

Imaging Auditory Representations of Song and Syllables in Populations of Sensorimotor Neurons Essential to Vocal Communication

Wendy Y.X. Peh, Todd F. Roberts, and Richard Mooney

Department of Neurobiology, Duke University School of Medicine, Durham, North Carolina 27710

Vocal communication depends on the coordinated activity of sensorimotor neurons important to vocal perception and production. How vocalizations are represented by spatiotemporal activity patterns in these neuronal populations remains poorly understood. Here we combined intracellular recordings and two-photon calcium imaging in anesthetized adult zebra finches (*Taeniopygia guttata*) to examine how learned birdsong and its component syllables are represented in identified projection neurons (PNs) within HVC, a sensorimotor region important for song perception and production. These experiments show that neighboring HVC PNs can respond at markedly different times to song playback and that different syllables activate spatially intermingled PNs within a local ($\sim 100 \mu\text{m}$) region of HVC. Moreover, noise correlations were stronger between PNs that responded most strongly to the same syllable and were spatially graded within and between classes of PNs. These findings support a model in which syllabic and temporal features of song are represented by spatially intermingled PNs functionally organized into cell- and syllable-type networks within local spatial scales in HVC.

Key words: auditory encoding; calcium imaging; functional organization; sensorimotor area; songbird; vocal communication

Introduction

Complex behavioral and perceptual processes depend on coordinated patterns of activity in neuronal populations, and these patterns can vary over space and time. Consequently, a major challenge is to relate spatiotemporal activity patterns of neurons within functionally relevant circuits to units of behavior and perception. Birdsong provides a powerful context in which to explore this relationship (Doupe and Kuhl, 1999; Mooney, 2009). Songbirds learn to sing sequences of syllables, the spectral structure and order of which distinguish different individuals and species (Marler and Tamura, 1964; Immelmann, 1969; Konishi, 1985; Hultsch and Todt, 1989). Moreover, the songbird brain contains a well-described sensorimotor network important to song production and perception (Wild, 2004), providing an anatomical framework for exploring how neuronal populations encode communication signals.

The telencephalic nucleus HVC resides near the apex of this sensorimotor network, and HVC lesions can impair singing and

song perception (Nottebohm et al., 1976; Brenowitz, 1991; Vu et al., 1994; Gentner et al., 2000; Thompson et al., 2007). Consistent with a sensorimotor role in vocal communication, many HVC projection neurons (PNs) fire temporally precise action potential bursts during singing and during auditory presentation of the bird's own song (BOS). Certain HVC PNs display complex properties, including sensorimotor mirroring and auditory selectivity for songs, syllables, or syllable combinations, consistent with the idea that, as a population, HVC neurons encode song and syllable representations (McCasland and Konishi, 1981; Margoliash, 1983; Lewicki and Konishi, 1995; Prather et al., 2008; Prather et al., 2009).

Although HVC's song-related properties have been extensively analyzed using electrophysiological recordings (Margoliash, 1983; Mooney, 2000; Hahnloser et al., 2002; Long et al., 2010; Amador et al., 2013), these methods are poorly suited for relating a neuron's activity to the functional organization of surrounding neurons. Moreover, HVC contains at least three cell types, including one PN (HVC_{RA}) that innervates the song motor nucleus RA, another PN (HVC_{X}) that innervates a striatopallidal region (Area X), and interneurons (HVC_{I}) (see Fig. 1A). Because these three cell types are spatially intermingled and HVC_{I} dominate multiunit recordings (compare Yu and Margoliash, 1996; Hahnloser et al., 2002), electrophysiological recordings of HVC population activity are challenging to interpret. Lastly, although axonal connections and singing-related activity patterns may be organized preferentially along HVC's rostrocaudal axis (Stauffer et al., 2012; Day et al., 2013), how this organization relates to auditory representations of song remains unknown.

In vivo two-photon (2p) calcium imaging has been used in various systems to examine spatiotemporal activity patterns in

Received June 6, 2014; revised Feb. 25, 2015; accepted March 1, 2015.

Author contributions: W.Y.X.P. and R.M. designed research; W.Y.X.P., T.F.R., and R.M. performed research; W.Y.X.P. analyzed data; W.Y.X.P., T.F.R., and R.M. wrote the paper.

This work was supported by National Institutes of Health DC02524 and National Science Foundation IOS-0821914 and IOS-354962 to R.M. and Howard Hughes Medical Institute International Student Research Fellowship to W.Y.X.P. We thank Kosuke Hamaguchi, Ian Davison, and Colton Brown for help with developing the software for data analysis; and Katherine Tschida and David Schneider for providing useful comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Richard Mooney, Department of Neurobiology, Duke University School of Medicine, Durham, NC 27710. E-mail: mooney@neuro.duke.edu.

T.F. Roberts' present address: Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, TX 75390.

DOI:10.1523/JNEUROSCI.2308-14.2015

Copyright © 2015 the authors 0270-6474/15/355589-17\$15.00/0

neuronal populations with cellular resolution (Stosiek et al., 2003; Ohki et al., 2005; Kerr et al., 2007; Sato et al., 2007; Bandyopadhyay et al., 2010; Rothschild et al., 2010; Graber et al., 2013). Here we combined intracellular electrophysiological recordings with 2p calcium imaging in the anesthetized male zebra finch to explore how song and its component syllables are spatially represented in populations of identified HVC neurons. These experiments provide evidence that syllabic and temporal features of song are represented by spatially intermingled PN networks organized into cell- and syllable-type networks.

Materials and Methods

Subjects. All experiments were performed in accordance with a protocol approved by Duke University Institutional Animal Care and Use Committee. Results were collected from a total of 51 adult (>90 d post hatch) male zebra finches (*Taeniopygia guttata*).

Retrograde labeling of HVC projection neurons. For $n = 48$ birds, 7% AlexaFluor-594 dextran (10,000 MW, Invitrogen) or red fluorescent latex beads (Lumafuor) were injected into Area X or RA, respectively, 4–7 d before imaging experiments. For $n = 3$ birds, dextran was injected into RA to make soma area measurements of HVC_{RA} cells. Birds were anesthetized with isoflurane inhalation (2%) and placed in a stereotaxic apparatus. Injections were performed based on stereotaxic coordinates. A total volume of 161 nl of dextran or 97 nl of latex beads was injected into each site (divided into pulses of 32.2 nl delivered at 30 s intervals) using a glass pipette attached to a Nanojet II (Drummond Scientific). After the injection, the craniotomy was covered with bone wax (Ethicon) and the scalp wound was closed with cyanoacrylate (Vetbond, 3M). Birds were warmed under a heat lamp during the recovery from anesthesia.

Brain slice preparation. Birds ($n = 17$) were deeply anesthetized with isoflurane inhalation, decapitated, and their brains quickly removed and placed in oxygenated ice-cold sucrose ACSF. Brains were cut into 350–500 μm horizontal or sagittal slices using a vibrating microtome (Leica). Slices containing HVC were recovered in a submerged holding chamber containing oxygenated standard ACSF solution at room temperature. Standard ACSF solution (in mM) consists of the following: 119 NaCl, 2.5 KCl, 1.3 MgCl_2 , 2.5 CaCl_2 , 1 NaH_2PO_4 , 26.2 NaHCO_3 , and 11 glucose, equilibrated with 95% O_2 /5% CO_2 . For sucrose ACSF, NaCl was substituted with equimolar sucrose. After at least an hour of recovery, slices containing HVC were transferred to a submerged slice chamber (30°C) for simultaneous intracellular recordings and 2p calcium imaging.

Preparatory surgery for in vivo imaging. Birds were anesthetized with an intramuscular injection of 20% urethane (3 doses of 30 μl at 30 min intervals) and placed in a stereotaxic apparatus. A small amount of isoflurane anesthesia was used to supplement the urethane anesthesia as needed. In one bird, 45 μl of diazepam was used instead of urethane. As there was no observed difference in electrophysiological or imaging results collected, data from the diazepam-anesthetized bird were included in the full dataset. Lidocaine was first applied to the scalp, and excess scalp lateral to HVC was retracted and excised. Scalp margins were then secured to the skull with cyanoacrylate. Subsequently, a stainless steel head post was secured to the rostral part of the skull using dental cement. To facilitate 2p imaging using a water-immersion microscope objective lens, a well of dental cement was also created around the retracted scalp. A large area (3 \times 3 mm) of outer leaflet of the skull at the stereotaxic coordinates over HVC was removed, and HVC was located through the inner leaflet under a 10 \times objective lens (Carl Zeiss, Plan-Neofluar, 0.30 NA) of a microscope (Carl Zeiss, LSM510) via visualization of fluorescent retrograde labeling of HVC PN. A small craniotomy (\sim 800 μm \times 800 μm) was then made over HVC, and the dura was carefully removed. Exposed brain was kept hydrated by the application of saline. In most experiments, an imaging window was created using a small custom-cut glass coverslip (thickness #0, Electron Microscopy Sciences) that was placed over the craniotomy leaving a small caudal edge (\sim 200 μm in width) exposed to allow the entry of either a sharp intracellular electrode or a dye delivery pipette. The edges of the coverslip were covered with a small amount of bone wax before the application of gel cyanoacrylate and dental cement to secure the coverslip to the skull. In a subset of popula-

tion imaging experiments, the craniotomy was first covered by a small amount of Kwik-Sil (World Precision Instruments), which allowed for dye delivery pipette entry into the brain without clogging; following dye injection, the Kwik-Sil was removed and a coverslip was placed over the craniotomy for imaging. After surgery, birds were transferred to a customized head-post holder under the 2p microscope, leaving the ears unobstructed. A heating pad (37°C) was used to warm birds during experiments.

Intracellular sharp electrode recording and filling of HVC neurons with Oregon Green 488 BAPTA-1 (OGB-1). Sharp electrode pipettes (borosilicate glass, BF100–50–10; Sutter Instrument) were pulled to yield a resistance of 80–120 $\text{M}\Omega$ when the tips were filled with 5 mM OGB-1 (hexapotassium salt, Invitrogen) and 2 M potassium acetate. Pipettes were advanced into HVC using a hydraulic manipulator (Soma Scientific Instruments). Intracellular penetrations were achieved using previously described methods (Mooney, 2000). An AxoClamp 2B intracellular amplifier (Molecular Devices) was used in bridge mode to record intracellular membrane potentials, which were low-pass filtered at 3 kHz and digitized at 10 kHz. Neurons were filled with OGB-1 within 2–5 min after the initial penetration by the sharp electrode, and the extent of cell filling was visualized under 2p imaging. Different HVC cell types were identified based on the presence of retrograde labeling and/or their intrinsic electrophysiological properties for *in vitro* and *in vivo* experiments and, for *in vivo* experiments, by their song-evoked electrophysiological properties (Dutar et al., 1998; Mooney, 2000; Mooney and Prather, 2005). Electrophysiological data from each trial were recorded as an epoch beginning 2 s before the auditory stimulus and ending 3 s after the end of the auditory stimulus. For brain slice experiments, a family of 500 ms duration positive (50–2500 pA) current pulses was delivered through the recording electrode at \sim 5.5 to 7.5 s intertrial intervals, and the resultant membrane potential responses were collected using custom LabVIEW software. Time stamps of the current pulses from each trial were sent from LabVIEW to synchronize with the imaging software (Carl Zeiss AIM software, version 3.2). Data were subsequently analyzed using custom MATLAB software (written by K. Hamaguchi, R.M. laboratory).

Loose cell-attached recording. Patch electrode pipettes (borosilicate glass, G75150T-4; Warner Instruments) were pulled to yield a resistance of 5–9 $\text{M}\Omega$ when filled with an extracellular solution containing 150 mM NaCl, 2.5 mM KCl, and 10 mM HEPES, pH 7.4. To aid with visualizing the pipette tip, the pipette was either coated with gold nanoparticles (courtesy of Dr. Mladen Barbic, Howard Hughes Medical Institute Janelia Farms Research Campus) or 20 μM AlexaFluor-568 hydrazide solution was included in the pipette solution. The pipette was advanced into HVC using a motorized manipulator (Scientifica). Positive pressure (20–80 mbar) was applied using a syringe while the pipette was advanced toward an OGB-1 labeled HVC neuron in small increments ($<$ 5 μm). The resistance change at the pipette tip was measured with a voltage pulse of -5 mV using an AxoClamp 2B intracellular amplifier (Molecular Devices), and a 2- to 10-fold increase of the seal resistance was sufficient to record action currents when in voltage-clamp mode at 0 mV holding potential. Signals were low-pass filtered at 3 kHz, digitized at 10 kHz, and analyzed using custom MATLAB software (written by K. Hamaguchi and W.Y.X.P.).

“Bulk-loading” of HVC neurons with OGB-1 AM dye. To prepare the dye mixture for labeling populations of HVC neurons, 4 μl of 20% pluronic acid in DMSO (F-127, Invitrogen) was first added to one vial of OGB-1 AM (O-6807, 50 μg ; Invitrogen) and vortexed for 2–3 min. The mixture was then diluted with 26 μl of a solution containing 150 mM NaCl, 2.5 mM KCl, 10 mM HEPES, pH 7.4, and 0.8 μl of 2 mM AlexaFluor-568 hydrazide solution (for visual monitoring of the injection) to yield a final concentration of 1.32 mM OGB-1 AM. The mixture was further vortexed for 2–3 min before sonication in an ice water bath for 20 min. In one bird, the dye mixture was also supplemented with 0.2 μl of a surfactant-Cremaphor EL (C5135, Sigma). After sonication, the dye mixture was filtered through a 0.22- μm -pore centrifuge filter (Millipore, UFC30GV25) and kept on ice until use. Glass borosilicate pipettes (1B120F-4, World Precision Instruments) were pulled using a vertical pipette puller (700C, David Kopf Instruments) to achieve a long taper with a tip diameter of \sim 5 μm , and pipettes were back filled with the dye

mixture. Using a hydraulic manipulator, the pipette was advanced slowly into the HVC on the right hemisphere to the dye injection site, which was chosen based on the presence of red fluorescent retrograde labeling of HVC_x cells, an absence of major blood vessels, and a depth of <300 μm from the surface of the brain. Low-volume (<1 nl) injections were performed during pipette travel in the brain to prevent clogging. At the target area, a total of 180 pulses of dye (total injection volume was ~100–200 nl) were ejected from the pipette tip delivered at 15 s intervals with pulse durations varying between 10 and 100 ms at 4 psi using a Picrospritzer II (General Valve). The pipette was withdrawn from the brain 5 min after the end of the injection protocol to minimize efflux through the pipette track and facilitate dye absorption by surrounding tissue. Imaging of “bulk-loaded” cells was performed at least 1 h after the end of dye injection to allow for sufficient time for dye uptake by cells.

