Published in final edited form as:

Nat Neurosci. 2019 September ; 22(9): 1469–1476. doi:10.1038/s41593-019-0458-4.

Emergent tuning for learned vocalizations in auditory cortex

Jordan M. Moore[1,2], **Sarah M. N. Woolley**[1,2,3,4,*]

[1]Department of Psychology, Columbia University, New York, New York, USA

[2]Zuckerman Institute, Columbia University, New York, New York, USA

[3]Kavli Institute for Brain Science, Columbia University, New York, New York, USA

[4]Center for Integrative Animal Behavior, Columbia University, New York, New York, USA

Abstract

Vocal learners use early social experience to develop auditory skills specialized for communication. However, it is unknown where in the auditory pathway neural responses become selective for vocalizations or how the underlying encoding mechanisms change with experience. We used a vocal tutoring manipulation in two species of songbird to reveal that tuning for conspecific song arises within the primary auditory cortical circuit. Neurons in the deep region of primary auditory cortex responded more to conspecific songs than other species' songs and more to species-typical spectrotemporal modulations, but neurons in the intermediate (thalamorecipient) region did not. Moreover, birds that learned song from another species exhibited parallel shifts in selectivity and tuning toward the tutor species' songs in the deep but not intermediate region. Our results locate a region in the auditory processing hierarchy where an experience-dependent coding mechanism aligns auditory responses with the output of a learned vocal motor behavior.

INTRODUCTION

Animal communication relies on the transfer of information between senders and receivers via signals, and receivers have sensory capabilities specialized for encoding those signals to promote efficient social exchanges[1,2]. For example, vocal communicators have auditory perceptual skills that are tailored to the acoustic features of species-specific communication sounds[3–6]. In humans and songbirds, vocal communication signals are learned[7] and every stage of vocal development relies on hearing[5,8], suggesting that auditory perceptual and

* sw2277@columbia.edu.

AUTHOR CONTRIBUTIONS

J.M. designed experiments, collected data, analyzed data, and wrote the manuscript. S.W. designed experiments, analyzed data, and wrote the manuscript.

ACCESSION CODES

N/a

CODE AVAILABILITY

The code used to analyze data in this study is available from the corresponding author upon request.

COMPETING INTERESTS

The authors declare no competing interests.

vocal motor skills develop in concert. Humans develop life-long auditory skills based on infant experiences with native-language sounds[9,10], and childhood experience with a language predicts adult proficiency[11]. Similarly, songbirds develop life-long auditory preferences for[12] and vocal copies of[13] adult songs they hear as juveniles, and adults discriminate among songs of their own species better than those of other species[3].

In both humans and songbirds, adult perceptual skills reflect an individual's early experience with vocal signals, suggesting that experience-dependent plasticity permanently organizes perception for communication. Human auditory cortex (AC) responds more strongly to speech than to other sounds[14], and responses depend on learned phonetic boundaries[15,16] in a way that correlates with language fluency[17]. The avian auditory cortex is potentially homologous to mammalian AC[18] and exhibits similar anatomical connectivity[19] and coding properties such as spike rate, receptive field complexity, stimulus selectivity, and cell-type-specific connectivity[20]. Songbird AC responses are stronger to conspecific song than to synthetic sounds[4,21] if individuals experience adult song early in life[22], but whether learning refines auditory circuits to enhance the encoding of conspecific vocal sounds is unknown. Here, we tested the hypothesis that auditory cortical neurons are specialized to encode learned vocalizations and that tuning for vocal acoustics depends on which sounds are learned during development.

RESULTS

Juveniles learn diverse songs from social tutors

We first determined how well sibling songbirds that were reared in the same colony, with the same global exposure to song, would learn acoustically disparate songs. Some species will copy the songs of heterospecific (different species) adults if allowed to interact with them socially, even if conspecifics can be heard in the background[13,23]. The extent of learning or accuracy of song reproduction under these conditions is unclear, however. Here, all birds were housed in a single room that contained three estrildid finch species. Zebra finches (*Taeniopygia guttata*) and long-tailed finches (*Poephila acuticauda*) were transferred as eggs or nestlings (<10 d) into the nests of either conspecific or Bengalese finch (*Lonchura striata domestica*) foster parents (Fig. 1a, *left*) and allowed to mature in single-family enclosures that restricted close social interactions to foster parents and nestmates. These three species are closely related[24] but sing highly dissimilar songs (Fig. 1a, *right*). Zebra finch song (ZF) contains broadband syllables with dense (low-pitch) harmonics. Long-tailed finch song (LF) has syllables with low-density (high-pitch) harmonics that slowly sweep upward and downward in acoustic frequency. Bengalese finch song (BF) has syllables with harmonic densities intermediate to those of ZF and LF song syllables and fast downward frequency sweeps.

Every pupil learned song from its tutor regardless of whether the tutor was a conspecific or Bengalese finch or even whether the pupil's genetic father was in an adjacent cage. The rearing manipulation yielded four groups of adult subjects: two that learned conspecific song (normal: zf_{ZF}, lf_{LF}) and two that learned BF song (cross-tutored: zf_{BF}, lf_{BF}). Thus, some members of the same species developed markedly different songs (zf_{ZF} and zf_{BF}; lf_{LF} and lf_{BF}), and some members of different species developed highly similar songs (zf_{BF} and lf_{BF})

(Figs. 1, S1). The majority of pupil syllables were overt copies of tutor syllables, and groups did not differ in the proportions of their repertoire that were learned (Fig 1b; Tukey-Kramer *post hoc* tests, all $P > 0.06$; detailed statistics for all figures are provided in Table S1). Cross-tutored birds reproduced their tutor's syllables as accurately (lf_{BF}) or nearly as accurately (zf_{BF}) as did normal birds (Fig. 1c, filled boxplots; ANOVAs with bird identity as a random-effect nested covariate). Between cross-tutored bird groups (zf_{BF} and lf_{BF}), there was no difference in the number of unique syllable types produced (two-sided *t*-test, $P = 0.98$), the number of syllable types copied from a BF tutor ($P = 0.98$), or the proportion of a BF tutor's repertoire copied by pupils ($P = 0.61$). The two species did differ in how well they reproduced some acoustic features of BF songs, but it was unclear if these disparities reflected a difference in learning ability because they often depended on how closely BF songs resembled their own species' songs (Fig. S1d–g). Therefore, both zebra finches and long-tailed finches reproduced heterospecific tutor syllables with a high degree of accuracy.

Conspecific song selectivity emerges in primary auditory cortex

Vocal communicators possess neural specializations for processing the acoustic features of vocalizations. For example, regions in the human temporal lobe respond more to human vocal sounds than to environmental noises[14] or animal vocalizations[25], speech-evoked activity scales with language familiarity[15], and multi-neuron clusters respond selectively to categories of learned phonemes[16]. In nonhuman primates, regions of the auditory cortex and/or insula sparsely encode conspecific calls[26], respond more to forward than time-reversed calls[27], and respond selectively to conspecific calls compared to other animal vocalizations[28]. It remains unclear, however, where in the auditory pathway coding specializations for conspecific vocalizations emerge, how neuronal activity patterns relate to vocal acoustics (but see[29]), and what tuning mechanisms change through experience when vocalizations are learned.

To address these issues, we first determined where in the normal adult AC neurons respond selectively to conspecific song (Fig. 2). In normal adults (zf_{ZF} and lf_{LF}), we recorded responses to ZF and LF songs from 1550 single neurons in the intermediate (thalamorecipient), superficial, and deep regions of primary AC and a secondary AC region (Figs. 2a, S2, S3). We quantified song selectivity in each cell as the standardized difference in average spike rate evoked by ZF or LF syllables (Fig. 2b,c). The two species exhibited different patterns of selectivity across the cortical hierarchy. In zf_{ZF} birds, the three primary AC regions and secondary region all had higher responses, on average, to ZF than LF songs (Fig. 2d, repeated measures ANOVAs with bird identity as a covariate, orange stars). In lf_{LF} birds, intermediate-region neurons also had higher spike rates to ZF than LF songs (gray stars), while superficial-region neurons were not consistently selective for one species' songs over the other. In contrast, deep- and secondary-region neurons had higher average spike rates to LF than ZF songs. Responses were similar across individuals from each group for all brain regions (Fig. S4), and selectivity was not caused by higher spike rates to the tutor's syllable types (Fig. S5). Therefore, considering both zebra finches and long-tailed finches, spike rate selectivity for conspecific song emerged within the primary AC because only the deep and secondary regions had greater responses to conspecific songs in *both* species.

We compared three additional response metrics between species, only one indicated a specialization for conspecific song and its pattern across the AC hierarchy was correlated with spike rate. First, spike rate reliability (*i.e.* normalized variability of spike rates to the same syllable across trials) was not specialized for conspecific song because it was higher to ZF than LF syllables in both zf_{ZF} and lf_{LF} birds across all brain regions (Fig. S6a; repeated-measures ANOVA with bird identity as a covariate). The magnitude of the difference was greater in the intermediate region of zf_{ZF} than lf_{LF} birds (nested ANOVA, $P = 0.03$), however, and differences in the superficial and deep regions were nearly significant (both $P < 0.06$). Second, spike timing precision (*i.e.* spiking at the same time across trials) did not differ between ZF and LF songs for either zf_{ZF} or lf_{LF} birds in any brain region (Fig. S7a). Third, neural discrimination of songs did differ between song types within and between bird groups, and its variation across AC regions paralleled that of spike rate selectivity (Fig. S8a). In zf_{ZF} birds, single-neuron spike trains from all AC regions discriminated among ZF songs better than LF songs (repeated-measures ANOVA with bird identity as a covariate, all $P < 0.01$). In lf_{LF} birds, neurons in the intermediate region also discriminated among ZF songs better than LF songs ($P < 0.05$), but superficial- and deep-region neurons showed no consistent difference between song types, and secondary-region neurons discriminated among LF songs better than ZF songs ($P < 0.001$). Neural discrimination was strongly correlated with evoked spike rate (with bird identity as a covariate, all partial $r \geq 0.79$, all $P < 0.001$), and the difference in discrimination between song types was correlated with spike rate selectivity (Fig. S8d; 7/8 $r \geq 0.25$, $P < 0.01$).

