Supplemental Experimental Procedures:

Fly Stocks

We used the following GAL4 lines:

VT029306 (cell types AV1, AV5, and AV6)

VT011148 (cell type AV6)

VT050279 (cell types AV2 and AV3)

VT050245 (cell type AV4)

VT002042 (cell type V1)

NP1046 and VT049365 (cell type B2)

VT lines came from the Vienna Tiles GAL4 collection (BJD, unpublished, see http://stockcenter.vdrc.at for z-projections of these lines). NP1046 was a gift from A. Kamikouchi (Kamikouchi et al., 2006). Cell types AAL1 and V2 were identified by dye fills following patch clamp recordings (see below and Supplemental Movie M1) not by GFP labeling with a particular genetic enhancer line. Each GAL4 line was crossed to UAS-eGFP2x to target identifiable neurons for patch clamp recordings, and to UAS-DenMark2; UAS-synaptotagmin-GFP (Nicolaï et al., 2010; Zhang et al., 2002) for identifying dendrites and axons, respectively (see Table 1).

Immunohistochemistry

Brains were dissected in phosphate-buffered saline (PBS) and fixed for 45-50 minutes in paraformaldehyde (4%) at room temperature, and then blocked in PBS + 0.1% Triton + 5% normal goat serum. The following primary antibodies were used: mouse anti-nc82 (1:20) (DSHB nc82), rabbit anti-GABA (1:100) (Sigma-Aldrich A2052), mouse anti-DChAT (1:1000) (DSHB ChAT-4B1), rabbit anti-DsRed (1:500) (Clontech 63296), and mouse anti-GFP (1:1000) (Life Technologies A-11120). The following secondary antibodies were used: goat anti-mouse AlexaFluor594 (1:300) (Life Technologies A-11005), goat anti-rabbit AlexaFluor633 (1:400) (Life Technologies A-21070), goat anti-mouse AlexaFluor488 (1:200) (Life Technologies A-11029), goat anti-rabbit AlexaFluor555 (1:200) (Life

Technologies A-21429), and goat anti-mouse AlexaFluor568 (1:400) (Life Technologies A-11004). To detect biocytin from neuronal fills, we used Strepavidin AlexaFluor568 (1:300) (Life Technologies S-11226). Brains were mounted in Vectashield (Vector labs H-1000) and imaged under a confocal microscope.

Anatomy

From dye fills made during patch clamp recordings (see below), we were able to unambiguously identify the recorded neuron type in each VT GAL4 line confocal image stack (generated by the Dickson Lab; available at http://brainbase.imp.ac.at). We used the SimpleNeuriteTracer plugin in Fiji (Schindelin et al., 2012) to manually trace the specific neuron type in the VT line image that was morphed to a template brain (Longair et al., 2011). These traces where then registered to the FlyCircuit reference brain (Manton et al., 2014) (code from https://github.com/jefferislab/BridgingRegistrations, accessed 3/22/2015). We then searched the Flycircuit database (http://www.flycircuit.tw (Chiang et al., 2010)) with these traces either manually or using the nBlast software package (Costa et al., 2015). Manual searches were performed with region-based queries. nBlast searches used the traces as a query after bridge registration from the Vienna Tile template to the FlyCircuit template. Neuron skeletons were then selected based on comparison with the original fills and are plotted in Figs. 2A and 2C and Fig. S2. Queries and plotting were performed in R based on code modified from http://github.com/jefferislab/NBLAST_figures, accessed 3/22/2015). The fill for V2 did not allow us to identify it in the FlyCircuit database. We added the fill as a supplemental movie M1.

Neural Recordings and Analysis

Fly Preparation and Electrophysiology

Virgin female flies (1-2 days old) were mounted and dissected as described previously (Murthy and Turner, 2013; Tootoonian et al., 2012). For *in vivo* whole-cell patch clamp recordings, neuronal cell bodies were either accessed from the posterior surface of the head (cell types AV1, AV5, AV6 and V1) or

from the ventral surface, through the proboscis (cell types B2, AV2, AV3, AV4, AAL1 and V2). For both types of mounting, the antennae were kept dry and mobile for auditory stimulation. Whole-cell patch clamp recordings were performed as described previously (Murthy et al. 2008). We typically removed the perineural sheath with a suction pipette under visual control. We used patch clamp electrodes pulled from capillary glass (o.d., 1.5; i.d., 1.1) with resistances of ~5-6 MΩ. Only neurons with input resistances > 50mΩ were included for analysis. All neurons were filled with biocytin. Neuronal membrane potential was recorded in current clamp mode (neurons were held between –65 to –59 mV) with a Multiclamp 700B amplifier (headstage CV-7B) and digitized at 10 kHz in Igor Pro (Wavemetrics). Raw data were then further analyzed using Matlab (Mathworks). Recordings typically lasted for 30-60 min.

Sound delivery system

The sound delivery system consisted of i) a sound card (M-Audio Delta 44, 16 bits, 44100 Hz, 2-channel), ii) a 2-channel amplifier (Crown D-75A), iii) a headphone speaker (KOSS, 16 Ohm impedance; sensitivity, 112 dB SPL/1 mW), and iv) a coupling tube (12 cm, diameter: 1 mm). We compensated for distortions in the system as previously described (Tootoonian et al., 2012) by first calculating the transfer function of our system and then by multiplying each stimulus by the inverse of the transfer function before playing it back. To measure sound intensity, we first calibrated a pressure gradient microphone (NR23159, Knowles Electronics Inc., Itasca, IL, USA) under far field conditions in an anechoic chamber against a pressure microphone (4190-L-001, Brüel & Kjaer). Calibration was performed using pure tones (100-1000 Hz, each 2 seconds). The detailed procedures and cross-calibration between the pressure and the pressure gradient microphone were described in (Göpfert et al., 2006). To estimate the sound amplitude of each stimulus we placed the calibrated gradient microphone at the same position as the fly (2-3 mm from sound tube outlet) in separate experiments. The recorded voltage was then converted to particle velocity (with units mm/s).

Acoustic stimuli

The following acoustic stimuli were presented to flies on the electrophysiology rig: i) pure tones (8 frequencies: 100, 160, 200, 250, 300, 400, 550, and 700 Hz, amplitude: 3.7 mm/s, N=32 cells), ii) white noise (frequency range: 80-1000 Hz, amplitudes: 0.25, 0.4, 1.0, and 1.8 mm/s, N=32 cells), iii) synthetic pulses trains (20 pulses at interpulse intervals: 20, 30, 40, 60, 90, and 120 ms, carrier frequency: 200 Hz, amplitudes: 0.9, 1.4, and 5.5 mm/s, N=32 cells), iv) stimuli with a naturalistic bout structure, i.e. artificial pulse trains (IPI 40 ms, duration 1s) separated by pauses drawn from the distribution present in *D. melanogaster* courtship songs (overall stimulus duration 10 minutes, N=12 cells), and v) *D. melanogaster* songs recorded during courtship (10 stimuli, each duration 3 minutes, N=20 cells). Due to limited recording time, we recorded only single trails.

Carrier frequency and intensity tuning curves

We constructed carrier frequency tuning curves from voltage responses to 8 pure tones and intensity tuning curves from voltage responses to white noise delivered at 4 different intensities (see *Acoustic stimuli*). The membrane potential traces were first offset (from resting potential) to zero by subtracting the mean baseline activity 1 s before stimulus presentation. The steady-state response was computed as average Vm from 2-4s of the stimulus presentation). To compare across cell types, tuning curves of each recording were normalized to the maximum and then averaged across individual recordings.

