Systems/Circuits

Auditory Selectivity for Spectral Contrast in Cortical Neurons and Behavior

Nina L.T. So,[1,3] Jacob A. Edwards,[2,3] and Sarah M.N. Woolley[1,2,3]

[1]Program in Neurobiology and Behavior, [2]Psychology Department, and [3]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York 10027

Vocal communication relies on the ability of listeners to identify, process, and respond to vocal sounds produced by others in complex environments. To accurately recognize these signals, animals' auditory systems must robustly represent acoustic features that distinguish vocal sounds from other environmental sounds. Vocalizations typically have spectral structure; power regularly fluctuates along the frequency axis, creating spectral contrast. Spectral contrast is closely related to harmonicity, which refers to spectral power peaks occurring at integer multiples of a fundamental frequency. Although both spectral contrast and harmonicity typify natural sounds, they may differ in salience for communication behavior and engage distinct neural mechanisms. Therefore, it is important to understand which of these properties of vocal sounds underlie the neural processing and perception of vocalizations.

Here, we test the importance of vocalization-typical spectral features in behavioral recognition and neural processing of vocal sounds, using male zebra finches. We show that behavioral responses to natural and synthesized vocalizations rely on the presence of discrete frequency components, but not on harmonic ratios between frequencies. We identify a specific population of neurons in primary auditory cortex that are sensitive to the spectral resolution of vocal sounds. We find that behavioral and neural response selectivity is explained by sensitivity to spectral contrast rather than harmonicity. This selectivity emerges within the cortex; it is absent in the thalamorecipient region and present in the deep output region. Further, deep-region neurons that are contrast-sensitive show distinct temporal responses and selectivity for modulation density compared with unselective neurons.

Key words: **auditory cortex; harmonic sounds; perception; social communication; songbirds; vocalizations**

Significance Statement

Auditory coding and perception are critical for vocal communication. Auditory neurons must encode acoustic features that distinguish vocalizations from other sounds in the environment and generate percepts that direct behavior. The acoustic features that drive neural and behavioral selectivity for vocal sounds are unknown, however. Here, we show that vocal response behavior scales with stimulus spectral contrast but not with harmonicity, in songbirds. We identify a distinct population of auditory cortex neurons in which response selectivity parallels behavioral selectivity. This neural response selectivity is explained by sensitivity to spectral contrast rather than to harmonicity. Our findings inform the understanding of how the auditory system encodes socially-relevant signals via detection of an acoustic feature that is ubiquitous in vocalizations.

Introduction

Vocal communicators rely on auditory processing to acquire social information from others (Belin et al., 2004; Seyfarth and Cheney, 2017). Vocalizations are distinguished from other sounds by their acoustic structure (Rieke et al., 1995; Attias and Schreiner, 1998; Singh and Theunissen, 2003; Woolley et al., 2005). Because vocalizations are produced by periodic oscillations of vocal membranes, they are characterized by spectral contrast and harmonicity; sound energy fluctuates at regular intervals across the frequency axis (Riede and Goller, 2010; Titze, 2017). As the auditory equivalent of contrast in vision, spectral contrast is the difference between the peak and valley amplitudes of sound energy across the frequency spectrum (Singh and

Received May 23, 2019; revised Dec. 4, 2019; accepted Dec. 6, 2019.

Author contributions: N.L.T.S. and S.M.N.W. designed research; N.L.T.S. and J.A.E. performed research; N.L.T.S., J.A.E., and S.M.N.W. analyzed data; N.L.T.S., J.A.E., and S.M.N.W. wrote the paper.

This work was supported by a Croucher Scholarship to N.L.T.S.; N.L.T.S. was also supported by NIH Grant R01-DC009810, and J.A.E. was supported by NSF Grant IOS-1656825, both to S.M.N.W. We thank M.J. McPherson for generating inharmonic calls for behavioral experiments; V.F. Srey for assistance with behavioral data; S.A. Shamma for discussions on data analysis and interpretation; D.B. Kelley, W.B. Grueber, and S.A. Shamma for comments on the paper; and J.M. Moore for input on experiments.

The authors declare no competing financial interests.

Correspondence should be addressed to Sarah M. N. Woolley at sw2277@columbia.edu.

https://doi.org/10.1523/JNEUROSCI.1200-19.2019

Theunissen, 2003; Lewis et al., 2005). Harmonicity is the occurrence of frequencies at integer multiples of a fundamental frequency (F0; X. Wang and Walker, 2012; X. Wang, 2013). Spectral contrast and harmonicity are common in communication vocalizations (Soltis, 2010; Simmons and Megela Simmons, 2011; X. Wang, 2013), and are important for vocal perception (Vicario et al., 2001; Elliott and Theunissen, 2009; Oxenham, 2018). For example, degrading speech by reducing spectral resolution results in lower intelligibility (Shannon et al., 1995; Winn and Litovsky, 2015; Nogueira et al., 2016; Nourski et al., 2019) and difficulty identifying speaker identity and sex (Gonzalez and Oliver, 2005). This type of spectral degradation reduces both contrast and harmonicity, making it difficult to determine whether and how these features contribute to perception.

To enable effective neural encoding of vocalizations, the auditory cortex may be tuned to acoustic signatures of vocal sounds (Smith and Lewicki, 2006; Lewis et al., 2009; Perrodin et al., 2014; Shepard et al., 2015; Holdgraf et al., 2016; Allen et al., 2017). Because spectral contrast and harmonicity co-occur in vocalizations, determining whether auditory neurons encode one of these features or both is important for understanding the neural mechanisms of vocal communication. Evidence exists for cortical tuning to both features. The responses of cortical neurons are modulated by spectral peak placement and contrast (Schreiner and Calhoun, 1994; Shamma et al., 1994) and harmonic (integer ratio) relationships between frequency components (Feng and Wang, 2017). Whether both or one of these features drives response selectivity for vocal sounds is unknown, however. Spectral contrast and harmonicity covary in studies comparing neural responses to frequency combinations and stimuli with flatter spectra (Lewis et al., 2009; Norman-Haignere et al., 2013, 2016). Therefore, selectivity for the spectral structure of vocalizations could be due to tuning for either spectral contrast or harmonicity.

Here, we determined behavioral and neural sensitivity to spectral structure in the zebra finch (*Taeniopygia guttata*), a social songbird with high contrast, harmonic vocalizations (Zann, 1996; Brainard and Doupe, 2013) and a well characterized auditory system (Woolley, 2017). We measured perceptual sensitivity to spectral structure from behavioral responses to normal and vocoded calls. We measured neural sensitivity to spectral structure by recording the electrophysiological responses of single auditory cortex neurons to the same stimuli. We found that the neurons with response preferences that matched behavior were anatomically localized to the deep region of primary auditory cortex. We then tested whether these neurons were tuned to spectral contrast, harmonicity or both by recording their responses to synthetic sounds that varied in each feature independently. We found that neurons were tuned to spectral contrast rather than harmonicity. Because tuning results suggested that harmonicity was not a significant feature in response preference for vocal sounds, we tested birds' perceptual sensitivity to harmonic and inharmonic calls. Behavioral responses showed that birds were insensitive to harmonicity. Results indicate that neural tuning for the acoustic structure of vocalizations relies on spectral contrast and emerges within cortical circuits; it is present in neurons of the deep output region, but absent in the thalamorecipient region. Although both spectral contrast and harmonicity are ubiquitous in vocalizations, contrast appears to be the feature that drives auditory selectivity in both neural coding and behavior.

Materials and Methods

Stimuli

Nine natural distance calls served as stimuli in behavioral and physiological experiments and as stimulus templates to create vocoded calls (Fig. 1), harmonic and inharmonic synthetic calls, and tones at call fundamental frequencies (see Fig. 7). Calls were recorded from each of nine females housed alone in an anechoic sound-attenuation booth (Industrial Acoustics) through a microphone (Sennheiser, MKE 2–60) connected to an audio interface (Focusrite Saffire Pro 40), using Sound Analysis Pro software (Tchernichovski et al., 2000). Recorded calls were bandpass filtered between 300 and 8000 Hz before being used as templates to generate vocoded calls, ripples, and tones.

Vocoded calls were used in both behavioral and physiological experiments and were generated from natural calls using a vocoder implemented in MATLAB (Fig. 1; Gaudrain, 2016). The frequency axis of natural calls was divided into 16, 20, 27, 40, or 80 linearly-spaced bands by bandpass-filtering with 12th-order Butterworth filters, resulting in channel widths that decreased approximately linearly; channel widths were 481, 385, 285, 183, and 96 Hz, respectively. The amplitude envelope of each band was extracted by half-wave rectification and low-pass filtered at 150 Hz with fourth-order Butterworth filters. The extracted envelopes were used to modulate the amplitude of bandpass-filtered noise, and the modulated noise bands were combined to form a vocoded call.

Inharmonic calls were used in behavioral experiments and were generated from the nine natural calls using the modified STRAIGHT framework (McDermott et al., 2012) devised to synthesize inharmonic speech (McPherson and McDermott, 2018; Popham et al., 2018). Under the modified STRAIGHT framework, the input signal (a distance call) was decomposed into three time-varying components: spectral envelope, periodic excitation (voiced component), and aperiodic excitation (unvoiced component). The periodic excitation was modeled as a sum of sinusoids, and each sinusoid was then individually modified in frequency. These sinusoidal components were then recombined with the original aperiodic component and spectrotemporal envelope to produce an inharmonic call. Synthesized inharmonic calls had three degrees of inharmonicity (maximum frequency shifts). Frequencies of individual components in a distance call were shifted up or down by a random amount (i.e., jittered), and the amount of jitter was constrained within 10, 30, or 50% of the fundamental frequency. For each distance call and maximum jitter, we included three variants with different random jitter patterns. A constraint of 30 Hz was imposed on the minimum spacing between adjacent frequency components. Synthesized harmonic calls were generated with the same procedure except that frequency shifts were not introduced in the sinusoidal components before synthesis. Synthesized harmonic calls were generated to control for potential artifacts introduced by the synthesis procedure.