Early learning shapes adult song selectivity

Developmental manipulations such as continuous noise or tone exposure[22,30,31] and operant training[32] can impact AC processing into adulthood[33], but the effects of early exposure to vocalizations on neuronal responses throughout the adult AC are unknown. To test whether learning conspecific or heterospecific song affected neural song selectivity, we compared the responses of neurons in normal birds to those in cross-tutored birds (zf_{ZF} versus zf_{BF}, lf_{LF} versus lf_{BF}) across brain regions (Fig. 3a,c). Neurons in all four AC regions of zf_{ZF} birds had higher spike rates to ZF songs than to BF songs (Fig. 3a, repeated-measures ANOVAs with bird identity as a covariate, orange stars). In zf_{BF} birds, responses to ZF and BF songs were similar in the superficial and secondary regions, and selectivity was significantly shifted toward BF songs in the deep region (black star, nested ANOVA, $P = 0.02$). Cross-tutoring had a similar effect in long-tailed finches. While lf_{LF} neurons in the deep and secondary regions responded more strongly to conspecific than BF songs (Fig. 3c, gray stars), lf_{BF} neurons in those regions did not. Moreover, selectivity metrics for the superficial, deep, and secondary regions were all shifted significantly toward BF songs (black stars, nested ANOVAs, all $P < 0.05$). Individuals in the same rearing group exhibited similar experience-dependent effects across brain regions (Fig. S4b,c), and selectivity was not skewed by responses to the tutor's syllable types (Fig. S5b,c). Thus, only the deep and secondary regions exhibited conspecific-versus-BF spike rate selectivity in normal birds of both species, and the deep region exhibited significant shifts in selectivity toward BF songs in cross-tutored birds of both species.

Other response properties either did not differ between bird groups or did so in a way that was highly similar to the variation in spike rate selectivity. First, spike rate reliability differed by song type but did not differ consistently between bird groups. Neurons in all brain regions of all bird groups responded more reliably to conspecific than BF syllables (Fig. S6b,c; repeated-measures ANOVAs, all $P < 0.01$), and only the responses of deep region neurons in lf_{LF} and lf_{BF} birds were different (nested ANOVA, $P = 0.0085$). Second, spike timing was more precise to ZF than BF syllables in deep- and secondary-region neurons of both zf_{ZF} and zf_{BF} birds, but it was not consistently different between LF and BF syllables in lf_{LF} and lf_{BF} birds (Fig. S7b,c). There were no differences between normal and cross-tutored groups for either species. Finally, experience-dependent effects on neural discrimination were similar to those on spike rate selectivity. Deep- and secondary-region neurons in zf_{ZF} birds discriminated among ZF songs better than BF songs (Fig. S8b; repeated-measures ANOVAs, both $P < 0.01$), but neurons in zf_{BF} birds did not. In lf_{LF} birds, neural discrimination was better among LF songs than BF songs in the secondary region. By contrast, the intermediate-, deep-, and secondary-regions of lf_{BF} birds all discriminated among BF songs better than LF songs (Fig. S8c). Group-level differences existed between the deep-region responses of normal and cross-tutored birds for both species (nested ANOVAs, both $P < 0.05$). Finally, across all bird groups, brain regions, and song types, neural discrimination performance was strongly correlated with evoked spike rate (regressions with bird identity as a covariate; all partial $r \geq 0.78$, all $P < 0.001$), and within-neuron differences in discrimination between song types was correlated with spike rate selectivity (Fig. S8e,f; 14/16 partial $r \geq 0.23$, $P < 0.05$). Together, these results suggest spike rate is a principal coding mechanism through which AC representations are specialized for learned vocalizations. Species- and experience-dependent response patterns suggest that song selectivity emerges in the deep region of primary AC and is maintained in the secondary AC.

To understand how tutoring experience shaped AC responses to song, we identified segments of ZF, LF, and BF syllables that evoked significantly different population responses (population peri-stimulus time histograms, pPSTHs) between normal and cross-tutored birds (Fig. 3b,d, deep; Figs. S9–S11). Population responses were temporally precise and aligned with specific song segments. In the deep and secondary regions, syllable segments that evoked higher responses in one bird group than the other were more likely to be from the tutor species' songs (Figs. 3b,d and S11; bar graphs to the right of pPSTHs, paired *t*-tests, 3/4 $P < 0.05$ in both regions). Responses were also highly similar across renditions of the same syllable type; the mean absolute difference in pPSTHs to different utterances of the same syllable type was consistently smaller within a bird group than between bird groups (Fig. S12; Tukey-Kramer *post hoc* tests, 15/16 $P < 0.03$ across all AC regions). Thus, neurons in normal and cross-tutored birds were driven by different syllable segments, which suggests that song learning shaped AC tuning to encode specific acoustic features of learned songs.

Neuronal tuning for song acoustics

The emergence of song selectivity in deep AC neurons predicts that the tuning mechanism underlying song selectivity also emerges in deep AC. We first tested whether a neuron's

basic frequency tuning, measured from tone-evoked receptive fields, explained vocalization selectivity as well as it does in rats[29]. Results showed that neither the frequency evoking a neuron's strongest excitatory response (best frequency) nor the range of frequencies evoking significant responses (bandwidth) predicted song selectivity (Fig. S13).

The failure of acoustic frequency tuning to explain song selectivity led us to test whether song responses could be explained by tuning for spectrotemporal modulations. Animal vocalizations often have complex acoustic structure and are composed of acoustic frequency combinations that change together over time. The spectral and temporal modulations in human speech, for example, are critical for intelligibility[34]. We first generated a set of ripples (auditory equivalent of visual gratings) that spanned the range of spectral, temporal, and joint spectrotemporal modulations in estrildid songs[34–36]. We then found the best-fit ripples for each ZF, LF, and BF syllable in the stimulus set (Fig. 4a,b) and quantified the primary spectrotemporal modulations in ZF, LF, and BF songs (Fig. 4c). ZF syllables had high-density spectral modulations (>1 cyc/kHz) and slow, downward temporal modulations (0–30 Hz); LF syllables had low-density spectral modulations (0.2–1 cyc/kHz) and slow upward and downward temporal modulations (−20 to 20 Hz); and BF syllables had wide-ranging spectral modulation densities (0.2–1.8 cyc/kHz) that swept between slow upward and fast downward rates (−10 to 50 Hz). Next, we used the differences in spectrotemporal modulations across ZF, LF, and BF songs to test whether AC neurons were tuned to song modulations. We generated ripples that differed parametrically in spectral modulation density (Fig. 4b *y*-axis) and temporal modulation rate (Fig. 4b *x*-axis) across the range of modulations that are common in ZF, LF, and BF songs. We then tested if single-neuron responses to ripples were related to their responses to songs.

In parallel with song selectivity, neurons in the intermediate region were not tuned to the modulations in tutor species' songs in any bird group (Fig. 4d,e *top*). For example, in the intermediate region of normal zf_{ZF} birds, the average normalized spike rate to the ripples most commonly found in ZF songs (*i.e.* pixels inside the black contour line from Fig. 4c) was equivalent to the spike rate evoked by ripples outside that line. In contrast, neurons in the deep region were tuned to tutor song modulations in all bird groups (Fig. 4d,e *bottom*). For example, deep-region neurons in zf_{ZF} birds had consistently higher spike rates to ripples with the high-density spectral modulations that typify ZF syllables. Comparisons between normal and cross-tutored birds of the same species showed that early tutoring experience altered modulation tuning (Fig. 4d). Cross-tutored birds (zf_{BF} and lf_{BF}) exhibited subtle but clear shifts toward ripples more common in BF song than in their respective conspecific songs (lower density ripples in zf_{BF} birds, higher density ripples in lf_{BF} birds). In the deep region, the mean tuning matrices of zf_{BF} and lf_{BF} neurons were strongly correlated ($r = 0.92$) while the relationships between birds belonging to the same species but different rearing groups were weaker (zf_{ZF}–zf_{BF} and lf_{LF}–lf_{BF}, both $r \leq 0.81$). Therefore, tuning for the spectrotemporal modulations in learned songs emerged in parallel with song selectivity in the primary AC.

Modulation tuning predicts song selectivity

We identified tuning for spectrotemporal modulations as a mechanism underlying song selectivity by linking song-evoked responses to tuning at neural population and single-neuron levels. Because each bird group exhibited distinct population response patterns to song (Fig. 3), we tested if disparities between them could be explained by the same neurons' tuning for modulations. Temporal modulation rate and spectral modulation density vectors for each syllable were aligned with pPSTHs to identify the modulation frequencies associated with divergent pPSTHs (Fig. 5a–c). Next, the modulation frequencies of those syllable segments were compared to modulation tuning curves of the same neurons (Fig. 5d–f). In both the intermediate and deep regions, the spectral modulation densities of segments that evoked divergent pPSTHs closely corresponded to each group's ripple tuning. For example, syllable segments that evoked larger pPSTHs in zf_{ZF} deep-region neurons than in lf_{LF} neurons had dense spectral modulations, and zf_{ZF} neurons also had greater responses to high-density ripples (1.2–1.8 cyc/kHz; nested ANOVAs within each spectral modulation frequency, all $P < 0.01$). In contrast, syllable segments and ripples that evoked greater responses in lf_{LF} neurons both had low-density (0.4–0.8 cyc/kHz; all $P < 0.01$) spectral modulations.

