will evoke responses from many VPN classes, suggesting a dense population code. To what extent is it possible to decode stimulus identity based on the activity of a single VPN class? Our results show that population measurements are important to describe feature encoding by VPNs for two reasons. First, evaluation of stimulus decoding revealed that most visual features are encoded jointly across the population, not by single VPN types. This is because information about stimulus identity is contained not only in the responses of glomeruli that are strongly tuned to a particular feature but also in the weaker responses of glomeruli that have different tuning properties. Second, while responses in each VPN type showed high trial-to-trial variability, simultaneous measurements revealed that this variability was strongly correlated across the population, thus improving coding fidelity across the population relative to uncorrelated variability. This trial-to-trial variability was dominated by a single, shared population response gain that was associated with walking behavior, but only weakly (Fig. S3). Thus, this shared gain is likely modulated by other factors, for example, shared upstream noise (Ala-Laurila et al., 2011; Zylberberg et al., 2016) or other behavioral or physiological states that we did not measure. How downstream circuits combine signals across glomeruli may provide insight into how the brain decodes VPN population responses to encode local features, and available connectomic datasets can accelerate progress on this question (Klapoetke et al., 2022; Scheffer et al., 2020).

Natural locomotor behavior modulates the sensitivity of small object detectors

Behavior-associated gain changes are widespread in visual systems across phyla (Maimon, 2011; Maimon et al., 2010; McAdams and Maunsell, 1999; McBride et al., 2019; Niell and Stryker, 2010). Recent work demonstrates that locomotor signals are prevalent throughout the *Drosophila* brain, including in the visual system (Aimon et al., 2019; Brezovec et al., 2022; Schaffer et al., 2021), but has been examined most extensively in circuits involved in elementary motion detection and widefield motion encoding. Behavioral activity has been shown to modulate response gain in widefield motion detecting lobula plate tangential cells (LPTCs) and some of their upstream circuitry (Chiappe et al., 2010; Kohn et al., 2021; Maimon et al., 2010; Strother et al., 2018; Suver et al., 2012), and LPTC membrane potential tightly tracks walking behavior, even in the absence of visual stimulation (Cruz et al., 2021; Fujiwara et al., 2017; Fujiwara et al., 2022). During flight, efference-copy based modulation of LPTC membrane potential has been proposed to cancel expected visual motion due to self-generated turns (Fenk et al., 2021; Kim et al., 2015; Kim et al., 2017). In each of these cases, behavioral signals adjust response gain according to expected visual inputs, for example faster rotational speeds during flight. Conversely, the behavioral gain modulation we describe here selectively adjusts visual sensitivities to reflect the fact that specific visual inputs are particularly corrupted by self motion. Small object detection is an especially challenging task during self motion (Fig. 1), and consequently, gain modulation most strongly affects glomeruli involved in this task. Glomeruli that are tuned more strongly to looming visual objects were not modulated by walking behavior, suggesting that these larger visual features can be reliably extracted under walking conditions. During a high-velocity locomotor saccade, sensitivity to small objects is transiently decreased, and in the subsequent inter-saccade interval, small object detector gain is restored, allowing for selective encoding of visual features at different points in the locomotor cycle.

Is the behavioral modulation of small object detecting glomeruli related to the well studied modulation of widefield motion detecting circuits? A parsimonious explanation of both of these observations is that neurons in the elementary and widefield motion pathways feed into the suppressive surround of small object detecting glomeruli. This would endow optic glomerulus surrounds with both the widefield, coherent motion sensitivity as well as the behavioral modulation that we see. In support of this proposed mechanism, the glomeruli that show strong visual suppression are also subject to strong behavioral suppression (Fig. 8). This hypothesis further predicts that glomeruli which derive their excitatory center inputs from elementary motion detectors (e.g. the loom-selective LPLC2 (Klapoetke et al., 2017)) might be positively gain modulated under other behavioral conditions, such as flight. Taken together, these results demonstrate that understanding local feature detection during natural vision requires accounting for the structure of locomotion. More broadly, we have shown that walking behavior modulates a subset of glomeruli, raising the possibility that different behavioral states might selectively alter other glomeruli subsets, reshaping population coding of visual features to subserve different goals.

Motor signals and visual cues provide independent inputs to feature detectors

In addition to the motor-related gain modulation, small object detecting glomeruli are modulated by a visual surround that is tuned to widefield, coherent visual motion that would normally be associated with locomotion. Why would the fly visual system rely on these two seemingly redun-

dant cues to estimate self motion? One possibility is that either cue alone could be unreliable or ambiguous under some conditions. For example, a striking characteristic of natural scenes is their immense variability from scene to scene. As a result, detecting small moving objects could occur against a background of a dense, contrast-rich visual environment like a forest or a uniform, low-contrast background like a cloudy sky. These two scenes would be expected to be associated with very different wide field motion signals, even given the same self-motion. Because of this, relying on visual cues alone for evidence of self motion will be unreliable under the diversity of natural scenes. Thus, motor signals and visual cues characteristic of self motion work together to provide a robust estimate of self motion to feature detectors.

Our observation that small object detecting glomeruli are modulated by a visual surround tuned to widefield motion agrees with previous observations that flies use global motion as well as local figure information to support object tracking behavior during flight (Aptekar et al., 2012; Aptekar et al., 2015), where rotational velocities are much greater in magnitude than those associated with locomotor turns (Fry et al., 2003). How strategies for reliable object tracking during walking relate to flying conditions is not clear, and more work is needed to understand how small object detectors can support object tracking under these drastically different visual conditions.

