Singing-Related Activity of Identified HVC Neurons in the Zebra Finch

Alexay A. Kozhevnikov1,2 and Michale S. Fee1
1McGovern Institute and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts; and 2Department of Physics and Department of Psychology, Pennsylvania State University, University Park, Pennsylvania

Submitted 6 September 2006; accepted in final form 16 December 2006

Kozhevnikov AA, Fee MS. Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol 97: 4271–4283, 2007. First published December 20, 2006; doi:10.1152/jn.00952.2006. High vocal center (HVC) is part of the premotor pathway necessary for song production and is also a primary source of input to the anterior forebrain pathway (AFP), a basal ganglia-related circuit essential for vocal learning. We have examined the activity of identified HVC neurons of zebra finches during singing. Antidromic activation was used to identify three classes of HVC cells: neurons projecting to the premotor nucleus RA, neurons projecting to area X in the AFP, and putative HVC interneurons. HVC interneurons are active throughout the song and display tonic patterns of activity. Projection neurons exhibit highly phasic stereotyped firing patterns. X-projecting (HVC(X)) neurons burst zero to four times per motif, whereas RA-projecting neurons burst extremely sparsely—at most once per motif. The bursts of HVC projection neurons are tightly locked to the song and typically have a jitter of <1 ms. Population activity of interneurons, but not projection neurons, was significantly correlated with syllable patterns. Consistent with the idea that HVC codes for the temporal order in the song rather than for sound, the vocal dynamics and neural dynamics in HVC occur on different and uncorrelated time scales. We test whether HVC(X) neurons are auditory sensitive during singing. We recorded the activity of these neurons in juvenile birds during singing and found that firing patterns of these neurons are not altered by distorted auditory feedback, which is known to disrupt learning or to cause degradation of song already learned.

INTRODUCTION

The avian song control system has emerged as an important model system for studying the generation and learning of vocal and motor sequences (for a review, see Brainard and Doupe 2002). Young birds listen to their tutor’s song and memorize a “song template.” By practicing the song and using auditory feedback to evaluate their own vocal performance, the birds can learn to faithfully reproduce the song of their tutor (Konishi 1965).

The extremely stereotyped vocal behavior of songbirds is implemented by a discrete and anatomically well-characterized neural circuit (for a review, see Brenowitz et al. 1997). The song control system consists of two major pathways: the descending motor pathway and the anterior forebrain pathway (AFP) (Fig. 1A). The descending motor pathway consists of brain nuclei necessary for song production: HVC (used as a proper name), robust nucleus of the arcopallium (RA) and the brain stem motor nucleus innervating the muscles of the syrinx (Nottebohm et al. 1976; Vicario 1991; Vicario and Nottebohm 1988). Nucleus HVC also projects to the anterior forebrain pathway (AFP). The AFP consists of basal ganglia (area X), the medial nucleus of the dorsolateral thalamus (DLM), and lateral magnocellular nucleus of the anterior nidopallium (LMAN), which projects to RA (Bottjer et al. 1989; Luo and Perkel 1999a,b; Okuhata and Saito 1987). Lesions of the AFP leave song production intact in adult birds but cause severe deficits of vocal learning in juvenile birds (Bottjer et al. 1984; Scharff and Nottebohm 1991; Sohrabji et al. 1990). Neurons in HVC and the AFP exhibit auditory responses in anesthetized birds and have especially strong responses to the bird's own song (BOS) (Doupe and Konishi 1991; Margoliash 1986). Evidence available to date suggests that auditory responses of AFP neurons are driven by X-projecting HVC (HVC(X)) neurons (Doupe and Konishi 1991; Vicario and Yohay 1993; Williams and Nottebohm 1985).

To understand the neural mechanisms underlying singing behavior, it is critical to characterize the activity of identified neurons in the song control system when the bird is singing. Neural recordings in singing birds are technically challenging; however, several important steps toward understanding the singing-related neural dynamics in the song control system have been made (Hahnloser et al. 2002; Hessler and Doupe 1999; Leonardo 2004; Leonardo and Fee 2005; McCasland 1987; Yu and Margoliash 1996). Despite the progress made, we do not yet have a complete picture of the operation of this neural circuit.

Because of the crucial importance of nucleus HVC for song production, and the possible role of the HVC projection to the AFP for song learning, HVC is one of the most-studied nuclei in the song control system. Neural types and synaptic connections within HVC have been characterized (Mooney 2000; Mooney and Prather 2005), and both multi- and single-unit recordings of singing-related activity in HVC have been obtained (Hahnloser et al. 2002; McCasland 1987; Yu and Margoliash 1996). However, systematic characterization of the singing-related activity of identified HVC neurons has not been done. In this work, we identified three different classes of HVC neurons in zebra finches by antidromic stimulation and recorded their patterns of activity during singing. We characterized the patterns of activity of each class of neurons and analyzed the relation of the neural activity to the acoustic structure of the song.

Because HVC(X) neurons are implicated in providing auditory input to the AFP, we asked a question whether their singing-related activity has an auditory component. Auditory activity of HVC(X) neurons could constitute a vocal evaluation signal, which is used for vocal learning and error correction. To elucidate an auditory-related component in the activity of HVC(X) neurons, we recorded the firing patterns of HVC(X) neurons in the zebra finch. We test whether HVC(X) neurons are auditory sensitive during singing. We recorded the activity of these neurons in juvenile birds during singing and found that firing patterns of these neurons are not altered by distorted auditory feedback, which is known to disrupt learning or to cause degradation of song already learned.

1988). Nucleus HVC also projects to the anterior forebrain pathway (AFP), a basal ganglia-related circuit essential for vocal learning. We have examined the activity of identified HVC neurons of zebra finches during singing. Antidromic activation was used to identify three classes of HVC cells: neurons projecting to the premotor nucleus RA, neurons projecting to area X in the AFP, and putative HVC interneurons. HVC interneurons are active throughout the song and display tonic patterns of activity. Projection neurons exhibit highly phasic stereotyped firing patterns. X-projecting (HVC(X)) neurons burst zero to four times per motif, whereas RA-projecting neurons burst extremely sparsely—at most once per motif. The bursts of HVC projection neurons are tightly locked to the song and typically have a jitter of <1 ms. Population activity of interneurons, but not projection neurons, was significantly correlated with syllable patterns. Consistent with the idea that HVC codes for the temporal order in the song rather than for sound, the vocal dynamics and neural dynamics in HVC occur on different and uncorrelated time scales. We test whether HVC(X) neurons are auditory sensitive during singing. We recorded the activity of these neurons in juvenile birds during singing and found that firing patterns of these neurons are not altered by distorted auditory feedback, which is known to disrupt learning or to cause degradation of song already learned.

INTRODUCTION

The avian song control system has emerged as an important model system for studying the generation and learning of vocal and motor sequences (for a review, see Brainard and Doupe 2002). Young birds listen to their tutor’s song and memorize a “song template.” By practicing the song and using auditory feedback to evaluate their own vocal performance, the birds can learn to faithfully reproduce the song of their tutor (Konishi 1965).