The same relationship among learned song, population responses to song, and spectral modulation tuning held across normal and cross-tutored bird groups. Syllable segments that drove larger deep-region pPSTHs in zf_{BF} than zf_{ZF} birds had lower modulation densities than those that drove larger responses in zf_{ZF} birds, and zf_{BF} neurons were tuned more strongly to low-density ripples (Fig. 5e *bottom*). Segments that evoked larger responses in lf_{LF} neurons than in lf_{BF} neurons had low spectral modulation densities, and lf_{LF} but not lf_{BF} neurons were tuned to ripples with low-density spectral modulations (Fig. 5f *bottom*). Notably, the spectral modulations associated with divergent pPSTHs in the intermediate region of zf_{ZF} and zf_{BF} birds did not match the overall difference between ZF and BF songs (Fig. 5e *top*), but song-evoked spike rates still matched tuning. Here, syllable segments that evoked greater pPSTHs in cross-tutored zf_{BF} birds than in normal zf_{ZF} birds had high-density spectral modulations, and zf_{BF} neurons had higher spike rates to high-density ripples (Fig. 5e *bottom*).

Finally, we tested whether modulation tuning explained song selectivity at the level of single neurons. Across all bird groups and brain regions, neurons that were selective for a particular species' songs were tuned to the ripples that were more prevalent in those songs (Fig. S14; regressions with bird identity as a covariate, all partial $r \geq 0.35$, all $P < 0.001$). Furthermore, neurons that were selective for the same songs had highly similar modulation tuning regardless of bird group or brain region (Fig. 6, S15). Neurons selective for ZF song were tuned to dense spectral modulations, neurons selective for LF song were tuned to low-density spectral modulations and slow frequency sweeps, and neurons selective for BF song were tuned to low/intermediate-density modulations and fast downward frequency sweeps. Thus, across the AC circuit, zebra finches and long-tailed finches both possess neurons tuned to ZF modulations and neurons tuned to LF modulations; but in the deep and secondary regions, each has a larger proportion of neurons tuned to the modulations in conspecific song. Likewise, cross-tutored birds possessed a larger proportion of neurons

tuned to BF song modulations than did normal birds (Fig. S15). These results demonstrate that tuning for spectrotemporal modulations is a mechanism for generating experience-dependent response selectivity for vocalizations in auditory cortex.

DISCUSSION

Here, we show that early vocal learning aligns auditory cortical selectivity with adult vocal behavior by tuning neurons to the spectrotemporal modulations in learned songs. This finding demonstrates how three principles of auditory coding combine to meet the sensory processing demands of learned vocal communication. First, response magnitude and temporal precision scale with behavioral significance[37–40]. In normal birds, neurons in the deep primary and secondary regions of auditory cortex responded with higher spike rates to conspecific songs than to heterospecific songs. In cross-tutored birds, neurons showed significant shifts in selectivity toward their heterospecific tutor species' songs compared to neurons in normal birds. These differences within and between groups were broad in that disparate spike rates were evoked by multiple syllable types and across multiple songs, but selectivity was also highly specific in that response variability was consistently aligned to specific syllable segments. The selective neural encoding of learned vocal features may subserve behavioral skills crucial for effective communication. For example, perceptual learning of specific sound cues facilitates their detection in noisy backgrounds[41], and higher spike rates lead to improved neural discrimination between complex natural stimuli (Fig. S8)[42] and increase information coding capacity[43,44]. Our observations in songbirds suggest that birds and mammals share the capacity for experience-dependent auditory cortical specializations that could underlie the perception of behaviorally relevant sounds, including native-language speech sounds[14–16].

The second principle is that auditory cortical neurons are tuned to the second-order spectral and temporal patterns of behaviorally relevant sounds[34,36]. The precise temporal patterning and acoustic specificity of population responses to song indicate that tuning for modulations is a fundamental mechanism underlying vocalization selectivity. Tuning for the spectrotemporal modulations in songs that birds learned to sing emerged in the deep region of primary AC, in parallel with the emergence of song selectivity. Tuning for modulations predicted syllable selectivity at single-neuron and population levels, whereas tuning for basic frequency did not. Similar observations have been made in mammals. For example, in human AC, response selectivity for phonetic categories is explained better by spectrotemporal modulations than by tone frequencies[16,36], and speech perception depends critically on spectrotemporal modulations. An important area for future research will be to identify circuit and synaptic mechanisms that create tuning for the modulations important for vocal perception.

The third principle is that social learning in the early postnatal period exerts long-lasting effects on auditory coding and perception[30,31,33,45,46]. In rodents, extreme environmental manipulations during development such as continuous noise[31] or single-tone[30,47] exposure cause significant changes to auditory cortical circuit organization. Moreover, temporary hearing loss in juveniles leads to long-term deficits in the coding and perception[48], and postnatal experience can reverse deficits that arise from early hearing loss[49]. In songbirds,

also, sensory deprivation[50] and continuous noise exposure[22] change the nature in which sounds are represented in auditory cortex. Our results add to these findings by showing that the coding properties of songbird auditory cortical neurons are aligned to the songs they learn early in life. Here, all birds had access to the same enriched acoustic environment but differed in the specific tutor with which they interacted socially. This difference in the local environment was sufficient to shape tuning toward the acoustics of their tutor species' songs. Because the effects of song learning persist into adulthood, auditory tuning for specific spectrotemporal features could impact whether birds effectively navigate social interactions throughout life. Tuning for modulations is therefore a robust neural mechanism through which experience couples auditory coding and vocal communication behavior. Similar processes could explain why early exposure to language-specific phonemes predicts adult speech perception[9,10].

METHODS

Animals.

Birds were raised in single-family enclosures in an open room. Five adult pairs of three estrildid finch species (zebra finch, *Taeniopygia guttata*; long-tailed finch, *Poephila acuticauda*; and Bengalese finch, *Lonchura striata domestica*) were given nesting materials and allowed to breed. Zebra finch and long-tailed finch eggs and nestlings (≤10 d) were transferred to nests of conspecifics or Bengalese finches (whose eggs were removed; each clutch was a single species) and raised by foster parents to maturity (zebra finches, ≥120 d; long-tailed finches, ≥180 d). At maturity, each cohort was moved to its own cage in the same room. Birds could hear vocalizations of all three species and see birds in other enclosures at a distance (>1 m). To record song, a tutor or pupil was temporarily moved to a sound-attenuating booth (Industrial Acoustics MAC-1) outfit with recording equipment (Sennheiser MKE2 microphone, Focusrite Saffire Pro 40 audio interface, Sound Analysis Pro software). All procedures were approved by the Columbia University Institute for Animal Care and Use Committee.

Song Analysis.

Undirected songs of each tutor and adult pupil ($n \geq 20$ bouts per bird, ~6–10 s each) were analyzed to compare learning accuracy in normal (zf_{ZF}, $n = 25$; lf_{LF}, $n = 12$; bf_{BF}, $n = 3$) and cross-tutored birds (zf_{BF}, $n = 12$; lf_{BF}, $n = 10$). Audio files were bandpass-filtered (300 Hz to 8 kHz), and syllable boundaries were defined with an amplitude threshold applied to the log-transformed envelope (Fig. 2b). Syllable type labels were assigned from visual inspection of the spectrograms and verified by comparing the similarity of syllables within and across types.

The similarity between different syllable renditions (both pupil self-comparisons and tutor-pupil copies) was quantified by cross-correlating the spectrograms of up to 20 (when possible, otherwise no fewer than 5) randomly selected examples of each syllable type. First, multi-taper spectrograms (log-transformed) were computed for each syllable with 2-ms temporal resolution and 100-Hz spectral resolution (Sound Analysis Tools for Matlab, http://soundanalysispro.com/matlab-sat). Second, spectrograms were re-scaled to range from 0 to

1 and pixel values below 0.5 were set to 0. Third, spectrograms for every pair of syllables were cross-correlated: the shorter syllable was convolved with the longer syllable in 2-ms steps and the peak correlation coefficient was used as a similarity metric. This method outperformed other common approaches (*e.g.*, acoustic feature correlation) in its ability to discriminate between syllables of the same versus different type [contrast index: $(r_{same} - r_{diff}) / (r_{same} + r_{diff})$][51]. Finally, several acoustic features were measured from each syllable, including pitch, mean frequency, frequency modulation, and Wiener entropy. To estimate how well each feature was learned, each vector was averaged over the duration of a syllable and the mean values of copied syllable types were correlated between tutors and pupils (Fig. S1d–g).

Stimuli.

Auditory stimuli were short bouts of each tutor's song (2–4 s each from five zebra finches, ZF; long-tailed finches, LF; and Bengalese finches, BF), ripples (1 s), and pure tones (200 ms). They were delivered through a speaker (JBL Control 1 Pro) with a flat frequency response (±5 dB) placed 20 cm in front of the bird. Songs and ripples were bandpass-filtered between 0.3 and 8 kHz, root-mean-square power-matched, and delivered at 60 dB SPL. Ripples were generated with custom software (S. Andoni, Univ. of Texas) to cover temporal modulation rates (−50 to 50 Hz in 10 Hz steps) and spectral modulation densities (0–1.9 cyc/kHz in 0.2 or 0.3 cyc/kHz steps) in estrildid finch songs (Fig. 4a–c). Tones ranged from 0.5–8 kHz in 500 Hz steps and spanned 10–90 dB SPL in 10 dB steps. Ripples and tones had 10 ms sinusoidal onset and offset amplitude ramps. Stimulus order was pseudorandom over 10 trial blocks, and inter-stimulus intervals were drawn randomly from a range of 500–750 ms.

The primary spectrotemporal modulation frequencies in song syllables were measured by tracing adjacent harmonics in spectrograms (1 ms, 50 Hz bins), calculating their separation in frequency, and computing their average rate of change. Each time bin was assigned a temporal modulation rate (Hz) and spectral modulation density (cyc/kHz) rounded to the resolution of ripple stimuli (Fig. 4c). Syllable segments without clear harmonics were excluded. The spectrotemporal composition of each song was computed as the fraction of total syllable duration matching each ripple, and contours were fit to the predominant modulation frequencies comprising 90% of the total duration of all syllables. For clarity, heatmaps showing the relative ripple proportions were log-scaled (Fig. 4c).

Electrophysiology.

