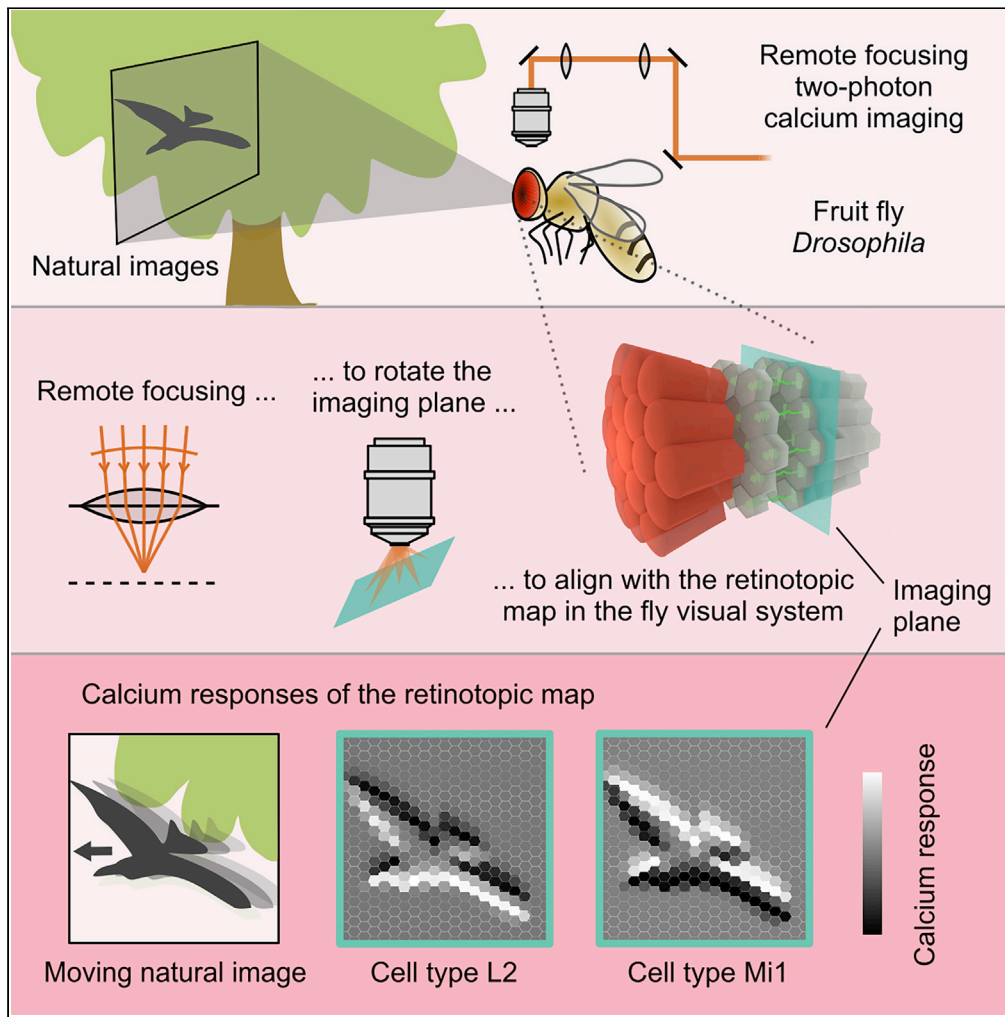**Article**

# Seeing Natural Images through the Eye of a Fly with Remote Focusing Two-Photon Microscopy

Anna Schuetzenberger, Alexander Borst

schuetzenberger@neuro.mpg.de (A.S.)
borst@neuro.mpg.de (A.B.)

HIGHLIGHTS

A remote-focusing two-photon module for real-time rotation of the imaging plane

Simultaneously recording >100 cells of the 2D retinotopic map in *Drosophila*

Calcium imaging of L2 and Mi1 shows their dynamic representation of natural images

Linear properties of L2 and Mi1 largely predict dynamic natural scene responses

# iScience

**Article**

# Seeing Natural Images through the Eye of a Fly with Remote Focusing Two-Photon Microscopy

Anna Schuetzenberger[1,2,3,*] and Alexander Borst[1,2,*]

## SUMMARY

**Visual systems of many animals, including the fruit fly *Drosophila*, represent the surrounding space as 2D maps, formed by populations of neurons. Advanced genetic tools make the fly visual system especially well accessible. However, in typical *in vivo* preparations for two-photon calcium imaging, relatively few neurons can be recorded at the same time. Here, we present an extension to a conventional two-photon microscope, based on remote focusing, which enables real-time rotation of the imaging plane, and thus flexible alignment to cellular structures, without resolution or speed trade-off. We simultaneously record from over 100 neighboring cells spanning the 2D retinotopic map. We characterize its representation of moving natural images, which we find is comparable to noise predictions. Our method increases throughput 10-fold and allows us to visualize a significant fraction of the fly's visual field. Furthermore, our system can be applied in general for a more flexible investigation of neural circuits.**

## INTRODUCTION

Vision is a crucial sense for many animals, including humans. Through vision, we gather abundant information about our surroundings, which is processed in the visual system to extract relevant features that then guide behavioral choices. In the initial stages of processing, visual information is represented as a set of 2D maps of the surrounding space (Chklovskii and Koulakov, 2004; Land and Fernald, 1992). In the fruit fly *Drosophila melanogaster*, the visual system collects information through 750 ommatidia (Ready et al., 1976), each containing a lens and eight photoreceptors. Correspondingly, the subsequent neuropils, lamina, medulla, lobula, and lobula plate are each composed of 750 columns. In the lamina, photoreceptor inputs from neighboring ommatidia converge such that one lamina column receives information from one point in visual space ("neural superposition principle," Braitenberg, 1967; Kirschfeld, 1967). Neighboring columns process information from neighboring points in space (Land, 1997). Together, they thus retinotopically map the surrounding visual scene. Accordingly, many cell types, which code for a specific visual feature like local contrast or motion, come in 750 unicolumnar copies (for recent reviews, see, e.g., Mauss et al., 2017; Song and Lee, 2018).

Likewise, the vertebrate retina is composed of several layers of neurons that form retinotopic maps, each of which processes certain visual features (for a comparative review, see, e.g., Borst and Helmstaedter, 2015; Sanes and Zipursky, 2010). Importantly, due to the retina's anatomical separation from the rest of the brain, it can be mounted on a flat surface and the retinotopic array of neurons can be recorded from by means of microelectrode arrays or two-photon calcium imaging (Baden et al., 2016; Denk and Detwiler, 1999; Meister et al., 1994; Pillow et al., 2008; Segev et al., 2004). This ease of access has probably made the retina one of the best-understood neural circuits (for reviews, see Demb and Singer, 2015; Diamond, 2017; Dowling, 2012).

The fruit fly poses many advantages for the study of neural circuits. Morphologically and genetically defined cell types, stereotyped anatomy, and refined genetic tools for targeted expression, recording, and manipulation, together with two-photon imaging, resulted in an unparalleled understanding of the *Drosophila* visual system (for reviews, see Aptekar and Frye, 2013; Mauss et al., 2017; Song and Lee, 2018). However, due to the differences in their anatomy, the retinotopic maps cannot be recorded from as in the mammalian retina. The retinotopic plane in the *Drosophila* visual system, as typically mounted for *in vivo* calcium imaging and visual stimulation, is an arbitrarily rotated plane with respect to the imaging system
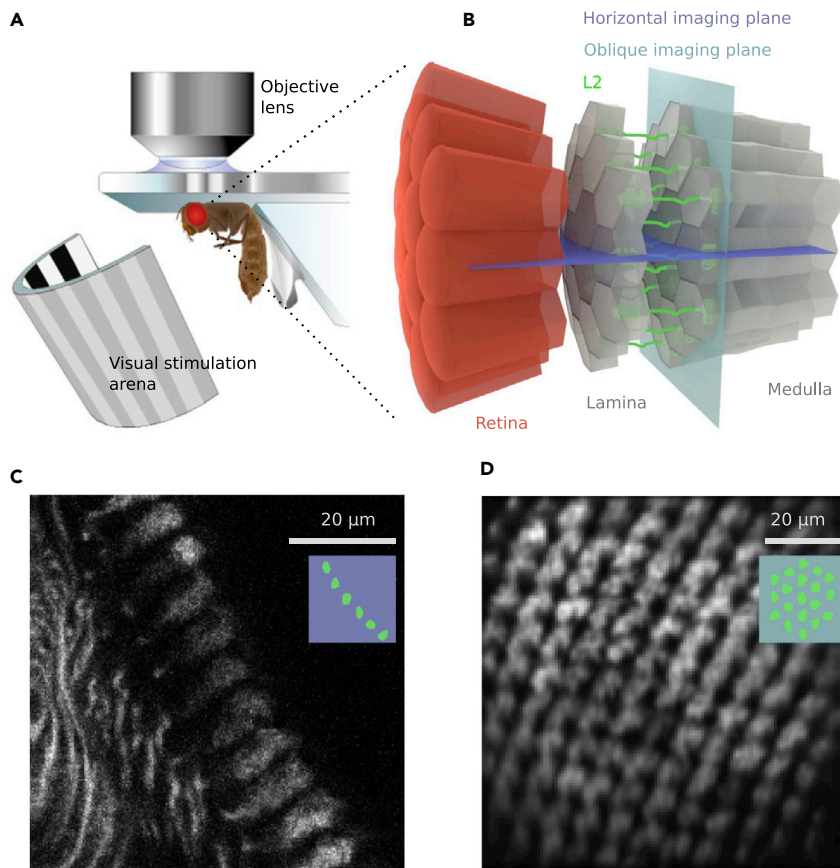
[1]Department Circuits - Computation - Models, Max-Planck-Institute of Neurobiology, 82152 Planegg, Germany

[2]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, 82152 Planegg, Germany

[3]Lead Contact

*Correspondence: schuetzenberger@neuro.mpg.de (A.S.), borst@neuro.mpg.de (A.B.)

**Figure 1. Imaging the *Drosophila* Visual System**

(A) Setup for *in vivo* calcium imaging of the *Drosophila* visual system during visual stimulation. The fly is fixed in a holder while the brain is viewed from the dorso-posterior side. Visual stimuli are shown on a cylindrical screen surrounding the fly. Schematic not to scale.

(B) The visual system of *Drosophila* consists of retinotopically arranged layers of neuropil. The horizontal imaging plane (purple) sections the visual system in a particular orientation. An oblique imaging plane (turquoise) can be aligned to the retinotopic plane of the visual system. The axon terminals of the cell type L2 lie in layer 2 of the medulla. For simplicity, the chiasm between lamina and medulla is omitted in this schematic.

(C) Horizontal plane of L2 expressing GCaMP6f.

(D) Oblique, retinotopic plane of axon terminals of L2 in the medulla layer 2 expressing GCaMP6f.

(Figures 1A and 1B, Reiff et al., 2010; Seelig et al., 2010). Its precise orientation depends on the specifics of the mounting and setup arrangement. With a scanning two-photon microscope, only the cells that lie in the focal plane can be recorded from at fast scan speeds (Figure 1C). Thus, functional analysis of the visual system is restricted to few cells at a time, which limits the throughput and analysis of more complex stimuli, for which recording from a larger population would be required.

Recent advances in imaging technologies are opening doors to a more flexible investigation of neural structures (for reviews. see Ji et al., 2016; Ronzitti et al., 2018). For example, random-access two-photon microscopy using acousto-optic modulators enables fast access to sparse cell locations (Katona et al., 2012), and fast z-scanning is possible among others with electrically tunable lenses or remote focusing (Botcherby et al., 2008; Grewe et al., 2011). Employing the remote-focusing principle, the focal spot of the laser is moved fast along the optical axis by means of moving a third, remote mirror (Botcherby et al., 2008). When certain criteria are met, imaging can be aberration free in a large volume (Botcherby et al., 2007, 2012). Remote focusing has primarily been applied to skip to several horizontal planes (parallel to the focal plane) with fast imaging rates (Rupprecht et al., 2016; Sofroniew et al., 2016). In principle, however, it should enable other scanning trajectories as well. Here, we present a near-aberration-free remote-focusing module, which can be added onto conventional two-photon microscopes

and which can be controlled by an extension to the open source software ScanImage 5.1 (Pologruto et al., 2003). Our software extension facilitates online rotation of the imaging plane in 3D in any arbitrary direction. The plane can thus be aligned with cellular structures of interest during the imaging session, and oblique planes can be imaged without a trade-off in resolution or speed.

We apply this method to simultaneously image over 100 axon terminals in the retinotopic plane of the uni-columnar cell types L2 (Figure 1D) and Mi1 and thus reconstitute the 2D representation of the visual scene as represented by these cell types. Both cell types play a crucial part in the circuit for motion vision (Ammer et al., 2015; Joesch et al., 2010; Strother et al., 2017; Tuthill et al., 2013). L2 receives direct input from photoreceptors (Rivera-Alba et al., 2011) and codes for local light decrements when probed with artificial stimuli (Clark et al., 2011; Freifeld et al., 2013; Reiff et al., 2010). Mi1 codes for local light increments (Arenz et al., 2017; Behnia et al., 2014) and provides output to direction-selective T4 cells (Strother et al., 2017; Takemura et al., 2017). Here, we characterize the 2D population representation of natural images by these cell types. Based on their functional properties derived from experiments with artificial stimuli we expect that they code for local luminance changes of natural scenes.

