

We acquired *GMR-hid* and *UAS-shi<sup>ts</sup>* flies from the Bloomington Stock Center, *orco-* flies from Leslie Vosshall, *fru P1-GAL4*, *fru P1-FLP*, *VT41688*, and *VT40556* flies from Barry Dickson, NP2631 flies from the Kyoto stock center, and *DSX-GAL4* flies from Stephen Goodwin. Michael Reiser provided all split-GAL4 lines for labeling lamina neurons [84], and *Cont<sup>Dm</sup>* [91], *Phos* [89], and *Dual* [90], lines were provided by Jim O. Vigoreaux.

For arista- and tarsi- cutting paradigms (AC and TCBL), to render flies deaf (AC) or remove gustatory cues (TCBL), WT1 (for AC) or BL (for TCBL) flies were anesthetized with CO<sub>2</sub>, and aristae or tarsi were removed with fine scissors >20 hours prior to recording. To generate SP flies, *Drosophila melanogaster* sex peptide [56] was injected into PIBL virgin females 12-24 hours prior to recording. These flies were anesthetized on ice during injections. For recordings from headless FRU<sup>Act</sup> flies, heads were removed during cold anesthesia < 5 minutes before recording.

### 3.4.2 Behavioral chamber

The recording chamber used for all untethered experiments was as described in [77] with minor alterations. We used a sectional LED ring light (edmund optics) for experiments with constant or light/dark illumination, and a 3" white LED ring light (edmund optics) when switching between 4 light intensity levels. Finally, we used an external triggering system to record video at 60Hz.

### 3.4.3 Air-supported ball

Designs for the ball holder were supplied by Vivek Jayaraman, and 6mm diameter HDPE balls (Craig Technologies) were roughened slightly with sand-paper before use. Flies were tethered to a 0.005" tungsten rod (A-M Systems) using a small spot of ultraviolet-cured glue (Norland Optical and Electronic Adhesives) on the anterior mesothorax under cold anesthesia. All flies were tethered < 30 minutes prior to recording.

Song signals were recorded with 2 NR23158 microphones placed either side of the fly (Knowles Electronics). Both microphones were connected to a custom-built amplifier [70]. Stimuli were presented on an XL2411T monitor (BenQ) at 144Hz using Psychtoolbox-3, and synchronized with microphone recordings via a light sensor (GA1A12S202, Adafruit). A 808nm laser (Dragon Lasers), passing through a 780nm collimation fiber (Thor labs) and focused on the fly head, was used to heat flies. For each recording, laser power was increased until flies began singing, and then held constant throughout the recording. All recordings lasted 15 minutes.

#### **3.4.4 Courtship assay**

When recording male/female pairs, the chamber was prepared as previously describe [77]. Flies were loaded individually into the chamber using a custom-built aspirator. To optimize for peak fly activity, we began recordings within 150 minutes of the behavioral incubator lights switching on. Recordings were terminated when copulation occurred or after 30 minutes (classified as no copulation). Recordings with no song in the first 5 minutes were terminated early and disregarded.

#### **3.4.5 Wing Measurement**

Wings were measured after behavioral experiments with wing length defined as previously described [79]: the distance from the intersection of the anterior cross vein and longitudinal vein 3 (L3) to the intersection of L3 with the distal wing margin. All wing lengths were measured at 5x magnification using a reticule inserted into a microscope eyepiece.

#### **3.4.6 Thermogenetic activation in the behavioral chamber**

For thermogenetic activation of FRU, DSX, P1, pIP10, and dPR1 neurons in males, the chamber was heated with a coil (HSTAT, BriskHeat) to 26°C (for FRU activation), 27°C

(for DSX activation), or 32-34°C (for P1, pIP10, and dPR1 activation). For experiments in which a heat activated fly was placed with a female (see [Figure 3.12](#)), activated song was confirmed prior to female introduction, and after female removal at the end of the 30 minute trial. If either of these criteria were not met, the trial was discarded. When heat-activated flies were recorded in the absense of a female (see [Figure 3.4f](#)), recordings were not timed to peaks in circadian activity.

### 3.4.7 Dark/Light transition assay

For acute transitions between dark/light conditions (see [Figure 3.5b](#) and [Figure 3.7a](#)), the chamber was covered completely with BK5 Blackout Fabric (Thor Labs), and flies were loaded in darkness. A custom built voltage switch was used in combination with a Master 8 (A.M.P.I.) pulse generator to cycle power to the sectional LED ring light (Edmund Optics) with a period of 30s (15s on/off). On/off periods were detected post-experiment using the light-intensity during each camera frame. As flies were activated in isolation, recordings were not timed to peaks in circadian activity.