Ripple stimuli used in both behavioral and physiological experiments were 200 ms in duration, including 10 ms linear onset and offset ramps. Spectral ripples are broadband sounds consisting of sinusoidal modulations along the frequency axis that do not change along the time axis, and are the auditory equivalent of visual gratings (deCharms et al., 1998; Theunissen et al., 2004; Woolley et al., 2005). We generated ripple stimuli that parametrically varied in spectral contrast and phase. Spectral contrast is the difference in amplitude (dB) between peaks and valleys of spectral energy. Spectral contrast was varied by adjusting the amplitude of the sinusoidal envelope to values of 5, 10, 20, 40, and 80 dB. The harmonicity of ripples can be parametrically varied by shifting the phase at which frequency peaks occur. In each ripple, the phase at which frequency peaks aligned with integer multiples of the F0 was defined as zero. Phase-shifted ripples were defined by the amount of shift (in proportion of a cycle) relative to the phase-aligned ripple. Harmonicity was highest for the phase-aligned ripple (phase = 0) and decreased as phase deviated from zero (see Fig. 5*A*). Spectral modulation density is the frequency of the sinusoidal spectral envelope in units of cycles per kilohertz, and determines how close the spectral peaks are to one another. It is inversely proportional to the fundamental frequency, which in zebra finch females ranges from 400 to 700 Hz (Elie and Theunissen, 2016) and in males from

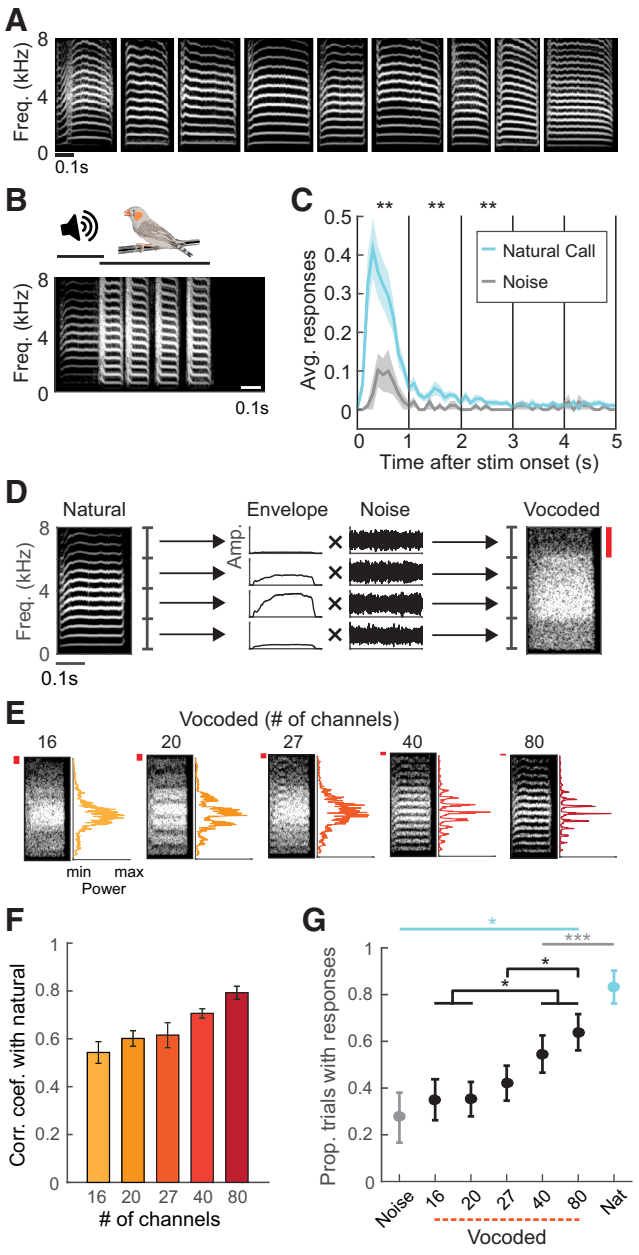

Figure 1. Spectral degradation decreases vocal responses to playback of calls. ***A***, Original nine female distance calls used as stimuli. ***B***, Example of one trial in the call-response behavior paradigm, in which presentation of a natural call evoked four response calls. ***C***, Average number of response calls in the first 5 s following the onset of natural call and noise presentation, shown in 100 ms time bins. Asterisks indicate significant differences between responses to natural calls and noise for each 1 s period (*$p < 0.01$, Wilcoxon signed rank tests). ***D***, Schematic showing how vocoded calls were generated from natural calls. The example depicts generation of a four-channel vocoded call. The natural call (left spectrogram) was decomposed into four spectral bands, and the amplitude envelope of each was extracted. Red bar indicates the width of one spectral channel. ***E***, Example spectrograms and frequency power spectra of vocoded versions of a call showing the differences in acoustic structure across test stimuli. Red bars indicate the widths of spectral channels. ***F***, Correlation coefficients between the frequency power spectra of vocoded calls and those of their natural call templates (mean ± SEM; $N = 9$). The spectral similarity of vocoded calls with natural calls increased with channel number. ***G***, Proportion of trials in which birds produced at least one response call for each stimulus type (mean ± SEM; $N = 10$). Cyan and gray asterisks indicate significant differences from natural calls and noise respectively in stimulus-evoked responses. Black asterisks and brackets denote significant differences between responses evoked by vocoded calls with different numbers of channels. *$p < 0.05$, ***$p < 0.001$, rank-transformed repeated-measures ANOVA with Dunn–Sidak tests.

600 to 1000 Hz (Vignal et al., 2008). We generated ripples with spectral modulation densities of 1.2, 1.6, and 2.0 cyc/kHz, which match the spectral modulation densities of zebra finch calls (1.2 cyc/kHz for males, 1.6 and 2.0 cyc/kHz for females) and song syllables (Vignal et al., 2008; Elie and Theunissen, 2016; Moore and Woolley, 2019). Eight ripple phases were spaced evenly across a cycle for each modulation density. Ripples were generated using custom software (S. Andoni, University of Texas).

Pure tones used in physiological experiments were 200 ms duration at frequencies ranging from 500 to 8000 Hz in 500 Hz intervals, and intensities ranging from 30 to 70 dB SPL, in 10 dB steps. Tones used in behavioral experiments were at frequencies and durations that matched the fundamental frequencies (F0s) and durations of natural calls, between 445 and 610 Hz and between 220 and 400 ms.

Behavioral experiments

Vocal responses to the presentation of natural calls, vocoded calls, noise, ripples and tones were recorded from adult male zebra finches (Fig. 1; >120 d old). Before testing, a bird was isolated for 3–5 d in an anechoic sound-attenuation booth (Industrial Acoustics) with *ad libitum* access to food and water. Experimental sessions for all birds began within 3 h of the onset of the light phase of the light/dark cycle and lasted ~80 min. For vocoded call experiments ($n = 10$), each bird's stimulus set consisted of four natural calls (selected at random from the set of 9 calls), the five vocoded versions of each natural call (with varying channel numbers), and a white noise sample that matched the average duration of the four natural calls. Stimuli were presented in pseudorandom order. Ten repetitions of each stimulus were presented. Interstimulus intervals were sampled from a uniform distribution between 15 and 22 s. Stimuli sampled at 44.1 kHz were delivered in the free field at 60 dB SPL through a speaker (Kenwood, KFC-1377) placed ~23 cm away from the perch in the booth. With each onset of stimulus presentation, audio recording was initiated to record vocal responses of the subject bird to presented stimuli.

For inharmonic call experiments, each bird's stimulus set consisted of three natural calls (from the same set of 9 as in vocoded call experiments), nine synthesized inharmonic versions of each natural call (10, 30, and 50% maximum shift, including three different shift patterns for each maximum shift value), three synthesized harmonic versions of each natural call, and a white noise sample that matched the average duration of the natural calls. Eight repetitions of each unique stimulus were presented in pseudorandom order. Interstimulus intervals and audio recording methods followed those of vocoded call experiments. Three of the eight birds in the inharmonic call experiments were also used in the vocoded call experiments.

For tone and ripple experiments, a separate cohort of birds ($n = 4$) was presented with a stimulus set containing the same natural calls used in previous experiments, plus either tones or ripples. Tone and ripple experiments were conducted 4 weeks apart. Stimuli were presented and behavioral responses were recorded exactly as in the other behavioral experiments.

Experimental design and statistical analysis: behavioral testing

Audio recordings of the 5 s following stimulus onsets were analyzed and distance calls were extracted. The probability of at least one distance call response to each stimulus was computed from the time stamps of extracted distance calls. Only birds that produced ≥1 distance call(s) in ≥10% of all trials and produced ≥1 distance call(s) in ≥50% of all trials for at least one specific stimulus were included in the final analysis. Because of individual variability, 50–60% of birds typically meet inclusion criteria (Vicario et al., 2001).

To test the statistical significance of differences in response call behavior by stimulus, we rank-transformed the data and used repeated-measures ANOVAs with an α of 0.05 to test the main effect of stimulus type on response behavior, and Dunn–Sidak tests for multiple comparisons between stimulus types. Wilcoxon signed rank tests were used to test differences between response rates along the time course of trials.

Electrophysiology

Recordings were made in 6 adult male zebra finches (>120 d old). For surgeries, birds were anesthetized with 0.5–2% isoflurane continuously

delivered through a custom inhalation device, and placed in a stereotaxic holder. Then, 2.5 by 2.5 mm bilateral craniotomies were made, centered at a point 1.25 mm lateral and 1.25 mm anterior from the bifurcation of the midsagittal sinus. Using dental acrylic, a metal pin was attached to the skull. A ground wire was inserted beneath the skull and affixed with dental acrylic at a position ~0.2 mm caudal to the bifurcation of the midsagittal sinus. After surgeries, lidocaine was applied to the head and birds recovered for 2 d before the first recording session. Before the first recording session and between sessions, craniotomies were covered with Kwik-Cast Sealant (World Precision Instruments).

Recordings of single neuron responses were made throughout auditory cortex (Figs. 2, 3). Recordings were made in awake, head-fixed birds inside a walk-in sound-attenuating booth (Industrial Acoustics). One to two electrode array penetrations were made per day, with probes oriented along the rostral-caudal axis. Each probe had 32 channels, with 8 conductive contacts on each of 4 shanks (NeuroNexus, A32). The spacing was 200 μm between shanks and 100 μm between contacts on the same shank. Neural responses were recorded at three or four depths along the same penetration, with the base of the probe positioned at 1.2, 2.0, and 2.8 mm, or 0.8, 1.6, 2.4, and 3.2 mm below the surface of the brain. Stimuli were sampled at 24.4 kHz and delivered through a speaker (JBL Control I) placed 23 cm in front of the bird. All non-tone stimuli were delivered at 60 dB SPL. During recording sessions, 10 repetitions of each stimulus were presented in pseudorandom order, with interstimulus intervals sampled from a uniform distribution spanning 0.75–1 s. Daily recording sessions lasted up to 5 h. Continuous voltage traces were amplified, bandpass filtered between 300 and 5000 Hz, digitized at 24.4 kHz (RZ5, Tucker-Davis Technologies), and stored for subsequent data analysis. Before each penetration, electrode arrays were coated with CM-DiI (C7000, Invitrogen) or SP-DiO (D7778, Invitrogen) dissolved in 100% ethanol. The DiI and DiO were alternated between adjacent penetrations along the medial-lateral axis so that they could be resolved in subsequent histological analysis.

