Cortical Processing of Conspecific Vocalizations in Zebra Finches Depends on the Early Acoustical Environment

Authors: Samantha M Moseley[1] and C Daniel Meliza[1,2,*]

Affiliations: [1] Department of Psychology, [2] Neuroscience Graduate Program, University of Virginia, Charlottesville VA 22904, USA

[*]**Corresponding author:** cdm8j@virginia.edu

Keywords: experience-dependent plasticity, auditory processing, zebra finch, inhibition, coding

Sensory experience during development has lasting effects on perception and neural processing. Exposing animals to artificial stimuli early in life influences the tuning and functional organization of the auditory cortex, but less is known about how the rich acoustical environments experienced by vocal communicators affect the processing of complex vocalizations. Here, we show that in zebra finches (*Taeniopygia guttata*), a colonial-breeding songbird species, exposure to a naturalistic social-acoustical environment during development has a profound impact on cortical-level auditory responses to conspecific song. Compared to birds raised by pairs in acoustic isolation, birds raised in a breeding colony had higher average firing rates, selectivity, and discriminability, especially in the narrow-spiking, putatively inhibitory neurons of a higher-order auditory area, the caudomedial nidopallium (NCM). Neurons in colony-reared birds were also less correlated in their tuning and more efficient at encoding the spectrotemporal structure of conspecific song. These results suggest that the auditory cortex adapts to noisy, complex acoustical environments by strengthening inhibitory circuitry, functionally decoupling excitatory neurons while maintaining overall excitatory-inhibitory balance.

Introduction

Early experience is critical to the development of auditory perception. Human infants learn phonetic categories in the first year of life by hearing speech[1], resulting in changes to perception[2] paralleled

by neural responses to speech sounds[3,4]. In rodents, auditory experience during a developmental critical period has long-lasting effects on the functional organization of the auditory cortex[5–11]. In most songbird species, early experience not only provides a model for the bird to copy[12,13] but shapes how the auditory system processes conspecific vocalizations[14–17]. Because auditory perception is the foundation for vocal production and for higher-order aspects of communication, it is important to understand how natural experience shapes its development.

In many cases, the effects of the early environment on auditory processing are consistent with Hebbian plasticity, which predicts that sensory neurons will become tuned to frequently experienced acoustical features or combinations of features. For example, in the auditory cortex of rats raised hearing pulses of white noise at a rate of 4 Hz, more neurons are tuned to temporal modulation rates around 4 Hz compared to controls[8], and in zebra finches (*Taeniopygia guttata*) fostered by other finch species, cortical-level neurons are preferentially tuned to the spectrotemporal features of the foster species' song[18]. In other cases, tuning shifts away from prevalent features of the acoustical environment, such as when rats are reared in spectrotemporally modulated noise[11]. This anti-Hebbian plasticity may support the ability of animals to hear and discriminate vocalizations in natural environments, which are characterized by broadband, acoustically complex sounds from other animals and non-biological sources that can mask signals of interest[19,20], but the underlying mechanisms remain poorly understood.

For vocal communicators that raise their young in groups, the sounds produced by conspecific individuals present opportunities and challenges. An enriched early sensory environment is broadly beneficial to neuronal development[17], and hearing many different individuals may be essential for learning the statistical invariants of song or speech[21]. However, when multiple individuals vocalize simultaneously, it creates an acoustical background with identical statistics to the signals of interest[22], which is a particularly challenging source of interference that the developing brain has to learn to filter out.

In this study, we examined how a complex acoustical environment affects the development of auditory processing in the zebra finch, a social songbird that raises its young in dense colonies of tens to hundreds of individuals[23]. Zebra finches communicate vocally using calls and songs with rich and varied spectrotemporal structure[24], and cortical-level auditory areas play a critical role in decoding individual identity and other messages from these signals[25–27]. Raising zebra finches in

continuous white noise results in abnormal cortical-level auditory responses[15], and exposing adults to the acoustical background of a canary colony transiently affects lateralization and frequency tuning[28]. Exposure to conspecific colony noise has a profound effect on intrinsic dynamics in the auditory system during the critical period for song memorization period[29,30], but the long-term effects of the early social-acoustical environment of the colony have not been characterized. Here, we compared the cortical-level auditory responses of birds reared by pairs in an indoor breeding colony (colony-reared, CR) to birds raised by pairs in acoustic isolation (pair-reared, PR). Compared to the PR birds, neurons in CR birds had higher firing rates, selectivity, and discriminability, especially in the narrow-spiking, putatively inhibitory neurons of the caudomedial nidopallium (NCM). There were also effects on the relationship between signal and noise correlations that suggest exposure to a complex acoustical environment promotes development of inhibitory circuitry that more effectively decorrelates auditory activity, leading to increases in coding efficiency.

Results

The zebra finch colony is a complex acoustical environment that masks individual songs

This study tested how the early acoustical environment impacts auditory processing by comparing two rearing conditions. In the colony-reared (CR) condition, birds were housed with their families in individual cages situated in our breeding colony, where they were able to hear the songs and calls of many other individuals. In the pair-reared (PR) condition, birds were housed with their families in acoustic isolation chambers, where they could hear only the songs and calls of their parents and siblings. To characterize the differences between these environments, we collected acoustic recordings from one nestbox in each condition when juveniles were between 26–40 dph, a critical period for auditory development. The CR environment was much louder than the PR environment (Fig. 1A,D). On average, the median acoustic amplitude during the birds' daytime hours was 50 ± 2 dB_A SPL (mean ± SD, $n = 723$ intervals of 10 min over 6 d) in the CR recording and 28 ± 5 dB_A SPL in the PR recording ($n = 722$). The average maximum amplitude was 88 ± 4 dB_A SPL in the CR recording and 77 ± 7 dB_A SPL in the PR recording.

The CR environment was also more complex than the PR environment. There was a nearly

