Linear and nonlinear auditory response properties of interneurons in a high-order avian vocal motor nucleus during wakefulness

Jonathan N. Raksin,1 Christopher M. Glaze,2 Sarah Smith,2 and Marc F. Schmidt1,2

1Neuroscience Graduate Group and 2Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania

Submitted 16 November 2009; accepted in final form 23 December 2011

Raksin JN, Glaze CM, Smith S, Schmidt MF. Linear and nonlinear auditory response properties of interneurons in a high-order avian vocal motor nucleus during wakefulness. J Neurophysiol 107: 2185–2201, 2012. First published December 28, 2011; doi:10.1152/jn.01003.2009.—Motor-related forebrain areas in higher vertebrates also show responses to passively presented sensory stimuli. However, sensory tuning properties in these areas, especially during wakefulness, and their relation to perception, are poorly understood. In the avian song system, HVC (proper name) is a vocal-motor structure with auditory responses well defined under anesthesia but poorly characterized during wakefulness. We used a large set of stimuli including the bird’s own song (BOS) and many conspecific songs (CON) to characterize auditory tuning properties in putative interneurons (HVCINT) during wakefulness. Our findings suggest that HVC contains a diversity of responses that vary in overall excitability to auditory stimuli, as well as bias in spike rate increases to BOS over CON. We used statistical tests to classify cells in order to further probe auditory responses, yielding one-third of neurons that were either unresponsive or suppressed and two-thirds with excitatory responses to one or more stimuli. A subset of excitatory neurons were tuned exclusively to BOS and showed very low linearity as measured by spectrotemporal receptive field analysis (STRF). The remaining excitatory neurons responded well to CON stimuli, although many cells still expressed a bias toward BOS. These findings suggest the concurrent presence of a nonlinear and a linear component to responses in HVC, even within the same neuron. These characteristics are consistent with perceptual deficits in distinguishing BOS from CON stimuli following lesions of HVC and other song nuclei and suggest mirror neuronlike qualities in which “self” (here BOS) is used as a referent to judge “other” (here CON).

While it has been proposed that such responses facilitate self-referential understanding of the goal-directed motor actions of others (Rizzolatti et al. 2001), no functional evidence exists as of yet in support of this notion (Hickock 2009). The avian song system circuits underlying a bird’s unique vocal output may also underlie the ability to perform perceptual discrimination tasks (Gentner et al. 2000; Scharff et al. 1998). The forebrain nucleus HVC forms part of a dedicated neural circuit for vocal production in songbirds (Ashmore et al. 2005; Hahnloser et al. 2002). In addition to its role in motor production, HVC also exhibits auditory responses during wakefulness (Cardin and Schmidt 2003; Rauske et al. 2003; Sakata and Brainard 2008), and lesions to HVC or song motor nuclei receiving auditory input from HVC result in discrimination task-related perceptual deficits involving species-relevant stimuli including song (Brenowitz 1991; Del Negro et al. 1998; Gentner et al. 2000; Halle et al. 2002; Scharff et al. 1998) and calls (Vicario et al. 2001). Because auditory processing and perceptual function are colocalized in discrete structures, the song system is ideally suited for the study of the specific nature and influence of sensory signals in motor systems.

HVC receives auditory input from two primary sources, the caudal mesopallium (CM) and the nucleus interface of the nidopallium (NIF) (Bauer et al. 2008; Cardin et al. 2005; Yates et al. 1996). CM receives auditory input both directly and indirectly from the avian primary auditory cortical analog field L (Yates et al. 1996), and limited studies of single units during wakefulness suggest a mix of broadly responsive cells, some with modest selectivity for the bird’s own song (BOS) and others with more broadly tuned auditory properties (Bauer et al. 2008). NIF receives auditory input from CM (Yates et al. 1996) and possibly also from nucleus avalanche (Av), a recently characterized nucleus imbedded within CM that is reciprocally connected with NIF and HVC (Akutagawa and Konishi 2010). Multiunit recordings from NIF suggest a general lack of selectivity for BOS during wakefulness (Cardin and Schmidt 2004b). Auditory response properties in HVC of zebra finches have been recorded primarily during nonwaking (sleep, anesthesia) states, and responses are strongly selective for BOS (Cardin and Schmidt 2003; Margoliash 1986; Mooney 2000; Rauske et al. 2003). During wakefulness, however, auditory responses are much more variable in amplitude (Cardin and Schmidt 2003; Schmidt and Konishi 1998; Nick and Konishi 2001), and recordings obtained from multiunit electrodes suggest that HVC also lacks response selectivity for BOS (Cardin and Schmidt 2003, 2004b). Interestingly, single-unit recordings in HVC suggest that a subset of interneurons are weakly selective to BOS (Rauske et al. 2003). Similar selectivity for BOS in HVC during wakefulness has also been observed in a small number of studies recording in other songbird species (Margoliash 1986; Prather et al. 2008, 2009; Nealen and Schmidt 2006; Sakata and Brainard 2008). The tuning property of HVC neurons, while unambiguously selective to BOS during nonwaking states, is therefore more ambiguous during wakefulness, with an apparent bias toward BOS but with responsiveness to other auditory stimuli. Factors including extremely limited stimulus sets, response analyses restricted to firing rate metrics, and, with only a few exceptions (Bauer et al. 2008; Prather et al. 2008, 2009), a lack of nonbehaviorally disruptive single-unit recording techniques have hindered a detailed characterization of awake auditory
response properties in both HVC and its known auditory afferents.

In the present study, we probed the auditory properties of a large population (n = 233 neurons, 17 birds) of putative HVC interneurons (HVC\textsubscript{IN}) with a comprehensive set of conspecific and BOS stimuli in awake and freely moving adult male zebra finches. Interneurons are known to precisely shape the output of projection neurons in a multitude of systems (Ferster and Miller 2000; Merchant et al. 2008; Murayama et al. 2009; Sprio et al. 1999) and are likely to do so in HVC, given that they interact reciprocally with neurons projecting to the robust nucleus of the arcopallium (RA) in the primary song production axis and those projecting to the avian basal ganglia (area X) (Mooney and Prather 2005). The large sample of single units and comprehensive stimulus set in the present study allow the most thorough assay of BOS selectivity and multiple other basic response features yet carried out in any neural population in the song system during wakefulness. We also quantify linear receptive field properties of song system neurons for the first time during wakefulness, with the goal of evaluating whether HVC\textsubscript{IN} are sensitive to features likely to be of use during song-specific perceptual discriminations. To this end, we employ spectrottemporal receptive field analysis (STRF), a method that has proven successful in defining linear receptive field features in the ascending auditory pathway (Gill et al. 2006, 2008; Graña et al. 2009; Sen et al. 2001; Theunissen et al. 2000; Woolley et al. 2006, 2009).

MATERIALS AND METHODS

Animals

Experimental subjects were adult male zebra finches (Taeniopygia guttata) obtained from a local supplier (Canary Bird Farm, Old Bridge, NJ). All birds were at least 120 days of age at the time of experimentation. Birds were fed ad libidum and kept on a 12:12-h light-dark cycle in a colony room until several days before implantation of a chronic recording device (see below for details). All procedures documented here were approved by an Institutional Animal Care and Use Committee at the University of Pennsylvania.

Sound Recording and Presentation

For several days before surgery, the songs of birds were continuously recorded with high-quality microphones (Earthworks SRO) in a sound attenuation chamber (Acoustic Systems, Austin, TX) with custom software (Sound Analysis Pro, D. Swigger and O. Tchernichovski). Songs were recorded at a sampling rate of 44.1 kHz and played back at 20 kHz with a peak intensity level of 70 dB SPL. A stimulus set consisting of 2 or 3 motifs of song from each of 10 conspecific birds (CON), along with a recent version of the bird’s own song back at 20 kHz with a peak intensity level of 70 db SPL. A stimulus set consisting of 2 or 3 motifs of song from each of 10 conspecific birds (CON), along with a recent version of the bird’s own song was well represented (Woolley et al. 2005). To demonstrate the relative spectral and temporal modulation power inherent to our song ensemble, we calculated the modulation power spectrum (MPS) (Fig. 1E). This MPS was obtained by first decomposing the 10 CON songs into their ripple components, which are essentially the acoustic analog of visual gratings. The power density of each ripple component was then estimated and plotted on a two-dimensional (2D) Cartesian grid (Singh and Theunissen 2003; Theunissen et al. 2004a). When recordings were stable, we presented up to 50 repetitions of each stimulus in the ensemble, which guaranteed a robust data set for STRF estimation (Theunissen et al. 2000). Stimuli were presented in pseudorandom order with a random interstimulus interval between 10 and 20 s.