Saccade suppression as a general visual strategy

Visual motion is a prominent feature of realistic retinal inputs for both flies and vertebrates. Primates make frequent eye movements at different spatial scales during free viewing which can rapidly translate the image on the retina (Rucci and Victor, 2015; Van Der Linde et al., 2009; Zuber et al., 1965). Eye movements in primates are dominated by saccades, large movements that can shift the image on the retina by up to tens of degrees of visual angle. Walking flies perform locomotor saccades, which similarly rapidly shift the image impinging on the retina in a short time period (Fig. S4, (Cruz et al., 2021; Geurten et al., 2014)). We found that the responses of some small object detecting glomeruli were suppressed around the time of a simulated visual saccade, while others (LC18, LC6, and LC26) showed no visual saccade suppression. Similarly, saccades in primates induce variable changes in response gain across different brain regions, a physiological effect thought to underlie the perceptual phenomenon of saccadic suppression (Binda and Morrone, 2018; Bremmer et al., 2009; Wurtz, 2018; Thiele et al., 2002). Our data show that in flies, a similar form of saccade related suppression can be recruited selectively to circuit elements whose feature selectivity is most sensitive to the corrupting effect of self motion on the visual input. More broadly, this work suggests that a saccade-and-sample visual strategy is shared between flies and primates.

Acknowledgments

We thank Estela Stephenson for excellent technical support. Steven Herbst designed the original version of flystim, of which an updated version was used for visual stimulation for this work. We thank Fred Rieke and members of the Clandinin lab for helpful feedback on this manuscript. This project was supported by NIH grants F32-MH118707 (MHT), K99-EY032549 (MHT), R01 EY022638 (TRC), R01NS110060 (TRC), the NSF GRFP (AK), and an NDSEG fellowship (MMP).

Author contributions

MT conceived the project, designed and performed the experiments, and performed the analysis and modeling. AK collected the fly walking trajectory data. MMP designed syt1GCaMP6F. TRC supervised all aspects of the project. MT and TRC wrote the manuscript.

Methods

Data and code availability

All data collected for this study will be posted to a public data repository upon final publication. All software and analysis code used for this study can be found on GitHub. Of particular note, the analysis code used to analyze these data and generate the figures presented here, can be found on github at `https://github.com/mhturner/glom_pop`.

Fly lines and genetic constructs

We generated the UAS-syt1GCaMP6F construct by cloning the cDNA sequence of Drosophila synaptotagmin 1, a 3x GS linker, and the GCaMP6F sequence into the pJFRC7-20XUAS vector (Pfeiffer et al., 2010) (Genscript Biotech). The GS linker connects the C-terminus of syt1 to the N-terminus of GCaMP6F (after Cohn et al., 2015). Transgenic flies were generated by PhiC31-mediated integration of the construct into the attP40 landing site (BestGene).

The genotype of flies used for pan-glomerulus imaging was the following:

$$\frac{w^+}{w^-}; \frac{UAS{-}syt1GCaMP6F}{+}; \frac{UAS{-}myr{::}tdTomato}{TI\{2A{-}GAL4\}ChAT\{2A{-}GAL4\}}$$

For Split-Gal4 imaging (Fig. 3), we used the following genotype:

$$\frac{w^+}{w^-}; \frac{UAS{-}syt1GCaMP6F}{LCxx{-}p65.AD}; \frac{UAS{-}myr{::}tdTomato}{LCxx{-}GAL4.DBD}$$

Where LCxx corresponds to a pair of LC subtype-specific hemidrivers from (Wu et al., 2016).

Animal preparation and imaging

Female flies, 2-7 days post eclosion, were selected for imaging. Flies were cold anesthetized and mounted in a custom-cut hole in an aluminum shim at the bottom of an imaging chamber before being immobilized with UV curing glue. The front left leg was removed to prevent occluding the left eye, and the proboscis was immobilized using a small drop of UV curing glue. The cuticle covering the left half of the posterior head capsule was removed using a fine dissection needle, and fat bodies and trachea covering the brain were removed. The prep was continuously perfused with room temperature, carbogen-bubbled fly saline throughout the experiment. We imaged the left optic glomeruli in each fly.

For in vivo imaging, we used a two-photon resonant scanning microscope (Bruker) with a 20x 1.0 NA objective (Leica) and a fast piezo-driven Z drive to control the focal plane during volumetric imaging. Two photon laser wavelength was 920 nm and post-objective power was ∼15 mW. We collected red and green channel fluorescence to image myr::tdTomato and syt1GCaMP6F, respectively. For functional scans, to record GCaMP responses, we collected volumes with voxel resolution

1x1x4 μm (x, y, z) at a sampling frequency of 7.22 Hz. For high resolution anatomical scans, voxels were 0.5x0.5x1 μm. The imaging volume for glomerulus imaging was 177 x 101 x 45 μm. Each fly was typically imaged for approximately 30-45 minutes. For Split-Gal4 imaging, we used the same imaging parameters that we did for the pan-glomerulus imaging experiments.