2p calcium imaging. Imaging of OGB-1-labeled HVC cells was performed at 810–820 nm with a 2p laser scanning microscope (Carl Zeiss LSM 510, Carl Zeiss, 40 \times IR-Achroplan 0.80 NA water-immersion objective lens) with a mode-locked titanium:sapphire laser system (Tsunami, Spectra-Physics). Calcium signals were bandpass-filtered between 470 and 560 nm. AlexaFluor-594 dextrans or red dextran beads were simultaneously excited and bandpass-filtered between 570 and 640 nm before detection by a second photomultiplier tube. Laser power and detector gain were adjusted accordingly to prevent saturation of fluorescent signals. Small-field images (38 \times 38 μm fields, 128 \times 128 pixels) were collected at 10 Hz (dwell time per pixel: 2.56 μs), whereas large field images (115 \times 115 μm fields, 256 \times 256 pixels) were collected at 2.5 Hz (dwell time per pixel: 2.56 μs). Imaging data were collected continuously until all trials of a given set of auditory stimuli or all pulses in a current pulse family were delivered. Imaging data were subsequently transferred to ImageJ (National Institutes of Health) and custom-written MATLAB programs (K. Hamaguchi and W.Y.X.P.) for analysis.

Location of imaged region within HVC. At the end of population calcium imaging experiments, birds were deeply anesthetized with isoflurane inhalation, decapitated, and their brains were rapidly removed from the skull and mounted in an upright position for slicing in standard ACSF solution at room temperature to obtain horizontal tissue slabs containing HVC (>800 μm thickness). Slabs containing HVC were placed on a glass slide for imaging of red retrograde tracer labeling and green OGB-1 labeling using a fluorescence microscope (Carl Zeiss Axioskop, 5 \times Plan-Neofluar, 0.15 NA objective lens), allowing us to determine the location of the imaged areas relative to HVC's boundaries. Images from each HVC were then uniformly scaled to a standard size in order to compare the position of imaged regions across birds.

Calcium imaging data analysis. Calcium imaging data were converted into tiff stacks in ImageJ and imported into MATLAB for motion correction, selection of ROIs, and calculation of $\Delta\text{F}/\text{F}$ in response to auditory stimuli or, for brain slice experiments, current injection pulses. For motion correction, x and y translational shifts between individual frames of an imaging session at 10 frame intervals were first determined by an image correlation algorithm based on signals from either the red or green channel (MATLAB, written by I. Davison). Motion correction between frames to a reference image was then applied to both the green and red channels when pixel shifts were >3 pixels in either the x or y direction. The reference image for correction was obtained by averaging frames (typically 100–1000 frames) from a scanning period with no overt movement during the imaging session. Imaging sessions with gross movement artifact of pixel shifts of >10 pixels were excluded from any subsequent analysis.

Cell bodies were manually selected as ROIs based on visualized boundaries of red and/or green channels. The fluorescence signal of each ROI in a frame was obtained by averaging all the pixels within the ROI. To average calcium responses across trials and to align imaging data with electrophysiological recordings, calcium signals from each trial were up-sampled by linear interpolation to the same sampling rate as the electrophysiological recordings (i.e., 10 kHz). The calcium responses or $\Delta\text{F}/\text{F}$ values for each ROI were calculated by the following: $(F_t - F_{\text{baseline}}) / F_{\text{baseline}} \times 100\%$, where F_t is the fluorescence at each point in time and F_{baseline} is the average fluorescence during a baseline period before the delivery of each auditory stimulus or current injection pulse. The base-

line period was set to equal the duration of the ensuing auditory stimulus or, in the case of brain slice experiments, a 500 ms period before each current pulse. For the playback of a single song motif or syllable, the average calcium response to a stimulus was defined as the mean $\Delta\text{F}/\text{F}$ across trials during a 600 ms period that spanned from 100 ms before the peak response to 500 ms after the peak response. This averaging of signals around the peak response was implemented to avoid any underestimation of response due to the sparse nature of action potential activity of HVC PNs. The peak response was determined as the maximum value of the average calcium response within a time window that spanned from the stimulus onset to 300 ms after stimulus offset. In those instances where the auditory stimulus consisted of multiple song motifs or multiple iterations of a single syllable, the average calcium response was defined as the mean $\Delta\text{F}/\text{F}$ across trials from stimulus onset to 300 ms following stimulus offset. For the playback of multiple song motifs, the peak response was determined as the maximum value of the average calcium response within a time window that included the entire stimulus and ended 300 ms after stimulus offset. For the calculation of the decay rate of calcium responses to the playback of multiple motifs, the response peak was defined as the maximum value of the average calcium response in a window that began with the onset of the last motif and ended 300 ms after stimulus offset. For brain slice experiments, the average calcium response was defined as the mean $\Delta\text{F}/\text{F}$ across trials measured from the start of the current pulse to 300 ms after the end of the current pulse. The peak response was determined as the maximum value of the average calcium response during a time window that began with the start of the current pulse and ended 300 ms after the end of the current pulse. To determine the statistical significance of calcium responses, the average $\Delta\text{F}/\text{F}$ response was compared with the $\Delta\text{F}/\text{F}$ measured during the baseline period before the start of the stimulus using a two-tailed paired t test. A response was deemed significant if $p < 0.05$ and if, at any time, the stimulus response was at least 3 SDs greater than the average baseline $\Delta\text{F}/\text{F}$.

The decay rate of BOS-evoked calcium signals was quantified as the percentage decrease in the calcium signal (ΔF , %) divided by the time (Δt , seconds) required for a response peak to either return to the baseline value or, in those cases in which the response remained elevated beyond the response window, the time from the response peak to the end of the recording epoch, whichever occurred first. For the playback of multiple BOS motifs, the response peak for the calculation of the decay rate was identified from the onset of the last motif to 300 ms following stimulus offset.

To calculate the difference in timing of the peak BOS-evoked calcium responses among PNs imaged within the same small field (38 \times 38 μm), we restricted peak time measurements to BOS-selective PNs that showed one distinct peak in their BOS-evoked response. Distinct calcium response peaks were identified as instances when the first derivative of the averaged $\Delta\text{F}/\text{F}$ trace crossed a defined threshold (mean baseline + 3 SDs) and aligned with the maximum value in the $\Delta\text{F}/\text{F}$ trace. This calcium response peak time measurement criterion was also applied to PNs to determine the correspondence between the timing of the peak calcium response and the peak action potential firing rate.

To calculate the trial-to-trial noise pairwise correlation coefficient in BOS-evoked calcium signals from pairs of PNs, the average BOS-evoked $\Delta\text{F}/\text{F}$ calculated from all playback trials was subtracted from the $\Delta\text{F}/\text{F}$ in each individual trial. The mean-subtracted traces (spanning a time window from the onset of the BOS stimulus to 300 ms after stimulus offset) were compared on a trial-by-trial basis between PNs using MATLAB's `corrcoef` function (MathWorks) and averaged to obtain an average noise correlation coefficient value for each pair of PNs. For controls, spatial positions were randomly shuffled among PNs within each image field before the calculation of the interpair distance and correlation coefficient. Random resampling of noise correlation coefficients for pairs of PNs was performed using MATLAB's `datsample` function in the Statistics Toolbox (MathWorks).

Quantification of soma area. Soma areas were measured (ImageJ) based on the boundaries of retrograde tracer labeling within identified HVC_x or HVC_{RA} cells visualized at a single plane within a z -stack of images (1024 \times 1024 pixels) collected from *in vivo* 2p imaging. The

chosen plane was the one that showed the largest amount of dextran labeling within the soma. Soma area measurements of retrograde tracer-identified PNs showed a largely bimodal size distribution between the 2 PN types, which is similar to measurements from prior studies (Wild et al., 2005; Tschida and Mooney, 2012). Putative PNs with soma area of $<100 \mu\text{m}^2$ are classified as putative HVC_{RA} , whereas those $>165 \mu\text{m}^2$ are classified as putative HVC_{X} cells. Based on these measurements, 95% of all sampled cells with soma areas $>165 \mu\text{m}^2$ are HVC_{X} cells (i.e., false positive rate $<5\%$).

Firing rate analysis. A neuron's average firing rate in response to a particular auditory stimulus or current pulse (in the case of brain slice experiments) was calculated based on the number of action potentials evoked by the playback or current pulse, respectively, divided by the duration of the stimulus or current pulse. Peristimulus time histograms (PSTHs) of the average firing rates were calculated from all trials of a given stimulus type using 10 ms bins. Peak timings of BOS-evoked firing activity were measured from PSTHs calculated using 100 ms bins. To test the statistical significance of changes in firing rates, the average firing rate during the stimulus period was compared with a baseline period before the stimulus (two-tailed paired *t* test, α level 0.05). A baseline period had equal duration as the ensuing auditory stimulus or current pulse.

Auditory stimuli. Before each *in vivo* experiment, songs were recorded from each bird in isolation in a sound-isolated chamber. Recordings were amplified and low-pass filtered at 10 kHz, digitized at 20 kHz, and further bandpass filtered between 350 and 10,000 Hz before being digitally edited to create the different song and syllable stimuli. A 10 ms ramp function was applied at the beginning and end of each stimulus using a cosine function (custom MATLAB, custom LabVIEW) to suppress acoustical transients. Song stimuli included the BOS, BOS played in reverse (REV), and BOS in which the syllable order had been reversed (REVSO). In those experiments where we examined syllable-evoked responses, individual syllables, or 10 concatenated repeats of individual syllables were presented as stimuli. Stimuli were presented in a fixed order and usually consisted of 15 trials of each stimulus, ranging from 7 to 40 trials for each stimulus. For small ($38 \times 38 \mu\text{m}$) field imaging experiments, either one or two song motifs, individual syllables, or 10 repeats of individual syllables were presented in each trial with an interstimulus interval that varied between 5.5 and 7.5 s. The onset of each auditory playback was slightly jittered from trial to trial relative to the continuous acquisition of imaging frames. For large ($115 \times 115 \mu\text{m}$) field imaging experiments, either 3 motifs of a song (150 ms intermotif interval) or 10 repeats of individual syllables (150 ms interrepeat intervals) were presented in each trial, with an interstimulus interval of 12.5 to 15.5 s. Stimuli were presented at ~ 70 dB sound pressure level, from a speaker located ~ 20 cm in front of the bird, measured with a sound level meter (A-weighted). Time stamps of auditory stimuli playbacks from each trial were sent from LabVIEW to synchronize with the imaging software (Carl Zeiss AIM software, version 3.2).

Estimating auditory selectivity. The auditory selectivity of an HVC cell for BOS over REV or BOS over REVSO was calculated using the *d'* metric (Green and Swets, 1966). Only HVC cells with significant BOS responses were included in this analysis. The *d'* values based on electrophysiological recording (*d'* FR) or calcium transients (*d'* Cal) were calculated based on the formulas: $d' \text{FR} = \frac{2(\bar{R}_{\text{BOSFR}} - \bar{R}_{\text{OtherFR}})}{\sqrt{\sigma_{\text{BOSFR}}^2 + \sigma_{\text{OtherFR}}^2}}$ or $d' \text{Cal} = \frac{2(\bar{R}_{\text{BOS}\Delta F/F} - \bar{R}_{\text{Other}\Delta F/F})}{\sqrt{\sigma_{\text{BOS}\Delta F/F}^2 + \sigma_{\text{Other}\Delta F/F}^2}}$, where "Other" indicates either REV or REVSO, \bar{R} is the mean song response (as described above), and σ^2 is its variance. A *d'* value of >0.5 was used as the criterion for identifying a cell as "BOS-selective" (Solis and Doupe, 1997). Auditory selectivity of an HVC neuron for a particular syllable over another syllable in the BOS was calculated using: $d' \text{SYLL} = \frac{2(\bar{R}_{\text{syll}\Delta F/F} - \bar{R}_{\text{Othersyll}\Delta F/F})}{\sqrt{\sigma_{\text{syll}\Delta F/F}^2 + \sigma_{\text{Othersyll}\Delta F/F}^2}}$, where \bar{R} is the mean syllable response and σ^2 is its variance. A *d'* value of >0.5 for a syllable over all other syllables was the criterion for identifying a cell's preferred syllable.

Results

Our overarching goal was to determine how song and its component syllables are spatially encoded in populations of HVC PNs. To this end, we first established the correspondence between action potential activity and calcium signals in identified HVC cell types. Using these data, we identified criteria for distinguishing different PN types and interneurons based on quantifiable differences in calcium signals as well as soma size measurements and confirmed that these distinctions held for bulk-loaded populations of HVC neurons imaged *in vivo*. We then imaged auditory-evoked activity in populations of bulk-loaded HVC neurons, allowing us to determine how song and syllable information is spatially encoded in populations of HVC PNs. Finally, we examined the HVC network for noise correlations that could reflect a greater abundance of shared inputs and/or synaptic interconnections between specific anatomical and/or functional classes of HVC PNs.

The correspondence between calcium transients and action potential activity in different HVC cell types

One challenge to interpreting calcium-imaging data is to understand the relationship between changes in fluorescence signals and changes in membrane potential of the imaged cell. To examine the relationship between auditory-evoked electrophysiological activity and calcium transients in identified HVC cell types, we combined sharp electrode intracellular current-clamp recordings and 2p imaging of the calcium-sensitive dye OGB-1 in anesthetized adult male zebra finches. We used these approaches to compare electrophysiological activity and calcium signals in individual neurons in response to auditory presentation of the BOS, REV, and REVSO. We identified the borders of HVC under a 2p microscope by visualizing PNs that had previously been retrogradely labeled with AlexaFluor-594 dextran or latex beads injected into either Area X or RA, respectively. We then positioned an intracellular electrode containing OGB-1 under the cranial window to target individual HVC neurons. After impaling a neuron with the pipette, we filled the cell with OGB-1 (Fig. 1B) and classified it as HVC_{X} , HVC_{RA} , or HVC_{I} based on its spontaneous, DC-evoked, and auditory-evoked electrophysiological properties and by the presence or absence of retrograde label (see Materials and Methods) (Dutar et al., 1998; Mooney, 2000; Mooney and Prather, 2005). We then obtained simultaneous intracellular current-clamp recordings and 2p images of calcium signals from the impaled cell (Fig. 1B).