Chronic Recordings

Chronic recording device implantation has been described in detail previously (Nealen and Schmidt 2006). Briefly, birds had food and water removed for 1 h before an acute preparatory surgery. They were administered an intramuscular injection of a ketamine (35 mg/kg)-xylazine (7 mg/kg) mixture and placed in a stereotaxic apparatus. Before scalp incision, feathers were removed and a topical anesthetic (1% lidocaine; Copley Pharmaceutical, Canton, MA) was applied along the midline incision site. The scalp was retracted, and a custom-built (Fred Letterio, INS Machine Shop, Univ. of Pennsylvania), remotely controllable microdrive was implanted on the skull such that one to three tungsten microelectrodes (2–4 MΩ; FHC, Bowdoinham, ME) rested just above the stereotactic coordinates for HVC. Differential recordings were made between these leads and either a custom ground electrode implanted just outside of HVC (FHC) or a silver ground wire implanted between the lower skull layer and the dura. To ensure that the ground wire did not dry out, a mixture of equal parts mineral oil and paraffin was applied. This mixture was also placed in the craniotomy above HVC to prevent the exposed surface of the brain from drying out. Recording locations were verified by multiple methods. These included 1) the presence of characteristic burst activity during initial drive implantation under anesthesia, 2) the presence of premotor activity during chronic recording sessions, and 3) histological verification of electrode track location and extent within the cresyl violet-defined boundaries of HVC (Cardin and Schmidt 2003).

Immediately after implant surgery, birds were placed on a tether and habituated to the chronic recording apparatus and sound attenuation chamber for 24–48 h. Once birds were fully recovered and displayed normal feeding, perching, and vocal behaviors, chronic recordings were initiated. Electrodes were moved remotely with micrometer resolution using customized hardware and software (RP Metrix, Princeton, NJ). Neurons were not selected based on response to auditory search stimuli, thus reducing bias in terms of response tendencies and potentially allowing for highly heterogeneous sampling (see RESULTS and DISCUSSION). Because neural isolation had to be stable for long enough periods to present all of the auditory stimuli, our methods did, however, bias toward neuron types, in this case interneurons, that were stable over these time periods (Rauske et al. 2003). Online unit isolation was achieved with a sound monitor (Grass Telefactor, West Warwick, RI) and oscilloscope mode of custom neural acquisition software (A. Leonardo). Further unit isolation was performed offline with spike detection and classification algorithms described below. Spontaneous and auditory evoked neural signals were amplified 100 times (custom-built headstage) and band-pass filtered between 300 and 5,000 Hz (Brownlee 440 Amplifier; Brownlee Precision, San Jose, CA). Signals were then digitized at 25 kHz and saved for later analysis with Spike 2 (version 6; CED, Cambridge, UK). Auditory presentation always occurred with the cage lights on. A camera was used to verify that birds were awake and active during auditory presentation. Trials were discarded if any of these criteria were met: 1) birds closed their eyes for >2 s, 2) birds were not in an upright position and assumed one of several characteristic sleep postures (Low et al. 2008), and 3) birds vocalized within 5 s of an auditory presentation. These criteria are very similar to those previously employed in chronic auditory recording experiments during wakefulness (Cardin and Schmidt 2003).
Data Analysis

Spike detection and classification. All spike detection and classification were performed off-line. For each neural recording site, spike waveforms were detected and compiled into a single matrix with a Spike 2 threshold-based algorithm with a sliding 1.6-ms window. Spike waveforms were then exported to MATLAB (version 7; The MathWorks, Natick, MA), in which all subsequent data analysis occurred.

Recording sites often appeared to contain spikes from more than one neuron, which we classified into different units with the following procedure written in MATLAB. First, waveforms were interpolated by a factor of 4 with cubic splines and aligned to the peaks. Next, a random sample of 10,000 waveforms was selected. Principal components analysis (PCA) was then used to reduce the sample data to the three dimensions associated with the three largest eigenvalues of the data covariance matrix, resulting in visible clusters of data points (Fig. 1B). The reduced sample data were then automatically clustered with a Bayesian latent variable model based on a mixture of Student’s t distributions (Svensén and Bishop 2005), in which each cluster was defined by a distribution defined by a unique set of parameters. The remaining data (i.e., waveforms not included in the sample) were classified by projecting those waveforms onto the same subspace defined in the PCA and applying the clustering parameters generated by the Bayesian mixture model. Often, waveforms from electrical artifact in the recordings would be grouped into one or more clusters, so we then manually screened clusters to eliminate those that appeared to represent this noise; elimination was based on visual inspection of the mean waveform and 10–20 exemplars from each cluster. The resulting set of units had a median 0.026% (range 0–1.882%) inter-spike intervals < 1 ms, a typical refractory period for a neuron.

To verify that each cell was an HVCIN, we measured the width of the largest positive going spike peak of each isolated waveform at 25% of peak amplitude. We only accepted neurons with action potential widths that were <0.3 ms. Previous work has shown that widths below 0.35 ms are uniformly associated with a lack of antidromic activation from the two afferent targets of HVC, area X and RA (Rauske et al. 2003).

Basic auditory response characterization. Spike times were used for basic statistical characterization of auditory responsiveness. These methods were previously used to characterize multunit auditory responses in HVC (Cardin and Schmidt 2004a). Briefly, on a per stimulus basis, single auditory trial firing rate (FR) was calculated for equal duration stretches of immediate prestimulus baseline (FRBASE) and stimulus period (FRSTIM). We defined whether the response to a given stimulus was significant based on a paired t-test between FRBASE and FRSTIM measurements across all trials. For all cells we calculated change in mean firing rate (spikes/s) between the prestimulus and stimulus periods as defined above. This response strength (RS) metric was calculated on a trial-by-trial basis as FRSTIM - FRBASE; then the average value across trials was taken as RS for a given stimulus.

While RS is informative in that it reports the raw change in spike rate for a given stimulus between baseline and stimulus periods, we also wanted to report a metric that normalized spike rate in a way that...
facilitated comparison across cells with different baseline and peak firing rates. The metric we used for this purpose, RS_{INDEX}, was calculated as follows:

\[
RS_{INDEX} = \frac{FR_{TIM} - FR_{BASE}}{FR_{TIM} + FR_{BASE}}
\]

The advantage of RS_{INDEX} is that it restricts the dynamic range of response to between \(-1\) and \(1\) for all cells. An RS_{INDEX} of \(0\) indicates no difference in firing rate between prestimulus and stimulus periods. A positive RS_{INDEX} indicates an increase in firing rate during the stimulus period, and a negative RS_{INDEX} indicates a decrease in firing rate during the stimulus period. We report RS_{INDEX} for BOS as well as for CON stimuli. Along with response strength metrics, we also calculated \(d^*\) values to report bias in responsive cells for BOS relative to CON (Green and Swets 1966; Theunissen and Doupe 1998):

\[
d_{BOS-CON} = \frac{2(RS_{BOS} - RS_{CON})}{\sqrt{\sigma^2_{BOS} + \sigma^2_{CON}}}
\]

where \(\sigma^2_{BOS}\) and \(\sigma^2_{CON}\) are the variance of RS_{BOS} and RS_{CON}, respectively, across trials.

**STRF Analysis**

**STRF generation and response prediction.** For the subset of cells showing significant excitatory responses to at least one auditory stimulus based on the response strength criteria described above, we generated spectrotemporal receptive fields (STRFs) in the auditory domain (STRFPACK version 5.3, J. Gallant and F. Theunissen, http://strfpak.berkeley.edu). Briefly, the STRF in the auditory domain is the optimal linear filter that transforms a representation of a time-varying stimulus into a prediction of the firing rate as estimated by the peristimulus time histogram (PSTH) of the neuron (Theunissen et al. 2000). Auditory STRF methodology is well described elsewhere (deChamps et al. 1998; Hsu et al. 2004b; Sen et al. 2001; Theunissen et al. 2000, 2004a; Woolley et al. 2006, 2009). STRF generation, briefly, includes three steps: 1) generation of a spike-triggered average (STA) via cross-correlation between log spectrograms of sound (here CON song stimuli) and averaged time-varying spiking responses, 2) removal of stimulus autocorrelations from the STA, and 3) a regularization-cross-validation step to effectively reduce the number of parameters used to estimate the STRF (i.e., to avoid overfitting of the data). For the initial spectrographic representation of song, we chose spectral and temporal filter widths of 125 Hz (from 250 Hz to 8 kHz, which is the audible range for zebra finch hearing) and 1.27 ms, respectively, because these values have previously yielded good predictions in auditory midbrain and forebrain neurons of anesthetized birds (Singh and Theunissen 2003; Woolley et al. 2006).