Three adult males of each pupil group were used in electrophysiology experiments [zf_{ZF} (295, 696, 791 d); zf_{BF} (322, 400, 712 d); lf_{LF} (224, 258, 359 d); and lf_{BF} (423, 757, 1105 d)]. Each bird in a pupil group had a different tutor, and the three cross-tutored birds of each species learned the same three BF songs. Two days before their first recording sessions, birds underwent a preparatory surgery. They were anesthetized with an intramuscular injection of Equithesin (0.0025 mL/g), placed in a custom stereotaxic holder, and given bilateral craniotomies (~2 mm^2) centered 1 mm lateral of the midline and 1 mm rostral of the bifurcation of the midsagittal sinus. Ink markings were made at known coordinates along the edges of the craniotomies and used as a reference for electrode placement during

experiments. Then, a copper ground wire was inserted between the skull and cerebellum, a small metal post was set atop the skull, and both objects were affixed to the skull with dental cement. Finally, the dura was retracted, and the craniotomies were sealed with silicone between recording sessions.

Recordings were made from awake, restrained birds using 16-channel electrode arrays (NeuroNexus Technologies, A4×4, 177 μm^2 site area). Birds were wrapped in a custom cloth jacket, placed in a sound-attenuating chamber (ETS-Lindgren) lined with anechoic foam (SonexOne), and their heads were immobilized in a stereotaxic device with the beak pointing downward by 45 degrees. Typically, a single electrode penetration was made per day and responses were recorded at three to five non-overlapping depths (1.0–2.8 mm beneath the dorsal surface); recording sessions lasted approximately 6 h. Usually, five penetrations that were separated mediolaterally by 300 μm were made in each hemisphere. Prior to each electrode pass, the shanks were painted along the back with either CM-DiI (C7000, Molecular Probes) or DiO (D275) dissolved in 100% ethanol to reconstruct electrode locations (Fig. S2). Continuous voltage signals were amplified, bandpass filtered (300 to 5000 Hz), digitized at 25 kHz (Tucker-Davis Technologies, RZ5), and stored for analysis offline. A total of 4392 single neurons were recorded from three adult males in each pupil group (normal, zf_{ZF} and lf_{LF}; cross-tutored, zf_{BF} and lf_{BF}).

Histology.

One day after the final recording session, birds were injected with Equithesin and perfused transcardially with 0.9% saline followed by 10% formalin in saline. Brains were extracted, postfixed for 24 h, cryoprotected in 30% sucrose in formalin for 48 h, embedded in gelatin, and stored in cryoprotectant for >5 d. Brains were then sectioned at 40 μm thickness in the sagittal plane on a freezing microtome, and sections were mounted onto gelatin-coated slides. Immediately after mounting, wet sections were imaged with a fluorescent microscope to view DiI- and DiO-labeled tracks. Background images were taken using a blue filter and brightfield backlight to view myelin-rich structures that served as landmarks, particularly the mesopallial and pallio-subpallial laminae and subregion L2a. Sections were then left to dry for >3 d, stained with cresyl violet and coverslipped, and imaged again in the brightfield.

Recording-site depths were reconstructed along fluorescently labeled tracks, and electrode positions with respect to AC subregions were assigned by viewing the reconstructed array on Nissl-stained images of the same sections. Cortical regions were delineated on the basis of cytoarchitecture and thalamic fiber terminations (Fig. S2)[52]. Intermediate subregions L2a and L2b are the primary thalamorecipient areas and contain small, densely packed cell bodies and heavy myelination. The superficial region L1 is dorso-rostral to L2 and ventral to the mesopallial lamina, and the caudal mesopallium (CM) is bounded ventrally by the mesopallial lamina and dorsally by the lateral ventricle. The deep region L3 is ventral to L2 and characterized by large, sparsely packed cells; region L is caudal to L3 and ventral to L2b and contains small, densely packed cells. The secondary region caudal nidopallium (NC) is caudal to L2b and L.

Electrophysiological Data Analysis.

Spikes were detected and sorted using an automated clustering algorithm and user-controlled interface (WaveClus)[53] that performs well compared to other commonly used programs[54]. First, continuous voltage traces from each channel were transformed with a nonlinear filter that selectively enhanced large-amplitude, high-frequency deflections; this procedure improved detection of small action potentials (Fig. S3a). Second, instances where the transformed signal exceeded a noise threshold were detected, corresponding snippets of the original voltage trace were stored, and spike waveforms were grouped using an unsupervised clustering algorithm. Third, clusters were refined manually based on spike waveform shape and magnitude (Fig. S3b,c). Fourth, single units were identified as those with: stable isolation throughout all trials (Fig. S3d), a large signal-to-noise ratio (standard separation D: difference between mean spike and baseline amplitudes divided by the geometric mean of their SDs; Fig. S3e), spike magnitude variance approximately equal to or exceeding baseline variance (Fig. S3f), and inter-spike interval distributions with few refractory period violations (Fig. S3g). Fifth, single units were classified as fast-spiking (n = 1382) or regular-spiking (n = 3010) by fitting a mixture-of-Gaussians model to a bimodal distribution of spike width (measured at half-height of each major voltage deflection, Fig. S3a). However, song selectivity and tuning were both similar between unit types and the data were combined. Trials contaminated by movement artifacts were identified by large perturbations in the raw voltage traces and excluded from all analyses.

Units were used in analyses only if they surpassed minimum response criteria for all relevant stimuli. Each unit's spontaneous spike rate was measured in the 200 ms preceding every trial, and stimulus-evoked responses exceeding the mean spontaneous rate by at least 2 SDs were considered significant. For songs, spike rates were quantified on a syllable-by-syllable basis to control for species differences in the proportion of songs that were sound versus silence. The number of spikes occurring within windows spanning 10 ms after syllable onset to 40 ms after syllable offset (or to the subsequent syllable onset, if shorter) was divided by the product of window duration and number of trials. Asymmetric windows were used to accommodate variation in response latencies and temporal profiles across bird groups (zebra finches had ~5 ms shorter latencies than long-tailed finches), brain regions (intermediate-region neurons had 5–20 ms shorter latencies than deep- and secondary-region neurons), and in some cases within neurons (responses could vary between phasic/sustained or onset/offset depending on syllable features). A unit was considered responsive to a song type (*i.e.* ZF, LF, or BF) if it had a significantly elevated spike rate to ≥5% of those syllables. The same procedure was used to identify units driven by ripples.

For tones, spike rate alone was an unreliable indicator of unit responsiveness, therefore we devised a metric that weighted spike time precision. Spike times from each trial were binned at 1 ms, smoothed with a 10-ms Hanning kernel, and averaged to yield a peri-stimulus time histogram (PSTH) for each frequency-level combination. Response magnitude was computed as the square root of the product of the mean and peak PSTH value. The same metric was computed for each pre-stimulus time window (to measure spontaneous spike rate), and evoked values that were greater than the mean spontaneous value by at least 2 SDs were considered significant. Units responding to ≥2% of tones were considered responsive.

Not all units responded significantly to all song types or stimulus types, therefore sample sizes vary slightly between some analyses.

Song selectivity for each unit was quantified as the difference in mean spike rate evoked by two song types standardized by their unpooled variance,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

Where $\bar{x}$ is the mean syllable-evoked spike rate for species 1 and 2, s^2 is the respective distribution variance, and n is the respective number of syllables. The statistical significance of within-neuron differences in average spike rate to two song types (*e.g.*, colored stars in Figs. 2d; 3a,c) was tested using repeated-measures ANOVAs with bird identity as a covariate. Comparisons of selectivity (t-statistics) between two groups (*e.g.*, black stars in Figs. 2d; 3a,c) used nested ANOVAs with bird identity random-effect covariate. All tests were two-sided.

To analyze temporal response dynamics, population peri-stimulus time histograms (pPSTHs) were created for each pupil group/brain region combination: single-unit PSTH vectors were created as described above, normalized (z-scored) across all songs relevant to a particular comparison, and then averaged across all song-responsive units. The pPSTHs were aligned with spectrograms according to each region's latency, which was measured to each song as the first pPSTH sample after the onset of the first syllable to exceed 5 SDs of the mean spontaneous pPSTH (200 ms preceding each song). The mean of the five shortest delays was used as the latency for each pupil group/brain region combination. Response similarity within and between groups was measured by computing the mean absolute difference between pPSTHs to different renditions of the same syllable type (Fig. S12). Finally, sustained differences between the pPSTHs of different pupil groups were identified as periods in which the pPSTHs were significantly different (two-sided t-tests) for >80% of the samples in a ≥10 ms segment.

Several response properties in addition to spike rate were analyzed. First, response reliability across trials was characterized by the Fano factor (*i.e.* coefficient of variation for syllable-evoked spike rates). Smaller values indicate greater reliability, so differences in reliability between pupil groups was calculated as the difference between–CVs (Fig. S6). Second, spike timing precision was quantified by the correlation index (CI), a metric based on shuffled autocorrelograms that measures the tendency for spikes to occur at the same time to repeated presentations of the same stimulus (Fig. S7)[55]. Briefly, for a given syllable, inter-spike intervals were measured between each spike in a trial and all spikes that occurred coincidentally or later in all other trials. Larger values reflected greater temporal precision. Finally, neural discrimination measured the tendency for spike trains to a given song stimulus, over multiple trials, to resemble one another more closely than spike trains to the other songs of the same species (Fig. S8). Because this study used only five songs per

species, performance was assessed using a method that was not constrained by ceiling effects (*d*-prime)[42].

Ripple-evoked responses were organized into two-dimensional modulation response area (MRA) plots with temporal modulation rate (Hz) varying along the *x*-axis and spectral modulation density (cyc/kHz) along the *y*-axis. To describe the general tuning pattern of a brain region, individual MRAs were normalized (*z*-scored spike rates) and then averaged. Spectral modulation tuning curves were computed as the mean response to each spectral modulation density across all temporal modulation rates.