For both HVC PN types, BOS playback evoked parallel increases in action potential firing rates and somatic calcium signals (Fig. 1B, C). Playback trials that evoked action potentials in HVC PNs always resulted in mean $\Delta F/F$ signals $>1\%$, whereas playback trials that evoked only subthreshold membrane depolarizations usually (21 of 23 PNs) failed to evoke any significant change in fluorescence and never evoked mean $\Delta F/F$ signals $>1\%$. In contrast to PNs, HVC_{I} cells showed only small changes in calcium signals despite firing at high rates to BOS playback (Fig. 1C, right, *D*; $n = 13 \text{ HVC}_{\text{X}}$, $n = 3 \text{ HVC}_{\text{RA}}$, and $n = 7 \text{ HVC}_{\text{I}}$). Indeed, 3 of 7 HVC_{I} cells did not show significant calcium responses to BOS playback, despite exhibiting significant firing rate changes. Within-cell normalizations revealed that average song-evoked action potential firing rates and calcium signals were strongly correlated in both HVC_{RA} and HVC_{X} cells but not for HVC_{I} cells (Fig. 1E; HVC_{X} : $r = 0.91$, $p = 2.4 \times 10^{-14}$; HVC_{RA} : $r = 0.75$, $p = 0.02$, linear regression; HVC_{I} : data not shown, $r = 0.29$, $p = 0.49$). Although peak BOS-evoked calcium signals were larger in HVC_{RA} cells than in HVC_{X} cells, the decay rates of the peak calcium

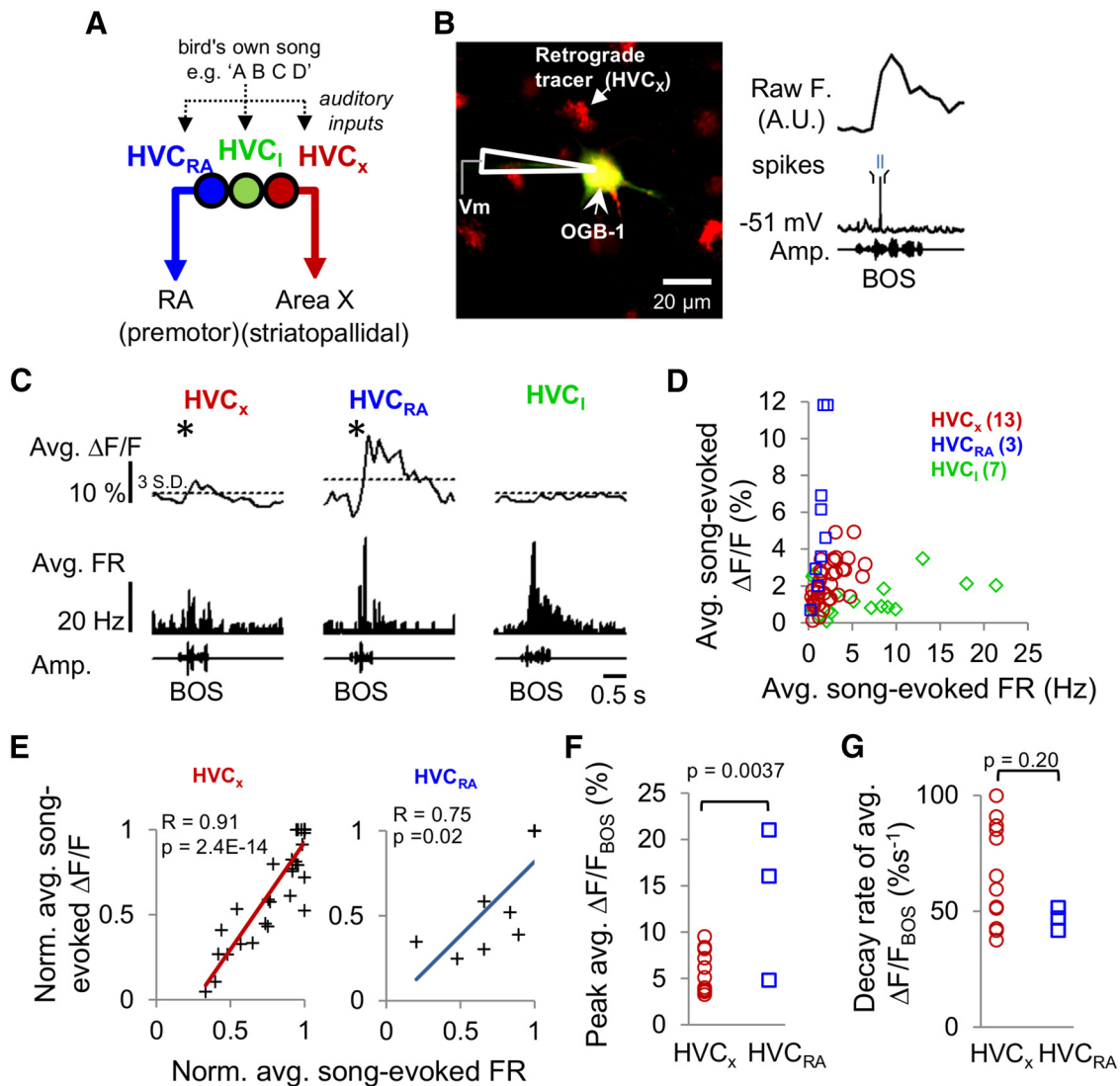


Figure 1. Calcium signals in HVC PN types are correlated with action potential activity in anesthetized zebra finches. **A**, HVC contains several distinct cell types that are spatially intermingled and synaptically interconnected, all of which receive auditory input from extrinsic sources. HVC_{RA} and HVC_X cells send long-range axonal projections to the robust nucleus of the arcopallium (RA) and the striatopallidal region Area X, respectively, whereas axons of inhibitory HVC interneurons (HVC_I cells) are restricted to HVC. **B**, Left, A 2p image showing a retrogradely labeled HVC_X cell filled with the calcium-sensitive dye OGB-1 (green) via a sharp electrode recording pipette. HVC_X cells were identified by retrograde labeling from Area X with AlexaFluor-594 dextran (red). Right, Intracellular fluorescence and membrane potential traces obtained from simultaneous calcium imaging and intracellular recording of an HVC_X cell during auditory playback of the BOS. **C**, Average somatic calcium responses ($\Delta F/F$) and average FR PSTHs (10 ms bin) evoked by BOS playback in three different HVC cell types: HVC_X cell (left), HVC_{RA} cell (middle), and HVC_I cell (right). Horizontal dashed line indicates 3 SDs of the baseline period measured before stimulus playback. *Significant responses. **D**, Average song-evoked $\Delta F/F$ plotted against average song-evoked firing rates for all three cell types (HVC_X : $n = 13$; HVC_{RA} : $n = 3$; HVC_I : $n = 7$). **E**, Normalized auditory-evoked calcium responses were positively correlated with normalized average auditory-evoked firing rates in HVC_X cells (left, $r = 0.91$, $p = 2.4E-14$, linear regression) and HVC_{RA} cells (right, $r = 0.75$, $p = 0.02$). Lines indicate linear fits. **F**, Mean peak BOS-evoked calcium responses of HVC_{RA} cells were significantly larger than those of HVC_X cells (HVC_X : $5.6 \pm 0.6\%$; HVC_{RA} : $14.0 \pm 4.8\%$; mean \pm SEM, two-tailed t test, $p = 0.00037$). **G**, Decay rates of peak BOS-evoked calcium signals were not significantly different between the two PN types (HVC_X : $64.3 \pm 6.1\%s^{-1}$; HVC_{RA} : $46.8 \pm 2.8\%s^{-1}$; mean \pm SEM, $p = 0.20$, two-tailed t test).

response were similar between the two PN types (Fig. 1 F, G; average peak calcium signal HVC_X : $5.6 \pm 0.6\%$, HVC_{RA} : $14.0 \pm 4.8\%$, $p = 0.00037$; decay rates HVC_X : $64.3 \pm 6.1\%s^{-1}$, HVC_{RA} : $46.8 \pm 2.8\%s^{-1}$, mean \pm SEM, $p = 0.20$, two-tailed t test).

Because *in vivo* imaging can be less sensitive because of light scattering and higher spontaneous activity, we also conducted a separate set of experiments using intracellular recordings and 2p-imaging methods in brain slices to further explore the correspondence between action potential activity and calcium-related changes in fluorescence in identified HVC neurons (Fig. 2 A, B). These experiments showed that DC-evoked action potential activity and calcium signals were positively correlated in both HVC PN types (Fig. 2 C; HVC_X : $r = 0.92$, $p = 2.7E-48$; HVC_{RA} : $r =$

0.86 , $p = 1.7E-11$, linear regression). The peak calcium signal in response to a single action potential and the decay rates of this calcium signal were not significantly different between HVC_{RA} and HVC_X cells (Fig. 2 D, E; average peak calcium signal HVC_X : $14.6 \pm 4.6\%$, HVC_{RA} : $21.7 \pm 9.2\%$, $p = 0.47$, decay rates HVC_X : $87.6 \pm 18.3\%s^{-1}$, HVC_{RA} : $56.7 \pm 15.0\%s^{-1}$, mean \pm SEM, $p = 0.31$, two-tailed t test). Similar to the *in vivo* experiments, these *in vitro* experiments also revealed that HVC_I cells show little or no change in somatic fluorescence even when driven with current injection to fire action potentials at rates ranging from 100 to 200 Hz (Fig. 2 B). These *in vivo* and *in vitro* experiments establish that 2p calcium imaging methods afford a suitable proxy for measuring action potential firing rates in

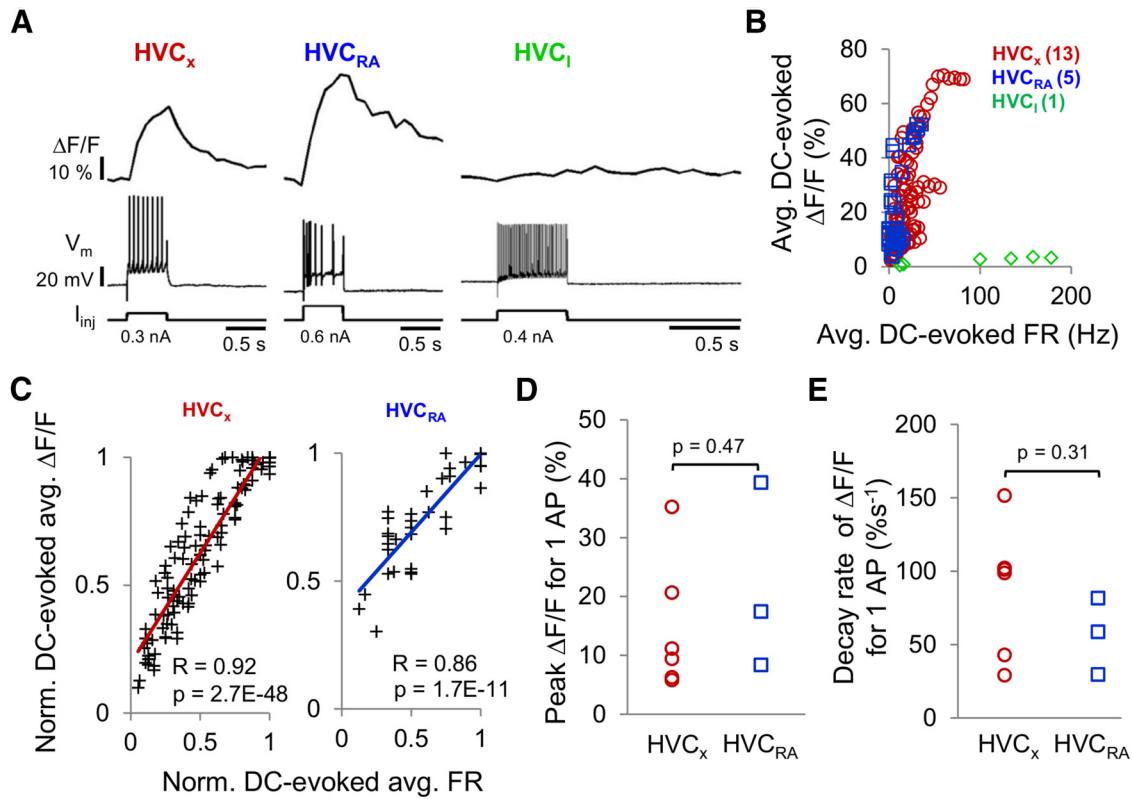


Figure 2. Calcium signals in HVC PN types are correlated with action potential activity in brain slice preparations. **A**, Examples of simultaneously recorded calcium signals ($\Delta F/F$) and action potentials evoked by positive, intrasomatic current injection into an HVC_x cell (left), HVC_{RA} cell (middle), and HVC_I cell (right). **B**, Average DC-evoked $\Delta F/F$ plotted against average DC-evoked firing rates for all three cell types (HVC_x: $n = 13$ cells, red circles; HVC_{RA}: $n = 5$ cells, blue squares; HVC_I: $n = 1$ cell, green diamonds). **C**, Normalized average DC-evoked $\Delta F/F$ was significantly correlated with normalized average FR in HVC_x cells (left, $r = 0.92$, $p = 2.7E-48$, linear regression) and in HVC_{RA} cells (right, $r = 0.86$, $p = 1.7E-11$). Calcium responses and firing rates were each normalized to their respective maximum values obtained within each cell. Lines indicate linear fits. **D**, **E**, Mean peak responses and decay rates of the peak calcium signals associated with DC-evoked single action potentials were not significantly different between the two PN types (peak $\Delta F/F$ -HVC_x: $14.6 \pm 4.6\%$, HVC_{RA}: $21.7 \pm 9.2\%$; decay rates HVC_x: $87.6 \pm 18.3\%s^{-1}$, HVC_{RA}: $56.7 \pm 15.0\%s^{-1}$; mean \pm SEM, two-tailed t test, both $p > 0.05$).

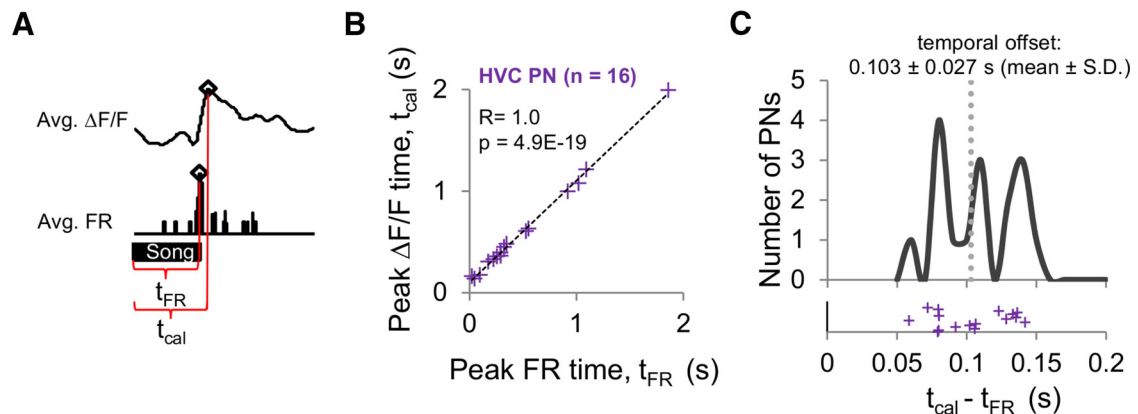


Figure 3. Timing of peak calcium responses was highly correlated with the timing of peak firing rates in HVC PN types. **A**, Measurements of when the peak in the average BOS-evoked calcium response (i.e., t_{cal}) and firing rate (i.e., t_{FR}) occurred were made relative to the onset of the song stimulus. The example calcium and electrophysiological responses shown here are from an HVC PN. Open diamonds represent peak values. **B**, Peak times of BOS-evoked calcium responses (t_{cal}) were linearly correlated with peak times measured from the average firing rate (t_{FR} ; PSTH: 100 ms bin, $n = 16$ HVC PN types, $r = 1.00$, $p = 4.9E-19$, linear regression). Dashed black line indicates the linear fit. **C**, Timing of peaks in BOS-evoked calcium transients lagged behind peaks in firing rates of HVC PN types by a mean temporal offset of 103 ms (gray vertical dashed line). Crosses below histogram indicate temporal offset values of individual PN types.

identified HVC PN types but not in HVC_I cells. These experiments also indicate that it is difficult to distinguish between HVC PN types by relying on calcium response magnitudes or decay kinetics.

