Supporting Information

Calabrese and Woolley 10.1073/pnas.1408545112

SI Materials and Methods

Stimuli. Stimuli were 10 samples of modulation-limited (ml) noise. Each sample was 2 s in duration. ml noise is correlated Gaussian noise designed to match song in frequency range (250–8,000 Hz) and maximum spectral and temporal modulations. Stimuli were delivered free-field through a flat frequency response speaker positioned 20 cm in front of the bird, at a mean intensity of 65 dB sound pressure level (SPL). Between 30 and 40 response trials were recorded for each of the 10 ml noise stimuli. Stimuli were presented in pseudorandom order. Silent periods between trials (intertrial interval) were between 2 and 3 s in duration, determined randomly by sampling from a uniform distribution of durations between 2 and 3 s.

Electrophysiology. Two days before initial recording sessions, birds were anesthetized with single intramuscular injections of 0.04 mL of Equithesin and placed in a custom stereotaxic holder. Craniotomies were made bilaterally and were centered at stereotaxic coordinates over the auditory cortex of male zebra finches (1.3 mm lateral and 1.3 mm anterior from the bifurcation of the midsagittal sinus). Craniotomy size was determined by the anterior to posterior (A–P) and medial to lateral (M–L) extent of planned recording penetrations (1.5 mm in the M–L axis and 1.5 mm in the A–P axis). Fig. 2*B* and Fig. S2 show the coordinates of individual polytrode penetrations. A metal pin was then affixed to the skull to keep the head still during recording, and a ground wire was cemented in place, with its end just beneath the skull, 5–10 mm lateral to the junction of the midsagittal sinus. Birds recovered from surgery for 2 d.

Recordings were made in the primary auditory cortex (CML, L1, L2, and L3) and secondary region (NC) in head-fixed, unanesthetized, male zebra finches (*Taeniopygia guttata*, $n = 6$) using a planar multichannel silicon polytrode (4 × 4 electrode layout, 177 μm^2 contact surface area, 100 μm intercontact distance in the dorsal-ventral direction and 125 μm intercontact distance in the anterior-posterior direction; NeuroNexus Technologies). In each bird, 50% of penetrations were made in the left hemisphere.

On each recording day, a polytrode coated with DiI was positioned vertically in the brain, with the bird's beak pointing 45 degrees downward. All penetrations were made with the same anterior-to-posterior (A–P) orientation. DiI–labeled electrode tracks (one per recording session) marked the trajectories of the four closely spaced (125 μm apart) shanks of the polytrode. Fig. 2*B* and Fig. S1 show an example of DiI tracks labeling the location of one polytrode penetration. Penetrations on different days were made at M–L coordinates that were at least 300 μm apart to use the DiI tracks to determine polytrode trajectories from different days in the same bird. This procedure ensured clear resolution of polytrode tracks from different days. For anatomical localization of recorded neurons, we used the stereotaxic coordinates from each recording day to assign each polytrode track to a specific set of recordings. Recordings from each region were made, on average, in five out of six birds. For each bird, recordings were made daily over ∼1 wk, and each recording session lasted ∼6 h. Signals were amplified and bandpass filtered between 300 and 5,000 Hz, digitized at 25 kHz (RZ5; Tucker-Davis Technologies), and stored for off-line processing.

Histology. To label recording sites, polytrodes were coated with DiI at the beginning of each recording session. After final recording sessions, birds were given an overdose of Nembutal (0.1 cc) and transcardially perfused with heparinized saline followed by 10% formalin. Brains were cryoprotected in 30% (wt/vol) SDS sucrose formalin (2–4 d) and sectioned at 40 μm thickness using a freezing microtome. Immediately after mounting, sections were imaged with a blue filter to localize fluorescent DiI tracks. After imaging polytrode tracks, sections were imaged with a white-light filter to visualize fiber tracts. The same sections were then Nissl-stained and imaged to define the borders of cortical regions by visualizing laminae and cell bodies.