## RESULTS

### Imaging Oblique Planes with Remote Focusing Two-Photon Microscopy

We first designed a remote-focusing module that is easily integrated into a typical two-photon microscope and provides near-homogenous resolution throughout a large volume (Figure 2A). The remote-focusing principle is based on creating a beam with a convergent or divergent waveform, which, upon passing through a lens, will focus before or beyond the lens' focal plane, respectively (Figure 2B, ; Botcherby et al., 2008). To create such a beam and to easily modulate the beam vergence, we introduced a second objective lens (remote objective lens), and a mirror (Z-mirror, Figure 2A inset, Figure 2B, left), moveable with a piezo motor. The collimated laser beam passes through the remote lens, is reflected by the z-mirror, and is projected back through the lens. Depending on the position of the z-mirror relative to the focal plane of the lens, the resulting waveform is convergent or divergent (Figure 2B). To image the back aperture of the remote lens onto the back focal plane of the imaging objective lens, we used two telescopes (Figures 2A and 2C), with the intermediate conjugate plane at the lateral scan unit. The magnifications of the telescopes were matched such that the overall magnification between the remote and imaging lens was close to the ratio between the refractive indices of water and air, $n_2/n_1$ (for details, see Transparent Methods). This way, the aberrations deliberately introduced by the remote lens are to a large extent canceled out by the imaging objective lens (Botcherby et al., 2008). As a result, the optical resolution is near-diffraction-limited throughout the imaging volume, and the relationship between mirror position and position of the laser focal spot in z is linear (Botcherby et al., 2012).

To quantify the optical resolution across the imaging volume in our setup, we measured the point spread function of fluorescent beads at different positions in a volume of 100 $\mu m^3$. The lateral resolution, defined as the full width at half maximum of the bead point spread function, lies below 0.75 $\mu m$ at all positions in the volume (Figures 2D, 2E, S1, and S2), close to the theoretical estimate (for details see Transparent Methods). The axial resolution is on average 3.6 $\mu m$ (ranging from 3 to 5 $\mu m$ at all positions in the volume; Figures 2D, 2E, S1, and S2). The mirror movement causes near-linear displacement of the focal spot in the axial direction (Figure S1I) across 90 $\mu m$.

To utilize the remote-focusing module for imaging of oblique planes, we extended the functionality of the open source software ScanImage 5.1 (Pologruto et al., 2003) to control the movement of the motorized z-mirror. In ScanImage, the scanned imaging plane is represented as a sequence of 2D vectors, each dimension corresponding to the movement of one mirror, the lateral scan mirrors (x- and y-mirrors). We extended this representation to a third dimension, reflecting the movement of the z-mirror. Applying geometric operations, the imaging plane can then be rotated and translated arbitrarily in 3D, including simultaneous rotations about several axes. Finally, additional user controls in the ScanImage GUI enable plane rotation in real time, such that the user can align the plane to cellular structures during imaging.

A custom-build piezo motor was used to move the z-mirror (for details see Transparent Methods). Together with two galvanometer motors moving the mirrors in the lateral scan unit, this enabled imaging of arbitrarily rotated, oblique planes of size 90 × 90 $\mu m$ at frame rates of up to 15 Hz, or smaller planes (50 × 50 $\mu m$) at frame rates up to 20 Hz. These frame rates are typical for recordings of the visual system in *Drosophila* with fluorescent indicators.
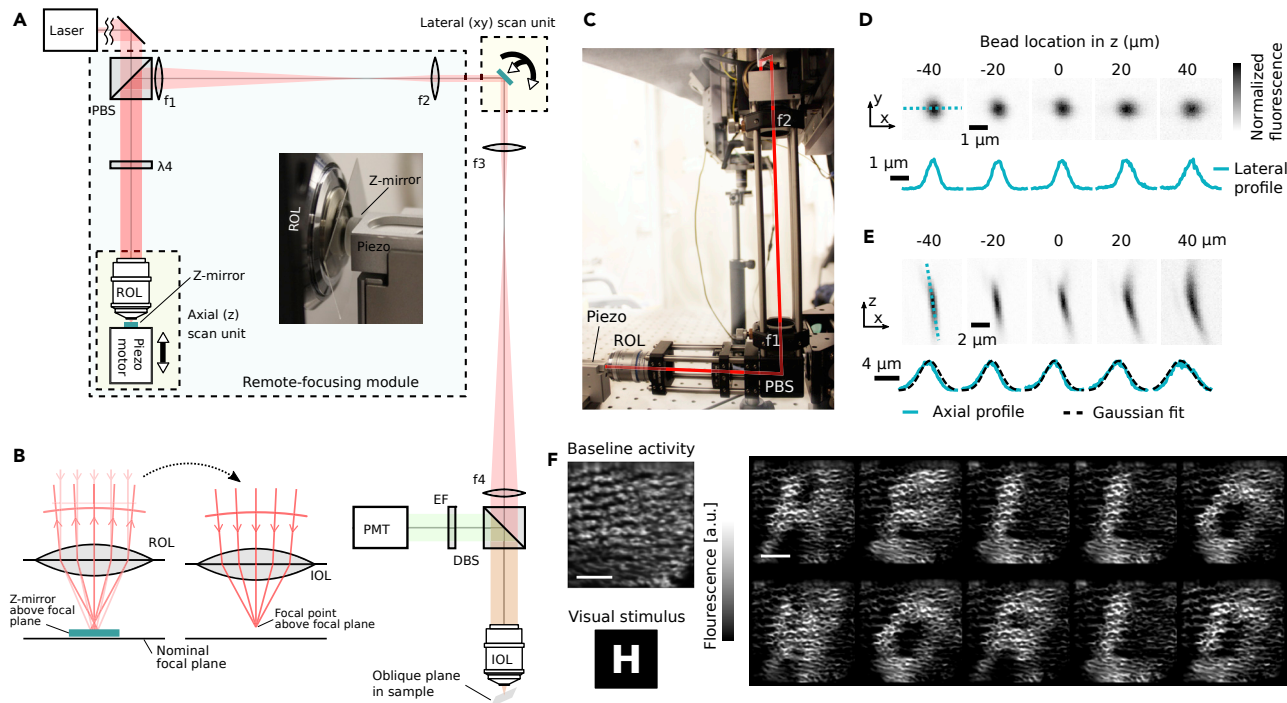
**Figure 2. Remote Focusing Two-Photon Microscope**

(A) The remote-focusing module in a standard two-photon microscope. A polarization-based beam splitter (PBS) and λ/4 plate guide the beam to the axial scan unit, which is composed of the remote objective lens (ROL) and the piezo-motorized z-mirror. Two lenses (f1 and f2, focal lengths of 125 and 40 mm, respectively) in 4f configuration relay the beam from the back aperture of the ROL to the lateral scan unit. Scan and tube lens (f3 and f4, focal lengths of 50 and 200 mm, respectively) then relay the beam to the imaging objective lens (IOL). Inset: Photograph of the axial scan unit.

(B) Schematic of the beam waveform at the ROL and IOL. A collimated beam entering the ROL is reflected by the z-mirror. Depending on the mirror position relative to the nominal focal plane, the exiting beam at the back focal plane of the ROL will have non-zero vergence. The beam relayed to the IOL results in a focal spot outside the nominal focal plane of the IOL.

(C) Photograph of the remote-focusing module.

(D) Point spread function (PSF) of 0.1 μm-diameter fluorescent beads at five different positions in z in their lateral (x-y) profile. Top: Bead images, average of three measurements. Bottom: 1D profiles along the x axis.

(E) As in (D) for the axial (x-z) profile of the PSF. Top: Bead images, average of three measurements. Bottom: 1D profiles along a tilted axis corresponding to the widest point of the PSF, determined by fitting a 2D Gaussian to the bead images. 1D Gaussian fit along the tilted axis in dashed black lines.

(F) Calcium imaging of the retinotopic plane of Mi1 axon terminals in layer 10 of the medulla. Top left: Calcium baseline level without a stimulus. Bottom left: A letter is displayed on the stimulus screen centered on the receptive fields of the cells in view. Right: Fluorescence increase in response to different letters. Scale bar 20 μm.

To demonstrate the applicability of our method, we performed calcium imaging experiments from the medulla interneuron Mi1. We expressed the calcium indicator GCaMP6f in Mi1 and aligned the oblique imaging plane to the retinotopic map spanned by the cells' axon terminals, viewing the 2D array of axon terminals in the imaging plane (Figure 2F, top left). We then presented different letters in the fly's field of view, each for 1 s (Figure 2F, bottom left). Cells with a receptive field within the pixels of the letter responded with an increase in calcium levels (Figure 2F, right). As a result, the activity map of Mi1 clearly reflects the respective letter shown to the fly.

With our method, we can simultaneously image up to 100 axon terminals in a single field of view, which is a substantial fraction of the fly's visual field. Moreover, this constitutes a 10-fold increase in throughput compared with imaging in the horizontal plane, without compromising spatial resolution or scan speed.

### Linear Receptive Field Properties of L2

We next utilized this technique to map the 2D arrangement of receptive fields of the lamina cell type L2. This will subsequently allow us to assign each cell to the location in space it is centered at, and to compare its linear response properties to its responses to natural scene stimulation. To this end, we expressed the calcium sensor GCaMP6f in L2 and aligned the imaging plane with the retinotopic plane spanned by L2 axon terminals
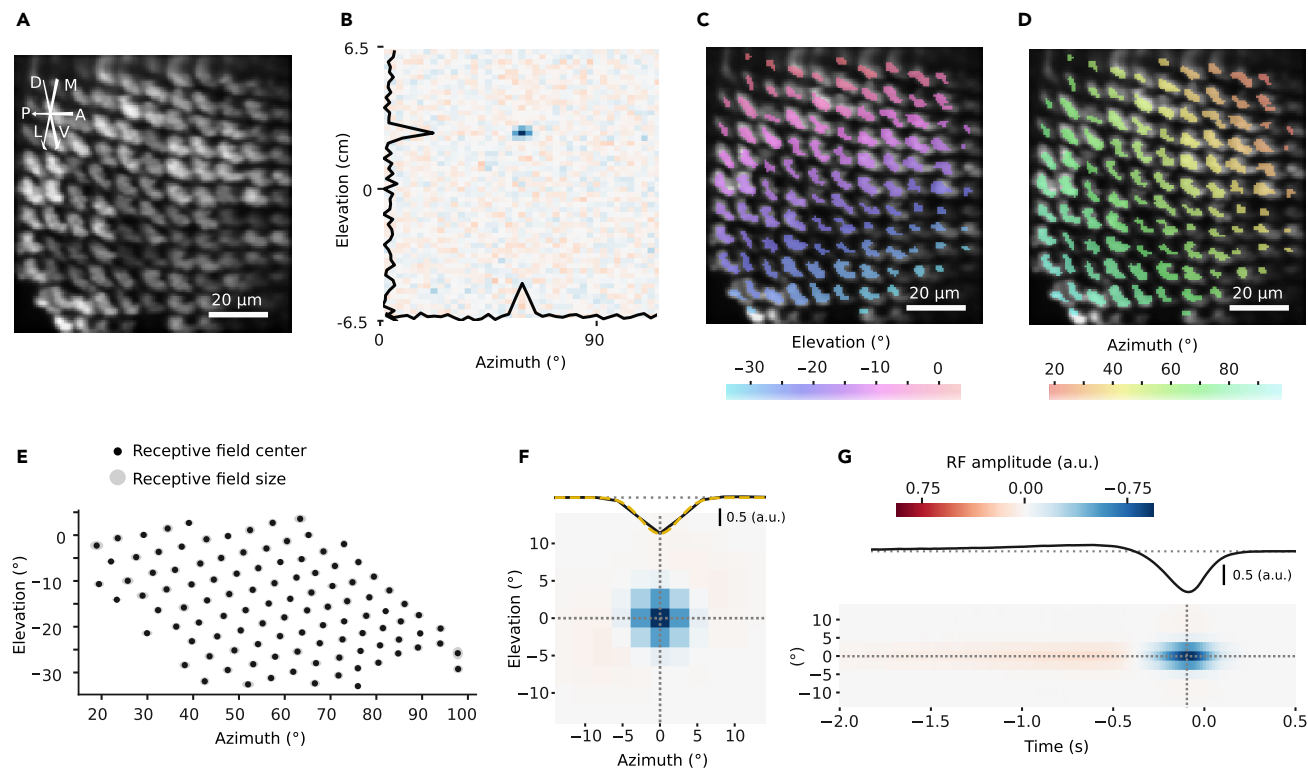
**Figure 3. Receptive Field Analysis of L2 with White Noise**

(A) Mean image of a recording of L2 axon terminals in layer 2 of the medulla expressing GCaMP6f. Approximate orientation in the fly brain as indicated with dorsoventral (D-V), anterior-posterior (A-P), and medial-lateral (M-L) axes.

(B) Reverse correlation of the white noise stimulus with the calcium signal of an example cell. The heatmap shows a cross section of the receptive field at the time point with the absolute peak amplitude. Red corresponds to a positive correlation with the stimulus luminance, and blue corresponds to a negative correlation (a.u.). Black traces show 1D cross sections.