Anatomical localization of recorded neurons. After the last recording session, a bird was given an overdose of Euthasol and transcardially perfused with saline followed by 10% formalin. The brain was extracted and postfixed in 10% formalin. After at least 24 h, the brain was transferred to 30% sucrose-formalin solution for cryoprotection. After cryoprotection, 40 μm parasagittal sections were made using a freezing microtome (American Optical). Sections were mounted and imaged (Olympus America) under CY3 and FITC filters to localize fluorescent DiI and DiO tracks. A bright-field image was taken for each brain section to delineate the borders of the thalamorecipient regions L2a, where dark fibers are visualized (Calabrese and Woolley, 2015; Moore and Woolley, 2019). Sections were then dried, stained for Nissl bodies and imaged to delineate borders between cortical regions based on their cytoarchitectural features (Fortune and Margoliash, 1992).

Boundaries between regions were visualized based on laminae, cytoarchitecture and thalamic fibers (Fig. 2). The intermediate region L2a (intermediate-a) is characterized by small, densely packed cells and by the termination of dark thalamic fibers. The intermediate region L2b (intermediate-b) is a population of densely packed, darkly Nissl-stained cells dorsal to the tip of intermediate-a. The deep region, L3, is caudal to intermediate-a, and ventral to intermediate-b. Neurons in the deep region are larger and less densely packed than those in intermediate-a, intermediate-b, and L. The ventral border of the deep region is the dorsal medullary lamina (LMD). The secondary region, caudal nidopallium (NC) is caudal to intermediate-b and L3.

To create anatomical maps of single unit locations in the auditory cortex (Fig. 3), we estimated the coordinates of each unit by measuring the location of recording sites visible in the DiI and DiO electrode tracks relative to anatomical reference points. Reference points were anatomical landmarks that could be identified in each hemisphere for each bird, allowing us to standardize recording location estimates across birds. To estimate the medial-lateral coordinates of all units recorded from a single electrode penetration, we determined the reference plane, which is the parasagittal section at which the ventral tip of L2a first intersects with the LMD when moving laterally from the midline. For each identified DiI or DiO track, the medial-lateral coordinate was determined based on the relative position of the DiI or DiO track compared with the reference plane. To estimate the caudal-rostral and dorsal-ventral positions within a parasagittal section, the reference point was defined as the point at which the ventral tip of L2a is closest to, or intersects with, the LMD. The position of the center of the probe base was measured relative to the reference point, and a coordinate was assigned to each unit based on the known positional difference between the recording site from which the unit was recorded, and the base of the probe.

Table 1. Number and proportion of all recorded units across auditory cortical regions that did (responsive) and did not meet inclusion criteria across stimulus types

Region	Call-responsive	Ripple-responsive	Tone-responsive	Total responsive	Total recorded
L2a (Int-a)	111 (89)* (26)†	108 (86)* (25)†	116 (93)* (27)†	125 (29)†	429
L2b (Int-b)	189 (66)* (45)†	164 (57)* (39)†	241 (84)* (58)†	287 (67)†	418
L3 (Deep)	411 (54)* (48)†	315 (41)* (37)†	537 (70)* (63)†	768 (90)†	852
NC (Sec)	136 (21)* (20)†	87 (13)* (13)†	290 (45)* (43)†	645 (95)†	681
All regions	847 (46)* (26)†	674 (37)* (21)†	1184 (65)* (36)†	1825 (56)†	3245

Numbers are counts of recorded units.

*Percentage of responsive neurons out of total responsive.

†Percentage of responsive neurons out of total recorded.

Experimental design and statistical analysis: electrophysiology

Data analysis was conducted as by Calabrese and Woolley (2015) and Moore and Woolley (2019). Spikes were detected and sorted offline using the WaveClus automated sorting algorithm followed by manual refinement (Quiroga et al., 2004). To detect spikes, we applied a nonlinear filter on the bandpass-filtered voltage trace, which emphasized high-amplitude and high-frequency voltage deflections. This maximized the accuracy with which the time-stamps of spike events were determined (Kim and Kim, 2000; Calabrese and Woolley, 2015; Moore and Woolley, 2019). We fed the voltage waveforms into the WaveClus algorithm to automatically sort spikes on the raw voltage traces from each channel, and manually refined the output by inspecting the waveform shape and amplitude of each cluster. Last, single units were identified based on signal-to-noise ratio (the difference between mean of spike amplitudes and noise amplitudes divided by the geometric mean of their SDs; 95% CI: 6.64–7.14), interspike interval distribution (the percentage of ISIs $<$1 ms; 95% CI: 0.04–0.06%), and stability of recordings across trials. This procedure identified a total of 1825 single units across L2a, L2b, L3, and NC.

Units were included in the analysis of call (vocoded and natural), ripple, and tone responses if they showed significant responses to at least 5% of all stimuli in each respective set. Significant responses were determined as follows. Each unit's spontaneous firing rate was computed from the 200 ms period preceding each trial. Driven firing rates were computed with spikes occurring between stimulus onset and 20 ms after stimulus offset. Onset firing rates were computed with spikes occurring within the first 50 ms following stimulus onset. Evoked responses were considered significant if either the driven firing rate or the onset firing rate was significantly higher or lower than spontaneous rates at $p < 0.05$. This procedure yielded 847 call-responsive, 674 ripple-responsive, and 1184 tone-responsive units across L2a, L2b, L3, and NC (Table 1). Response selectivity for call stimuli was defined as the proportion of natural and vocoded call stimuli that did not evoke significant driven firing rates from a given unit (Fig. 2).

Like mammalian cortex, the avian auditory cortex (AC) has two major physiological cell types, which differ in action potential waveform shape and in average spontaneous and stimulus-driven firing rate (Meliza and Margoliash, 2012; Harris and Mrsic-Flogel, 2013; Calabrese and Woolley, 2015; Araki et al., 2016). Because the correspondence between physiological cell type and morphological or biochemical features has not been established in the songbird, these types are referred to as putative excitatory principal cells (pPCs) and putative inhibitory interneurons (pINs). Each single unit was classified as either a pPC or a pIN based on spike waveform shape, as by Calabrese and Woolley (2015) (see Fig. 3). Briefly, we computed the average waveform for each unit using all de-

tected waveforms for that unit. We then computed the width of each unit's average waveform (the width at half-height of the negative peak, plus the width at half-height of the positive peak) and used a Mixture of Gaussians clustering algorithm to classify each unit as either a pPC or a pIN.

Spectral preference index. A spectral preference index (SPI) was used to compare units' responses to high-resolution vocoded calls (40 and 80 channels) and low-resolution vocoded calls (16 and 20 channels). The SPI was defined as the difference between average firing rates evoked by high-resolution vocoded calls and low-resolution vocoded calls, normalized by their sum. SPI ranged from −1 to 1, with more negative values indicating stronger responses to low-resolution calls, a value of zero indicating the same firing rates to high- and low-resolution vocoded calls, and more positive values indicating stronger responses to high-resolution calls. The SPI (see Figs. 3, 4, 6) was computed with the following formula:

$$\mathrm{SPI} = \frac{\mathrm{FR}_{40,80} - \mathrm{FR}_{16,20}}{\mathrm{FR}_{40,80} + \mathrm{FR}_{16,20}},$$

where $\mathrm{FR}_{40,80}$ represents the average firing rate evoked by 40- and 80-channel vocoded calls, and $\mathrm{FR}_{16,20}$ represents the average firing rate evoked by 16- and 20-channel vocoded calls. Neurons with SPI > 0.2, indicating 50% response enhancement to high- over low-resolution calls, were classified as high-resolution-selective (High). Neurons with SPI < -0.2, indicating 50% response enhancement to low- over high-resolution calls, were classified as low-resolution-selective (Low). The remaining units with $-0.2 <$ SPI < 0.2 were classified as unselective (Un).

Temporal response properties. Single-unit peristimulus time histograms (PSTHs; see Fig. 4) were constructed by calculating the trial-averaged instantaneous firing rates in 1 ms bins and smoothing the responses with a 5 ms Hanning window. Population PSTHs (pPSTHs) were computed by averaging the min-max normalized PSTHs across a population of single units. For visualization purposes, pPSTHs were smoothed by applying a 10 ms moving average. Latency of neural response (see Fig. 4) was calculated using established methods (Chase and Young, 2007; Schumacher et al., 2011) by identifying the first time after stimulus onset at which spiking activity significantly deviated from spontaneous activity ($p < 0.05$), assuming that the neuron was firing spontaneously with Poisson statistics. Call response latency was taken as the shortest latency among those computed for all natural and vocoded call stimuli. Tone response latency was calculated by averaging latencies across sound intensities and taking the shortest average latency across tone frequencies.

Onset index was calculated from responses to natural and vocoded calls using the following formula:

$$\text{Onset Index} = \frac{\mathrm{FR}_{\mathrm{onset}} - \mathrm{FR}_{\mathrm{sustained}}}{\mathrm{FR}_{\mathrm{onset}} + \mathrm{FR}_{\mathrm{sustained}}},$$

where $\mathrm{FR}_{\mathrm{onset}}$ represents the average firing rate during the first 50 ms after stimulus onset, and $\mathrm{FR}_{\mathrm{sustained}}$ represents the average firing rate during the subsequent period until stimulus offset. For each neuron, onset indices were averaged across all call stimuli that elicited a significant response (either the driven firing rate or the onset firing rate was significantly higher or lower than spontaneous rates at $p < 0.05$).

Sensitivity to spectral contrast and harmonicity. Responses to ripples were analyzed to determine neural sensitivity to spectral contrast and/or phase (see Figs. 5, 6). For ripples at each spectral modulation density (1.2, 1.6, and 2.0 cyc/kHz), a unit's driven firing rate was computed for each contrast-phase combination and used to construct a response matrix with contrast depth on the *y*-axis and phase on the *x*-axis. To examine ripple responses at the population level, the depth-phase matrices of a neuron's responses to ripples at the three spectral modulation densities were *z*-scored and averaged. Single neuron matrices were then averaged to make the population response depth-phase matrix.

To investigate whether spectral contrast and/or harmonicity (phase) explained the variance in firing rates to ripple stimuli, we used partial *F* tests to compare nested regression models. We began with a full model, including modulation density, spectral contrast, and phase as predictor variables. We then tested the contributions of spectral contrast and phase by constructing two reduced models, one without spectral contrast, and one without phase. *F* tests were used to compare each reduced model to the full model.

Modulation density and contrast in single neurons. Best spectral modulation density (see Fig. 6) was determined for each neuron by averaging the firing rates evoked by all ripples, across phases and depths, for each density, and selecting the modulation density with the highest average firing rate. Best phase was determined for each modulation density by averaging firing rates across contrasts, and selecting the phase evoking maximal average firing rate. Spectral contrast dependency (see Fig. 6) of single-unit responses was computed via Spearman's tests of correlation between driven firing rates and ripple contrasts at each tested modulation density and at a neuron's preferred phase (the phase eliciting maximal average response). A ρ of 1 indicates that firing rate increases monotonically with contrast, and a ρ of −1 indicates that firing rate decreases monotonically with contrast.