Boundaries between regions were determined by visualizing laminae and region-specific cytoarchitecture (Nissl stain) and axons (white light). CML is dorsal to the lamina mesopallialis (Fig. 1 and Fig. S1), L2a and L2b are characterized by neurons with small cell bodies that are grouped at high density, and by the termination of dark, thalamic fibers (Fig. S1 and Fig. 1). L is also characterized by a dark, densely packed cells although cells in L2b stain more darkly and are more densely packed than those in L (Fig. 1 and Fig. S1). Region L1 is dorsal to L2a and ventral to CML. Region L3 lies ventral to the border of L2a, is posterior and dorsal to LMD, and is surrounded dorsally and caudally by L and L2b. Cells in L3 are larger and less densely packed than in L2a, L2b, and L and are arranged into large fusiform clusters that give L3 a punctate appearance in Nissl-stained sections. The large region NC is posterior to L2b and L (Fig. 1 and Fig. S1).

Data Analysis. Data analysis was carried out in MATLAB (Mathworks). Spikes were sorted offline with an automated sorting algorithm (WaveClus). First, a nonlinear filter that increases the signal-to-noise ratio (SNR) by emphasizing voltage deflections that are both large in amplitude and high in frequency content was applied to the bandpass filtered voltage trace recorded on each channel. Second, spikes were detected and sorted using WaveClus. Third, the output of this algorithm was refined by hand for each electrode channel, taking into account the waveform shape and interspike interval distribution. For a candidate unit, we computed the SNR as the ratio of the average amplitude of detected action potential waveforms to the SD of the noise. Candidate units that had SNRs greater than 3.5 were considered single units and were included in the analyses (median SNR across all included units = 7.47). This procedure yielded a total of $n = 823$ units recorded during 37 recording sessions from six birds.

Each recorded single unit was classified as either a putative principal cell (pPC) or a putative narrow-spiking interneuron (pIN) using two methods that yielded virtually identical results. In the first method, we computed the average waveform for each unit using all detected waveforms for that unit. We then computed the width of each unit's average waveform and the average peak amplitude assymetry for each unit as shown in Fig. S3*A*. We then used a mixture-of-Gaussians clustering algorithm with action potential width and peak amplitude assymetry as inputs to classify all single units as either pPCs or pINs (Fig. S3 *A* and *B*). In the second method, we used only the average spike waveform width of each unit to classify it as either a pPC or a pIN (Fig. 3*A* and Figs. S3 *A* and *C* and S4). Both methods yielded nearly the same results, with only 4 units out of 823 units classified differently between methods. Based on these results, we used the average action potential (AP) widths to classify units in further analysis. We used AP width because it led to the best separation, providing a clear bimodal distribution. A second feature, spike-peak asymmetry, was also examined and added no extra information, resulting in a nearly identical clustering solution (Fig. S3).

Spectrotemporal receptive fields (STRFs) were computed from ml noise responses as described in ref. 1. Briefly, STRFs from sound-evoked responses were calculated by fitting a generalized linear model (GLM) with two parameters: a 2D linear filter in frequency and time (k), and an offset term (b). The conditional spike rate in the model was given by $I(t) = f(b + k*x)$, were $k*x$ is the convolution between the spectrogram of the stimulus and the linear filter, and f is a linear rectification function. The accuracy of the STRF as a model of the neuron's tuning was measured by using the STRF to predict responses to a test stimulus that was excluded from STRF estimation. We compared the predicted and observed responses to the test stimulus by computing Pearson's correlation coefficient (c.c.) (prediction performance) between the predicted and observed peristimulus time histograms (PSTHs) (5 ms smoothing). Only STRFs that were accurate models of tuning (c.c. ≥ 0.4) were used to compute latencies (superficial $n = 42$; intermediate $n = 33$; deep $n = 20$; secondary $n = 19$). Response latencies were computed by measuring the latency of the best excitatory frequency (BF) from the STRF. STRFs were upsampled (3×), and the BF was measured by setting all negative STRF values to zero and averaging along the time axis. The resulting spectral tuning curve was convolved with a five-point symmetric Hanning window, and the BF was taken to be the position (Hz) of the peak of the smoothed curve. Response latency was measured as the peak point on the time axis (ms). Receptive field linearity was computed as the prediction performance (c.c.) for neurons in each class and region. The STRF describes only the linear component of a neuron's responses to a sensory stimulus. Therefore, the accuracy of the STRF in predicting a neuron's responses to novel stimuli can be used to assess the linearity of a neuron's sensory-coding properties.