### 3.4.8 Multi-level light intensity assay

The intensity of a 3" white LED ring light (edmund optics) was adjusted using an analog voltage signal generated from the same data acquisition terminal (National Instruments) used to acquire amplifier signals from the microphones and trigger the camera. This ensured synchrony between all components. Lux levels were calibrated by placing a light sensor (GA1A12S202, Adafruit) in the behavioral chamber and recording the output over the entire range of light intensities. For each recording, the sequence of light intensities presented was randomly generated using a truncated 4-level maximum length sequence; this eliminated any confounds of periodicity while ensuring an even distribution of transitions [\[96\]](#).

In experiments where 0.5lux was one of the light intensities used, the chamber was covered completely with BK5 Blackout Fabric (Thor Labs) to avoid light contamination from the

computer monitor. If 650lux was the lowest light intensity, the shutter time of the camera was continually adjusted throughout the experiment to prevent image saturation at higher light intensities and allow for fly tracking. As flies were activated in isolation, recordings were not timed to peaks in circadian activity.

### 3.4.9 Inactivation of motion-sensitive neurons

Chronic silencing of motion-sensitive lamina neurons with TNT and Kir2.1 (see [Figure 3.9](#)), and chronic silencing of T4 and T5 lobula plate neurons with Kir2.1, (see [Figure 3.10](#)), were achieved using the genetic crosses described above.

We performed acute inactivation of T4 and T5 cells using temperature sensitive shibire. Control flies (Cont<sup>shi</sup>) were paired with PIBL females at a permissive temperature ( $\sim 22^{\circ}\text{C}$ ). Experimental flies were placed at a non-permissive temperature ( $\sim 27^{\circ}\text{C}$ )  $\sim 30$ mins prior to being paired with a PIBL female, and the entire courtship bout was recorded at the non-permissive temperature.

### 3.4.10 Data processing and analysis

All data processing and analysis was conducted in MATLAB (MathWorks, Inc.).

#### Significance Testing

Unless otherwise specified in the text, significance tests were as follows: To determine significant differences between means, we used one-way ANOVA or KruskalWallis analyses, depending on whether the distributions in question were non-normal (as identified by the Jarque-Bera test for normality). For confidence intervals in [Figure 3.1e-f](#), non-overlap was used as a conservative estimate of significant differences. For normally distributed data, variance was estimated to be similar between groups, as required for applying ANOVA. Given variability in behavior, we sampled a large number of flies for each wild type strain ( $\sim 40$ ) before performing any statistical tests. After observing the effect size in wild type data, we

then collected between 5 and 40 flies (depending on the analysis) for other genotypes and manipulations. Blinding was not necessary for these experiments because all data analyses were automated. As all statistical tests were performed between different fly strains, or between flies of the same strain with an internal control, randomization of animal groups was not necessary.

### **Normalizing pulse amplitude in the behavioral chamber**

We defined the raw amplitude of each pulse as its maximum noise-subtracted amplitude averaged across the three microphones with the largest signal. We took a total of 600,000 pulse amplitudes produced by our 8 wild type strains when courting PIBL, SP, or AC females. To avoid normalizing for differences in Dis throughout the chamber, we ordered these pulses by the Dis at which they were produced and split them into groups of 20,000 (all results in this paper are effectively unchanged without this step). For each group of 20,000 pulses, we divided the pulse amplitude produced at each point within the chamber (bin-width 0.5mm) by the mean amplitude produced by that particular fly (within that distance-based group of 20,000 pulses). All pulses produced >11mm from the center of the chamber were excluded due to data limitations (~10% of pulses), and flies which did not produce >20 pulses within the group of 20,000 were not included.

We then recombined the 600,000 pulses and calculated the fractional deviation from the mean (across all locations) at each position in the chamber with 2-dimensional kernel density estimation (see [Figure 3.1b](#)). Using this map of fractional deviation, all pulse amplitudes were scaled depending on the position of the male at the time of pulse production. These values were then used to compare the song amplitudes produced by different strains and genetic manipulations (see [Figure 3.1c-d](#), [Figure 3.5a](#), [Figure 3.11a-b](#), and [Figure 3.12b](#)).

