Supplemental Materials and Methods

Songs

The auditory stimuli used in this study were derived from the songs of three male starlings, drawn from a library of starling vocalizations. For recordings a bird was housed separately in a cage suspended from the ceiling of a 2 m^3 double-walled sound isolation chamber booth (Industrial Acoustics Corporation; IAC). Recordings were made with an AT4071a directional microphone (Audio-Technica, Stow, OH) and amplified with a DMP3 microphone preamplifier (M-Audio, Irwindale, CA). Some recordings were digitized with a DB2000 PCI digital acquisition board (Measurement Computing, Norton, MA) with a sampling rate of 20 kHz and resolution of 16 bits per sample, without an antialiasing filter; others were digitized at 48 kHz and 16 bits per sample with a sound card with an integrated antialiasing filter. Each song was digitally highpass filtered (12 dB/octave) at 100 Hz and scaled to 70 dB RMS. Between 100 and 300 song bouts were recorded from each bird over the course of several days.

Electrophysiology

Prior to recording, birds were implanted with an annular metal chamber, used for head fixation, under equithesin anesthesia (3.75 mL/kg, I.M.). The scalp and first layer of skull were removed over the location of CMM, and the implant was affixed to the skull using dental acrylic. Birds were allowed to recover fully for several days before beginning recording, during which period their daylight cycle was shifted gradually so they would be asleep during the day. Birds were kept in isolation in IAC chambers after surgery and between recording sessions.

Recordings from CMM were made using 16-channel single shank silicon multielectrode arrays with 413 μm^2 (impedance 250 kΩ) recording sites separated by 50 μm (model A1x16-5mm50-413, NeuroNexus Technologies, Ann Arbor, MI) or custom glass-coated etched Pt-Ir electrodes (0.13 mm diameter, impedance 700–1400 kΩ). Note that impedance values for the NeuroNexus probes do not necessarily correspond to single-electrode impedances; our ability to isolate single units was about the same with both kinds of electrodes. Signals were amplified and bandpass filtered between 300–3000 Hz (Model 15 Neurodata, Grass Instruments, West Warwick, RI), digitized at 20 kHz (DB3000, Measurement Computing), and stored to computer disk. Spike events were detected online with a simple window discriminator to give feedback during the experiment, but re-sorted offline. Candidate spike events were those that crossed a threshold of 4.5 times the RMS amplitude of the signal in each episode (i.e. stimulus presentation). The spike waveforms were aligned by their peaks, and the projections onto the first three principal components calculated. Spike clusters were first calculated automatically using KlustaKwik and then manually refined with Klusters (both programs by K. Harris, L. Hazan, Buzsáki lab, Rutgers, Newark NJ). A unit was considered to be well isolated only if there were no spikes with a refractory time of less than 1 ms, and the cluster was significantly separated in the principal component space from all other clusters and the unsorted noise ($P < 0.05$, MANOVA). We never observed spikes on multiple channels, presumably due to the spacing of the recording sites.

Stimuli were presented free-field in an anechoic chamber (IAC-3) from a speaker positioned in front of the bird, at an RMS amplitude of 67–70 dB SPL, measured from the position of the bird's head. Because many CMM cells can be extremely selective, we used a large proportion of the motifs to search for responsive cells. In the first set of experiments (72 units), once sufficient isolation of a single unit was achieved on at least one channel, single motifs (39–54 of the motifs described above) were presented to the animal to test for selectivity. We presented these motifs randomly without replacement with at least 5 repeats of each motif, which required the neuron to be stable for at least 18 min. Responses to each motif were monitored, and if the unit remained isolated and had responded robustly to at least one stimulus, we then examined the response to one or more of the motifs by presenting isolated notes, note deletions, and note reconstructions as described in the previous section. For each motif tested, we collected 6–10 repeats (median 10) of each of the stimuli, which required the neuron to remain stable for an additional 4–15 min (depending on the number of notes). In the second set of experiments, we did not test for selectivity but immediately presented well-isolated neurons with a stimulus set comprising an unmodified 10 s song segment, reconstructions of the song from notes and fragments, and 5 different permutations of the notes and fragments. We collected at least 5 repetitions of each stimulus (at least 10 of the original song, to aid in validation), which required the unit to remain stable for at least 20 min. If the neuron remained stable we presented up to two additional stimulus sets derived from different song segments.

At the end of the final recording session, one or two fiduciary lesions were made, and birds were given an overdose of Nembutal (250 mg/kg) and transcardially perfused with heparinized saline followed by 10% formalin. We cryoprotected the brains in 30% sucrose formalin until saturated (2–4 days). Tissue was sectioned at 50 μm parasaggitally using a cryostat and then stained with cresyl violet. The location of recording sites was determined based on their location relative to the lesions (Supplemental Fig. 2).