(C) Selected regions of interest encompassing single axon terminals colored by their receptive field location in elevation.

(D) Axon terminals colored by their receptive field location in azimuth.

(E) Receptive field locations (black dots) with receptive field size (sigma of Gaussian fit, gray shaded areas) of all cells in the example recording also shown in (A, C, and D).

(F) Average spatial receptive field of all cells (N = 12 flies, n = 711 cells). The profile in azimuth (black line) was fitted with a Gaussian (dashed yellow line).

(G) 1D and 2D temporal profiles of the receptive field.

(Figure 3A). Each cell in the imaging field of view was easily distinguishable due to its compact axon terminal. We then showed a white noise stimulus (for details see Transparent Methods and Arenz et al., 2017) to the fly while recording fluorescence signals. From each terminal, we extracted the calcium signal and reverse-correlated it to the stimulus, thus obtaining its spatiotemporal receptive field (Dayan and Abbott, 2005). The receptive field location of a cell, the point in visual space it responds to, is then given by the coordinates of the maximum absolute correlation (Figure 3B). The hexagonal and retinotopic coverage of visual space by L2 is visualized in Figures 3C–3E and reflects the ommatidial layout of the *Drosophila* eye (Buchner, 1971). As expected from previous studies that probed L2 response properties with flashes, gratings, and noise (Drews et al., 2020; Freifeld et al., 2013; Reiff et al., 2010), L2 responded to light decreases with an increase in calcium levels, as is shown by the negative receptive field center (Figure 3F, N = 12 flies, n = 711 cells). In addition, the spatial receptive field showed an aniso-tropic, antagonistic surround (Figures 3F and S4), which has been suggested by Freifeld et al. (2013). In the temporal domain, L2 showed band-pass characteristics (Figure 3G). The receptive field size was on average 1.91 ± 0.24° (σ of a Gaussian fit, corresponding to 4.51 ± 0.57° full width at half maximum), and we did not observe systematic differences across the eye.

## L2 Responses to Natural Scenes Encode Luminance

We next asked how a population of L2 cells responds to natural images, and whether local image luminance can account for its responses. To this end, we showed natural images to the fly while recording from the
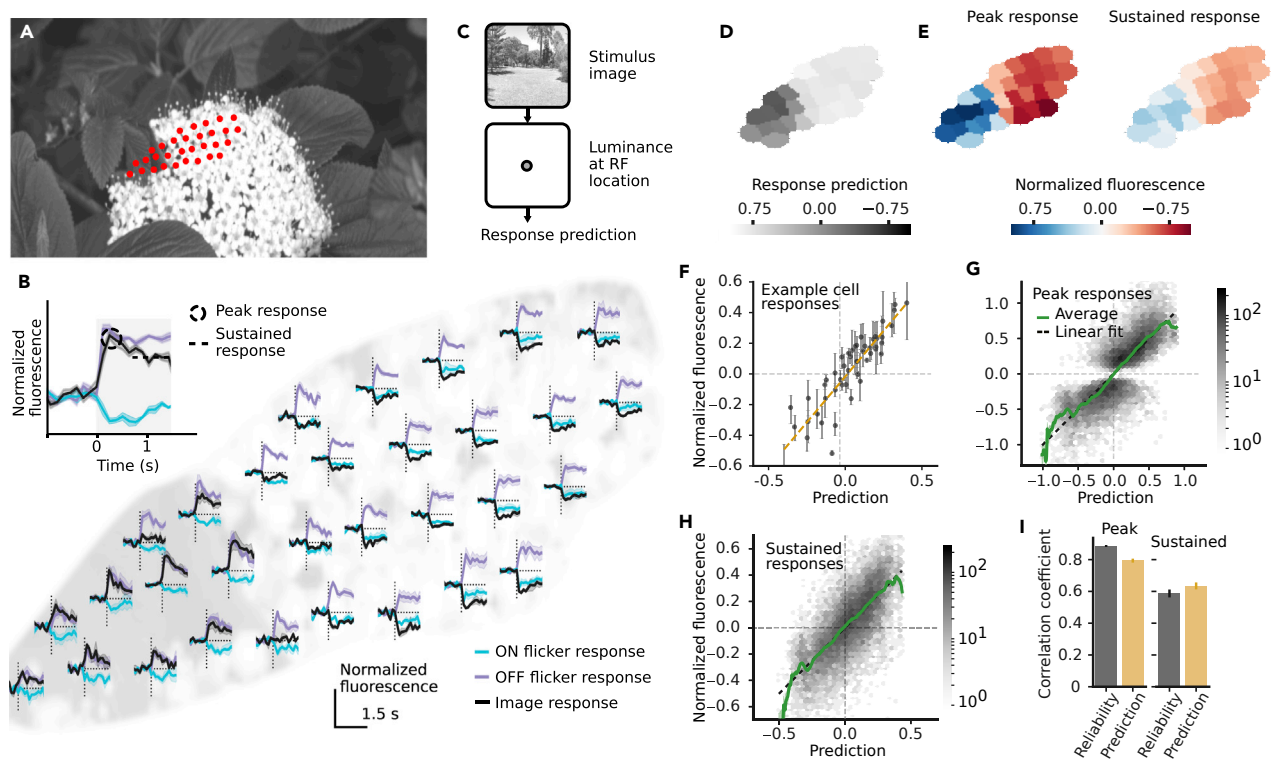
**Figure 4. L2 Responses to Natural Scenes**

(A) Example image with the receptive field location of all cells in an example recording (red circles).

(B) Top left: Example response of a cell to ON and OFF flicker stimuli and an image. Average over three trials. Shaded areas indicate standard deviation. The gray shaded box indicates the time during which the image was shown. Remainder of figure: Example responses of all cells in an example recording (same example as in A). The background shows the image stimulus, and individual plots are positioned on the corresponding receptive field locations of each cell.

(C) Schematic of the model. Stimulus images were first blurred (filtered with a 2D Gaussian kernel, σ = 1.91°). The luminance values of the filtered images at the receptive field coordinates of each cell then constituted the prediction. Prediction values were scaled for each cell to fit the normalized fluorescence values of the data.

(D) Response predictions of the example recording in (A) and (B).

(E) Left: Peak response of all cells in the example recording. The peak response was defined as the largest absolute response during the time the image was shown. Right: Sustained responses of all cells in the example recording. The sustained response was defined as the average response over the last second during which the image was shown.

(F) Sustained responses of an example cell (cell #13) to all 50 images, plotted against the model predictions. Error bars denote the lowest and highest response out of three trials (range). Dashed yellow line represents the linear fit.

(G) Peak responses in dependency of the response prediction (N = 8 flies, n = 290 cells). Hexagon color represents the number of data points in each hexagon. Green line represents the average; shaded area indicates 95% confidence interval (over binned data). Dashed black line represents linear fit.

(H) Sustained responses as for peak responses in (G).

(I) Correlation between trials (black) and between prediction and data (yellow), expressed as the Pearson's correlation coefficient r. Error bars denote 95% confidence intervals.

retinotopic plane of L2 axon terminals. The stimulus set consisted of 50 images from the van Hateren Image Database (van Hateren and van der Schaaf, 1998), featuring a variety of natural environments. Each image was displayed for 1.5 s, following a gray screen. In addition, an ON and an OFF flicker stimulus, a completely white or black image, respectively, were included in the stimulus set. Before the natural image stimulation, receptive field coordinates of the cells were determined with white noise analysis (Figure 4A). The calcium signals from all cells were baseline subtracted and normalized (for details see Transparent Methods). Responses from three trials were averaged. The population response of L2 cells in an individual recording shows that L2 cells respond to local image luminance with a calcium increase in dark areas and decrease in bright areas of the image (Figure 4B).

To quantitatively compare the cellular response to the local luminance of the image at its receptive field center, we blurred the images with a 2D Gaussian, the standard deviation and sign of the filter given by

the receptive field size from the white noise analysis (Figure 4C). This simple model yields a prediction of the cells' responses should they solely respond according to local image luminance (Figure 4D). We evaluated two response features: the peak response was defined as the largest absolute response during the time the image was shown (Figures 4B and 4E, left) and the sustained response was defined as the average response over the last second during which the image was shown (Figures 4B and 4E, right). We then tested whether the quantified responses correlate with model predictions. We found a significant linear dependence of L2 responses to local image luminance (Figures 4F–4H, Bonferroni-corrected Student's t test: p < 0.05 for 286/290 cells). Thus, L2 encodes brightness increments with a calcium decrease and brightness decrements with a calcium increase.

L2 codes for local luminance, yet its responses might in addition depend on other features of the natural scene. Thus, we next asked whether the luminance-based model can account for all response variance. To this end, we first quantified the reliability of the responses as the Pearson's correlation coefficient r between trials (Figure 4I, for details see Transparent Methods). The reliability determines an upper bound for the explainable variance (expressed in $r^2$) by the model; the remaining variance derives from cellular and experimental noise. The model's correlation with the data was comparable to the response reliability, with only a small difference for peak responses (Figure 4I, for statistics see Table S1). Thus, the linear model can explain most of the predictable variance of the responses. We conclude that the retinotopic map of L2 cells, when probed with static natural images, largely represents their local luminance.

## L2 and Mi1 Responses to Moving Natural Scenes

L2 represents the major lamina input to the OFF pathway of motion vision (Joesch et al., 2010). Thus, its responses to moving stimuli are particularly relevant to the downstream circuitry. Therefore, we measured the dynamic representation of moving images by the 2D array of L2 axon terminals. We used 10 panoramic images (Figure 5A, Brinkworth and O'Carroll, 2009) and moved them horizontally across the screen at constant speed. Before presenting the moving images, we mapped the receptive fields of individual cells using white noise analysis as done previously. Figure 5B displays example responses of L2 to a selection of moving images (black traces). Figure 5D shows all cells' responses of an example recording at a particular time point during the recording together with the image section viewed at that time point (see also Video S1). To quantify the response properties, we generated response predictions based on three models of increasing complexity (Figure 5F). The first model reports the blurred image luminance at the receptive field center of a cell at a given time point (Luminance model, pink). This model corresponds to the model we used to predict the static natural scene responses. The second model's predictions were generated by filtering the image sequences with the 3D spatiotemporal receptive field (RF model, brown). Compared with the first model, it adds the cells' temporal properties as well as the spatial surround. The third model adds a static nonlinearity to the RF model (L-NL model, blue; for details see Transparent Methods). Exemplary model predictions are shown for the cells in Figure 5B. The correlation between the RF model and the data is shown in Figure 5G. To obtain the static nonlinearity for the L-NL model, we fitted a logistic function to the input-output curve of the RF model and the data (Figures 5G and 5H). The luminance model could predict the data to a certain degree, whereas the RF and L-NL models fared significantly better than the first (Figure 5I, $p = 1.49 \times 10^{17}$ and $p = 1.19 \times 10^{17}$, respectively). We compared the performance of each model with the reliability of the data (determined as for the static natural scenes as the correlation between different trials, as a proxy for the maximally predictable variance). Neither of the models was able to predict the data fully. Thus, although the representation of moving natural images by L2 to a large part reflects its linear response properties, a small part of the response remains unexplained.