Inter-pulse interval tuning curve

IPI tuning curves were calculated for spiking responses (only for cell types AV1 and AV3, Fig. S3) and for subthreshold Vm responses (all AMMC/VLP neurons, Fig. 2E). For spiking responses, tuning curves were based on the number of spikes produced during the response to each pulse train. For subthreshold responses, we calculated the integral voltage in response to each pulse train, after subtracting the baseline voltage. All raw tuning curves were normalized by the duration of the pulse train to yield a response rate in units of ‘Vm per second’. To compare across cell types, tuning curves of each recording were normalized to the maximum and then averaged across individual recordings.

Behavioral Recordings and Analysis

We used a previously published dataset of natural courtship song and accompanying male and female movements, from pairs of wild type males (from 8 geographically diverse strains) and PIBL (pheromone-insensitive and blind) females (Coen et al., 2014).

We used the following wild type strains:

- Wild type from Nairobi, Kenya; collected by Andolfatto & Bachtrog (2006)
- Canton-S laboratory strain
- Wild type from San Diego, California; collected by Andolfatto (2006)
- Wild type from Cartagena, Colombia; collected by Andolfatto (2009)
- Wild type from the Netherlands; collected by Davis (2000)
- Wild type from Zanzibar, Tanzania; collected by Andolfatto & Bachtrog (2006)
- Wild type from Harare, Zimbabwe; collected by Begun (1993)
- Wild type from Victoria Falls, Zimbabwe; collected by Ballard (2002)

Methods for tracking fly position and song segmentation are described in (Arthur et al., 2013; Coen et al., 2014). A pulse train was defined as a consecutive sequence of pulses with intervals smaller than 200ms. Likewise, sine song was defined as a consecutive stretch of sinusoidal modulation in the song with a pause smaller than 200ms. Bouts were defined as continuous sequences of pulse or sine song or combined pulse and sine song, with intervals shorter than 200 ms. Pauses were defined as intervals between song or pulse trains longer than 200ms. We described song structure using three types of summary statistic (amount, onset, and duration), calculated within sliding time windows of length 5-120s, with 50% overlap between windows. ‘Amount’ was defined as the total amount of each song type (pulse, sine, bout=all song). ‘Number of onsets’ corresponded to the total number of segments for each song type (pulse, sine, bout). ‘Duration’ was the average duration of each song type (pulse, sine, bout, pause). We also quantified the average duration of IPIs. This gave us a total of 11 song features (see Fig. 1G).

The choice of time window only affected the overall magnitude of correlations between song features and female behavior, not the relative performance of individual features.

The main immediate effect of song on female behavior is a reduction in speed, dependent on her receptivity state (Bussell et al., 2014; Coen et al., 2014). We therefore used the average female speed during each time window as a behavioral readout. Female behavior (including her speed) influences male song choice (pulse versus sine) on timescales of approximately 100-200 ms (Coen et al., 2014). To avoid any confounding effects associated with the male actively patterning his song in response to female speed, we omitted female speed data during (and 300ms before and after) song. We also excluded time windows with less than 5% and more than 95% song.

To visualize the relation between female behavior and song features we binned song feature values by percentiles (~100 data points per bin) and plotted the mean song feature value and mean female speed for each bin. Error bars correspond to the s.e.m. of the female speed for each bin. These plots are termed 'preference functions.'

To determine statistical associations between song features and female speed, we calculated Spearman's rank correlation using the whole data set. We chose rank – instead of the linear – correlation to quantify the association between song features and female speed since preference functions were often nonlinear but monotonous (e.g. saturating). To avoid spurious correlations due to strain-specific differences in speed or song features, we z-scored values for speed and song features for each wild type male strain (all females were of the same genotype).

To ensure that bout duration was sufficient to predict female speed, we determined the performance of bivariate models. We fitted linear models of the form $v' = a_0 + f_1 + a_2 f_2 + a_3 f_1 f_2$, where v' is the female speed, f_1 and f_2 are the two song features and a_i are the model coefficients (a_3 is the interaction term). In the case of $f_1 = f_2$, this corresponds to a univariate quadratic model. We then computed the rank correlation between v' and the actual female speed.

Computational modeling - encoder models

We used encoder models to quantitatively describe how sound is represented in AMMC/VLP neurons and as a tool to predict responses to courtship song.

Linear-nonlinear (LN) models

LN models describe the transformation from the stimulus to the neural response using a linear filter and a static nonlinearity function. The linear filter $F(\tau)$ in our case corresponds to the cell's preferred temporal feature, while the nonlinearity *N(g)* transforms the filtered stimulus *g* to the predicted response $\hat{r}$ and accounts for rectification or saturation in the response: $\hat{r}(t) = N(g(t)) = N(\int d\tau F(\tau)s(\tau - t))$, where $s(t)$ is the stimulus envelope (extracted using the root mean square method, sigma=16 ms). We chose the stimulus envelope as the input to the LN model since none of the AMMC/VLP neurons we recorded resolved the carrier frequency of the elements of courtship song, but rather represent the smoothed, time-varying amplitude of the stimulus (Fig. 2D, 3E, Fig. S5A,B).

The standard method for fitting such models is by stimulus-response correlation (Schwartz et al., 2006). However, this correlation yields an unbiased estimate of the filter only if the stimulus is Gaussian. Since our stimuli are non-Gaussian, we fitted the model by directly minimizing the mismatch (mean-squared error) between the actual and the predicted response: $argmin_{F,N}[(r(t) - \hat{r}(t))^2]$. This objective function was minimized using ridge regression, which adds a penalty term to the objective function that suppresses filter components that do not contribute to the model performance. We used the empirical Bayes implementation described in (Park and Pillow, 2011) with code publicly available from http://pillowlab.princeton.edu/code/code_ALD.html.

The membrane voltage and the stimulus envelope were down sampled to 250 Hz. For fitting we used the responses to pulse train stimuli of intensity of 5.5 mm/s, since they were presented for all recordings and responses exhibited the best signal-to-noise ratio. We did not observe any strong intensity dependence of the model properties. Filter duration was 2048 ms and filters were estimated on a basis of 16 raised cosine bumps with nonlinear spacing (Fig. S4A). The input-output function N was

estimated using smoothing-spline interpolation between the linear prediction and the actual response (Matlab's fit function with the smoothingspline option and the smoothing parameter set to 0.5).

Adaptive LN models

All recordings exhibited a prominent decrease of response amplitude over the duration of the pulse trains. Since the standard LN could not account for this response property of AMMC/VLP neurons, we explicitly added adaptation to the model. Adaptation can have multiple sources that are hard to isolate experimentally. Possible sources in an AMMC/VLP neuron are: JON (Johnston's Organ neuron or auditory receptor neuron) adaptation, synaptic depression at JO-AMMC or AMMC-VLP synapse, adaptive conductances in dendrites and axons of AMMC neurons, pre or post-synaptic inhibition. Since the identification of the biophysical bases of adaptation was beyond the scope of this study, we sought a phenomenological description that could characterize the stimulus transformation associated with adaptation while remaining agnostic as to the specific source(s) of adaptation. We found that a model based on a depressing synapse acting divisively on the input worked well for reproducing the adaptation in our recordings (Tsodyks et al., 1998).

$$d(t; v, \tau) = d(t-1) + s(t-1) \cdot [1 - d(t-1)] \cdot v - \frac{d(t-1)}{\tau}$$

Here *d* is the depression status, *s* is the stimulus envelope, *v* is the rate of vesicle depletion per unit of presynaptic potential, and τ is the recovery time constant.