Spectrotemporal isolation of notes from motifs

For each motif a threshold was set to isolate regions of high power from the surrounding noise (typically at 45 dB, which was about 10 dB above the noise floor of the spectrogram). Spectrograms were calculated using an adaptive multitaper method (Thomson 1982), with a taper size of 320 samples (16 ms), a frame shift of 10 samples (0.5 ms), and a time-frequency product of 3.5. These parameters were chosen to optimize our ability to visually resolve a wide variety of starling notes in both time and frequency. We identified all points in the spectrogram where the power was above the threshold and grouped them into connected components. These connected components satisfy the basic definition of a note as a region of spectrotemporally contiguous power, but we found that some degree of manual intervention was necessary at this stage. Specifically, we grouped components that were harmonics of each other, and split components that consisted of two notes that were not spectrotemporally disjoint (for instance, when a click overlapped with a tonal note). Whenever the overlap between notes was sufficient to prevent a clean separation, we left the notes grouped together. An example of the result of this segmentation process is shown in Figure 2*A* (main text) and Supplemental Figure 4 (left panels), where the notes have been numbered and labeled in different colors.

Each note identified in this way consisted of a set of K points in the time-frequency grid of the spectrogram, $\Gamma = (\omega_1, \tau_1), \ldots, (\omega_K, \tau_K)$ (ω, frequency coordinate; τ, time coordinate). We applied these masks to the short-time Fourier transform (STFT) of the motif to extract the complex Fourier coefficients associated with the note. The STFT was computed using a Hamming window and the same frequency and time resolution as the multitaper spectrogram. To reduce edge effects, masks were smoothed with a Gaussian roll-off filter, which was defined as follows:

$$\mathbf{W}_\Gamma(\omega,\tau) = \max\left\{\exp\left[\frac{-(\omega-\omega_k)^2}{2\sigma^2_\omega} + \frac{-(\tau-\tau_k)^2}{2\sigma^2\tau}\right] : (\omega_k,\tau_k) \in \Gamma\right\}$$

where σ_τ and σ_ω are the temporal and frequency bandwidth of the Gaussian filter. We used values that corresponded to 2 ms and 512 Hz. The mask has a value of 1 for all points in the note, and decreases outside the note as a function of the distance to the nearest point in the note. Although notes were defined as spectrotemporally disjoint, the roll-off filter sometimes caused the masks of nearby notes to overlap. To avoid oversampling from these regions, we normalized each point in the mask so that the total contribution from all the notes in the motif was no more than unity. The STFT coefficients corresponding to each note are given by

$$\chi_\Gamma(\omega,\tau) = \begin{cases} \mathbf{W}_\Gamma(\omega,\tau)\chi(\omega,\tau), & 0 \le \omega < N/2 \\ \mathbf{W}_\Gamma(N-\omega,\tau)\chi(\omega,\tau), & N/2 \le \omega < N \end{cases}$$

where $\chi(\omega,\tau)$ is the STFT of $x(t)$ and N is the number of frequency points. Note that because the original signal is real-valued, $\chi(\omega,\tau)$ is the complex conjugate of $\chi(N-\omega,\tau)$. This symmetry explains why we identified notes using only the lower half of the spectrogram, and used a mirror image of the mask to isolate the STFT coefficients from the upper half. This guarantees that each column of $\chi(\omega,\tau)$ is a Hermitian series, and thus the inverse of the STFT, which was computed using a weighted overlap-and-add method (Feichtinger and Strohmer 2001), is real-valued.

Note fragments

To divide motifs into note fragments, we used the following procedure. For each note onset, t_n, we found the next note onset, t_{n+1}, that was at least 20 ms later. We then counted N, the number of notes in that interval and divided the interval into $N+1$ parts of equal duration. This process was repeated for t_{n+1} until all the notes were used. The original onsets were not included, so the new intervals spanned the original temporal boundaries between notes. Most of the cut points were at the midpoints of the original notes, but if the bird sang multiple notes in that interval then the notes were divided more than one time. We then reassigned all of the points in the original notes to the new intervals. Examples of the resulting segmentation are shown in Supplemental Figure 4; the effect is to combine some notes spectrally while splitting almost all of them temporally.

Inter-trial coherence

Coherence quantifies the synchronization of two processes as a function of frequency. At each frequency, the inter-trial coherence indicates how reliable the neuronal spike patterns are at that timescale. Neurons with higher spike precision have higher inter-trial coherences at higher frequencies (Fig. 5*B*, main text). Consequently, the spike precision of the neuron is indicated by the highest frequency at which the neuron has significant inter-trial coherence. For each motif presented to a given neuron, we calculated an unbiased estimate of the inter-trial coherence, $\gamma^2_{AR}(\omega_i)$ (Hsu et al. 2004). Multiple tapers were used to calculate 95% jackknife confidence intervals for (Percival and Walden 1993; Bokil et al. 2007). We identified a cutoff frequency, ω_M , which was the highest frequency (in the band starting at 0 Hz) where the coherence remained significantly greater than zero. The spike precision for the neuron was defined as the median ω_M for all the motifs presented to the neuron, excluding motifs that elicited on average less than two spikes per trial (γ^2_{AR} is poorly defined for extremely low spike counts).