The separability index (SI) of the STRF was computed by performing a singular value decomposition. The SI was defined as $SI = s_1^2 / S_i\, s_i^2$, where s_1 is the largest singular value. The SI ranges between 0 and 1, with 1 indicating that the function can be described as a product of two 1D functions. For STRFs, the SI quantifies how well frequency and time can be dissociated in the receptive field. For a neuron with a symmetric or circular excitatory region, frequency and time are relatively independent in the receptive field, and the SI is near 1. For a neuron with more diagonally oriented tuning, or with asymmetric excitatory or inhibitory regions, the SI is closer to 0.

Population sparseness was computed by measuring the fraction of cells in a population that elicit a significant response to a given stimulus. Stimulus selectivity was computed by counting the number of different stimuli that elicit a significant response for a given neuron. To compute significance of the evoked response for a given stimulus, we first used a two-sample Kolmogorov–Smirnov test to compare evoked response to 500 ms of spontaneous activity previous to stimulus onset on a given trial. A significant response to a given stimulus was defined as an evoked spike rate that was significantly above the spontaneous rate ($P < 0.01$) in at least half of the trials.

To compute spike-count correlations, we first used a 50-ms sliding window to compute spike counts from the spike trains of all recorded units. Each neuron was then paired with all other simultaneously recorded neurons, excluding pairs from the same electrode contact. Pearson's correlation coefficients between spike counts were computed for all resulting pairs. Correlations in sound-evoked activity were computed from responses to the entire duration of the stimulus (although nearly identical results were obtained when the first 200 ms of the evoked response were discarded). Correlations in spontaneous activity were computed using a 1-s period before the onset of a stimulus. All correlations were corrected for correlations induced by the stimulus by subtracting a shift predictor, calculated from trials 1 to $n - 1$, with an offset of one trial. For Fig. 4*C*, correlation coefficients of all cell pairs from each region were first z-scored; each value was normalized by subtracting the population median for the region and dividing by the variance. This normalization was done to put correlations in spontaneous and evoked activity on the same scale.

Multiunit activity rasters in Fig. 5*A* were obtained by placing a threshold on the bandpass-filtered voltage trace of each polytrode channel at 4 SDs. Spike count correlations for local pairs were computed by including only neuron pairs recorded from contacts that were ≤150 μm apart. Spike count correlations for distal pairs were computed by including only neuron pairs recorded from contacts that were ≥300 μm apart. To compare correlations across regions, all correlations within a region were normalized so that local correlations in each region had a median of 1.

1. Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. *PLoS ONE* 6(1):e16104.

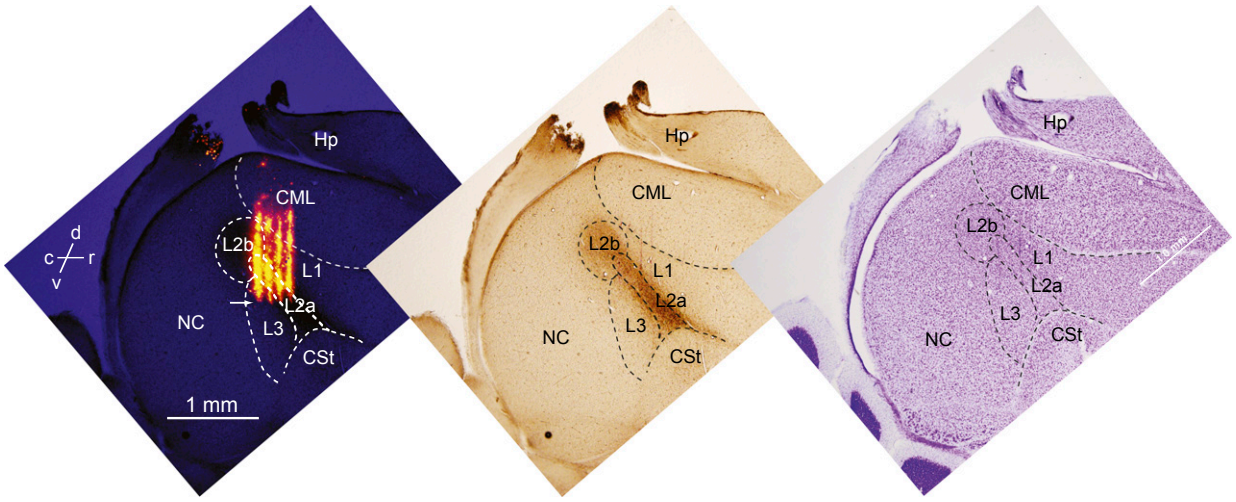

Fig. S1. Histological identification of electrode tracks and recording sites. Neurons were assigned to CML, L1, L2, L3, or NC by comparing fluorescently labeled electrode tracks (*Left*), axon fibers (*Center*), and Nissl-stained cell bodies and lamina (*Right*) in the same section to reconstruct anatomical locations of recordings.