Validation of the STRF involves a jackknife procedure in which the model is tested on untouched data. One point with critical relevance to our analyses was that the jackknife procedure used to generate predicted responses to individual stimuli always left the specific stimulus being predicted out of the STRF used for prediction of its response. Thus, although we used CON-derived STRFs to predict responses to BOS stimuli (see next section), this prediction was an equally valid and no more stringent test of linearity than using CON-derived STRF to predict responses to a novel CON stimulus. We quantified prediction quality with the CC ratio, averaged across stimuli (Gill et al. 2006), which we commonly refer to as response linearity. This measure is based on the correlation between predicted and actual PSTH, divided by the correlation between actual PSTH with itself trial to trial; the latter term represents a noise correction and effectively normalizes prediction quality by how predictable the spike train is irrespective of the STRF analysis. We also report non-noise-corrected CC ratios (raw CC values) for evaluation of how well the STRF model predicts responses to novel CON and BOS stimuli (see next section). We chose a Hanning window width of 21 ms for smoothing actual PSTHs, a value that has been used in previous studies with which we compare some of our results (Gill et al. 2006; Woolley et al. 2006).

To evaluate how well BOS responses were predicted by CON-derived STRF relative to novel CON responses, we calculated a z-score metric for the raw CC value (BOS CC-Z) to indicate the number of standard deviations (i.e., \(z\) values) that separate the average CON raw CC value from the BOS raw CC value. This provides a normalized statistic to evaluate whether CON-derived STRFs predict BOS responses as well as they predict CON responses across our population of recorded neurons. Negative values indicate BOS predictions that were lower than the average CON prediction, while positive values indicate BOS predictions that were higher than the average CON prediction.

\[
BOS\;CC-Z = \frac{BOS\;raw\;CC - CON\;raw\;CC}{\sigma_{CON\;raw\;CC}}
\]

where CON and BOS raw CC are the raw cross-correlation values between predicted and actual PSTHs for individual stimuli and \(\sigma_{CON\;raw\;CC}\) is the standard deviation of the individual CON raw CC values.

**STRF feature extraction.** The STRF provides a powerful tool for measuring a host of temporal, spectral, and spectrotemporal receptive field features in response to complex stimuli (Theunissen et al. 2004a). We measured excitatory and inhibitory latencies, best inhibitory and excitationary latencies, and temporal and spectral response bandwidths. Our STRFs were divided into 1-ms bins in the temporal domain and 125-Hz bands in the frequency domain between 250 Hz and 8 kHz. These values delimited the precision with which we could extract features in the temporal and spectral domains, respectively. We used a large time window (601 ms) for STRF computation because it was a value previously used in the ascending auditory pathway (Woolley et al. 2006) and because of previously demonstrated long-reaching correlations between stimulus features and response properties in song system neurons during nonwaking states (Dave and Margoliash 2000; Fortune and Margoliash 1992). All feature measurements made from STRFs are represented in Fig. 9E. Green areas on the STRF represent mean firing rate. Red and blue areas represent increases and decreases, respectively, in firing rate relative to the mean firing rate. Increasing color intensity indicates greater change in firing rate from baseline values.

Excitatory and inhibitory best frequencies describe the sound frequencies most likely to elicit a spike based on their presence (excitatory) or absence (inhibitory) prior to a spike. Excitatory and inhibitory latencies describe the amount of time between the presence of a stimulus feature (excitatory) or the absence of a stimulus feature (inhibitory) most reliably associated with a spike (Woolley et al. 2005). It helps here to think of the mirror-reversed STRF as a spectrogram of the stimulus most likely to elicit a spike at time zero. Excitatory best frequency and latency were defined as the frequency and time of peak amplitude in a 2D frequency profile through the STRF. Inhibitory best frequency and latency were defined as the frequency and time of minimum amplitude in a 2D frequency profile through the STRF.

Temporal bandwidth relates the temporal precision between the occurrences of specific frequencies and changes in firing rate. Neurons with narrower temporal bandwidths are well situated to respond to transient sound features, such as onsets and offsets, while those with broader temporal bandwidths may be well situated to encode more slowly evolving stimulus features (Woolley et al. 2009). We defined temporal bandwidth as the full width at 50% peak amplitude of the peak excitatory and inhibitory regions in a 2D temporal profile of the STRF at best frequency (Nagel and Doupe 2008).

Spectral bandwidth defines the range of frequencies that are reliably associated with increases, in the case of excitation, or decreases,
in the case of inhibition, of spike rate relative to mean firing rate. We measured spectral bandwidth separately for excitation and inhibition and defined it as the full width at half peak amplitude in a 2D frequency profile through the STRF at the time of peak excitation or inhibition, respectively (Nagel and Doupe 2008; Sen et al. 2001).

RESULTS

The high degree of auditory selectivity to the BOS and the virtual omnipresence of stereotyped auditory responses in the zebra finch song system during nonwaking states is well established (Margoliash 1986; Mooney 2000; Sutter and Margoliash 1994). While it is also well established that auditory responses in the zebra finch are suppressed for some time upon arousal (Cardin and Schmidt 2003; Nick and Konishi 2001; Schmidt and Konishi 1998), there are conflicting reports concerning the degree of BOS selectivity, consistency, and vigor of responses when present during wakefulness (Cardin and Schmidt 2003; Nick and Konishi 2001; Schmidt and Konishi 1998). Previous studies have been limited in terms of stimuli presented (usually just BOS and 1 or 2 other complex stimuli), small numbers of single units that, when present, were usually recorded with techniques disruptive to ongoing behavior, and analysis of responses limited to spike rate-derived measures.

In an effort to provide a rigorous characterization of awake auditory response properties in adult zebra finch HVC at the single-neuron level, we sampled from a large number of neurons (n = 233 single units from 103 recording sites in 17 birds) and used a comprehensive stimulus set for each recorded neuron that consisted of 10 conspecific song (CON) stimuli and 1 BOS stimulus. CON stimuli used in the present study were collected from our colony several years before the present study was initiated. Although we cannot rule out prior exposure to these songs, subjects were unlikely to have been exposed to these stimuli. Recordings in HVC focused entirely on responses in neurons that were defined as putative interneurons, a class of neuron well known to shape the output of projection neurons across many neural systems. All neurons in the present study had waveform widths at 25% peak amplitude that were smaller than 0.35 ms (0.176 ± 0.003 ms) (Fig. 1D; Table 1). Based on a previous study, this cutoff is sufficient to provide a clear separation from antidromically identified projection neurons, which have much wider waveforms (Rauske et al. 2003). Figure 1D shows the complete lack of overlap between the distribution of spike widths for the cells included in the present study and those obtained in the Rauske et al. study for both types of HVC projection neurons. From this point forward, we refer to these neurons as putative HVC interneurons or HVCIN.

Throughout, we characterize overall responsiveness using response strength (RS), which is the increase in spike rate to a given stimulus over prestimulus rate, and RSINDEX, which is normalized to fall between −1 and 1 (see MATERIALS AND METHODS). To provide a qualitative description of RSINDEX, the distribution of BOS RSINDEX values for our population of cells is displayed in Fig. 2, along with example raster plots from three different neurons to illustrate the range of recorded responses to BOS. We use the term “bias” simply to refer to a relatively greater spike rate in response to one stimulus over another and the term “selectivity” to denote a significant response to one stimulus type only. For a given unit, we quantify responsiveness to conspecific song by averaging RS

![Fig. 2. HVCIN show three general classes of responses to song stimuli. Raster plot (top) and peristimulus time histogram (PSTH, middle) for each of 3 example neurons to illustrate the general types of responses obtained from HVCIN during the presentation of a song stimulus in awake birds. In all 3 examples, the stimulus was the BOS. Example on left ("Suppressive") represents a neuron that shows a significant decrease in spike rate during presentation of the song stimulus. This class of neurons showed suppressive responses to all song stimuli. Example at center ("Unresponsive") represents a neuron lacking responsiveness to the song stimulus. This class of neurons did not respond to any of the song stimuli that were presented. Example on right ("Excitatory") represents neurons showing excitatory responses to 1 or more song stimuli. Bottom: distribution of RSINDEX values following presentation of BOS for all 233 neurons recorded in this study. Arrows illustrate the RSvalues for the 3 example cells. These values were ~0.44, ~0.05, and 0.71 for suppressive, unresponsive, and excitatory neurons, respectively.](http://jn.physiology.org/)

### Table 1. Nonauditory characteristics of HVC interneurons

<table>
<thead>
<tr>
<th>Condition</th>
<th>Unresponsive (n = 28)</th>
<th>Suppressive (n = 51)</th>
<th>Excitatory (n = 154)</th>
<th>Total Population (n = 233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spike width, ms</td>
<td>0.200 ± 0.008</td>
<td>0.181 ± 0.005</td>
<td>0.171 ± 0.003</td>
<td>0.176 ± 0.003</td>
</tr>
<tr>
<td>Spontaneous spike rate, Hz</td>
<td>3.767 ± 1.048</td>
<td>4.163 ± 0.543</td>
<td>4.066 ± 0.312</td>
<td>4.052 ± 0.268</td>
</tr>
</tbody>
</table>

Values are means ± SE.
and RS_INDEX across all 10 CON stimuli. We characterize bias for BOS over CON by comparing RS and RS_INDEX values for the respective stimuli. For direct measurement of BOS bias and comparison to other studies, we use d', a measure that is normalized by variability in responses (see MATERIALS AND METHODS). We report means ± SE and use Student’s t-test for statistical significance among response measurements.