We next investigated the representation of natural scenes by Mi1, a cell that is one synapse downstream in the motion vision pathway, postsynaptic to lamina neuron L1, and provides direct input to the direction-selective cell type T4. Mi1 responds to brightness increments with an increase in calcium levels (Arenz et al., 2017; Behnia et al., 2014; Strother et al., 2017), which we confirmed by the white noise analysis (Figures S3 and S4). We then measured Mi1 responses to moving natural scenes as for L2 (Figures 5C and 5E and Video S1) and created predictions of three different models. The static nonlinearity for Mi1 shows a pronounced saturation for low and high luminances (Figures 5J and 5K). As for L2, all models are able to explain significant variance of the data, yet none of them is able to fully predict the explainable variance (Figure 5L). Thus, although Mi1 represents a dynamic natural stimulus to a large degree as determined by its static properties, additional, likely dynamic, features add to this representation that are not captured by either of the models derived from white noise analysis.
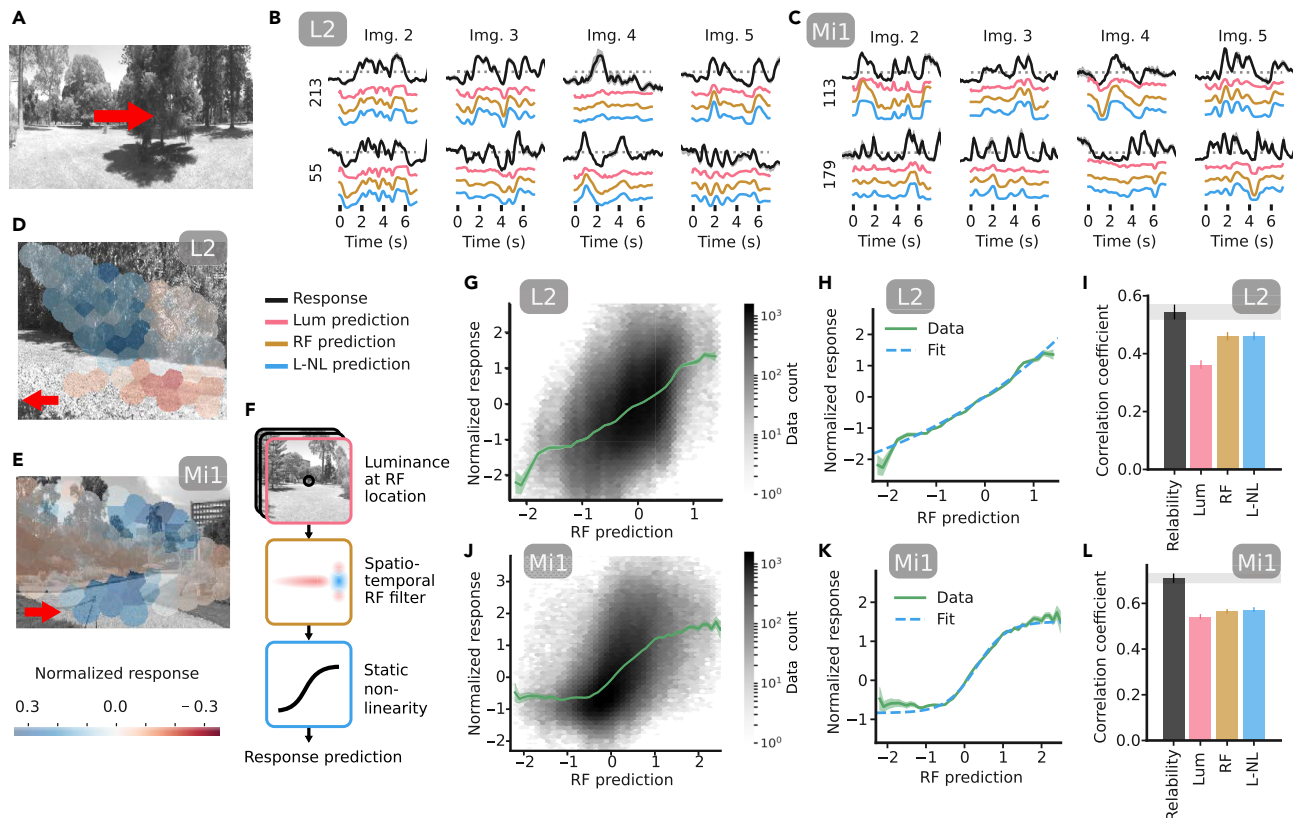
**Figure 5. L2 and Mi1 Responses to Moving Natural Scenes**

(A) Illustration of the stimulus. 360° panoramic images moved to the left for 7 s, followed by motion to the right for 7 s.

(B) Example responses of L2 cells (y axis label: cell number) to a selection of images. Solid black line represents the mean response over three trials; shaded black areas represent the 95% confidence interval. Pink, brown, and blue traces represent the corresponding model responses according to the legend below and the models illustrated in (F). More example responses are shown in Figure S5.

(C) Analogous to (B) but for Mi1.

(D) L2 example of all cells' responses at a single point in time during an individual recording, with the image frame at the cells' receptive fields in the background. The response dynamics to the moving image is shown in the Video S1.

(E) Analogous to (D) but for Mi1.

(F) Illustration of the models. The model output of the simplest model is the image luminance over time at the receptive field centers of the cells (pink, Luminance model). The second model's responses are generated by filtering the image sequences with the spatiotemporal receptive field kernel (from the white noise analysis) for each cell (brown, RF model). The last model adds a static nonlinearity to the RF model (L-NL model, blue).

(G) Pooled responses of all L2 axon terminals (N = 5 flies, n = 389 cells) in all recordings to all images at all time points, plotted against the RF model prediction (gray scale hexagons). Data count denotes the number of data points in each bin on a log-axis. The green trace represents the average, and the green shaded area represents the 95% confidence interval (over binned data).

(H) Mean response as in (G) and a least-squares logistic function fit (dashed blue line). This function is the static nonlinearity applied to generate the L-NL model predictions.

(I) Cellular reliability (black), described as the Pearson's correlation coefficient r between different trials (for details see Transparent Methods), and correlation between the different models and the data (pink, brown, blue). Error bars denote the 95% confidence interval.

(J) Analogous to (G) but for Mi1 (N = 4 flies, n = 351 cells).

(K) Analogous to (H) but for Mi1.

(L) Analogous to (I) but for Mi1.

Overall, we could demonstrate that the retinotopic populations of L2 and Mi1 code for local luminance changes of static and dynamic natural scenes, with an additional dynamic component not described by linear or static nonlinear receptive field models.

## DISCUSSION

Here, we present a remote-focusing module for two-photon microscopy to image oblique planes with uncompromised resolution and frame rate. We used the technique to simultaneously image over 100 cells in

the retinotopic plane of the *Drosophila* visual system, a significant portion of its visual field. This represents a large increase compared with the small number of simultaneously recorded cells (∼10) in similar preparations. Showing natural scenes to the cell types L2 and Mi1, we could demonstrate that the linear receptive field as measured with white noise can predict the cells' responses to natural scenes to a large extent.

In the original publication on remote focusing, and some in subsequently published adaptations, the remote focusing unit follows the lateral scan unit in the beam path, or is embedded between scan and tube lens (Botcherby et al., 2012; Colon and Lim, 2015; Rupprecht et al., 2016). Typical two-photon microscopes do not allow for additional optical elements between lateral scan unit and imaging objective. In this article, we presented a system in which the remote-focusing module is placed ahead of the lateral scan unit in the beam path. This way, standard two-photon microscopes can be equipped with this module without changes in the arrangement and design of the original optical elements. Sofroniew et al. (2016) also placed a remote-focusing module ahead of the lateral scan units in their system. Their lateral scan optics comprised three lateral scan mirrors and several relay telescopes, optimized for large field-of-view imaging, which makes it hard to compare the performance of their remote focusing with a remote-focusing module in front of a conventional two-photon microscope.

The distance between two columns in the visual system of the fruit fly is about 5 μm. Therefore, we aimed for a near-diffraction-limited focal spot size (0.5 × 0.5 × 3 μm) throughout the volume. A typical sine-corrected objective lens forms a diffraction-limited focal spot in the focal plane, but aberrations are introduced when focusing above or below the focal plane (Botcherby et al., 2008, 2007; Conrady, 1905). By introducing a second objective lens (the remote lens), aberrations are introduced on purpose to be then canceled out by the imaging lens. Full cancellation is achieved with a 4f telecentric system and an overall magnification of $n_2/n_1$ between the two objectives (Botcherby et al., 2007). In similar system (Rupprecht et al., 2016), in which either the telecentric relay or magnification criterion were omitted, the axial resolution depended on the axial position of the mirror, reaching 7 μm, and the relationship between mirror displacement and focus displacement was no longer linear. Depending on the size of the structures of interest, this resolution loss can be neglected. For small biological structures, however, a system with higher resolution is preferable. In our system, the relationship between mirror movement and focal spot displacement was linear, and the magnification was close to the ideal magnification (1.42 instead of 1.33), resulting in an overall better resolution. The remaining aberrations in the axial point spread function are likely due to the remaining inaccuracy in magnification, due to the short distance between tube lens and imaging objective lens, or may stem from alignment imprecision. For a magnification even closer to ideal, lenses in the remote-focusing module could potentially be chosen to precisely match the magnification of tube and scan lens.

Most other studies have used remote focusing to jump to different planes along the z axis. Our approach of tilting the imaging plane provides a simple and efficient solution to maximize information content in one, oblique plane (see also: Colon and Lim, 2015) without reducing the imaging rate. A comparable result could be attained by volume-imaging approaches using acousto-optic deflector systems (e.g., Grewe et al., 2010; Katona et al., 2012) or electro-tunable lenses (Grewe et al., 2011). Acousto-optic deflectors require beam stabilization and dispersion control (Iyer et al., 2003), and both would involve substantial modifications to the microscope setup. Thus, our system provides a simple, low-maintenance alternative that utilizes existing two-photon hardware. Although we used the module only to image oblique planes, arbitrary line scans or other 2D manifolds, for example, spherical surfaces, can be achieved without changing the hardware, by software extensions only.

We demonstrated the applicability of our technique by imaging the responses of the cell types L2 and Mi1 in the visual system of *Drosophila* to natural images, within the retinotopic plane spanned by their axon terminals. We first characterized their linear receptive fields and mapped their spatial distribution. As expected from the layout of the eye, receptive field locations formed a regular, hexagonal, and retinotopic grid, mapping the 2D visual space. We did not observe a systematic difference in receptive field structure across the range of locations we measured (70 × 90°, elevation × azimuth). The response properties of L2 to moving gratings, edges, flashes, and white noise stimulation have been characterized in detail (Arenz et al., 2017; Clark et al., 2011; Drews et al., 2020; Freifeld et al., 2013; Reiff et al., 2010), suggesting an OFF center, ON surround, and band-pass receptive field with an anisotropic structure, which we confirmed by white noise analysis. We also confirmed the receptive field of Mi1 as previously described (Arenz et al., 2017) and in addition identified a spatial anisotropy as for L2.

Our results show that the dynamic responses to natural scenes are largely, but not completely, predictable using classical, static linear-nonlinear models. This indicates dynamic nonlinear properties of the system. Lamina cells in blowflies have been shown to dynamically adapt to luminance (Laughlin and Hardie, 1978) and to natural, temporal sequences of luminances (van Hateren, 1997). The photoreceptors in *Drosophila* also adapt to and compress luminance levels (Juusola and Hardie, 2001), and similar phenomena have been described for other lamina neurons (Tuthill et al., 2014) and also recently for L2 (Ketkar et al., 2020). This dynamic adaptation could cause the variance not explained by the L-LN model for L2. Mi1 has recently been shown to exhibit spatial normalization to surround contrast, as well as temporal adaptation to variations in input contrast (Drews et al., 2020; Matulis et al., 2020). Contrast in natural scenes is highly variable, and these dynamic adaptation processes could therefore be responsible for the residual variance in Mi1.

Several studies in other model organisms have found both similarities and differences in visual response properties when comparing artificial with naturalistic stimuli. For example, complex cells of V1 in cats respond to natural scenes as predicted from their linear receptive field (Touryan et al., 2005), despite their known nonlinear properties (Hubel and Wiesel, 1962). Natural scenes have been used for reverse correlation analysis, which generally yielded receptive fields comparable to the ones obtained with white noise stimuli (Ringach et al., 2002; Theunissen et al., 2001; Touryan et al., 2005; Vance et al., 2016; Willmore et al., 2010). On the other hand, a recent study highlights the dynamic nonlinearities during natural scene stimulation in salamander retinal ganglion cells (Maheswaranathan et al., 2019), where white-noise-derived models are not able to explain their data well. Overall, whereas many response properties of cells can be probed with artificial stimuli, the use of natural stimuli can further our understanding of a dynamic neural circuit that has evolved in natural conditions.

In summary, with remote-focusing microscopy, we could effectively visualize the retinotopic maps in the *Drosophila* visual system. We showed that two early visual cell types form 2D feature maps of the fly's environment. We expect that our method will provide insights on spatial relationships of the cells in the visual system of the fly. Moreover, our module for imaging oblique planes can be implemented in any conventional two-photon system, which opens possibilities to investigate neural structures from new angles.

### Limitations of the Study

The distance between the imaging objective and the tube lens $f_4$ in conventional two-photon microscopes, including ours, is shorter than required for a complete 4f telecentric coupling between the remote objective and imaging objective lens. Therefore, our system could not be constructed entirely aberration-free. In addition, the magnification between remote objective and imaging objective lens is 1.42, whereas ideally it should be 1.33, which is another factor that introduces aberrations. Nevertheless, for the imaging volume required for our application, the aberrations are negligible. We are using gray scale natural images that are compressed in their luminance range, and that exhibit only global motion at constant speed. Although these images represent many parameters of the natural environment, they do not fully reproduce it. We are using the calcium sensor GCaMP6f to measure calcium signals of visual neurons. The dynamic properties of the sensor are far slower than the calcium dynamics in neurons, which limits the explanatory power of the measured signals, especially for the comparison between white noise and natural scene responses.

### Resource Availability

#### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Anna Schuetzenberger (schuetzenberger@neuro.mpg.de).

#### Materials Availability

This study did not generate new unique reagents.

#### Data and Code Availability

The modified version of ScanImage 5.1 created in this study is available at GitHub (https://github.com/borstlab/Scanimage_Planes3D). The datasets and the data analysis code generated during this study is available at GitHub (https://github.com/borstlab/obique-planes).

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101170.

## AUTHOR CONTRIBUTIONS

Conceptualization, A.S. and A.B.; Methodology, A.S. and A.B.; Investigation, A.S.; Software and Formal Analysis, A.S.; Writing – Original Draft, A.S.; Writing – Review & Editing, A.B.; Supervision, A.B.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

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

Aptekar, J.W., and Frye, M.A. (2013). Higher-order figure discrimination in fly and human vision. Curr. Biol. 23, R694–R700.

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

Baden, T., Berens, P., Franke, K., Roman-Roson, M., Bethge, M., and Euler, T. (2016). The functional diversity of retinal ganglion cells in the mouse. Nature 529, 345–350.