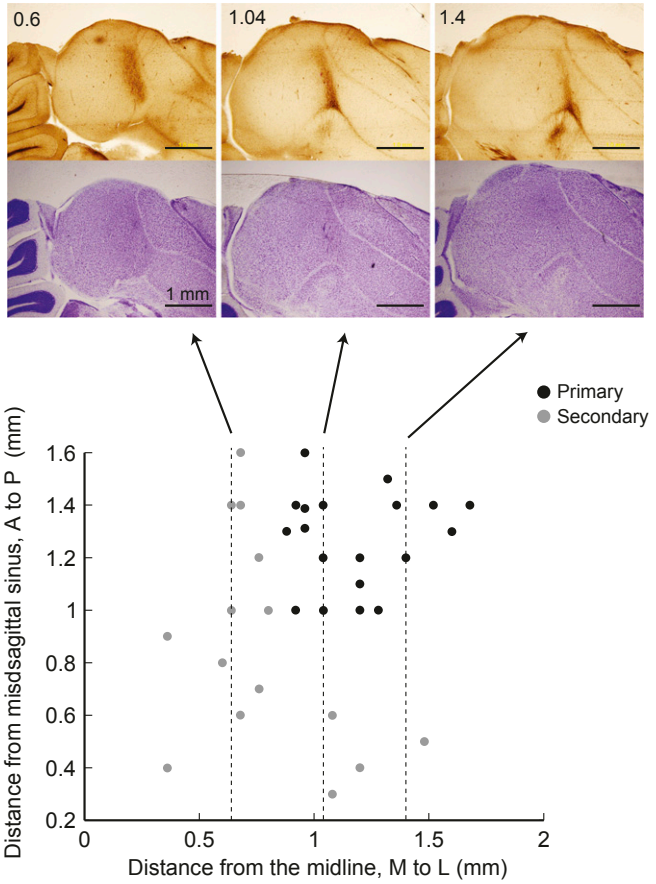

Fig. S2. Recording locations. Each dot corresponds to the medial-lateral (M–L) and anterior-posterior (A–P) coordinates of the center of the polytrode in one penetration. (*Top*) Three parasagittal sections from one brain, at 0.6, 1.04, and 1.4 mm (M–L). The label "Primary" refers to CML, L1, L2, and L3 and "Secondary" refers to NC.