The extremely stereotyped vocal behavior of songbirds is implemented by a discrete and anatomically well-characterized neural circuit (for a review, see Brenowitz et al. 1997). The song control system consists of two major pathways: the descending motor pathway and the anterior forebrain pathway (AFP) (Fig. 1A). The descending motor pathway consists of brain nuclei necessary for song production: HVC (used as a proper name), robust nucleus of the arcopallium (RA) and the brain stem motor nucleus innervating the muscles of the syrinx (Nottebohm et al. 1976; Vicario 1991; Vicario and Nottebohm 1988). Nucleus HVC also projects to the anterior forebrain pathway (AFP). The AFP consists of basal ganglia (area X), the medial nucleus of the dorsolateral thalamus (DLM), and lateral magnocellular nucleus of the anterior nidopallium (LMAN), which projects to RA (Bottjer et al. 1989; Luo and Perkel 1999a,b; Okuhata and Saito 1987). Lesions of the AFP leave song production intact in adult birds but cause severe deficits of vocal learning in juvenile birds (Bottjer et al. 1984; Scharff and Nottebohm 1991; Sohrabji et al. 1990). Neurons in HVC and the AFP exhibit auditory responses in anesthetized birds and have especially strong responses to the bird’s own song (BOS) (Doupe and Konishi 1991; Margoliash 1986). Evidence available to date suggests that auditory responses of AFP neurons are driven by X-projecting HVC (HVC(X)) neurons (Doupe and Konishi 1991; Vicario and Yohay 1993; Williams and Nottebohm 1985).

To understand the neural mechanisms underlying singing behavior, it is critical to characterize the activity of identified neurons in the song control system when the bird is singing. Neural recordings in singing birds are technically challenging; however, several important steps toward understanding the singing-related neural dynamics in the song control system have been made (Hahnloser et al. 2002; Hessler and Doupe 1999; Leonardo 2004; Leonard and Fee 2005; McCasland 1987; Yu and Margoliash 1996). Despite the progress made, we do not yet have a complete picture of the operation of this neural circuit.

Because of the crucial importance of nucleus HVC for song production, and the possible role of the HVC projection to the AFP for song learning, HVC is one of the most-studied nuclei in the song control system. Neural types and synaptic connections within HVC have been characterized (Mooney 2000; Mooney and Prather 2005), and both multi- and single-unit recordings of singing-related activity in HVC have been obtained (Hahnloser et al. 2002; McCasland 1987; Yu and Margoliash 1996). However, systematic characterization of the singing-related activity of identified HVC neurons has not been done. In this work, we identified three different classes of HVC neurons in zebra finches by antidromic stimulation and recorded their patterns of activity during singing. We characterized the patterns of activity of each class of neurons and analyzed the relation of the neural activity to the acoustic structure of the song.

Because HVC(X) neurons are implicated in providing auditory input to the AFP, we asked a question whether their singing-related activity has an auditory component. Auditory activity of HVC(X) neurons could constitute a vocal evaluation signal, which is used for vocal learning and error correction. To elucidate an auditory-related component in the activity of HVC(X) neurons, we recorded the firing patterns of HVC(X) neurons in the zebra finch.

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Address for reprint requests and other correspondence: M. S. Fee, MIT, 46-5133A, 77 Massachusetts Ave., Cambridge, MA 02139.
Single-unit recordings were made with a three-channel miniature surgery, anesthesia was produced with 1–3% isofluorane in oxygen. Collected from eight adult and five juvenile zebra finches. Before the local Institutional Animal Care and Use Committee. Data were National Institutes of Health and has been reviewed and approved by the animals was carried out in accordance with guidelines of the posthatch) zebra finches. The care and experimental manipulation of Subjects, surgery, and recording technique

METHODS

Subjects, surgery, and recording technique

Subjects were adult (>90 days posthatch) and juvenile (70–90 days posthatch) zebra finches. The care and experimental manipulation of the animals was carried out in accordance with guidelines of the National Institutes of Health and has been reviewed and approved by the local Institutional Animal Care and Use Committee. Data were collected from eight adult and five juvenile zebra finches. Before surgery, anesthesia was produced with 1–3% isofluorane in oxygen. Single-unit recordings were made with a three-channel miniature motorized microdrive using previously described techniques (Fee and Leonardo 2001). Bipolar stimulating electrodes were implanted in RA and Area X for antidromic stimulation of HVC (Fig. 1A). Antidromic identification was done by stimulation RA or Area X just above the threshold for eliciting a reliable evoked spike. Classification of neurons as RA-projecting (HVC(\text{RA})), X-projecting (HVC(\text{X})), or putative interneurons (HVC(\text{ix}), neurons not projecting either to RA or X) was done by measuring the latency variability of the first evoked spike and was verified using collision protocol. Detailed description of the antidromic identification protocol is presented elsewhere (Fee et al. 2004; Hahnloser et al. 2006).

Alignment of acoustic and neural patterns

If one makes a raster plot of the firing pattern of a neuron by aligning the spike trains at the beginning of the song motif, variability in the duration of the motif results in typically 10–20 ms jitter in the spike trains at the end of the motif. To take into account this source of variability, spike trains were aligned using each syllable as a time reference. A single song motif was chosen as a template for time alignment. All other motifs were aligned to the template motif using syllable onsets and offsets as reference points and using piecewise-linear time warping (Glaze and Troyer 2006; Leonardo 2002). Using the same piecewise-linear time-warping, spike times were mapped onto the template motif.

This alignment procedure dramatically reduced the spike time variability due to the variations in the song motif duration. In the rest of the paper, unless noted otherwise, all the data analysis was done using time-warped spike trains aligned to the song template.

Discretized firing patterns, population activity of neurons and characterization of syllable song pattern

Discretized mean firing patterns $n_i(t)$ were computed for each neuron by counting the mean number of spikes occurring during nonoverlapping time windows $\Delta t = 2$ ms long

$$n_i(t) = (\text{mean number of spikes } i^{th} \text{ neuron fires at } t_i - \Delta t/2 < t_i < t_i + \Delta t/2)$$

Population activities of the three classes of neurons [RA$_{pop}(t)$, X$_{pop}(t)$, and I$_{pop}(t)$] were computed by summing the discretized mean firing patterns of all the neurons of a given class within each bird. To characterize the temporal pattern of syllables during the song motif, we used a discrete “syllable-interval” variable

$$S(t) = \begin{cases} 1 & \text{during the song syllable} \\ 0 & \text{during the interval between syllables} \end{cases}$$

Syllable onsets and offsets were determined as threshold crossing times of the logarithm of acoustic power.

Calculation of coherence

To characterize the relation between population activity of interneurons and the syllable structure of the song, we computed the coherence between the syllable pattern $S(t)$ and the population average of interneurons $I_{pop}(t)$. The advantage of working with the coherence versus time-domain correlation function (i.e., working in frequency domain vs. time domain) is the simplicity of the estimating confidence levels for the null hypothesis (Jarvis and Mitra 2001).

Direct spectral estimators of $S(t)$ and $I_{pop}(t)$ were computed using the multi-taper method (Thompson 1982) with time-bandwidth product $NW = 7/2$. Functions $S(t)$ and $I_{pop}(t)$ were convolved with taper windows and zero-padded to $n_m = 512$ points, and then fast Fourier transforms $S(\omega)$ and $I_{pop}(\omega)$ were computed. Analysis was done for five birds in which five or more interneurons per bird were recorded. Data from different birds were regarded as different samples. The coherence $\gamma(\omega)$ was computed by summing cross-spectral products
over the samples and over the tapers and normalizing (Thompson 1982). The coherence \( C(\omega) \) is defined as \( C(\omega) = |\gamma(\omega)|^2 \). The \((p \times 100)\)th percentile confidence intervals for the null hypothesis [i.e., that the mean correlation functions between the activity of a neuron and the syllable temporal pattern, we computed the mean correlation functions where \( N_{\text{free}} = 2N_{\text{shp}}N_{\text{sample}} = 4NW\ N_{\text{birds}} \) is the number of degrees of freedom (Jarvis and Mitra 2001).