To account for the fact that adaptation is likely mediated by several sources and hence occurs on multiple time scales, we implemented adaptation as a bank of depressing synapses (David and Shamma, 2013).

$$d(t) = \sum_{v,\tau} \mathrm{w}(v, \tau) \cdot d(t; v, \tau)$$

$$s_d(t) = s(t) \cdot [1 - d(t)]$$

d(t;v,τ) is the 'depression trace' for each individual time scale tau, *w(v, τ)* the weight of each time scale, *d(t)* the total depression trace, and $s_d(t)$ is the depressed stimulus that forms the input to the standard LN model.

Preliminary analyses indicated that eight different time scales τ (32, 64, 128, 256, 512, 1024, 2048 ms) with a single depletion rate v (0.2) plus one non-adapting input synapse was sufficient to reach optimal model performance.

Fitting this equation directly to the data involves solving a nonlinear regression problem. We hence applied a 'kernel trick' as in (David and Shamma, 2013), by pre-computing synaptic depression traces for all parameter combinations (v,τ) and representing the depressed stimulus as a weighted sum of these 'basis traces' – one for each adaptation time constant. Thereby, the problem is reduced to a linear, constrained (all $w(v,\tau) \geq 0$) optimization problem. This approach thus constitutes a computationally efficient way to incorporate adaptation on multiple timescales.

The complete model (adaptation stage followed by the LN) is fitted iteratively in two steps:

0. Initialize parameters with filter and nonlinearity from LN model without adaptation

1. Optimize adaptation parameters (weights for each adapted 'basis function') by minimizing the mean squared error between prediction and response, while holding the filter fixed but using an optimal nonlinearity.

2. Find filter and nonlinearity that transform the adapted stimulus to the Vm as described for the LN models above while holding the adaptation parameters fixed.

Repeat until convergence.

This procedure usually converges to a stable set of adaptation parameters and filter/NL after two cycles. Filter shapes change only little when adding adaptation.

We tried several alternatives to this divisive input gain control model (subtractive input gain control, a mixture of subtractive and divisive input gain control, subtractive or divisive output gain control (acting on the output of a standard LN model)). Neither of the variants significantly improved prediction

performance, but they all made the fitting procedure computationally more demanding and unstable. Note that subtractive output gain control is implemented through the filter's negative lobe.

Model evaluation

Model fit was estimated using cross-validation. That is, we used 3/4 of the data from each recording for fitting the responses and the remaining 1/4 to estimate the linear coefficient of determination r^2 between predicted and actual neuronal responses. We did not use rank correlation since the nonlinearity stage in the aLN models accounted for weak nonlinearities in the relation between model output and neuronal response. To quantify how well aLN models fitted to pulse trains generalized to courtship song, we compared the performance of these models to those directly fitted to responses to courtship song excerpts (see *Acoustic Stimuli*). The performance of the latter model variant constitutes an upper bound for the models fitted to pulse train responses when tested on courtship song responses. We also tested the models fitted to pulse train responses to responses to longer pulse train stimuli with a naturalistic bout structure (see *Acoustic Stimuli*).

To determine whether filter shapes and adaptation parameters (weights for individual time constants) were cell-type specific, we compared the similarity of these parameters across and within cell types. As a similarity measure we used Pearson's correlation coefficient *r*.

Computational modeling - decoder models

We sought to relate song and female behavior to the neural representation in the brain. To that end, we first predicted AMMC/VLP responses to the same natural courtship songs used in the behavioral analysis and designed a simple readout that could explain the observed female speed.

Generation of neuronal surrogate responses for natural song using the aLN encoder model

Since the original recordings did not allow a reliable estimate of the song amplitude as received by the female – which changes with the male's distance and angle to the female - we generated naturalistic

song segments from the original recording by assuming a pulse amplitude of 1 mm/s and a sine song amplitude of 0.25 mm/s. This choice of amplitude values reflects the fact that sine song is usually softer than pulse song (see Fig. 1A). Results were robust to the specific choice of absolute and relative amplitudes of sine and pulse.

Prediction of female speed from surrogate responses

The Vm predicted by the aLN model for each cell was transformed using a piecewise linear approximation to a sigmoidal nonlinearity with a threshold θ and a saturation term σ:

$$\hat{r}_T(t) = \begin{cases} 0 \; if \; \hat{r}(t) < \theta \\ 1 \; if \; \hat{r}(t) > \sigma \\ \hat{r}(t) \; otherwise \end{cases}$$

The output $\hat{r}_T(t)$ of this nonlinearity was then integrated to yield a prediction of the female speed *v'* for each time window. We selected the optimal parameters of the nonlinearity (θ, σ) for each cell (i.e., those that maximize Spearman's rank correlation between female speed and decoder prediction) using a grid search with values spanning the range of aLN response values in 41 steps (820 parameter combinations after excluding all parameter combinations where the threshold is larger than the saturation point). For the modified decoder, we integrated only the positive or negative parts (relative to baseline) of the nonlinearly transformed aLN responses. The parameters of the nonlinearity were the same as those for the original decoder. The predicted female speed was then given by the ratio of the integrated positive and integrated negative parts of the aLN response.

To assess the role of filtering and adaptation in predicting female speed from aLN responses, we ran the same analysis using predicted responses either lacking a filter (aN models) or lacking the adaptation term (LN models). Spearman's rank correlation between the decoder prediction and female speed or song features as well as tuning curves for visualization were computed as for the behavioral analysis.