For analyses involving “Normalized amplitude” (throughout the paper), pulse amplitudes were further normalized before combining individual flies. Flies were required to sing a minimum number of pulses (specified in each figure legend) and the amplitudes were then

z-scored (mean subtracted and then divided by the STD). Thus, values of “Normalized amplitude” represent the number of standard deviations from the mean amplitude and can take positive and negative values.

### **Normalizing pulse amplitude for tethered flies**

To control for small differences in the size and position of each fly, and differences in the sensitivity of the two microphones, we separately normalized the amplitude of pulses produced by each wing. Pulses which were recorded with approximately equal amplitude on each microphone (within 25%) were excluded as these could not be attributed to a particular wing. The remaining pulses were then grouped by wing and z-scored separately before being combined for analysis. Thus, the tethered version of “Normalized amplitude” is also measured in standard deviations from the mean amplitude.

### **Fly tracking and song segmentation**

Both tracking and segmentation were as previously described [77].

### **Quantification of movement parameters**

Males can approach the female from either side during courtship. As a result, distinctions between left/right lateral movement and anticlockwise/clockwise rotations or angles are not meaningful for the analyses we performed. For this reason, only forward velocities were allowed to take negative values and we use absolute values for lateral movements and rotations.

### **GLM implementation**

For this work, we used GLMs only for model selection, and so used 100% of the data for model fitting (see [Figure 3.1e-f](#), [Figure 3.2a](#)). We fit all models using the method described in [49], which uses a sparseness prior to penalize redundant features. Movement features were z-scored for each fly before running the model to account for any trivial differences in

behavior. Given the non-binary nature of our model output (normalized pulse amplitudes) we used a Gaussian link function. To calculate the deviance reduction for each fit, data were randomly split into 4 equal subgroups for cross-validation, and this process was repeated 1000 times to estimate 95% confidence intervals.

## Model selection

When determining which movement features were most predictive of normalized pulse amplitude (see [Figure 3.1e-f](#)), we considered a stimulus history of 1s prior to pulse production for all features. We first selected the single feature which produced the largest deviance reduction (Dis), and then re-ran models with paired inputs (e.g. Dis and mFV), to determine if adding a second feature improved model performance. Subsequent features were added until the relative improvement from an additional feature was not significantly different from zero, but adding a third feature beyond Dis and mFV produced no significant improvement in deviance reduction.

## Natural autocorrelation in movement features

We extracted movement features from all wild type flies courting PIBL females whose trials lasted longer than 20 minutes (to obtain sufficient data for the autocorrelation). For each delay, the autocorrelation was normalized by a correlation with shuffled data. The mean correlation across flies was scaled to have a minimum of zero and maximum of 1 for ease of visualization (see [Figure 3.2b](#)).

## Amplitude modulation curves

When testing for amplitude modulation (for example, see [Figure 3.2c](#)) we binned the normalized pulse amplitudes produced by all flies of a particular strain/manipulation according to the pre-Dis (or Dis) for each pulse. We used a bin-width of 0.2mm unless data-limited (for example, when analyzing pulses produced when facing away from the female, [Figure](#)

3.5d). After combining flies, bins with  $<100$  pulses were excluded (unless otherwise specified in figure legends), before  $r^2$  values were calculated. SEM values are included only for illustrative purposes as subsequent pulses cannot be considered independent samples.

All pulses produced at  $\text{Dis} < 1.99\text{mm}$  were excluded as tracking errors because fly centers are very unlikely to be this close given their body length ( $\sim 0.5\%$  of pulses). For non-heat-activated flies, pulses produced at  $\text{Dis} > 12.01\text{mm}$  were also excluded as segmentation errors since flies rarely naturally sing beyond that distance ( $\sim 0.8\%$  of pulses).

When testing for amplitude modulation in tethered flies, we used the same methodology, but binned hAng or vAng, rather than pre-Dis, with a bin-width of  $1^\circ$ .

### Testing for correlations between pulse parameters

For Figure 3.3b, e, and h, empirical joint probability density functions (PDFs) were calculated by binning data in 2 dimensions. For Figure 3.3c, f, and i, independent PDFs were calculated as the outer product of the empirical marginal PDFs. In all cases, PDFs were calculated using only values within 1 STD of the mean, with bin-width 0.1 STDs, because the sparsity of values prevented accurate empirical PDF estimation beyond this range. Since two pulses are required to calculate an IPI, the initial pulses of each train were not included when testing for a correlation between IPI and amplitude.