Context-dependent suppression and facilitation of note responses

A context dependence index (*CD*) was defined to quantify the difference in the excitatory responses to notes in isolation and in the context of the motif. We used the linear model described in the main text to calculate the responses to notes in the context of the motif by fitting the model to the note deletions. Letting $r_i(t)$ equal the isolated response to a note, and $R_i(t)$ the context-dependent response, we summed the excitation for each, with $r_i^{\mathrm{ex}} = \int \max(r_i(t), 0)\ dt$ and R_i^{ex} likewise for $R_i(t)$. *CD* was defined as $(R^{\mathrm{ex}} - r^{\mathrm{ex}})/\max(R^{\mathrm{ex}}, r^{\mathrm{ex}})$. It is positive for notes that elicit stronger responses in the motif context (facilitation) and negative for notes that elicit stronger responses in isolation (suppression), with the magnitude of *CD* indicating the ratio between the smaller and larger responses. *CD* was considered to be zero for notes where there was no significant difference between and (two-tailed t test, using ordinary least squares estimates of standard error pooled across time points; $\alpha = 0.05$). We also calculated *CD* for each motif-neuron pair by summing the excitatory responses across notes, with $r^{\mathrm{ex}} = \sum_i \int \max(r_i(t), 0)\ dt$ and R^{ex} likewise.

Spectrotemporal RF estimates

For comparison with the FRF model, we calculated receptive fields for each of the neurons in the second experiment using the maximally informative dimensions (MID) method (Sharpee et al. 2004; Atencio et al. 2008). Earlier methods for calculating STRFs use ordinary least squares (Aertsen and Johannesma 1981; Eggermont et al. 1983), which assumes that the stimulus ensemble is Gaussian distributed. For natural stimulus ensembles, some degree of regularization is required (Theunissen et al. 2001). MID overcomes this limitation by maximizing the mutual information between the spiking response of the neuron and the projection of the stimulus onto the spectrotemporal filter. MID estimates the linear spectrotemporal filter (i.e. STRF) as well as a static nonlinearity that translates the projections of the stimulus onto the STRF into firing rates. We calculated the first MID for each neuron using the note and fragment noise, and used it to predict responses to the unmodified song. The algorithm was adapted from code by Tatyana Sharpee and Minjoon Kouh (http://www.cnl-t.salk.edu). The STRF was an average of four jackknife estimates. Each jackknife estimate was used independently to calculate the static nonlinearity of the neuron, and these estimates were combined and fit with a loess smoothing filter. The response to the original song was calculated by convolving the spectrogram with the STRF and using the fitted loess function to determine the firing rate at each point in time.

References

Aertsen AM, Johannesma PI. The spectro-temporal receptive field. a functional characteristic of auditory neurons. *Biol Cybern* 42:133–143, 1981.

Atencio CA, Sharpee TO, Schreiner CE. Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58:956–966, 2008.

Bokil H, Purpura K, Schoffelen JM, Thomson D, Mitra P. Comparing spectra and coherences for groups of unequal size. *J Neurosci Methods* 159:337–345, 2007.

Eggermont JJ, Aertsen AM, Johannesma PI. Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. *Hear Res* 10:167–190, 1983.

Feichtinger H, Strohmer T. Numerical harmonic analysis and image processing In: *Digital Image Analysis*, edited by Kropatsch WG, Bischof H. Springer, 2001.

Hsu A, Borst A, Theunissen FE. Quantifying variability in neural responses and its application for the validation of model predictions. *Network* 15:91–109, 2004.

Percival DB, Walden AT. *Spectral Analysis for Physical Applications,Multitaper and Conventional Univariate Techniques* Cambridge University Press, Cambridge, UK, 1993.

Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Comput* 16:223–250, 2004.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatiotemporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316, 2001.