Visual stimulation

We back-projected visual stimuli from two LightCrafter 4500 projectors onto a fabric screen covering the front visual field of the animal. The screens subtended approximately 60° in elevation and 140° in azimuth. We used the blue LED of the projectors and a 482/18nm bandpass spectral filter to limit bleedthrough into our green PMT channel. Visual stimuli were generated using a python and OpenGL-based, open source software package we have developed in the lab, called flystim `https://github.com/ClandininLab/flystim`. Flystim renders three-dimensional objects in real time and computes the required perspective correction based on the geometry of the screen and animal position in the experimental setup to generate perspective-appropriate virtual reality stimuli. Rotating stimuli (e.g. gratings, images) were rendered as textures on the inside of virtual cylinders. Small spot stimuli were rendered as patches moving on cylindrical or spherical trajectories. Another custom, open-source software package, visprotocol `https://github.com/ClandininLab/visprotocol`, was used to control visual stimulation protocols and handle experimental metadata.

Stimulus code for every stimulus used here can be found in the github repositories for flystim and visprotocol. Below we describe some of the key visual stimulus parameters. For the synthetic visual stimulus suite, we presented 32 distinct stimulus parameterizations. All stimuli were presented from a mean gray background that remained on, between trials, throughout the entire experiment. Each stimulus presentation period was 3 seconds long, and was preceded and followed by 1.5 seconds of pre- and tail- time with a mean gray background. Note that we also presented uniform flashes of +/-100% contrast, but these stimuli did not drive responses in any glomerulus so we have excluded these stimuli from this paper.

For natural image experiments (Figs. 1, 7 & 8), we used grayscale natural images from the van Hateren database (van Hateren and van der Schaaf, 1998). When presenting filtered versions of natural images, we rescaled the filtered images such that they had the same mean and standard deviation pixel values as the original images. We scaled the whitened images to have the same peak pixel intensity as the original image.

For the saccade stimulus (Fig. 8), we used a van Hateren natural image as the background while a small, dark probe stimulus (15°in diameter) moved across the screen at 100°/sec. The background image was translated by 70°over 200 msec to mimic fly walking saccades (Cruz et al., 2021).

Virtual reality stimuli (Fig. 6) consisted of a 3D environment with a gaussian-smoothed random noise texture on the "floor" and a collection of randomly-located vertical, dark, cylinders. To simulate the visual input that would be generated from *Drosophila* walking through such an environment, we moved the camera through the scene according to measured fly walking trajectories. Trajectories of female flies walking in the dark were measured in a 1 square meter arena with automatic locomotion tracking, as described previously (York et al., 2022). 20 second snippets from measured trajectories were selected to include periods of locomotor movement, and to exclude long stationary periods. Each fly was presented with five walking trajectories, each with its own randomly-generated pattern of cylinder locations, and five trials of each trajectory were shown.

Behavior tracking

For experiments with behavior tracking, we raised a patterned, air-suspended ball underneath the fly to monitor its fictive walking behavior, as in Brezovec et al., 2022. We monitored the fly and ball movement using IR illumination and a camera triggered by our imaging acquisition software at 50Hz frame rate.

Alignment between *in vivo* functional imaging data and glomerulus map

To assign voxels in a single fly's functional in vivo image to an optic glomerulus of interest, we generated a chain of image registrations using ANTsPy (Avants et al., 2014; Tustison et al., 2021). First, each volumetric image series, including both functional and anatomical scans, was motion corrected using the myr::tdTomato signal. We then created a "mean brain" using high resolution anatomical scans from 11 different animals, which we aligned to one another using the myr::tdTomato channel, and averaged iteratively until a clean, crisp mean brain of the PVLP/PLP was produced. The syt1GCaMP6F channel of the mean brain was then used to register the mean brain to a hand-cropped subregion of the JRC2018 template brain (Bogovic et al., 2020). To generate glomerulus masks, we first extracted the presynaptic T-bar locations in the PVLP/PLP for all LC and LPLC neurons using the *Drosophila* hemibrain connectome (Scheffer et al., 2020) and custom written R code relying on the natverse suite of registration tools (Bates et al., 2020). We used a published transformation between JRC2018 space and the Drosophila hemibrain connectome space (Scheffer et al., 2020), as a start to map hemibrain synapse locations to JRC2018 space, but we also computed a small additional transformation between VPN T-Bar density and JRC2018 to improve alignment at the glomerulus level. This yielded masks for each glomerulus in our *in vivo* mean brain space. Finally, each fly's functional image was registered to that fly's own high resolution anatomical scan, and this anatomical scan was aligned to the mean brain. We could then bring each glomerulus mask into the functional image space of each individual fly. These masks were used to collect voxels corresponding to each distinct glomerulus, and the included voxel signals were averaged over space to yield the glomerulus response. For Split-Gal4 imaging data, we hand-drew ROIs in the glomerulus.

Analysis of visually-evoked calcium signals

Glomerulus responses from imaging series were aligned to visual stimulus onset times using a photodiode tracking the projector timing. We used a window of time before stimulus onset (typically 1-2 seconds) to measure a baseline fluorescence for each trial. Using this baseline, we converted trial responses to reported dF/F values. For the functional clustering presented in Fig. 3, we used a complete linkage criterion.

Small object discriminability analysis

For the small object discrimination task in Fig. 1, we moved a 15° dark patch across a grayscale natural image and through a "receptive field" similar in size to small object detecting VPNs. For each time point, we defined the local luminance as the average pixel intensity within the receptive field and the local spatial contrast as the variance of pixel intensities normalized by the mean pixel intensity within the receptive field. We quantified discriminability between the "spot present" and "spot absent" conditions using d', defined below:

$$d' = \frac{mean_{present} - mean_{absent}}{\sqrt{(var_{present} + var_{absent})/2}}$$

Where mean and var represent the mean and variance of luminance or contrast within the time window when the patch passed through the receptive field. For luminance-based discrimination we inverted the sign of d' because the presence of the patch was indicated by a decrease in local luminance.