Auditory Responses to BOS and CON Across the HVC_IN Population

The HVC_IN population as a whole exhibited several notable characteristics (Fig. 3, Table 2). First, as a population neurons responded to both BOS and CON, with an average RS to BOS of 2.220 ± 0.317 spikes/s and a RS to CON of 0.579 ± 0.174 spikes/s. The RS_INDEX values for BOS and CON were 0.145 ± 0.018 and 0.044 ± 0.013, respectively. Second, the population was biased toward BOS, with average BOS RS_INDEX being significantly greater than for CON (P < 0.0001, paired t-test), while average d' was 0.640 ± 0.073, significantly above 0 (P < 0.0001). Third, there was a significant positive correlation between BOS RS_INDEX and mean CON RS_INDEX (Pearson’s r = 0.605, P < 0.0001), suggesting a general excitability (or responsiveness) property that varied across the population and affected responses to both BOS and CON (Fig. 3A). We pursued this last characteristic further by randomly picking a CON stimulus for each cell (“Random CON”) and computing the correlation between that RS_INDEX and the average for the remaining CON set (“Remaining CON”). This yielded an even stronger correlation (Fig. 3B; 0.755, P < 0.0001), suggesting a responsiveness property that was especially determinant of CON responses (albeit BOS response as well); we will occasionally refer to this characteristic as the population’s “response correlation.”

These results suggest that both BOS and CON responses scale in a relatively linear manner with the overall excitability of the unit at the time it is recorded.

HVC_IN in Awake Birds Show a Heterogeneity of Auditory Responses

To further probe the response properties of individual neurons, we applied a t-test to each unit’s response to a given stimulus across trials and classified each unit according to whether it responded significantly (at α = 0.05) to either BOS or at least one CON. To facilitate the analysis of neuronal response properties in our sample, we used this statistical cutoff to divide all 233 units into 5 classes: 1) no significant increase or decrease in spike rate to any stimulus (“unresponsive,” n = 28 units); 2) no statistically significant excitatory responses and at least one significant decrease in spike rate (“suppressive,” n = 51); 3) excitatory responses to BOS and no CON stimuli (“BOS-ONLY,” n = 15); 4) excitatory responses to at least one CON and not BOS (“CON-ONLY,” n = 64), and 5) significant excitatory responses to both BOS and at least one CON (“BOS-CON,” n = 75). It is important to note that we present this classification in order to more thoroughly describe the data along the spectrum of responses; qualitative examination of the data does not suggest distinct neural classes per se (Fig. 3).

It was not uncommon for us to simultaneously record cells from different classes at the same electrode site (see Fig. 1B), highlighting the heterogeneity of auditory response types in awake HVC suggested in a previous study (Rauske et al. 2003). Of the 103 recording sites, we identified 80 that yielded >1 unit from the clustering. Of these, 56 sites had units from >1 of the 5 classes. Importantly, a number of suppressive neurons came from sites without any excitatory cells, indicating that

Table 2. Auditory characteristics of HVC interneurons

<table>
<thead>
<tr>
<th>Class</th>
<th>Unresponsive (n = 28)</th>
<th>Suppressive (n = 51)</th>
<th>BOS-ONLY Excitatory (n = 15)</th>
<th>CON-ONLY Excitatory (n = 64)</th>
<th>BOS-CON Excitatory (n = 75)</th>
<th>Total Population (n = 233)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOS RS_INDEX</td>
<td>0.074 ± 0.027</td>
<td>−0.108 ± 0.026</td>
<td>0.396 ± 0.055</td>
<td>0.028 ± 0.021</td>
<td>0.394 ± 0.023</td>
<td>0.145 ± 0.018</td>
</tr>
<tr>
<td>CON RS_INDEX</td>
<td>0.010 ± 0.023</td>
<td>−0.193 ± 0.016</td>
<td>−0.084 ± 0.027</td>
<td>0.072 ± 0.010</td>
<td>0.219 ± 0.020</td>
<td>0.044 ± 0.013</td>
</tr>
<tr>
<td>d’</td>
<td>0.186 ± 0.091</td>
<td>0.328 ± 0.093</td>
<td>2.139 ± 0.291</td>
<td>−0.167 ± 0.079</td>
<td>1.398 ± 0.139</td>
<td>0.640 ± 0.075</td>
</tr>
<tr>
<td>Linearity (mean CC ratio)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.193 ± 0.025</td>
<td>0.320 ± 0.019</td>
<td>0.262 ± 0.016</td>
</tr>
</tbody>
</table>

Values are means ± SE. BOS, bird’s own song; CON, conspecific song; RS_INDEX, normalized response strength; d’, bias toward BOS.

J Neurophysiol • doi:10.1152/jn.01003.2009 • www.jn.org
suppression was not simply an artifact caused by large increases in spike rates in excitatory neurons obscuring the detection of spikes in those cells.

HVC comprises a heterogeneity of classes of neurons (Fortune and Margoliash 1995; Nixdorf et al. 1989), and a previous study (Rauske et al. 2003) suggested that nonresponsive interneurons in HVC tended to have smaller spike waveforms (0.13–0.16 ms) than neurons that showed excitatory responses to BOS (0.18–0.23 ms). We therefore measured spike waveform and spontaneous activity for the suppressive (n = 51) and unresponsive (n = 28) groups of units. We also grouped together all of the “excitatory” neurons that showed a response to at least one auditory stimulus (n = 154) and measured their spike width and spontaneous activity. Analysis of variance (ANOVA) revealed no significant difference between categories in terms of spontaneous activity (P = 0.525), but the ANOVA did reveal a significant difference in spike waveform width, specifically between unresponsive neurons (0.200 ± 0.008 ms) and excitatory neurons (0.171 ± 0.003 ms; P < 0.05) by the multiple-comparisons test (see Table 1). Suppressive neurons (0.181 ± 0.005 ms) were not significantly different from either of the two other groups. The significance of these differences is presently unclear given that the Rauske et al. (2003) study showed a trend that was opposite to ours.

Response Properties of Unresponsive and Suppressive HVC<sub>IN</sub> Units

While we failed to determine statistically significant excitatory responses for 79 unresponsive and suppressive units, qualitative examination of the data suggested that the two basic properties observed across neurons as a whole, BOS bias and response correlation, appeared to extend across even this population (Fig. 3). Thus, while most of this study focused on the attributes of cells demonstrating excitatory responses to one or more BOS and CON stimuli, these cells merit a quantitative description of their properties.

As a group, unresponsive neurons showed a weak but significant increase in spike rate to BOS (RS = 0.342 ± 0.156 spikes/s, RS<sub>INDEX</sub> = 0.074 ± 0.027; P < 0.05), although, by definition, no individual neuron in this category showed a statistically significant firing rate change. Interestingly, despite the response to BOS, as a group these cells did not show a significant response to CON (RS = 0.087 ± 0.063 spikes/s, RS<sub>INDEX</sub> = 0.010 ± 0.023; P = 0.174), suggesting a subtle BOS bias (see Fig. 3A). Consistent with this observation, the difference in RS<sub>INDEX</sub> values between BOS and CON was significant (P < 0.05), as was the average d′, which was small but nevertheless significantly positive (0.186 ± 0.090; P < 0.05). Thus, even though these units failed to show statistically significant excitatory responses to any one stimulus, as a group they did show a weak but significant BOS response and selectivity. Unsurprisingly, given the lack of CON response, the group failed to show response correlations between BOS and CON, as well as between CON and Random CON (respective Pearson’s r = 0.232 and 0.149, P = 0.234 and 0.450).