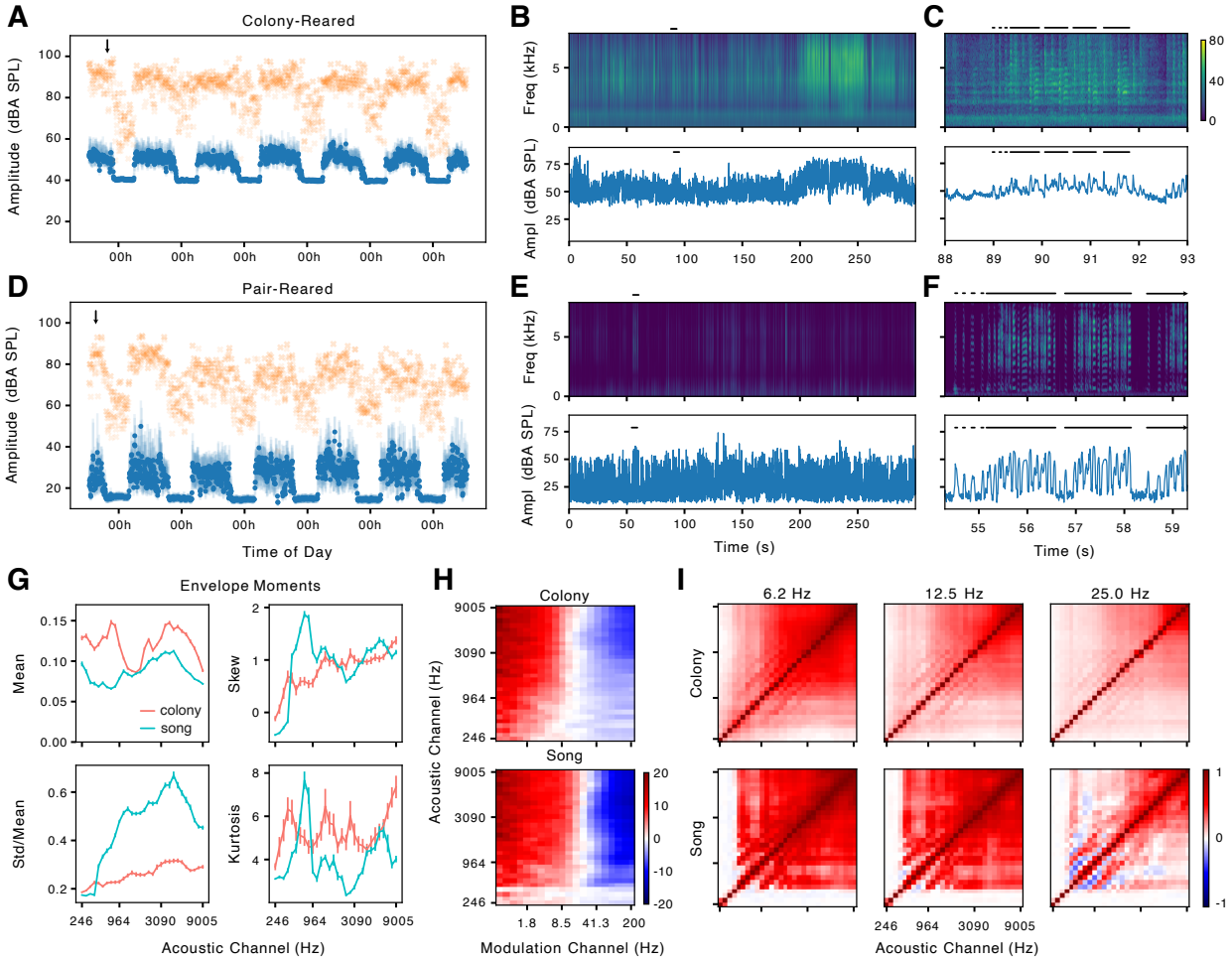

Fig. 1. Statistics of acoustical environment for colony-reared (top row) and pair-reared (bottom row) chicks. (**A**) Sound pressure level (A-weighted) in 10-min intervals across several days in a recording from a nestbox in the colony. The clutch comprised four chicks that were 26–29 dph at the start of the recording, and the colony contained 70–72 animals (median 71). Blue dots, median; blue bars, interquartile range; yellow crosses, maximum. (**B**) Spectrogram (top) and amplitude (bottom) of colony noise during the interval marked by an arrow in A. The black bar marks a song by the father of this clutch. The loud sound between 200–250 s is a series of begging calls. (**C**) Detail of the interval marked by a bar in B. The bars mark introductory notes and syllables of the father's song. (**D–F**) Same format as A–C but for a family housed in an acoustic isolation chamber. The clutch comprised six chicks that were 27–34 dph at the start of the recording. The color scale (dB SPL) is the same for all the spectrograms in the figure. The arrow in F indicates that the song continues beyond the displayed time period. (**G**) Marginal moments of the amplitude envelopes for log-spaced acoustic channels corresponding to locations on the basilar membrane. Lines show mean ± SE for samples from the colony recording in A–C (salmon; $n = 66$) and for samples from individual zebra finch songs (cyan; $n = 62$). (**H**) Modulation power in each acoustic channel, normalized by the variance of the envelope, expressed in dB relative to the same statistic for pink noise. (**I**) Correlations between the envelopes of the acoustic channels at different modulation rates.

continuous background of sounds from birds and artificial sources in the colony (Fig. 1B,C). The frequencies below 1 kHz were dominated by wide-band, incoherent noise that sounded like it was produced by the room's ventilation system. The higher frequencies also included some incoherent noise but were dominated by overlapping vocalizations. These were easily identified as zebra finch songs and calls by their characteristic harmonic structure, but it was rarely possible to distinguish specific types of vocalizations or distinct individuals. Even the song of the father, who was never more than 60 cm from the microphone, was almost impossible to identify except by having an experienced observer listen to the recording, because the spectrographic features of the song were heavily obscured by noise, and the amplitude modulations were only 10–15 dB above the noise floor (Fig. 1C). In contrast, the PR environment had long periods of silence during which the background amplitude was as low as 20 dB_A SPL. It was trivial to identify songs and calls, and the spectrographic features of songs were clear and distinct (Fig. 1E,F).

Thus, although a bird raised in the colony is exposed to a larger number and variety of conspecific vocalizations, at the same time this complex background is likely to mask the features of individual songs, including the song of a male juvenile's tutor. Masking can occur not only at the level of the periphery (i.e., energetic masking) but at higher levels of the auditory system as well, when the background has similar higher-order spectrotemporal structure (i.e., informational masking[31]). To characterize these higher-order features, we computed a series of statistics developed by McDermott and Simoncelli[32] to characterize and synthesize sound textures. The statistics are based on a model of the auditory periphery that decomposes sound into logarithmically spaced frequency bands corresponding to locations along the cochlea (or basilar papilla, in the case of birds), followed by filtering of the amplitude envelope for each band, which reflects the tuning of more central auditory neurons to different modulation rates[33]. As seen in Figure 1G–I, the statistics of these higher-order features in our colony were broadly similar to the statistics for clean recordings of individual zebra finch song. In the colony, the envelopes for the acoustic frequency bands around 850 and 3500 Hz had the largest means, corresponding to the low-frequency noise and the higher-frequency vocalizations (Fig. 1B,C). The standard deviation of the envelopes was relatively low, but greater for higher frequencies, indicating that the amplitude of the vocalizations varied more over time than the incoherent noise. The skew and kurtosis were high across almost all of the frequency bands, suggesting that the recording had rare periods of high amplitude, perhaps

corresponding to the vocalizations of individual animals. By comparison, the clean song recordings had less power at low frequencies, consistent with the absence of ventilation noise. The variance in frequencies spanning from about 500 to 7000 Hz was larger than in the colony recording, consistent with the lower noise floor in the sound isolation chamber. Skew and kurtosis were comparable to the colony recording, but there were large peaks around 850 Hz. Both colony noise and song had slow temporal modulations across a broad range of acoustic frequencies (Fig. 1H) and high correlations between acoustic frequency bands over a wide range of modulation rates (Fig. 1I), consistent with the broadband character of zebra finch song. There was a banding pattern in the cross-channel correlations for both colony noise and song that reflects harmonic structure. Overall, although this is not an exhaustive analysis, it illustrates that the CR birds in this study were exposed to a much louder and more complex acoustical environment than PR birds, and that the statistics of the background are likely to present a challenging "cocktail-party" problem for the developing zebra finch auditory pallium to solve.

Birds raised in a complex acoustical environment have a more active auditory pallium

To examine the effects of the acoustical environment on auditory processing, we recorded single-unit responses to unfamiliar songs from the auditory pallium in CR and PR birds. The avian auditory pallium consists of several quasi-laminar regions organized into columns with connectivity resembling the canonical neocortical circuit (Fig. 2A). Recordings were grouped into three broad subdivisions: the thalamorecipient areas L2a and L2b, the superficial areas L1 and CM (including medial and lateral), and the deep/secondary areas L3 and NCM. As has been previously reported[35,36], units in each of these regions had extracellular waveforms that were either narrow with larger peaks or broad with shallower peaks (Fig. 2B). Narrow-spiking (NS) cells are thought to be fast-spiking inhibitory interneurons, whereas broad-spiking (BS) cells are thought to be excitatory[36,37]. Table 1 shows the sample sizes from this experiment, including the number of cells of each type in each area. Auditory responses to song were diverse and complex (Fig. 2C), with different neurons responding to different songs with tightly time-locked peaks of activity or with slower modulations of firing rate. Across all three areas, BS neurons tended to have lower firing rates and to respond to fewer songs, whereas NS neurons tended to have higher firing rates and

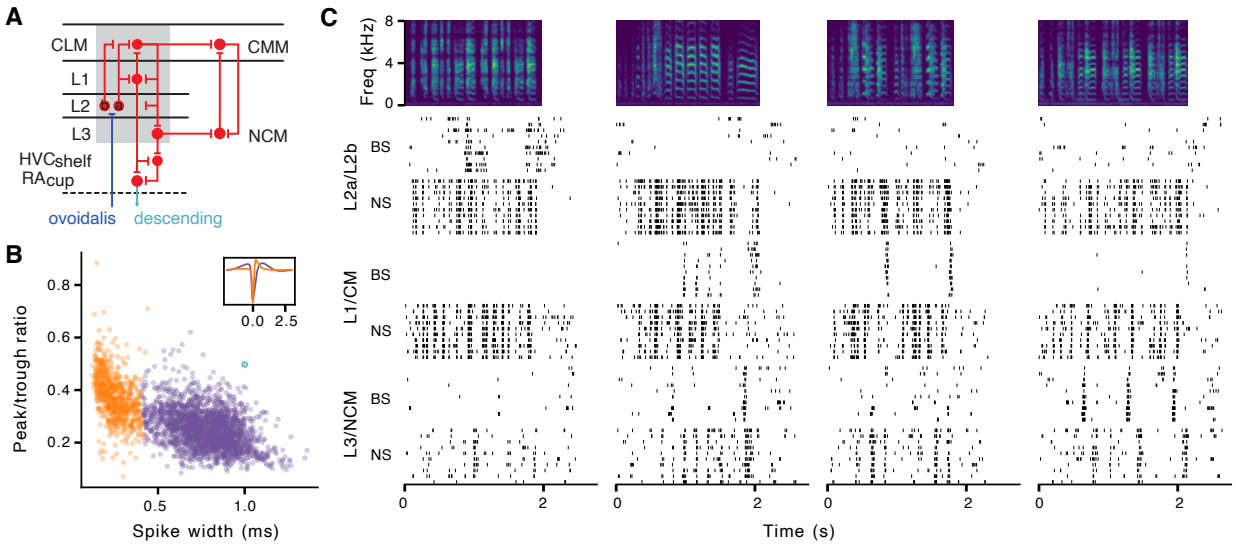

Fig. 2. Example responses from zebra finch auditory pallium. (**A**) Architecture and connectivity of the auditory pallial areas examined in this study (adapted from Wang et al.[34]). Recordings were grouped into intermediate, thalamorecipient areas (L2a, L2b), superficial areas (L1, CLM, CMM), and deep/secondary output areas (L3, NCM). (**B**) Peak-to-trough ratio and spike width for all units in the study. Units were classified as broad-spiking (BS, purple) or narrow-spiking (NS, orange). Inset shows average waveforms for the two classes. (**C**) Raster plots of example BS and NS units from each brain area responding to zebra finch songs (spectrograms at top).

broader selectivity (Fig. 3A,B).

Rearing condition had a large effect on firing rates. Overall, CR neurons tended to have higher spontaneous and evoked activity than PR neurons, but the magnitude of the difference varied by area and cell type. The effect of rearing condition on spontaneous rate (Fig. 3A) was larger than it was on evoked rate (Fig. 3B), and the differences in spontaneous rate and evoked rate were largest for NS neurons in L3/NCM. We also found that we had recorded from fewer units in PR birds despite having a similar number of animals and recording sites (Fig. 3C; Table 1). This suggests that there were a large number of neurons in PR birds with firing rates below our threshold for inclusion (at least 100 spikes during the 58-minute recording session).