To relate song selectivity to modulation tuning in single units, we computed an index that reflected how well MRAs overlapped with disparities in two species' ripple composition spectra (Fig. S14). First, a single neuron's MRA was normalized (*z*-scored; higher-than-average spike rates were positive and lower-than-average spike rates were negative). Second, the difference between ripple spectra of two species' songs was computed; pixels of ripples common in only the first song type were positive, pixels of ripples common in only the second song type were negative, and pixels of uncommon or equally common ripples were near zero. Third, the two matrices were multiplied elementwise (*i.e.* Hadamard product) and the resulting values were added. This index separated units based on the shape of their MRA relative to the ripples in two species' songs: units tuned to ripples mostly in the first song type had a positive value, units tuned to ripples mostly in the second song type were negative, and units with weak tuning or tuning for ripples in both song types had values near zero. Finally, the index was correlated with spike rate selectivity.

Tone-evoked responses were assessed by response strength (evoked – spontaneous PSTH products; described above) and organized into two-dimensional frequency response area (FRA) plots with frequency varying along the *x*-axis and sound level (dB) along the *y*-axis (Fig. S13). A frequency response curve was calculated as the mean response to each frequency across levels and used to calculate two basic tuning features: best frequency (Bf) was the frequency eliciting the maximum response, and bandwidth (Bw) was the width of the response curve measured at half-height.

Randomization.

Eggs were transferred to different tutors' nests at random. Pupils were chosen for electrophysiology to maximize diversity of tutor song learning within groups [each bird within a group (*e.g.*, zf_{ZF}) had a different tutor] and maximize consistency among cross-tutored pupils (*i.e.* the three zf_{BF} birds and three lf_{BF} birds learned the same three songs). Electrode penetrations were spaced evenly throughout the caudal telencephalon to record from as many AC subregions as possible in each bird. Stimulus presentation within each experiment was varied pseudorandomly across ten blocks.

Statistics.

When possible, statistical tests accounted for the non-independence of song learning and neuronal data by including bird identity as a covariate. Within-neuron differences in responses to different songs (*e.g.*, Fig. 2d, 3a, 3c) or ripples (Fig. 4e) were assessed with repeated-measures ANOVAs that included bird identity as a covariate; between-group

differences in syllable learning accuracy (Fig. 1c), neural song selectivity (*e.g.*, Fig. 2d, 3a, 3c), and modulation tuning (*e.g.*, Fig. 5d–f) were assessed with hierarchical ANOVAs that included bird identity as a random-effect, nested covariate; and regressions between metrics included bird identity as a categorical covariate (*e.g.*, Figs. S5, S8, S13, S14). Comparisons of pPSTHs used paired *t*-tests (*e.g.*, Fig. 3b,d). Modulation tuning maps were correlated using Pearson correlations (*e.g.*, Fig. 4d), and their significance was determined with permutation tests to account for autocorrelation. Over 1000 iterations, each matrix was randomly permuted, smoothed to approximate the autocorrelation structure of the original maps, and then correlated. Correlation coefficients of the original maps were then compared to this distribution to estimate a *p*-value. Detailed statistical results for all figures are provided in Table S1.

Data collection was not performed blind to the conditions of the experiment, but all spike sorting and determinations of unit inclusion were made without knowledge of the stimulus evoking particular responses or the brain region in which a unit was located. All animals were used in all relevant analyses; units were included in relevant analyses if they responded above a minimum activity threshold (see above). Responses from the same units were analyzed in multiple ways (*e.g.*, spike rate selectivity and pPSTHs to songs), and responses of the same units to different stimulus types were related directly (*e.g.*, Fig. S14). Data distributions were inspected visually and assumed to be normal, but this was not formally tested. No statistical methods were used to pre-determine sample sizes. All tests were two-sided unless otherwise noted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank J. Schumacher, A. Calabrese, D. Schneider, H. Brew, S. Rosis, and N. So for suggestions on design and analysis. We are grateful to N. Mesgarani, M. Long, N. So, and E. Perez for comments on previous versions of the manuscript. Funding was provided by NIH grant DC009810 (SW) and NSF grant IOS-1656825 (SW).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

1. Bradbury JW & Vehrencamp SL Principles of Animal Communication. 2 ed, (Sinauer Associates, 2011).
2. Ord TJ & Stamps JA Species identity cues in animal communication. Am. Nat 174, 585–593 (2009). [PubMed: 19691435]
3. Dooling RJ, Brown SD, Klump GM & Okanoya K Auditory perception of conspecific and heterospecific vocalizations in birds: evidence for special processes. J. Comp. Psychol 106, 20–28 (1992). [PubMed: 1555398]
4. Woolley SMN, Fremouw TE, Hsu A & Theunissen FE Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat. Neurosci 8, 1371–1379 (2005). [PubMed: 16136039]
5. Saffran JR, Werker JF & Werner LA in Handbook fo Child Development (eds Seigler R & Kuhn D) Ch. 2, 58–108 (Wiley, 2006).

6. Poremba A, Bigelow J & Rossi B Processing of communication sounds: contributions of learning, memory, and experience. Hearing Res 305, 31–44 (2013).
7. Doupe AJ & Kuhl PK Birdsong and human speech: common themes and mechanisms. Annu. Rev. Neurosci 22, 567–631 (1999). [PubMed: 10202549]
8. Konishi M The role of auditory feedback in birdsong. Ann. NY Acad. Sci 1016, 463–475 (2004). [PubMed: 15313790]
9. Werker JF & Tees RC Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev 7, 49–63 (1984).
10. Kuhl PK, Williams KA, Lacerda F, Stevens KN & Lindblom B Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608 (1992). [PubMed: 1736364]
11. Johnson JS & Newport EL Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. Cogn. Psychol 21, 60–99 (1989). [PubMed: 2920538]
12. Riebel K Song and female mate choice in zebra finches: a review. Adv. Study Behav 40, 197–238 (2009).
13. Immelmann K in Bird Vocalizations (ed Hinde RA) 61–74 (Cambridge Univ. Press, 1969).
14. Belin P, Zatorre RJ, Lafaille P, Ahad P & Pike B Voice-selective areas in human auditory cortex. Nature 403, 309–312 (2000). [PubMed: 10659849]
15. Näätänen R et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434 (1997). [PubMed: 9009189]
16. Mesgarani N, Cheung C, Johnson K & Chang EF Phonetic feature encoding in human superior temporal gyrus. Science 343, 1006–1010 (2014). [PubMed: 24482117]
17. Winkler I et al. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36, 638–642 (1999). [PubMed: 10442032]
18. Butler AB, Reiner A & Karten HJ Evolution of the amniote pallium and the origins of mammalian neocortex. Ann. N.Y. Acad. Sci 1225, 14–27 (2011). [PubMed: 21534989]
19. Wang Y, Brzozowska-Prechtl A & Karten HJ Laminar and columnar auditory cortex in avian brain. Proc. Natl Acad. Sci. USA 107, 12676–12681 (2010). [PubMed: 20616034]
20. Calabrese A & Woolley SMN Coding principles of the canonical cortical microcircuit in the avian brain. Proc. Natl Acad. Sci. USA 112, 3517–3522 (2015). [PubMed: 25691736]
21. Grace JA, Amin N, Singh NC & Theunissen FE Selectivity for conspecific song in the zebra finch auditory forebrain. J. Neurophysiol 89, 472–487 (2003). [PubMed: 12522195]
22. Amin N, Gastpar M & Theunissen FE Selective and efficient neural coding of communication signals depends on early acoustic and social environment. PLoS ONE 8, e61417 (2013). [PubMed: 23630587]
23. Eales LA Do zebra finch males that have been raised by another species still tend to select a conspecific song tutor? Anim. Behav 35, 1347–1355 (1987).
24. Gomes ACR, Sorenson MD & Cardoso GC Speciation is associated with changing ornamentation rather than stronger sexual selection. Evolution 70, 2823–2838 (2016). [PubMed: 27718251]
25. Fecteau S, Armony JL, Joanette Y & Belin P Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage 23, 840–848 (2004). [PubMed: 15528084]
26. Remedios R, Logothetis NK & Kayser C An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J. Neurosci 29, 1034–1045 (2009). [PubMed: 19176812]
27. Wang X, Merzenich MM, Beitel R & Schreiner CE Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J. Neurophysiol 74, 2685–2706 (1995). [PubMed: 8747224]
28. Perrodin C, Kayser C, Logothetis NK & Petkov CI Voice cells in the primate temporal lobe. Curr. Biol 21, 1408–1415 (2011). [PubMed: 21835625]
29. Carruthers IM, Natan RG & Geffen MN Encoding of ultrasonic vocalizations in the auditory cortex. J. Neurophysiol 109, 1912–1927 (2013). [PubMed: 23324323]