Correlation coefficients and time-domain correlation functions

Pairwise correlations between the firing patterns of different neurons were characterized by computing Pearson correlation coefficients between the discretized mean firing patterns of neurons (defined in the preceding text). When computing correlation coefficients between the activity of an individual interneuron \( n(t) \) and the population activity of interneurons, we did not include the interneuron itself in the estimator of the population activity to avoid artificial increase of the correlation coefficient.

For computing time-domain correlation functions, all data (discretized mean firing patterns, population activities and syllable temporal patterns) were wrapped around. This ensures large overlap even for large time lags and is justifiable because song motifs are periodically repeated during the song bout.

We also analyzed the correlations between the population activities of HVC\(_{\text{RA}}\) and HVC\(_{\text{X}}\) neurons. We used two different methods for combining data from different birds: we computed correlation coefficients between the populations of HVC\(_{\text{X}}\) and HVC\(_{\text{RA}}\) neurons and averaged them over birds, and we computed correlation coefficients between the populations of HVC\(_{\text{X}}\) and HVC\(_{\text{RA}}\) neurons and averaged over birds weighting the correlation coefficients by the number of HVC\(_{\text{RA}}\) neurons recorded in each bird—this way the birds in which more neurons were recorded received more weight. To assess the significance of the computed correlations, we randomly time-shifted the activity of individual neurons to construct shuffled population activities of HVC\(_{\text{X}}\) and HVC\(_{\text{RA}}\) neurons. The shuffling procedure was repeated \( n = 10,000 \) times, and the \( P = 0.05 \) confidence intervals for the correlations of the shuffled data were computed.

To study whether the activity of the projection neurons is modulated coherently with the syllable temporal pattern, we computed the mean correlation functions between the activity of a neuron and the syllable patterns of the song: \( C_{\text{RA-S}}(r) = \langle \text{corr}[\text{RA}(r), S_i(t - \tau)] \rangle \) and \( C_{\text{X-S}}(r) = \langle \text{corr}[X_i(t), S_i(t - \tau)] \rangle \) (averaging is over all neurons in all birds).

To assess the significance of this correlation, we generated randomized data sets by introducing random shifts \( t_{\text{rand}} \) to the mean firing patterns of projection neurons and computing the mean correlation functions between the syllable pattern \( S(t) \) and the shifted neural firing patterns: \( C_{\text{RA,S}}^{\text{rand}}(r, \tau) = \langle \text{corr}[\text{RA}(r, t + t_{\text{rand}}), S_i(t - \tau)] \rangle \) and \( C_{\text{X-S}}^{\text{rand}}(r, \tau) = \langle \text{corr}[X_i(t + t_{\text{rand}}), S_i(t - \tau)] \rangle \). Because the relative timing between the neural activity and the syllable pattern is a priori unknown, we used a fairly conservative criterion for statistical significance: the mean correlation functions \( C_{\text{RA-S}}(\tau) \) and \( C_{\text{X-S}}(\tau) \) were considered significant if they exceeded the \( P = 0.01 \) confidence intervals for the distribution of maxima or if they were below the \( P = 0.01 \) confidence interval for the distribution of minima of the functions \( C_{\text{RA-S}}^{\text{rand}}(r, \tau) \) and \( C_{\text{X-S}}^{\text{rand}}(r, \tau) \) in the interval \(-200 \text{ ms} < t < 200 \text{ ms}\).

Correlations between multiple bursts of X-projecting neurons and sound: burst-triggered similarity

We analyzed whether multiple bursts of HVC\(_{\text{X}}\) neurons are related to the occurrence of similar sounds in the song. To analyze whether repeating bursts of individual HVC\(_{\text{X}}\) neurons are preceded or followed by similar sounds, we computed the burst-triggered similarity function averaged over all pairs of bursts of individual HVC\(_{\text{X}}\) neurons

\[
C(\tau) = \frac{1}{N_{\text{pairs}}} \sum_{i=1}^{N_{\text{pairs}}} C(t_i^1 + \tau t_i^2 + \tau)
\]

where \( t_i^1 \) and \( t_i^2 \) are times at which bursts of an HVC\(_{\text{X}}\) neuron occurred, and \( C(t_i^1, t_i^2) \) is an acoustic similarity matrix that characterizes the similarity of the acoustic structure of the song at times \( t_i^1 \) and \( t_i^2 \) (see following text).

We used three different methods for constructing the similarity matrix between sounds in the song: syllable-syllable similarity, which only takes into account syllable pattern of the song and ignores fine-scale acoustic correlation; similarity of fine acoustic structure correlation using sound spectrograms; and similarity using acoustic features of the song (Tchernichovski et al. 2000).

SYLLABLE-SYLLABLE SIMILARITY. To take into account similarity associated with syllable structure of the song only, we used the following rule for constructing the syllable-syllable acoustic similarity matrix

\[
C_{\text{syll}}(t_i^1, t_i^2) = \begin{cases} 1 & \text{if } S(t_i^1) = S(t_i^2) \\ -1 & \text{if } S(t_i^1) \neq S(t_i^2) \end{cases}
\]

The burst-triggered syllable-syllable similarity function averaged over multiple bursts of HVC\(_{\text{X}}\) neurons is defined as

\[
C_{\text{syll}}(\tau) = \frac{1}{N_{\text{pairs}}} \sum_{i=1}^{N_{\text{pairs}}} C_{\text{syll}}(t_i^1 + \tau t_i^2 + \tau)
\]

SPECTRAL SIMILARITY. Song spectrograms were computed over time window \( \Delta t = 8 \text{ ms} \) using the multi-taper method with \( NW = 3/2 \).

If \( P(t,f) \) is time- and frequency-dependent acoustic power, the logarithmic spectrogram (called for shortness “spectrogram” in the following text) is \( B(t,f) = \log P(t,f) \). One can define the matrix of acoustic similarities by using the spectra at times \( t_1 \) and \( t_2 \): \( C(t_1, t_2) = \text{corr}[B(t_1,f), B(t_2,f)] \). The acoustic similarity matrix \( C(t_1, t_2) \) defined this way has a lot of structure due to the syllable pattern of the song. The syllable-pattern structure in \( C(t_1, t_2) \) makes it difficult to analyze the similarities in fine acoustic structure (i.e., similar notes). To circumvent this problem, we calculated the mean temporal and spectral modes of the spectrogram \( B(t,f) \), normalized the spectrum at each time by the total acoustic energy and then subtracted the projection of the spectrum onto the mean spectral mode (Leonardo and Fee 2005)

\[
B_{\text{norm}}(t,f) = B(t,f) - \bar{B}(t) - \int_0^{t_{\text{max}}} \left( B(t',f) - \bar{B}(t') \right) B(t',f) \text{d}t'
\]

where \( \bar{B}(t) = \log P'(P(t,f)) \) is the mean temporal mode, \( \bar{B}(f) = \log \langle P'(t,f) \rangle \) is the mean spectral mode of the song and \( f_{\text{max}} = 8 \text{ kHz} \). Using the normalized spectrogram \( B_{\text{norm}}(t,f) \), we calculated the spectral similarity matrix [this quantity was named correlation matrix in (Leonardo and Fee 2005)]:

\[
C_{\text{spec}}(t_1, t_2) = \text{corr}[B_{\text{norm}}(t_1,f), B_{\text{norm}}(t_2,f)]
\]