In response to BOS playback, most HVC PN types in anesthetized zebra finches display phasic bursts of action potentials that are both temporally sparse and precise (Mooney, 2000). As a prelude

to population imaging in anesthetized birds, we therefore addressed whether peaks in BOS-evoked calcium transients could be used to accurately estimate peaks in action potential responses of HVC PN types. Indeed, by combining 2p imaging in small fields ($38 \times 38 \mu m$, 10 Hz scan rate) with sharp electrode recordings from identified HVC PN types, we determined that peaks in BOS-evoked calcium transients were strongly and linearly correlated

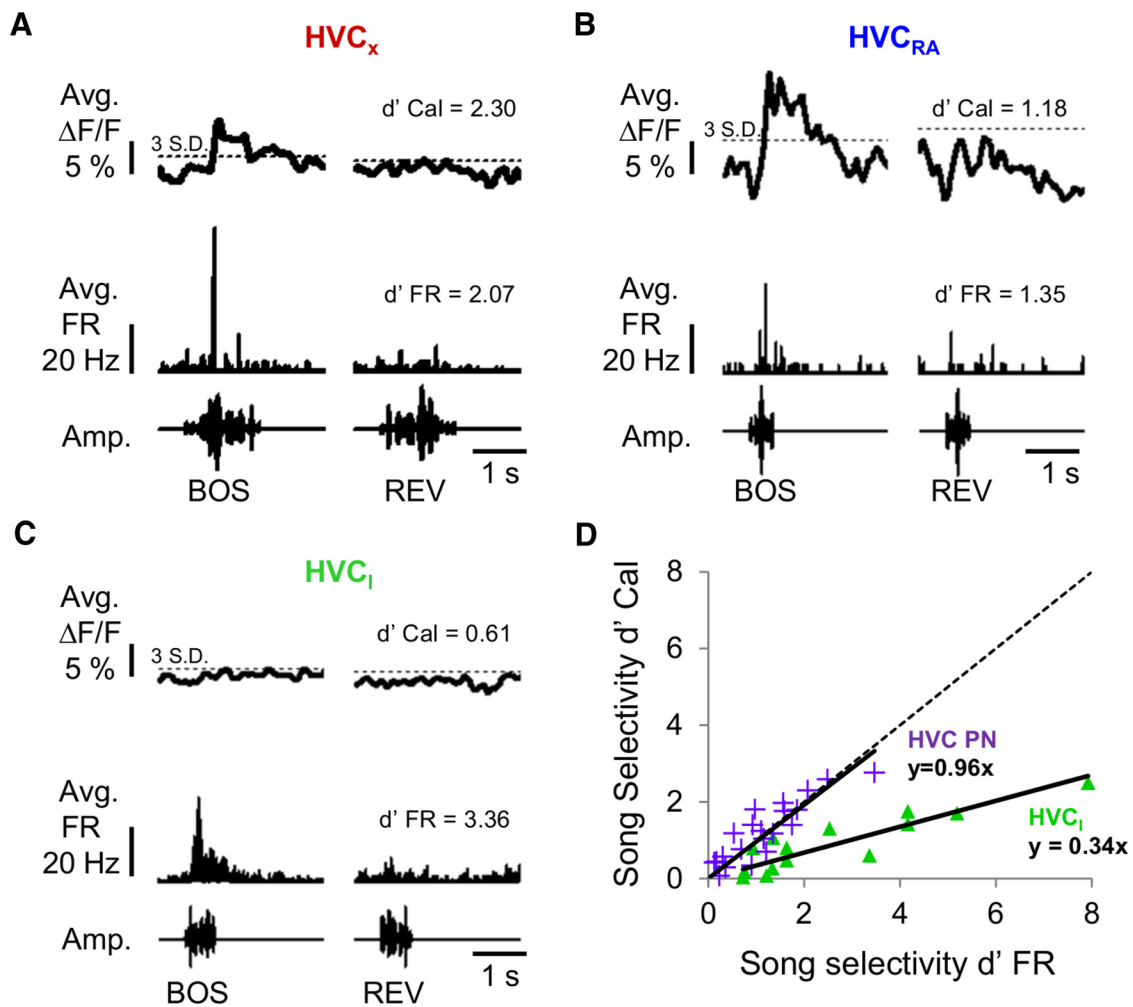


Figure 4. Electrical and optical measures of song selectivity were linearly correlated for HVC cells. **A–C**, Examples of corresponding average calcium responses and average firing rates evoked by BOS and REV playbacks obtained from each of the three cell types: (**A**) HVC_x cell, (**B**) HVC_{RA} cell, and (**C**) HVC_i cell. **D**, HVC PNs displayed a strong linear relationship (slope of linear fit = 0.96) between d' values calculated using either action potential activity (d' FR) or calcium response selectivity (d' Cal), whereas corresponding d' values from HVC_i cells exhibited a shallower linear relationship (slope of linear fit = 0.34). Dashed line indicates unity.

with peaks in firing rate (Fig. 3*A,B*; $r = 1.0$, $p = 4.9E-19$, linear regression). Peaks of BOS-evoked calcium transients (t_{cal}) lagged behind peaks in BOS-evoked action potential activity (t_{FR}) with a mean temporal offset of 103 ms and an SD of 27 ms, which is substantially less than the average syllable duration measured in birds from our colony (Fig. 3*C*; mean syllable duration = 111.0 ± 43.3 ms; $n = 30$ birds, 124 syllables). Therefore, 2p calcium imaging methods can be useful for detecting the phasic action potential response peaks that are often evoked in HVC PNs by BOS playback.

A hallmark of HVC neurons in the anesthetized zebra finch is that they typically fire more action potentials in response to playback of the BOS than to REV or REVSO (Fig. 4*A–C*) (Lewicki and Arthur, 1996; Mooney, 2000; Coleman and Mooney, 2004). We explored whether 2p calcium imaging in HVC could be used to detect this form of auditory selectivity, which is typically estimated using a d' metric (Green and Swets, 1966; Solis and Doupe, 1997; Theunissen and Doupe, 1998; Janata and Margoliash, 1999; Mooney, 2000; Rosen and Mooney, 2003). We found that both PN types displayed a linear relationship between d' values calculated using either action potential activity (d' FR; see Materials and Methods) or calcium transients (d' Cal; see Materials and Methods) (Fig. 4*D*; slope of linear fit = 0.96). In contrast, in-

terneurons exhibited a shallower linear relationship between electrical and optical estimates of selectivity (Fig. 4*D*; slope of linear fit = 0.34). Because the d' metric depends in part on response strength measurements, the diminished sensitivity of calcium imaging methods for estimating auditory selectivity in interneurons presumably stems from the smaller response magnitudes exhibited by this cell type. In summary, 2p calcium imaging affords a useful method for estimating the magnitude, timing, and selectivity of song-evoked action potential activity in HVC PNs but not in HVC_i cells.

Extending calcium imaging to populations of identified HVC PNs

We next sought to simultaneously monitor activity in populations of PNs to explore how song is represented spatially and temporally within HVC. We imaged populations of HVC cells bulk-loaded with the membrane-permeant calcium-sensitive dye OGB-1 AM in anesthetized adult male zebra finches in which HVC_x or HVC_{RA} cells had previously been labeled with retrograde tracers (Fig. 5*A*). Combining these methods with song playback, we could detect BOS-selective calcium responses in bulk-loaded HVC PNs containing retrograde tracers as well as other HVC neurons lacking the tracer (Fig. 5*B,C*).

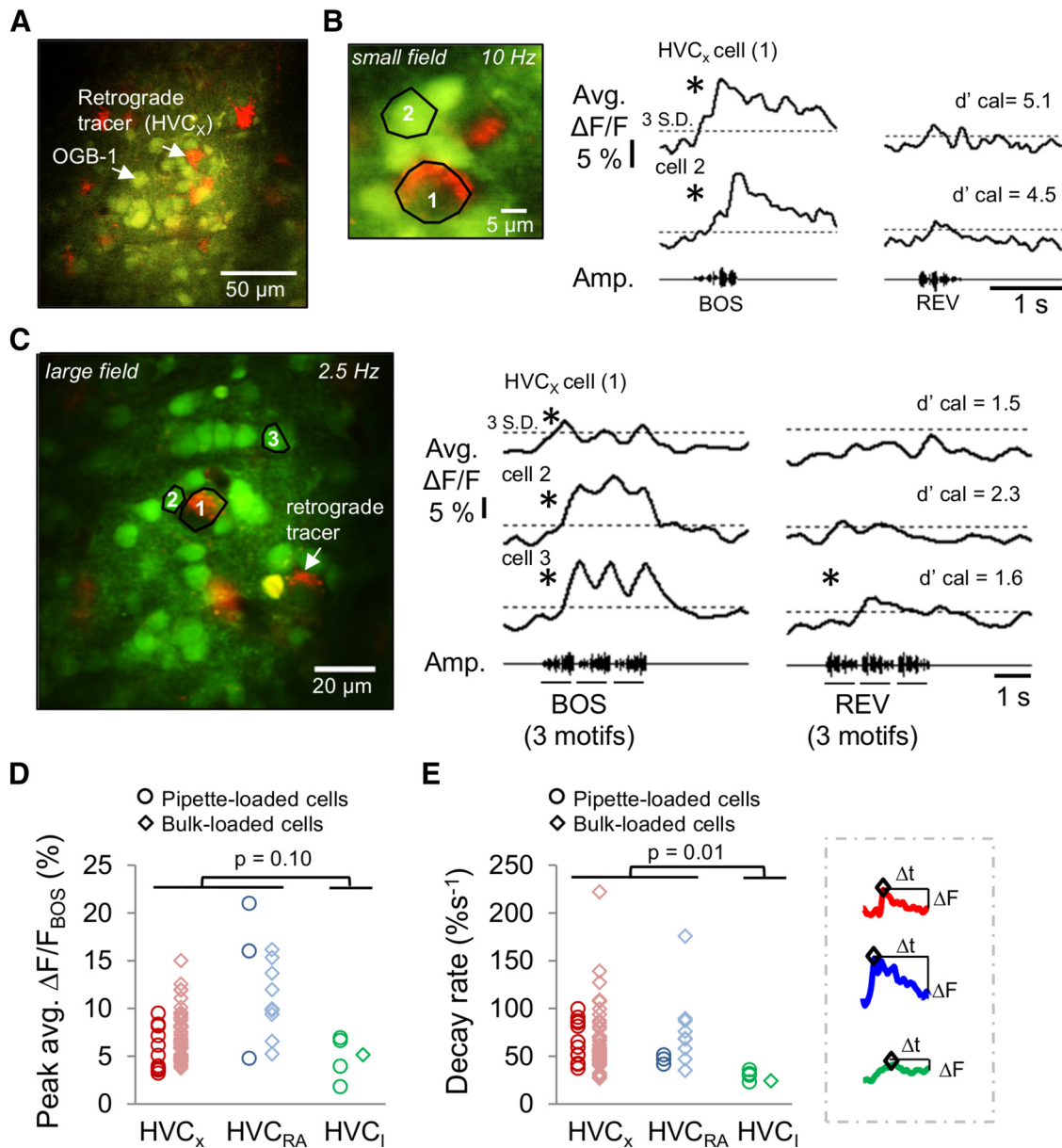


Figure 5. Population calcium imaging reveals auditory-evoked calcium signals that distinguish HVC PN from HVC_x cells. **A**, Example 2p image showing populations of HVC cells labeled with OGB-1 (green) via the bulk-loading method. **B**, **C**, BOS-selective calcium responses were obtained from identified (i.e., retrogradely labeled) and unidentified HVC cells simultaneously imaged within a small field (**B**, $38 \times 38 \mu\text{m}$, 10 Hz scan rate) or within a large field (**C**, $115 \times 115 \mu\text{m}$, 2.5 Hz scan rate). *Significant responses. **D**, **E**, Both bulk-loading and pipette loading methods reveal a trend toward larger and faster decaying BOS-evoked calcium transients in HVC PN than in HVC_I cells (mean peak responses HVC PN: $7.6 \pm 0.4\%$, HVC_I: $4.9 \pm 0.9\%$; $p = 0.10$; decay rates HVC PN: $65.1 \pm 3.7\% \text{ s}^{-1}$, HVC_I: $29.0 \pm 2.4\% \text{ s}^{-1}$, mean \pm SEM, two-tailed t test, $p = 0.01$). The bulk-loaded HVC_I cell was identified by loose cell-attached recording from a cell that lacked retrograde label and exhibited sustained BOS-evoked action potential responses characteristic of this cell type. **E**, Inset, Examples of different decay rates of BOS-evoked calcium responses from an HVC_x cell (red trace), an HVC_{RA} cell (blue trace), and an HVC_I cell (green trace).

Because retrograde tracers do not label all PN that target the tracer injection site, unlabeled cells could comprise different PN types, interneurons, or even glial cells (Graber et al., 2013). Therefore, we sought to develop functional and morphological criteria that could be used to further determine the identity of cells that were stained with OGB-1 but were not labeled with retrograde tracers. As a first step, we determined that BOS-evoked calcium transients in bulk-loaded HVC_{RA} or HVC_x cells, as identified by the presence of retrograde tracer, exhibited peaks and decay rates that overlapped with those obtained from their pipette-loaded counterparts (Fig. 5D,E). We also obtained a loose cell-attached recording from a bulk-loaded cell that we putatively identified as an HVC_I cell, based on sustained spiking

activity elicited by BOS playback and the absence of any retrograde label (data not shown). This putative bulk-loaded HVC_I cell displayed calcium response peaks and decay rates similar to those obtained from pipette-loaded HVC_I cells. We then identified functional differences in the BOS-evoked responses of HVC PN versus HVC_I cells. Both intracellular and bulk-loading methods indicated a trend toward larger and faster decaying BOS-evoked calcium transients in HVC PN than in HVC_I cells (Fig. 5D,E; mean peak responses HVC PN: $7.6 \pm 0.4\%$, HVC_I: $4.9 \pm 0.9\%$, mean \pm SEM; $p = 0.10$; decay rates HVC PN: $65.1 \pm 3.7\% \text{ s}^{-1}$, HVC_I: $29.0 \pm 2.4\% \text{ s}^{-1}$, $p = 0.01$ two-tailed t test).