Statistical tests were nonparametric or on rank-transformed data. We used Kruskal–Wallis ANOVAs at an α of 0.05 to test for effects of brain region or SPI on the response metrics described above, with Dunn–Sidak tests for multiple comparisons. We used Wilcoxon signed-rank tests to test for effects of cell types on responses. To control for temporal response properties in some analyses, we performed ANCOVAs on rank-transformed response data. We used Pearson's χ^2 test to test distributions of neurons' preferred ripple phases.

Results

Spectral degradation decreases behavioral responses to calls

We first tested how spectral degradation affects the behavioral relevance of vocal communication sounds. Like humans, zebra finches rely on hearing to learn, produce and perceive the vocalizations they use for social communication. Zebra finches exchange distance calls when visually separated (Zann, 1996), and birds recognize their mates' voices (Vignal et al., 2004). Socially-isolated birds readily respond to distance call playbacks by vocalizing (Vicario et al., 2001; Vignal and Mathevon, 2011; Perez et al., 2015). We conducted a "call and response" behavioral test to assess how birds' responses differed with variations in the spectral properties of acoustic stimuli (Fig. 1). Stimuli were natural calls, noise and vocoded versions of the natural calls. First, adult males were housed alone in sound-isolated booths and presented with the distance calls of other birds and filtered noise stimuli (Fig. 1*A*,*B*). Vocal responses to call and noise presentations were recorded, analyzed and compared between stimulus types (Fig. 1*C*). Birds reliably responded to presentations of natural calls by producing their own distance calls, and the occurrence of response calls peaked within 1 s of stimulus call onset (Fig. 1*B*,*C*). The average strength of responses to noise presentation was significantly lower than the average strength of responses to calls (Fig. 1*C*; Wilcoxon signed rank tests, 1 s: $W = 55$, $p = 0.0020$; 2 s: $W = 55$, $p = 0.0020$; 3 s: $W = 36$, $p = 0.0078$). Responses to noise and natural calls provided the behavioral baseline against which the effects of spectral manipulations on response calls could be assessed.

To determine whether spectral degradation of call stimuli affected birds' vocal responses, we created vocoded calls that varied in spectral resolution (Fig. 1*D*,*E*). Vocoded calls were composed of linearly spaced spectral channels. Each channel consisted of bandpass filtered noise whose amplitude modulation matched that of the corresponding natural call (Fig. 1*D*). With each incremental increase in number of channels, the width of each channel decreased linearly. The presence of distinct and evenly spaced frequency components was apparent in vocoded calls with 40 and 80 channels; spectral channels were narrow enough for adjacent

frequency components to fall into different channels (Fig. 1*E*). The acoustic similarity of natural and vocoded calls increased with the number of channels in vocoded calls (Fig. 1*F*).

Using the same call and response paradigm, we tested whether natural and vocoded calls differed in behavioral salience (Materials and Methods). Vocal responses significantly differed across stimulus type and across vocoded calls with different numbers of channels (Fig. 1*G*). The probability of a bird responding increased with increasing spectral resolution (rank-transformed repeated-measures ANOVA, $F_{(6,54)} = 25.59$, $p = 4.10^{-14}$). Birds responded to a greater proportion of vocoded calls with 40 and 80 spectral channels than to noise (all $p < 5.10^{-5}$, Dunn–Sidak test). These results showed that birds responded more to vocoded calls with higher spectral resolution, indicating that the presence of discrete frequency components was necessary for eliciting behavioral responses.

Sensitivity to spectral degradation is anatomically localized in the auditory cortex

Because birds' responses to calls decreased with spectral degradation (Fig. 1), we hypothesized that AC neurons would show similar sensitivity to spectral degradation. The structure and function of the zebra finch auditory cortex are well studied (Woolley, 2017) and avian cell types (Dugas-Ford et al., 2012), connectivity patterns (Y. Wang et al., 2010), and information-processing principles (Schneider and Woolley, 2013; Calabrese and Woolley, 2015; Moore and Woolley, 2019) parallel those of mammalian cortex. Like mammalian cortex, zebra finch AC processes sound hierarchically (Fig. 2*A*). The intermediate subregions, L2a and L2b (hereafter intermediate-a and intermediate-b), receive input from the auditory thalamus and relay information to the superficial (L1/CM) and deep (L3) regions; the superficial regions also project to the deep region. The deep region is a major source of projections to the secondary auditory cortex (NC) and subcortical regions (Mello et al., 1998). To test whether the same AC neurons responded differently to natural and vocoded calls, we recorded the extracellular activity of single auditory neurons while birds were presented with the same natural and vocoded calls used in behavioral experiments (Materials and Methods; Table 1; Fig. 2*B*).

Single neurons' spiking responses to calls were progressively lower and more selective along the cortical processing pathway, in agreement with previous reports of song encoding in these regions (Meliza and Margoliash, 2012; Calabrese and Woolley, 2015; Moore and Woolley, 2019). Stimulus-evoked firing rates differed significantly across brain regions (Fig. 2*C*; Kruskal–Wallis ANOVA, $\chi^2_{(843)} = 62.62$, $p = 1.10^{-13}$); neurons in the intermediate-a region fired at higher rates than did neurons in all other regions (Dunn–Sidak test, all $p < 0.0079$), and secondary-region neurons fired at lower rates than did neurons in other regions (Dunn–Sidak test, all $p < 5.10^{-5}$). Response selectivity, defined as the proportion of calls that failed to evoke significant driven firing rates, was significantly lower in the intermediate-a region than in all other regions (Fig. 2*D*, left; Kruskal–Wallis ANOVA, $\chi^2_{(843)} = 31.78$, $p = 6.10^{-7}$; Dunn–Sidak test, all $p < 0.00014$). Response selectivity differed across vocoded calls only in the deep region, the avian parallel of mammalian layer 5. The proportion of high-resolution calls that evoked responses was larger than the proportion of low-resolution calls that evoked responses, in the same neurons (Fig. 2*D*, right; deep region: Kruskal–Wallis ANOVA, $\chi^2_{(2870)} = 64.08$, $p = 7.10^{-12}$; Dunn-Sidak test, all $p < 0.0067$). Response selectivity did not differ across stimuli in other AC regions (intermediate-a: Kruskal–Wallis ANOVA, $\chi^2_{(770)} = 1.77$, $p = 0.94$; intermediate-b: Kruskal–Wallis ANOVA, $\chi^2_{(1316)} = 9.28$, $p = 0.16$; secondary: Kruskal–Wallis ANOVA, $\chi^2_{(945)} = 3.55$, $p = 0.74$; data not shown). Additionally, average firing rates across vocoded calls differed only in the deep region (Kruskal–Wallis ANOVA, $\chi^2_{(2870)} = 17.42$, $p = 0.0079$; intermediate-a: $\chi^2_{(770)} = 0.6$, $p = 0.99$; intermediate-b: $\chi^2_{(1316)} = 3.69$, $p = 0.72$; secondary: $\chi^2_{(945)} = 0.58$, $p = 0.99$).

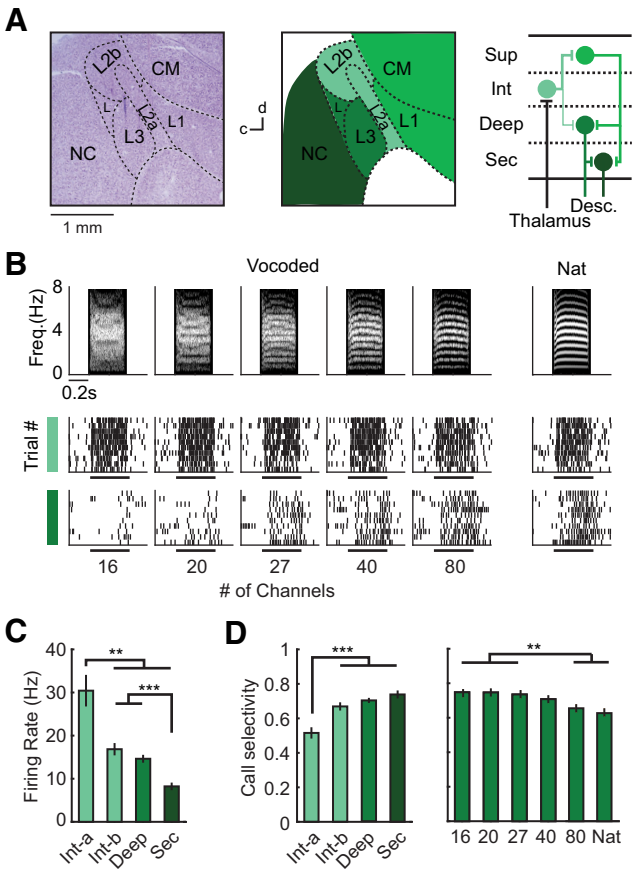

Figure 2. Auditory cortical organization and response selectivity among call stimuli. ***A***, Left, Cresyl violet-stained parasagittal section of the songbird brain. Dashed lines delineate major anatomical subdivisions. Middle, Traced diagram of the same section, with color and labeling denoting cortical regions. Right, Circuit diagram of major projections in the songbird AC (Y. Wang et al., 2010; Calabrese and Woolley, 2015). L2a, L2b, L1, L3, L, Subdivisions of the Field L complex; CM, caudal mesopallium. ***B***, Example spike trains of single neurons in intermediate and deep regions illustrating responses evoked by vocoded and natural calls. Spectrograms of stimuli are shown above the spike trains. ***C***, Average driven firing rates in different regions computed from single neuron responses to natural and vocoded calls (Kruskal–Wallis ANOVA with Dunn–Sidak tests; intermediate-a, $N = 111$; intermediate-b, $N = 189$; deep, $N = 411$; secondary, $N = 136$ call-responsive units). ***D***, Selectivity, defined as the proportion of stimuli that did not elicit a significant response, for different regions (left) and different stimulus categories in deep region neurons (right). Bar graphs show mean ± SEM. Asterisks denote significant differences revealed by Dunn–Sidak tests following Kruskal–Wallis ANOVA. For all panels, $^{**}p < 0.01$, $^{***}p < 0.001$.