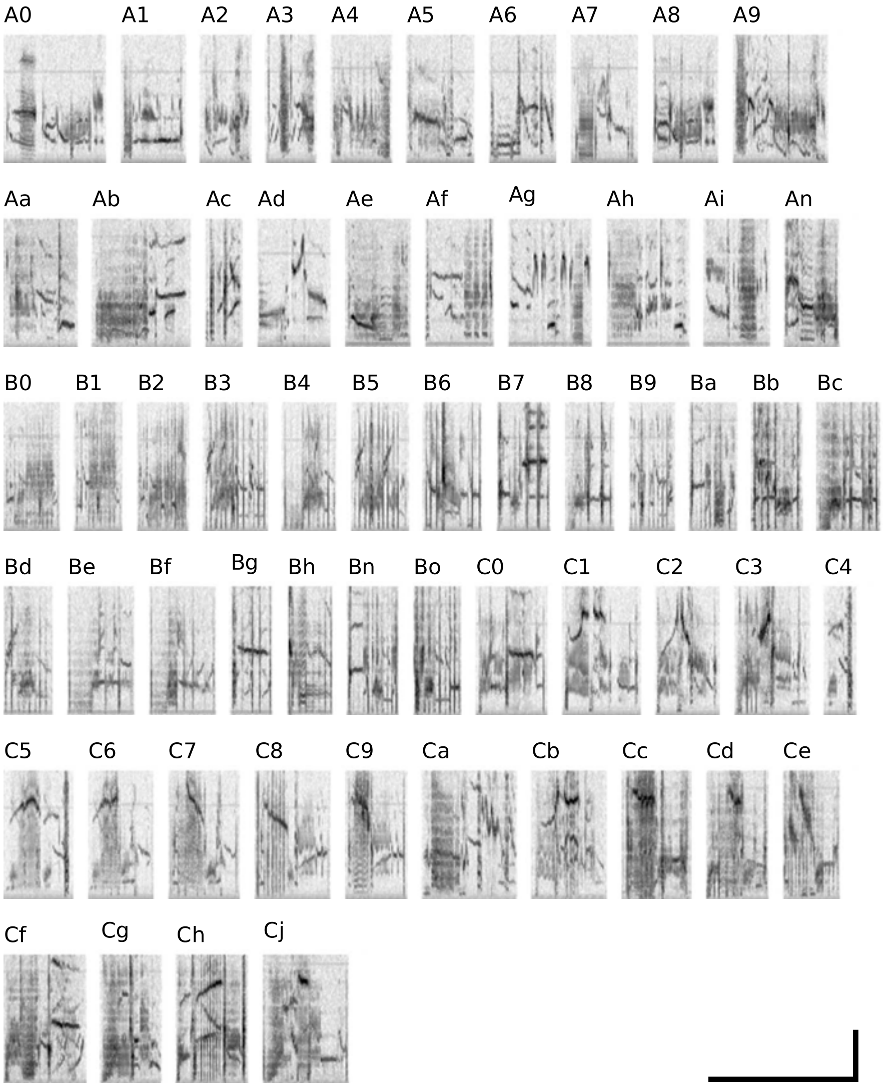
Thomson DJ. Spectrum estimation and harmonic analysis. *Proc IEEE* 70:1055–1096, 1982.

	unanesthetized	anesthetized	P
spontaneous rate (Hz)	1.77	2.83	0.18^{*}
auditory units (% of total)	85 (23/27)	89 (40/45)	$0.65^{\dagger}$
maximum response (Hz)	10.9	12.2	0.19^{*}
selectivity (SI)	0.37	0.16	0.19^{*}
spike precision (Hz)	13.1	7.1	0.004^{*}
intertrial correlation	0.86	0.67	0.0002^{*}
reconstruction quality (CCR)	0.95	0.90	0.014^{*}
motif prediction from notes (CCR)	0.67	0.56	0.17^{*}
context dependence (motif)	-0.16	-0.17	0.44^{*}

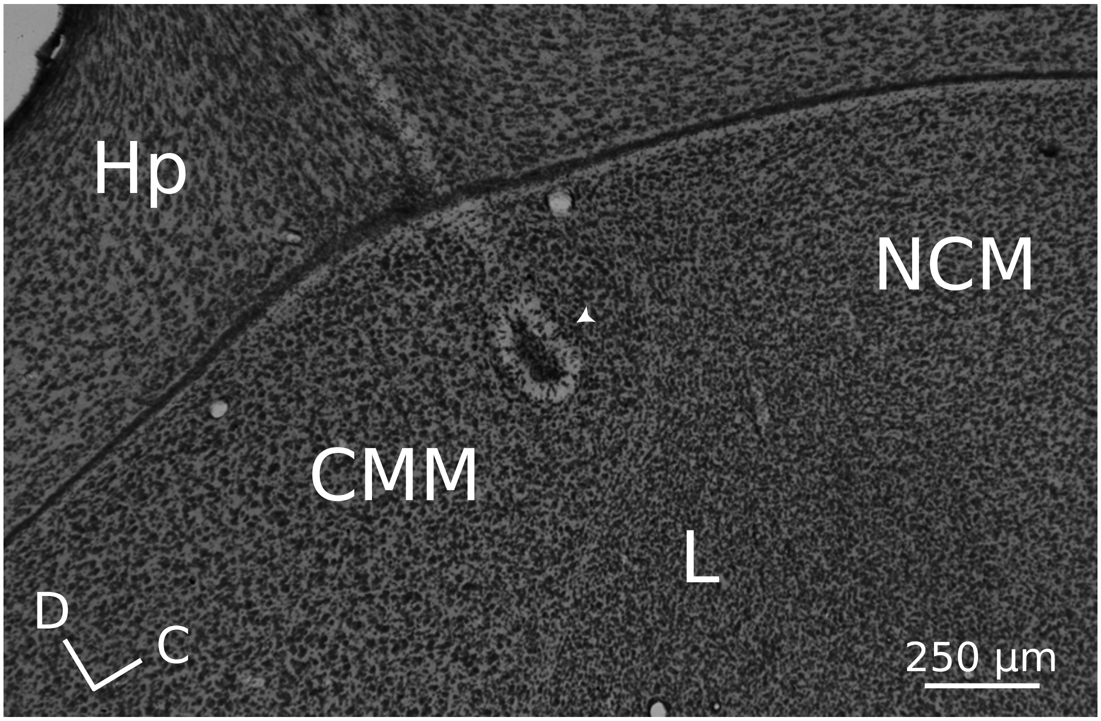
Supplemental Table 1. Comparison of response properties for neurons recorded under restrained, unanesthetized conditions and neurons recorded under urethane anesthesia. Except for proportions, all values are medians. Statistical tests: (∗) two sample Wilcoxon rank-sum test; (†) chi-squared test.