Single trial stimulus decoding model

For the single trial decoding model presented in Fig. 4, we used a multinomial logistic regression model to predict stimulus identity using a vector of glomerulus response amplitudes for each trial. For the decoding model, responses for each glomerulus were z-scored to standardize the mean and variance across glomeruli. To train the model, we used 90% of trials, and the remaining 10% of trials were used to test performance. We iterated training/testing 100 times and we present averages across all iterations. For the trial shuffling analysis in Fig. 4, we shuffled response amplitudes across trials of the same stimulus identity independently for each glomerulus, such that the stimulus-dependent means and variances of responses were the same, but the covariance structure was removed.

Analysis of behavior data

To measure fictive walking behavior from video recordings of flies on an air suspended ball, we used FicTrac (Moore et al., 2014) to process videos post-hoc. To measure walking amplitude, at each point in time, we calculated the magnitude of the total rotation vector, using the ball rotation over all three axes of rotation, i.e. walking amplitude $= \sqrt{rot_x^2 + rot_y^2 + rot_z^2}$. To classify trials as walking vs. not walking, a threshold was automatically determined for each walking amplitude trajectory, using the Li minimum cross entropy method (Li and Lee, 1993). A trial was classified as walking if the walking amplitude exceeded this threshold for at least 25% of the time points in that trial.

References

Ache JM, Polsky J, Alghailani S, Parekh R, Breads P, Peek MY, Bock DD, von Reyn CR, Card GM (2019) Neural Basis for Looming Size and Velocity Encoding in the Drosophila Giant Fiber Escape Pathway. *Current Biology* 29:1073–1081.e4.

Aimon S, Katsuki T, Jia T, Grosenick L, Broxton M, Deisseroth K, Sejnowski TJ, Greenspan RJ (2019) Fast near-whole–brain imaging in adult Drosophila during responses to stimuli and behavior. *PLOS Biology* 17:e2006732 Publisher: Public Library of Science.

Ala-Laurila P, Greschner M, Chichilnisky EJ, Rieke F (2011) Cone photoreceptor contributions to noise and correlations in the retinal output. *Nature Neuroscience* 14:1309–1316.

Angelaki DE, Hess BJM (2005) Self-motion-induced eye movements: effects on visual acuity and navigation. *Nature Reviews Neuroscience* 6:966–976 Number: 12 Publisher: Nature Publishing Group.

Aptekar JW, Keleş MF, Lu PM, Zolotova NM, Frye MA (2015) Neurons Forming Optic Glomeruli Compute Figure–Ground Discriminations in Drosophila. *Journal of Neuroscience* 35:7587–7599 Publisher: Society for Neuroscience Section: Articles.

Aptekar J, Shoemaker P, Frye M (2012) Figure Tracking by Flies Is Supported by Parallel Visual Streams. *Current Biology* 22:482–487.

Avants BB, Tustison N, Johnson H (2014) Advanced Normalization Tools (ANTS) p. 41.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nature Reviews Neuroscience* 7:358–366.

Averbeck BB, Lee D (2006) Effects of Noise Correlations on Information Encoding and Decoding. *Journal of Neurophysiology* 95:3633–3644.

Baccus SA, Olveczky BP, Manu M, Meister M (2008) A Retinal Circuit That Computes Object Motion. *Journal of Neuroscience* 28:6807–6817.

Bates AS, Manton JD, Jagannathan SR, Costa M, Schlegel P, Rohlfing T, Jefferis GS (2020) The natverse, a versatile toolbox for combining and analysing neuroanatomical data Publisher: eLife Sciences Publications Limited.

Binda P, Morrone MC (2018) Vision During Saccadic Eye Movements. *Annual Review of Vision Science* 4:193–213.

Bogovic JA, Otsuna H, Heinrich L, Ito M, Jeter J, Meissner G, Nern A, Colonell J, Malkesman O, Ito K, Saalfeld S (2020) An unbiased template of the Drosophila brain and ventral nerve cord. *PLOS ONE* 15:e0236495.

Borst A, Haag J, Reiff DF (2010) Fly Motion Vision. *Annual Review of Neuroscience* 33:49–70.

Bremmer F, Kubischik M, Hoffmann KP, Krekelberg B (2009) Neural Dynamics of Saccadic Suppression. *Journal of Neuroscience* 29:12374–12383.

Brezovec LE, Berger AB, Druckmann S, Clandinin TR (2022) Mapping the Neural Dynamics of Locomotion across the Drosophila Brain Technical report, bioRxiv Section: New Results Type: article.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience* 12:4745–4765 Publisher: Society for Neuroscience Section: Articles.

Britten KH (2008) Mechanisms of Self-Motion Perception. *Annual Review of Neuroscience* 31:389–410 _eprint: https://doi.org/10.1146/annurev.neuro.29.051605.112953.

Chiappe ME, Seelig JD, Reiser MB, Jayaraman V (2010) Walking Modulates Speed Sensitivity in Drosophila Motion Vision. *Current Biology* 20:1470–1475.

Cohn R, Morantte I, Ruta V (2015) Coordinated and Compartmentalized Neuromodulation Shapes Sensory Processing in Drosophila. *Cell* 163:1742–1755.