To assess sensitivity to spectral resolution in each single neuron, we calculated a SPI value from each neuron's responses to vocoded calls. The SPI quantifies a neuron's response preference for low versus high spectral resolution (Materials and Methods; Fig. 3). To determine whether spectral preference was spatially organized within the AC, we anatomically mapped SPI by plotting the locations of recorded neurons based on reconstructed recording coordinates (Fig. 3*A*). Maps showed that SPI was spatially organized; neurons with positive SPI (blue), indicating

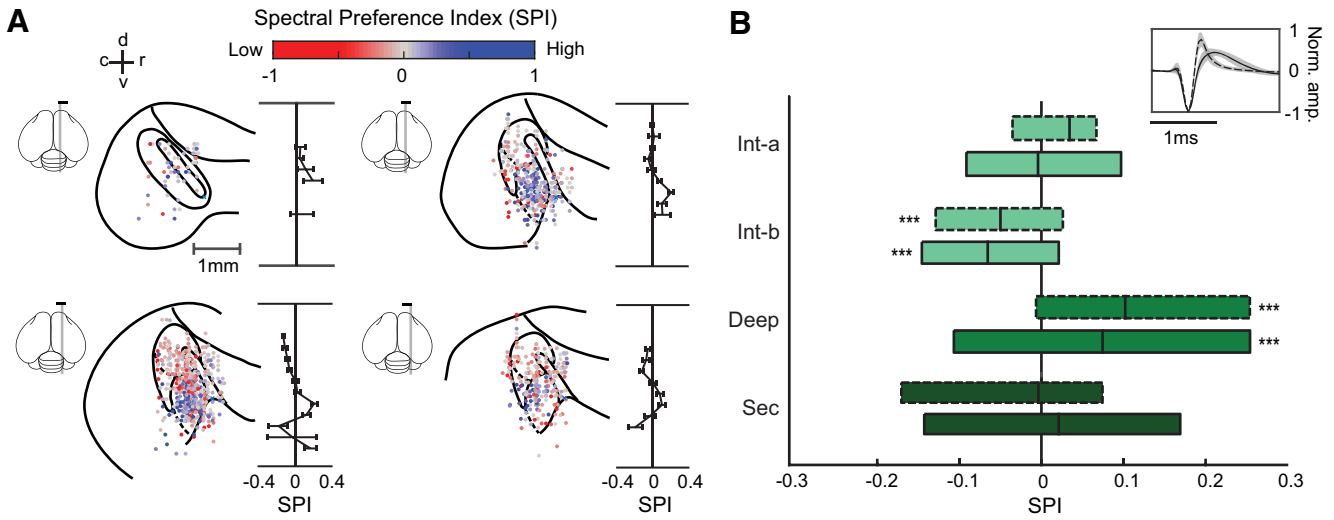

Figure 3. Neurons selective for high spectral resolution were localized in the deep region. ***A***, Spatial organization of SPI in the AC ($N = 1154$ call-responsive units). Each of the four sagittal brain diagrams show single units within a 0.3 mm range on the medial-lateral axis (estimated medial-lateral coordinates: 0.7–1.0, 1.0–1.3, 1.3–1.6, and 1.6–1.9 mm), plotted according to their rostral-caudal and dorsal-ventral coordinates. Each data point represents a single unit and is color-coded according to its SPI. To the right of each brain section diagram, average SPIs (mean ± SEM) for each dorsal-ventral position bin are plotted. Note that average SPI was only computed for bins with at least five single units recorded. ***B***, Boxplots showing SPI distribution of pINs (dashed outlines) and pPCs (solid outlines) in each auditory region. Asterisks indicate significant differences from zero. ***$p < 0.001$, Wilcoxon signed rank test; intermediate-a, N(pIN, pPC) = 39, 72; intermediate-b, N(pIN, pPC) = 54, 135; deep, N(pIN, pPC) = 85, 326; secondary, N(pIN, pPC) = 28, 108 call responsive-units.

preference for high-resolution calls, were concentrated in the deep region.

Because the songbird AC is comprised of two major cell types (Meliza and Margoliash, 2012; Harris and Mrsic-Flogel, 2013; Calabrese and Woolley, 2015; Araki et al., 2016), we tested whether each AC region's pINs or pPCs differed in spectral preference by determining whether their SPIs differed (Fig. 3*B*). We found no significant differences between pINs and pPCs in any region. In the deep region, both pINs and pPCs were selective for high-resolution calls (Wilcoxon signed rank test, pIN: $W = 2975$, $p = 5.10^{-7}$; pPC: $W = 2625$, $p = 3.10^{-6}$). In the intermediate-b region, both pINs and pPCs in were selective for low-resolution calls (Wilcoxon signed rank test, pIN: $W = 390$, $p = 0.0024$; pPC: $W = 2625$, $p = 2.10^{-5}$). SPIs in the intermediate-a region and secondary region did not differ significantly from zero (Wilcoxon signed rank tests, intermediate-a pIN: $W = 495$, $p = 0.14$; pPC: $W = 1274$, $p = 0.82$; secondary pIN: $W = 180$, $p = 0.60$; pPC: $W = 3208$, $p = 0.42$). Based on these results, the responses of pINs and pPCs were grouped in subsequent analyses of response properties.

Temporal response properties of deep region neurons vary with spectral structure preference

While the deep region contained a high density of neurons that were selective for more natural (high-resolution) calls (Figs. 2, 3), SPI did vary across deep-region neurons (Fig. 4*A*). To identify response features that covaried with selectivity in deep-region neurons, we tested how the temporal dynamics of call responses differed with SPI. We compared High, Un, and Low units' responses to calls using three measures of response dynamics: (1) latencies to first spike, (2) pPSTHs, and (3) onset index (OI). The responses of High ($N = 120$), Un ($N = 239$), and Low ($N = 52$) neurons differed significantly. First spike latencies differed between groups, even after controlling for onset firing rates (Fig. 4*B*, left; rank-transformed ANCOVA, $F_{(2,405)} = 6.86$, $p = 0.0012$). Latencies were longer in High and Low units than in Un units (Dunn–Sidak test, High-Un $p = 0.013$; Low-Un $p = 0.0075$). A similar pattern of latency differences was apparent in responses to pure tones, but differences were not significant after controlling for onset responses (Fig. 4*B*, right; rank-transformed ANCOVA, $F_{(2,351)} = 1.53$, $p = 0.22$). Response latencies to calls were correlated with response latencies to pure tones (Pearson's $r = 0.59$, $p = 3.10^{-35}$) and ripples (Pearson's $r = 0.75$, $p = 6.10^{-52}$).

Responses of High, Un, and Low neurons also differed over the course of a call (Fig. 4*C*,*D*). The population of Un neurons showed a strong response at stimulus onset, followed by a weaker, sustained response thereafter. In contrast, the population responses of High and Low neurons were strongest in the sustained portion of the response, after stimulus onset. To quantify the relationship between spectral preference and the temporal responses in each neuron, we compared OI (Materials and Methods) and SPI (Fig. 4*E*). Like SPI, OI differed across neurons, ranging from nearly 1 to −1. This analysis showed that OI values were significantly higher in Un neurons than in High or Low neurons (Fig. 4*E*; Kruskal-Wallis ANOVA, $\chi^2_{(408)} = 37.33$, $p = 8.10^{-9}$; Dunn–Sidak test, High-Un $p = 7.10^{-9}$; Low-Un $p = 0.023$). These results showed that response selectivity based on spectral structure was correlated with temporal response pattern; neurons that were sensitive to spectral structure had stronger sustained responses, whereas those that were insensitive to spectral structure had stronger onset responses.

Sensitivity to call structure is explained by spectral contrast

To understand the specific acoustic features driving sensitivity to spectral structure, we tested whether tuning for spectral contrast or harmonicity, or both, explained response preference for high-resolution calls, in deep-region neurons (Figs. 5, 6). As described in Materials and Methods, spectral degradation decreases both contrast and harmonicity, and each acoustic property has been proposed to be important for vocal perception. AC neurons that are sensitive to amplitude differences across frequencies have been identified in guinea pigs (Catz and Noreña, 2013) and marmosets (Barbour and Wang, 2003), and enhanced neural repre-

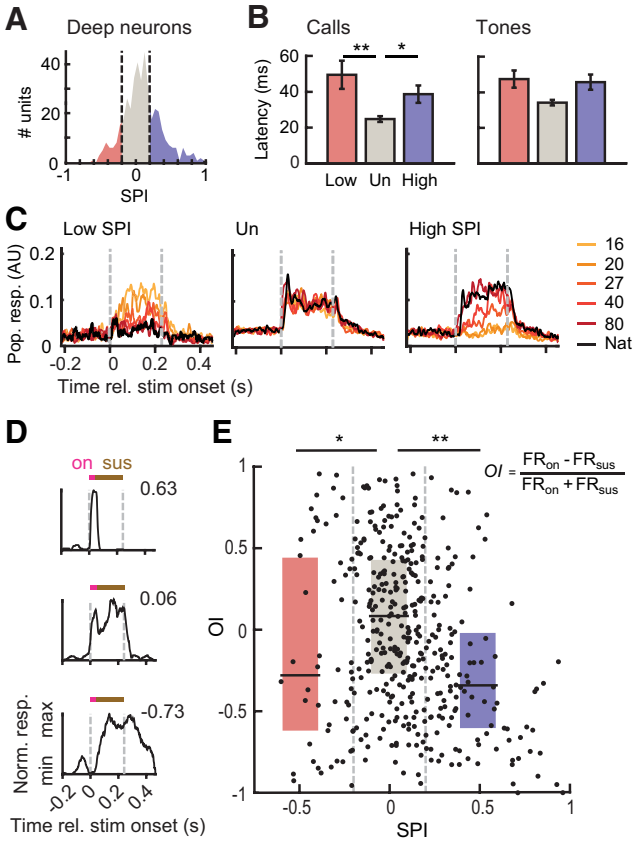

Figure 4. Temporal response patterns differ with spectral preference index (SPI) in the deep region of auditory cortex ($N = 411$ call-responsive units). ***A***, Segmentation of deep region units into low-resolution-selective (Low, red), unselective (Un, gray), and high-resolution-selective (High, blue) groups (Low, $N = 52$; Un, $N = 239$; High, $N = 120$). ***B***, First-spike latencies in response to natural and vocoded calls (left) and tones (right), shown for Low, Un, and High neurons (rank-transformed ANCOVA with Dunn–Sidak tests). ***C***, Example call-evoked pPSTHs for Low, Un, and High neurons. AU, Arbitrary units. pPSTHs to different stimulus groups are color-coded according to the legend. Responses of Un neurons peak at stimulus onset, whereas High and Low responses increase after stimulus-onset and are sustained over the duration of the stimulus. ***D***, Example PSTHs for neurons with negative (top), near-zero (middle), and positive OI (bottom). Pink and brown bars show time periods where onset and sustained firing rates were measured. ***E***, Scatter plot and box plots showing the relationship between OI and SPI. Onset responses are stronger in Un neurons than in High or Low neurons (Kruskal–Wallis ANOVA with Dunn–Sidak tests). For all panels, $*p < 0.05$, $**p < 0.01$.