30. Zhang LI, Bao S & Merzenich MM Persistent and specific influences of early acoustic environments on primary auditory cortex. Nat. Neurosci 4, 1123–1130 (2001). [PubMed: 11687817]
31. Chang EF & Merzenich MM Environmental noise retards auditory cortical development. Science 300, 498–502 (2003). [PubMed: 12702879]
32. Sarro EC & Sanes DH The cost and benefit of juvenile training on adult perceptual skill. J. Neurosci 31, 5383–5391 (2011). [PubMed: 21471373]
33. Sanes DH & Woolley SMN A behavioral framework to guide research on central auditory development and plasticity. Neuron 72, 912–929 (2011). [PubMed: 22196328]
34. Chi T, Ru P & Shamma SA Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am 118, 887–906 (2005). [PubMed: 16158645]
35. Shamma SA, H. V & N. K Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. Aud. Neurosci 1, 233–254 (1995).
36. Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE & Chang EF Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J. Neurosci 36, 2014–2026 (2016). [PubMed: 26865624]
37. Bizley JK, Walker KM, Nodal FR, King AJ & Schnupp JW Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr. Biol 23, 620–625 (2013). [PubMed: 23523247]
38. Fukushima M, Saunders RC, Leopold DA, Mishkin M & Averbeck BB Differential coding of conspecific vocalizations in the ventral auditory cortical stream. J. Neurosci 34, 4665–4676 (2014). [PubMed: 24672012]
39. Harris KD & Thiele A Cortical state and attention. Nat. Rev. Neurosci 12, 509–523 (2011). [PubMed: 21829219]
40. Kato HK, Gillet SN & Isaacson JS Flexible sensory representations in auditory cortex driven by behavioral relevance. Neuron 88, 1027–1039 (2015). [PubMed: 26586181]
41. Caras ML & Sanes DH Top-down modulation of sensory cortex gates perceptual learning. Proc Natl Acad Sci USA 114, 9972–9977 (2017). [PubMed: 28847938]
42. Schneider DM & Woolley SMN Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. J. Neurophysiol 103, 3248–3265 (2010). [PubMed: 20357062]
43. Mesgarani N, David SV, Fritz JB & Shamma SA Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol 102, 3329–3339 (2009). [PubMed: 19759321]
44. Sun W & Barbour DL Rate, not selectivity, determines neuronal population coding accuracy in auditory cortex. PLoS Biol. 15, e2002459 (2017). [PubMed: 29091725]
45. Razak KA, Richardson MD & Fuzessery ZM Experience is required for the maintenance and refinement of FM sweep selectivity in the developing auditory cortex. Proc. Natl Acad. Sci. USA 105, 4465–4470 (2008). [PubMed: 18334643]
46. Schreiner CE & Polley DB Auditory map plasticity: diversity in causes and consequences. Curr. Opin. Neurobiol 24, 143–156 (2014). [PubMed: 24492090]
47. Han YK, Köver H, Insanally MN, Semerdjian JH & Bao S Early experience impairs perceptual discrimination. Nat. Neurosci 10, 1191–1197 (2007). [PubMed: 17660815]
48. Caras ML & Sanes DH Sustained perceptual deficits from transient sensory deprivation. J. Neurosci 35, 10831–10842 (2015). [PubMed: 26224865]
49. Green DB, Mattingly MM, Ye Y, Gay JD & Rosen MJ Brief stimulus exposure fully remediates temporal processing deficits induced by early hearing loss. J. Neurosci 37, 7759–7771 (2017). [PubMed: 28706081]
50. Cousillas H et al. Experience-dependent neuronal specialization and functional organization in the central auditory area of a songbird. Eur. J. Neurosci 19, 3343–3352 (2004). [PubMed: 15217389]
51. Mandelblat-Cerf Y & Fee MS An automated procedure for evaluating song imitation. PLoS ONE 9, e96484 (2014). [PubMed: 24809510]

52. Fortune ES & 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). [PubMed: 1447407]

53. Quiroga RQ, Nadasdy Z & Ben-Shaul Y Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural. Comput 16, 1661–1687 (2004). [PubMed: 15228749]

54. Wild J, Prekopcsak Z, Sieger T, Novak D & Jech R Performance comparison of extracellular spike sorting algorithms for single-channel recordings. J. Neurosci. Methods 203, 369–376 (2012). [PubMed: 22037595]

55. Joris PX, Louage DH, Cardoen L & van der Heijden M Correlation index: a new metric to quantify temporal coding. Hear. Res 216–217, 19–30 (2006).

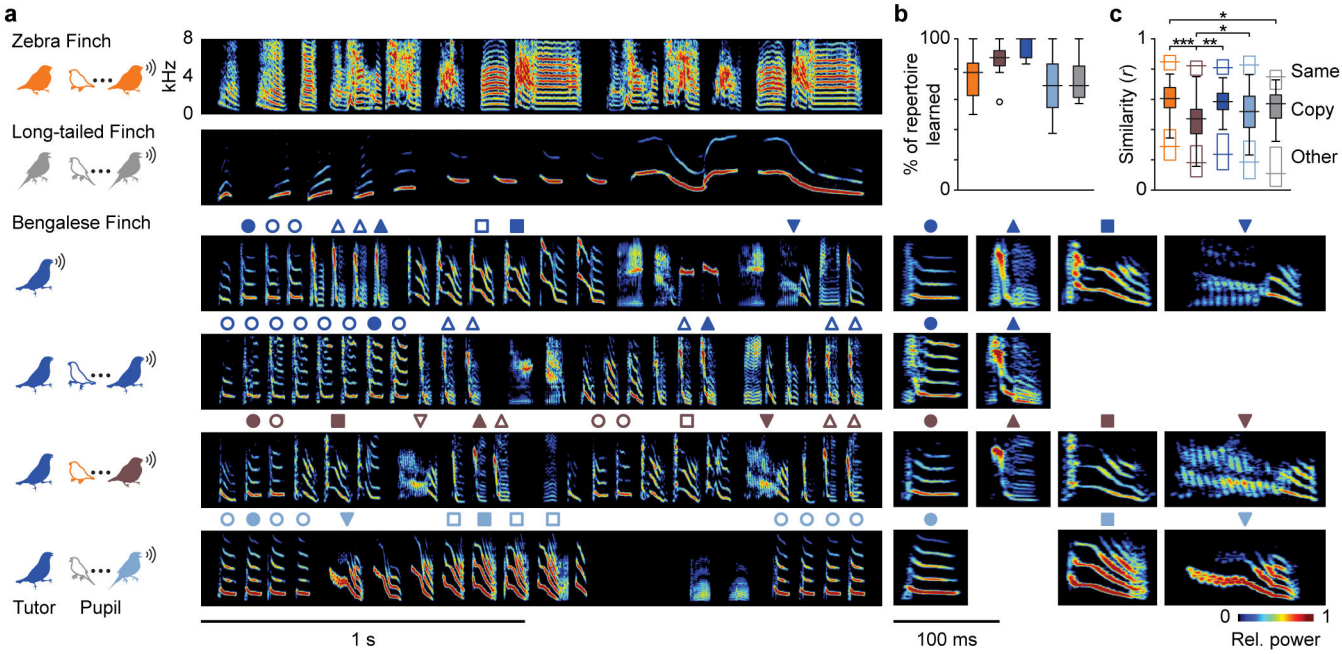

Fig. 1.

Juvenile songbirds learn song from conspecific or heterospecific tutors. **a**, Spectrograms of song segments from an adult zebra finch and adult long-tailed finch tutored by conspecifics (zf_{ZF} and lf_{LF}, respectively; tutor songs in Fig. 2), a Bengalese finch tutor (BF), and its adult Bengalese finch (bf_{BF}), zebra finch (zf_{BF}, brown) and long-tailed finch (lf_{BF}, light blue) pupils. Symbols denote BF syllable types and corresponding pupil copies, with high-magnification spectrograms of examples shown to the right. **b**, Both normal and cross-tutored birds learned most of their syllable repertoire from their tutor [$n = 25$ (zf_{ZF}), 11 (zf_{BF}), 3 (bf_{BF}), 10 (lf_{BF}), and 12 (lf_{LF}) birds; Tukey-Kramer *post hoc* tests, all $P > 0.06$]. **c**, Pupils in all groups reproduced their tutor's syllables accurately (filled box-and-whisker plots), though zf_{BF} birds produced syllables that were less similar to their tutors' than did zf_{ZF}, bf_{BF}, or lf_{BF} pupils [$n = 192$ (zf_{ZF}), 113 (zf_{BF}), 32 (bf_{BF}), 97 (lf_{BF}), and 74 (lf_{LF}) syllable types; ANOVAs used bird identity as a nested covariate, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$]. For reference, open boxplots along the top show similarity between different renditions of the same syllable type within pupils [$n = 192$ (zf_{ZF}), 113 (zf_{BF}), 32 (bf_{BF}), 97 (lf_{BF}), and 74 (lf_{LF}) syllable types], and those along the bottom show similarity between different syllable types of pupils and tutors [$n = 1287$ (zf_{ZF}), 1733 (zf_{BF}), 488 (bf_{BF}), 1210 (lf_{BF}), and 424 (lf_{LF}) comparisons]. For **b** and **c**, the measure of center is the median, box limits show the 25th and 75th percentiles, whiskers extend up to 1.5× the interquartile range beyond the quartiles; and circles show outliers.

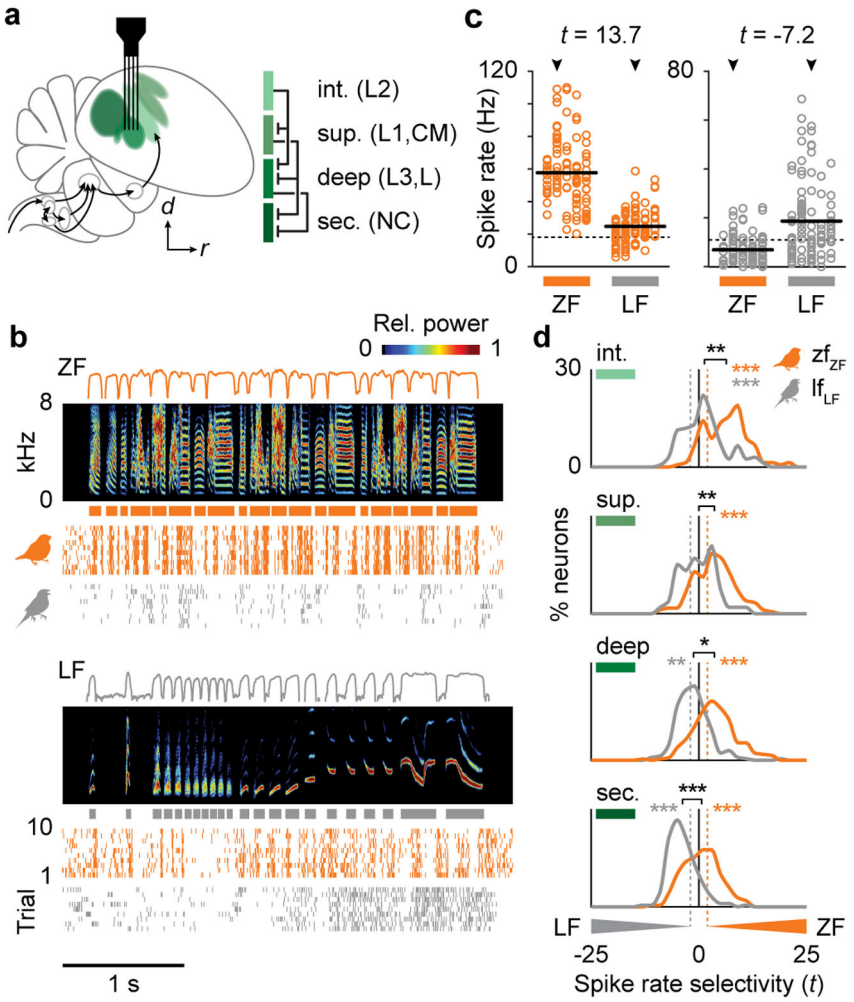

Fig. 2.