Supplemental Figures and Legends

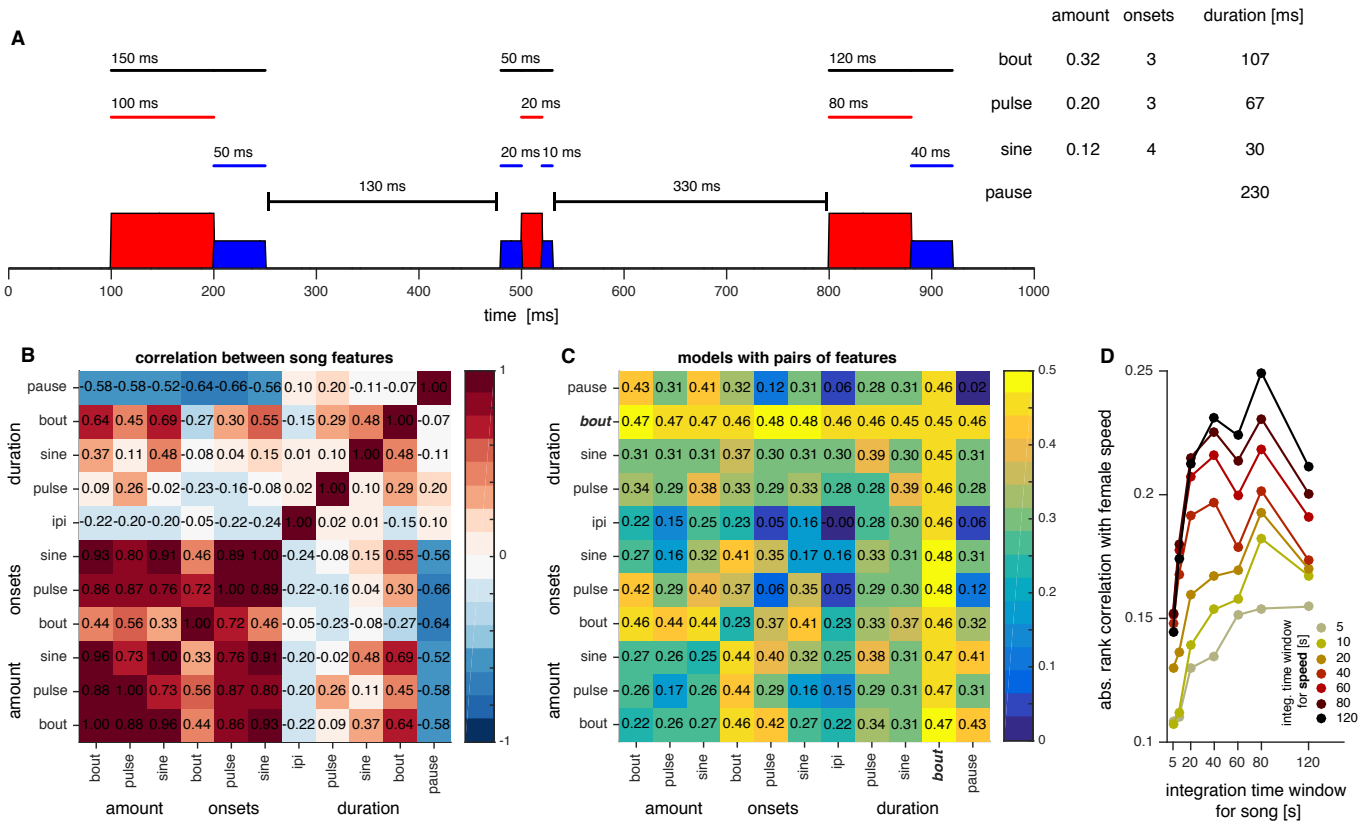

Fig S1 Song features and their correlation with behavior (refers to Fig. 1)

A Extraction of the song features from natural courtship song. Song is arranged in bouts which consist of pulse and sine mode (compare with Fig. 1A). We quantified the duration of individual modes and full bouts as well as the pause between bouts. Likewise, we counted the number of individual pulse, sine or bout onsets. The song features are then given as the average values of individual song elements over a time window of typically 60 seconds. Amount was given by the fraction of time window comprised of sine, pulse or song bouts.

B Matrix of rank correlations between pairs of song features (for color-coding see color bar).

C Performance of two-variable models. Rank correlation between female speed and the output of linear models whose input was pairs of song features, including an interaction term (see Experimental Procedures). The best two feature model only marginally outperforms correlations with bout duration (best two feature model: ρ=0.48 versus bout duration only model: ρ=0.45). All models that include bout

duration show similar performance (for color-coding see color bar). Thus, combining bout duration with other song features does not improve the rank correlation with female speed.

D Rank correlation between amount of song and female speed for non-overlapping windows of song and speed. Duration of the window for song integration is on the x-axis. Duration of the window for speed integration is color coded (see legend). The rank correlation saturates for time windows exceeding the duration of IPIs or individual song bouts, supporting our conclusion that female flies are sensitive to song structure on long timescales. N=3896 minutes of song.

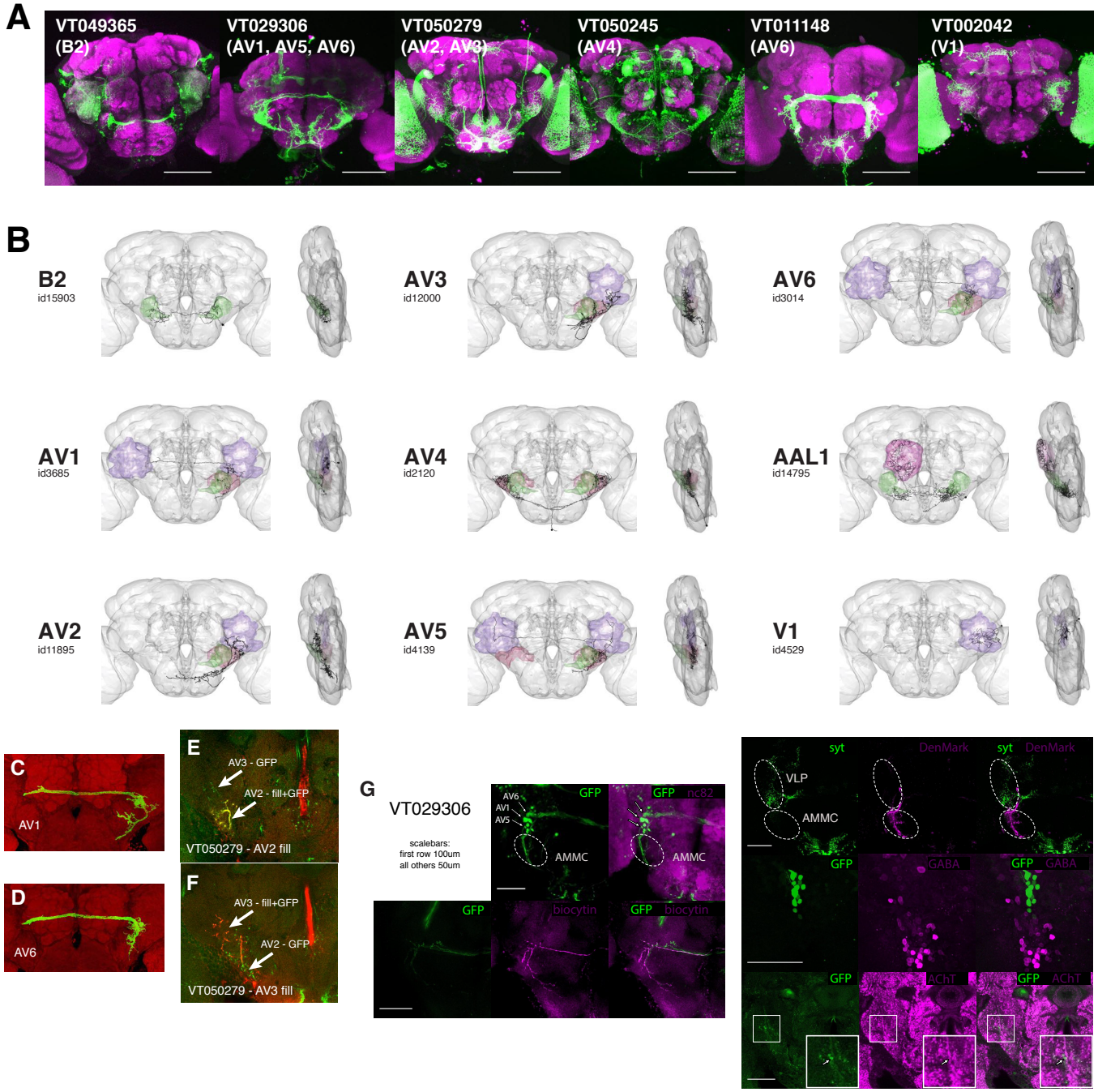

Fig. S2 Morphology and projections of recorded neurons (refers to Fig. 2)

A We recorded from neurons in the following VT lines (VT049365 – B2; VT029306 – AV1, AV5, AV6; VT050279 – AV2, AV3; VT050245 – AV4; VT011148 – AV6; VT002042 – V1). Scale bar in each panel corresponds to 100 microns.

B Frontal and lateral brain views of single neuron traces from the FlyCircuit database (Chiang et al., 2011, http://flycircuit.tw). Each recorded neuron was filled post-hoc. These fills were used to inform digital

tracing of the recorded neuron in the VT line image, which was warped to a standard reference brain (see Experimental Procedures). We then used NBlast (Costa et al., 2015) and region-based queries to search the FlyCircuit database for the recorded neuron type (see FlyCircuit id for each trace). Anatomical structures: AMMC – green, iVLP – red, pVLP – blue, AL – purple.