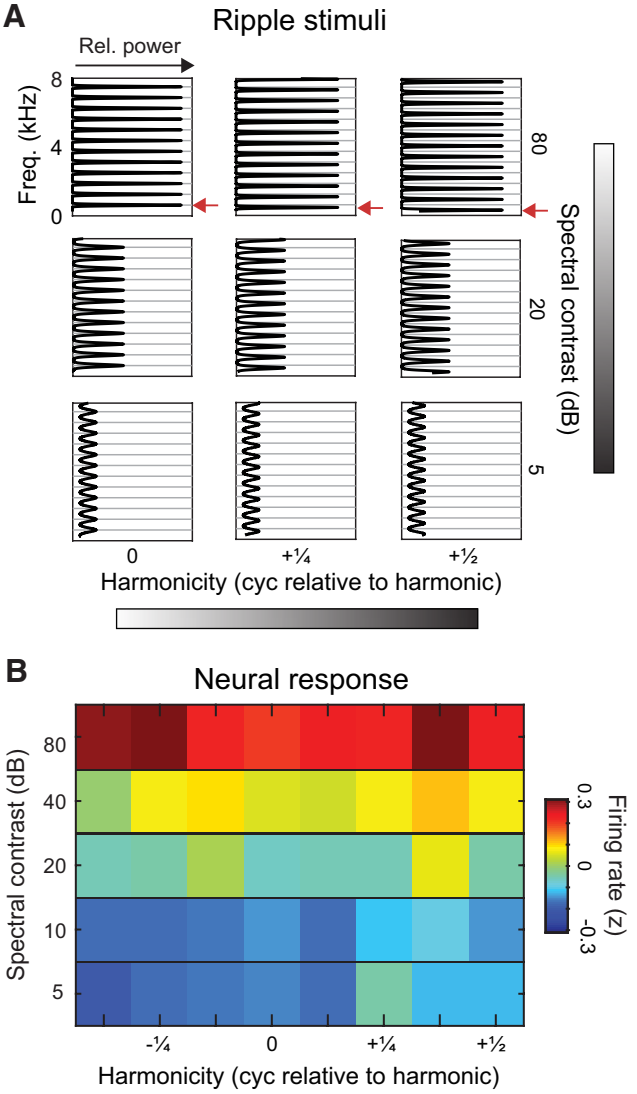

Figure 5. Selectivity for behaviorally relevant sounds is explained by tuning for spectral contrast rather than harmonicity. ***A***, Spectral profiles of ripples varied in spectral contrast and phase relative to harmonic phase. Gray lines indicate the peaks of frequency components if they were integer multiples of an F0 (F0 = 1000 Hz/1.6 = 625 Hz). ***B***, Heatmap of population firing rates in response to ripples with varying phase and spectral contrast. Each pixel shows the mean *z*-scored firing rate across all deep region units ($N = 315$ ripple-responsive units).

sentation of spectral peak-to-valley differences is predicted by computational models to result from lateral inhibition (Shamma, 1985; Yost, 1986). The detection of harmonically related frequency components is a proposed mechanism in spectral models of pitch extraction (Duifhuis et al., 1982; Scheffers, 1983).

To test neuronal sensitivity to spectral contrast and harmonicity, we recorded the responses of single deep-region neurons to a set of spectral ripples (Shamma et al., 1994), in which spectral contrast and harmonicity were independently varied (Fig. 5). We analyzed responses first by constructing contrast-phase matrices of the average firing rate evoked by each ripple. Responses were higher to ripples with the largest contrast, regardless of phase, and decreased with decreasing contrast (Fig. 5*B*). To determine the respective contributions of spectral contrast and harmonicity, we tested whether removing contrast level or phase value as predictor variables from a full multiple linear regression model would reduce its ability to predict firing rates. The full model, including spectral contrast, phase, and modulation density as predictors, explained 68% of population firing rate variance (adjusted $R^2 = 0.68$). Removing phase from the model did not change its predictive power ($F_{(1,116)} = 1.86$, $p = 0.18$; adjusted $R^2 = 0.68$). Removing spectral contrast, however, decreased the model's predictive power ($F_{(1,116)} = 226.87$, $p = 5.10^{-29}$; adjusted $R^2 = 0.065$). Results showed that deep-region neurons were sensitive to spectral contrast, and that sensitivity to harmonicity was not significant, at the population level. We then analyzed each neuron's preferred phase, and found that preferred phases were distributed evenly at modulation densities and SPIs (Table 2).

We then analyzed tuning to modulation density and spectral contrast in Low, Un, and High neurons. Figure 6*A* shows the contrast-phase matrices for each neuron group for the three modulation densities examined. Each neuron's matrix was centered such that the phase evoking the highest firing rate (preferred phase) was equal to zero. Each neuron included in the analysis was responsive to both calls and ripples.

Low, Un, and High units showed distinct tuning profiles to modulation density and contrast. Low neurons largely preferred

Table 2. Neurons' best phases are distributed evenly in High, Un, and Low SPI neurons and at each spectral modulation density

SPI	1.2 cyc/kHz	1.6 cyc/kHz	2.0 cyc/kHz
High	$\chi^2 = 9.20$	$\chi^2 = 9.20$	$\chi^2 = 7.20$
	$p = 0.24$	$p = 0.24$	$p = 0.41$
Un	$\chi^2 = 9.33$	$\chi^2 = 10.67$	$\chi^2 = 9.71$
	$p = 0.23$	$p = 0.15$	$p = 0.21$
Low	$\chi^2 = 13.79$	$\chi^2 = 8.45$	$\chi^2 = 5.55$
	$p = 0.055$	$p = 0.29$	$p = 0.59$

Results of Pearson's χ^2 tests. There are 7 degrees of freedom for all tests.

ripples with the lowest modulation density of 1.2 cyc/kHz (69.7%; Fig. 6*B*, left), and fired less to ripples with deeper modulations (Fig. 6*C*, left). Un neurons showed a relatively even distribution of preferred modulation density (Fig. 6*B*, middle), and moderately increased firing with increasing contrast (Fig. 6*C*, middle). High neurons predominantly preferred ripples with higher modulation densities of 1.6 or 2.0 cyc/kHz (58.8 and 31.3%, respectively; Fig. 6*B*, right), and fired more to ripples with greater contrast (Fig. 6*C*, right). These results showed that neurons with response preferences for calls with natural-like structure (High neurons) also showed response preferences for modulation densities typical of female zebra finch calls (Vignal et al., 2008; Elie and Theunissen, 2016).

Sensitivity to spectral contrast was a strong predictor of SPI (Fig. 6*D*). Spectral contrast dependency (Spearman's ρ) was used to quantify the monotonicity of the association between firing rate and spectral contrast at a neuron's preferred phase, for a given modulation density. Positive values indicate that firing rates increase with contrast, while negative values indicate that firing rates decrease with increases in contrast. High neurons had significantly more strongly positive ρ than Low and Un neurons for all spectral modulation densities examined (Kruskal–Wallis ANOVA, 1.2 cyc/kHz: $\chi^2_{(278)} = 30.35$, $p = 3.10^{-7}$; 1.6 cyc/kHz: $\chi^2_{(278)} = 74.21$, $p = 8.10^{-17}$; Fig. 6*D*, left; 2.0 cyc/kHz: $\chi^2_{(278)} = 58.91$, $p = 2.10^{-13}$; Dunn–Sidak tests, $p < 0.0058$ for all pairwise comparisons and all modulation densities). Contrast dependency was significantly correlated with call SPI at all modulation densities examined (Pearson correlations, 1.2 cyc/kHz: $r = 0.35$, $p = 1.10^{-9}$; 1.6 cyc/kHz: $r = 0.56$, $p = 2.10^{-24}$; Fig. 6*D*, right; 2.0 cyc/kHz: $r = 0.50$, $p = 3.10^{-19}$). For all three modulation densities, a majority of units (68.3, 71.2, and 74.0%) had contrast dependencies that matched the sign of selectivity for vocoded calls (i.e., positive ρ and positive SPI, or negative ρ and negative SPI). Based on the tuning for high contrast characterizing High neurons and the close relationship between contrast tuning and SPI, results indicate that sensitivity to spectral contrast rather than harmonicity underlies neural response preference of deep output neurons for the spectral structure of communication calls.

Behavioral responses to calls do not vary with harmonicity

The electrophysiological experiments showed that AC neurons were sensitive to spectral contrast rather than harmonicity. We reasoned that, if tuning predicted perceptual behavior, then birds would not be sensitive to harmonicity. To test this hypothesis, we conducted further behavioral experiments, measuring vocal responses to harmonic and inharmonic calls. Synthesized calls varied in harmonicity but not contrast (Fig. 7*A*). For synthesized calls, each frequency component was randomly shifted up or down by a maximum amount of 0, 10, 30, or 50% of the F0 (Materials and Methods).

Using the same call and response paradigm used with natural and vocoded calls, we measured vocal responses to harmonic and

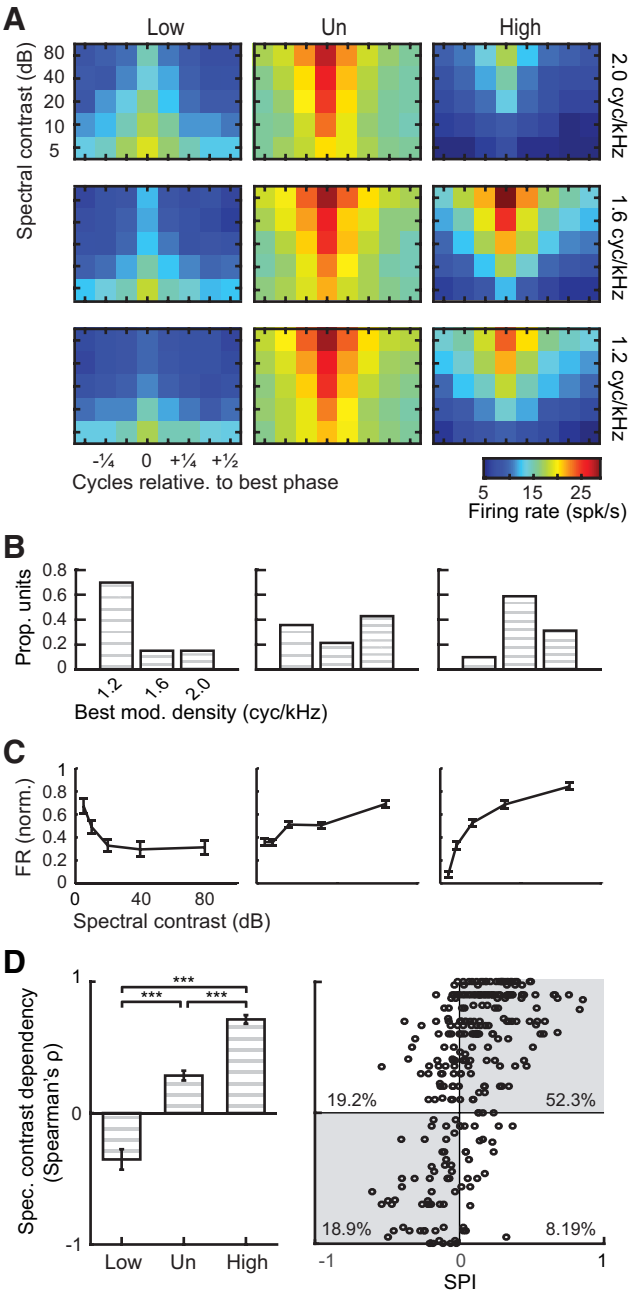

Figure 6. Spectral tuning aligns with response selectivity for calls ($N = 281$ ripple- and call-responsive units for all panels). ***A***, Heatmaps of average firing rates of Low ($N = 33$), Un ($N = 168$), and High ($N = 80$) neurons in response to ripples with varying depth and phase at three modulation densities. Neurons are aligned by best phase. ***B***, Proportions of Low, Un, and High neurons showing maximal responses to each modulation density. ***C***, Normalized firing rates (mean ± SEM) of Low, Un, and High neurons evoked by each contrast, measured at a modulation density of 1.6 cyc/kHz and at each neuron's best phase. Firing rates were normalized to range from 0 to 1 for each unit. ***D***, Left, Spectral contrast dependencies (mean ± SEM) of Low, Un, and High units, measured at a modulation density of 1.6 cyc/kHz and at each neuron's best phase (Kruskal–Wallis ANOVA with Dunn–Sidak tests). ***$p < 0.001$. Positive ρ indicates that driven responses increase with spectral contrast, and negative ρ indicates that driven responses decrease with increases in spectral contrast. ***D***, Right, Scatter plot showing the relationship between spectral contrast dependency (measured at 1.6 cyc/kHz) and spectral preference for calls (SPI). Shaded quadrants include units whose direction of spectral contrast preference matched the sign of call SPI.