	wide spikes	narrow spikes	P
spontaneous rate (Hz)	1.50	5.95	$1.4 \times 10^{-6}\ ^{*}$
auditory units (% of total)	85 (45/53)	94 (17/18)	$0.29^{\dagger}$
maximum response (Hz)	10.4	19.3	$6.4 \times 10^{-4}\ ^{*}$
selectivity (SI)	0.31	0.095	$3.6 \times 10^{-5}\ ^{*}$
spike precision (Hz)	8.11	13.1	0.03^{*}
intertrial correlation	0.77	0.68	0.65^{*}
reconstruction quality (CCR)	0.91	0.92	0.28^{*}
motif prediction from notes (CCR)	0.59	0.57	0.80^{*}
context dependence (motif)	-0.16	-0.21	0.48^{*}

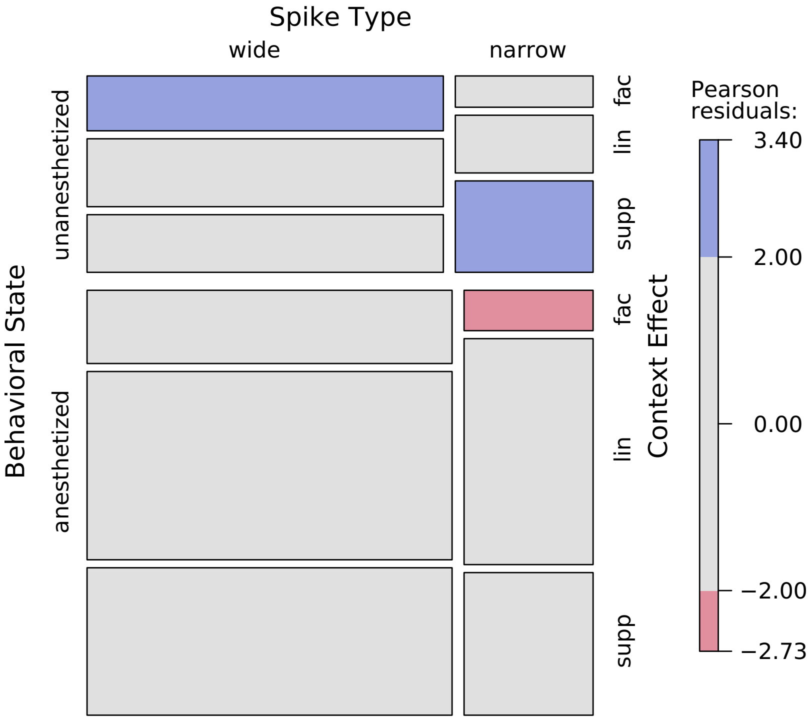
Supplemental Table 2. Comparison of response properties for neurons with wide and narrow spikes. The total unit count (71) is different from in Supplemental Table 1 because one unit had an undefined spike type. Statistical tests: (∗) two sample Wilcoxon rank-sum test; (†) chi-squared test.



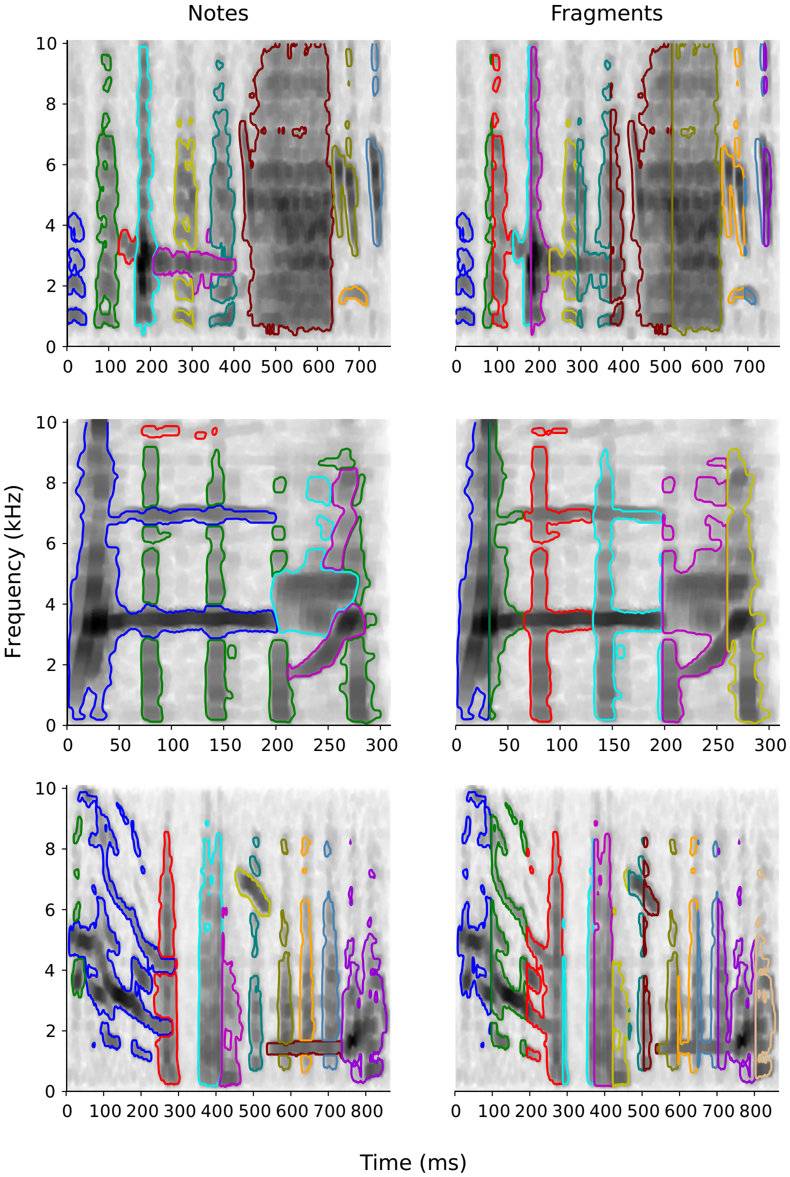
Supplemental Figure 1. Spectrograms of the 54 motifs used to probe selectivity of CMM neurons. Labels above the motifs indicate the motif class (A, variable motif; B, rattle; C, high frequency motif). Vertical and horizontal scalebars are 4 kHz and 2000 ms, respectively.