Assuming that NS neurons are inhibitory, these results suggest that a complex acoustical environment increases both the resting inhibitory tone and the recruitment of inhibitory activity during auditory stimulation, or conversely, that an impoverished acoustical environment results in much weaker inhibitory responses. It is interesting to note that the 3- to 10-fold greater firing rates of NS neurons in CR birds were not accompanied by a decrease in BS firing rates, as might be expected if the network connectivity were the same in both groups. This suggests that the

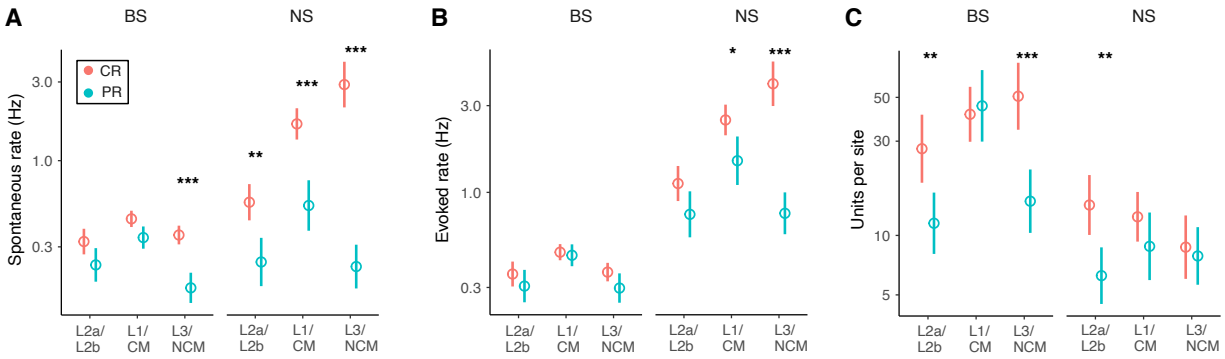

Fig. 3. Effects of the acoustical environment on firing rates. (**A**) Average spontaneous firing rates of BS and NS neurons in each brain area for CR (salmon) and PR (cyan) adults. Circles indicate means; whiskers indicate 90% credible intervals for the mean (GLMM; see Materials and Methods). Asterisks indicate whether there is a significant post-hoc difference between means for PR and CR units in each condition (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). BS cells in CR birds had higher spontaneous rates compared to PR birds in L1/CM (log ratio [LR] $= 0.26 \pm 0.11$; $z = 2.2$; $p = 0.025$) and L3/NCM (LR $= 0.74 \pm 0.15$; $z = 4.9$; $p < 0.001$). NS cells in CR birds had higher spontaneous rates in L2a/L2b (LR $= 0.83 \pm 0.26$; $z = 3.2$; $p = 0.0011$), L1/CM (LR $= 1.14 \pm 0.25$; $z = 4.5$; $p < 0.001$), and L3/NCM (LR = 2.5 ± 0.27; $z = 9.5$; $p < 0.001$). (**B**) Average evoked firing rates of BS and NS neurons in each brain area. NS neurons in CR birds had higher evoked rates compared to PR birds in L1/CM (LR $= 0.52 \pm 0.22$; $z = 2.3$; $p = 0.02$) and L3/NCM (LR $= 1.64 \pm 0.23$; $z = 7.0$; $p < 0.001$). There was no significant difference between PR and CR birds in the evoked rate of BS neurons (all $z < 1.53$; $p > 0.13$). (**C**) Average number of of BS and NS neurons recorded per site in each brain area. CR birds had more BS units per site in L2a/b (LR $= 0.87 \pm 0.32$; $z = 2.6$; $p = 0.007$) and L3/NCM (LR $= 1.22 \pm 0.33$; $z = 3.7$; $p = 0.002$) and more NS units per site in L2a/b (LR $= 0.82 \pm 0.29$; $z = 2.8$; $p = 0.005$).

differences in firing rates were indicative of broader effects on the structure of inhibitory and excitatory networks.

Noise and signal correlations depend on the early acoustical environment

One method of probing the functional connectivity within sensory systems is to measure the correlation of activity in pairs of simultaneously recorded neurons. We hypothesized that increased inhibition in CR birds would decorrelate auditory responses, leading to sparser and more invariant representations of song. We calculated signal correlations, which measure how similarly tuned two neurons are, and noise correlations, which measure how much activity covaries when the stimulus is the same[38]. Consistent with our hypothesis, we found that signal correlations were lower in CR birds compared to PR birds, although the magnitude and significance of the effect varied by brain region and cell type (Fig. 4A). BS neurons were less correlated in CR birds at earlier stations of the auditory hierarchy (L2a/L2b and L1/CM), whereas NS neurons were less correlated in CR birds at

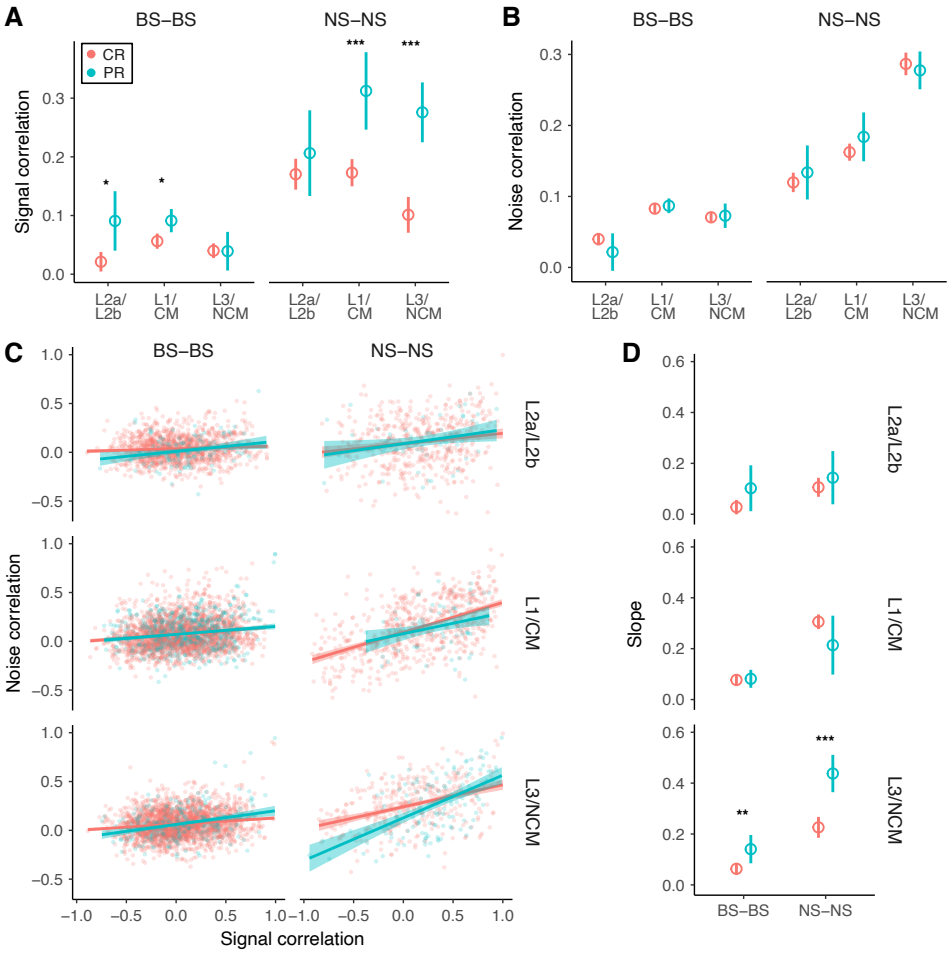

Fig. 4. Pairwise correlations of auditory responses. (**A**) Signal correlations between pairs of BS neurons and pairs of NS neurons in each brain area for CR and PR birds. On average, BS neurons were less correlated with each other than NS neurons ($\beta = -0.15 \pm 0.01$; $t(8237) = -10.7$; $p < 0.001$). Pairs of BS neurons in CR birds had lower signal correlations compared to PR birds in L2a/L2b ($\beta = -0.07 \pm 0.03$; $t(8237) = -2.1$; $p = 0.03$) and L1/CM ($beta = -0.03 \pm 0.01$; $t(8237) = -2.4$, $p = 0.02$). Pairs of NS neurons in CR birds had lower signal correlations compared to PR birds in L1/CM ($\beta = -0.14 \pm 0.04$; $t(8237) = -3.3$; $p = 0.001$) and L3/NCM ($\beta = 0.17 \pm 0.03$; $t(8237) = -4.8$; $p < 0.001$). (**B**) Noise correlations between pairs of BS neurons and pairs of NS neurons in each brain area for CR and PR birds. There was a significant interaction between area and cell type ($F(2, 8237) = 34.8$; $p < 0.001$), but none of the interactions involving rearing condition were significant ($F < 0.8$; $p > 0.37$). (**C**) Scatter plots of signal and noise correlations between pairs of BS neurons and pairs of NS neurons in each brain area for CR and PR birds. Colored lines are linear regressions with standard errors. (**D**) Regression slopes for C with 90% credible intervals. The slopes in CR birds are lower compared to PR birds in L3/NCM for both BS pairs ($\beta = -0.08 \pm 0.02$; $t(8225) = -2.5$; $p = 0.014$) and NS pairs ($\beta = -0.21 \pm 0.04$; $t(8225) = -4.9$, $p < 0.001$). All of the slopes were significantly greater than zero except for BS pairs from CR birds in L2a/L2b.

later stations (L1/CM and L3/NCM). We did not consider BS–NS pairs because of the difficulty of interpreting correlations between putatively excitatory and putatively inhibitory neurons.