A quadratic discriminant analysis was then applied to the magnitudes and decay rates of BOS-evoked calcium responses of

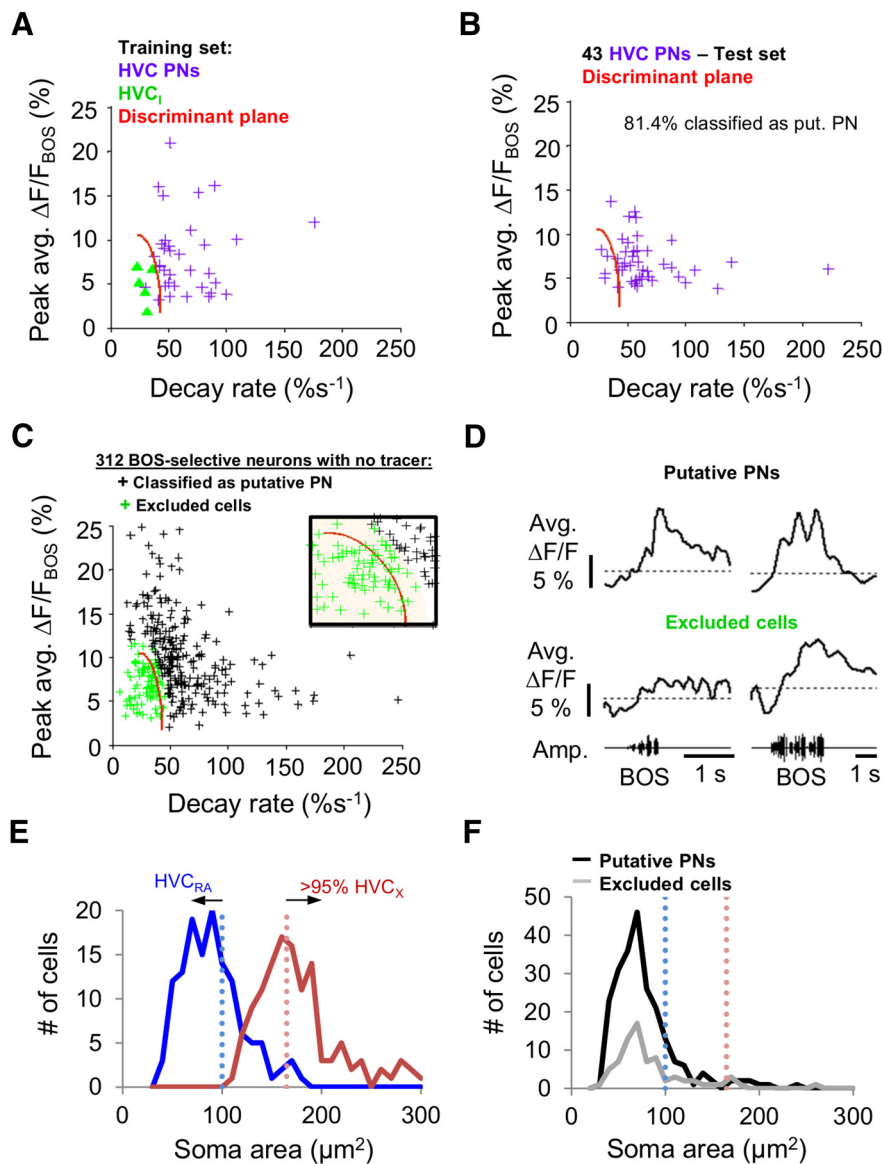


Figure 6. Classification of bulk-loaded HVC PNs based on BOS-evoked calcium signals and soma size. **A**, Quadratic discriminant analysis was applied to peak magnitudes and decay rates of average BOS-evoked calcium response in identified PNs (purple crosses) and HVC_I cells (green triangles) to obtain a discriminant plane (red) for classifying putative PNs. Datasets from small (38 × 38 μm) imaged fields were used to establish the discriminant criteria. **B**, The majority (81.4%) of identified HVC PNs from another dataset, obtained from large imaged fields, were correctly classified as PNs using this discriminant plane and using a criterion where the posterior probability is >0.95. Magnitude and decay rates of calcium responses of PNs from large imaged fields were not significantly different from small imaged field data (two-tailed *t* test, *p* > 0.05). **C**, Classification of non-tracer-labeled HVC cells (*n* = 312 cells) with BOS-selective responses (mean average ΔF/F > 2% and BOS-REV *d'* > 0.5) into either putative PNs (*n* = 226 cells, posterior probability > 0.95) or excluded cells (*n* = 86 cells). Inset, Magnified view of the discriminant plane boundary (red). Cells that were classified as non-PNs (green crosses) or where the posterior probability of being classified as a PN was <0.95 (shaded region) were excluded from further analysis. Black crosses represent cells that were classified as PNs with a posterior probability of >0.95. **D**, Examples of BOS-evoked calcium responses of putative PNs or excluded cells. **E**, Soma area measurements of retrogradely labeled HVC cells show the largely bimodal size distribution between the two PN types. Blue and pink dotted line indicates the 100 μm² and 165 μm² boundaries for classifying PNs into HVC_{RA} and HVC_X cells with <5% false positives, respectively. **F**, Most putative PNs identified had small soma areas and were therefore classified as HVC_{RA} cells.

identified HVC cells from small (38 × 38 μm) imaging fields to determine a discriminant plane that divided HVC PNs from HVC_I cells (Fig. 6A). To estimate the performance of the discriminant plane derived from the training set of PNs, BOS-evoked calcium imaging data from a test set of retrogradely labeled HVC PNs was applied (data were from large imaging fields, 115 × 115 μm, 2.5 Hz scan rate; the two response features were not significantly

different from data collected from small imaging fields, *p* > 0.05, two-tailed *t* test). Greater than 80% (81.4%) of the identified HVC PNs in the test set were correctly classified as PNs with high confidence (posterior probability > 0.95; Fig. 6B). We then used the quadratic discriminant plane we obtained from the training set to classify BOS-selective, bulk-loaded HVC cells that did not contain retrograde tracers into either putative PNs or excluded cells (Fig. 6C). Analysis was restricted to BOS-selective cells (mean ΔF/F > 2% and BOS-REV *d'* > 0.5) to minimize inclusion of astrocytes, which do not exhibit BOS-selective activity (Graber et al., 2013). The majority of HVC cells with significant BOS-selective calcium responses were classified as putative PNs (226 of 312 cells, 72.4%). These putative PNs displayed average BOS-evoked calcium responses that resembled those of identified HVC PNs, whereas excluded cells typically showed a “sustained” calcium response similar to those of identified HVC_I cells (Fig. 6D). Finally, we used differences in soma size to further classify the putative PNs as either HVC_{RA} or HVC_X cells (Fortune and Margoliash, 1995; Dutar et al., 1998; Mooney, 2000; Tschida and Mooney, 2012) (Fig. 6E). Most of the putative PNs had soma areas <100 μm² and were therefore classified as HVC_{RA} cells (197 of 226 cells; Fig. 6F).

Auditory representations of song’s temporal features are spatially distributed

A zebra finch song motif comprises a highly stereotyped sequence of syllables controlled with millisecond precision (Immelmann, 1969; Hahnloser et al., 2002; Glaze and Troyer, 2006; Long and Fee, 2008). Likewise, HVC PNs can display temporally sparse and precise bursts of action potentials in response to song playback and, in HVC_X cells, these auditory responses can closely mirror their singing-related activity (Mooney, 2000; Prather et al., 2008; Fujimoto et al., 2011; Hamaguchi et al., 2014). The extent to which these patterns of activity are spatially organized within HVC remains unknown. To begin to explore this issue, we imaged neighboring HVC PNs bulk-loaded with OGB-1 in small (38 × 38 μm) fields and measured their responses to BOS playback. We observed that neighboring HVC PNs, including both retrogradely labeled and putative HVC_{RA} and HVC_X cells, could exhibit BOS-evoked calcium responses with temporally distinct peaks (Fig. 7A). Because of the strong correlation between peaks in calcium responses and peaks in firing rate, these findings indicate that peaks in BOS-

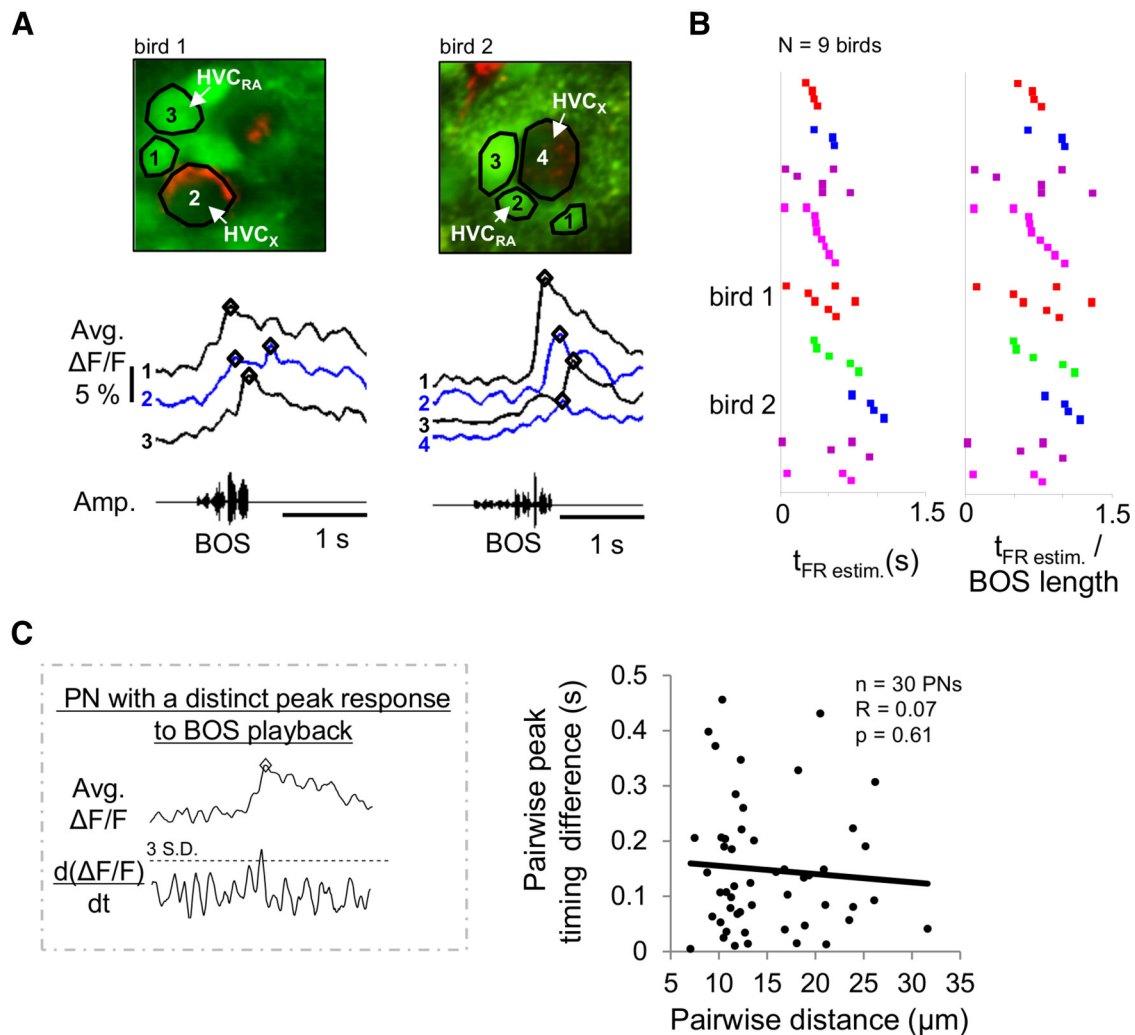


Figure 7. Neighboring HVC PN can respond with different temporal profiles to BOS playback. **A**, Examples of average BOS-evoked calcium responses from neighboring HVC PN in two different birds. Open diamonds represent local peaks of calcium responses from the different PN. Left, Only a subset of PN is shown (bird 1) to aid visualization. Right, Raster plots of estimated times of peak firing ($t_{FR\ estimated} = t_{cal} - \text{mean temporal offset}$; see Fig. 3C) of neighboring HVC PN during BOS playback. Squares of the same color represent different PN imaged from a single bird. Different colors represent different birds. Right, Estimated peak firing times normalized to the length of the BOS of different birds; responses could occur at >1 BOS when responses occurred after the end of the stimulus. **C**, Calculations of differences in peak times were restricted to PN that showed a distinct peak response during BOS playback (top). Distinct calcium response peaks were identified as instances when the first derivative of the averaged $\Delta F/F$ trace exceeded a defined threshold (mean baseline + 3SD) and aligned with the maximum value in the $\Delta F/F$ trace. Differences in peak times between HVC PN did not vary significantly with the distance between the PN ($r = 0.07$, $p = 0.61$, linear regression).

evoked firing patterns of neighboring neurons can be asynchronous. Across birds ($n = 9$), the times of the estimated peak firing rates (i.e., t_{FR}) of neighboring HVC PN could be distributed across the duration of the BOS, with a bias toward the latter half of the song (Fig. 7B; peak responses could occur after the end of the motif, which can account for peaks occurring at >1 BOS length). To further explore whether peak timing differences in neuronal responses to BOS playback were related to spatial positions of neurons within small regions of HVC, we plotted peak timing differences between PN with distinct peak calcium responses as a function of their pairwise distance. Peak time differences were not correlated with the relatively small distances between neighboring cells (Fig. 7C; $r = 0.07$, $p = 0.61$, linear regression). These observations indicate that small regions of HVC contain PN that respond at markedly different times in response to BOS playback, consistent with a distributed auditory representation of song within small distances ($<40 \mu m$).