Matrix \( C_{\text{spec}}(t_1, t_2) \) has most of its structure due to similar notes (e.g.,
The burst-triggered spectral similarity function averaged over multiple bursts of HVC(X) neurons is defined as

$$C_{\text{spect}}(\tau) = \frac{1}{N_{\text{pairs}}} \sum_{i=1}^{N_{\text{pairs}}} C(t_i^1 + \tau t_i^2 + \tau)$$

**FEATURE-BASED SIMILARITY.** We also computed acoustic features of the songs and constructed a feature-based similarity matrix

$$C_{\text{features}}(t_i^1, t_i^2) = \tilde{F}(t_i^1) \cdot \tilde{F}(t_i^2) \sqrt{[\tilde{F}(t_i^1)][\tilde{F}(t_i^2)]}$$

where the components of vector $\tilde{F}(t)$ are time-dependent scaled acoustic features (Tchernichovski et al. 2000). Because some acoustic features are defined only during song syllables, the feature-based similarity matrix is also correlated with the syllable-saylable similarity matrix. To suppress the syllable pattern-related structure of matrix $C_{\text{features}}(t_i^1, t_i^2)$, we modified the feature similarity matrix as follows: when either $S(t_i^1) = 0$ or $S(t_i^2) = 0$, we set the similarity matrix value to the mean over all times when $S(t_i^1) = S(t_i^2) = 1$

$$C_{ij} = \begin{cases} C_{\text{features}}(t_i^1, t_i^2) & \text{if } S(t_i^1) = S(t_i^2) = 1 \\ \langle C_{\text{features}}(t_i^1, t_i^2) \rangle & \text{otherwise} \end{cases}$$

With this definition of the acoustic similarity matrix $C_{ij}(t_i^1, t_i^2)$, the similarities due to the fine acoustic structure are preserved, but the syllable pattern-related structure in the matrix $C_{ij}(t_i^1, t_i^2)$ is suppressed. Similarly to syllable-saylable similarity and spectral similarity, the burst-triggered feature-based similarity function was obtained by averaging over pairs of multiple bursts of HVC(X) neurons

$$C_{ij}(\tau) = \frac{1}{N_{\text{pairs}}} \sum_{i=1}^{N_{\text{pairs}}} C_{ij}(t_i^1 + \tau t_i^2 + \tau)$$

**ASSESSMENT OF SIGNIFICANCE.** To assess the significance of the burst-triggered similarities, we generated pairs of bursts distributed uniformly during the song motif and computed the burst-triggered similarity functions using simulated bursts $C_{\text{rand}}(\tau)$.

For the burst-triggered syllable-saylable similarity, we repeated the procedure $n = 3,000$ times and computed the distribution of maxima and minima of $C_{\text{rand}}(\tau)$ in the interval $-200$ ms $< \tau < 200$ ms. Values of observed burst-triggered similarity functions were considered significant when they either exceeded the $P = 0.01$ confidence interval for the distribution of maxima of $C_{\text{rand}}(\tau)$ or were below the $P = 0.01$ confidence interval for the distribution of minima of $C_{\text{rand}}(\tau)$.

For the burst-triggered spectral similarity and the feature-based similarity, we computed the distribution of $C_{\text{rand}}(\tau = 0)$. The values of observed burst-triggered spectral and feature-based similarity functions were considered significant when they were outside the $P = 0.01$ confidence intervals for the distribution of $C_{\text{rand}}(\tau = 0)$. This significance criterion is less stringent than the one used for burst-triggered syllable-saylable similarity, which is justified because the observed burst-triggered spectral and feature similarities are not significant as assessed by confidence intervals for the distribution of $C_{\text{rand}}(\tau = 0)$.

**Correlation times of vocal dynamics and neural dynamics in HVC**

To compute acoustic correlation times we used the spectral similarity matrix $C_{\text{spect}}(t_i^1, t_i^2)$ defined in the preceding text. For computing the neural similarity matrix, we constructed a neural activity vector $\hat{N}(t)$ (the $i$th component of this vector is the mean firing pattern of the $i$th interneuron $n_i(t)$) and computed Pearson correlation coefficient between the components of vector $\hat{N}(t)$ at different times: $C_{\text{corr}}(t_i^1, t_i^2) = \text{corr}(\hat{N}(t_i^1), \hat{N}(t_i^2))$.

The procedure for computing the correlation time for matrices $C_{\text{spect}}(t_i^1, t_i^2)$ and $C_{\text{corr}}(t_i^1, t_i^2)$ is as previously described (Leonardo and Fee 2005). For each point $(t_{\text{corr}}, t_i^1)$ on the matrix diagonal, we determined the maximal size of the square around this point for which all matrix values inside the square are larger than the threshold value $C = 0.5$. The size of that square is the characteristic time $\tau_{\text{peak}}$. We then searched for the peaks of the function $\tau(t_{\text{corr}})$, starting from the largest peak, and for each peak occurring at time $t_{\text{peak}}$ we extended the peak value of the characteristic time $\tau_{\text{peak}}$ over the time interval $\tau_{\text{peak}}$ long so that the correlation time $\tau_{\text{corr}}(t_{\text{corr}}) = \tau_{\text{peak}}$. Such procedure results in the correlation time being constant over the length of the interval where the values of the matrix are high (for example, sound correlation time $\tau_{\text{sound}}(t)$ is constant during a harmonic stack and equal to the duration of the harmonic stack).

**Distorted auditory feedback**

For the generation of DAF during neural recordings, loud bursts of broadband noise were played back to the bird during 50% of song renditions (sound volume: 80–90 dB SPL, the burst duration typically 70–100 ms). Using the Golay pulse calibration (Golay 1961), noise cancellation was performed and song signals were recovered (Leonardo 2002).

**Assessments of effects of DAF**

Characterization of systematic changes in song structure as a result of DAF exposure was done by computing a spectral similarity score between each rendition of a syllable and a template syllable produced 3–4 days before the onset of DAF. The similarity score between two syllables was defined as a Pearson correlation coefficient between the multi-taper spectrograms (Thompson 1982) of the syllables (time-bandwidth product: $NW = 3/2$). Spectra were averaged over three tapers and FFTs were computed over time windows 8-ms-long overlapping by 7 ms. Spectral components between 0.3 and 8 kHz were used to compute the similarity scores. This procedure for computing similarity scores typically yielded values of 0.8–0.9 for different renditions of the same syllable and values of 0.4–0.7 for different syllables (depending on how similar the different syllables are).

To assess whether DAF caused significant degradation of the acoustic structure of the song syllable, we computed similarity scores of each syllable to the template syllable and separated the scores in two groups: similarity scores on catch trials on the days when DAF was presented, starting on the second day of DAF presentation and through the first day after DAF presentation stopped, and similarity scores on the days when no DAF was present, both before and after the period of DAF presentation. The distributions of scores in these two groups were compared with each other, and the degradation of acoustic structure was considered significant if the similarity scores during the period of DAF presentation were significantly ($P < 0.05$, 1-sided KS test) lower than when no DAF was presented.

**Comparison of firing patterns of HVC(X) neurons with and without DAF**

To compare the firing patterns of HVC(X) neurons with and without DAF, for each burst of each HVC(X) neuron recorded, we computed several burst parameters: burst onset times during the song motif $t_{\text{burst}}$, burst duration $t_{\text{dura}}$, the number of spikes in a burst $n_{\text{spikes}}$, the mean firing rate during the burst and the interspike intervals in a burst (for the calculation of interspike intervals, we used nontime-warped spike times). We compared the distributions of each of these parameters during normal singing and during singing with DAF. We used the two-sided KS test and the one-way ANOVA to assess statistical differences in the distributions of the burst parameters.
RESULTS

Activity patterns of identified HVC neurons during singing

We antidromically identified and obtained single-unit recordings from 28 HVC(RA) neurons (8 birds), 103 HVC(X) neurons (7 birds), and 52 putative HVC interneurons (8 birds) during singing. The three types of neurons have very distinct patterns of activity during singing (Fig. 2).