We did not observe a significant effect of acoustical environment on noise correlations (Fig. 4B), but there was a difference in the relationship between signal and noise correlations. As has been commonly observed in the cortex[39,40], noise correlations tended to increase with signal correlations in all three brain areas and both cell types (Fig. 4C,D). This relationship between noise and signal correlations is bad for coding efficiency, because similarly tuned neurons with correlated noise provide more redundant information and less reduction in uncertainty compared to neurons with independent or negatively correlated noise[38]. In European starlings, learning to recognize the songs of other individuals inverts the typical relationship between signal and noise correlations, increasing the information ensembles of neurons carry about song identity[41]. We observed a similar though smaller effect in L3/NCM (Fig. 4C,D), such that pairs of BS neurons and pairs of NS neurons with similar tuning (high signal correlations) had lower noise correlations in CR birds compared to PR birds.

Auditory responses are more discriminable and selective in colony-reared birds

To characterize the effects of the acoustical environment on the functional properties of auditory neurons, we examined two complementary measures of sensory coding. The first measure, discriminability, assesses how reliably and distinctively a neuron responds to different stimuli. Neurons with more discriminable responses carry more information about which stimulus was presented. In this study, discriminability was quantified by using a timescale-free measure of spike train similarity[42] to make pairwise comparisons between the responses of the neuron in ten presentations of 9 songs (one song was excluded from this analysis because it was much shorter than the others). Examples of the similarity matrix that results from this comparison are shown for two example cells in Figure 5A. The similarity matrices were used with a k-nearest neighbors classifier to predict which stimulus was presented (Fig. 5B). Discriminability was taken as the proportion of trials that were correctly matched to the presented stimulus. The second measure, selectivity, assesses how narrowly a neuron is tuned. In this study we used a simple definition of selectivity based on the proportion of stimuli that evoked a response with a rate significantly greater than the

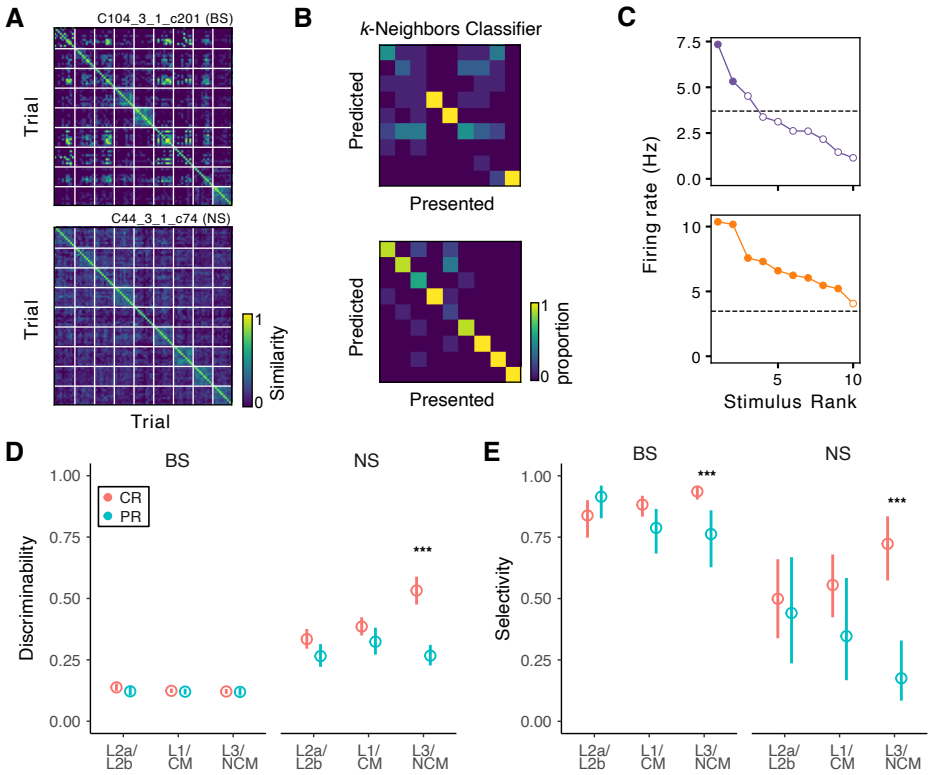

Fig. 5. Discriminability and selectivity of auditory responses. (**A**) Spike-train similarity matrices between all pairs of trials for two exemplar units. The trials have been sorted by song, as indicated by the grid of white lines. Blocks along the diagonal correspond to comparisons between trials of the same song. (**B**) Confusion matrices from the predictions of a classifier model for each of the exemplars shown in A. Correct predictions correspond to the diagonal. The proportion of correct predictions was higher when spike train similarity was high between trials for the same song and low between trials for different songs. (**C**) Average firing rates of the two exemplar neurons in A–B evoked by each of the songs. The dashed line indicates the average spontaneous rate for that unit. Filled circles indicate responses significantly greater than the spontaneous rate (GLM; see Materials and Methods). (**D**) Average discriminability (proportion of trials correctly classified) for BS and NS neurons in PR and CR birds. The responses of BS neurons were less discriminable on average than NS responses (LOR $= -1.3 \pm 0.06$; $z = -21.5$; $p < 0.001$). Discriminability of BS responses did not differ between PR and CR birds (LOR $= 0.06 \pm 0.06$; $z = 1.0$; $p = 0.31$). Among NS cells, CR responses were more discriminable than PR responses in L3/NCM (LOR $= 1.14 \pm 0.19$; $z = 6.0$; $p < 0.001$). (**E**) Average selectivity for BS and NS neurons in PR and CR birds. BS neurons were more selective on average than NS neurons (LOR $= 2.0 \pm 0.3$; $z = 8.4$; $p < 0.001$). Selectivity was higher in CR birds compared to PR birds in L3/NCM for both BS (LOR $= 1.5 \pm 0.4$; $z = 3.2$; $p = 0.0014$) and NS neurons (LOR $= 2.5 \pm 0.7$; $z = 3.9$; $p < 0.001$).

neuron's spontaneous firing (Fig. 5C). Selectivity and discriminability are complementary because they measure functional properties with an inherent tradeoff[27]. As seen in the examples, highly selective neurons tend to have low discriminability scores, because a large proportion of stimuli evoke weak and inconsistent responses that are difficult to discriminate. Conversely, neurons that give distinctive responses to many different motifs will tend to have high discriminability scores but poor selectivity. Our results show evidence of this tradeoff at the level of cell types: averaging across areas and rearing conditions, BS neurons had lower discriminability compared to NS neurons (Fig. 5D) but greater selectivity (Fig. 5E).

Discriminability and selectivity tended to increase at later stages of the processing hierarchy, but only in colony-reared birds (Fig. 5D,E). This was observed as a statistically significant difference between CR and PR birds in L3/NCM, with CR birds having more discriminable NS neurons and more selective BS and NS neurons in this area. If we continue to assume that NS neurons are inhibitory interneurons, this result indicates that inhibition in CR birds is not only stronger, but more reliable and more selective. This could contribute to more selective responses in the BS cells and to the decoupling of noise and signal correlations (Fig. 4).

Stimuli are decodable from smaller numbers of neurons in colony-reared birds

We hypothesized that the less correlated and more discriminable responses in CR animals would lead to higher coding efficiency at the population level. To test this hypothesis, we measured how well the spectrotemporal structure of song stimuli could be decoded from neural activity in CR or PR birds. We used a simple decoder that was trained to predict the spectrum of the stimulus at a given time t as a linear function of the population response over the subsequent 200 ms (Fig. 6A). The decoder performance was quantified using 10-fold cross-validation, with 9 songs used for training and 1 song for testing in each fold. When trained with the entire set of CR neurons, the decoder was able to make a prediction of the spectrotemporal structure of the test stimulus at a high level of detail and accuracy (Fig. 6B). The predictions from the PR neurons were not as good, but there were also fewer PR neurons. To control for this confound, we tested decoder performance on randomly selected subsets ranging in size on a logarithmic scale from 40 to 927 units, with 100 replicates per size. Decoder performance indeed increased with population size, but subsets of

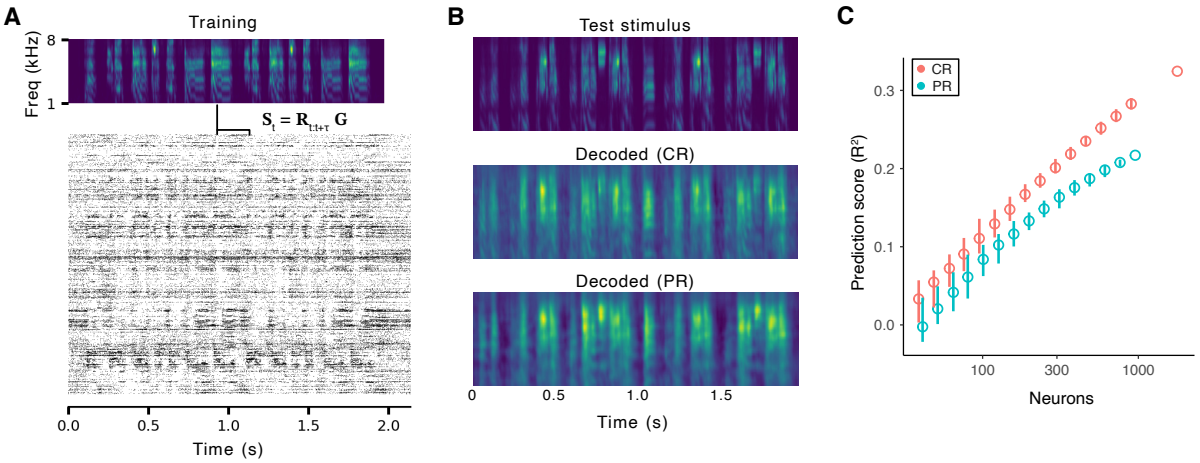

Fig. 6. Decoding spectrographic features of song from population responses. (**A**) Raster plot of responses from all CR neurons to a zebra finch song. A linear decoder (G) was trained on 9/10 songs using ridge regression (see Methods) to predict the spectrum of the stimulus at each time point (S_t) from the responses of the population in an interval immediately following that time ($R_{t:t+\tau}$). (**B**) To test decoder performance, the presented stimulus was decoded from the population response to the held-out song. Top spectrogram shows one test stimulus, middle shows the decoder prediction for the CR units, bottom shows the prediction for the PR units. (**C**) Decoder performance depends on the number of units used to train the model. The decoder was fit using subsets of the PR or CR units, sampled without replacement ($n = 100$ replicates per population size). Circles indicate median prediction score, whiskers the range between the 25% and 75% quantiles. Population sizes were matched across conditions, but the symbols are offset for clarity. The difference between median scores for CR and PR is significant for every population size (Wilcoxon rank-sum test, $p < 0.05$).

the CR neurons always outperformed equally sized subsets of PR neurons (Fig. 6C). In terms of coding efficiency, smaller numbers of CR neurons were able to match the performance of larger numbers of PR neurons; for example, around 300 CR neurons carried as much information about the spectrotemporal structure of the stimulus as 927 PR neurons.