### Responses to multi-level changes in light intensity

For Figure 3.7b-c, peaks were defined as significant if the mean normalized amplitude produced by flies at the peak was significantly different from the mean normalized amplitude produced during the 10 time-points preceding the intensity transition. This analysis was aimed at identifying the speed of intensity responses and necessitated narrow bins within the time domain. For later analyses (see Figure 3.7d-e and Figure 3.8) a “response” is calculated from the difference in normalized amplitude between pulses produced up to 45ms



before the switch, and between 15ms and 60ms after the switch. Thus, a positive response indicates that males produced louder pulses after the light intensity transition.

### **Visual stimuli presented to tethered flies**

Visual stimuli were generated using tracking statistics from  $P1^{Act}$  males paired with PIBL females. We concatenated the data from 11 flies and eliminated all sections where the female moved more than  $45^\circ$  from the males line of sight (i.e.  $Ang2 > 45^\circ$ , see [Figure 3.5d](#)). This produced  $>115$  minutes of natural variations in  $Ang2$  from which we randomly selected a 15 minute segment for each trial. To replicated naturalistic variations in the size of the stimulus, we extracted 15 minutes of  $Dis$  values from the same sample of time-points. The random selection of  $Ang2$  and  $Dis$  sections were independent, so the azimuthal location of the stimulus was uncorrelated with its size. Relative changes in  $Dis$  were used to control the dimensions of the stimulus. For example, if  $Dis$  doubled, the vertical and horizontal dimensions of the stimulus halved. This naturalistic stimulus was used for [Figure 3.13b-c](#).

For experiments where the stimulus was stationary in the azimuth, variations in size were still produced in the same way ([Figure 3.13d](#)). To independently vary the stimulus size along horizontal and vertical dimensions, we randomly selected two different 15 minute segments of  $Dis$  ([Figure 3.13e-f](#)).

- [11] Gidon Felsen and Yang Dan. A natural approach to studying vision. *Nature neuroscience*, 8(12):1643–1646, 2005.
- [12] Alex Gomez-Marin, Joseph J Paton, Adam R Kampff, Rui M Costa, and Zachary F Mainen. Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, 17(11):1455–1462, 2014.
- [13] David J Anderson and Pietro Perona. Toward a science of computational ethology. *Neuron*, 84(1), 2014.
- [14] Yi Deng, Philip Coen, Mingzhai Sun, and Joshua W Shaevitz. Efficient multiple object tracking using mutually repulsive active membranes. *PloS one*, 8(6), 2013.
- [15] Fabrice de Chaumont, Renata Dos-Santos Coura, Pierre Serreau, Arnaud Cressant, Jonathan Chabout, Sylvie Granon, and Jean-Christophe Olivo-Marin. Computerized video analysis of social interactions in mice. *Nature Methods*, 9(4):410–417, 2012.
- [16] Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99), 2014.
- [17] Joshua T Vogelstein, Youngser Park, Tomoko Ohyama, Rex A Kerr, James W Truman, Carey E Priebe, and Marta Zlatic. Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science*, 344(6182):386–392, 2014.
- [18] Eviatar Yemini, Tadas Jucikas, Laura J Grundy, André E X Brown, and William R Schafer. A database of caenorhabditis elegans behavioral phenotypes. *Nature Methods*, 10(9):877–879, 2013.
- [19] Adam J Calhoun, Sreekanth H Chalasani, and Tatyana O Sharpee. Maximally informative foraging by caenorhabditis elegans. *eLife*, 3, 2014.
- [20] C C Ioannou, V Guttal, and I D Couzin. Predatory fish select for coordinated collective motion in virtual prey. *Science*, 337(6099):1212–1215, 2012.
- [21] M Mller and R Wehner. The hidden spiral: systematic search and path integration in desert ants, cataglyphis fortis. *Journal of Comparative Physiology A*, 175(5):525–530, 1994.
- [22] Rüdiger Wehner and Bruno Lanfranconi. What do the ants know about the rotation of the sky? *Nature*, 293(5835):731–733, 1981.
- [23] Matthias Wittlinger, Rüdiger Wehner, and Harald Wolf. The ant odometer: stepping on stilts and stumps. *Science*, 312(5782):1965–1967, 2006.
- [24] M Wittlinger, R Wehner, and H Wolf. The desert ant odometer: a stride integrator that accounts for stride length and walking speed. *Journal of Experimental Biology*, 210(2):198–207, 2007.