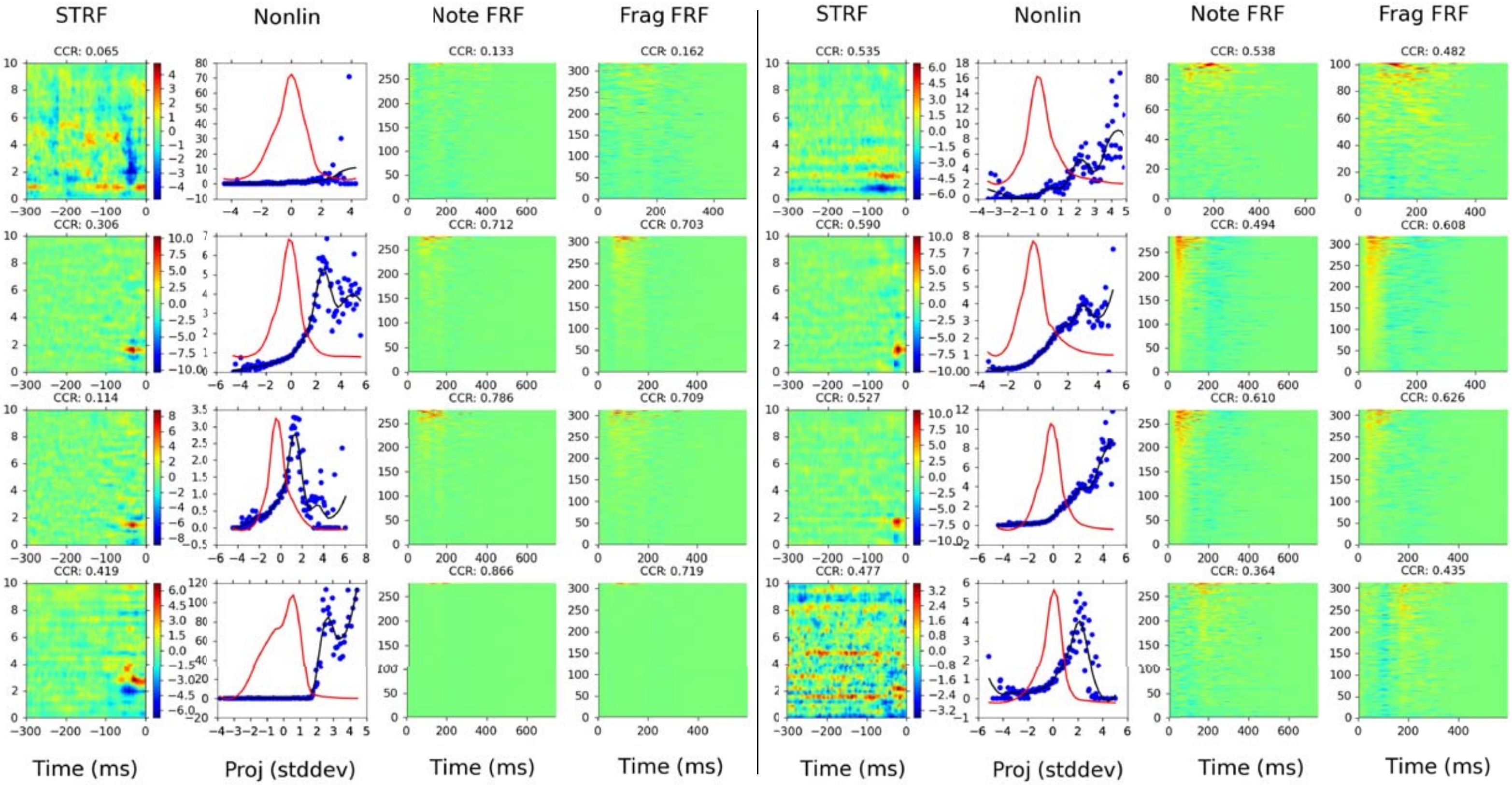
Supplemental Figure 2. Photograph of a Nissl-stained parasaggital section in the starling forebrain, approximately 0.9 mm from the midline. Abbreviations: NCM, caudomedial nidopallium; CMM, caudomedial mesopallium; L, field L; Hp, hippocampus; D, dorsal; C, caudal. The arrowhead indicates a fiduciary electrolytic lesion.