Suppressive neurons, as expected by their definition, showed significant decreases in spike rate (P < 0.0001). RS and RS<sub>INDEX</sub> averages to BOS were −0.326 ± 0.188 spikes/s and −0.108 ± 0.026, respectively. RS and RS<sub>INDEX</sub> averages to CON were −1.188 ± 0.197 spikes/s and −0.193 ± 0.016, respectively. RS<sub>INDEX</sub> to BOS was significantly greater than to CON (P < 0.0005), suggesting that neurons were less suppressed to BOS than they were to CON. This feature is also apparent from the significant and positive d′ measure of 0.328 ± 0.093 (P < 0.0005). Interestingly, this BOS bias was stronger than what we measured among unresponsive cells. This group also showed significant response correlations between BOS and CON RS<sub>INDEX</sub>, as well as between a randomly chosen CON and the remaining CON stimuli (r = 0.510 and 0.617, respectively, P < 0.0001 in both cases). Thus the data suggest a common responsiveness property linked with the degree of responses to all stimuli presented across suppressive neurons. Coupled with the BOS bias, suppressive neurons therefore show the same two basic response properties that we observe across the neural population as a whole.

Response Properties of BOS-ONLY Neurons

The typical response properties of BOS-ONLY cells are well represented by the example in Fig. 4A. This cell had relatively modest, phasic excitatory peaks in response to BOS (RS<sub>INDEX</sub> = 0.455) that were reliably present from trial to trial at specific time points throughout the stimulus. These phasic excitatory peaks were sometimes followed by transient suppression below baseline levels, a common response property of cells in this class. This example cell also clearly showed response peaks during the second and third motifs (M2 and M3 in Fig. 4A, top) that were larger than during the first motif (M1). This increase in response magnitude across motifs was a response feature that we sometimes observed in BOS-ONLY cells. In contrast to its response to BOS, this cell showed no response peaks to any of the representative CON stimuli.

Overall, BOS-ONLY neurons showed an increase in spike rate during BOS stimulation that was on par with the population as a whole, with an average RS of 3.267 ± 0.857 spikes/s (range: 0.757 to 14.159). By definition, cells in this group did not show significant responses to CON stimuli, with a mean conspecific RS of −0.463 ± 0.243 spikes/s. Average RS<sub>INDEX</sub> values for BOS and CON were 0.396 ± 0.055 and −0.084 ± 0.027, respectively. For the BOS vs. mean CON response comparison, mean d′ was significantly above zero, 2.139 ± 0.291, (P < 0.001), and over three times the population average (0.640). Importantly, the RS<sub>INDEX</sub> for BOS in this population was similar to what we found for BOS-CON (mean 0.396 vs. 0.394, see below), and the small difference failed to reach significance (P = 0.972).

Thus BOS-ONLY neurons have strong selectivity for BOS by definition, but increases in spike rate to this stimulus are relatively modest.

Response Properties in CON-ONLY Neurons

An example of a CON-ONLY neuron is shown in Fig. 4B. Unlike BOS-ONLY cells, these units tended to respond weakly to their preferred stimuli, with an average RS and RS<sub>INDEX</sub> to CON of only 0.413 ± 0.067 spikes/s and 0.072 ± 0.010, respectively. Furthermore, unlike BOS-CON cells (see below), the majority (36/64) showed statistically significant responses to just one conspecific song (Fig. 5B). Even when analysis was restricted to conspecific stimuli that elicited the strongest response from cells, CON-ONLY units showed significantly lower RS and RS<sub>INDEX</sub> values than BOS-CON units (see...
Fig. 4. Auditory response profile of typical BOS-ONLY and CON-ONLY neurons: 2 examples from different classes of neurons (see text for how classes were defined). A: like the majority of BOS-ONLY neurons, this exemplar shows a modest, phasic excitatory response to BOS ($RS\text{INDEX} = 0.12$, top). The BOS stimulus used in this experiment consisted of 4 individual motifs (M1–M4). The BOS response of this neuron was stronger for M2–M4 than for M1, a feature sometimes observed in BOS-ONLY neurons. Response to CON stimuli (bottom) tended to be slightly suppressed relative to baseline ($RS\text{INDEX} = -0.093$). Moderate suppression to CON stimuli was common in BOS-ONLY neurons. B: CON-ONLY neuron demonstrating a moderate response increase to CON ($RS\text{INDEX} = 0.379$, bottom) and a slightly suppressed response to BOS ($RS\text{INDEX} = -0.136$, top).

below). This group also failed to show a response correlation among conspecific stimuli ($r = 0.0005$, $P = 0.997$), unlike the suppressive and BOS-CON groups (see below).

Thus CON-ONLY neurons constitute a group that respond weakly to conspecific song overall. By definition, this group does not significantly respond to BOS. Thus it may be that these neurons are not selective for CON over BOS per se, but rather are weakly responding cells with receptive fields that by chance do not match features in BOS as much as they do features in one of the conspecific stimuli we present.

**Response Properties in BOS-CON Neurons**

Besides responding to BOS and at least one CON stimulus, response properties of BOS-CON neurons tended to be diverse. To capture the range of responses observed in these neurons, we depict two examples in Figs. 6 and 7 that represent most of the response features observed in this class of HVC$_{\text{IN}}$. Both cells showed BOS RS values (34.891 and 27.884 spikes/s, respectively) that were substantially higher than average, including BOS-ONLY cells (see previous section). The cell depicted in Fig. 6 showed reliable phasic response peaks to both BOS (Fig. 6B, top) and CON stimuli (Fig. 6B, middle and bottom). This particular cell did not show a strong preference for BOS compared with CON stimuli ($d' = 0.592$; BOS vs. mean CON). The cell depicted in Fig. 7 had somewhat different, yet still common to the BOS-CON class, response characteristics. Responses to CON stimuli (Fig. 7B, middle and bottom) were well above baseline but nevertheless substantially weaker than those elicited by BOS (Fig. 7B, top). This selectivity for BOS was supported by a moderately high $d'$ value relative to the mean CON response ($d' = 3.601$). Finally, while there were clearly consistent response peaks to both BOS and CON stimuli, this cell had higher background activity and slightly less precision than the cell shown in Fig. 6.

As a population, BOS-CON neurons showed a diversity of response characteristics in terms of both the degree of BOS bias as well as the overall strength of response. The response strength to BOS ranged from 0.212 to 34.891 spikes/s with a mean of 6.104 spikes/s, while mean $RS\text{INDEX}$ to BOS was 0.394 ± 0.023. Responses to CON were significant by definition but were weaker on average than to BOS, with a mean RS of 2.381 ± 0.439 spikes/s and $RS\text{INDEX}$ of 0.219 ± 0.019 ($P < 0.0001$). BOS bias at the population level for this group was also verified by measurement of $d'$, with an average of $1.398 ± 0.139$ ($P < 0.0001$). In contrast to the CON-ONLY group, the majority of BOS-CON neurons (43/75) responded significantly to at least 5 conspecific stimuli (Fig. 5A), with 22 units responding to all 10 stimuli. BOS-CON neurons were also more responsive than CON-ONLY neurons even when comparisons were restricted to the CON stimulus that elicited the strongest increase in spike rate. RS average was 3.937 ± 0.520 for BOS-CON neurons and 1.600 ± 0.182 spikes/s for CON-ONLY neurons. $RS\text{INDEX}$ values were 0.367 ± 0.021 and 0.281 ± 0.019 ($RS\text{INDEX}$ comparison $P < 0.005$, paired $t$-test) for BOS-CON and CON-ONLY neurons, respectively.

Finally, we found significant response correlations between BOS and CON, as well as between Random CON and Remaining CON ($r = 0.654$ and 0.765, $P < 0.0001$; Fig. 3). This suggests that the responsiveness property of the cells drove the overall excitatory response to all stimuli.
The STRF analysis has not been used previously to analyze auditory responses and thus show an extreme degree of nonlinearity, compared with field L of anesthetized birds, where neurons tend to have properties. Alternatively, it could be that this is a population response only to BOS. This feature strongly distinguished them from BOS-ONLY cells, which showed excitatory responses only to BOS, and on average, than BOS-CON cells. Note that the mode of the distribution is 1, which is markedly different from the distribution of BOS-CON cells.

BOS-CON neurons thus show the two basic properties we observe across the population, BOS bias and the correlation between BOS and CON responses. This group has relatively strong increases in spike rate in response to CON, which contrasts with the generally weak responses that are observed in CON-ONLY neurons.

A Subpopulation of CON- Responding Neurons Showed Responses Well Predicted by Linear STRF

While BOS bias was evident at the population level, 72 of 154 cells with excitatory responses showed d’ values below 0.5 relative to the mean CON response (many of these were CON-ONLY cells). In previous studies, cells with d’ below 0.5 were deemed to be unselective with respect to BOS (Cardin and Schmidt 2003; Solis and Doupe 1997). The significant CON response, the small BOS bias, and the response correlations all suggested the possibility that the receptive fields of these neurons had “low-order” properties that could be captured by established linear analysis methods.