Cruz TL, Pérez SM, Chiappe ME (2021) Fast tuning of posture control by visual feedback underlies gaze stabilization in walking Drosophila. *Current Biology* 31:4596–4607.e5.

Deng B, Li Q, Liu X, Cao Y, Li B, Qian Y, Xu R, Mao R, Zhou E, Zhang W, Huang J, Rao Y (2019) Chemoconnectomics: Mapping Chemical Transmission in Drosophila. *Neuron* 101:876–893.e4.

Fenk LM, Kim AJ, Maimon G (2021) Suppression of motion vision during course-changing, but not course-stabilizing, navigational turns. *Current Biology* 31:4608–4619.e3.

Fischbach KF, Dittrich A (1989) The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell and Tissue Research* 258.

Franke F, Fiscella M, Sevelev M, Roska B, Hierlemann A, Azeredo da Silveira R (2016) Structures of Neural Correlation and How They Favor Coding. *Neuron* 89:409–422.

Fry SN, Sayaman R, Dickinson MH (2003) The Aerodynamics of Free-Flight Maneuvers in Drosophila. *Science* 300:495–498 Publisher: American Association for the Advancement of Science.

Fujiwara T, Brotas M, Chiappe ME (2022) Walking strides direct rapid and flexible recruitment of visual circuits for course control in Drosophila. *Neuron* 110:2124–2138.e8 Publisher: Elsevier.

Fujiwara T, Cruz TL, Bohnslav JP, Chiappe ME (2017) A faithful internal representation of walking movements in the Drosophila visual system. *Nature Neuroscience* 20:72–81.

Geurten BRH, Jahde P, Corthals K, Gopfert MC (2014) Saccadic body turns in walking Drosophila. *Frontiers in Behavioral Neuroscience* 8.

Hardcastle BJ, Krapp HG (2016) Evolution of Biological Image Stabilization. *Current Biology* 26:R1010–R1021.

Hindmarsh Sten T, Li R, Otopalik A, Ruta V (2021) Sexual arousal gates visual processing during Drosophila courtship. *Nature* 595:549–553 Number: 7868 Publisher: Nature Publishing Group.

Juusola M, Dau A, Song Z, Solanki N, Rien D, Jaciuch D, Dongre SA, Blanchard F, de Polavieja GG, Hardie RC, Takalo J (2017) Microsaccadic sampling of moving image information provides Drosophila hyperacute vision. *eLife* 6:e26117 Publisher: eLife Sciences Publications, Ltd.

Keleş MF, Frye MA (2017) Object-Detecting Neurons in Drosophila. *Current Biology* 27:680–687.

Keleş MF, Hardcastle BJ, Städele C, Xiao Q, Frye MA (2020) Inhibitory Interactions and Columnar Inputs to an Object Motion Detector in Drosophila. *Cell Reports* 30:2115–2124.e5.

Kerschensteiner D (2022) Feature Detection by Retinal Ganglion Cells. *Annual Review of Vision Science* 8:annurev–vision–100419–112009.

Kim AJ, Fenk LM, Lyu C, Maimon G (2017) Quantitative Predictions Orchestrate Visual Signaling in Drosophila. *Cell* 168:280–294.e12.

Kim AJ, Fitzgerald JK, Maimon G (2015) Cellular evidence for efference copy in Drosophila visuomotor processing. *Nature Neuroscience* 18:1247–1255.

Klapoetke NC, Nern A, Peek MY, Rogers EM, Breads P, Rubin GM, Reiser MB, Card GM (2017) Ultra-selective looming detection from radial motion opponency. *Nature* 551:237–241.

Klapoetke NC, Nern A, Rogers EM, Rubin GM, Reiser MB, Card GM (2022) A functionally ordered visual feature map in the Drosophila brain. *Neuron* p. S0896627322001787.

Kohn JR, Portes JP, Christenson MP, Abbott LF, Behnia R (2021) Flexible filtering by neural inputs supports motion computation across states and stimuli. *Current Biology* 31:5249–5260.e5.

Land M (1999) Motion and vision: Why animals move their eyes. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology* 185:341–352.

Land M (2019) Eye movements in man and other animals. *Vision Research* 162:1–7.

Leong JCS, Esch JJ, Poole B, Ganguli S, Clandinin TR (2016) Direction Selectivity in Drosophila Emerges from Preferred-Direction Enhancement and Null-Direction Suppression. *Journal of Neuroscience* 36:8078–8092 Publisher: Society for Neuroscience Section: Articles.

Lettvin JY, Maturana HR, McCulloch WS, Pitts WH (1959) What the Frog's Eye Tells the Frog's Brain. *Proceedings of the IRE* 47:1940–1951 Conference Name: Proceedings of the IRE.

Li CH, Lee CK (1993) Minimum cross entropy thresholding. *Pattern Recognition* 26:617–625.

Maimon G (2011) Modulation of visual physiology by behavioral state in monkeys, mice, and flies. *Current Opinion in Neurobiology* 21:559–564.

Maimon G, Straw AD, Dickinson MH (2010) Active flight increases the gain of visual motion processing in Drosophila. *Nature Neuroscience* 13:393–399 Number: 3 Publisher: Nature Publishing Group.

Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, Schilling T, Bahl A, Rubin GM, Nern A, Dickson BJ, Reiff DF, Hopp E, Borst A (2013) A directional tuning map of Drosophila elementary motion detectors. *Nature* 500:212–216.