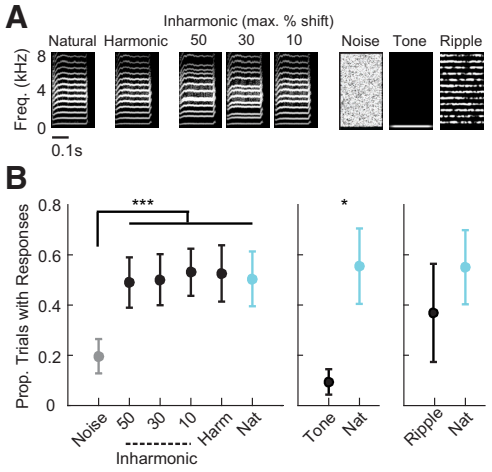

Figure 7. Behavioral responses to call playback are not affected by inharmonicity. ***A***, Example spectrograms of a natural call, a synthesized harmonic call, synthesized inharmonic calls, noise, a tone and a ripple. ***B***, Left, Proportion of trials in which birds produced response calls (mean ± SEM; $N = 8$). Middle, Proportion of trials in which birds produced response calls to tones versus natural calls (mean ± SEM; $N = 4$). Right, Proportion of trials in which birds produced response calls to ripples versus natural calls (mean ± SEM; $N = 4$). Asterisks indicate significant differences in stimulus-evoked responses (for all panels: rank-transformed repeated-measures ANOVA with Dunn–Sidak tests). $^{*}p < 0.05$, $^{***}p < 0.001$.

inharmonic calls (Fig. 7*A*,*B*). We presented natural calls, synthesized harmonic calls, synthesized inharmonic calls and filtered noise segments. As in the previous behavioral experiment, stimulus type had a significant effect on birds' responses (rank-transformed repeated-measures ANOVA, $F_{(5,35)} = 10.35$, $p = 4.10^{-6}$). Responses to all inharmonic calls were significantly above noise-evoked responses (Fig. 7*B*, left; Dunn–Sidak tests, all $p < 0.00017$). But responses to inharmonic calls did not differ from responses to harmonic and natural calls, at any frequency shift (Fig. 7*B*, left; Dunn–Sidak tests, all $p > 0.99$). Finally, we conducted two additional behavioral experiments to test whether the F0s of calls or phase-shifted ripples differed from natural calls in behavioral salience. With the same call and response approach used previously, we tested birds' vocal responses to tones at frequencies matching call F0s (Fig. 7*B*, middle) and to the phase-shifted ripples presented during electrophysiological recording (Fig. 7*B*, right). Birds were significantly less responsive to tones than to natural calls (Dunn–Sidak test, $p = 0.013$), and equally responsive to phase-shifted ripples and natural calls (Fig. 7*C*; Dunn–Sidak test, $p = 0.33$). These results confirmed that birds' behavioral responses did depend on spectral contrast and did not depend on harmonicity.

Discussion

We found that spectral contrast is a behaviorally-relevant vocalization feature that is represented by a distinct neuronal population within the AC. A robust neural representation of contrast emerges within cortical circuitry, first evident in deep primary auditory cortex. Spectral selectivity of neurons in the deep output region may subserve the perceptual identification of vocalizations in the environment and the extraction of social information carried in those signals. Compared with unselective neurons, contrast-tuned neurons have longer first spike latencies and more sustained responses. Finally, we distinguish spectral contrast from harmonicity as a driver of behavioral and neural selectivity. Neurons that are selective for behaviorally salient vocal sounds are characterized by a preference for high contrast, and lack specific tuning to sounds with harmonically-related frequency components.

The spectral structure of speech contributes to the extraction of social information from voices and speech intelligibility in the presence of interfering signals (Popham et al., 2018). Listeners' abilities to correctly identify speaker identity and sex decrease with reduction of number of channels in noise-vocoded speech (Gonzalez and Oliver, 2005). Noise-excited speech, which mimics whispered speech and lacks spectral modulations and harmonicity, is more difficult for listeners to understand in speech mixtures than is voiced speech (Popham et al., 2018). Our experiments demonstrate that for zebra finches, the spectral structure of calls must be preserved to elicit vocal responses. To elicit responses, calls must contain distinct spectral peaks and valleys, but the spectral peaks need not be harmonically related.

Our results add to existing knowledge on acoustic features that drive social responses to vocalizations in the zebra finch. In the spectral domain, call stimuli that have F0s within the range of 550–750 Hz evoke the strongest behavioral responses (Vicario et al., 2001). Wideband calls are preferred to narrowband calls, with at least four frequency components required to elicit typical behavioral responses (Vignal and Mathevon, 2011). In our study, only 40- and 80-channel vocoded calls had distinct frequency components, and only those vocoded calls evoked significant vocal responses. This selectivity is likely not a response to coarse spectral shape and amplitude envelope, as vocoded calls with differing channel numbers have similar coarse spectral shapes and amplitude envelopes that are highly correlated to those of natural calls. We also found that birds responded similarly to inharmonic and harmonic calls, despite the ability to detect fine frequency shifts (discrimination thresholds <1 Hz) in harmonic sounds with training (Lohr and Dooling, 1998). While zebra finches are likely able to discriminate between inharmonic and harmonic calls, harmonicity does not appear to be the acoustic feature that cues a communication response, in the absence of training. The acoustic structure of zebra finch vocalizations makes this species particularly suitable for testing sensitivity to spectral contrast and harmonicity. Some species use more tonal vocalizations, with narrower frequency bandwidths than those of zebra finch calls and speech, however (Podos, 1997). In such species, acoustic features other than spectral contrast may be salient. For example, some bats use social calls with prominent frequency-modulated sweeps and have inferior colliculus neurons that are sensitive to spectral motion (Andoni and Pollak, 2011). A current hypothesis is that vocal acoustics and auditory tuning coevolve to align communication receivers' encoding mechanisms with senders' signals (Woolley and Moore, 2011; Moore and Woolley, 2019).

The importance of spectral structure could differ between types of auditory tasks. A previous study showed that birds trained to recognize harmonic musical tone sequences recognized noise-vocoded versions that lack deep spectral modulations and harmonicity, as long as the coarse spectral shape was preserved (Bregman et al., 2016). Here, we avoided assigning salience to one or multiple acoustic feature(s) through training to measure natural salience. Zebra finches use calls to monitor each other's locations when socially isolated in the wild (Zann, 1996). Although some evidence suggests that call rates may be modulated by social context or the need to identify individuals (Vignal et al., 2004, 2008), zebra finches reliably produce response vocalizations without reinforcement (Zann, 1996). Further studies are needed to determine the significance of spectral contrast and harmonicity for communication tasks such as extracting signals

from sound mixtures or evaluating social context, which may require training.

The identification of high-resolution-selective neurons in the deep region of primary AC suggests that the processing of behaviorally relevant sounds engages specialized neural pathways. Specifically, high-resolution-selective neurons preferred sounds with deeper and denser modulations, but showed no preference for harmonic placement of spectral peaks. Consistent with the tuning properties of deep region neurons, birds responded similarly to harmonic and inharmonic calls, indicating that the preservation of deep spectral modulations was sufficient to elicit behavioral responses. Previous studies have characterized AC neurons by their responses to spectral modulation density, depth, and phase (Schreiner and Calhoun, 1994; Shamma et al., 1994). And other studies have reported anatomical organization of complex response properties, such as preferred bandwidth (Rauschecker et al., 1995), spectrotemporal modulation tuning (Hullett et al., 2016), and F0-specific responses (Bendor and Wang, 2005). Our study builds on previous work by providing evidence for robust anatomical grouping of neurons tuned to high contrast, and further establishing that contrast tuning contributes to the selective representation of behaviorally-salient sounds. Although deep-region neurons reside within primary auditory cortex and are therefore unlikely to directly control behavioral output, neurons selective for contrast may send information to downstream regions that drive behavioral sensitivity. Interestingly, it is the deep-region neurons that project to brainstem regions surrounding the vocal control nuclei, in songbirds (Kelley and Nottebohm, 1979; Vates et al., 1996; Mello et al., 1998). In the mouse AC, response sensitivity to stimulus contrast is also strongest in neurons in the deep layers (Cooke et al., 2018). Future experiments in both systems could specifically target contrast-sensitive neurons to identify and manipulate their inputs.

Our results can also be discussed with regard to contrast gain control and existing theories of natural sound processing. A previous study showed that AC neurons can dynamically adjust gain to compensate for changes in spectral contrast but this compensation is incomplete (Rabinowitz et al., 2011). In our dataset, deep region responses scaled with spectral contrast, indicating that if compensatory gain control were present, it did not result in invariance to contrast. We characterized ripple responses at three spectral modulation densities, and found that High neurons preferred higher densities (1.6–2.0 cyc/kHz), which correspond to the spectral modulation densities of female zebra finch calls (Vicario et al., 2001; Singh and Theunissen, 2003; Woolley et al., 2005; Mouterde et al., 2014). These findings support the hypothesis that auditory neurons are sensitive to the spectral properties of communication sounds (Singh and Theunissen, 2003).

One interpretation of the observed longer call response latencies in High and Low neurons compared with Un neurons is that input projections to these neurons may differ. In the songbird AC, deep-region neurons receive input from multiple pathways: (1) the superficial region (Wild et al., 1993; Vates et al., 1996); (2) the thalamorecipient regions (Y. Wang et al., 2010); and (3) the shell region of the auditory thalamus, a relatively sparse projection (Vates et al., 1996). One possibility is that High and Low neurons receive input from the superficial region, undergoing more intracortical processing than Un neurons. Intracortical processing and feedback connections from higher cortical areas have been proposed to result in sustained firing in the AC (X. Wang et al., 2005). Because High and Low neurons show a more sustained firing profile than Un neurons, intracortical processing and feedback from superficial regions could contribute to their selectivity. Alternatively, differences in excitation/inhibition balance may result in selectivity, as has been shown in zebra finch secondary AC (Yanagihara and Yazaki-Sugiyama, 2016). Further, an analysis of primary AC responses to modulated noise in ferrets shows that neurons with sustained responses are more likely to be selective for certain acoustic features than are those with onset responses, potentially through circuit-level inhibition (Lopez Espejo et al., 2019).