HVC(RA) neurons fired at most one stereotypical burst during the song motif. Of 28 cells, 21 cells produced a single burst during the song motif, 4 cells produced bursts during the call vocalizations but not during song motifs, and 2 cells generated bursts only during introductory notes but not during song motifs or calls. In addition, one cell produced sparse irregular single spikes during the song motif (<0.25 spikes per song motif). Analysis of the bursting properties of the majority of HVC(RA) neurons was presented previously (Hahnloser et al. 2002). HVC(RA) neurons were virtually silent when the birds were awake but not singing (spontaneous firing rate: <0.001 spike/s).

An additional eight units were identified by antidromic stimulation as HVC(RA) neurons, but these were not spontaneously active nor were they active during singing or during any calls that could be elicited. Because spikes were only observed from these neurons during antidromic stimulation, there is less certainty about the classification of these units.

The burst duration of HVC(RA) neurons was 5.1 ± 1.8 ms (minimum = 3.0 ms, maximum = 9.1 ms), and the bursts contained 4.3 ± 1.3 spike (minimum = 2.0, maximum = 6.8;
ranges are means ± 1 SD unless specified otherwise). The average firing rate during the bursts was 650 ± 150 Hz (minimum = 330 Hz, maximum = 890 Hz). The bursts were extremely reliable and were tightly locked to the vocalization—the root mean square (RMS) jitter of the first spike is 0.73 ± 0.3 ms (minimum = 0.4 ms, maximum = 2.5 ms). The ratio of the SD of the number of spikes in a burst to the mean number of spikes in a burst was 0.10 ± 0.05.

HVC(X) neurons also fired sparse bursts during singing. Most HVC(X) neurons (72 cells) exhibited stereotyped high-frequency bursts during singing that were tightly locked to the vocalization. The bursts of these HVC(X) neurons were similar to the bursts of HVC(RA) neurons, but HVC(X) neurons fired up to 4 bursts per motif (1.3 ± 0.9 burst/motif). The burst duration was 6.3 ± 3.3 ms (minimum = 2.7 ms, maximum = 14 ms), the number of spikes in a burst was 3.7 ± 1.4 (minimum = 1.75, maximum = 7.0), and the average firing rate during the burst was 434 ± 130 Hz (minimum = 110 Hz, maximum = 750 Hz). The rms temporal jitter of the first spike in a burst was 1.0 ± 0.5 ms (minimum = 0.3 ms, maximum = 2.3 ms; 114 bursts in 72 neurons). The ratio of the SD of the number of spikes in a burst to the mean number of spikes in a burst is 0.16 ± 0.12. In awake nonsinging birds, HVC(X) neurons fired single spikes at an average rate of 1.5 ± 0.5 Hz.

About 10% of HVC(X) cells (n = 13) exhibited bursts with lower firing rate and higher variability in a number of parameters. These cells burst up to three times per motif and their bursts consisted of 3.2 ± 1.6 spikes. However, the bursts of these neurons were longer (burst duration: 10 ± 5 ms) and had a lower firing rate (150 ± 72 Hz), and the number of spikes in a burst was more variable: the ratio of the SD of the number of spikes to the mean number of spikes is 0.64 ± 0.32. The bursts of these HVC(X) cells were less tightly locked to the vocalization; the rms jitter of the first spike in the burst was 8.2 ± 7.8 ms. A scatter plot of the rms jitter of the first spike in a burst versus average firing rate during the burst for HVC(X) neurons is shown in Fig. 2B. Bursts with higher temporal jitter tended to have lower firing rates. Although the two groups of HVC(X) neurons had significantly different firing patterns, the sample size is too small to determine if this less stereotyped group of neurons form a distinct cluster in the space of burst characteristics.

In addition, 18 HVC(X) neurons were either completely silent during singing or produced sparse, irregular single spikes (<1 spike/s). The distribution of the number of bursts generated by HVC(X) neurons per song motif is presented in Fig. 2C.

Similarly to HVC neurons previously described (Yu and Margoliash 1996) HVC interneurons (HVCi) were active throughout the vocalization and had tonic patterns of activity. Their firing patterns were much less stereotyped than the firing patterns of projection neurons, but there is a stable underlying time-dependent firing rate pattern (Fig. 2A). Interneurons generated 69 ± 31 spike/song motif (minimum = 20, maximum = 208). During singing they had average firing rates of 95 ± 40 Hz (minimum = 25 Hz, maximum = 262 Hz). The SD of the number of spikes of an interneuron per song motif is 6 ± 3 spike. In an awake nonsinging bird, HVC interneurons had an average spontaneous firing rate of 8 ± 6 Hz.

Population activity of different classes of HVC neurons

We asked whether the activity of different classes of HVC neurons is synchronized during singing. Because of the stereotypical patterns of activity, the firing patterns of different neurons are strongly correlated but this does not necessarily imply synchrony. Multimunit data suggest that there is a synchronized pattern of population activity in HVC (Vu et al. 1998). However, only single-unit recordings from identified neurons can tell us whether neurons of different classes tend to fire in synchrony.

The average correlation coefficient between the activity of an interneuron and the population activity of other interneurons was 0.14 ± 0.11 (see METHODS), larger than the pairwise correlations between individual interneurons (1-way KS test, P < 2 × 10⁻⁵). These data show that the firing patterns of individual interneurons are weakly synchronized.

To address the question of whether neurons of different classes are synchronized with each other, we analyzed pairwise zero time-lag correlations between the firing patterns of HVC neurons. The mean and the SD of the pairwise correlation coefficients for different types of neurons are presented in Table 1. To test whether the synchrony in the firing patterns of individual neurons is statistically significant, we introduced random time shifts to the spike trains and calculated pairwise correlation coefficients for the shifted spike trains. We used KS test to test the hypothesis that the distributions of correlation coefficients for nonshifted and shifted spike trains are different. Statistically significant correlations were found between pairs of interneurons (KS test, P < 10⁻⁸) and between interneurons and HVC(X) neurons (P < 5 × 10⁻⁸). Thus activity of HVC(X) neurons is synchronized with the coherent modulation of interneuron activity in HVC.

Our analysis of the pairwise correlation coefficients does not reveal significant synchrony between HVC(RA) neurons and interneurons (P = 0.051) or between HVC(RA) and HVC(X) neurons (P = 0.65). Likewise, analysis of correlations between the population averaged activities of the projection neurons did not reveal significant synchrony. This could be due to sparse firing patterns of projection neurons—even in the presence of synchrony at a population level, synchrony between individual HVC(RA) and other neurons would be difficult to detect unless very strong. It is possible, however, that significant synchrony between HVC(RA) and HVC(X) neurons would be revealed with a larger sample of neurons.

| TABLE 1. Pairwise correlations between firing patterns of different HVC neuron classes |
|--------------------------|--------------------------|--------------------------|
|                          | RA Projecting            | X Projecting             | Interneurons               |
| RA projecting            | −0.012 ± 0.006 (34 pairs)| 0.02 ± 0.15 (263 pairs)  | 0.015 ± 0.09 (156 pairs)  |
| X projecting             | 0.004 ± 0.11 (737 pairs)  | 0.009 ± 0.08 (852 pairs) |                |
| Interneurons             | 0.05 ± 0.10 (242 pairs)  |                          |                |

Statistically significant (P < 0.05, one-way Kolmogorov-Smirnov test) correlations are in bold.
Relation of activity of HVC neurons to syllable patterning

We analyzed the relation between the population activity of the three classes of HVC neurons and the syllable pattern of the song. At the simplest level, zebra finch song consists of song syllables separated by intervals of silence. We asked whether firing patterns of HVC neurons are related to the pattern of intervals of sound and silence in the song motif. For this question, we disregard the spectral content of the song and characterize the song with syllables represented by 1 and silences by 0. We refer to this pattern of sounds and silences as the syllable pattern of the song. The syllable pattern \( S(t) \), and the population averages of the three classes of neurons \( [\text{RA}_p(t), \text{X}_{\text{pop}}(t), \text{I}_{\text{pop}}(t)] \), see METHODS] for one of the birds (bird 2) are shown in Fig. 3A.