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

Borst, A., and Helmstaedter, M. (2015). Common circuit design in fly and mammalian motion vision. Nat. Neurosci. 18, 1067–1076.

Botcherby, E.J., Juškaitis, R., Booth, M.J., and Wilson, T. (2008). An optical technique for remote focusing in microscopy. Opt. Commun. 281, 880–887.

Botcherby, E.J., Juškaitis, R., Booth, M.J., and Wilson, T. (2007). Aberration-free optical refocusing in high numerical aperture microscopy. Opt. Lett. 32, 2007–2009.

Botcherby, E.J., Smith, C.W., Kohl, M.M., Débarre, D., Booth, M.J., Juškaitis, R., Paulsen, O., and Wilson, T. (2012). Aberration-free three-dimensional multiphoton imaging of neuronal activity at kHz rates. Proc. Natl. Acad. Sci. U S A 109, 2919–2924.

Braitenberg, V. (1967). Patterns of projection in the visual system of the fly. I. Retina-lamina projections. Exp. Brain Res. 3, 271–298.

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

Buchner, E. (1971). Dunkelanregung des stationaeren Flugs der Fruchtfliege Drosophila (University Tuebingen).

Chklovskii, D.B., and Koulakov, A.A. (2004). Maps in the brain: what can we earn from them? Annu. Rev. Neurosci. 27, 369–392.

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

Colon, J., and Lim, H. (2015). Shaping field for 3D laser scanning microscopy. Opt. Lett. 40, 3300–3303.

Conrady, A.E. (1905). The optical sine-condition. R. Astron. Soc. 515, 501–509.

Dayan, P., and Abbott, L.F. (2005). Theoretical Neuroscience (MIT Press).

Demb, J.B., and Singer, J.H. (2015). Functional circuitry of the retina. Annu. Rev. Vis. Sci. 1, 263–289.

Denk, W., and Detwiler, P.B. (1999). Optical recording of light-evoked calcium signals in the functionally intact retina. Proc. Natl. Acad. Sci. U S A 96, 7035–7040.

Diamond, J.S. (2017). Inhibitory interneurons in the retina: types, circuitry, and function. Annu. Rev. Vis. Sci. 3, 1–24.

Dowling, J. (2012). The Retina. An Approachable Part of the Brain, 2nd ed. (Belknap Press).

Drews, M.S., Leonhardt, A., Pirogova, N., Richter, F.G., Schuetzenberger, A., Braun, L., Serbe, E., and Borst, A. (2020). Dynamic signal compression for robust motion vision in flies. Curr. Biol. 30, 209–221.

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

Grewe, B.F., Langer, D., Kasper, H., Kampa, B.M., and Helmchen, F. (2010). High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. Nat. Methods 7, 399–405.

Grewe, B.F., Voigt, F.F., van't Hoff, M., and Helmchen, F. (2011). Fast two-layer two-photon imaging of neuronal cell populations using an electrically tunable lens. Biomed. Opt. Express 2, 2035–2046.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.

Iyer, V., Losavio, B.E., and Saggau, P. (2003). Compensation of spatial and temporal dispersion for acousto-optic multiphoton laser-scanning microscopy. J. Biomed. Opt. *8*, 460–471.

Ji, N., Freeman, J., and Smith, S.L. (2016). Technologies for imaging neural activity in large volumes. Nat. Neurosci. *19*, 1154–1164.

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

Juusola, M., and Hardie, R.C. (2001). Light adaptation in Drosophila photoreceptors: I. Response dynamics and signaling efficiency at 25°C. J. Gen. Physiol. *117*, 3–25.

Katona, G., Szalay, G., Maák, P., Kaszás, A., Veress, M., Hillier, D., Chiovini, B., Vizi, E.S., Roska, B., and Rózsa, B. (2012). Fast two-photon in vivo imaging with three-dimensional random-access scanning in large tissue volumes. Nat. Methods *9*, 201–208.

Ketkar, M.D., Sporar, K., Gür, B., Ramos-Traslosheros, G., Seifert, M., and Silies, M. (2020). Luminance information is required for the accurate estimation of contrast in rapidly changing visual contexts. Curr. Biol. *30*, 657–669.

Kirschfeld, K. (1967). Die Projektion der optischen Umwelt auf das Raster der Rhabdomere im Komplexauge von Musca. Exp. Brain Res. *3*, 248–270.

Land, M., and Fernald, R. (1992). The evolution of eyes. Ann. Rev. Neurosci. *15*, 1—29.

Land, M.F. (1997). Visual acuity in insects. Annu. Rev. Entomol. *42*, 147–177.

Laughlin, S.B., and Hardie, R.C. (1978). Common strategies for light adaptation in the peripheral visual systems of fly and dragonfly. J. Comp. Physiol. A *128*, 319–340.

Maheswaranathan, N., Mcintosh, L.T., Tanaka, H., Grant, S., Kastner, D.B., Melander, J., Nayebi, A., Brezovec, L., Wang, J., Ganguli, S., and Baccus, S.A. (2019). The dynamic neural code of the retina for natural scenes. bioRxiv. https://doi.org/10.1101/340943.

Matulis, C.A., Chen, J., Gonzalez-Suarez, A.D., Behnia, R., and Clark, D.A. (2020). Heterogeneous temporal contrast adaptation in Drosophila direction-selective circuits. Curr. Biol. *30*, 222–236.

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

Meister, M., Pine, J., and Baylor, D.A. (1994). Multi-neuronal signals from the retina: acquisition and analysis. J. Neurosci. *51*, 95–106.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature *454*, 995–999.

Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online *9*, 13.

Ready, D.F., Hanson, T.E., and Benzer, S. (1976). Development of the Drosophila retina, a neurocrystalline lattice. Dev. Biol. *53*, 217–240.

Reiff, D.F., Plett, J., Mank, M., Griesbeck, O., and Borst, A. (2010). Visualizing retinotopic half-wave rectified input to the motion detection circuitry of Drosophila. Nat. Neurosci. *13*, 973–978.

Ringach, D.L., Hawken, M.J., and Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. J. Vis. *2*, 12–24.

Rivera-Alba, M., Vitaladevuni, S.N., Mischenko, Y., Lu, Z., Takemura, S., Scheffer, L., Meinertzhagen, I.A., Chklovskii, D.B., De Polavieja, G.G., and Gonzalo, G. (2011). Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain. Curr. Biol. *21*, 2000–2005.

Ronzitti, E., Emiliani, V., and Papagiakoumou, E. (2018). Methods for three-dimensional all-optical manipulation of neural circuits. Front. Cell. Neurosci. *12*, 469.

Rupprecht, P., Prendergast, A., Wyart, C., and Friedrich, R.W. (2016). Remote z-scanning with a macroscopic voice coil motor for fast 3D multiphoton laser scanning microscopy. Biomed. Opt. Express *7*, 1656–1671.

Sanes, J.R., and Zipursky, S.L. (2010). Design principles of insect and vertebrate visual systems. Neuron *66*, 15–36.

Seelig, J.D., Chiappe, M.E., Lott, G.K., Dutta, A., Osborne, J.E., Reiser, M.B., and Jayaraman, V. (2010). Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behaviour. Nat. Methods *7*, 535–540.

Segev, R., Goodhouse, J., Puchalla, J., and Berry, M.J. (2004). Recording spikes from a large fraction of the ganglion cells in a retinal patch. Nat. Neurosci. *7*, 1154–1161.

Sofroniew, N.J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. Elife *5*, e14472.

Song, B.-M., and Lee, C.-H. (2018). Toward a mechanistic understanding of color vision in insects. Front. Neural Circuits *12*, 16.

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

Takemura, S., Nern, A., Chklovskii, D.B., Scheffer, L.K., Rubin, G.M., and Meinertzhagen, I.A. (2017). The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila. Elife *6*, e24394.

Theunissen, F.E., David, S.V., Singh, N.C., Hsu, A., Vinje, W.E., and Gallant, J.L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Netw. Comput. Neural Syst. *12*, 289–316.

Touryan, J., Felsen, G., and Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. Neuron *45*, 781–791.

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

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

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

van Hateren, J.H., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. Biol. Sci. *265*, 359–366.

Vance, P., Das, G.P., Kerr, D., Coleman, S.A., Mcginnity, T.M., 2016. Refining receptive field estimates using natural images for retinal ganglion cells., in: International Conference on Advanced Cognitive Technologies and Applications. pp. 77–82.

Willmore, B.D.B., Prenger, R.J., and Gallant, J.L. (2010). Neural representation of natural images in visual area V2. J. Neurosci. *30*, 2102–2114.
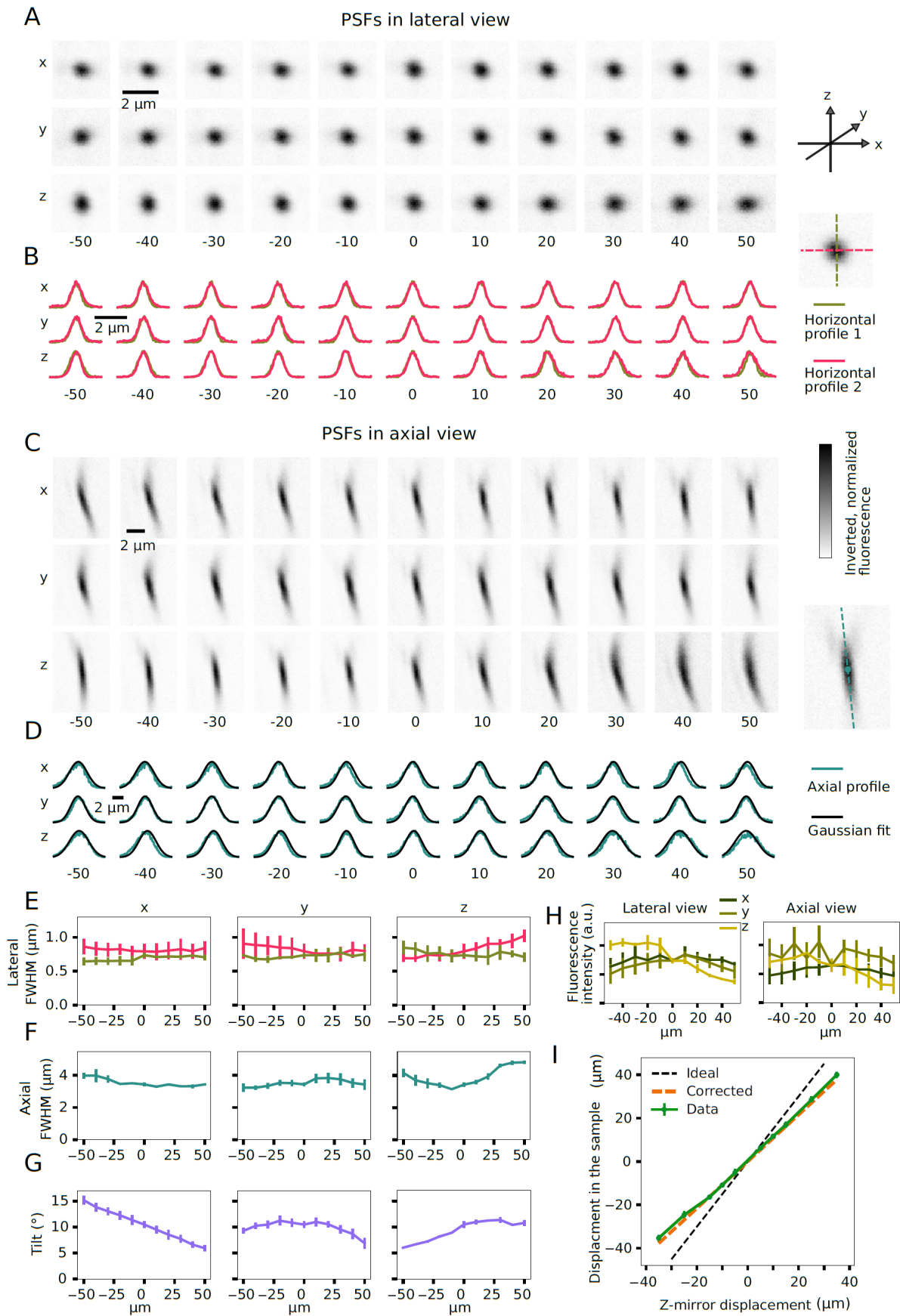
**Supplemental Information**

# Seeing Natural Images through the Eye

# of a Fly with Remote Focusing

# Two-Photon Microscopy

Anna Schuetzenberger and Alexander Borst

# Supplemental Information

## Supplemental Figures

A  PSFs in lateral view

x
y
z

-50  -40  -30  -20  -10  0  10  20  30  40  50

z
y
x

Horizontal
profile 1

Horizontal
profile 2

B

x
y
z

-50  -40  -30  -20  -10  0  10  20  30  40  50

C  PSFs in axial view

x
y
z

-50  -40  -30  -20  -10  0  10  20  30  40  50

Inverted, normalized
fluorescence

D

x
y
z

-50  -40  -30  -20  -10  0  10  20  30  40  50

Axial profile

Gaussian fit

E  Lateral FWHM (μm)    x    y    z

F  Axial FWHM (μm)

G  Tilt (°)