The PR population showed evidence of a sublinear relationship between population size and decoder performance, consistent with diminishing returns from combining neurons with similar tuning. In contrast, the CR population showed a more linear trend, consistent with lower signal correlations and higher discriminability. With the caveat that this is a linear model of a highly nonlinear system, if these trends continued with larger sample sizes, we might expect the difference in information encoded by the whole auditory pallium in CR and PR birds to be quite large indeed.

Discussion

Species that communicate vocally require auditory systems that can reliably decode signals with complex acoustic features. In this study, we have shown that for a social songbird that breeds and raises its young in colonies, early exposure to the social-acoustical environment of the colony has wide-ranging effects on how conspecific song is processed at the cortical level. Compared to birds raised by pairs in acoustic isolation, colony-reared birds have higher but less correlated firing rates, more discriminable firing patterns, more selective tuning, and greater population-level coding efficiency for the spectrotemporal structure of song. These results are consistent with the idea that enriched experience is broadly beneficial to the development of sensory processing and support a model in which the developing brain adapts to the highly variable environment of many overlapping vocalizations by sharpening and strengthening inhibitory circuitry.

The effects of colony-rearing were seen in multiple brain regions and cell types, but the largest and most consistent differences were in the narrow-spiking cells of NCM and L3, so this discussion will focus primarily on those areas. NCM and L3 are located next to each other in the auditory pallium, sharing an indistinct boundary[43]. They are both considered higher-order areas because neither receives direct input from the core of the auditory thalamus[44] and because neurons in both areas have nonlinear receptive fields and sharp tuning for small numbers of songs[35,36,45]. L3 projects to NCM, which additionally exhibits learning-dependent responses consistent with a role

in recognizing familiar individuals by their songs[26,35,46,47] and in guiding sensorimotor learning for song production[48–50]. Thus, experience-dependent changes to the functional properties of NCM neurons may have a significant impact on perception and perceptual learning, although this remains to be demonstrated.

Although NS firing rates were higher on average in CR birds, so was their selectivity, and their firing patterns were more distinctive. This indicates that exposure to a complex social-acoustical environment not only increases the overall inhibitory response to conspecific song but also sharpens its tuning for specific songs and the component features. Sharpened inhibitory tuning may explain why the evoked firing rates of BS cells were unaffected even as NS firing rates increased more than 3-fold (Fig. 3B). Indeed, we recorded from over three times as many BS units per NCM/L3 site in CR birds as we did in PR birds (Fig 3C). Because extracellular recordings are biased for units that spike often enough to be isolated during spike sorting, the most likely explanation for this large difference in the number of units is that there were many cells in the PR birds with extremely low firing rates. This is the opposite of what we would expect to see if increased inhibitory activity in CR birds led to a general dampening of responses, providing further evidence that exposure to the colony has fine-grained, specific effects on auditory processing. Our results strongly suggest that these effects are related to the reconfiguration of functional connectivity for greater coding efficiency. NS neurons in NCM/L3 had lower signal correlations with each other, which is suggestive of mutual inhibition. For both BS and NS neurons, the relationship between signal and noise correlations was altered, such that noise correlations were lower in similarly tuned pairs. This effect, which enhances coding efficiency[38], has also been reported in European starlings for stimuli that birds had learned to discriminate in an operant task[41]. Confirming a link between functional connectivity and coding efficiency, we found that randomly selected populations of CR neurons carried more information about the spectrotemporal structure of song than PR populations of the same size (Fig. 6).

To synthesize these results, we propose a conceptual model based on mutual inhibition of excitatory ensembles[51]. NCM neurons receive excitatory inputs from upstream areas that are tuned to spectrotemporal features found in the songs of many individuals[35,36,52,53]. NCM neurons are sensitive to the temporal sequencing of these features[26], which could result from the formation of Hebbian ensembles through repeated exposure to the unique songs of individual birds. Whereas CR birds are exposed to songs and calls from dozens of individuals, PR birds

only hear the calls of their immediate family, the song of their father, and the (very similar) songs of their siblings. This impoverished experience would lead to less plasticity and the formation of fewer ensembles, explaining why there were so many fewer responsive neurons in PR birds. In contrast, the richer experience of CR birds would drive the formation of more ensembles and more robust excitatory connectivity. However, this rich experience comes with a cost: there is a nearly continuous background of overlapping vocalizations (Fig. 1), which would activate multiple ensembles simultaneously. Interference between ensembles would also result because songs share similar notes and syllables (e.g., the short harmonic stacks seen in all but one of the exemplar songs in Fig. 2C). One mechanism for achieving pattern separation with correlated noise and overlapping features is mutual inhibition, which creates winner-take-all dynamics that suppress activity from features in the background[54]. In support of this model, blocking inhibition in NCM unmasks peaks of excitation in a context- and learning-dependent manner[26,55]. The experience-dependent increases in NS firing rates, selectivity, and discriminability we observed in this study are all consistent with stronger drive from excitatory ensembles onto inhibitory neurons, and the reductions in signal correlations among NS neurons and noise correlations among similarly tuned BS and NS neurons are consistent with a pattern of mutual inhibition. A testable prediction of this model is that BS neurons in CR birds will be more invariant to a background of colony noise due to stronger and more specific mutual inhibition.

The results of this study are consistent with an extensive literature showing how the statistics of early experience shape development of the functional organization of the auditory cortex[7,8,11,15,56]. With a few exceptions[10,18], most of the prior work employed artificial stimuli, so the current study represents an advance in understanding how principles of experience-dependent plasticity apply to natural development in social, vocal communicators. However, a limitation of the experimental design is that there are multiple differences between the CR and PR conditions that are difficult to dissociate. Most notably, CR birds experience more signal and more noise: they hear many more songs than PR birds, but in an acoustical environment that masks the features of individual songs (Fig. 1). CR birds also have the opportunity for vocal and visual interactions with individuals outside their family, and there may be more diffuse effects of social density on brain development[57]. It also remains to be seen to what extent the effects we observed are the result of an early sensitive period[58] or more recent plasticity[28]. Future work could begin to address these questions by

controlling the timing and statistics of the acoustical background presented via loudspeaker to PR clutches.

Humans are social animals, but the range of social and acoustical environments they experience in infancy is enormous. Our results demonstrate that a complex early environment has a profound impact on the functional properties and connectivity of a higher-order auditory area, consistent with the well-known link between an infant's early language experience and phonetic processing[1,2] but further suggesting that the development of auditory processing has a broader dependence on the diversity and complexity of the acoustical environment in which this experience occurs. Understanding the mechanisms of this plasticity and its effects on perception and auditory learning may shed light on a broad range of language-learning impairments that are associated with deficits in auditory processing[59–62].

Methods

Animals

All procedures were performed according to National Institutes of Health guidelines and protocols approved by the University of Virginia Institutional Animal Care and Use committee. Zebra finches (*Taeniopygia guttata*) were bred from our local colony from 10 different pairs. All birds received finch seed (Abba Products, Hillside, NJ) and water ad libitum and were kept on a 16:8 light-dark schedule in temperature and humidity-controlled rooms (22–24 °C).

Experimental Rearing Conditions

Zebra finches were reared in individual cages that were initially placed in the colony room, which houses around 70 male and female finches of varying ages. In the colony-reared (CR) condition, the subjects remained in the colony room until they were used in an experiment. In the pair-reared (PR) condition, families were moved to an acoustic isolation chamber (Eckel Industries) five days after the first chick hatched. Hatchlings were fed by hand after the move as needed. In both conditions, the parents were separated from their chicks at 35 dph (days post hatch). CR juveniles continued to be housed in the colony with other juveniles or in large single-sex flight aviaries. PR

chicks remained with their siblings in the acoustic isolation chamber until they were used in an experiment.