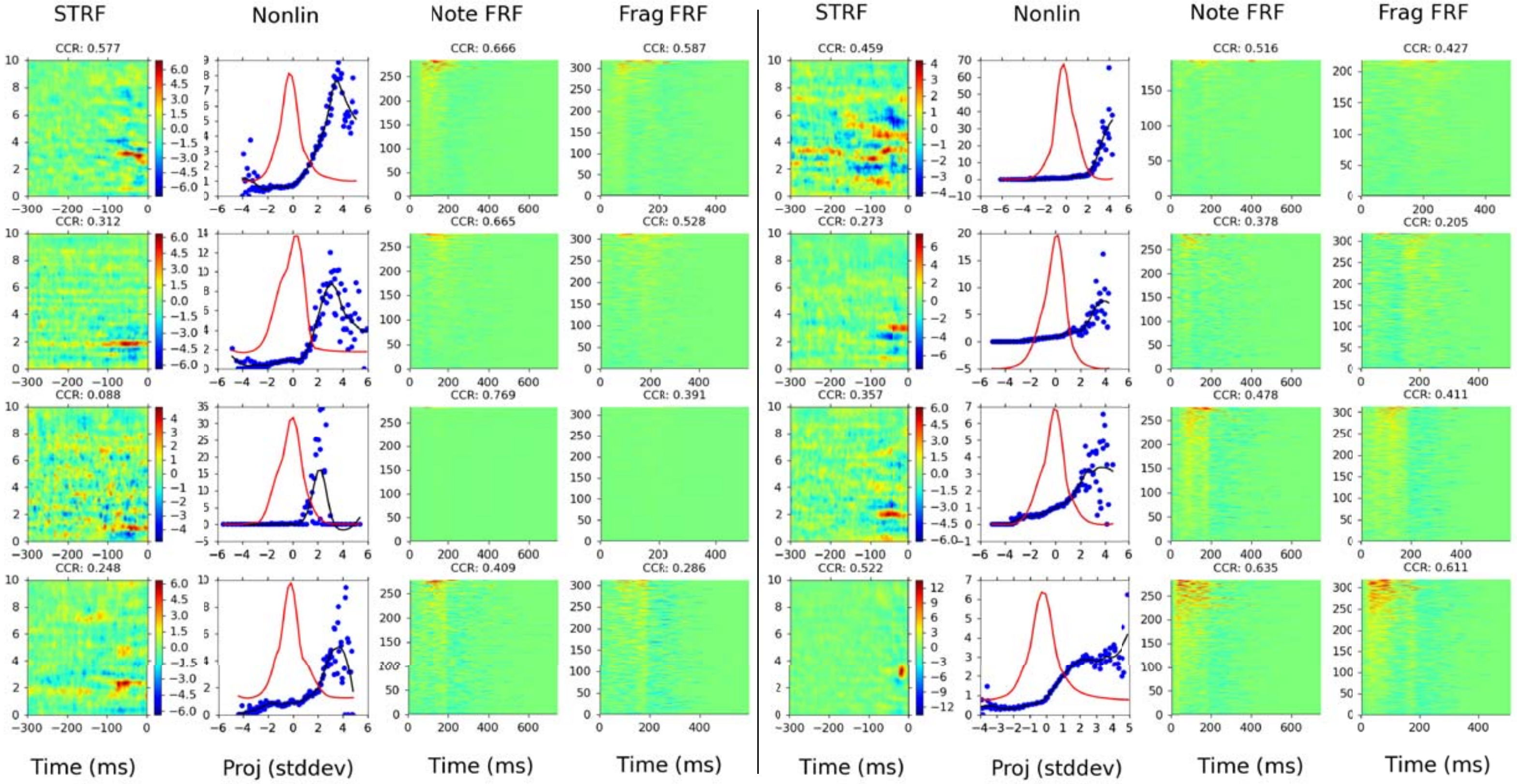
Supplemental Figure 3. Mosaic plot of context-dependent note responses. Notes were categorized as facilitated (fac), suppressed (supp), or linear (lin) depending on whether they elicted significantly larger or smaller excitatory responses in the context of the motif. The area of each cell indicates the proportion of each note type for neurons recorded under unanesthetized or anesthetized conditions, and for neurons with wide and narrow spikes. Colors of the cells indicate significant Pearson residuals; i.e. if the data deviate from the null hypothesis that spike type and behavioral state have no effect. Chi-squared test for independence of the factors: $P = 1.4 \times 10^{-5}$. N is 875 note-neuron pairs.