STRF is a method by which the linear portion of responses to complex, time-varying stimuli, such as songs and other natural sounds, can be estimated and used to predict responses to other such stimuli. Introduction of static nonlinearities in the initial spectrographic representation of song has aided response prediction in areas of the songbird auditory forebrain, where neurons are tuned to features inherent to CON song compared with complex modulated noise stimuli (Grace et al. 2003; Theunissen et al. 2000). However, because auditory responses in HVC under anesthesia are highly biased toward the BOS, and thus show an extreme degree of nonlinearity, CON-based STRF analysis has not been used previously to analyze auditory responses in this area (Theunissen and Doupe 1998). The existence of a relatively large population of HVC neurons that respond to CON stimuli in awake birds motivated us to employ STRF-based methods to describe the linear tuning properties of these cells. We measured mean CC ratio (response linearity) as the normalized (see MATERIALS AND METHODS for details) average cross-correlation between the actual time-varying response of neurons to CON stimuli and the response predicted by STRF generated from responses to other CON stimuli.

Figure 8A shows the distribution of response linearity (CC ratio) across the population of BOS-CON and CON-ONLY cells. The population showed a mean response linearity of 0.262 ± 0.016, which was only slightly lower than what has been reported in various areas of the ascending auditory pathway in anesthetized birds with similar CON stimulus sets, spectrogram time-frequency parameters, and STRF smoothing parameters (Gill et al. 2006). In contrast to CON-ONLY and BOS-CON neurons, BOS-ONLY cells showed much lower linearity in response to CON stimuli (0.181 ± 0.044), reminiscent of the nonlinearity that is observed in HVC responses under anesthesia. Because of their lack of significant excitatory response to the CON stimuli we used to generate our STRFs, BOS-ONLY cells were not included in the distribution depicted in Fig. 8 or in any further STRF-based analyses.

To extract spectral and temporal response parameters from the STRFs of individual neurons, we chose to focus on neurons whose response linearity scores were at least 0.3. This criterion is similar to that previously used in a study of neural properties under anesthesia in the ascending auditory pathway (Woolley et al. 2009). Using this criterion, we were left with a subgroup of cells (n = 56/139) that we refer to as linear neurons from this point forward.

BOS Bias in Linear Cells Is Linked with a Nonlinear Component

Many HVC neurons responding to BOS also responded to CON, but with a bias toward BOS. In fact we found such a bias for BOS in every cell class except CON-ONLY, which by definition must show a bias against BOS. Combining neurons from the BOS-CON and CON-only groups (n = 139), we found the average RS INDEX for BOS and CON to be 0.226 ± 0.022 and 0.151 ± 0.013 (P < 0.0001), respectively, with an average d’ of 0.677 ± 0.106 (P < 0.0001). This BOS bias was also present among cells we defined as linear (i.e., the subset of these cells whose CC ratio > 0.3, see above). For these linear cells, RS INDEX was 0.312 ± 0.031 for BOS and 0.200 ± 0.031 for CON (P < 0.0001) with a d’ of 0.990 ± 0.186 (P < 0.0001). This bias for BOS in these linear cells contrasts with properties of linear responsive neurons in primary auditory area field L of anesthetized birds, where neurons tend to have an average d’ <0 and therefore a slight bias against BOS (Theunissen et al. 2004b).

The existence of BOS bias in HVC neurons begs the question of how much of the BOS bias in linear HVC cells is due to a nonlinear component similar to what could be driving responses in BOS-ONLY cells; under this scenario, BOS responses could reflect a mixture of linear and nonlinear properties. Alternatively, it could be that this is a population more similar to field L, with a linear component that underlies BOS responses as much as it does CON, but with receptive...
fields that are simply more oriented toward BOS. Here, the increased neuronal response to BOS would be simply the consequence of BOS containing more spectrotemporal features in the cell’s receptive field.

To distinguish between these two factors, we compared the raw CC scores for BOS and CON (Fig. 8B) to measure how well predicted a cell’s response is to each individual stimulus on a song-by-song basis (see MATERIALS AND METHODS). If BOS responses were generally less linear, we would expect CC scores to be lower for BOS than for CON. On the other hand, if BOS responses were just as linear and the bias was due to disproportionate representation of BOS features in the STRFs, we would expect BOS CC scores to be the same or stronger. Indeed, we found that the average BOS CC score (0.167 ± 0.023) was significantly lower than the average CON CC score (0.259 ± 0.021) (P < 0.0001). On a neuron-by-neuron basis, BOS CC scores were smaller than CON CC scores for 45 of 56 linear neurons. Importantly, despite their relative weakness, BOS CC scores were still strongly correlated with CON CC scores (Pearson’s r = 0.766, P < 0.0001), suggesting that, in addition to the nonlinearity, at least some portion of the BOS response was driven by a linear property shared with CON responses.

To examine whether the observed BOS bias in response strength was linked directly with the nonlinear component (Fig. 8C), we computed, on a neuron-by-neuron basis, the correlation between the strength of BOS bias (d’) and the relative weakness of BOS CC scores, which we measured using the BOS CC-Z score (see MATERIALS AND METHODS). Consistent with a BOS bias being linked to the degree of nonlinearity, we found a negative correlation (Pearson’s r = −0.607, P < 0.0001) between d’ (i.e., BOS excitation relative to CON) and the ability to predict BOS relative to a CON-based STRF (i.e., BOS CC-Z score).

We next asked whether the linear component of each cell’s receptive field was more oriented toward BOS than to CON, in addition to that cell’s response being more nonlinear overall. Here, BOS bias could be explained by both hypothetical factors: a nonlinearity in the response as well as a representation in the STRF that was disproportionate for BOS features. If there was a disproportionate representation of BOS features in the STRF, cells with a CC-Z score ≈ 0, i.e., where BOS and CON were equally well predicted by the STRF, would be predicted to have a significant positive d’, i.e., cells with comparably predictable BOS responses would have larger increases in spike rate because BOS contained more features from those STRFs than CON overall. We examined this possibility by generating a regression between d’ and BOS CC-Z score and testing the location of the y-intercept, which effectively estimates average d’ when BOS CC-Z score = 0. The regression yielded an intercept of 0.315 ± 0.192 (estimate ± SE), which indicated a trend toward an intercept > 0 that failed...
to reach significance at the $\alpha = 0.05$ level by a $t$-test on the intercept estimate ($P = 0.053$; Devore 2004). This result suggested that any bias toward BOS in STRF features was likely to be small or nonexistent.

Taken together, the data as a whole indicate that cells with linear receptive fields for conspecific song also possess an additional nonlinear component in their responses to BOS that explains almost all of the bias in response strength for those stimuli. Qualitative examination of the data suggests that the degree of nonlinear bias varies continuously across these cells; Fig. 8 shows the distribution, while Figs. 6 and 7 show examples of cells from extreme ends.

**STRFs and Receptive Field Feature Extraction**

We next examined in more detail the spectrotemporal features represented by neurons with responses well predicted by STRFs. Analysis of 2D temporal and spectral slices through the STRF allows extraction of “classical” receptive field properties including, but not limited to, spike latencies, precision of the relationship between specific stimulus features and modulation of spike rate, and the spectral frequencies that a neuron is reliably sensitive to. We obtained receptive field features from all linear neurons (56 nonselective neurons with mean CC ratios >0.3). The extraction of all features is depicted on the example STRF in Fig. 9E, and the extracted values are summarized in Table 3.

**Spectral Receptive Field Features in Linear Neurons**

We found that the population of linear neurons showed best excitatory (range: 0.250 kHz to 7.065 kHz) and inhibitory (range: 0.918 kHz to 7.733 kHz) frequencies that nearly spanned the entire range of frequencies represented in our STRF analysis (0.250 kHz to 8 kHz). This range of best frequencies also covered the range of frequencies audible to the zebra finch (0.05 kHz to 7 kHz) (Okanoya and Dooling 1987) (Fig. 9A). There was a significant tendency for best excitatory frequencies (2.412 ± 0.220 kHz) to be lower than best inhibitory frequencies (4.101 ± 0.272 kHz) (paired $t$-test; $P < 0.0001$). The STRFs in Fig. 9E and Supplemental Fig. S1A2 are representative of this trend in the population.\(^1\) In addition, individual neurons tended to show broadband frequency sensitivity. As with best frequency, mean spectral bandwidth in linear HVC$_{\text{IN}}$ differed significantly between excitation (1.732 ± 0.187 kHz) and inhibition (2.352 ± 0.247 kHz) ($P < 0.01$) (Fig. 9B). The tendency for inhibition to have wider spectral bandwidth than excitation is well illustrated by the STRF shown in Supplemental Fig. S1C1.