The timing of song-evoked responses may reflect syllable selectivity

Prior studies have shown that individual HVC neurons can respond to specific syllables within the BOS (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995). The finding that neighboring HVC PN could respond at different times during BOS playback raises the possibility that nearby cells respond to different syllables. To examine the relationship between BOS- and syllable-evoked responses within small ($38 \times 38 \mu m$) regions of HVC, we presented the BOS and its component syllables while imaging the calcium responses of HVC PN bulk-loaded with OGB-1 (syllables were presented either individually or as 10 iterations of a given syllable within each trial; see Materials and Methods). For HVC PN with distinct peak calcium responses evoked by BOS playback, we observed variable levels of responsiveness to different syllables presented in isolation (Fig. 8A). To quantify syllable selectivity, we calculated a d'_{SYLL} value for all pairs of syllables and classified a syllable with a $d'_{SYLL} > 0.5$ when

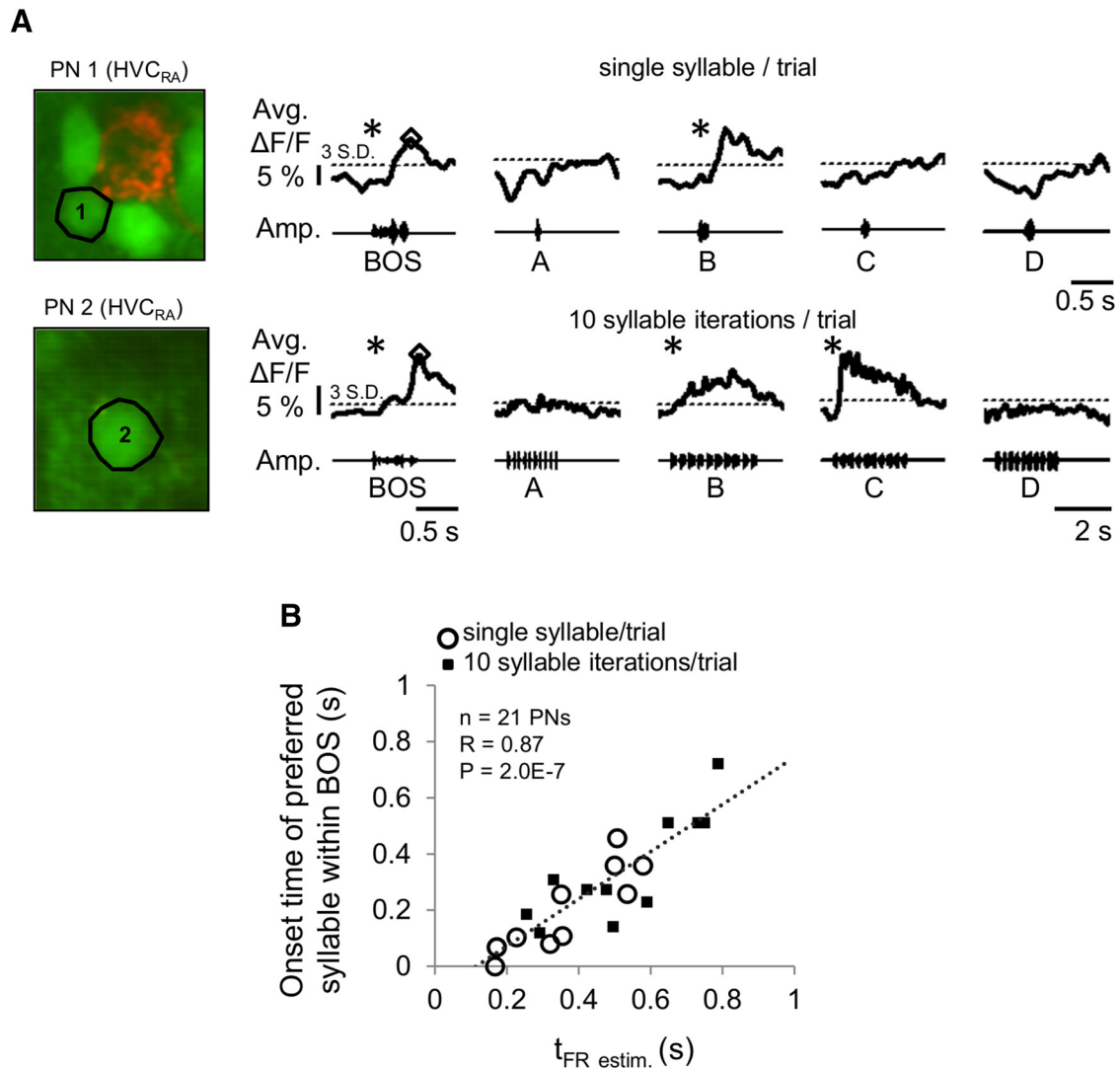


Figure 8. The timing of song-evoked responses may reflect selectivity for single syllables. **A**, Examples of average BOS- and syllable-evoked calcium responses from HVC_{RA} cells to the playback of a single syllable presented per trial (top) and to playback of 10 iterations of a given syllable per trial (bottom). Both cells responded to BOS playback with distinct peaks in their average calcium response (open diamonds) and responded selectively to the auditory playback of syllable B and C, respectively. *Significant responses. **B**, Estimated timing of peak firing during BOS playback measured from the majority of HVC PNs correlated positively with the onset time of their preferred syllable (linear regression, $r = 0.87$, $p = 2.0E-7$; d'_{SYLL} of the preferred syllable over all other syllables > 0.5). Data shown are from 21 of 32 HVC PNs tested (11 of 32 PNs had distinct peak responses to BOS playback but either did not have syllable-evoked responses or did not have a preferred syllable).

compared against all other syllables as a cell’s “preferred” syllable (see Materials and Methods). We then compared the onset time of the preferred syllable relative to the beginning of the BOS and plotted this preferred syllable onset time as a function of the timing of the estimated peak firing rate in the BOS-evoked response. Notably, we found that, for PNs with preferred syllable responses (21 of 32 PNs tested, $n = 14$ HVC_{RA} and $n = 7$ HVC_X), the best syllable onset time positively correlated with the BOS-evoked peak response time (Fig. 8B; $r = 0.87$, $p = 2.0E-7$, linear regression). We also observed that some HVC PNs with distinct peak responses evoked by BOS playback did not respond to syllables presented individually or responded to syllables in a non-selective manner ($n = 6$ HVC_{RA} and $n = 5$ HVC_X cells). These findings indicate that, for most HVC PNs, the peak in the BOS-evoked response may reflect their auditory selectivity for individual syllables, whereas a smaller fraction of PNs may either integrate information across syllables or respond broadly to many syllables in the BOS (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995).

Different syllables activate distinct and spatially intermingled PNs

To further examine how syllables are represented across larger populations of PNs in HVC, we imaged syllable-evoked calcium responses in larger ($115 \times 115 \mu m$) fields of bulk-loaded cells. Similar to the results we obtained with smaller imaging fields, PNs could exhibit syllable-selective responses (Fig. 9A). To compensate for the slower (2.5 Hz) scan rates used, we presented 10 iterations of each syllable within each trial. In most imaged fields ($n = 11$ of 16 fields from 11 birds), each syllable in the bird’s motif evoked a significant response in at least one PN, suggesting that local regions of HVC contained representation of all syllables within the BOS. Moreover, some PNs in each field responded only to a particular syllable, whereas others responded to multiple syllables (Fig. 9B). In several birds, we were able to image two fields located at different medial–lateral or dorsal–ventral positions of the same HVC (Fig. 9C; $n = 5$ birds). For each image field, we calculated a population syllable-evoked response and plotted this response as a function of syllable order (Fig. 9D). This

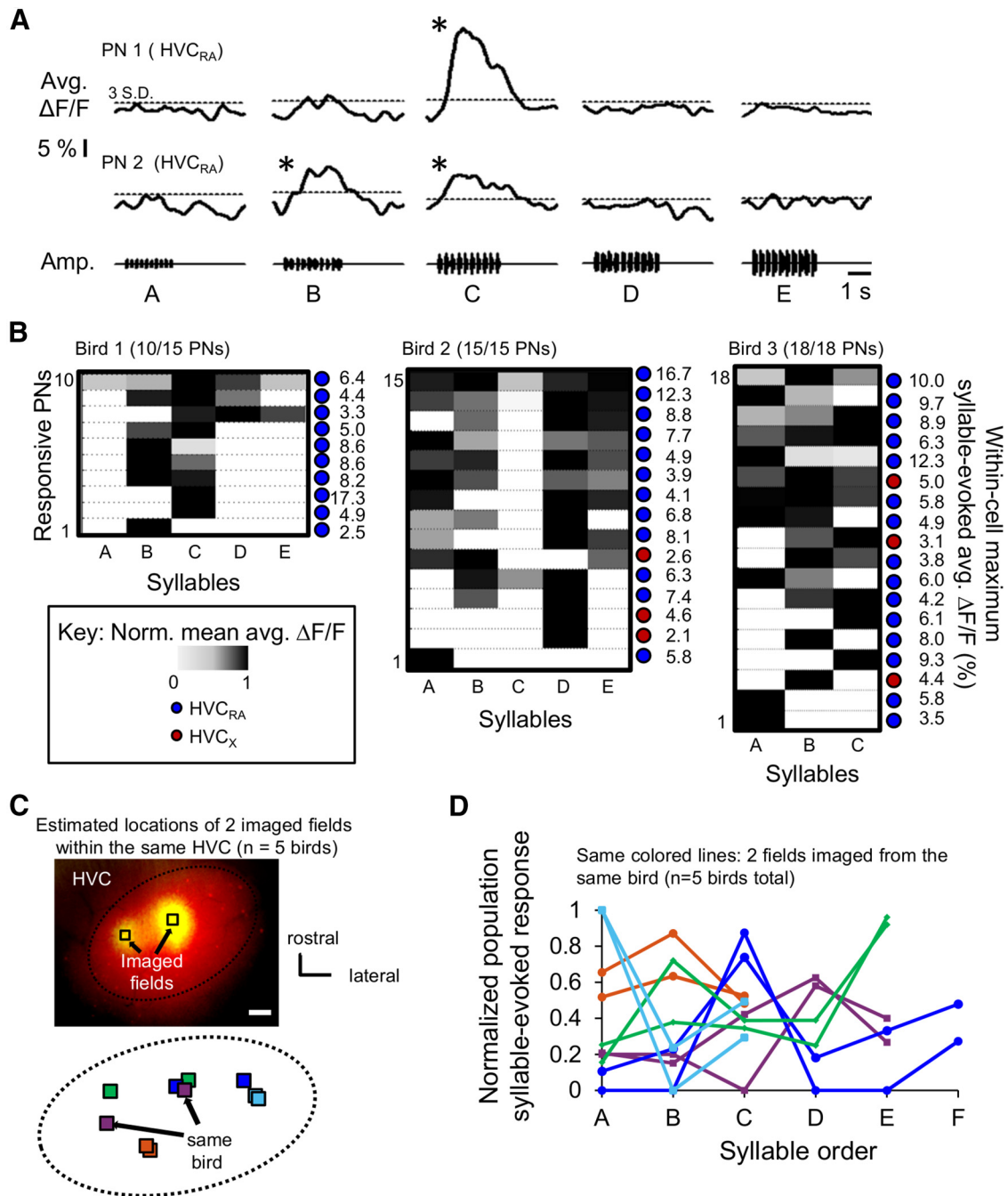


Figure 9. Different syllables evoke different population calcium responses in HVC. **A**, Average calcium responses from two example HVC_{RA} cells imaged within the same field evoked by playback of 10 iterations of a given syllable in each trial. *Significant responses. **B**, Matrices of normalized syllable-evoked calcium responses of HVC PNs within three imaged fields, each from a different bird. Only PNs with significant syllable-evoked responses are shown in these examples, and the proportion of syllable-responsive PNs within each field are indicated in parentheses above each matrix. **C**, Top, Fluorescent image where HVC’s boundaries were estimated by the extent of retrogradely labeled cells (red) and the two imaged fields are the bulk-loaded sites (green) within the same HVC. Black boxes represent the estimated locations of the imaged fields. Bottom, Schematic of estimated locations of two imaged fields from different birds mapped onto a “standardized” horizontal representation of HVC (n = 5 birds, see Materials and Methods). Same colored squares represent imaged locations from the same HVC. **D**, Normalized population syllable-evoked responses plotted against syllable order. The normalized population response is defined as the multiplication of the proportion of responsive PNs to a particular syllable and their summed normalized average $\Delta F/F$. The same color lines indicate imaged fields from the same bird (total n = 10 fields, 5 birds; imaged locations are shown in **C** with corresponding colored squares).

analysis revealed that syllable representations at the population level were qualitatively similar within different local sites in the same HVC (Fig. 9D).

To better understand the spatial organization of syllable-responsive PNs in HVC, we generated maps where the syllable that evoked the strongest mean calcium response from a given cell is color-coded and described as a PN’s “best” syllable (Fig.

10A). We pooled data across imaging fields and determined whether PNs that share the same best syllable are spatially located closer to each other than PNs that responded best to different syllables. Pairs of cells that shared the same best syllable and pairs that did not were separated by similar distances, regardless of whether the data were analyzed according to cell type (Fig. 10B; two-tailed *t* test, all *p* > 0.05) or when all types were pooled

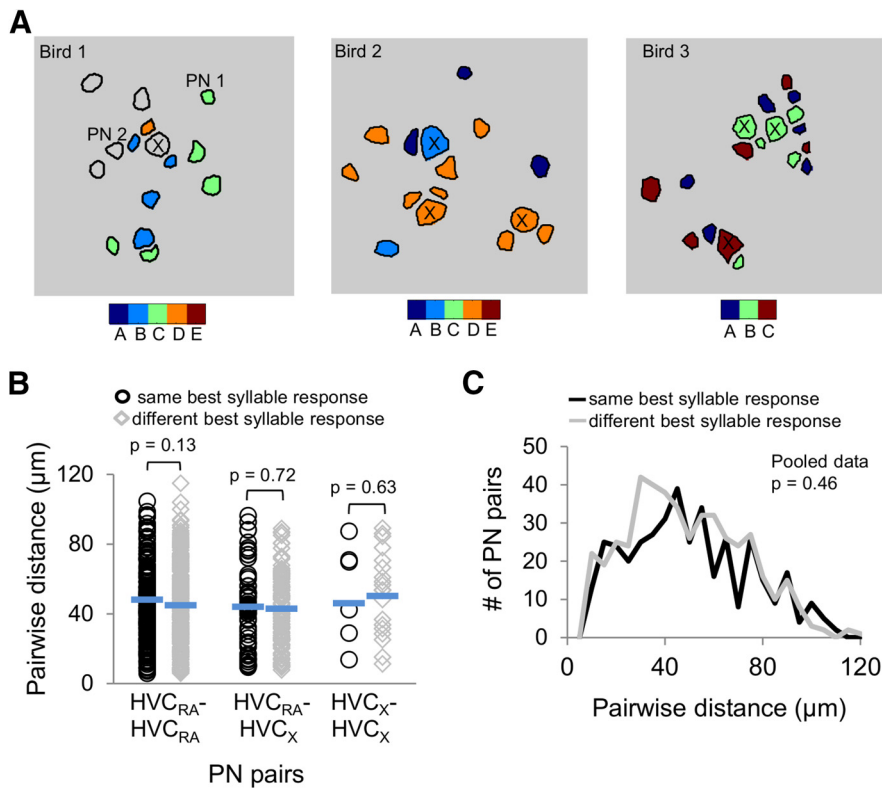


Figure 10. Different syllables activate spatially intermingled PNs. **A**, Spatial distributions of best syllable responses in HVC PNs are shown for the example imaging fields from Figure 9B. Each black line encircles a ROI and is color-coded to represent the PN's best syllable. HVC_X cells are marked with crosses, whereas unmarked cells are HVC_{RA} cells. **B**, Pairs of PNs that shared the same best syllable (black circles) and pairs that did not (gray diamonds) were separated by similar distances (two-tailed *t* test, all $p > 0.05$). Blue horizontal lines indicate mean values. **C**, Histogram showing similar distributions of pairwise distances for all PN pairs with the same (black line) or with different (gray line) best syllables (two-tailed *t* test, $p = 0.46$).

together (Fig. 10C; two-tailed *t* test, $p = 0.46$). Together, these analyses support the idea that HVC PNs displaying different syllable response preferences are spatially intermingled at a local ($\sim 100 \times 100 \mu\text{m}$) scale within HVC.