C, D Cell types AV1 and AV6 appear similar in the FlyCircuit single neuron traces (see **B**) – however, each cell type exhibits characteristic anatomical features that discriminate them as different morphological types. Single cell MARCM clones of AV1 (**A**) and AV6 (**B**) neurons from the FlyCircuit database (AV1 FlyCircuit id 3685, AV6 FlyCircuit id 3014). AV1 projects laterally into the VLP whereas AV6 has a relatively straight ventral branch. In contrast with the MARCM clone, fills of AV6 always label neurons in both hemispheres along with the giant fiber neuron (GFN), due to gap junction coupling (Phelan et al., 2008). AV1 fills only label a neuron in one hemisphere (see Supplemental Movies of neuronal fills S2_AV1.avi and S5_AV6.avi).

E, F Cell types AV2 and AV3 also appear similar. AV2 projects more superficially (anterior) and ventrally into the VLP versus AV3 (**E**). AV3 projects more dorsally into the VLP than AV2 (**F**). See also Supplemental Movies of neuronal fills S3_AV2.avi and S4_AV3.avi.

G Example of anatomical characterization performed on each VT GAL4 line (see Table 1). **(left)** Brain of female carrying VT029306, UAS-eGFP2x transgenes. The soma positions of the three recorded cell types from this GAL4 line are indicated along with the location of the AMMC. Cell type AV1 is filled with biocytin. **(right**) Brain of female carrying either VT029306, UAS-syt-GFP, UAS-DenMark transgenes (top row) or VT029306, UAS-eGFP2x transgenes (bottom two rows). Axons (green, syt-GFP) project to the VLP, whereas dendrites (magenta, DenMark) innervate the AMMC. **A**nti-GABA antibody labels the somata of GABAergic neurons, while anti-AChT antibody labels the processes of cholinergic neurons.

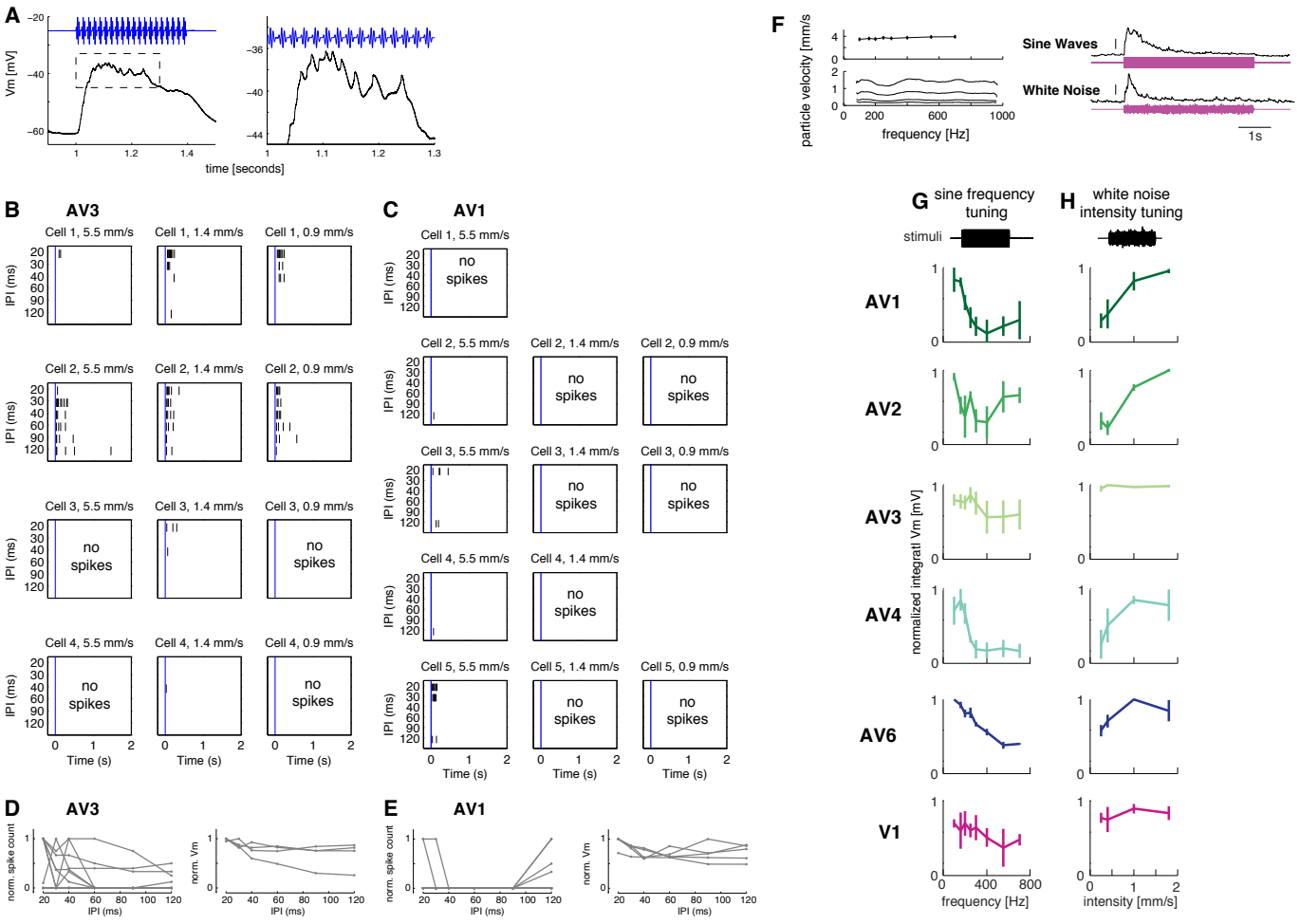

Fig. S3 Spiking responses and tuning of AMMC/VLP neurons (refers to Fig. 2)

A Spikes can be detected in recordings of AMMC/VLP neurons. AV3 responses (black) to pulse train stimuli (blue). Plot at right is blow up of region marked by dashed rectangle.

B,C AV3 and AV1 neurons spike infrequently and unreliably in response to pulse trains. Spike raster plots for different recordings of AV3 (**B**) and AV1 (**C**) – each cell was recorded in a different animal. Rows correspond to individual cells, stimulated with pulse trains of different IPI and intensity (measured in mm/s). Vertical blue line marks stimulus onset.

D, E IPI tuning curves for spike count versus Vm for AV3 (**D**) or AV1 (**E**) . To facilitate comparison, curves of individual recordings were normalized. Tuning based on Vm is more reliable across cells.

F Intensity of sinusoidal and band-limited (80-1000Hz) white noise stimuli. See Experimental Procedures for information on calibration of sound delivery system. (right) Example Vm responses of neuron type AV1 to a 300 Hz tone at 3.7 mm/s or white noise at 1.8 mm/s.

G Normalized integral Vm in response to sinusoids of different frequencies (intensity 3.7mm/s).

H Normalized integral Vm in response to white noise stimuli presented at different intensities.

Plots in **G** and **H** are mean +/- s.e.m. across recordings of a given cell type; each recording is from a different animal. Tuning curves of individual recordings were normalized before averaging. Color follows the scheme in Figs. 2 and 3 of the main mansucript.