H  Fluorescence intensity (a.u.)    Lateral view    Axial view    x  y  z

I  Displacement in the sample (μm)    Ideal  Corrected  Data

Z-mirror displacement (μm)

2

**Figure S1 | Point spread function measurements. Related to Figure 2.** a) Lateral (xy) view of 0.1 µm fluorescent beads at ten positions along each axis (x, y and z). Mean images of three measurements. b) 1D profiles of the images in a) A 1D Gaussian function was fitted to each bead's profile to determine its full width at half maximum (FWHM). c) Axial (xz) view of 0.1 µm fluorescent beads at ten positions along each axis (x, y and z). Mean images of three measurements. d) A 2D Gaussian was fitted to each bead image. The mean profile along the elongated axis (turquoise) and the Gaussian fit along this axis (black) are shown. e) Lateral FWHM of the beads in a). Left, middle and right panel show the measurements along the x-, y- or z-axis, respectively. f) Axial FWHM of the beads in c). Left, middle and right panel show the measurements along the x-, y- or z-axis, respectively. f) Angle relative to 90° of the 2D Gaussian fit of the beads in c). Left, middle and right panel show the measurements along the x-, y- or z-axis, respectively. h) Relative fluorescence intensity of the beads. Left: Intensity of the bead center along different axes as in a). Right: Intensity of the bead center along the different axes as in c). i) Displacement of a bead in the imaging volume upon z-mirror displacement. The empirical measurements (green) were fitted with a linear model (orange). The slope of the linear fit rather than the ideal slope was used as a parameter in our imaging software to ensure accurate lengths and plane transformations.
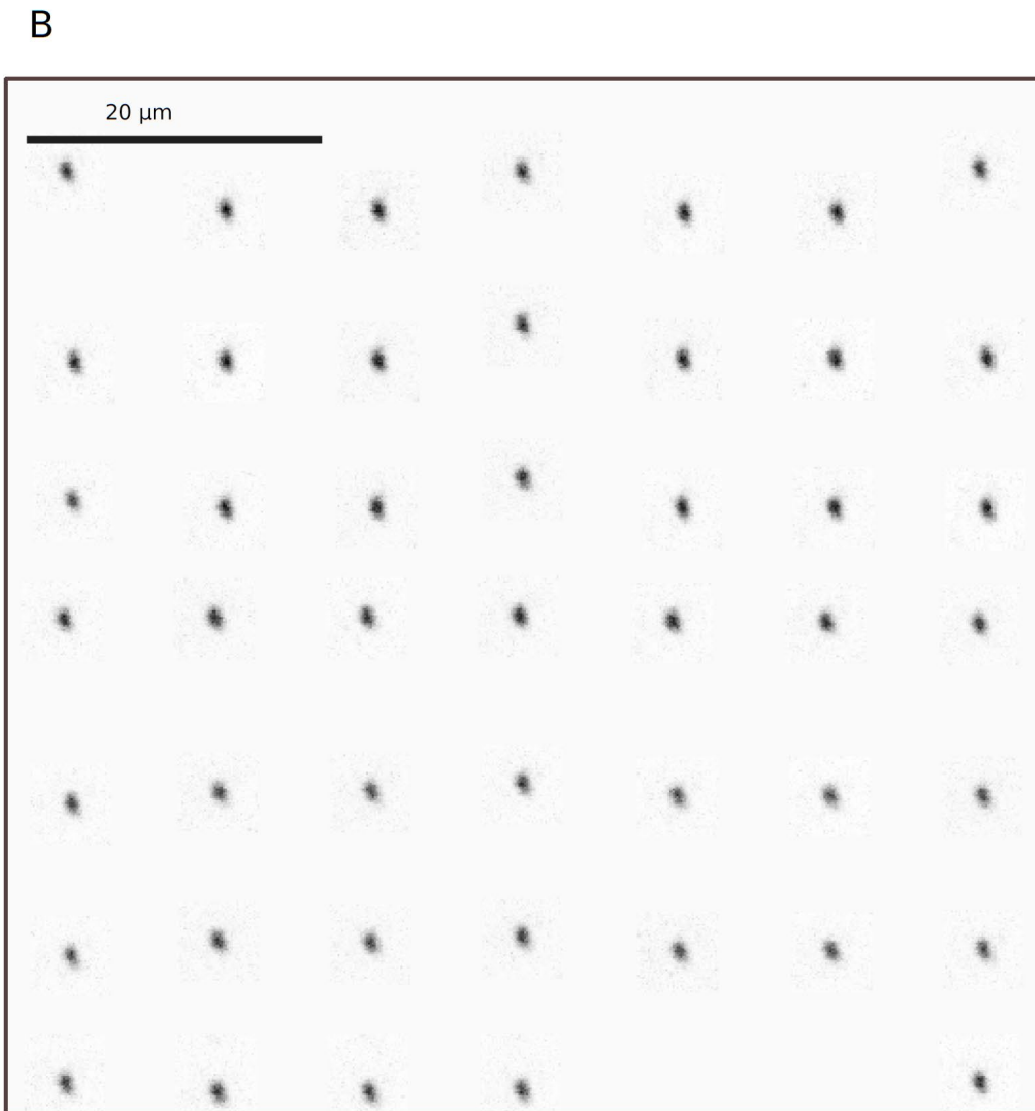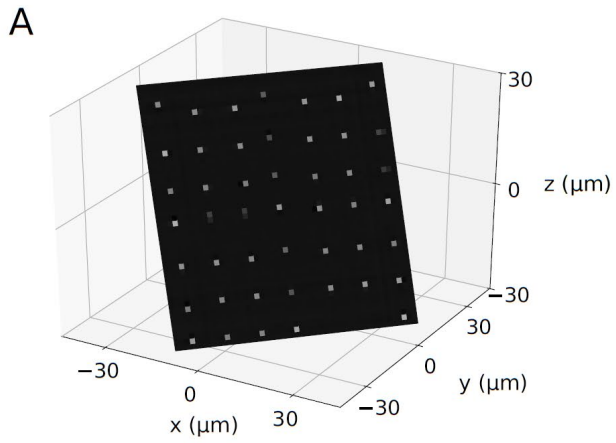
**Figure S2 | Point spread functions in an oblique cross-section. Related to Figure 2.** Composite image of 0.1 µm fluorescent beads imaged in different positions in an oblique imaging plane. This plane is tilted by 45° around the z-axis, followed by 55° around the x-axis, as illustrated in the top panel. It corresponds to a typical rotation and zoom we used to image planes in the fly visual system.
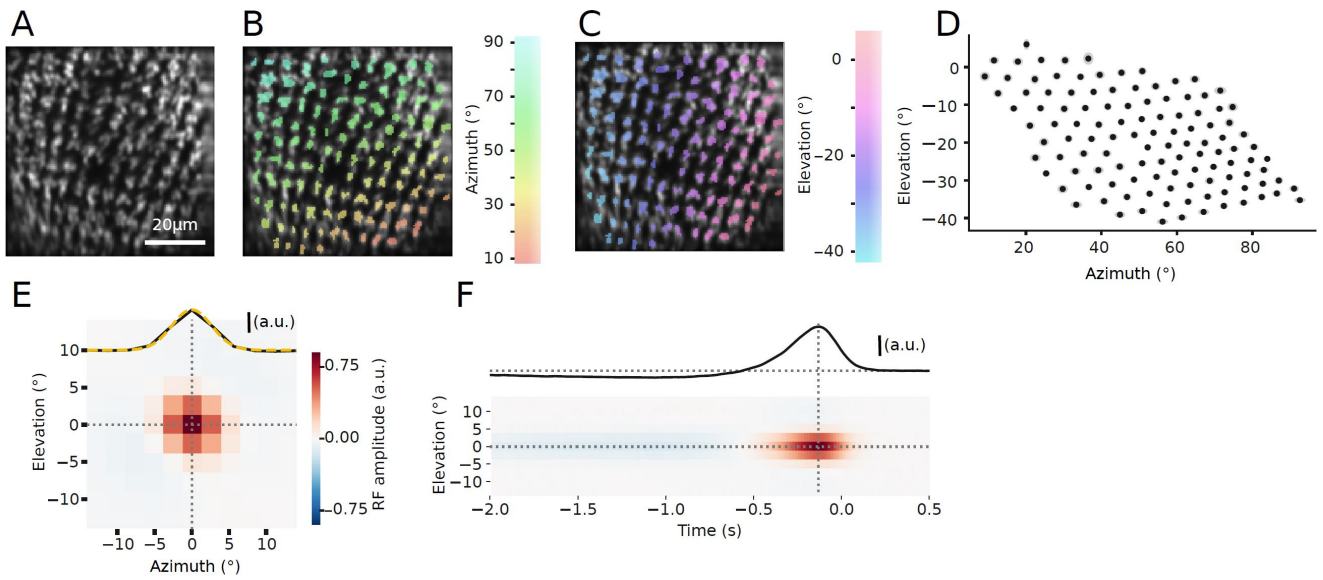
**Figure S3 | Receptive field of Mi1. Related to Figure 3.** a) Mean image of a recording of Mi1 axon terminals in layer 10 of the medulla expressing GCaMP6f. b) Axon terminals colored by their receptive field location in azimuth. c) Axon terminals colored by their receptive field location in elevation. d) Receptive field locations (black dots) with receptive field size (sigma of Gaussian fit, grey shaded areas) of all cells in the example recording also shown in a)-c). e) Average spatial receptive field of all cells (N = 8 flies, n = 540 cells). The profile in azimuth (black line) was fitted with a Gaussian (yellow). Average receptive field size was 2.02 ± 0.3° (σ of the Gaussian fit, corresponding to 4.77 ± 0.7° FWHM). f) Temporal profile of the receptive field. Grey dashed lines indicate the maximum receptive field amplitude.
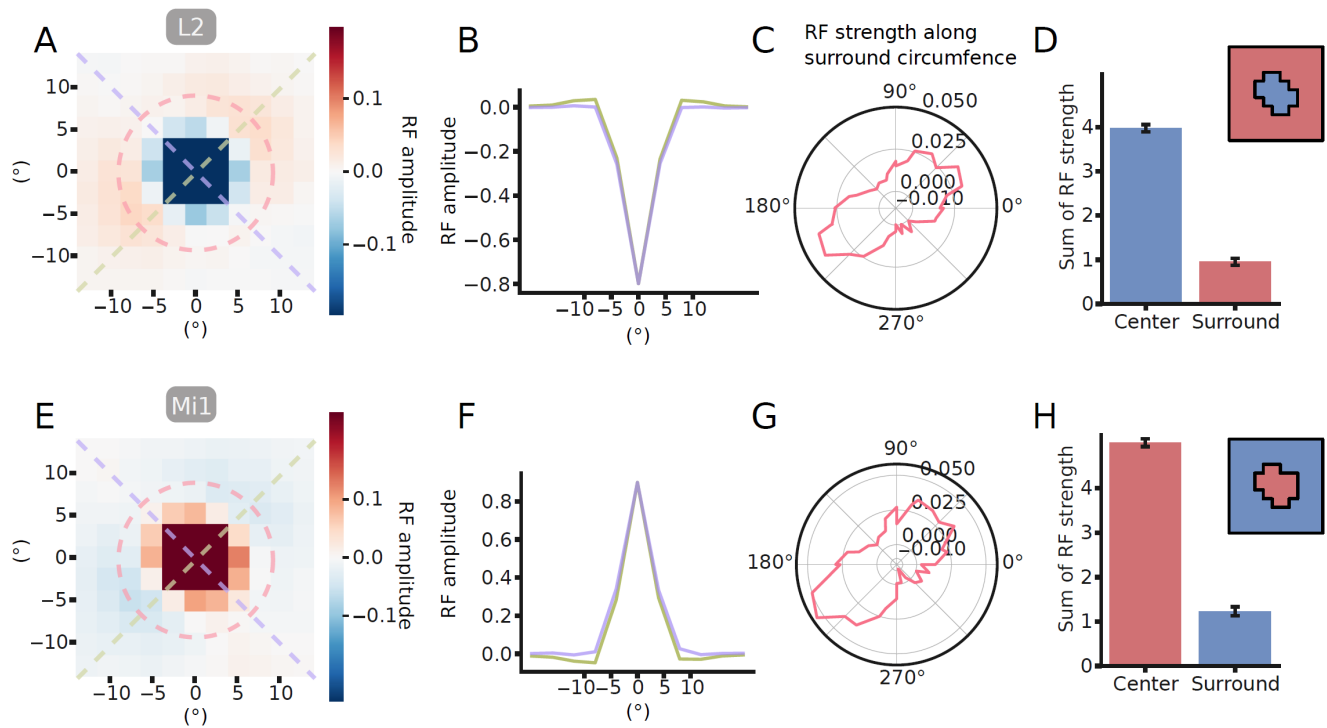
**Figure S4 | Anisotropic spatial receptive fields. Related to Figure 3.** a) Mean L2 receptive field (RF) with cropped color map, showing the anisotropic surround. The profiles along the colored dashed lines are plotted in b and c. b) 1D profile through the diagonal axes of the RF, as indicated in a. c) RF strength along the circumference of the RF indicated in a. A clear anisotropy can be observed. d) Relative RF strength of the center and the surround. All values of the RF center (blue area in inset) summed up result in the center RF strength. All values of the RF surround (red area in inset) summed up result in the surround RF strength. The surround accounts for 1/5th of the RF strength. e-f) Analogous for Mi1 as a-d) for L2. For better comparison, the data in the polar plot in g) was multiplied by -1.
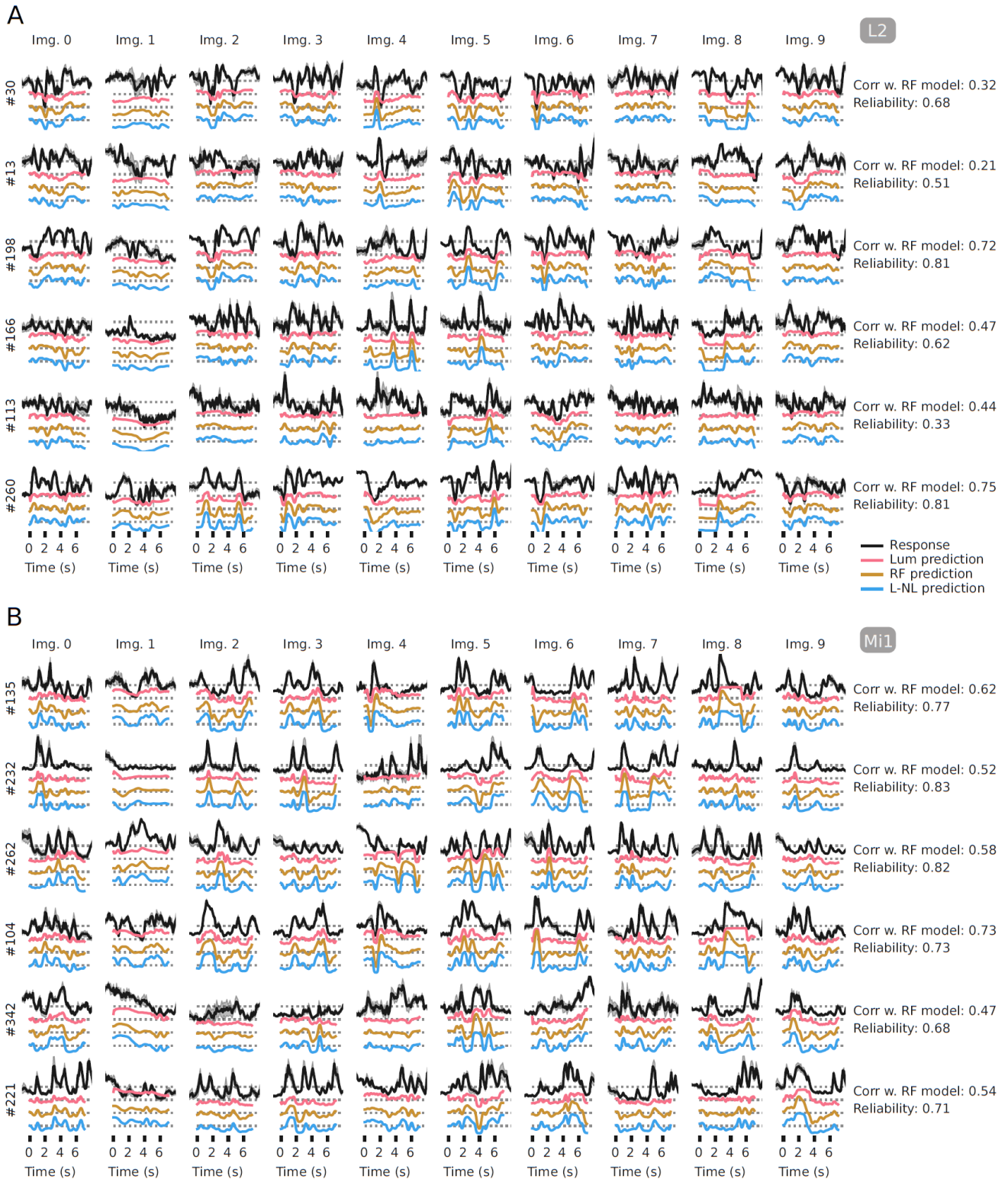
**Figure S5 | Moving image example responses. Related to Figure 5.** a) More example responses of L2 to moving images together with the corresponding model responses as in Fig. 5b. The y-axis label indicates the cell number. The dashed grey lines indicate the relative zero lines for the respective models. b) Analog to a) for Mi1.