Mann K, Gallen CL, Clandinin TR (2017) Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in Drosophila. *Current Biology* 27:2389–2396.e4.

Martinez-Conde S, Otero-Millan J, Macknik SL (2013) The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience* 14:83–96 Number: 2 Publisher: Nature Publishing Group.

McAdams CJ, Maunsell JHR (1999) Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *Journal of Neuroscience* 19:431–441 Publisher: Society for Neuroscience Section: ARTICLE.

McBride EG, Lee SYJ, Callaway EM (2019) Local and Global Influences of Visual Spatial Selection and Locomotion in Mouse Primary Visual Cortex. *Current Biology* 29:1592–1605.e5.

Moore RJD, Taylor GJ, Paulk AC, Pearson T, van Swinderen B, Srinivasan MV (2014) FicTrac: A visual method for tracking spherical motion and generating fictive animal paths. *Journal of Neuroscience Methods* 225:106–119.

Niell CM, Stryker MP (2010) Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* 65:472–479.

Otsuna H, Ito K (2006) Systematic analysis of the visual projection neurons ofDrosophila melanogaster. I. Lobula-specific pathways. *The Journal of Comparative Neurology* 497:928–958.

Pacheco DA, Thiberge SY, Pnevmatikakis E, Murthy M (2021) Auditory activity is diverse and widespread throughout the central brain of Drosophila. *Nature Neuroscience* 24:93–104 Number: 1 Publisher: Nature Publishing Group.

Pasupathy A, Connor CE (2001) Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation. *Journal of Neurophysiology* 86:2505–2519.

Pfeiffer BD, Ngo TTB, Hibbard KL, Murphy C, Jenett A, Truman JW, Rubin GM (2010) Refinement of tools for targeted gene expression in Drosophila. *Genetics* 186:735–755.

Piscopo DM, El-Danaf RN, Huberman AD, Niell CM (2013) Diverse Visual Features Encoded in Mouse Lateral Geniculate Nucleus. *Journal of Neuroscience* 33:4642–4656.

Rabinowitz NC, Goris RL, Cohen M, Simoncelli EP (2015) Attention stabilizes the shared gain of V4 populations. *eLife* 4:e08998.

Reynolds AM, Frye MA (2007) Free-Flight Odor Tracking in Drosophila Is Consistent with an Optimal Intermittent Scale-Free Search. *PLoS ONE* 2:e354.

Ribeiro IM, Drews M, Bahl A, Machacek C, Borst A, Dickson BJ (2018) Visual Projection Neurons Mediating Directed Courtship in Drosophila. *Cell* 174:607–621.e18.

Romo R, Hernández A, Zainos A, Salinas E (2003) Correlated Neuronal Discharges that Increase Coding Efficiency during Perceptual Discrimination. *Neuron* 38:649–657.

Rucci M, Victor JD (2015) The unsteady eye: an information-processing stage, not a bug. *Trends in Neurosciences* 38:195–206.

Schaffer ES, Mishra N, Whiteway MR, Li W, Vancura MB, Freedman J, Patel KB, Voleti V, Paninski L, Hillman EMC, Abbott LF, Axel R (2021) Flygenvectors: The spatial and temporal structure of neural activity across the fly brain Pages: 2021.09.25.461804 Section: New Results.

Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura Sy, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S, Clements J, Hubbard PM, Katz WT, Umayam L, Zhao T, Ackerman D, Blakely T, Bogovic J, Dolafi T, Kainmueller D, Kawase T, Khairy KA, Leavitt L, Li PH, Lindsey L, Neubarth N, Olbris DJ, Otsuna H, Trautman ET, Ito M, Bates AS, Goldammer J, Wolff T, Svirskas R, Schlegel P, Neace E, Knecht CJ, Alvarado CX, Bailey DA, Ballinger S, Borycz JA, Canino BS, Cheatham N, Cook M, Dreher M, Duclos O, Eubanks B, Fairbanks K, Finley S, Forknall N, Francis A, Hopkins GP, Joyce EM, Kim S, Kirk NA, Kovalyak J, Lauchie S, Lohff A,

Maldonado C, Manley EA, McLin S, Mooney C, Ndama M, Ogundeyi O, Okeoma N, Ordish C, Padilla N, Patrick CM, Paterson T, Phillips EE, Phillips EM, Rampally N, Ribeiro C, Robertson MK, Rymer JT, Ryan SM, Sammons M, Scott AK, Scott AL, Shinomiya A, Smith C, Smith K, Smith NL, Sobeski MA, Suleiman A, Swift J, Takemura S, Talebi I, Tarnogorska D, Tenshaw E, Tokhi T, Walsh JJ, Yang T, Horne JA, Li F, Parekh R, Rivlin PK, Jayaraman V, Costa M, Jefferis GS, Ito K, Saalfeld S, George R, Meinertzhagen I, Rubin GM, Hess HF, Jain V, Plaza SM (2020) A connectome and analysis of the adult Drosophila central brain. *eLife* 9:e57443.

Strother JA, Wu ST, Rogers EM, Eliason JLM, Wong AM, Nern A, Reiser MB (2018) Behavioral state modulates the ON visual motion pathway of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 115:E102–E111.

Städele C, Keleş MF, Mongeau JM, Frye MA (2020) Non-canonical Receptive Field Properties and Neuromodulation of Feature-Detecting Neurons in Flies. *Current Biology* .