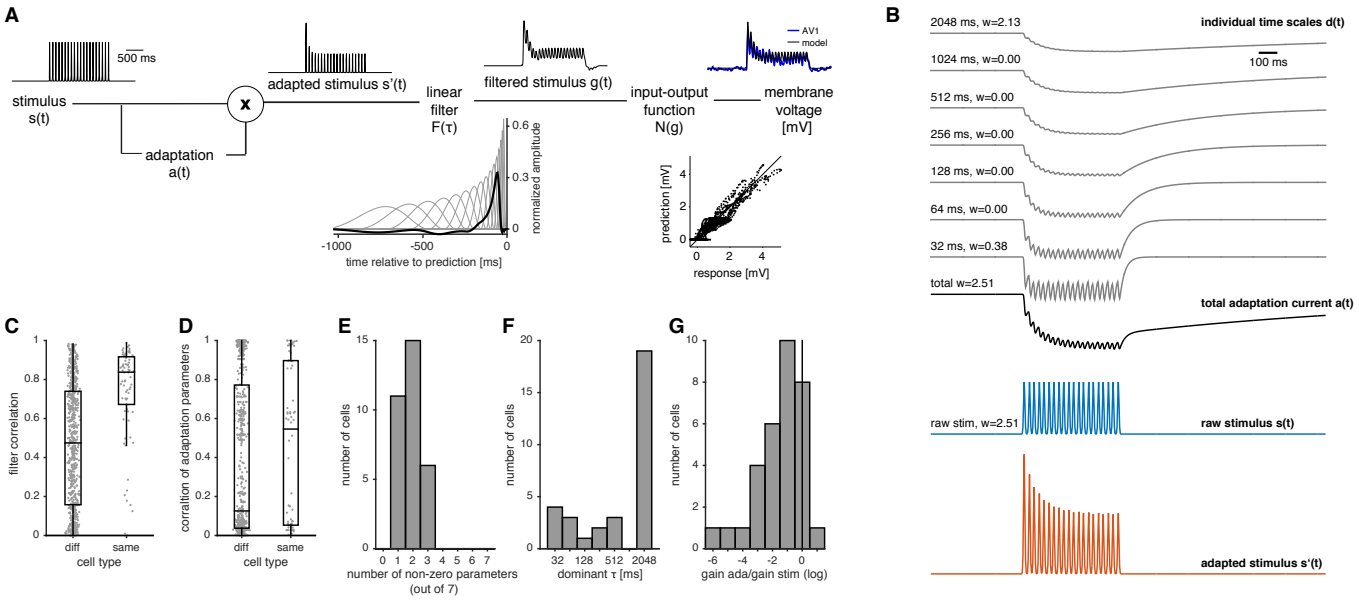

Fig. S4 Modeling AMMC/VLP responses to synthetic pulse trains (refers to Fig. 3)

A Detailed schematic of the adaptive linear-nonlinear model (aLN): The stimulus envelope s(t) is pre-processed in an adaptation stage: The output of a bank of depressing synapses with different adaptation timescales is weighted to yield the total depression term a(t) that is multiplied with the stimulus (see **B** for details on the adaptation stage). This adapted stimulus s'(t) forms the input for a standard linear-nonlinear model with a filter F(τ) and an input-output function N(g). Model parameters are fitted by minimizing the squared error. The resulting model prediction matches the measured membrane voltage (Vm) very well (r^2=0.88, small plot at bottom right of this panel). The filter F(τ) is defined as a weighted sum of basis functions spread nonlinearly across the duration of the filter (inset below F(τ)).

B Traces for the individual adaptation time scales d(t) included in the model. The adaptation current a(t) (red) in (**A**) is given by one minus the weighted sum of individual traces.

C, D Cell-type specificity of model parameters. Correlation between filter shapes (**C**) and adaptation parameters (**D**) when comparing within vs. across cell types. High cell-type specificity corresponds to high correlation between parameters within but not across cell types. This is true for the filter shapes but not for adaptation parameters, indicating that the adaptation parameters are relatively unspecific to cell type (r^2 of filters across type 0.47(0.58) (median(IQR)) versus within type 0.84(0.24), $p<6\ 10^{-14}$, rank sum; r^2 of adaptation parameters across type 0.13(0.73) versus within type 0.55(0.84), p=0.11, rank sum)

E The number of adaptation timescales with non-zero weight for each cell.

F The dominant adaptation timescale (given by largest weight) for each cell in the data set.

G Adaptation strength of all cells in the data set. Adaptation strength was defined as the ratio of the weight of all 7 adapting synapses and the weight of the non-adapting synapse ('raw stim' in (**B**)).

N=32 cells for C-G.

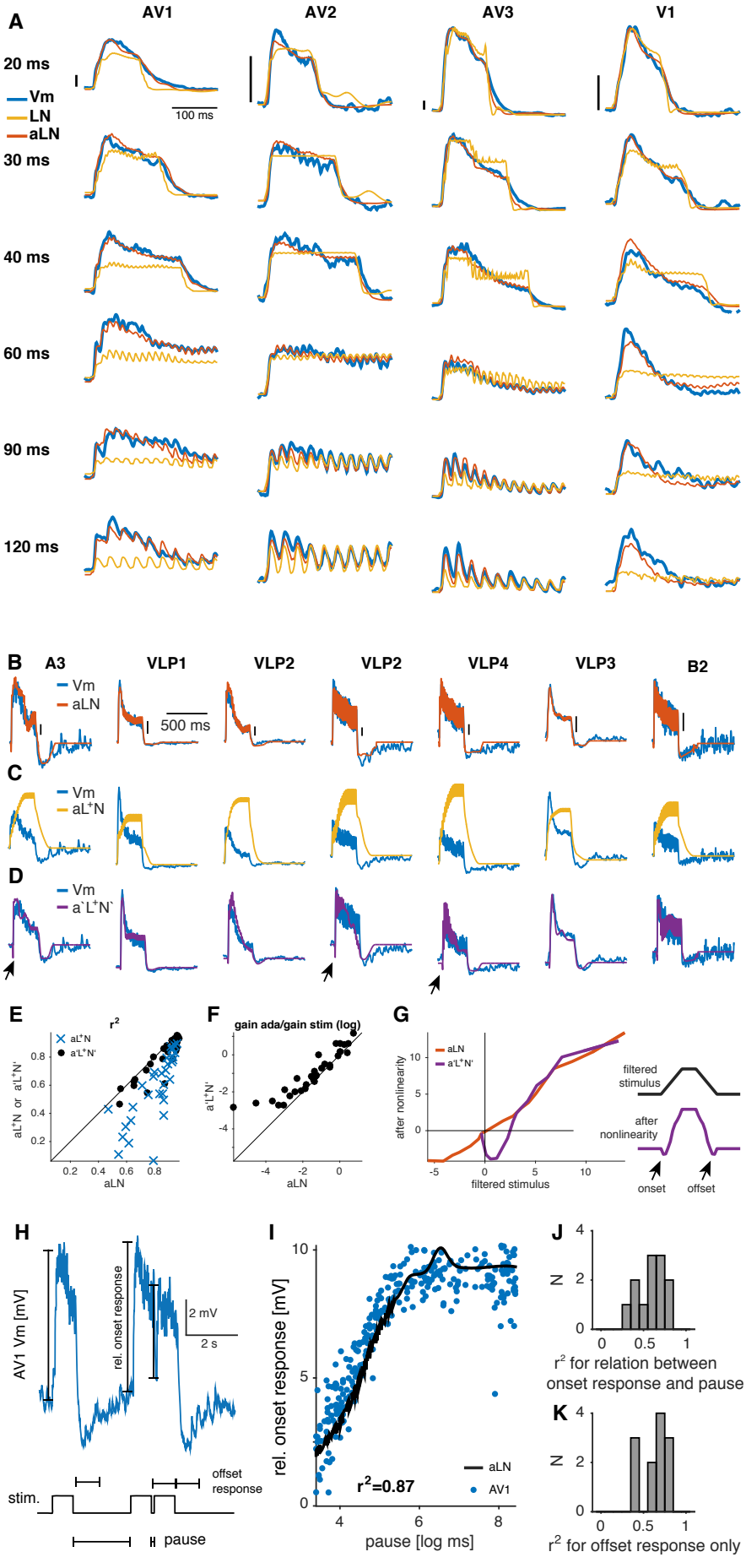
A
AV1
AV2
AV3
V1
20 ms
Vm
LN
aLN
100 ms
30 ms
40 ms
60 ms
90 ms
120 ms
B
A3
VLP1
VLP2
VLP2
VLP4
VLP3
B2
Vm
aLN
500 ms
C
Vm
aL+N
D
Vm
a`L+N`
E
r2
aL+N
a'L+N'
aL+N or a'L+N'
aLN
F
gain ada/gain stim (log)
a'L+N'
aLN
G
aLN
a'L+N'
after nonlinearity
filtered stimulus
filtered stimulus
after nonlinearity
onset
offset
H
AV1 Vm [mV]
rel. onset response
2 mV
2 s
stim.
offset response
pause
I
rel. onset response [mV]
r2=0.87
aLN
AV1
pause [log ms]
J
N
r2 for relation between onset response and pause
K
N
r2 for offset response only