To characterize the correlations between the population activity of interneurons and the syllable pattern of the song, we computed the coherence between \( S(t) \) and the population average of interneurons \( \text{I}_{\text{pop}}(t) \) (Fig. 3B). The coherence has a pronounced peak in the frequency interval between 8 and 15 Hz, reaching a value of \(-0.25 (P < 0.01)\). The frequency band between 8 and 15 Hz corresponds to the time scale of the song syllables \((\sim 70–130 \text{ ms})\), showing that the population activity of interneurons was modulated coherently with the syllable pattern of the song.

The correlation function between individual interneurons and \( S(t) \), averaged over all recorded interneurons from all birds, is shown in Fig. 3C. On average, the activity of an individual interneuron was weakly correlated with the syllable pattern of the song (correlation coefficient \(-0.08)\) and leads the song syllable pattern by 30–50 ms. The peak of the correlation between \( I_{\text{pop}}(t) \) and \( S(t) \) (at a lag of 45 ms) was 0.2, averaged over birds. Thus the population activity of interneurons was modulated by \(-20\%\) and was largest preceding song syllables by \(-45 \text{ ms}\).

To assess the syllable-related modulation of projection neurons, we computed the mean correlation between the firing pattern of a neuron of a given class and the syllable pattern and assessed its significance (see METHODS). The mean correlation functions of \( \text{HVC}_{(RA)} \) and \( \text{HVC}_{(X)} \) neurons with the syllable pattern of the song are presented in Fig. 3C. Although the mean correlation function for \( \text{HVC}_{(RA)} \) neurons exhibits a peak at \(-40 \text{ ms}\), the value of the peak does not reach statistical significance. It is possible that this peak is real, but a larger dataset would be needed to establish significance.

In summary, the population activity of HVC interneurons exhibited syllable-related modulation, and lead the syllable pattern of the song by \(-45 \text{ ms}\). The population activity of \( \text{HVC}_{(RA)} \) and \( \text{HVC}_{(X)} \) neurons was not significantly modulated with the song syllable pattern in our dataset.

Relation of activity of \( \text{HVC}_{(X)} \) neurons to acoustic structure of the song

Given the importance of the AFP for song learning and the proposed role of the \( \text{HVC}_{(X)} \) neurons in the vocal error correction, we asked whether the activity of \( \text{HVC}_{(X)} \) neurons is related to the acoustic structure of the vocalizations. Although we saw no effects of the song syllable pattern on the population activity of \( \text{HVC}_{(X)} \) neurons, the occurrence of multiple bursts from \( \text{HVC}_{(X)} \) neurons allows us to test the possibility that repeated bursts of a given \( \text{HVC}_{(X)} \) neuron occur in some definite relation to repeated sounds in the song. For example, if the activity of \( \text{HVC}_{(X)} \) neurons were auditory-related, one might expect that multiple bursts of a given neuron would be preceded by similar sounds. Alternatively, if activity of \( \text{HVC}_{(X)} \) neurons has a premotor relation to a particular sound in the song, one might expect multiple bursts to precede similar sounds during the song motif (Fig. 4A).

To test this hypothesis, we computed the burst-triggered similarity function averaged over all pairs of bursts of individual \( \text{HVC}_{(X)} \) neurons.
Time scales of vocal dynamics and neural activity in HVC

Vocal output during the bird’s song can change on varying time scales—from <10 ms during rapid acoustic transitions to ~100 ms during harmonic stacks. It has been shown that neural dynamics in RA evolve on fixed time scale (~8 ms) that is not related to the time scale of modulation in the vocalization (Leonardo and Fee 2005). Consistent with this observation, and with experiments suggesting that RA bursts are driven from HVC (Hahnloser et al. 2002), we expect that the time scale of HVC dynamics at all times during the song motif.

In contrast, HVC interneurons are active continuously, and provide a means of estimating the time scale of neural activity in HVC. Correlation matrices of sound and interneuron firing patterns (see METHODS) for one of the birds are presented in Fig. 5, A and B. Figure 5C shows time dependence of the correlation time of the song $\Delta t_{\text{acoustic}}(t)$ and the correlation time of activity of interneurons $\Delta t_{\text{neural}}(t)$. The values of the correlation time $\Delta t_{\text{neural}}(t)$ are consistent with the width (FWHM = 3.6 ms) of the central peak of the average autocorrelation function of the individual interneurons (Fig. 5C, inset). The

In summary, multiple bursts of a given HVC(X) neuron are correlated with the syllable structure of the song; Bursts of a given HVC(X) neuron tend to precede similar syllable pattern elements (either sounds or silences) by ~40 ms. We find no significant correlations of repeated bursts of individual HVC(X) neurons with repeated spectral structure in the song.

$$C(\tau) = \frac{1}{N_{\text{pairs}}} \sum_{i=1}^{N_{\text{pairs}}} C(t_i^1, t_i^2 + \tau)$$

where $t_i^1$ and $t_i^2$ are times at which two bursts of an HVC(X) neuron occur, and $C(t_1, t_2)$ is the acoustic similarity matrix that characterizes how similar sounds at times $t_1$ and $t_2$ are. We used three different choices for computing the similarity matrix: a syllable-syllable similarity measure that only distinguishes between syllables and intersyllable intervals, a spectral similarity measure, and a similarity measure based on acoustic features of the song (see METHODS for details of the definitions of different similarity measures). In Fig. 4B, we present the burst-triggered syllable-syllable similarity function $C_{\text{syl}}(\tau)$. The function $C_{\text{syl}}(\tau)$ has a peak at $t \approx 40$ ms significant at $P < 0.01$. This indicates that multiple bursts of an HVC(X) neuron tend to precede similar elements of the song syllable pattern (either sounds or silences). Thus their temporal relation to the syllable pattern is premotor-like.

We next turned to the question of whether multiple HVC(X) bursts are related to similar spectral structure in the song. We computed the burst-triggered spectral similarity function averaged over pairs of bursts of HVC(X) neurons, $C_{\text{spec}}(\tau)$, and assessed its significance (Fig. 4C, see METHODS). The spectral similarity did not have any pronounced structure as a function of $\tau$, i.e., multiple bursts of an individual HVC(X) neuron were not preceded or followed by similar spectral structure in the song. Analysis of feature-based similarity likewise shows that multiple bursts of HVC(X) neurons are neither preceded nor followed by similar acoustic features in the song.