References

Allen EJ, Burton PC, Olman CA, Oxenham AJ (2017) Representations of pitch and timbre variation in human auditory cortex. J Neurosci 37:1284–1293.

Andoni S, Pollak GD (2011) Selectivity for spectral motion as a neural computation for encoding natural communication signals in bat inferior colliculus. J Neurosci 31:16529–16540.

Araki M, Bandi MM, Yazaki-Sugiyama Y (2016) Mind the gap: neural coding of species identity in birdsong prosody. Science 354:1282–1287.

Attias H, Schreiner CE (1998) Coding of naturalistic stimuli by auditory midbrain neurons. In: Advances in neural information processing systems, pp 103–109. Cambridge, MA: MIT.

Barbour DL, Wang X (2003) Contrast tuning in auditory cortex. Science 299:1073–1075.

Belin P, Fecteau S, Bédard C (2004) Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165.

Brainard MS, Doupe AJ (2013) Translating birdsong: songbirds as a model for basic and applied medical research. Annu Rev Neurosci 36:489–517.

Bregman MR, Patel AD, Gentner TQ (2016) Songbirds use spectral shape, not pitch, for sound pattern recognition. Proc Natl Acad Sci U S A 113:1666–1671.

Calabrese A, Woolley SM (2015) Coding principles of the canonical cortical microcircuit in the avian brain. Proc Natl Acad Sci U S A 112:3517–3522.

Catz N, Noreña AJ (2013) Enhanced representation of spectral contrasts in the primary auditory cortex. Front Syst Neurosci 7:21.

Chase SM, Young ED (2007) First-spike latency information in single neurons increases when referenced to population onset. Proc Natl Acad Sci U S A 104:5175–5180.

Cooke JE, King AJ, Willmore BD, Schnupp JW (2018) Contrast gain control in mouse auditory cortex. J Neurophysiol 120:1872–1884.

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. Science 280:1439–1443.

Dugas-Ford J, Rowell JJ, Ragsdale CW (2012) Cell-type homologies and the origins of the neocortex. Proc Natl Acad Sci U S A 109:16974–16979.

Duifhuis H, Willems LF, Sluyter RJ (1982) Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. J Acoust Soc Am 71:1568–1580.

Elie JE, Theunissen FE (2016) 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.

Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. PLoS Comput Biol 5:e1000302.

Feng L, Wang X (2017) Harmonic template neurons in primate auditory cortex underlying complex sound processing. Proc Natl Acad Sci U S A 114:E840–E848.

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

Gaudrain E (2016) Vocoder v1.0. Available at https://github.com/egaudrain/vocoder.

Gonzalez J, Oliver JC (2005) Gender and speaker identification as a function of the number of channels in spectrally reduced speech. J Acoust Soc Am 118:461–470.

Harris KD, Mrsic-Flogel TD (2013) Cortical connectivity and sensory coding. Nature 503:51–58.

Holdgraf CR, de Heer W, Pasley B, Rieger J, Crone N, Lin JJ, Knight RT, Theunissen FE (2016) Rapid tuning shifts in human auditory cortex enhance speech intelligibility. Nat Commun 7:13654.

Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF (2016) Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J Neurosci 36:2014–2026.

Kelley DB, Nottebohm F (1979) Projections of a telencephalic auditory nucleus, field L, in the canary. J Comp Neurol 183:455–469.

Kim KH, Kim SJ (2000) Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. IEEE Trans Biomed Eng 47:1406–1411.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158.

Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA (2009) Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. J Neurosci 29:2283–2296.

Lohr B, Dooling RJ (1998) Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). J Comp Psychol 112:36–47.

Lopez Espejo M, Schwartz ZP, David SV (2019) Spectral tuning of adaptation supports coding of sensory context in auditory cortex. PLoS Comput Biol 15:e1007430.

McDermott JH, Ellis DP, Kawahara H (2012) Inharmonic speech: a tool for the study of speech perception and separation. In: SAPA-SCALE Conference.

McPherson MJ, McDermott JH (2018) Diversity in pitch perception revealed by task dependence. Nat Hum Behav 2:52–66.

Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. J Neurosci 32:15158–15168.

Mello CV, Vates GE, Okuhata S, Nottebohm F (1998) Descending auditory pathways in the adult male zebra finch (*Taeniopygia guttata*). J Comp Neurol 395:137–160.

Moore JM, Woolley SMN (2019) Emergent tuning for learned vocalizations in auditory cortex. Nat Neurosci 22:1469–1476.

Mouterde SC, Theunissen FE, Elie JE, Vignal C, Mathevon N (2014) Acoustic communication and sound degradation: how do the individual signatures of male and female zebra finch calls transmit over distance? PLoS One 9:e102842.

Nogueira W, Rode T, Büchner A (2016) Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants. J Acoust Soc Am 139:728–739.

Norman-Haignere S, Kanwisher N, McDermott JH (2013) Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. J Neurosci 33:19451–19469.

Norman-Haignere SV, Albouy P, Caclin A, McDermott JH, Kanwisher NG, Tillmann B (2016) Pitch-responsive cortical regions in congenital amusia. J Neurosci 36:2986–2994.

Nourski KV, Steinschneider M, Rhone AE, Kovach CK, Kawasaki H, Howard MA 3rd (2019) Differential responses to spectrally degraded speech within human auditory cortex: an intracranial electrophysiology study. Hear Res 371:53–65.

Oxenham AJ (2018) How we hear: the perception and neural coding of sound. Annu Rev Psychol 69:27–50.

Perez EC, Elie JE, Boucaud IC, Crouchet T, Soulage CO, Soula HA, Theunissen FE, Vignal C (2015) Physiological resonance between mates through calls as possible evidence of empathic processes in songbirds. Horm Behav 75:130–141.

Perrodin C, Kayser C, Logothetis NK, Petkov CI (2014) Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. J Neurosci 34:2524–2537.

Podos J (1997) A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution 51:537–551.

Popham S, Boebinger D, Ellis DPW, Kawahara H, McDermott JH (2018) Inharmonic speech reveals the role of harmonicity in the cocktail party problem. Nat Commun 9:2122.

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

Rabinowitz NC, Willmore BD, Schnupp JW, King AJ (2011) Contrast gain control in auditory cortex. Neuron 70:1178–1191.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111–114.

Riede T, Goller F (2010) Peripheral mechanisms for vocal production in birds: differences and similarities to human speech and singing. Brain Lang 115:69–80.

Rieke FB, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc Biol Sci 262:259–265.

Scheffers MT (1983) Simulation of auditory analysis of pitch: an elaboration on the DWS pitch meter. J Acoust Soc Am 74:1716–1725.

Schneider DM, Woolley SM (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. Neuron 79:141–152.

Schreiner CE, Calhoun BM (1994) Spectral envelope coding in cat primary auditory cortex: properties of ripple transfer functions. Audit Neurosci 1:39–61.

Schumacher JW, Schneider DM, Woolley SM (2011) Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. J Neurophysiol 106:500–514.

Seyfarth RM, Cheney DL (2017) The origin of meaning in animal signals. Anim Behav 124:339–346.

Shamma SA (1985) Speech processing in the auditory system II: lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J Acoust Soc Am 78:1622–1632.

Shamma SA, Versnel H, Kowalski N (1994) Ripple analysis in ferret primary auditory cortex: 1. Response characteristics of single units to sinusoidally rippled spectra. Audit Neurosci 1:233–254.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Shepard KN, Lin FG, Zhao CL, Chong KK, Liu RC (2015) Behavioral relevance helps untangle natural vocal categories in a specific subset of core auditory cortical pyramidal neurons. J Neurosci 35:2636–2645.

Simmons JA, Megela Simmons A (2011) Bats and frogs and animals in between: evidence for a common central timing mechanism to extract periodicity pitch. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 197:585–594.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am 114:3394–3411.

Smith EC, Lewicki MS (2006) Efficient auditory coding. Nature 439:978–982.

Soltis J (2010) Vocal communication in African elephants (*Loxodonta africana*). Zoo Biol 29:192–209.

Tchernichovski O, Nottebohm F, Ho CE, Bijan P, Mitra PP (2000) A procedure for an automated measurement of song similarity. Animal Behaviour 59:1167–1176.

Theunissen FE, Woolley SM, Hsu A, Fremouw T (2004) Methods for the analysis of auditory processing in the brain. Ann N Y Acad Sci 1016:187–207.

Titze IR (2017) Human speech: a restricted use of the mammalian larynx. J Voice 31:135–141.

Vates GE, Broome BM, Mello CV, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taeniopygia guttata*). J Comp Neurol 366:613–642.

Vicario DS, Naqvi NH, Raksin JN (2001) Sex differences in discrimination of vocal communication signals in a songbird. Anim Behav 61:805–817.

Vignal C, Mathevon N (2011) Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*). J Comp Psychol 125:150–161.

Vignal C, Mathevon N, Mottin S (2004) Audience drives male songbird response to partner's voice. Nature 430:448–451.

Vignal C, Mathevon N, Mottin S (2008) Mate recognition by female zebra finch: analysis of individuality in male call and first investigations on female decoding process. Behav Processes 77:191–198.

Wang X (2013) The harmonic organization of auditory cortex. Front Syst Neurosci 7:114.

Wang X, Walker KM (2012) Neural mechanisms for the abstraction and use of pitch information in auditory cortex. J Neurosci 32:13339–13342.

Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435:341–346.

Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. Proc Natl Acad Sci U S A 107:12676–12681.

Wild JM, Karten HJ, Frost BJ (1993) Connections of the auditory forebrain in the pigeon (*Columba livia*). J Comp Neurol 337:32–62.

Winn MB, Litovsky RY (2015) Using speech sounds to test functional spectral resolution in listeners with cochlear implants. J Acoust Soc Am 137:1430–1442.

Woolley SM, Moore JM (2011) Coevolution in communication senders and receivers: vocal behavior and auditory processing in multiple songbird species. Ann N Y Acad Sci 1225:155–165.

Woolley SMN (2017) Early experience and auditory development in songbirds. In: Auditory development and plasticity, pp 193–217. Cham, Switzerland: Springer.

Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci 8:1371–1379.

Yanagihara S, Yazaki-Sugiyama Y (2016) Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. Nat Commun 7:11946.

Yost WA (1986) Processing of complex signals and the role of inhibition. In: Auditory frequency selectivity, pp 361–370. Boston: Springer.

Zann R (1996) The zebra finch: A synthesis of field and laboratory studies. New York: Oxford UP.