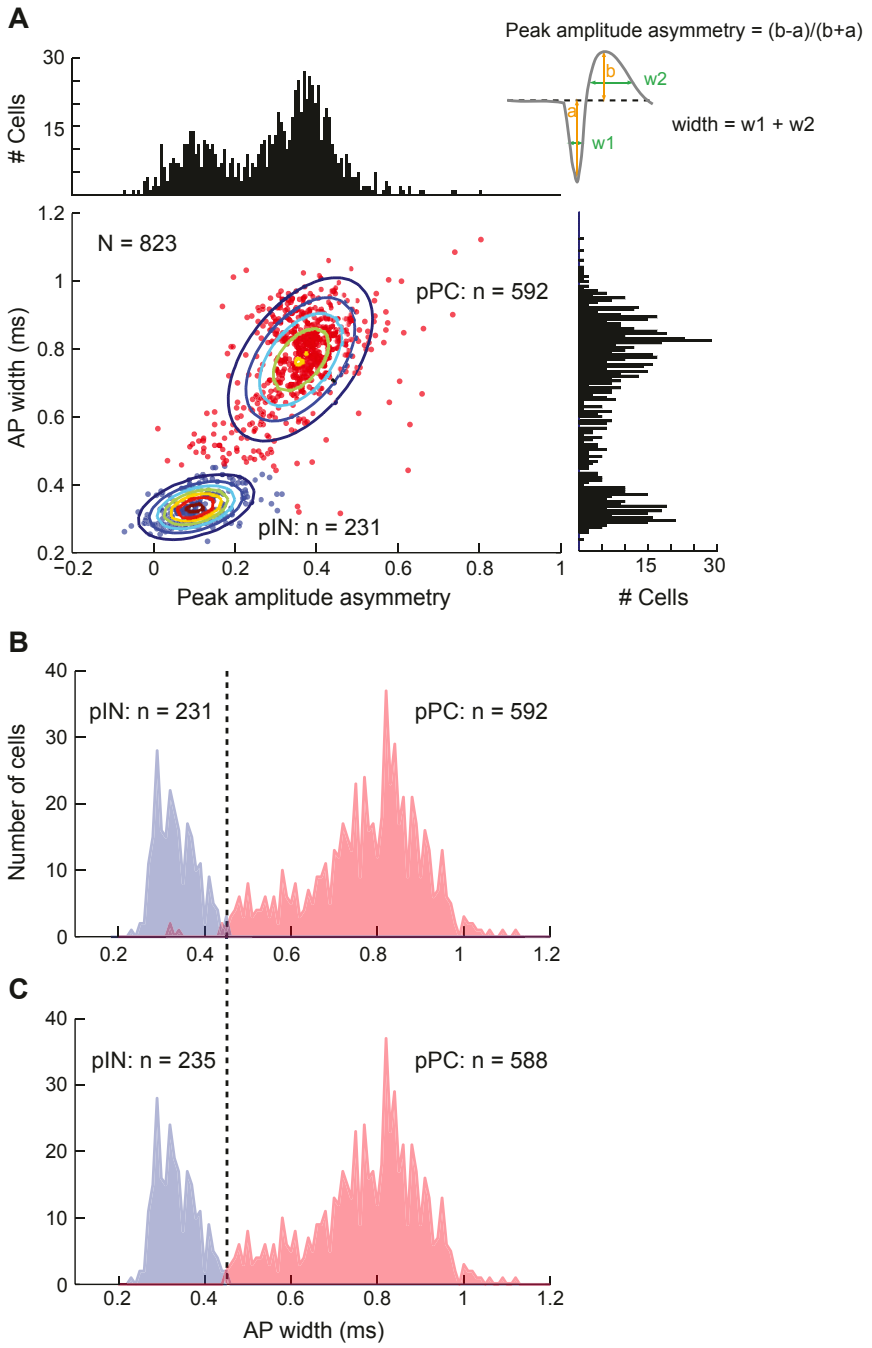

Fig. S3. Classification of putative principal cells (pPCs) and putative fast-spiking interneurons (pINs) using spike waveform shape. Spike-sorted units were classified based on two or one parameter(s) extracted from the spike waveform, spike width, and peak amplitude asymmetry or spike width. The extracted parameters were then grouped using a Gaussian mixture model, yielding two groups. (*A*) A 2D clustering based on spike width and peak amplitude asymmetry. Each dot represents a single unit. Putative INs were characterized by narrower spike waveforms and relatively lower peak amplitude asymmetry. Putative PCs were characterized by wider spike waveforms and relatively higher peak amplitude asymmetry. (*B*) Histogram of spike widths for all recorded units, color-coded according to the clustering solution in *A*. (*C*) One-dimensional clustering of the same data based on spike width only. Comparison between 1D and 2D spike classification solutions (*B* and *C*) shows that both methods lead to almost identical classification, with a difference in only 4 units out of 823. These units were cells with narrow spikes but relatively high spike amplitude asymmetry.