FIG. 4. Multiple bursts of individual X-projecting neurons and their relation to acoustic structure of the song. A: song spectrogram and firing patterns of X-projecting neurons bursting more than once per song motif (bird 4). B: burst-triggered syllable-syllable similarity averaged over all the pairs of multiple bursts of individual X-projecting neurons $C_{\text{syl}}(\tau)$ (see METHODS). Dotted line, $P = 0.01$ confidence intervals (see METHODS). C: burst-triggered spectral similarity averaged over all the pairs of multiple bursts of individual X-projecting neurons $C_{\text{spec}}(\tau)$. Dotted line, $P = 0.01$ confidence intervals (see METHODS).
linear regression of neural correlation time shifted forward by $\tau = 20 \text{ ms } \Delta t_{\text{neural}}(t + \tau)$ versus acoustic correlation time $\Delta t_{\text{acoustic}}(t)$ (Fig. 5D, straight line) had essentially a zero slope, showing that premotor activity in HVC changed with a characteristic time uncorrelated with the characteristic timescales of the vocal output. We varied delay time $\tau$ between $-70 \text{ ms}$ (which may correspond to auditory delay) to $+70 \text{ ms}$ (which exceeds delay time between premotor activity in HVC and vocal output). At all values of $\tau$ between $-70$ and $+70 \text{ ms}$, there was no significant positive correlation between $\Delta t_{\text{neural}}(t + \tau)$ and $\Delta t_{\text{acoustic}}(t)$; at all values $-70 \text{ ms} < \tau < 70 \text{ ms}$ the slope of the linear fit of $\Delta t_{\text{neural}}(t + \tau)$ versus $\Delta t_{\text{acoustic}}(t)$ was between $-0.015$ and $0.015$.

In summary, neural dynamics in HVC occurred on a fast fixed time scale, whereas vocal output changed over a wide range of time scales ($<10$ to $>100 \text{ ms}$). We found no correlation between characteristic times of neural dynamics in HVC and characteristic timescales of vocal output.

**Effects of distorted auditory feedback on the firing patterns of HVC(X) neurons**

The most direct test of auditory sensitivity of HVC(X) neurons during singing is to alter what the bird hears while it sings and look for changes in the firing patterns of these neurons. We recorded the firing patterns of HVC(X) neurons in juvenile birds during normal singing and during singing in the presence of DAF (Leonardo 2002; Leonardo and Konishi 1999). By comparing the firing patterns of HVC(X) neurons during singing with and without DAF, we directly tested the hypothesis that these neurons carry auditory signal during singing.

Singing-related single-unit activity of a total of 30 HVC(X) neurons (3 juvenile birds) was recorded in the presence and in the absence of distorted auditory feedback. Data were taken when the birds were 70–90 days posthatch. The firing patterns of HVC(X) neurons were quite similar with and without distorted feedback, and there were no apparent changes in the activity of these cells (Fig. 6C). The highly stereotypical and temporally precise firing patterns of HVC(X) neurons should make detectable even slight activity changes related to DAF, such as $\sim 1$ extra spike in a burst, or change of the burst onset time by $\sim 1 \text{ ms}$, or changes in the interspike interval during the burst by $\sim 300 \text{ us}$. For each burst of each HVC(X) neuron studied, we computed several burst parameters: burst onset times during the song motif, burst duration, the number of spikes in a burst, and the mean firing rate during the burst and interspike intervals in a burst. A summary of the analysis of the burst onset times, burst durations and number of spikes in the bursts of HVC(X) neurons are presented in Fig. 7. Of the 43 bursts analyzed, only 2 bursts showed a statistically significant difference in burst onset time or burst duration ($P < 0.05$, 2-sided KS test). This would be expected by chance—if there is no change in the firing patterns, analysis of $\sim 5\%$ of the bursts would yield statistical significance at $P < 0.05$ level. None of the 43 bursts analyzed showed statistical differences at $P < 0.01$ level. We conclude that the singing related firing patterns of HVC(X) neurons were not altered by on-line distorted auditory feedback.

**Effects of distorted auditory feedback on the song of juvenile birds**

In zebra finches, playback of loud sounds to the bird during singing does not immediately change the song but leads to gradual song degradation on a time scale from weeks to months (Leonardo and Konishi 1999). Such degradation has been observed for playback of song syllables or bursts of broadband
noise, both referred to as DAF. This suggests that the auditory information acquired during singing leads to plastic changes in the motor program.

We tested the effects of broadband noise DAF on the song of juvenile birds (n = 5 birds). The DAF was triggered by the birds’ vocalizations with high probability (P = 0.98) for several days, and changes in the songs were analyzed on catch trials.

After several days of distorted feedback exposure, the vocalizations of all five birds changed markedly (Fig. 6, A and B; supplementary Fig. 1). The changes observed included syllable stuttering (1 bird), abnormal repetition of an introductory note (2 birds), and changes in the acoustic structure of the song syllables (3 birds); one bird exhibited both introductory note stuttering and syllable structure degradation. Every bird studied exhibited some disruptive effect of DAF. In the three birds that showed degradation of the syllable structure, we found significant (P < 0.05) degradation of 4 syllables of total of 12 syllables in these three birds.

The observed changes in the song were reversible—both the syllable sequence and the syllable structure recovered to near prefeedback levels after the distorted feedback had been removed. Observation of the chronic and reversible effects of the distorted feedback on the birds’ song indicates that the distorted feedback protocol used to test the auditory responses of HVC(X) neurons was effective in causing plastic changes in the song control system.

These changes in the vocalizations were similar to those reported previously for adult birds (Leonardo and Konishi 1999) but occurred much faster—over the course of several days. The higher rate of song degradation is likely due to the fact that we used juvenile birds, consistent with previous observations that song degradation following deafening occurs faster in younger birds (Lombardino and Nottebohm 2000).

1 The online version of this article contains supplemental data.

FIG. 6. A and B: reversible chronic effects of distorted auditory feedback (DAF) on the vocalization of juvenile birds. A: reversible changes in the acoustic structure of the song syllable (syllable B, bird 17). Top: before DAF exposure; middle: after 7 days of continuous DAF exposure; bottom: 10 days after removal of DAF. B: changes in the structure of song syllables for a juvenile bird (bird 13) quantified by the syllable spectral similarity scores (see METHODS). Days on which DAF was presented are indicated by the gray shading. Statistically significant reversible changes in syllable structure were observed in syllables B and C. Error bars are ± SE. Although this bird’s (bird 13) song stabilized early (the similarity scores are ∼0.9 around posthatch day 55), the syllable structure underwent rapid and reversible degradation following DAF exposure. C: firing patterns of X-projecting HVC (HVC(X)) neurons in a juvenile zebra finch (bird 16) aligned to the song motif. For each neuron, trials without DAF are shown above the horizontal line and trials in the presence of DAF are shown below. Green lines, times during which distorted auditory feedback was presented during each motif. Also shown are microphone signal (top trace) and sound spectrograms without DAF (top) and with DAF (bottom).
DISCUSSION

Given the importance of HVC in song generation, it is crucial to characterize the singing-related neural activity in this nucleus to understand how the motor program underlying singing is generated. Moreover, because there are at least three distinct types of neurons in HVC with different postsynaptic targets, it is imperative to identify the neurons from which the recordings are made. This study provides the first detailed description of singing-related activity of HVC interneurons and both classes of projection neurons.

The three classes of HVC neurons have distinct firing patterns during singing: whereas interneurons have tonic firing patterns, the firing patterns of projection neurons are phasic and sparse. Taking advantage of the stereotyped song motif to align the firing patterns of neurons, we addressed the questions of synchrony among neurons and the relation between neural activity in HVC and vocal output.