Suver MP, Mamiya A, Dickinson MH (2012) Octopamine Neurons Mediate Flight-Induced Modulation of Visual Processing in Drosophila. *Current Biology* 22:2294–2302.

Tanaka R, Clark DA (2020) Object-Displacement-Sensitive Visual Neurons Drive Freezing in Drosophila. *Current Biology* .

Tanaka R, Clark DA (2022) Neural mechanisms to exploit positional geometry for collision avoidance. *Current Biology* p. S0960982222005929.

Thiele A, Henning P, Kubischik M, Hoffmann KP (2002) Neural mechanisms of saccadic suppression. *Science (New York, N.Y.)* 295:2460–2462.

Turner MH, Mann K, Clandinin TR (2021) The connectome predicts resting-state functional connectivity across the Drosophila brain. *Current Biology* 31:2386–2394.e3.

Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, Duda JT, Das SR, Cullen NC, Gillen DL, Yassa MA, Stone JR, Gee JC, Avants BB (2021) The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports* 11:9068.

Van Der Linde I, Rajashekar U, Bovik AC, Cormack LK (2009) DOVES: a database of visual eye movements. *Spatial Vision* 22:161–177.

van Hateren JH, van der Schaaf A (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265:359–366.

Walls GL (1962) The evolutionary history of eye movements. *Vision Research* 2:69–80.

Wu M, Nern A, Williamson WR, Morimoto MM, Reiser MB, Card GM, Rubin GM (2016) Visual projection neurons in the Drosophila lobula link feature detection to distinct behavioral programs. *eLife* 5:e21022.

Wurtz RH (2018) Corollary Discharge Contributions to Perceptual Continuity Across Saccades. *Annual Review of Vision Science* 4:215–237 _eprint: https://doi.org/10.1146/annurev-vision-102016-061207.

York RA, Brezovec LE, Coughlan J, Herbst S, Krieger A, Lee SY, Pratt B, Smart AD, Song E, Suvorov A, Matute DR, Tuthill JC, Clandinin TR (2022) The evolutionary trajectory of drosophilid walking. *Current Biology* .

Zuber BL, Stark L, Cook G (1965) Microsaccades and the Velocity-Amplitude Relationship for Saccadic Eye Movements. *Science* 150:1459–1460 Publisher: American Association for the Advancement of Science.

Zylberberg J, Cafaro J, Turner MH, Shea-Brown E, Rieke F (2016) Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code. *Neuron* 89:369–383.

Ölveczky BP, Baccus SA, Meister M (2003) Segregation of object and background motion in the retina. *Nature* 423:401–408 Number: 6938 Publisher: Nature Publishing Group.

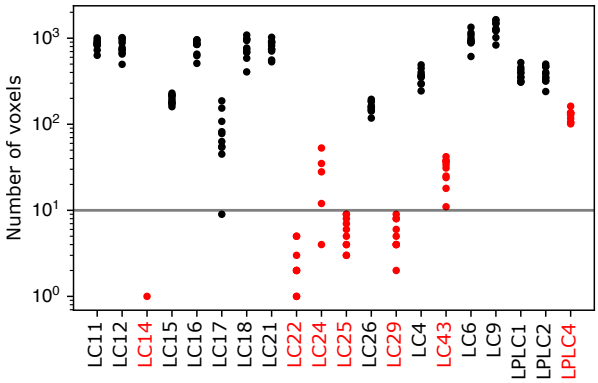

Figure S1: Number of voxels in each LC/LPLC glomerulus For each LC/LPLC glomerulus in or near the imaging volume, the number of voxels in each transformed glomerulus mask that lies within the imaging volume is shown for 10 flies. Glomeruli highlighted in red were excluded from this analysis because they were consistently small, or often not included in the volume. The LPLC4 glomerulus was excluded because this glomerulus is co-extensive with the LC22 glomerulus. The horizontal line indicates a threshold we set for further excluding glomeruli from individual flies for being small in that fly's imaging volume. For example, the LC17 glomerulus was sometimes excluded from individual flies because it lies at the bottom of our imaging volume, and brain motion or slight changes in the imaging depth caused the imaged portion of the LC17 glomerulus to be very small.