Acoustic Recordings and Analysis

Sound recordings were taken of the acoustical environment for one CR clutch and one PR clutch. A small hole was drilled in the lid of the nest box, a piece of foam rubber padding with a cavity for a microphone was glued above the hole, and an omnidirectional lavalier microphone (Shure SM93) was placed in the cavity. The signal was amplified and digitized at 48 kHz using a Focusrite Scarlett 2i2 and recorded directly to disk using custom software (jrecord v2.1.5 or v2.1.9, https://github.com/melizalab/jill). Before installing the lid on the nestbox, a recording was made of a 1 kHz calibration tone emitted by an R8090 (Reed Instruments) calibrator, which was placed immediately above the hole on the nest side of the lid. The amplitude of this tone at the opening of the R8090 was 80 dB SPL as measured by a calibrated sound meter (NTi XL2 with an M4261 microphone). The gain on the amplifier/digitizer was kept constant throughout the recording. The CR clutch comprised four chicks and was recorded for 7 days starting when the chicks were 26–29 dph. The PR clutch comprised six chicks and was recorded for 33 days starting when the chicks were 4–11 dph; however, only the recordings made when the chicks were a comparable age to the CR birds was used in this analysis.

To convert the units of the recordings to dB SPL, the power spectrum of the calibration recording was computed using Welch's method with a 10 ms flat-top window, and a scaling factor was determined for the 80 dB SPL peak at 1 kHz. The amplitude statistics of the nestbox recordings were analyzed in segments of 10 min that overlapped by 50%. For each segment, the calibration scaling factor was applied, followed by an A-weighting digital filter. The segments were further subdivided into 10 ms frames with 50% overlap, and a Fourier transform with a Hanning window was used to estimate the power spectral density (PSD). The PSD in each frame was integrated and log-transformed to obtain amplitude on the dB_A SPL scale. The 0% (minimum), 25%, 50%, 75%, and 100% (maximum) quantiles of the amplitude were then computed for each 10-min segment.

The acoustical structure of the colony noise was analyzed using sound texture statistics based on a model of auditory processing that separates acoustic signals into log-spaced frequency bands

corresponding to locations on the basilar membrane, extracts the amplitude envelope in each band on a nonlinearly compressed scale, and then further analyzes each acoustic frequency band into channels corresponding to different modulation rates[32]. Statistics for the colony were computed from sixty-six 10 s segments sampled from the CR nestbox recording. The recording was divided into hours, and one sample was chosen at a random time from each of 66 daylight hours. One sample was excluded because it included sounds of the animal care staff physically manipulating the cage, which was a relatively rare occurrence. For comparison, texture statistics were calculated for 62 samples of zebra finch song from 12 individual males (5 or 6 songs per bird), drawn from the lab's collection of recordings. These recordings were obtained in acoustic isolation chambers using the same digitization software and hardware but a different microphone (AudioTechnica Pro70). Both the SM93 and the Pro70 have frequency response curves that are flat to within ± 10 dB between 100–14,000 Hz. Song samples were chosen to be 10 s long; for some birds with shorter songs, these samples comprised bouts of 2–3 motifs separated by gaps of 1–2 seconds. For both colony noise and song, texture statistics were calculated separately for each segment and then averaged.

Texture statistics were calculated using the same algorithm and parameters reported by McDermott and Simoncelli[32], except that our acoustic frequency channels spanned from 246–9005 Hz, as the lower frequencies in our recordings of colony noise or individuals did not contain any vocal signals. We reimplemented the algorithm in Python so that we could efficiently perform analysis on the large number of samples used in this study (https://github.com/melizalab/colony-noise, to be released on acceptance of the manuscript). We confirmed that the output of our implementation matched that of the MATLAB code provided by McDermott and Simoncelli (http://mcdermottlab.mit.edu/downloads.html).

Electrophysiology

Surgery

Three days before recordings, adult birds (90–110 dph) were anesthetized with 3.5% isoflurane (Southmedic). Feathers were removed from the top of the head and the bird was placed in a stereotaxic holder. The scalp was prepared with 7.5% povidone iodine (McKesson), Neosporin

(Johnson & Johnson), and 2% Lidocaine (Henry Schein) prior to incision. The recording site was identified using stereotaxic coordinates relative to the Y-sinus. A metal pin was affixed to the skull rostral to the recording site with dental acrylic, and the first layer of skull over the recording site was removed with a scalpel. Subjects were allowed to recover between three to four days before recordings.

On the day of recording, the bird was given three intramuscular injections of 20% urethane (0.5 mL/g; Sigma-Aldrich), separated by 20 min. The bird was restrained in a padded 50 mL conical tube, and the head pin was attached to a stand in the recording chamber (Eckel Industries). For most birds, a craniotomy was performed adjacent to the recording site for ground wire placement; in other recordings the ground wire was placed in the well surrounding the recording site. The second layer of skull was removed over the recording site, the dura was thinned using an Ultra Needle (Electron Microscopy Sciences), and a well was formed with Kwik-Cast around the recording site to hold agarose and phosphate-buffered saline.

Stimulus Presentation

Acoustic stimuli were presented with custom Python software (https://github.com/melizalab/open-ephys-audio; v0.1.7) through a Focusrite 2i2 analog-to-digital converter, a Samson Servo 120a amplifier, and a Behringer Monitor Speaker 1C. The stimuli comprised songs selected from recordings of 10 male zebra finches in our colony. Songs included 1–2 motifs and were 2.52 ± 0.28 s in duration (mean ± SD; range 1.92–2.88 s). The stimuli were filtered with a 4th-order Butterworth highpass filter with a cutoff of 150 Hz, a 2 ms squared cosine ramp was applied to the beginnings and the ends, and the waveforms were rescaled to have an RMS amplitude of 60 dB_A SPL at the position of the bird's head. The songs were combined into sequences, each consisting of 10 permutations of the songs, following a balanced Latin square design. The songs within each sequence were separated by gaps of 0.5 s, and the sequences were presented with a gap of 5 s. The stimuli and responses analyzed in this study were part of a larger study in which we embedded the sequences in a background of varying amplitude; the background consisted of synthetic colony noise generated by the McDermott and Simoncelli[32] algorithm. The stimuli used in this study included this background noise, but at an amplitude of –10 dB_A SPL, which is well below the threshold of hearing. Therefore, we treated the stimuli as unmodified.

Data Acquisition

Neural recordings were made using 128-channel silicon electrodes (128 AxN Sharp, Masmanidis Lab) connected to an Intan RHD 128-channel headstage. Data were collected at 30 kHz by the Open Ephys Acquisition Board connected to a computer running Open Ephys GUI software (v0.5.3.1). The electrode was coated with DiI (ThermoFisher Scientific) and inserted at a ventro-caudal to dorso-rostral angle that confined the pentrations to one of three groups of auditory areas: the thalamorecipient areas L2a and L2b; the superficial areas L1, CLM (caudolateral mesopallium), and CMM (caudomedial mesopallium); and the deep/secondary areas L3 and NCM (caudomedial nidopallium). Recordings were only included if at least 80% of the contacts on the electrode were in one of these areas. After contacting the brain surface, the probe was advanced at 1 µm/s. A separate collection of zebra finch songs was played during electrode advancement. Electrode movement was stopped once the local field potentials and spiking indicated reliable auditory responses across the whole probe. Once the electrode was in position, the well was filled with 2% agarose (Sigma-Aldrich) and phosphate-buffered saline.

Histology

Immediately after recording, animals were administered a lethal intramuscular injection of Euthasol (Vibrac) and perfused transcardially with 10 U/mL solution of sodium heparin in PBS (in mM: 10 478 Na2HPO4, 154 NaCl, pH 7.4) followed by 4% formaldehyde (in PBS). Brains were removed from the skull, postfixed overnight in 4% paraformaldehyde, blocked along the midline, and embedded in 3% agar (Bio-Rad). Sections were cut at 60 µm on a vibratome and mounted on slides. After drying for 30 min, the sections were coverslipped with Prolong Gold with DAPI (Thermo Fisher, catalog P36934; RRID:SCR_05961). The sections were imaged using epifluorescence with DAPI and Texas Red filter cubes to locate the DiI-labeled penetrations.

Spike Sorting and Classification

Spikes were sorted offline with Kilosort 2.5[63]. Clusters were automatically excluded if they had a contamination score higher than 10 or contained fewer than 100 spikes (corresponding to a rate of 0.029 Hz over the 58 min session), as it was not possible to assess quality with so few

events. Clusters were further curated by visual inspection (phy 2.0; https://github.com/cortex-lab/phy) for spheroid distribution in the feature space, very low refractory period violations in the autocorrelogram, stable firing rate throughout the recording, and a typical average spike waveform. Individual spikes were excluded as artifacts if their amplitude was more than 6 times the amplitude of the average spike from that unit. To classify units as narrow-spiking (NS) or broad-spiking (BS), the spike waveforms for each unit were averaged after resampling to 150 kHz using sinc interpolation. Taking the time of the maximum negative deflection (trough) as 0, the spike width was defined as the time of the maximum positive deflection (peak) following the trough, and the peak/trough ratio was measured from the (absolute) magnitude of the peak and trough. A Gaussian mixture model with two groups was fit to these features, and units were classified as narrow-spiking if they were assigned to the group with the lower spike width and higher peak/trough ratio. Nearly identical results were obtained if separate classifiers were fit for each brain area.

Data Analysis

Data were analyzed using custom Python and R code (https://github.com/melizalab/cr_pr_adults, to be released on manuscript acceptance). Spike trains were separated into non-overlapping intervals corresponding to songs, beginning at the start of the song and ending 0.5 s after the end of the song. As noted above, we only analyzed responses to songs where the background noise was neglible, for a total of 10 trials per 10 songs per unit. For spontaneous firing, we used 1 s intervals taken from the periods of silence between all the motif sequences in the stimulus set, for a total of 110 trials per unit.