## Supplemental tables

| Static images | Mean | STD | 95 % confi. Interv. | Student's t-test p-value |
|---|---|---|---|---|
| Reliability peak | 0.89 | 0.05 | 0.01 | 8.14E-34 |
| Lum model peak | 0.79 | 0.11 | 0.01 | |
| Reliability sust. | 0.59 | 0.22 | 0.02 | 0.005220031 |
| Lum model sust. | 0.63 | 0.19 | 0.02 | |

**Table S1 | Statistics related to Figure 4.**

**L2**

| Model correlations | Mean* | STD | 95 % confi. Interv. |
|---|---|---|---|
| Reliability | 0.54 | 0.26 | 0.03 |
| Lum model | 0.36 | 0.15 | 0.02 |
| RF model | 0.45 | 0.15 | 0.01 |
| L-NL model | 0.45 | 0.15 | 0.01 |
| | | | |
| **Student's t-test p-value** | | | |
| Lum vs RF model | 1.49E-17 | | |
| RF vs. L-LN model | 0.95646 | | |

**Mi1**

| Model correlations | Mean* | STD | 95 % confi. Interv. |
|---|---|---|---|
| Reliability | 0.71 | 0.19 | 0.02 |
| Lum model | 0.54 | 0.11 | 0.01 |
| RF model | 0.57 | 0.10 | 0.01 |
| L-NL model | 0.57 | 0.10 | 0.01 |
| | | | |
| **Student's t-test p-value** | | | |
| Lum vs RF model | 0.003096 | | |
| RF vs. L-LN model | 0.303808 | | |

\* Pearson's correlation coefficient r

| Nonlinearity: logistic function fitted parameters | | | |
|---|---|---|---|
| $b_u$ | $b_l$ | $x_0$ | k |
| 174.33 | -4.18 | 14.72 | 0.25 |

| Nonlinearity: logistic function fitted parameters | | | |
|---|---|---|---|
| $b_u$ | $b_l$ | $x_0$ | k |
| 1.50 | -0.84 | 0.30 | 2.54 |

**Table S2 | Detailed statistics and model parameters related to Figure 5.**

## Transparent Methods

### Experimental animals

*Drosophila melanogaster* were kept on standard cornmeal-agar medium at 25° and at 60 % relative humidity in a 12/12 h light/dark cycle. The Gal4-UAS system (Brand and Perrimon, 1993) was used to express GCaMP6f (Chen et al., 2013) in the lamina cell type L2 and the medulla cell type Mi1. The resulting genotype of the L2 experimental flies was $w^+/w^-$; R53G02-AD/UAS-GCaMP6f; R29G11-DBD/UAS-GCaMP6f, and of the Mi1 experimental flies was $w^+/w^-$; R19F01-AD/UAS-GCaMP6f; R71D01-DBD/UAS-GCaMP6f. Female flies 1-7 days after eclosion were used for the experiments.

### Remote-focusing two-photon microscopy

We designed a remote-focusing two-photon microscope based on the two-photon microscope described in Euler et al. (2009) and the principle of remote focusing described in Botcherby et al. (2008).

#### Basic two-photon system

A MaiTai eHP DS Ti:Sapphire oscillator was used as a laser source. A beam attenuator (Newport VA-BB-2-CONEX) was used to control the laser intensity, and a telescope initially collimated the beam and widened it approximately 4-fold. The lateral scan unit comprised of two galvanometer optical scanners (Cambridge Technologies 6215H). A telescope then magnified the beam with magnification $M_1 = 4$ and imaged it to the back aperture of the sample objective lens. The water immersion Olympus Lumplfln 40x w with an NA of 0.8, back aperture of 9 mm and working distance of 3.3 mm was used as sample objective lens. A Photomultiplier Tube (Hamamatsu H10770-PB40) collected the emitted fluorescence signal, which was then digitized with a NI PCI-6110 data acquisition board. A band-pass filter (BrightLine 514/30) and a short-pass filter ensured that light from the visual stimulation arena and the laser could not reach the photomultiplier tube.

#### Remote focusing module

The remote focusing module was positioned in the beam path after the attenuator and the initial telescope, and before the lateral scan unit. The beam enters the module at a polarization-based beam splitter (PBS). It then passes through a λ/4 plate, enters the remote objective lens, is reflected by the z-mirror, and travels back through the remote objective lens and the λ/4 plate. It is in turn reflected by the PBS and passed though the telescope to the lateral scan unit.

The position of the z-mirror with respect to the focal plane of the remote objective lens was controlled by a piezoelectric actuator and a closed-loop DSP controller. The actuator was manufactured customized (nanoFaktur) to have a fast settling time (1 ms at 1 μm precision) at a large traveling range (280 μm) and a position sensor. A DSP Controller (nPoint LC.402 with extended current) was used to control the actuator. The

z-mirror was glued on the actuator. We chose a lightweight mirror (Thorlabs PF05-03-M01, 12mm diameter) that can tolerate the laser intensities needed for calcium imaging. Since the laser beam diameter on the mirror is very small (down to <1 µm if the mirror position is exactly in the focal plane), the energy per area is large. In addition, for mode-locked laser sources with femtosecond pulse length, damage thresholds were not determined by the manufacturer. Therefore, we tested several different mirrors for applicability in this specific setting. Metal-coated mirrors were found to be superior to dielectric mirrors. The actuator together with the z-mirror were mounted on a XYZ translation stage (Thorlabs PT3/M). This enabled a simple manual translation of the z-mirror and actuator in the case of local mirror damage. We tuned the PDI controller of the piezo to optimize for scanning distances of around 90 µm, which it could achieve with sufficient precision up to 15 Hz. Alternatively, we used it to scan smaller size planes (50x50 µm) at up to 20 Hz. Depending on the use, the PDI controller could also be tuned to optimize for shorter scanning distances, where it then could achieve still higher scan rates.

As the remote objective lens (ROL), we chose a Leica HC PLAN APO objective with 40x magnification, 0.85 NA, 0.21 mm free working distance and transmission of about 70% at 920 nm. Since the sample objective lens is a water immersion objective, we could not use the exact same model for the remote focusing lens. To account for the cover slip correction of this lens, we mounted a cover slip to its front.

Two telescopes imaged the back aperture of the remote objective lens onto the back aperture of the imaging objective lens, with an intermediate conjugate plane close to the lateral scan unit. The focal lengths of the relay lenses were $f_1 = 125$ mm, $f_2 = 40$ mm, $f_3 = 50$ mm (scan lens), $f_4 = 200$ mm (tube lens). Together with the focal lengths of the objective lenses, 5 mm for the ROL and 4.5 mm for the IOL, this resulted in an overall magnification of

$$\frac{f_{ROL}}{f_1} * \frac{f_2}{f_3} * \frac{f_4}{f_{IOL}} = \frac{5}{125} * \frac{40}{50} * \frac{200}{4.5} = 1.42$$

which is close to the desired value for aberration-free imaging of $n_2/n_1 = 1.33$ (Botcherby et al., 2007).

The distances between the lenses in the relay system were: Back end of ROL to $f_1$: 125 mm; $f_1$ to $f_2$: 165 mm; $f_2$ to $f_3$: 90 mm; $f_3$ to $f_4$: 250 mm; $f_4$ to back end of IOL: 95 mm. The distance between $f_4$ and IOL could vary by few mm, depending on the precise positioning of the sample. Apart from the last distance, all distances were chosen as to ensure a 4f telecentric configuration. This last distance is often set in conventional two-photon microscopes to be smaller than the sum of the focal distances of tube lens and objective lens. While entirely aberration-free imaging is only possible with 4f relays (Botcherby et al., 2008), the optical resolution and linearity in our system across the relevant imaging range were only marginally affected, which is why we decided to keep this configuration.

*Validation*

To validate the resolution and scaling along the z-axis, we imaged 1 μm and 0.1 μm diameter fluorescent beads (Polysciences Fluoresbrite) embedded in agarose. The approximate theoretical resolution limit of an imaging system, i.e. the minimum distance between two points to be resolvable, is given by the Rayleigh criterion (adapted for two-photon microscopy, see Yang and Yuste, 2017):

$$R_{xy} = \frac{0.4\,\lambda}{NA} = 0.46\ \mu m$$

$$R_z = \frac{1.4\,\lambda\,n}{NA^2} = 2.7\ \mu m$$

Where NA is the numerical aperture of the objective lens, λ is the wavelength and n the refractive index of the imaging medium. The cross-sections of a bead were measured at different points in a 100x100x100 μm volume. A 2D Gaussian was fitted to each cross-section, and the full width at half maximum (FWHM) was defined as the resolution of the system.

To measure the scaling factor between mirror movement and focal spot displacement in z, we focused on an individual bead, moved the z-mirror by a set amount, and then measured the distance we needed to move the IOL with a micromanipulator to bring the bead back in focus (Fig. S2i). The resulting scaling factor was used by the software to ensure appropriate distances and transformations. The scaling factor for a perfect imaging system from air to water is 1.5 (Botcherby et al., 2012, 2008). The empirically determined factor for our system was 1.1. The transformation was near linear across 90 μm of focal spot displacement (Fig. S2i).