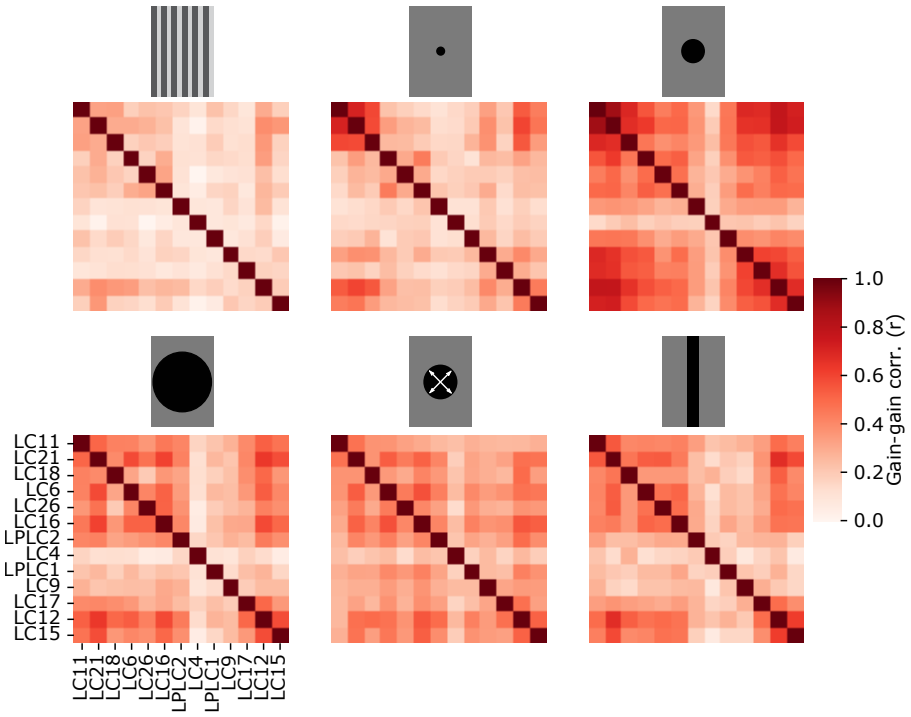

Figure S2: Trial correlation matrix for each stimulus class From the trial covariance analysis in Fig. 4, each panel shows the trial correlation matrix for each of the six stimuli shown in the reduced stimulus set for these experiments. Positive correlations are seen for all stimuli, but tend to be higher for stimuli that drive glomerulus responses strongly.

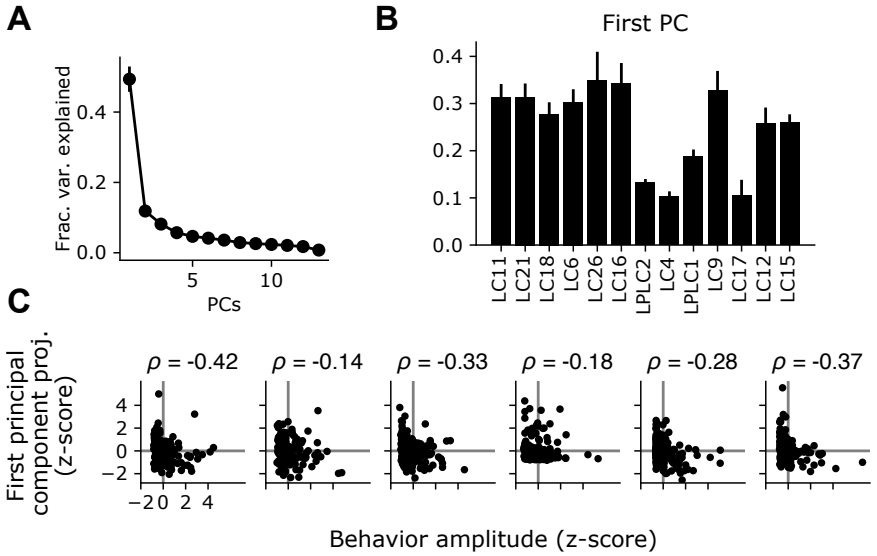

Figure S3: The shared glomerulus gain factor is negatively correlated with behavior. (A) Eigenvector decomposition of the covariance matrix reveals a dominant principal component which accounts for a large plurality of the trial-to-trial variance across the population. Points show mean +/- S.E.M. across flies (n=6 flies). (B) The first principal component of the response gain, averaged across flies. (C) For the variance analysis experiments in Fig. 4, we tracked behavior for a subset of the animals tested. For each fly, we z-scored the behavior amplitude across trials as well as the shared gain factor for each trial (equal to the projection of the first PC against the population response matrix). Each plot shows the relationship between the trial-by-trial shared gain factor and the average behavior amplitude for each trial. For every fly, there was a negative correlation between these two measures (mean ρ = -0.23, n=6 flies).

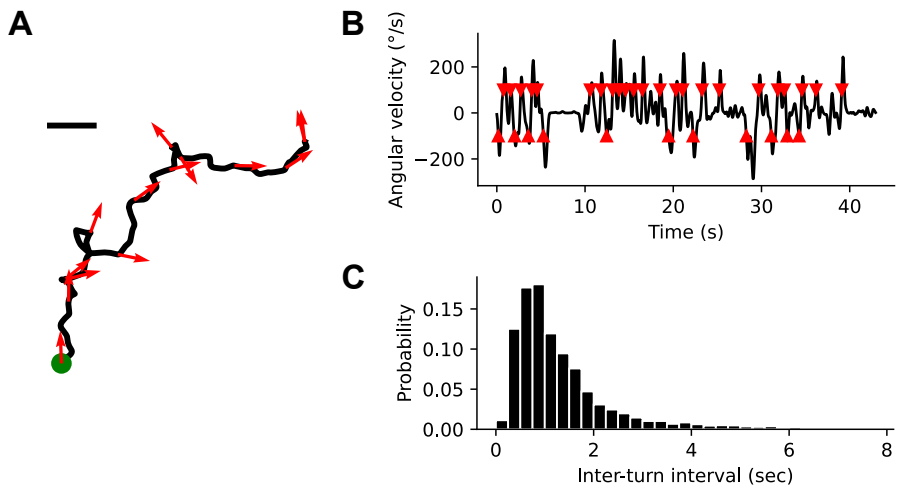

Figure S4: Natural fly walking consists of punctuated saccadic turns. (A) Example fly walking trajectory measured in a one square meter arena. Arrows show the fly's heading. Green point indicates start of walking trajectory. Scale bar = 1 cm. (B) Instantaneous angular velocity for the example walking trajectory in (A). Red arrowheads indicate turn events. (C) Histogram of inter-turn intervals shows a peak around 1 second and a long tail.