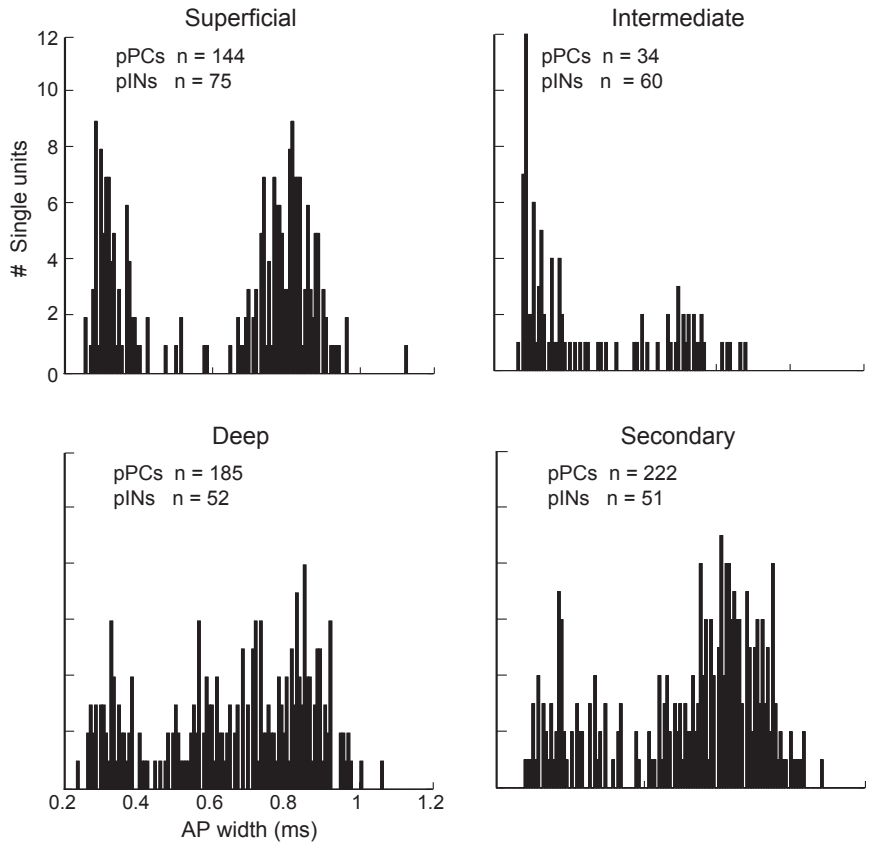

Fig. S4. Spike width distributions by region. Bimodal distributions of action-potential width were observed in all regions.