Spontaneous and evoked firing rates were quantified using generalized linear mixed-effects models (GLMMs) with spike count as the dependent variable with a Poisson distribution. The parameters of the model were estimated in R using lme4 (v1.1-35). The model specification was `n_events ~ area*spike*rearing + offset(duration) + (1|unit) + (1|motif)`, where `area*spike*rearing` defines fixed effects for brain region, cell type, rearing condition, and their interactions; offset(duration) defines a constant offset term for the duration of the interval in which the spikes were counted; and `(1|unit)` and `(1|motif)` define random intercepts for each unit

and stimulus (omitted from the model for spontaneous rate). Because the model is nonlinear and has many interactions, effects and confidence intervals are reported as estimated marginal means, calculated in R using the emmeans package (v1.10.1). Note that emmeans reports test statistics for post-hoc tests on GLMMs as z-scores, with infinite degrees of freedom. Contrasts are reported on the log ratio (LR) scale.

Signal and noise correlations were also computed using the number of spikes evoked in the interval between the start of each motif and 0.5 s after its end. Following Jeanne et al.[41], for each pair of simultaneously recorded neurons, the signal correlation was calculated as the Pearson correlation coefficient between the trial-averaged firing rates to the ten songs. The noise correlation was calculated by taking the Pearson correlation coefficient across trials of the same song and then averaging over songs. The noise correlation was corrected by subtracting the shift-predictor correlation, which is calculated the same way as the noise correlation but with trials that were not recorded simultaneously[64]. Only BS–BS and NS–NS pairs were considered, due to the difficulty of interpreting BS–NS correlations. Only auditory units (see below) were considered, and pairs recorded on the same electrode were excluded. The effect of rearing condition on signal correlations was estimated using linear regression in R, with area, cell type, rearing condition, and their interactions as predictors. The same model was used for noise correlations. To test whether the relationship between signal and noise correlations was affected by rearing condition, we used a linear model with noise correlation as the dependent variable and area, cell type, rearing condition, signal correlation, and their interactions as fixed effects. Contrasts for CR versus PR were computed for each area and cell type using emmeans.

Discriminability was calculated for each unit by computing the pairwise similarity between all of its song-evoked responses. Responses were truncated to the duration of the shortest stimuli to remove information encoded in the response length. One stimulus was excluded because it was much shorter than the others. We used the SPIKE-synchronization metric[42] implemented in the pyspike Python package (v0.7.0; http://mariomulansky.github.io/PySpike/), yielding a 90×90 matrix of similarity scores that were used to train a k-nearest neighbors classifier (scikit-learn v1.3.2). That is, each trial was classified by finding the k most similar trials (the neighbors) and assigning it the label that was the most common in that group. We used $k = 9$, but other values produced nearly identical results. This yielded a 9×9 confusion matrix with one dimension corresponding to the

stimulus that was actually presented and the other to the label that was assigned to the trial. Each unit received a score for the proportion of trials that were correctly classified. We used a GLMM with the number of correctly classified trials as the binomial dependent variable to quantify the effect of rearing condition. The model specification in R was `cbind(n_correct, 90 - n_correct) ~ area*spike*rearing + (1|unit)`. Contrasts were computed using emmeans and are reported on the log odds ratio (LOR) scale. Discriminability was also used to classify neurons as auditory: a neuron was deemed to be auditory if its average score for all the songs was more than 1.64 standard deviations (the one-sided 95% confidence level for a z-test) above the mean for 500 permutations in which the song labels were randomly shuffled.

Selectivity was calculated for each unit as 1 minus the proportion of songs that evoked a firing rate that was significantly greater than the spontaneous rate. We used a generalized linear model with spike count as the Poisson-distributed dependent variable to determine significance. For the rare neurons that did not spike at all during the spontaneous intervals (2%, $n = 72/2954$), a single spike was added to the spontaneous count to regularize the model estimates. To quantify the effect of rearing condition, we used a GLMM with the number of songs with rates above spontaneous as the binomial dependent variable. The model specification in R was `cbind(10 - n_responsive, n_responsive) ~ area*spike*rearing + (1|unit)`. Note that defining responsive motifs as "failures" gives a selectivity score where 0 is a unit that responds to all the stimuli and 1 is a unit that responds to none. Only auditory units were included in this analysis. Contrasts are reported on the LOR scale.

Stimulus reconstruction from population responses was implemented using a linear decoder. The model is similar to the spectrotemporal receptive field, in which the expected firing rate of a single neuron at a given time point t is modeled as a linear function of the stimulus spectrogram immediately prior to t. In the linear decoding model, the relationship is reversed, and the expected stimulus at time t is modeled as a linear function of the response that follows. Using a discrete time notation where s_t is the stimulus in the time bin around t, and r_t is the response of a single neuron in the same time bin, then the expected value of the stimulus is given by

$$\mathrm{E}\left[s_t\right] = \beta + r_t g_0 + r_{t+1} g_1 + \cdots + r_{t+k} g_k,$$

where k is the number of time bins one looks into the future, $\mathbf{g} = (g_0, g_1, \ldots, g_k)$ are the linear coefficients of the model, and β is a constant intercept term. If the errors are independent and normally distributed around the expectation with constant variance σ^2, then this is an ordinary linear model. If there are n time bins in the stimulus, then the stimulus is a vector $\mathbf{s} = (s_0, \ldots, s_n)$ drawn from a multivariate normal distribution. In vector notation,

$$\mathbf{s} | \mathbf{g}, \sigma^2, \mathbf{R}, \beta \sim N(\mathbf{R}\mathbf{g} + \beta, I\sigma^2),$$

where $\mathbf{R}$ is the $n \times k$ Hankel matrix of the response. Without any loss of generality, the model can be expanded to include the responses of multiple neurons. If there are p neurons, then r_t becomes a p-element vector $(r_{1,t}, \ldots, r_{p,t})$, $\mathbf{R}$ becomes a $n \times pk$ matrix formed by concatenating the Hankel matrices for each of the neurons, and $\mathbf{g}$ becomes a pk-element vector.

In this study, the response matrix was constructed from the peristimulus time histograms (PSTHs) averaged over 10 trials using a bin size of 2.5 ms. We fit the model using all the PR neurons, all the CR neurons, or subsets of the PR or CR populations randomly selected without replacement. We used a k of 80, corresponding to lags from 0–200 ms. The response matrix was projected into a basis set consisting of 20 nonlinearly spaced raised cosines[65]. The width of each basis function increased with lag, giving higher temporal resolution at short lags and lower resolution at long lags. This allowed the inclusion of longer lags without exploding the number of parameters.

Stimuli were converted to time-frequency representations using a gammatone filter bank, implemented in the Python package gammatone (version 1.0) with 40 log-spaced frequency bands from 1–8.5 kHz, a window size of 5 ms, and a step size of 2.5 ms. Power was log-transformed.

We used ridge regression to estimate the parameters of the model because the number of parameters pk was typically larger than the number of time bins n. The decoder performance was estimated using a 10-fold cross-validation strategy. There were 10 data folds generated by holding

out each of the songs. For each fold, the 9 training songs were used to estimate $\hat{g}_\lambda$, where λ denotes the ridge shrinkage penalty. We tested 30 λ values on a logarithmic scale from $10^{-1} - 10^{7}$. $\hat{g}_\lambda$ was used to decode the stimulus $\tilde{S}$ from the responses to the held-out motif $\tilde{R}$. The score for a given prediction was quantified as the coefficient of determination R^2. The cross-validated prediction score was computed by averaging scores across folds for each value of λ and then taking the best score. Ridge regression and cross-validation were performed in Python with the scikit-learn library (v1.3.2).

Table 1. Sample sizes for birds, recordings, units, and pairs of simultaneously recorded units

		L2a/L2b		L1/CM		L3/NCM		Total	
		CR	PR	CR	PR	CR	PR	CR	PR
Birds		6	6	5	4	4	4	9 (4♂)	7 (5♂)
Recordings		8	11	12	7	8	10	28	28
Units	BS	262	153	642	346	487	190	1391	689
	NS	132	76	174	69	81	93	387	238
	Total	394	229	816	415	568	283	1778	927
Pairs	BS	1188	129	1957	841	2111	305	5256	1275
	NS	479	62	619	76	355	127	1453	265
	Total	1667	191	2576	917	2466	432	6709	1540

Acknowledgements

We would like to thank Jonah Weissman and Christof Fehrman for assistance with coding an early version of the linear decoder. Bao Le, Ayush Sagar, and Crystal Gong assisted in developing and testing the operant apparatus used in this study. This work was supported by NSF IOS-1942480 and NIH R01DC018621

Author Contributions

SM and CDM conceived and planned the experiments. SM carried out the experiments. SM and CDM analyzed the data and prepared the figures. SM and CDM wrote the manuscript.

References

[1] Werker, J. F. & Lalonde, C. E. Cross-language speech perception: Initial capabilities and developmental change. *Dev Psychol* **24**, 672–683 (1988).

[2] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606–608 (1992).

[3] Bidelman, G. M., Moreno, S. & Alain, C. Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage* **79**, 201–212 (2013).

[4] Bosseler, A. N. *et al.* Theta brain rhythms index perceptual narrowing in infant speech perception. *Front Psychol* **4**, 690 (2013).

[5] Zhang, L. I., Bao, S. & Merzenich, M. M. Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience* **4**, 1123–1130 (2001).

[6] Chang, E. F. & Merzenich, M. Environmental noise retards auditory cortical development. *Science* **300**, 498–502 (2003).

[7] de Villers-Sidani, E., Chang, E. F., Bao, S. & Merzenich, M. M. Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *J Neurosci* **27**, 180–189 (2007).

[8] Zhou, X. & Merzenich, M. M. Enduring effects of early structured noise exposure on temporal modulation in the primary auditory cortex. *Proc Natl Acad Sci USA* **105**, 4423–4428 (2008).