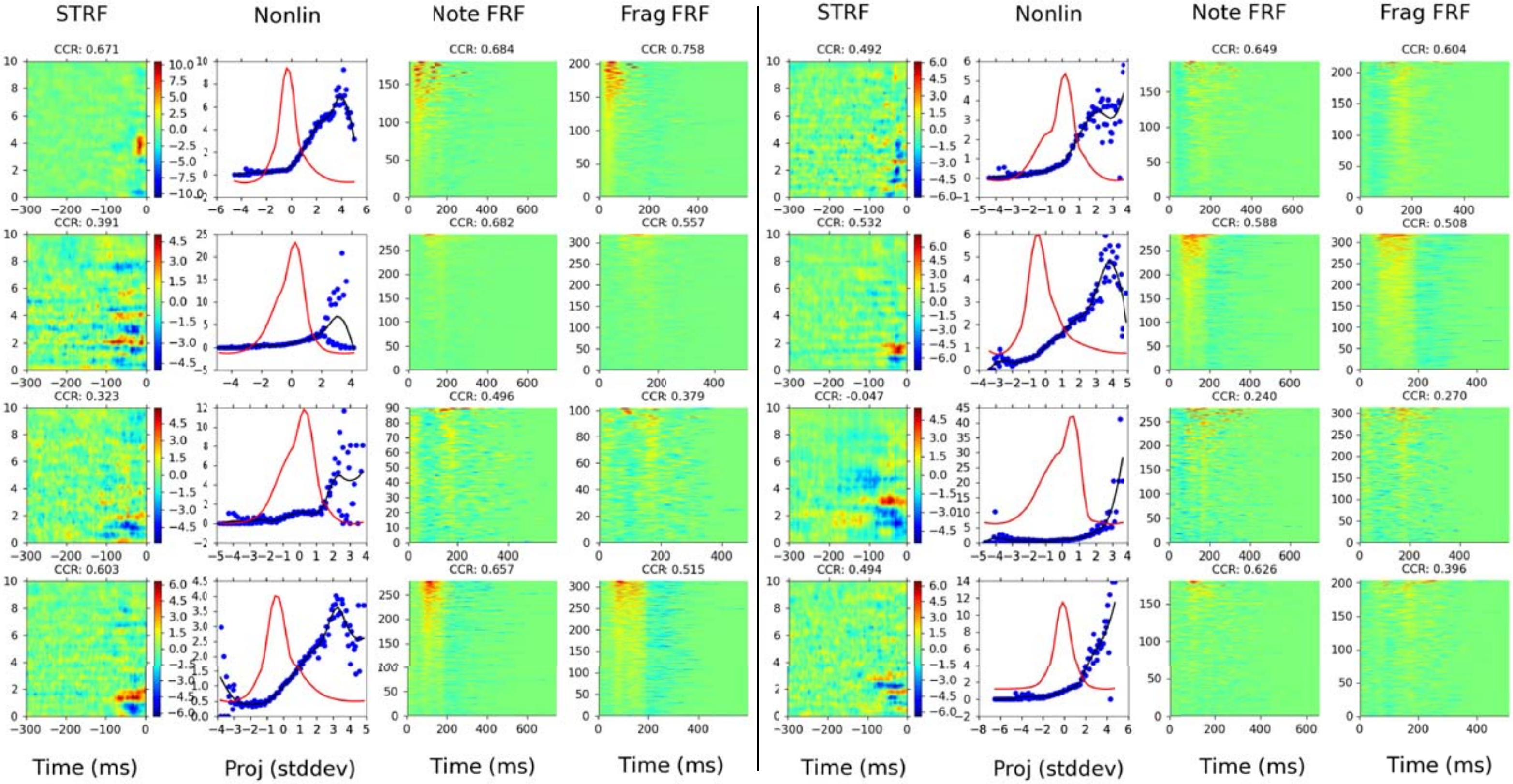
Supplemental Figure 4. Examples of note and note fragment segmentation of motifs. In each row, the spectrogram of a motif is plotted twice. In the left panels, the notes in each motif are indicated by colored outlines. In the right panels, the boundaries of the fragments are indicated. Notes and fragments are colored sequentially, according to onset time, and there is not necessarily any correspondence between the note and fragment that share a color. Fragments were defined using the onset times of the notes and dividing the intervals up into nonoverlapping segments (see Supplemental Materials and Methods for details).



Supplemental Figure 5. Comparison of response models for 8 of the 24 neurons in the second set of experiments. Each neuron is represented by four panels. The first two are the spectrotemporal filter and static nonlinearity for the STRF model. Colors in the STRF plot are scaled by the standard deviation of the STRF coefficients. The static nonlinearity is determined by measuring the distribution of projections of the stimuli onto the filter (red trace, arbitrary units) and calculating the distribution of projections conditional on observing a spike. Blue points are estimates of the static nonlinearity from four jackknife estimates of the STRF (units are expected firing rates for a given projection value), which were then interpolated using a loess smoother (black trace). The third and fourth panels show the note and fragment FRFs, as shown in Figure 12 in the main text. *CCR* of model predictions is given above each panel. Figure continues on next two pages.



Supplemental Figure 5 (continued).



Supplemental Figure 5 (continued).