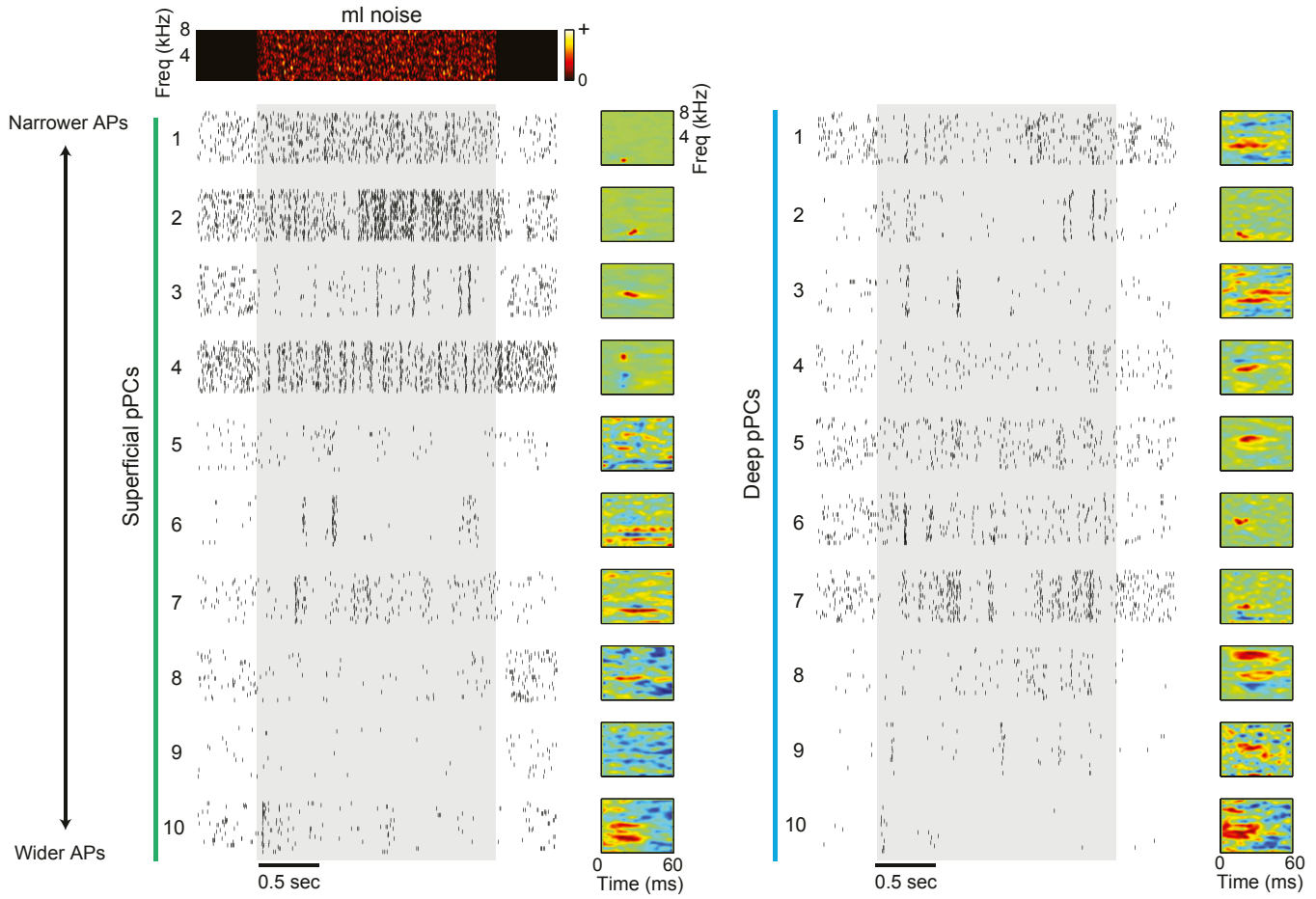

Fig. S5. Spike trains recorded from pPCs in the superficial (*Left*) and deep (*Right*) regions show examples of the relationship between action potential (AP) width and response sparseness. The spectrotemporal receptive field for each neuron is plotted next to the corresponding responses. Cells are organized vertically by AP width. Neurons with wider APs generally produced sparser responses.

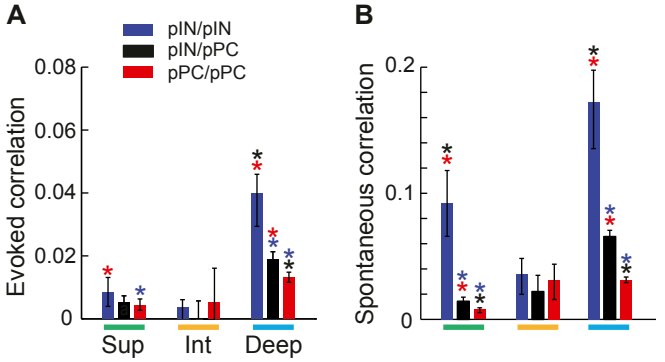

Fig. S6. Pairwise noise correlations plotted to compare correlations across the three pair types in each region. Bars show population medians, and error bars show 95% confidence intervals. (*A*) Evoked activity. (*B*) Spontaneous activity. Color asterisks indicate significant differences between pair types, in a region ($P < 0.05$, Kruskal–Wallis test with multiple groups comparison correction).

Table S1. Cortical regions recorded from in each of six birds

Bird	CML	L1	L2a	L2b	L3	L	NC
1	X	X	X	X	X		
2		X	X	X	X	X	X
3	X		X	X	X	X	X
4	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X
6				X		X	X

Table S2. Correlation magnitudes by brain region, pair type, and activity type

Pair type & region	Spontaneous correlation	Evoked correlation	*P* value
pPC/pPC			
Sup	0.0079 (0.0054, 0.0092)	0.0044 (0.0026, 0.0062)	0.0034
Int	0.0309 (0.0158, 0.0438)	0.0051 (−0.0070, 0.0160)	0.0068
Deep	0.0313 (0.0292, 0.0336)	0.0131 (0.0116, 0.0148)	5.34e−85
pIN/pIN			
Sup	0.0926 (0.0658, 0.1180)	0.0083 (0.0039, 0.0131)	1.453e−26
Int	0.0361 (0.0198, 0.0483)	0.0035 (0.0000, 0.0061)	3.915e−10
Deep	0.1721 (0.1976, 0.1354)	0.0398 (0.0459, 0.0294)	9.765e−20

Correlation values are population medians. The 95% confidence intervals for the medians are in parentheses. Correlations in spontaneous activity are significantly stronger than correlations in evoked activity for pPC/pPC pairs and pIN/pIN pairs, in all regions. In all cases, $P < 0.01$, Wilcoxon signed rank test. Int, intermediate; Sup, superficial.