Evidence of spatially graded and syllable-related noise correlations in HVC

Noise correlation analysis of calcium signals has been used to infer stronger synaptic connections or more abundant shared inputs in populations of neurons (Sato et al., 2007; Rothschild et al., 2010, 2013; Hofer et al., 2011). In the auditory cortex of the mouse, noise correlations in calcium transients of nearby neurons were found to decline significantly in strength over short distances ($\sim 100 \mu\text{m}$), which may indicate a spatial gradient of stronger local synaptic connections or more abundant shared inputs (Rothschild et al., 2010). To begin to explore whether such spatial gradients of connectivity exist within HVC, we calculated the correlation coefficient of the trial-to-trial deviations in BOS-evoked calcium responses (i.e., noise correlations) between different pairs of PNs and plotted this coefficient as a function of the cell-to-cell distance for each pair. This analysis revealed that trial-by-trial noise correlation coefficients between HVC_{RA} cells and between HVC_{RA} and HVC_X cells declined significantly over local ($\sim 100 \mu\text{m}$) distances (Fig. 11A,B; HVC_{RA} cells: $n = 687$ pairs, $r = 0.17$, $p = 5.3\text{E-}6$; HVC_X cells: $n = 411$ pairs, $r = 0.12$, $p = 0.02$, linear regression). Randomly shuffling the spatial locations among PNs within each field strongly reduced this distance effect (Fig. 11A,B; HVC_{RA} pairs: $r = 0.01$, $p = 0.73$; HVC_{RA}–HVC_X

pairs: $r = 0.02$, $p = 0.71$). By randomly down-sampling HVC_{RA} cell pair data to the same sample size as HVC_{RA}–HVC_X pairs, we observed that the distance effect of decreased noise correlation coefficient was stronger for HVC_{RA} pairs than HVC_{RA}–HVC_X pairs (range of *R* values for 411 pairs of HVC_{RA} cells: 0.13 to 0.24, 10 iterations of random sampling). In contrast, a distance-dependent effect on noise correlations was not detected between HVC_X cell pairs (Fig. 11C; $n = 62$ pairs, $r = 0.11$, $p = 0.42$). Although the sample size we obtained was too small to achieve statistical validity, random down sampling of our HVC_{RA}–HVC_{RA} and HVC_{RA}–HVC_X datasets to 62 pairs yielded distributions with negative slopes 9 of 10 times, suggesting that the positive slope of the correlations in HVC_X cell pairs may be distinct from the negative correlations we measured in HVC_{RA}–HVC_{RA} and HVC_{RA}–HVC_X cell pairs.

Neurons responding most strongly to the same syllable are more likely to display temporally coincident activity, a feature that could arise due to shared inputs and/or cause them to develop stronger synaptic connections (Hebb, 1949; Caporale and Dan, 2008). Therefore, we examined noise correlation coefficients in pairs of PNs with similar or different best syllable responses. Across all pairs, PN pairs with the same best syllable showed significantly higher pairwise noise correlations than PN pairs with different best syllables (Fig. 12A; two-way ANOVA, $p = 0.04$ for the main effect comparing PN pairs with the same or different best syllable responses). This effect was also observed for pairs of HVC_{RA} cells with similar syllable preferences but not for other cell pairs (Fig. 12B; two-tailed *t* test, $p = 0.03$ for HVC_{RA} pairs; $p = 0.23$ for HVC_{RA}–HVC_X pairs; $p = 0.33$ for HVC_X pairs). Overall, these findings provide evidence that spatially intermingled PNs are organized into cell- and syllable-type networks within HVC.

Discussion

Using *in vivo* 2p calcium imaging, we found that neighboring HVC PNs can respond at markedly different times to song playback and that different syllables activate spatially intermingled PNs within a local region of HVC. Moreover, noise correlation analysis reveals stronger synaptic connections or more abundant shared inputs between PNs that respond most strongly to the same syllable and also provides evidence of a spatial gradient of connectivity involving HVC_{RA} and HVC_{RA}–HVC_X cell pairs. These findings support a model in which features of song, including the identity and timing of individual syllables, are represented by spatially intermingled PNs functionally organized into cell- and syllable-type networks on local spatial scales ($\sim 100 \mu\text{m}$).

Population and single-cell auditory representations of song

Whereas multiunit mapping of BOS-evoked activity in HVC has detected highly synchronous activity across large extents of HVC

(Sutter and Margoliash, 1994), calcium imaging revealed remarkable heterogeneity in song- and syllable-evoked activity at the level of neighboring neurons. One explanation for this dissimilarity is that multiunit recordings in HVC mostly sample interneurons (compare Yu and Margoliash, 1996; Hahnloser et al., 2002), which fire tonically to BOS playback (Mooney, 2000; Rosen and Mooney, 2003, 2006; Coleman and Mooney, 2004), features that may contribute to synchronous multiunit activity. In contrast, the calcium imaging methods used here primarily sample PNs in HVC. A comparison of somatic membrane potential and calcium signals showed that our calcium imaging methods accurately estimate firing rate changes and stimulus selectivity in PNs but are less sensitive to these features in interneurons. Two factors that may account for the reduced sensitivity of calcium imaging methods in HVC_I cells are narrower action potentials, which may limit calcium entry, and high levels of calcium binding proteins, which may compete with the calcium indicator (Wild et al., 2005; Schwaller, 2010). This contrasting sensitivity of calcium imaging in PNs and interneurons suggests that this method may be broadly useful to selectively monitor information that PNs transmit to downstream structures important to song motor control, error correction, and perception (Nottebohm et al., 1976; Vu et al., 1994; Scharff et al., 1998; Brainard and Doupe, 2000; Ölveczky et al., 2005; Andalman and Fee, 2009).

The highly linear correspondence in the timing of peaks in action potential firing rates and calcium transients allowed us to detect timing differences in PN activity and relate these differences to their spatial locations within HVC. Although a recent calcium imaging study revealed spatial clustering of auditory responsiveness and selectivity in HVC_X neurons, the authors did not examine the temporal patterns of BOS-evoked activity (Graber et al., 2013). Also, although a previous study using dual intracellular recordings showed that pairs of HVC_{RA} cells could respond at different times to BOS playback (Rosen and Mooney, 2006), it lacked information about their spatial locations within HVC. Here we found that neighboring PNs could display asynchronous BOS response patterns and neurons with similar response times were not spatially clustered, supporting a distributed model for auditory representations of song in HVC. Such distributed representations may also extend to the motor domain, as HVC_X cells display auditory–vocal mirroring (Prather et al., 2008; Fujimoto et al., 2011; Hamaguchi et al., 2014) and single-unit recordings of neighboring HVC_{RA} neurons during sleep and singing revealed that these neurons do not fire action potential bursts at the same time (Hahnloser et al., 2002). Future experiments that image both auditory and singing-related activity in the HVC of individual birds can resolve whether auditory and motor representations of song are similar.

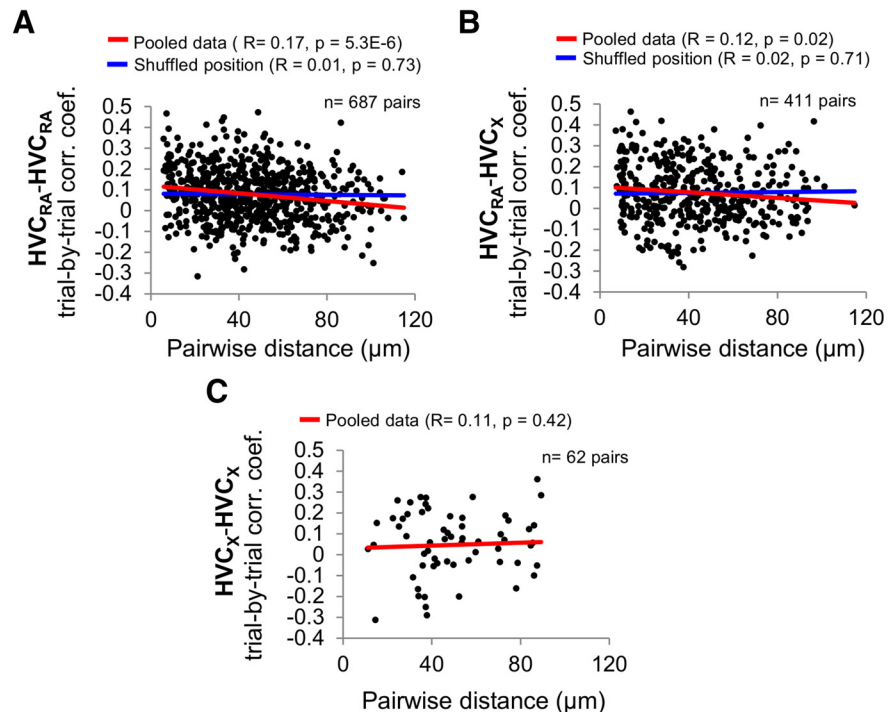


Figure 11. Spatial gradients of noise correlations were observed between HVC_{RA} cell pairs and HVC_{RA}-HVC_X cell pairs. **A**, Noise correlation coefficients between pairs of HVC_{RA} cells decline significantly with their pairwise distances ($r = 0.17$, $p = 5.3E-6$, linear regression). Noise correlation coefficients are calculated based on BOS-evoked calcium responses. Red lines indicate the linear regression fits of experimental data. Blue lines indicate the linear regression fit when spatial positions of PNs were randomly shuffled among PNs within each field (individual shuffled data points are not plotted). **B**, Noise correlation coefficients between pairs of HVC_{RA}-HVC_X cells decline significantly with their pairwise distances ($r = 0.12$, $p = 0.02$). **C**, Pairwise noise correlation coefficients of HVC_X-HVC_X cell pairs did not show significant changes with increasing pairwise distances ($r = 0.11$, $p = 0.42$).

Locally heterogeneous syllable representations in HVC

Here we found that neighboring neurons in HVC could prefer different syllables and that every region we imaged contained a near complete representation of all syllables in a bird's song, demonstrating functional heterogeneity and redundancy reminiscent of the frequency representations in the rodent auditory cortex (Bandyopadhyay et al., 2010; Rothschild et al., 2010). We also observed that, within any local region of HVC, individual syllables could be represented by the activity of some neurons that were highly selective for specific syllables or unique combinations of neurons that were broadly responsive to different syllables. This functional dichotomy within a local region of HVC could point to parallel and possibly hierarchical coding mechanisms within HVC, both of which are supported by experimental evidence. First, electrophysiological studies have detected HVC neurons with broader or narrower selectivity for features in the BOS, including cells that only respond to specific syllables or syllable combinations (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995; Prather et al., 2008, 2009). Furthermore, local circuit activity in HVC transforms tonically active and more broadly tuned synaptic inputs into more phasic and selective action potential output (Mooney, 2000; Rosen and Mooney, 2003, 2006; Coleman and Mooney, 2004). Finally, intracellular recordings from pairs of HVC neurons in brain slices revealed strong synaptic connections within and between HVC cell types (Mooney and Prather, 2005), which could enable more broadly responsive cells to transmit information to syllable-selective cells. An important next step will be to apply calcium-imaging methods to explore how higher-order features of song, such as syllable combinations, are represented within HVC and how these repre-

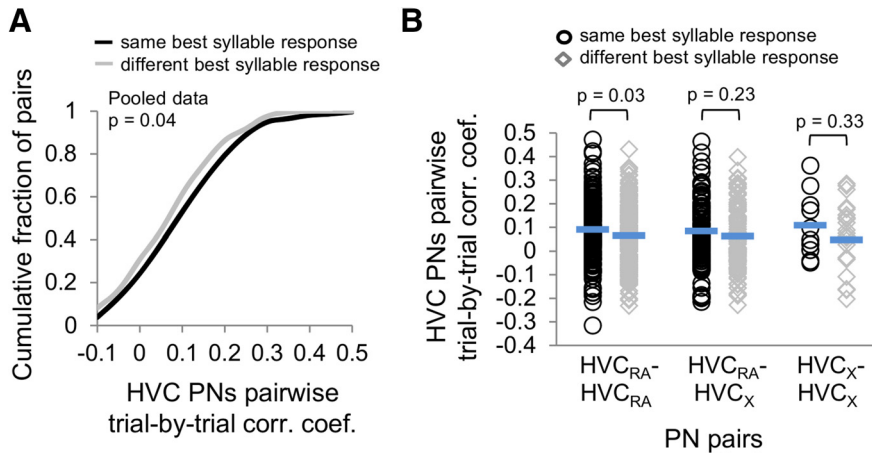


Figure 12. HVC PN pairs with the same best syllable have higher noise correlation coefficients. **A**, A cumulative distribution plot showing pairwise noise correlation coefficients from all PN pairs with either the same (black line) or different (gray line) best syllables. Pairs of PN with the same best syllable had significantly higher noise correlation coefficients than those with different best syllables (two-way ANOVA, $p = 0.04$ for main effect of PN pairs with the same vs different best syllables; $p = 0.91$ for main effect of cell pair group; $p = 0.73$ for interaction effect of the two factors). **B**, Significantly higher noise correlation coefficients were also observed for HVC_{RA} pairs sharing the same best syllable than for HVC_{RA} pairs that did not share the same best syllable (HVC_{RA} pairs: $p = 0.03$; HVC_{RA}–HVC_X pairs: $p = 0.23$; HVC_X pairs: $p = 0.33$; two-tailed t test). Blue lines indicate mean values.