Fig. S5 Adaptation within and across pulse trains, and role of the negative filter lobe (refers to Fig. 3)

A The adaptation term of the aLN model is necessary to reproduce observed response adaptation to pulse train stimuli. Recorded Vm response (blue, single-trial response each from one recorded cell of a given cell type) and model predictions (w/ adaptation (aLN) red, w/o adaptation (LN) orange) to pulse trains containing different IPIs (rows). Vertical scale bars in the first row correspond to 2 mV.

B The aLN model also reproduces offset responses. Example Vm response (blue, single-trial response each from one recorded cell) and aLN predictions (red) for seven different cell types.

C Removing the negative lobe of the linear filter (setting all negative weights to zero) in the aLN models (termed $aL^{+}N$ models) strongly affects model responses – both adaptation as well as negative offset responses disappear (yellow – model, blue – Vm, same cells as in B).

D Re-fitting the adaptation parameters and the nonlinearity for a`$L^{+}$N` models largely restores adaptation and negative offset responses (purple – model, blue – Vm, same cells as in B). However, errant negative responses at pulse onset are also present (see arrows and panel G).

E Performance for aLN models versus models with purely positive filters before ($aL^{+}N$, $r^2 = 0.65(0.36)$, blue crosses) and after (a`$L^{+}$N`, $r^2 = 0.86(0.18)$, black dots) re-fitting the adaptation parameters and the nonlinearity. In both cases, constraining the filter to be positive decreases performance ($p<3x10^{-7}$, sign test).

F Models with purely positive filters exhibit 20% stronger adaptation to compensate for the lack of the negative filter lobe.

G A small 'hook' in the nonlinearity (black arrow) aids in reproducing negative offset responses in the absence of any negative weights in the linear filter (red – nonlinearity of the original aLN model, purple – re-fitted nonlinearity for an a`$L^{+}$N` model). However, the shape of this nonlinearity also produces negative responses at the **onset** of the stimulus (see traces to the left of this panel and in panel D, arrows indicate negative onset responses produced by the nonlinearity). We never observed such a response feature in our recordings (see e.g. Fig. 3C,E). Note that the negative onset responses in panel D are shorter than

the offset responses because the filters are asymmetrical (Fig. 3G) and hence produce pulse responses with steep onsets and shallower offsets. Horizontal and vertical black lines correspond to y=0 and x=0, respectively.

H Calculation of the magnitude of the onset and offset responses. Onset response was defined as the difference in the average Vm in the 40ms before and 100 ms after a bout onset. The offset response was defined as the average Vm in the 1000ms following each bout. Note that for short pauses between bouts, the offset response is cut short by the response to the following bout. Vm trace (blue) shows one recording of cell type AV1.

I Dependence of onset response on bout pause for data (blue) and the aLN model fitted to short pulse train stimuli (black, r^2=0.87). Onset response is smallest for short pauses due to adaptation.

J aLN models reproduce the relationship between pause and onset response (r^2=0.63(0.23), median(IQR)).

K aLN models reproduce the offset responses. r^2 between response and aLN prediction in the 1000ms following each bout offset (r^2=0.71(0.26), median(IQR)).

N = 32 cells for E and F; N=12 cells for J and K.

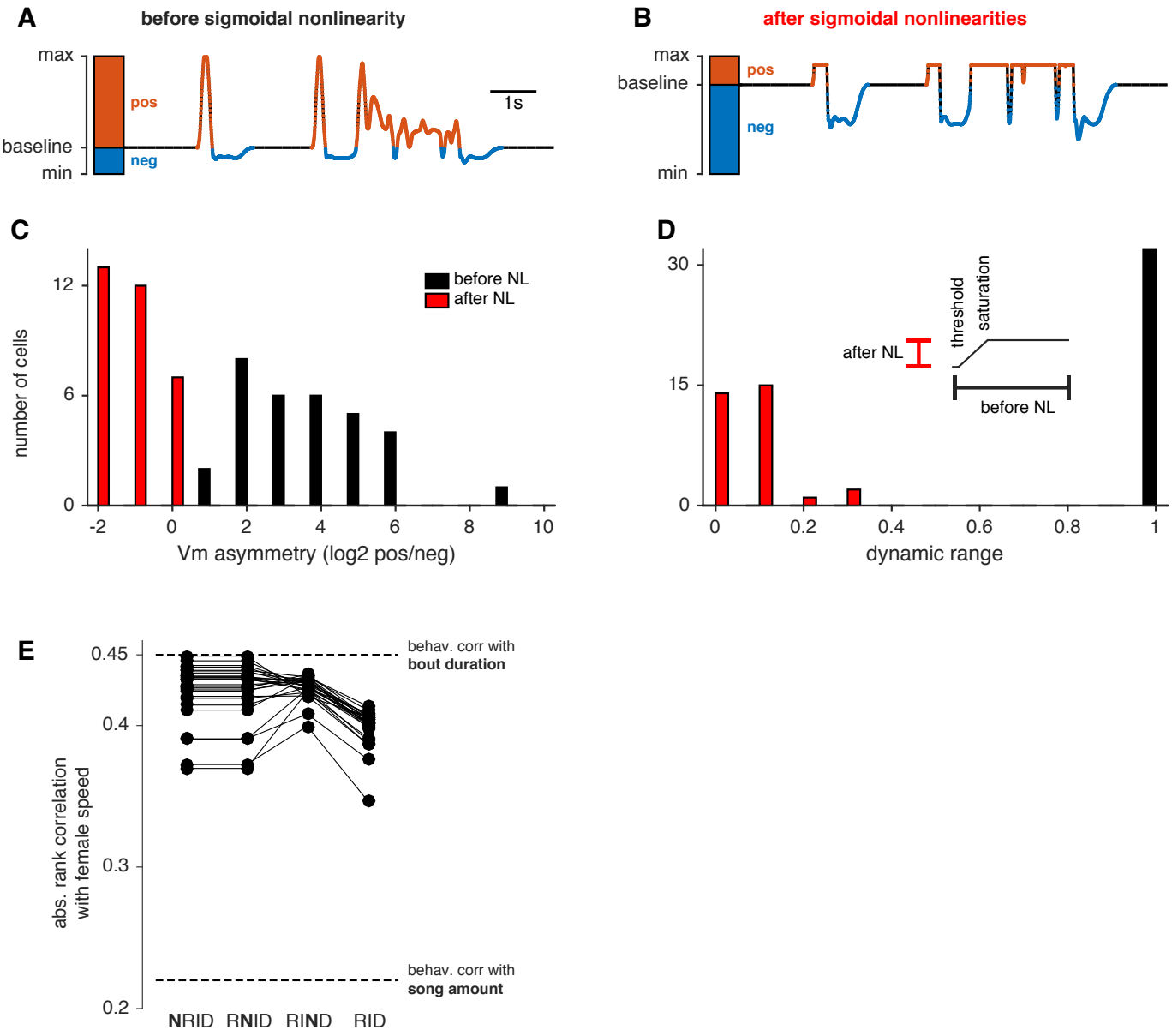

Fig. S6 The optimal sigmoidal nonlinearity in the decoder model reduces the dynamic Vm range and amplifies negative response components through strong saturation (refers to Fig. 4)

A AMMC/VLP model response to naturalistic song (same AV1 cell shown in Fig. 3). Baseline Vm is black, positive and negative response components are indicated in orange and blue, respectively. Orange/blue bar at the beginning indicates the range of values for the positive and negative Vm components. Note a strong asymmetry in the Vm towards positive response components.