---

\(^1\) Supplemental Material for this article is available online at the Journal website.
Values ranged from 0 to 0.85, and cells with a mean CC ratio
linearity scores (mean CC ratio) for all BOS-CON and CON-ONLY cells (n = 139). Values ranged from 0 to 0.85, and cells with a mean CC ratio > 0.3 (n = 56) were considered as linear and were used for further analyses. B: the ability of CON-based STRFs to predict BOS responses varied significantly among linear neurons. Neurons with mean CC ratios >0.3 were used to compare raw CC values for BOS with those for CON. As shown, neurons that were very good at predicting novel CON responses were also good at predicting BOS responses. The skew of the population under the unity line demonstrates that CON CC values were generally higher than BOS CC values. The 2 example neurons depicted in Figs. 6 and 7 are highlighted in this plot. C: BOS bias is linked to the nonlinear component of the response. Response selectivity (d') to BOS is plotted against BOS CC-Z score, which is the normalized difference between CON and BOS CC. A BOS CC-Z score value of 0 signifies that a CON-based STRF is equally able to predict the response to BOS and CON. Negative values signify that the STRF is better at predicting CON than BOS. Consistent with a BOS bias being linked to the degree of nonlinearity, we observed a negative correlation between d' and BOS CC-Z, indicating that BOS selectivity (a positive d') is associated with low linearity. The y-intercept of the regression line crosses near 0, suggesting that when BOS and CON are equally predictable any bias toward BOS in STRF features is likely to be minimal (see text for details). This suggests that the nonlinearity explains most, if not all, of the bias in response toward BOS. As in B, cells in Figs. 6 and 7 are highlighted in this distribution.

Temporal Receptive Field Features in Linear Neurons

Along with the spectral parameters described above, we extracted two types of temporal parameters from the STRFs of linear neurons: latency and temporal bandwidth.

Latency. Excitatory and inhibitory spike latencies indicate the most reliable time interval between the presence (excitation) or absence (inhibition) of a given frequency in the song stimulus and a spike event. We found that excitatory (17.929 ± 1.295 ms) and inhibitory (16.232 ± 1.002 ms) latencies were not significantly different at the population level (P = 0.283). The scatterplot in Fig. 9C illustrates that there was a near equal probability for each response component to lead the other (peak excitation led peak inhibition in 29/56 cells). This contrasts with latency properties recorded in lower regions of the auditory pathway, where most auditory neurons are thought to act as onset detectors, or have onset bias, and where inhibition typically lags excitation (Woolley et al. 2009).

Temporal bandwidth. This feature (Fig. 9D) measures the precision of the temporal relationship between specific spectral frequencies in the song stimuli and neural spiking. Low temporal bandwidths indicate a temporally precise relationship (i.e., low jitter) between a neuron’s response and the presence of a specific spectral frequency. It is important to note that the width of the smoothing window we used for STRFs (21 ms, see MATERIALS AND METHODS), will bias extraction away from more temporally precise relationships. Nonetheless, the population of neurons recorded in the present study showed a broad range of temporal bandwidths (4–38 ms for excitatory, 4–56 ms for inhibitory) with no significant difference in the mean bandwidths for excitatory and inhibitory regions (14.786 ± 1.097 ms for excitatory, 15.768 ± 1.299 ms for inhibitory; P = 0.361). The STRF of the cell shown in Supplemental Fig. S1D1 is an example of a cell with a narrow temporal bandwidth (9 ms excitatory, 8 ms inhibitory) at the low end of the range observed for linear nonselective neurons. This was typical for cells with a sharp relationship between stimulus and spiking response. In contrast, the STRF of the cell profiled in Supplemental Fig. S1D2 shows wide temporal bandwidth (21 ms excitatory, 27 ms inhibitory) near the high end of the range. This was typical for cells that had a relatively imprecise temporal relationship between stimulus features and spiking response.

DISCUSSION

Heterogeneity of Auditory Response Properties in HVC of Awake Songbirds

The present study is the first to record from a large number of single units in a song vocal motor nucleus during presentation of comprehensive sets of complex, natural stimuli in awake, freely behaving birds. The heterogeneity in response properties we observed in the population of HVC IN stands in sharp contrast to the vast majority of studies in HVC where auditory responses in nonwaking birds show vigorous, excitatory, BOS-selective responses that are highly homogeneous in their response properties across time and (at least in species with highly stereotyped songs) recording sites (Cardin and Schmidt 2003; Rauske et al. 2003; Sutter and Margoliash 1994). Furthermore, the wide range of response properties shown by HVC IN in the present study provides a possible
functional correlate to the diversity previously described for this cell type based on morphology, calcium-binding protein profile, and intrinsic firing properties (Mooney 2000; Nixdorf et al. 1989; Wild et al. 2005).

Previous studies have opened the possibility that factors such as repertoire sharing (Lehongre and Del Negro 2009) and territorial interactions requiring rapid song matching of neighboring birds (Margoliash 1986; Prather et al. 2008, 2009) may be positively correlated with responses during wakefulness in HVC to stimuli other than BOS. The results of the present study support previous multiunit studies showing that broad and non-BOS-selective responses also occur in the zebra finch (Cardin and Schmidt 2003, 2004b), a species with a highly stereotyped song and no overt repertoire sharing or territorial behavior (Zann 1993). Nonetheless, we also found cells showing a high degree of BOS selectivity, which corroborates the results of previous studies in species with diverse social ecology and repertoire sites (Margoliash 1986; Nealen and Schmidt 2006; Rauske et al. 2003; Sakata and Brainard 2008; Prather et al. 2008, 2009). In fact, the majority of neurons recorded had a bias for BOS even though many of these neurons also showed significant responses to CON.

In addition to the two-thirds of HVCIN showing excitatory responses to song stimuli, one-third of cells either showed only suppressive responses or were completely unresponsive, corroborating hints of their existence from a limited population recorded in a previous study (Rauske et al. 2003). The existence of HVCIN showing no auditory responses during wakefulness may be related to previous findings showing a complete suppression of responsiveness in HVC following arousal (Cardin and Schmidt 2003, 2004a; Nick and Konishi 2001; Schmidt and Konishi 1998). In the present study, many of the nonresponsive HVCIN appeared to retain this nonresponsive property stably across the period of stimulus presentation (60–100

Table 3. Summary of STRF-derived features of linear HVC interneurons

<table>
<thead>
<tr>
<th>Nonselective (CON-ONLY and BOS-CON) Linear Neurons (n = 56)</th>
<th>Excitatory</th>
<th>Inhibitory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best frequency, kHz</td>
<td>2.412 ± 0.220</td>
<td>4.101 ± 0.272</td>
</tr>
<tr>
<td>Latency, ms</td>
<td>17.929 ± 1.295</td>
<td>16.232 ± 1.002</td>
</tr>
<tr>
<td>Spectral bandwidth, kHz</td>
<td>1.732 ± 0.187</td>
<td>2.352 ± 0.247</td>
</tr>
<tr>
<td>Temporal bandwidth, ms</td>
<td>14.786 ± 1.097</td>
<td>15.768 ± 1.299</td>
</tr>
</tbody>
</table>

Values are means ± SE. STRF, spectrotemporal receptive field analysis.
properties in many HVCIN. Somewhat paradoxically, spike tent with providing the input that drives the linear response during "up" states. Because neurons were not followed longitudinally in the present study, it is conceivable that some of the diversity in response characteristics was caused by state-dependent modulations in response amplitude (Cardin and Schmidt 2004b; Shea and Margoliash 2003). However, the fact that both nonresponsive and depressed neurons could be recorded simultaneously on the same electrode with excitatory neurons (see Fig. 1B) strongly suggests that this cannot be the only factor in the diversity we observed. While state-dependent modulation might still play a significant role in influencing overall responsiveness, we observed many cases where nonresponsive or suppressed neurons could coexist with strongly responding neurons.