Auditory representation of BOS and syllables in HVC

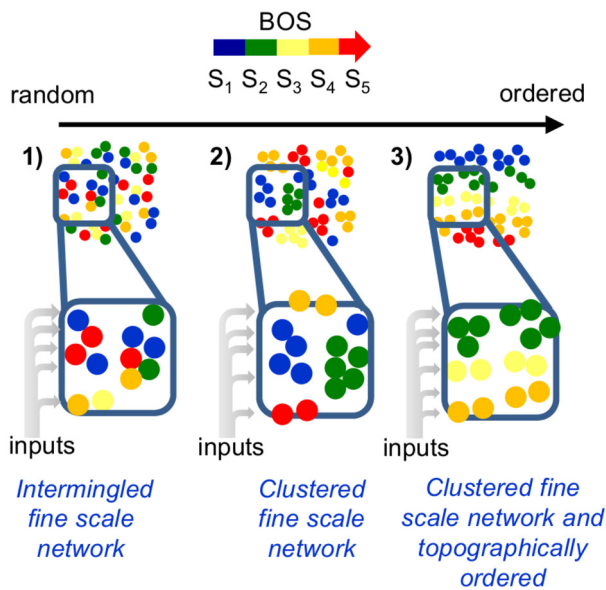


Figure 13. The current findings indicate that auditory representations of the BOS and its syllables are organized in a salt-and-pepper fashion in HVC. An illustration of different models of the spatial representation of BOS and its component syllables within HVC, including (1) a salt-and-pepper organization, (2) a spatially clustered organization, or (3) a topographically ordered organization. Different colors of circles indicate either different timing of peak responses during BOS playback or different song syllables that evoked the strongest responses within individual PNs.

sentations relate to the song and syllable representations studied here. One possibility is that combinations of syllables are represented by different organizational rules, possibly across larger spatial scales, than the single syllable representations studied here.

Cell-type and syllable-related functional networks in HVC

The current findings support a “salt-and-pepper” model of song representation within local regions of HVC, rather than a repre-

sentation involving either clustered or topographically organized networks (Fig. 13). Despite this salt-and-pepper song representation, we also found spatial gradients of noise correlations between HVC_{RA} cells and between HVC_{RA} and HVC_X cells and enhanced noise correlations between cells that responded most strongly to the same syllable. The spatial gradient of correlation we detected between pairs of HVC_{RA} cells and between HVC_{RA} and HVC_X cells could indicate that these cells share more common inputs or form stronger synapses with each other at a local scale (~100 μm). Indeed, neighboring HVC PNs can form strong monosynaptic and disinaptic connections, which could result in the enhanced noise correlations we observed at the local scale (Mooney and Prather, 2005). Moreover, assuming that cells that respond most strongly to auditory presentation of the same syllable also display coincident activity during singing, enhanced noise correlations

could reflect stronger synaptic connections or more abundant shared inputs between these cells that may arise as a result of synaptic strengthening driven by Hebbian or spike-timing-dependent plasticity mechanisms (Hebb, 1949; Caporale and Dan, 2008). Therefore, an important future goal will be to measure the synaptic connectivity between HVC neurons with similar or different response properties and to explore how the heterogeneous and redundant syllable representations we detected within small regions of HVC relates to the larger-scale circuit structure of this sensorimotor nucleus. Indeed, functional and anatomical evidence suggests that axonal connectivity is biased along HVC’s rostrocaudal axis (Stauffer et al., 2012; Day et al., 2013; Kosche et al., 2015), which may provide a circuit architecture that could organize these locally distributed representations into a larger scale pattern of organization. Indeed, future experiments that combine calcium imaging with direct measurements of synaptic connectivity can determine how activity propagates through the HVC network and whether such propagation involves a synaptic chain mechanism, as previously proposed (Abeles, 1982; Fee et al., 2004; Li and Greenside, 2006; Jin et al., 2007; Long et al., 2010).

Auditory encoding of learned vocalizations in vocal motor networks

This is the first study to image auditory representations of syllabic units with cellular resolution. The auditory representation of speech sounds in the human brain has been studied using fMRI (Binder et al., 2000; Wilson et al., 2004; Pulvermüller et al., 2006) and electrocorticography (Chang et al., 2010; Flinker et al., 2011; Mesgarani and Chang, 2012; Mesgarani et al., 2014), approaches that have the benefit of revealing regional patterns of brain activity, but with a relatively low (~1 mm) spatial resolution corresponding to the combined activity of thousands or even millions of neurons. In speech motor regions, which may be analogous to HVC, speech sounds that involve different articulatory gestures activate motor regions that control the corresponding articulators, such as the tongue or lips (Pulvermüller et al., 2006; D’Ausilio et al., 2009; Bouchard et al., 2013). In the songbird, topographic representations of the vocal musculature are present

at lower levels of the vocal motor system (Vicario and Nottebohm, 1988; Vicario, 1991), but not in HVC (Yip et al., 2012), which may instead encode information about song timing (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long and Fee, 2008) or vocal gestures (Amador et al., 2013). Therefore, the spatially intermingled organization we observed at local scales in HVC could be a consequence of mapping auditory representations of song onto a motor framework that encodes timing or gestural features of song.

References

- Abeles M (1982) Local cortical circuits: an electrophysiological study. Berlin: Springer.
- Amador A, Perl YS, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495:59–64. [CrossRef Medline](#)
- Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci U S A* 106:12518–12523. [CrossRef Medline](#)
- Bandyopadhyay S, Shamma SA, Kanold PO (2010) Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci* 13:361–368. [CrossRef Medline](#)
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and non-speech sounds. *Cereb Cortex* 10:512–528. [CrossRef Medline](#)
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. [CrossRef Medline](#)
- Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404:762–766. [CrossRef Medline](#)
- Brenowitz EA (1991) Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science* 251:303–305. [CrossRef Medline](#)
- Caporale N, Dan Y (2008) Spike timing-dependent plasticity: a Hebbian learning rule. *Annu Rev Neurosci* 31:25–46. [CrossRef Medline](#)
- Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13:1428–1432. [CrossRef Medline](#)
- Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J Neurosci* 24:7251–7265. [CrossRef Medline](#)
- D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. *Curr Biol* 19:381–385. [CrossRef Medline](#)
- Day NF, Terleski KL, Nykamp DQ, Nick TA (2013) Directed functional connectivity matures with motor learning in a cortical pattern generator. *J Neurophysiol* 109:913–923. [CrossRef Medline](#)
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631. [CrossRef Medline](#)
- Dutar P, Vu HM, Perkel DJ (1998) Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVC of the adult zebra finch. *J Neurophysiol* 80:1828–1838. [Medline](#)
- Fee MS, Kozhevnikov AA, Hahnloser RHR (2004) Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci* 1016:153–170. [CrossRef Medline](#)
- Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Subcentimeter language organization in the human temporal lobe. *Brain Lang* 117:103–109. [CrossRef Medline](#)
- Fortune ES, Margoliash D (1995) Parallel pathways and convergence onto HVC and adjacent neostriatum of adult zebra finches (*Taeniopygia guttata*). *J Comp Neurol* 360:413–441. [CrossRef Medline](#)
- Fujimoto H, Hasegawa T, Watanabe D (2011) Neural coding of syntactic structure in learned vocalizations in the songbird. *J Neurosci* 31:10023–10033. [CrossRef Medline](#)
- Gentner TQ, Hulse SH, Bentley GE, Ball GF (2000) Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *J Neurobiol* 42:117–133. [CrossRef Medline](#)
- Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. *J Neurosci* 26:991–1005. [CrossRef Medline](#)
- Graber MH, Helmchen F, Hahnloser RH (2013) Activity in a premotor cortical nucleus of zebra finches is locally organized and exhibits auditory selectivity in neurons but not in glia. *PLoS One* 8:e81177. [CrossRef Medline](#)
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.
- Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65–70. [CrossRef Medline](#)
- Hamaguchi K, Tschida KA, Yoon I, Donald BR, Mooney R, Calabrese RL (2014) Auditory synapses to song premotor neurons are gated off during vocalization in zebra finches. *eLife* 3.
- Hebb DO (1949) The organization of behavior: a neuropsychological theory. New York: Wiley.
- Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, Zeng H, Lein E, Lesica NA, Mrsic-Flogel TD (2011) Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat Neurosci* 14:1045–1052. [CrossRef Medline](#)
- Hultsch H, Todt D (1989) Memorization and reproduction of songs in nightingales (*Luscinia megarhynchos*): evidence for package formation. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 165:197–203. [CrossRef](#)
- Immelmann K (1969) Song development in zebra finch and other Estrildid finches. In: Bird vocalisations (Hinde RA, ed), pp 61–74. London: Cambridge UP.
- Janata P, Margoliash D (1999) Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *J Neurosci* 19:5108–5118. [Medline](#)
- Jin DZ, Ramazanoğlu FM, Seung HS (2007) Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *J Comput Neurosci* 23:283–299. [CrossRef Medline](#)
- Kerr JN, de Kock CP, Greenberg DS, Bruno RM, Sakmann B, Helmchen F (2007) Spatial organization of neuronal population responses in Layer 2/3 of rat barrel cortex. *J Neurosci* 27:13316–13328. [CrossRef Medline](#)
- Konishi M (1985) Birdsong: from behavior to neuron. *Annu Rev Neurosci* 8:125–170. [CrossRef Medline](#)
- Kosche G, Vallentin D, Long MA (2015) Interplay of inhibition and excitation shapes a premotor neural sequence. *J Neurosci* 35:1217–1227. [CrossRef Medline](#)
- Kozhevnikov AA, Fee MS (2007) Singing-related activity of identified HVC neurons in the zebra finch. *J Neurophysiol* 97:4271–4283. [CrossRef Medline](#)
- Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16:6987–6998. [Medline](#)
- Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci U S A* 92:5582–5586. [CrossRef Medline](#)
- Li M, Greenside H (2006) Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:011918. [CrossRef Medline](#)
- Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456:189–194. [CrossRef Medline](#)
- Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. *Nature* 468:394–399. [CrossRef Medline](#)
- Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3:1039–1057. [Medline](#)
- Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12:4309–4326. [Medline](#)
- Marler P, Tamura M (1964) Culturally transmitted patterns of vocal behaviour in sparrows. *Science* 146:1483–1486. [CrossRef Medline](#)
- McCasland JS, Konishi M (1981) Interaction between auditory and motor activities in an avian song control nucleus. *Proc Natl Acad Sci U S A* 78:7815–7819. [CrossRef Medline](#)
- Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485:233–236. [CrossRef Medline](#)
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010. [CrossRef Medline](#)

- Mooney R (2000) Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J Neurosci* 20:5420–5436. [Medline](#)
- Mooney R (2009) Neural mechanisms for learned birdsong. *Learn Mem* 16:655–669. [CrossRef Medline](#)
- Mooney R, Prather JF (2005) The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J Neurosci* 25:1952–1964. [CrossRef Medline](#)
- Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in the canary, *Serinus canarius*. *J Comp Neurol* 165:457–486. [CrossRef Medline](#)
- Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* 433:597–603. [CrossRef Medline](#)
- Ölveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* 3:e153. [CrossRef Medline](#)
- Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–310. [CrossRef Medline](#)
- Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R (2009) Neural correlates of categorical perception in learned vocal communication. *Nat Neurosci* 12:221–228. [CrossRef Medline](#)
- Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci U S A* 103:7865–7870. [CrossRef Medline](#)
- Rosen MJ, Mooney R (2003) Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* 39:177–194. [CrossRef Medline](#)
- Rosen MJ, Mooney R (2006) Synaptic interactions underlying song-selectivity in the avian nucleus HVC revealed by dual intracellular recordings. *J Neurophysiol* 95:1158–1175. [Medline](#)
- Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nat Neurosci* 13:353–360. [CrossRef Medline](#)
- Rothschild G, Cohen L, Mizrahi A, Nelken I (2013) Elevated correlations in neuronal ensembles of mouse auditory cortex following parturition. *J Neurosci* 33:12851–12861. [CrossRef Medline](#)
- Sato TR, Gray NW, Mainen ZF, Svoboda K (2007) The functional micro-architecture of the mouse barrel cortex. *PLoS Biol* 5:e189. [CrossRef Medline](#)
- Scharff C, Nottebohm F, Cynx J (1998) Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J Neurobiol* 36:81–90. [CrossRef Medline](#)
- Schwaller B (2010) Cytosolic Ca²⁺ buffers. *Cold Spring Harb Perspect Biol* 2:a004051. [CrossRef Medline](#)
- Solis MM, Doupe AJ (1997) Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning. *J Neurosci* 17:6447–6462. [Medline](#)
- Stauffer TR, Elliott KC, Ross MT, Basista MJ, Hyson RL, Johnson F (2012) Axial organization of a brain region that sequences a learned pattern of behavior. *J Neurosci* 32:9312–9322. [CrossRef Medline](#)
- Stosiek C, Garaschuk O, Holthoff K, Konnerth A (2003) In vivo two-photon calcium imaging of neuronal networks. *Proc Natl Acad Sci U S A* 100:7319–7324. [CrossRef Medline](#)
- Sutter ML, Margoliash D (1994) Global synchronous response to autogenous song in zebra finch HVC. *J Neurophysiol* 72:2105–2123. [Medline](#)
- Theunissen FE, Doupe AJ (1998) Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J Neurosci* 18:3786–3802. [Medline](#)
- Thompson JA, Wu W, Bertram R, Johnson F (2007) Auditory-dependent vocal recovery in adult male zebra finches is facilitated by lesion of a forebrain pathway that includes the basal ganglia. *J Neurosci* 27:12308–12320. [CrossRef Medline](#)
- Tschida KA, Mooney R (2012) Deafening drives cell-type-specific changes to dendritic spines in a sensorimotor nucleus important to learned vocalizations. *Neuron* 73:1028–1039. [CrossRef Medline](#)
- Vicario DS (1991) Organization of the zebra finch song control system: functional organization of outputs from nucleus robustus archistriatalis. *J Comp Neurol* 309:486–494. [CrossRef Medline](#)
- Vicario DS, Nottebohm F (1988) Organization of the zebra finch song control system: I. Representation of syringeal muscles in the hypoglossal nucleus. *J Comp Neurol* 271:346–354. [CrossRef Medline](#)
- Vu ET, Mazurek ME, Kuo Y (1994) Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci* 14:6924–6934. [Medline](#)
- Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. *Ann N Y Acad Sci* 1016:438–462. [CrossRef Medline](#)
- Wild JM, Williams MN, Howie GJ, Mooney R (2005) Calcium-binding proteins define interneurons in HVC of the zebra finch (*Taeniopygia guttata*). *J Comp Neurol* 483:76–90. [CrossRef Medline](#)
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702. [CrossRef Medline](#)
- Yip ZC, Miller-Sims VC, Bottjer SW (2012) Morphology of axonal projections from the high vocal center to vocal motor cortex in songbirds. *J Comp Neurol* 520:2742–2756. [CrossRef Medline](#)
- Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. *Science* 273:1871–1875. [CrossRef Medline](#)