B Same AMMC/VLP model response in (A), but following transformation by the sigmoidal nonlinearity optimized for decoder performance. Note a strong compression of the positive response component

(orange) and a relative amplification of the negative response components (blue). Vm asymmetry is now shifted toward negative response components. Bars indicate range over the full minute of song, although only a short segment is shown here. Trace in (A) and (B) are normalized to their respective min/max range.

C Vm asymmetry given as the log ratio between the range of positive and the range of negative response components (orange and blue in (A), (B)) before and after applying the sigmoidal nonlinearity (black and red, respectively). Note the shift of the asymmetry index from positive to negative values, indicating that negative response components tend to be stronger after sigmoidal transformation (median asymmetry before sigmoidal nonlinearity 3.4, after -1.9). N=26 since 6 low-performing cells did not exhibit robust negative response components (and exhibited poor readout performance).

D Vm dynamic range is reduced by the nonlinearity (black – before, median 1.0, red – after nonlinearity, median 0.06). Dynamic range is normalized to 1 for all cells before applying the nonlinearity. Inset illustrates the effect of the nonlinearity on input and output dynamic range mainly due to compression by early saturation. N= all 32 cells.

E Match with behavior for individual neurons put through 4 different decoder models. The ratio-based decoder model presented in Fig. 5E is termed NRID: “N” – nonlinearity, “R” – rectification, “I” – integration, and “D” – division. The threshold and saturation parameters of the sigmoidal nonlinearity (“N”) were optimized for each variant. Labels correspond to the order of operations in the decoder. Upper and lower dashed lines indicate the behavioral correlation with bout duration and song amount, respectively. Placing the nonlinearity before rectification (“NRID”) or after rectification (“RNID”) yields identical performance values for all cells (ρ=0.42(0.02) for both, median(IQR), p=1, sign test between NRID and RNID). These values are also identical to those in Fig 5H. Decoder performance is unaffected by placement of the sigmoidal nonlinearity before or after rectification because the threshold and saturation parts of the nonlinearity affect the positive and negative response components independently. Placing the sigmoidal nonlinearity after the integration step (“RIND”), leads to minor, non-significant changes in decoder performance (ρ = 0.42(0.1), median(IQR), p=1.0, sign test between NRID and

RIND). Removing the nonlinearity altogether ("RID") yields a small but significant decrease in performance ($\rho = 0.40$, $p = 8 \times 10^{-7}$, sign test between NRID and RID). This strongly contrasts with the sum-based decoder, where removing the nonlinearity led to a much stronger decrease in performance (w/o nonlinearity $\rho = 0.19$, Fig. 4C). This indicates that the sigmoidal nonlinearity plays a smaller role in the ratio-based than in sum-based decoder.

Supplemental Movies S1-S5 show images (full z-stack) of biocytin fills of recorded neurons (green – GFP, red – biocytin; Movie S1 – V2, Movie S2 - AV1, Movie S3 - AV2.avi, Movie S4 - AV3, Movie S5 - AV6.avi).

Supplemental References

Arthur, B.J., Sunayama-Morita, T., Coen, P., Murthy, M., and Stern, D.L. (2013). Multi-channel acoustic recording and automated analysis of Drosophila courtship songs. BMC Biol *11*, 11.

Bussell, J.J., Yapici, N., Zhang, S.X., Dickson, B.J., and Vosshall, L.B. (2014). Abdominal-B Neurons Control Drosophila Virgin Female Receptivity. Current Biology *24*, 1584–1595.

Chiang, A.-S., Lin, C.-Y., Chuang, C.-C., Chang, H.-M., Hsieh, C.-H., Yeh, C.-W., Shih, C.-T., Wu, J.-J., Wang, G.-T., Chen, Y.-C., et al. (2010). Three-Dimensional Reconstruction of Brain-wide Wiring Networks in Drosophila at Single-Cell Resolution. Current Biology *21*, 1–11.

Coen, P., Clemens, J., Weinstein, A.J., Pacheco, D.A., Deng, Y., and Murthy, M. (2014). Dynamic sensory cues shape song structure in Drosophila. Nature *507*, 233–237.

Costa, M., Manton, J.D., Ostrovsky, A.D., Prohaska, S., and Jefferis, G.S.X.E. (2015). NBLAST: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. bioRxiv 006346.

David, S.V., and Shamma, S.A. (2013). Integration over Multiple Timescales in Primary Auditory Cortex. Journal of Neuroscience *33*, 19154–19166.

Göpfert, M.C., Albert, J.T., Nadrowski, B., and Kamikouchi, A. (2006). Specification of auditory sensitivity by Drosophila TRP channels. Nature Neuroscience *9*, 999–1000.

Kamikouchi, A., Shimada, T., and Ito, K.E.I. (2006). Comprehensive Classification of the Auditory Sensory Projections in the Brain of the Fruit Fly Drosophila melanogaster. Comparative and General Pharmacology *356*, 317–356.

Longair, M.H., Baker, D.A., and Armstrong, J.D. (2011). Simple Neurite Tracer: Open Source software for reconstruction, visualization and analysis of neuronal processes. Bioinformatics *27*, btr390–btr2454.

Manton, J.D., Ostrovsky, A.D., Goetz, L., Costa, M., Rohlfing, T., and Jefferis, G.S.X.E. (2014). Combining genome-scale Drosophila 3D neuroanatomical data by bridging template brains. bioRxiv 006353.

Murthy, M., and Turner, G. (2013). Whole-Cell In Vivo Patch-Clamp Recordings in the Drosophila Brain. Cold Spring Harb Protoc *2013*, pdb.prot071704–pdb.prot071704.

Nicolaï, L.J.J., Ramaekers, A., Raemaekers, T., Drozdzecki, A., Mauss, A.S., Yan, J., Landgraf, M., Annaert, W., and Hassan, B.A. (2010). Genetically encoded dendritic marker sheds light on neuronal connectivity in Drosophila. Proc Natl Acad Sci U S A *107*, 20553–20558.

Park, M., and Pillow, J.W. (2011). Receptive Field Inference with Localized Priors. PLoS Comput Biol *7*, e1002219.

Phelan, P., Goulding, L.A., Tam, J.L.Y., Allen, M.J., Dawber, R.J., Davies, J.A., and Bacon, J.P. (2008). Molecular Mechanism of Rectification at Identified Electrical Synapses in the Drosophila Giant Fiber System. Current Biology *18*, 1955–1960.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schwartz, O., Pillow, J.W., Rust, N.C., and Simoncelli, E.P. (2006). Spike-triggered neural characterization. Journal of Vision *6*, 484–507.

Tootoonian, S., Coen, P., Kawai, R., and Murthy, M. (2012). Neural Representations of Courtship Song in the Drosophila Brain. Journal of Neuroscience *32*, 787–798.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. Neural Computation *10*, 821–835.

Zhang, Y.Q., Rodesch, C.K., and Broadie, K. (2002). Living synaptic vesicle marker: Synaptotagmin-GFP. Genesis *34*, 142–145.