*Software*

A customized version of the open-source software ScanImage 5.1 (Pologruto et al., 2003) was used to acquire the data. NI data acquisition boards were used to generate the output voltage for the mirrors (PCIe-6353 with two BNC 2110 accessories) and to collect the signal from the photomultiplier tube. In a second input channel, the trigger signal for synchronization of the visual stimulus with the imaging data was recorded. We extended the functionality of ScanImage (Matlab2013b) to control the movement of the piezo actuator for scanning in z in the same way as it controls the galvanometer motors for lateral scanning. A third output channel generated the voltage that defined the position of the z-mirror. A scaling parameter (0.0565 V/μm$_z$, determined empirically, see above) ensured the appropriate transformation of voltage to the axial displacement in the sample space. ScanImage represents an imaging plane as a time series of 2D vectors, the positions of the focal point in a virtual space. The first and second dimension correspond to the movement of the x and y scan mirrors, respectively. We added a third dimension to this representation. Simple rotation matrix transformations applied to the 3D vectors can then arbitrarily rotate the imaging plane in 3D space, about

several rotational axis simultaneously. In addition, since the piezo motor moving the z mirror is slower than the galvanometer motors, we ensured that the z-mirror receives only the slow part of scanning signal. With additional rotational controls in the graphical user interface, the user can rotate the imaging plane in the 3D imaging volume in all directions during imaging. Our modified version of ScanImage is publicly available on Github under https://github.com/borstlab/Scanimage_Planes3D.

## Visual stimulation arena

For the display of visual stimuli to the fly, we used a custom-build projector-based arena as described previously (Arenz et al., 2017). Briefly, visual stimuli were projected onto a semi-transparent cylinder that was positioned in front of the fly, spanning 180° in azimuth and 105° in elevation of the fly visual field. Stimuli were projected with a frame rate of 180 Hz in 255 brightness steps. The brightest absolute luminance of the arena reached approximately 250 candela/m2, and the wavelength was restricted to > 550 nm, to separate it from the GCaMP emission wavelength. Stimuli were written in Python 2.7 using the Panda3D software. To synchronize the visual stimulation with data acquisition, a small corner of the arena displayed a trigger signal that was recorded together with the imaging data.

## Visual stimuli and natural images datasets

### Letter stimulus
The letter stimulus consisted of 30x30° large, bright letters on a dark background. The letters were flashed following a completely dark background, for 1 second. Each letter was presented three times.

### White noise stimulus
To determine the linear receptive field of the neurons, a white noise stimulus was shown for 4-10 minutes. The stimulus consisted of a grid of 64x54 pixels, each pixel corresponding to approximately 2.8x1.9° in visual space. The luminance of each pixel changed every third frame according to values (0-255) drawn from a normal distribution. The luminance values were then slightly low-pass filtered in space and time ($\tau = 0.5$ s). For a subset of acquisitions, the luminance of each pixel could either take the value 0 or 255, set by a pseudo-random number generator every 0.5 s ("binary noise stimulus").

### Static natural scenes stimulus
We used images from the Van Hateren natural image dataset (van Hateren and van der Schaaf, 1998) for the static natural scenes stimulus. Of the 4000 images in the dataset, we selected 50 images that represented the image diversity and showed variable luminance and contrast characteristics. The images were then cropped, downsampled to 180x260 px (corresponding to approx. 1x0.4° in the fly's visual space) and gamma corrected ($\gamma=1.5$) to compress the luminance range to the 8 bit depth of the stimulation arena.

At the beginning of each trial, following a gray screen (luminance = 125), an image was displayed on the full screen. The image stayed for 1.5 seconds before the screen turned gray again. To compare the image responses to luminance steps, we added two additional stimulus conditions. Following a gray background, either a black screen (luminance = 0, OFF flicker stimulus) or a white screen (luminance = 255, ON flicker stimulus) was shown for 1.5 seconds. This resulted in a total of 52 stimulus conditions in the natural scenes stimulus. Each stimulus condition was repeated three times and trials were randomized for each acquisition.

*Moving natural scenes stimulus*
Ten 360° panoramic images from an HDR image dataset (Brinkworth and O'Carroll, 2009) were used for the moving natural scene stimulus. The images were gamma corrected (γ=2.2) to compress the luminance range to the 8 bit depth of the stimulation arena and slightly stretched along the elevation axis to cover the entire 105° vertical extend of the stimulus arena (original images spanned 75° in vertical extend). The images were displayed on the arena and, after two seconds, started moving first to the left for seven seconds, and then back (to the right) for another seven seconds (For one recording of Mi1, the image moved for only six seconds, stopped for two seconds and then moved back for six seconds. Only the first six seconds of all recordings were analyzed). Each image was shown in three trials, and all trials were randomized. For a new acquisition, the image phase in azimuth at which the image started to move was randomized to increase stimulus variability and avoid artificially introduced correlations between acquisitions. An example stimulus together with cellular responses is shown in the supplemental video.

## Calcium imaging procedure

Flies were prepared for imaging following a standard procedure (Maisak et al., 2013). Briefly, flies were anesthetized on ice and then glued to a holder with the posterior side of the head exposed to the objective lens through a small opening. Ringer's solution was applied, and the cuticle, fat and trachea at the back of the head were removed for optical access.

Initially, a moving grating was shown to locate areas of axon terminals with large calcium responses. The imaging plane was then rotated with the rotational controls in the ScanImage GUI until the plane spanned by the 2D array of axon terminals was visible. Then, the white noise stimulus was shown, followed by either the static or the moving natural scenes stimulus. For a subset of acquisitions, we showed a white noise stimulus again after the natural images stimulus, to verify that the receptive field locations of the cells had not changed. Recordings were terminated prematurely if cells stopped showing calcium responses, and data from these recordings were not used for further analysis. All data were recorded at a frame rate of 7.5 Hz and a zoom setting of 3 (for L2) or 4 (for Mi1), equivalent to a field of view size of about 90 or 70 μm, respectively.

For the letter stimulus, we first used a small, windowed grating stimulus to approximately locate the area on the screen the cell population responded to. We then placed the letter stimulus at that location. Fig. 2f shows average ΔF/F images over one second duration of the letter and over three trials.

## Data processing and analysis

Data analysis was performed in Python 2.7 with custom-written software.

### *Preprocessing*

Image sequences were registered to the mean image with a phase correlation algorithm described in detail in Arenz et al., (2017). For L2, individual cells' axon terminals were selected based on initial thresholding of the high-pass filtered mean image, followed by manual inspection and corrections. For Mi1, individual cells' axon terminals were selected using a custom-written algorithm based on correlations between neighboring pixels in the white noise experiment and thresholding, followed by manual inspection and correction.

### *White noise analysis*

For baseline subtraction, a dynamic baseline fluorescence level was determined by low-pass filtering (τ = 5 s) the raw fluorescence signal for each cell. The baseline was then shifted by in time by $\tau * framerate/2$ and subtracted from the raw fluorescence signal, resulting in the calcium signal ΔF. For each cell, reverse correlation of the calcium signal to the white noise stimulus was then performed, described in detail in Arenz et al. (2017), resulting in the spatio-temporal linear receptive field (RF). A 1D Gaussian function was fitted to the azimuth and the elevation axes of the spatial RF. The peaks of the curve fits were termed the RF location of each cell. RFs were normalized and cells with a z-score < 10 and recordings with the binary noise stimulus were not included in the white noise analysis (Fig. 3, Fig. S3 and S4). For the natural scenes analyses, cells with RFs with a z-score < 7 were not included (Fig. 4 and Fig. 5).

### *Static natural scenes analysis*

For each cell, the raw fluorescence signal in the second before the start of a trial was used as baseline and subtracted from the fluorescence signal during that trial. Subsequently, responses to three trials were averaged to obtain the mean response for each stimulus condition. Cells that exhibited very small responses to ON and OFF flicker stimuli (< 200 difference in baseline-subtracted raw fluorescence units), and cells with a standard deviation of trial-to-trial correlation coefficients of > 0.04 were excluded from further analysis. The peak and sustained responses, $R_{peak}$ and $R_{sust}$, were quantified as follows:

$$R_{peak} = \begin{cases} \min(\Delta F_s) & if \ |\min(\Delta F_s)| > |\max(\Delta F_s)| \\ \max(\Delta F_s) & if \ |\min(\Delta F_s)| < |\max(\Delta F_s)| \end{cases}$$

$$R_{sust} = mean(\Delta F_{s,t>1})$$

With $\Delta F_s$ being the calcium signal during the stimulus presentation, and $\Delta F_{s,t>1}$ the calcium signal during the last 0.5 s of stimulus presentation. The calcium signals and the response quantifications were then divided by a normalization factor. The factor was defined for each cell as 4 times the standard deviation of $R_{sust}$ over all stimulus conditions.

For each cell, we fitted a linear function to its mean image responses (50 data points) and the image luminance at its RF using the least squares method. How well the fit could predict the data was expressed as the mean of the Pearson's correlation coefficient ρ between fit and each of the three trials:

$$\rho_{fit} = \frac{\rho_{fit,0} + \rho_{fit,1} + \rho_{fit,2}}{3}$$

Where $\rho_{fit,1}$ denotes the Pearson's correlation coefficient between the fit and the 1$^{st}$ trial. The correlation coefficient between trials was determined as the mean of the correlation coefficient of each pair of trials, determined independently for each cell:

$$\rho_{trials} = \frac{\rho_{0,1} + \rho_{0,2} + \rho_{1,2}}{3}$$

Where $\rho_{0,1}$ denotes the Pearson's correlation coefficient between the first and second trial of all images. This measure defines an upper bound for the variance predictable by the fit.

*Moving natural scenes analysis and modeling*
Baseline subtraction was performed as for the white noise analysis. L2 cells with low average response dynamics (quantified as the standard deviation of response throughout a recording) were excluded from further analysis. The responses of each cell to three trials were averaged to obtain a mean response. Trial-to-trial variation ('reliability') was determined as for the static natural scenes for each time point (sampling rate 15 Hz). For all parts of Fig. 5 and S5, we analyzed only the first 6 seconds of the stimulus response (the image moving to the right), and we did not analyze the data of the image moving back.

The luminance model's responses were generated by taking, for each cell, the time series of image luminance values that appeared at the cell's receptive field location. Prior, the images were low-pass filtered with a 2D Gaussian filter with a sigma corresponding to the receptive field size of L2 or Mi1, respectively. In other words, the first model describes a cell with an instantaneous 2D Gaussian filter corresponding to only the center component of the receptive field. The RF model incorporated the 3D filter kernel obtained from the white noise analysis, and thus includes the temporal properties of the cell and the spatial surround in addition to the spatial center. For each cell, the image series was filtered with the 3D kernel to obtain the model responses. The L-NL model consisted of the RF model, and, in addition, a static nonlinearity. To obtain the model's

15

responses, the output of the RF model was passed through a logistic function. The parameters for the logistic function were determined empirically by fitting the logistic function $f(x) = b_l + \frac{b_u - b_l}{1 + \exp(-k(x - x_0))}$ to the output of the RF model and the data with the least-squares method (Fig. 5h for L2 and Fig. 5k for Mi1). In order to compare the model responses with the cells' responses, we calculated the Pearson's correlation coefficient between models and data as for the static natural scenes, each time point representing one data point.

## Supplemental references

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

Botcherby, E.J., Juškaitis, R., Booth, M.J., Wilson, T., 2008. An optical technique for remote focusing in microscopy. Opt. Commun. 281, 880–887.

Botcherby, E.J., Juškaitis, R., Booth, M.J., Wilson, T., 2007. Aberration-free optical refocusing in high numerical aperture microscopy. Opt. Lett. 32, 2007–2009.

Botcherby, E.J., Smith, C.W., Kohl, M.M., Débarre, D., Booth, M.J., Juškaitis, R., Paulsen, O., Wilson, T., 2012. Aberration-free three-dimensional multiphoton imaging of neuronal activity at kHz rates. Proc. Natl. Acad. Sci. USA 109, 2919–2924.

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

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

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., Looger, L.L., Svoboda, K., Kim, D.S., 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499, 295–300.

Euler, T., Hausselt, S.E., Margolis, D.J., Breuninger, T., Castell, X., Detwiler, P.B., Denk, W., 2009. Eyecup scope - optical recordings of light stimulus-evoked fluorescence signals in the retina. Eur. J. Physiol. 457, 1393–1414.

Maisak, M.S., Haag, J., Ammer, G., Serbe, E., Meier, M., Leonhardt, A., Schilling, T., Bahl, A., Rubin, G.M., Nern, A., Dickson, B.J., Reiff, D.F., Hopp, E., Borst, A., 2013. A directional tuning map of Drosophila elementary motion detectors. Nature 500, 212–216.

Pologruto, T.A., Sabatini, B.L., Svoboda, K., 2003. ScanImage: Flexible software for operating laser scanning microscopes. Biomed. Eng. Online 9, 13.

van Hateren, J.H., van der Schaaf, A., 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. Biol. Sci. 265, 359–366.

Yang, W., Yuste, R., 2017. In vivo imaging of neural activity. Nat. Methods 14, 349–359.