Possible Sources of Auditory Input to HVCIN During Wakefulness

The results of the present study suggest that HVCIN, as a population, might receive auditory inputs that exhibit a combination of linear and nonlinear response properties. HVC receives direct auditory input from NIf (Cardin et al. 2005; Coleman and Mooney 2004), CM (Bauer et al. 2008), and possibly Av (Akutagawa and Konishi 2010), with NIf and CM receiving direct input from the primary auditory forebrain (Vates et al. 1996) (see Fig. 1C). Data obtained from multunit recordings suggest that NIf, whose projection neurons drive all three classes of HVC neurons (Hahnloser and Fee 2007), lack BOS selectivity in awake birds (Cardin and Schmidt 2003). A limited amount of single-unit data obtained in CM of awake birds suggests a mix of broadly responsive cells, some with modest selectivity for BOS relative to other complex stimuli and others with no BOS selectivity. NIf and CM are therefore unlikely sources for the extremely BOS-selective excitatory responses shown by approximately one-third of auditory responsive HVCIN (BOS-ONLY cells). Nevertheless, the limited scope of these studies does not preclude that there may exist specific subpopulations of neurons in these structures that provide BOS-selective input during wakefulness, especially considering that auditory activity in nonwaking states is dominated by BOS-selective responses in both areas (Bauer et al. 2008; Cardin and Schmidt 2004a; Janata and Margoliash 1999). Alternatively, some of the nonlinear auditory response properties might arise from intrinsic network dynamics within HVC as has been previously proposed in anesthetized birds (Coleman and Mooney 2004).

While the basic response properties of CM and NIf neurons might not account for the extreme BOS selectivity recorded in some HVCIN during wakefulness, their properties are consistent with providing the input that drives the linear response properties in many HVCIN. Somewhat paradoxically, spike latencies derived from STRF analysis indicate that many of the linear neurons in HVC have input latencies similar to those obtained in primary and secondary auditory forebrain structures of anesthetized birds (Sen et al. 2001). Although anesthesia is known to significantly alter auditory spike latencies and other receptive field properties (Populin 2005; Wang et al. 2005), the shortest latencies observed in the present study (~10 ms) were only a few milliseconds greater than those recorded in nucleus MLD of the auditory midbrain (Woolley et al. 2006). While no evidence exists for direct auditory connectivity between MLD and HVC, HVC does receive a robust input from nucleus uvaformis (Uva), a thalamic nucleus that receives input directly from the ventral lateral lemniscus (LLV) in the auditory hindbrain (Coleman et al. 2007). Neurons in Uva show both BOS-selective and non-BOS-selective auditory responses during anesthesia and may therefore be a source of short-latency auditory inputs to HVC. Unfortunately, nothing is known about awake auditory properties in this structure.

Another possible contributor of short-latency auditory input to HVC is the primary auditory cortical analog field L, which has recently been shown to have some direct functional connectivity with HVC under anesthesia (Shaevitz and Theunissen 2007).

HVCIN Exhibit a Range of Linear and Nonlinear Response Properties and May Subserve Perceptual Discrimination of Song

To evaluate the possible link between auditory response properties in HVC and the song-related perceptual processes in which this structure is involved (Gentner et al. 2000), it is crucial to understand the nature of the song features that best drive neurons in HVC. One class of HVCIN we recorded from (BOS-ONLY cells) demonstrate the remarkable property of showing excitatory response only to BOS and not a single song from the large ensemble of CON stimuli. The low overall linearity of these responses supports previous work in anesthetized birds showing that these responses are exclusively driven by nonlinear inputs (Theunissen and Doupe 1998). A second class of neurons (BOS-CON and CON-ONLY cells) are responsive to CON stimuli. Many of these neurons have CC ratios >0.30, suggesting a degree of linearity approaching what is observed in field L (Gill et al. 2006) and MLD (Gill et al. 2006; Woolley et al. 2006), an area where neurons respond with high precision to the temporal features of sound (Woolley and Casseday 2004, 2005). These results imply that response linearity can propagate many synapses from the auditory periphery to the highest levels of the auditory system. The high degree of temporal precision exhibited by many highly linear HVCIN (see Fig. 6) is consistent with earlier work showing that relative time-varying phase across frequency bands of complex stimuli is preserved in HVC at the millisecond timescale (Theunissen and Doupe 1998). Interestingly, many cells in areas closer to the auditory periphery exhibit response linearity as low as that seen in BOS-ONLY interneurons (Sen et al. 2001; Woolley et al. 2006). Thus it is likely that linearity is established early in some auditory processing streams and is preserved to the highest levels of the system, while in other processing streams linearity is either rapidly lost or simply never established.

We described one prevalent group of nonselective cells as having a BOS bias because they have vigorous excitatory responses to CON stimuli but show responses to BOS that are poorly predicted by STRF. These neurons tended to have linearity scores in response to CON stimuli similar to the mean linearity value observed in CM of anesthetized birds (Gill et al. 2005).
2006) but nevertheless lower than the highly linear nonselective neurons. We propose that these neurons receive a combination of linear (making them respond to all song stimuli) and nonlinear (biasing them toward responding more strongly to BOS) inputs. Figure 10 depicts a hypothetical way in which these neurons might acquire their response properties and serves to summarize how the different response types may emerge in HVCIN.

The observed bias for BOS in HVC of waking birds and the exclusive selectivity for BOS in anesthetized or sleeping birds (Margoliash 1986; Mooney 2000; Rauske et al. 2003; Theunissen and Doupe 1998) suggest that HVC plays a fundamental role in processing BOS in adult birds. Our findings in zebra finches are therefore consistent with previous findings for auditory tuning to BOS in other species (Margoliash and Konishi 1986; Prather et al. 2008, 2009; Nealen and Schmidt 2006; Sakata and Brainard 2008). The additional property of many of the neurons in HVC to respond to CON make them particularly amenable to playing a fundamental role in perceptual discrimination. Previous studies, for example, have shown that lesions to song nuclei such as LMAN, which receive auditory input indirectly from HVC (Doupe and Konishi 1991), significantly affect the discrimination between CON and BOS (Scharff et al. 1998), and HVC-lesioned birds, while not explicitly tested for deficits in BOS vs. CON discrimination, show deficits in contingency reversals during perceptual discrimination tasks involving CON stimuli (Gentner et al. 2000). Because auditory responses during wakefulness in basal ganglia-projecting HVC neurons (Prather et al. 2008, 2009) show a remarkable temporal concordance with motor activity in the same neurons, BOS responses may therefore function as “self” in a motor-based comparison with “other” (CON responses). This notion is similar to what has been proposed for mirror neurons in primate and human premotor cortex (Rizzolatti et al. 2001). It is of interest that auditory responses recorded at the multiunit level in HVC of awake juvenile zebra finches, while responding to BOS and to some extent CON, show a bias for the tutor song (Nick and Konishi 2005). Rather than playing a role in discrimination, HVC in juvenile birds might be intimately involved in using the tutor song as a referent to judge motor performance (here BOS) during vocal learning.

The ability to record from a large number of single units while presenting a wide range of ethologically relevant auditory stimuli has allowed a characterization of the heterogeneity of auditory response properties in HVC of awake birds. Our findings reveal that HVCIN exhibit auditory-tuning properties that contain feature-based linearity and robust selectivity for BOS. A significant future challenge will be to understand how these linear and nonlinear, BOS-ONLY response properties are integrated to subserve perceptual discrimination. Recent evidence in secondary auditory forebrain suggests that these areas can encode the recent exposure history (Gill et al. 2008; Terleph et al. 2008), ethological relevance (George et al. 2008),

Fig. 10. Hypothetical model of how auditory inputs might shape response characteristics of HVCIN during wakefulness. Neurons showing excitatory responses to auditory stimuli fall into 2 broad categories: BOS-ONLY neurons respond selectively to BOS and often show response suppression to CON; BOS-CON neurons show excitatory responses to BOS as well as CON. Within this category, some neurons have highly linear STRFs that predict responses well to both BOS and CON stimuli (“highly linear neurons”). Other neurons have STRFs that predict responses to CON stimuli well but predict BOS responses poorly. Because these neurons respond more vigorously to BOS than they do to CON, these BOS-bias neurons are hypothesized to receive a linear input that drives the CON response and a nonlinear input that acts to boost the response to BOS. This scheme predicts that CON responses in BOS-bias neurons should be weaker than those in highly linear neurons. This trend is present in the data but does not achieve statistical significance. Likely input sources are shown on right. It should be noted that further response shaping is likely to take place within the HVC network itself, where HVCIN interact closely and reciprocally with other interneurons and both types of projection neurons (Mooney and Prather 2005). Plus symbols (+) denote gain estimates based on BOS and CON response strengths for each of the 3 classes of excitatory responders (see Figs. 3, 6C, 7C). Filled black squares represent functionally excitatory inputs. Red circles represent inputs that are functionally inhibitory. Relative size of symbols is scaled to input strength.
and behavioral salience (Gentner and Margoliash 2003) of complex acoustic stimuli. If HVC neurons are sensitive to these or related contextual stimulus features, it will provide strong evidence that the auditory forebrain and the song motor system constitute a unified network that functions in high-order, behaviorally relevant perception.

GRANTS
This work was supported by National Institute on Deafness and Other Communications Disorders Grants DC-006102 and DC-006453.

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

REFERENCES