Selectivity for conspecific song emerges in primary auditory cortex. **a**, Schematic of the songbird auditory system in which shades of green indicate cortical region (intermediate, superficial, deep, secondary) and lines show major projections between them. **b**, Spike rasters show song-evoked responses of a single neuron from a zf_{ZF} (orange, deep region) and a lf_{LF} (gray, secondary region) bird to ZF (top) and LF (bottom) songs. Lines above spectrograms show log-transformed amplitude envelopes used to delineate syllable boundaries (indicated by boxes below spectrograms), and rows in each raster show the spike times during an individual trial. **c**, Spike rates of the same two neurons shown in **b** to ZF versus LF syllables with responses to different songs organized in columns (arrows indicate the songs shown in **b**). Circles show the mean spike rates to each syllable (n = 13, 20, 17, 22, 22 syllables in ZF songs; n = 21, 33, 22, 13, 14 syllables in LF songs), solid black lines show the mean spike rates across all syllables per species, and dotted lines show spontaneous spike rates. Selectivity was computed as the difference in mean spike rate to the syllables of two species divided by their unpooled variance (t-statistic). **d**, Distributions of spike rate selectivity for ZF versus LF songs in zf_{ZF} (orange) and lf_{LF} (gray) neurons in each AC region. All regions in zf_{ZF} birds had, on average, higher spike rates to ZF song (n = 148,

179, 281, 217 neurons per intermediate, superficial, deep, and secondary regions, respectively). In lf_{LF} birds, intermediate-region neurons also had greater responses to ZF song, but the deep and secondary regions had greater responses to LF song (n = 168, 53, 237, 199 neurons per region). Thus, only the deep and secondary regions were selective for conspecific song in both species. Colored stars indicate a significant difference between song types within a group (repeated-measures ANOVAs with bird identity as a covariate) and are plotted on the side of the song that evoked a greater response. Black bars show the separation between distribution means, and black stars indicate a difference in selectivity between bird groups (ANOVAs with bird identity as a nested covariate). Dashed lines indicate the criteria for selectivity in single neurons (t = ±1.96).
*P < 0.05, **P < 0.01, ***P < 0.001.

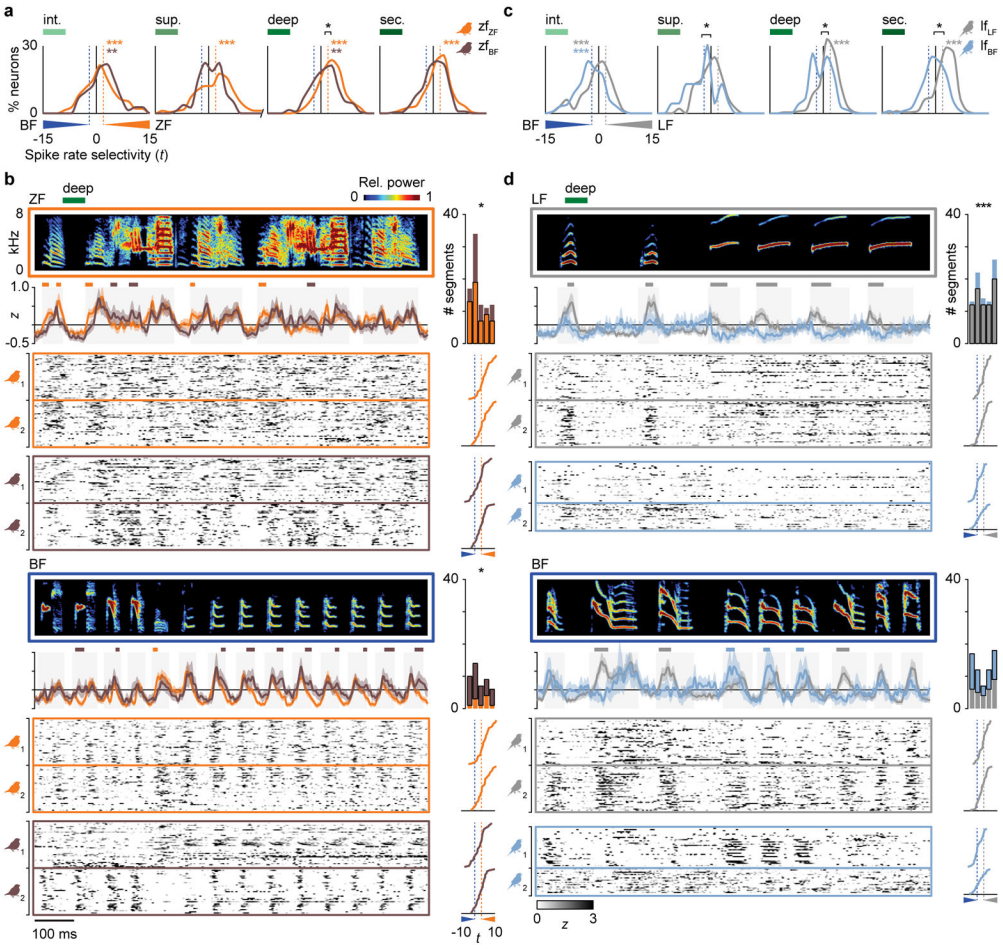

Fig. 3.

Song selectivity and population response dynamics are experience-dependent. **a**, Distributions of selectivity for ZF versus BF songs in zf_{ZF} birds (orange; n = 149, 181, 270, 197 neurons per AC region) and zf_{BF} birds (brown; n = 73, 80, 136, 250 neurons per region). In zf_{ZF} birds, all regions had higher spike rates to ZF songs. In zf_{BF} birds, the superficial and secondary regions exhibited no difference between songs, and selectivity in the deep region was shifted toward BF songs compared to normal birds. Colored stars indicate a significant response difference between song types within neurons (repeated-measures ANOVAs with bird identity as a covariate), black bars show the separation between distribution means, and black stars indicate a difference in selectivity between groups (ANOVAs with bird identity as a nested covariate). **b**, Spectrograms (0–8 kHz) of ZF (top) and BF (bottom) song segments plotted above deep-region pPSTHs (mean ± 95% C.I.) and neurograms (z-scored single-neuron PSTHs) from two birds in each group (n = 40 randomly selected neurons per bird). Colored lines above pPSTHs indicate sustained differences (≥10 ms) between groups, and bar graphs to the right show the number of segments in each ZF or BF stimulus that evoked a greater pPSTH in zf_{ZF} (orange) or zf_{BF} (brown) birds (two-sided paired t-tests, n = 5 songs for each species). Traces to the right of neurograms show the selectivity of each respective neuron (dashed lines are t = ±1.96). **c**, Same as in **a**, but showing distributions of spike rate selectivity for LF versus BF songs in lf_{LF} birds (gray, n = 164, 48, 224, 208

neurons per AC region) and lf_{BF} birds (light blue, n = 133, 23, 75, 190 neurons per region). Both groups had greater responses to BF song in the intermediate region, but only lf_{LF} birds had greater responses to LF songs in the deep and secondary regions. **d**, Same as in **b**, but showing spectrograms of LF and BF songs and deep-region pPSTHs and randomly selected neurograms from lf_{LF} birds (n = 40 neurons per bird) and lf_{BF} birds (n = 35 and 23 neurons per bird). For **b** and **d**, pPSTHs and neurograms were shifted in time by the average response latency of paired groups (zf_{ZF} and zf_{BF}, 15 ms; lf_{LF} and lf_{BF}, 22 ms).
*P < 0.05, **P < 0.01, ***P < 0.001.

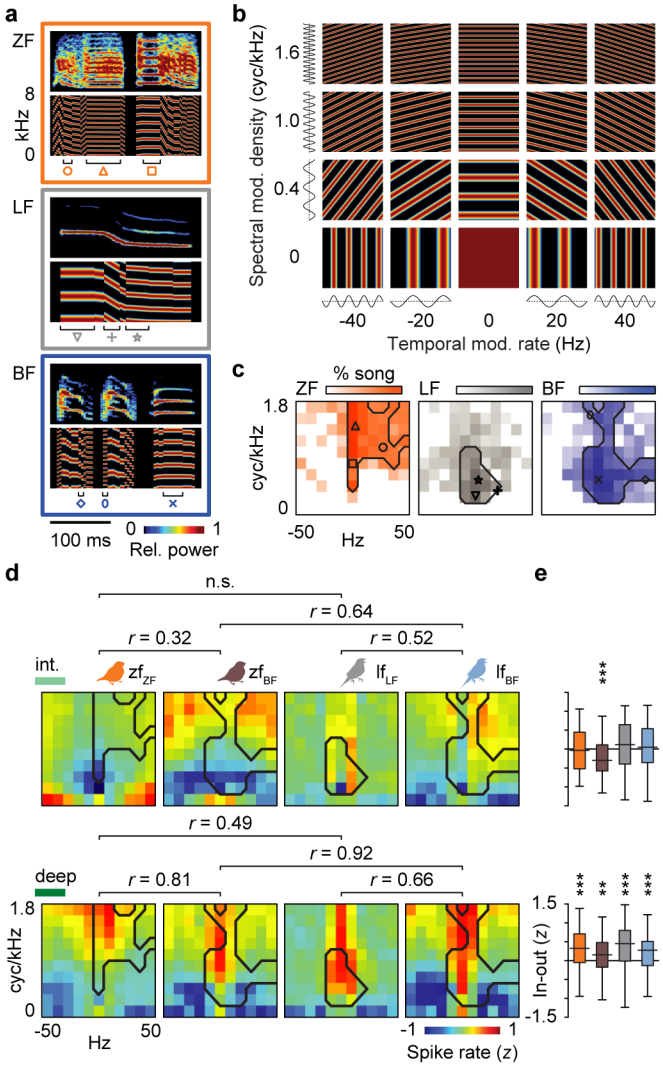

Fig. 4.