[9] Insanally, M. N., Kover, H., Kim, H. & Bao, S. Feature-dependent sensitive periods in the development of complex sound representation. *J Neurosci* **29**, 5456–5462 (2009).

[10] Bao, S., Chang, E. F., Teng, C.-L., Heiser, M. A. & Merzenich, M. M. Emergent categorical representation of natural, complex sounds resulting from the early post-natal sound environment. *Neuroscience* **248**, 30–42 (2013).

[11] Homma, N. Y., Hullett, P. W., Atencio, C. A. & Schreiner, C. E. Auditory cortical plasticity dependent on environmental noise statistics. *Cell Reports* **30**, 4445–4458 (2020).

[12] Marler, P. & Tamura, M. Culturally Transmitted Patterns of Vocal Behavior in Sparrows. *Science* **146**, 1483–1486 (1964).

[13] Gobes, S. M. H., Jennings, R. B. & Maeda, R. K. The sensitive period for auditory-vocal learning in the zebra finch: Consequences of limited-model availability and multiple-tutor paradigms on song imitation. *Behav Processes* **163**, 5–12 (2019).

[14] Sturdy, C. B., Phillmore, L. S., Sartor, J. J. & Weisman, R. G. Reduced social contact causes auditory perceptual deficits in zebra finches, Taeniopygia guttata. *Anim Behav* **62**, 1207–1218 (2001).

[15] Amin, N., Gastpar, M. & Theunissen, F. E. Selective and efficient neural coding of communication signals depends on early acoustic and social environment. *PLoS ONE* **8**, e61417 (2013).

[16] Chen, Y., Clark, O. & Woolley, S. C. Courtship song preferences in female zebra finches are shaped by developmental auditory experience. *Proc Biol Sci* **284**, 20170054 (2017).

[17] Woolley, S. M. N. Early experience shapes vocal neural coding and perception in songbirds. *Dev Psychobiol* **54**, 612–631 (2012).

[18] Moore, J. M. & Woolley, S. M. N. Emergent tuning for learned vocalizations in auditory cortex. *Nat Neurosci* **22**, 1469–1476 (2019).

[19] Waser, P. M. & Brown, C. H. Habitat acoustics and primate communication. *Am J Primatol* **10**, 135–154 (1986).

[20] Brumm, H. & Slabbekoorn, H. Acoustic communication in noise. *Adv Stud Behav* **35**, 151–209 (2005).

[21] Maye, J., Werker, J. F. & Gerken, L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, B101–11 (2002).

[22] Cherry, E. C. Some Experiments on the Recognition of Speech, with One and with Two Ears. *J Acoust Soc Am* **25**, 975–979 (1953).

[23] Zann, R. A. *The Zebra Finch: A synthesis of field and laboratory studies* (Oxford University Press, Oxford, 1996).

[24] Elie, J. E. & Theunissen, F. E. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim Cogn* **19**, 285–315 (2016).

[25] Wang, L., Narayan, R., Graña, G., Shamir, M. & Sen, K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci* **27**, 582–589 (2007).

[26] Schneider, D. & Woolley, S. Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* **79**, 141–152 (2013).

[27] Elie, J. E. & Theunissen, F. E. Meaning in the avian auditory cortex: neural representation of communication calls. *Eur J Neurosci* **41**, 546–567 (2015).

[28] Yang, L. & Vicario, D. Exposure to a novel stimulus environment alters patterns of lateralization in avian auditory cortex. *Neuroscience* **285**, 107–118 (2015).

[29] Chen, A. N. & Meliza, C. D. Experience- and sex-dependent intrinsic plasticity in the zebra finch auditory cortex during song memorization. *J Neurosci* **40**, 2047–2055 (2020).

[30] Lu, Y. *et al.* Rapid, activity-dependent intrinsic plasticity in the developing zebra finch auditory cortex. *J Neurosci* **43**, 6872–6883 (2023).

[31] Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J. & Durlach, N. I. Informational masking. In Yost, W. A., Popper, A. N. & Fay, R. R. (eds.) *Auditory Perception of Sound Sources*, 143–189 (Springer US, Boston, MA, 2008).

[32] McDermott, J. H. & Simoncelli, E. P. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron* **71**, 926–940 (2011).

[33] Woolley, S. M. N., Fremouw, T. E., Hsu, A. & Theunissen, F. E. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* **8**, 1371–1379 (2005).

[34] Wang, Y., Brzozowska-Prechtl, A. & Karten, H. J. Laminar and columnar auditory cortex in avian brain. *Proc Natl Acad Sci USA* **107**, 12676–12681 (2010).

[35] Meliza, C. D. & Margoliash, D. Emergence of selectivity and tolerance in the avian auditory cortex. *J Neurosci* **32**, 15158–15168 (2012).

[36] Calabrese, A. & Woolley, S. M. Coding principles of the canonical cortical microcircuit in the avian brain. *Proc Natl Acad Sci USA* **112**, 3517–3522 (2015).

[37] Bottjer, S. W., Ronald, A. & Kaye, T. Response properties of single neurons in higher level auditory cortex of adult songbirds. *J Neurophysiol* **121**, 218–237 (2019).

[38] Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat Rev Neurosci* **7**, 358–366 (2006).

[39] Bair, W., Zohary, E. & Newsome, W. T. Correlated firing in macaque visual area mt: Time scales and relationship to behavior. *J Neurosci* **21**, 1676–1697 (2001).

[40] Gu, Y. *et al.* Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**, 750–761 (2011).

[41] Jeanne, J. M., Sharpee, T. O. & Gentner, T. Q. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**, 352–363 (2013).

[42] Kreuz, T., Bozanic, N. & Mulansky, M. Spike-synchronization: a parameter-free and time-resolved coincidence detector with an intuitive multivariate extension. *BMC neuroscience* **16** (2015).

[43] Fortune, E. S. & Margoliash, D. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (Taenopygia guttata). *J Comp Neurol* **325**, 388–404 (1992).

[44] Vates, G. E., Broome, B. M., Mello, C. V. & Nottebohm, F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J Comp Neurol* **366**, 613–642 (1996).

[45] Sen, K., Theunissen, F. E. & Doupe, A. J. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* **86**, 1445–1458 (2001).

[46] Chew, S. J., Vicario, D. S. & Nottebohm, F. A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc Natl Acad Sci USA* **93**, 1950–1955 (1996).

[47] Thompson, J. V. & Gentner, T. Q. Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J Neurophysiol* **103**, 1785–1797 (2010).

[48] Bolhuis, J. J., Zijlstra, G. G., Boer-Visser, A. M. D. & Zee, E. A. V. D. Localized neuronal activation in the zebra finch brain is related to the strength of song learning. *Proc Natl Acad Sci USA* **97**, 2282–2285 (2000).

[49] Phan, M. L., Pytte, C. L. & Vicario, D. S. Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc Natl Acad Sci USA* **103**, 1088–1093 (2006).

[50] Adret, P., Meliza, C. D. & Margoliash, D. Song tutoring in presinging zebra finch juveniles biases a small population of higher-order song-selective neurons toward the tutor song. *J Neurophysiol* **108**, 1977–1987 (2012).

[51] Harris, K. D. & Mrsic-Flogel, T. D. Cortical connectivity and sensory coding. *Nature* **503**, 51–58 (2013).

[52] Woolley, S. M., Gill, P. R., Fremouw, T. & Theunissen, F. E. Functional groups in the avian auditory system. *J Neurosci* **29**, 2780–2793 (2009).

[53] Kozlov, A. S. & Gentner, T. Q. Central auditory neurons have composite receptive fields. *Proc Natl Acad Sci USA* **113**, 1441–1446 (2016).

[54] Espinoza, C., Guzman, S. J., Zhang, X. & Jonas, P. Parvalbumin+ interneurons obey unique connectivity rules and establish a powerful lateral-inhibition microcircuit in dentate gyrus. *Nat Commun* **9**, 4605 (2018).

[55] Thompson, J. V., Jeanne, J. M. & Gentner, T. Q. Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. *J Neurophysiol* **109**, 721–733 (2013).

[56] Zhang, L. I., Bao, S. & Merzenich, M. M. Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nat Neurosci* **4**, 1123–1130 (2001).

[57] Møller, A. Brain size, head size and behaviour of a passerine bird. *J Evol Biol* **23**, 625–635 (2010).

[58] Schroeder, K. M. & Remage-Healey, L. Adult-like neural representation of species-specific songs in the auditory forebrain of zebra finch nestlings. *Dev Neurobiol* **81**, 123–138 (2021).

[59] Mody, M., Studdert-Kennedy, M. & Brady, S. Speech perception deficits in poor readers: auditory processing or phonological coding? *J Exp Child Psychol* **64**, 199–231 (1997).

[60] Tsao, F.-M., Liu, H.-M. & Kuhl, P. K. Speech Perception in Infancy Predicts Language Development in the Second Year of Life: A Longitudinal Study. *Child Dev* **75**, 1067–1084 (2004).

[61] Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X. & Lorenzi, C. Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci USA* **102**, 14110–14115 (2005).

[62] Pennington, B. F. & Bishop, D. V. M. Relations among speech, language, and reading disorders. *Annu Rev Psychol* **60**, 283–306 (2009).

[63] Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* (2016).

[64] Ponce-Alvarez, A., Thiele, A., Albright, T. D., Stoner, G. R. & Deco, G. Stimulus-dependent variability and noise correlations in cortical mt neurons. *Proc Natl Acad Sci USA* **110**, 13162–13167 (2013).

[65] Pillow, J. W., Paninshi, L., Uzzell, V. J., Simoncelli, E. P. & Chichilnisky, E. J. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci* **25**, 11003–11013 (2005).