Tuning for the spectrotemporal modulations in learned song emerges in parallel with song selectivity. **a**, Spectrograms (0–8 kHz) of ZF, LF, and BF syllables are shown above spectrograms of their best-fit ripples. **b**, Spectrograms of some ripples used as stimuli, organized by spectral modulation (harmonic) density and temporal modulation rate. **c**, Song modulation heat maps show the log-transformed proportions of ZF, LF, and BF songs ($n = 5$ each) composed of each spectrotemporal modulation frequency. Symbols indicate the modulation frequencies of ripples shown in **a**; contour lines delineate the primary modulations constituting 90% of each species' songs. **d**, Neural response heat maps show the mean normalized spike rates to ripple stimuli from the intermediate (upper) and deep (lower) regions of each bird group (zf_{ZF}, $n = 167$, 296 neurons from intermediate and deep regions, respectively; zf_{BF}, $n = 89$, 186 neurons; lf_{LF}, $n = 192$, 255 neurons; lf_{BF}, $n = 162$, 111 neurons). Pearson correlation coefficients show the relationships between mean response maps of each bird group (for all shown, $P \leq 0.001$), and they were larger between birds that shared a tutor species (zf_{BF} and lf_{BF}) than between birds of the same species that had different tutor species (zf_{ZF} and zf_{BF}; lf_{LF} and lf_{BF}). The tutor species' song contour

lines from **c** are overlaid on the tuning response maps. **e**, Box-and-whisker plots of within-neuron differences in spike rates evoked by ripples inside versus outside the song contour lines. Sample sizes are the same as in **d**; the measure of center is the median, box limits show the 25^{th} and 75^{th} percentiles, whiskers extend to minimum/maximum values. Repeated-measures ANOVAs with bird identity as a covariate, $^{**}P < 0.001$, $^{***}P < 0.001$.

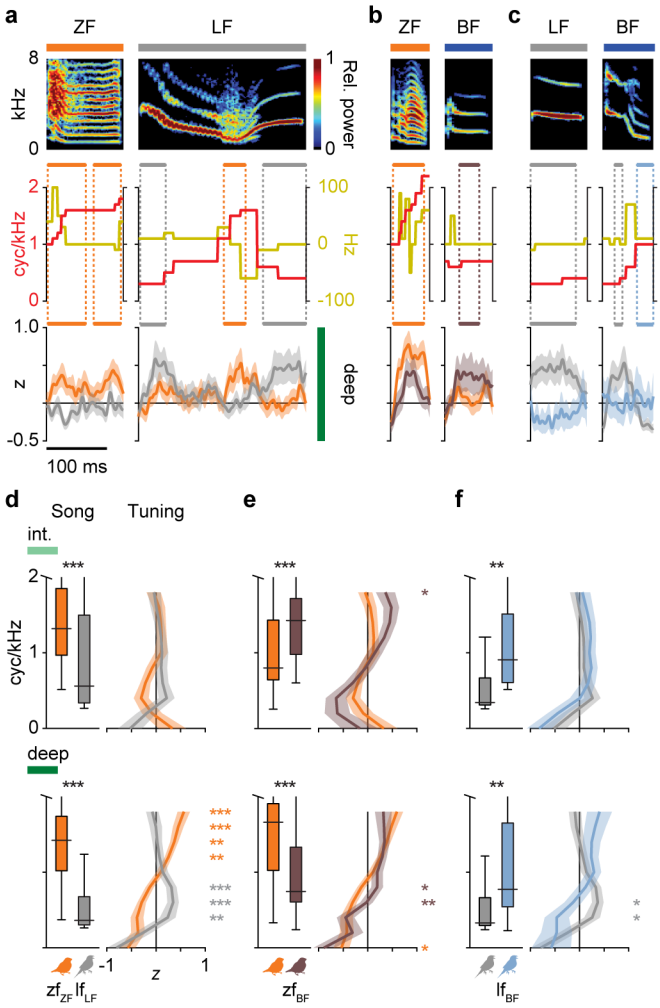

Fig. 5.

Neural population response dynamics to song reflect tuning for spectrotemporal modulations. **a**, Spectrograms (0–8 kHz) of different species' syllables are plotted above their corresponding spectral modulation density (red) and temporal modulation rate (yellow) vectors and above deep-region pPSTHs (mean ± 95% C.I.). Syllable segments that evoked a greater response in zf_{ZF} birds (orange, $n = 281$ neurons) or lf_{LF} birds (gray, $n = 237$ neurons) are indicated by horizontal lines. **b**, Same as **a**, but showing ZF and BF syllables and zf_{ZF} ($n = 270$ neurons) and zf_{BF} (brown, $n = 136$ neurons) pPSTHs. **c**, Same as **a**, but showing LF and BF syllables and lf_{LF} ($n = 224$ neurons) and lf_{BF} (light blue, $n = 75$ neurons) pPSTHs. **d**, *Left*, Box-and-whisker plots show the spectral modulation densities of syllable segments (from ZF and LF songs combined) that evoked sustained differences (≥10 ms) between zf_{ZF} (orange) and lf_{LF} (gray) pPSTHs. Top row shows data from intermediate-region pPSTHs [$n = 183$ (zf_{ZF}) and 127 (lf_{LF}) segments], and bottom row shows data from deep-region pPSTHs [$n = 188$ (zf_{ZF}) and 101 (lf_{LF}) segments]. *Right*, Spectral modulation tuning curves (mean ± 95% C.I.) of zf_{ZF} and lf_{LF} birds diverge at the same spectral modulation frequencies as those in syllable segments that drive distinct pPSTH responses [int., $n = 141$ (zf_{ZF}) and 151 (lf_{LF}) neurons; deep, $n = 242$ (zf_{ZF}) and 178 (lf_{LF}) neurons]. **e**, Same as **d** but

to ZF and BF songs and zf_{ZF} and zf_{BF} birds from the intermediate [n = 90 (zf_{ZF}) and 47 (zf_{BF}) syllable segments; n = 142 (zf_{ZF}) and 69 (zf_{BF}) neurons] and deep regions [n = 65 (zf_{ZF}) and 67 (zf_{BF}) segments; n = 234 (zf_{ZF}) and 123 (zf_{BF}) neurons]. **f**, Same as **d** but to LF and BF songs and lf_{LF} and lf_{BF} birds from the intermediate [n = 107 (lf_{LF}) and 73 (lf_{BF}) syllable segments; n = 149 (lf_{LF}) and 114 (lf_{BF}) neurons] and deep regions [n = 103 (lf_{LF}) and 50 (lf_{BF}) segments; n = 175 (lf_{LF}) and 54 (lf_{BF}) neurons]. For the boxplots in **d-f**, the measure of center is the median, box limits show the 25th and 75th percentiles, and whiskers extend to minimum/maximum values. Tests between syllable segments were ANOVAs with stimulus species as a covariate; tests between tuning curves were ANOVAs with bird identity as a nested covariate, *P < 0.05,
**P < 0.01, ***P < 0.001.

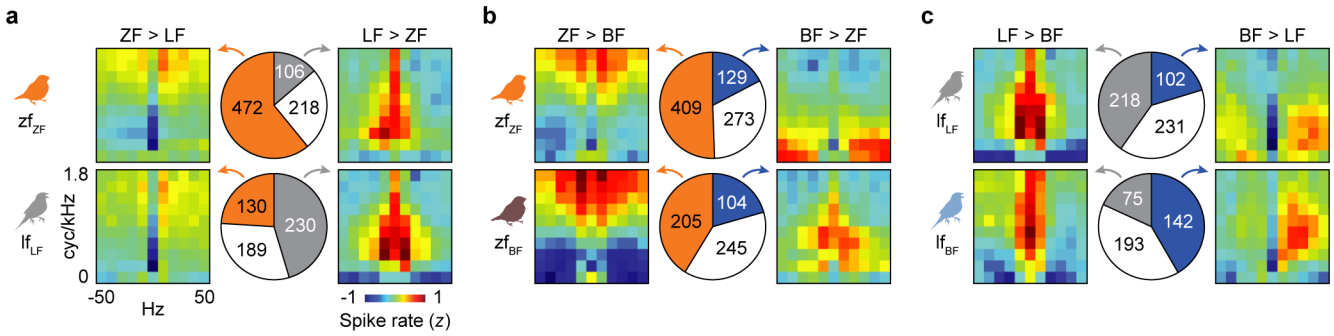

Fig. 6.

Neurons that respond selectively to same species' songs have highly similar modulation tuning regardless of species identity or tutoring experience. **a**, Pie charts show the proportions of neurons from all AC regions (with raw numbers superimposed) in zf_{ZF} (*top*) and lf_{LF} (*bottom*) birds that were selective for ZF song (orange), selective for LF song (gray), or not selective (open). Heat maps show the average modulation tuning maps of neurons selective for ZF (*left*) or LF (*right*) songs. **b**, Same as **a** but separating zf_{ZF} and zf_{BF} neurons based on selectivity for ZF or BF songs. **c**, Same as **a** but separating lf_{LF} and lf_{BF} neurons based on selectivity for LF or BF songs. For all comparisons, mean tuning maps of neurons with the same song selectivity were positively correlated (between groups: $0.35 \leq r \leq 0.92$, all $P < 0.001$), and maps of neurons from the same species but with different selectivity were not (within groups: $-0.70 \leq r \leq -0.02$). Modulation tuning maps for individual AC regions are shown in Fig. S15.