Motivated by the idea that HVC(X) neurons carry a vocal performance evaluation signal to the AFP during singing, we directly tested this possibility. We altered the acoustic environment of a juvenile bird in a way that causes gradual song deterioration and studied the effects of such distorted acoustic feedback on the firing patterns of HVC(X) neurons.

In the following text, we summarize and interpret our findings. Our findings present several lines of evidence suggesting that the activity of HVC(X) neurons during singing transmits information about timing of the premotor sequence, but not auditory information, to the AFP.

Singing-related spiking activity of identified HVC neurons

Projection HVC neurons generate sparse and stereotypical patterns of bursts during singing. Bursts of both HVC(RA) and HVC(X) neurons are extremely tightly locked to the vocalization and frequently have rms jitter of <1 ms. The sequence of bursts of HVC projection neurons is one of the most temporally precise neural sequences found in nature to date. In contrast, HVC interneurons have tonic and more variable patterns of activity. We find that there is significant, albeit weak, synchrony between individual interneurons. There is also statistically significant synchrony between the populations of interneurons and HVC(X) neurons.

Neural activity in HVC and the syllable pattern of the song

We find that the population activity of HVC interneurons is modulated coherently with the syllable structure of the song. The population activity of interneurons leads the syllable pattern by ~45 ms. It has been noted previously that the onset of HVC multi-unit activity precedes the onset of the contact calls (which are learned vocalizations similar in acoustic structures to song syllables) by ~50 ms (McCasland 1987). This time delay between the onset of neural activity and the onset of sound has been interpreted to correspond to a neural delay between activity in HVC and vocal output (Troyer and Doupe 2000). Estimation of the neural delay between the activity in HVC and the vocal output is important because this delay contributes to the total delay between premotor activity and auditory feedback, which, in turn, is regarded as a fundamental difficulty in motor learning (Dave and Margoliash 2000; Troyer and Doupe 2000). Paired recordings in HVC and RA suggest that the bursts in RA are driven from HVC and the delay between bursts of HVC(RA) neurons and RA neurons is ~5 ms (Hahnloser et al. 2002). Direct stimulation of RA during a harmonic stack results in the transient vocal output perturbation that starts ~10–15 ms after the stimulation pulse.
(Fee et al. 2004). These results suggest that the delay between bursts of HVC(RA) neurons and the vocal output is ~15–20 ms, substantially smaller than 50 ms previously inferred from the onsets of multi-unit HVC activity. The functional significance of the syllable-related modulation of neural activity in HVC is unclear. However, the modulation may reflect activity in HVC involved in preparing the premotor circuitry or vocal organ for production of the next song syllable. For example, it may reflect external input to HVC involved in interhemispherical synchronization of HVC activity (Schmidt 2003).

**Relation of bursts of HVC\(_{(X)}\) neurons to acoustic structure of the song**

The idea that HVC transmits auditory information to the AFP suggests that bursts of HVC\(_{(X)}\) neurons may be related to specific acoustic elements in the song. Our recordings of firing patterns of HVC\(_{(X)}\) neurons allow for testing of this hypothesis: auditory-related activity of HVC\(_{(X)}\) neurons implies that multiple bursts of an HVC\(_{(X)}\) neuron would follow similar acoustic elements. This idea is not supported by our observations: we find that multiple bursts of a given HVC\(_{(X)}\) neuron are not associated with similar notes in the song. However, multiple bursts of a given HVC\(_{(X)}\) neuron tend to precede similar elements of the syllable pattern (either sound or silence) by ~40 ms. The lack of relation of bursts of an individual HVC\(_{(X)}\) neuron to similar spectral structure in the song does not support the hypothesis that HVC\(_{(X)}\) neurons transmit auditory information to the AFP.

**Neural activity and vocal output evolve on uncorrelated time scales**

We use the population activity of interneurons to assess the time scales of neural dynamics and to study the relation of timescales of neural dynamics to the characteristic times of acoustic output. We find that neural dynamics in HVC occurs on a fast fixed (~5 ms) time scales, whereas vocal dynamics occur on variable time scales (from <10 to ~100 ms) uncorrelated with the time scales of neural dynamics. This observation is consistent with the idea that HVC codes for temporal order in the song rather than for sound: whereas vocal output can remain essentially constant for ~100 ms (e.g., during a harmonic stack), the premotor sequence in HVC evolves on a substantially shorter time scale.

**Absence of auditory responses of HVC\(_{(X)}\) neurons during singing**

Lesion studies, together with the auditory responses exhibited in the AFP under anesthesia, have led to the hypothesis that the AFP processes auditory information to produce an error signal useful to guide vocal learning. In a simple implementation of this model, HVC might provide auditory information to the AFP, which then induces a gradual modification of the motor program. In this study, we used DAF to directly test the idea that HVC\(_{(X)}\) neurons are responsive to auditory signals or errors during singing. Several previous experiments have addressed the issue of auditory feedback in the AFP (Hessler and Doupe 1999a; Leonardo 2004; McCasland and Konishi 1981), but none of them were carried out in juvenile birds, and none have directly examined the role of HVC\(_{(X)}\) neurons in the processing of singing-related auditory signals.

We recorded from HVC\(_{(X)}\) neurons during singing with and without distorted feedback in juvenile birds and found no difference in the activity of HVC\(_{(X)}\) neurons with and without DAF. Absence of auditory sensitivity of HVC\(_{(X)}\) neurons during singing is in surprising contrast with their robust responses to playback of the bird’s own song (Mooney 2000), which are highly sensitive to distortions of the auditory input (Theunissen and Doupe 1998). During singing, HVC\(_{(X)}\) neurons generate extremely stereotyped patterns of spikes, thus even subtle changes in their firing patterns should be easily detectable.

Absence of auditory responses of HVC\(_{(X)}\) neurons would be expected if our distorted auditory feedback were not effective in inducing song plasticity. To rule out this possibility, we carried out recordings in young (70–90 days posthatch) birds the songs of which can undergo systematic changes over a time course of several hours (Deregnaucourt et al. 2003; Tchernichovski et al. 2001). We also verified that our DAF protocol was efficient in causing plastic changes in the motor program. Despite the effectiveness of DAF in song degradation and young age of birds, we observe no changes in the firing patterns of HVC\(_{(X)}\) neurons in the presence of DAF.

Absence of auditory responses of HVC\(_{(X)}\) neurons during singing raises questions about the possible role of HVC input to the AFP in vocal learning. It is conceivable that auditory responses of HVC\(_{(X)}\) neurons during singing are too weak to be detected in the sample of spike trains which can be obtained with our technique (typically 15–20 song motifs). However, it is not clear why vocal error correction would be carried out with such small signals, i.e., with low signal to noise. Given the exquisite vocal learning capabilities of the songbird, it seems advantageous for the auditory and error-correction circuits to have high signal to noise and for these circuits to effect gradual changes in the vocal output. Another possibility is that HVC transmits auditory information about vocal performance to the AFP offline, i.e., not during singing. There is some evidence from behavioral experiments suggesting that song learning may be happening offline as well as during sleep (Deregnaucourt et al. 2005; Tchernichovski et al. 2001). Regardless of when plasticity in the motor circuit occurs, auditory information about the bird’s vocal performance must be collected during singing. Lack of auditory responses to DAF in HVC might indicate, for example, that song-related auditory information may be stored in auditory areas upstream from HVC (e.g., Field L, NCM, cmHV, NIf) and transmitted through HVC to the AFP some time after singing, for example, during sleep (Dave and Margoliash 2000).

**Activity of HVC\(_{(X)}\) neurons may constitute precise timing signal**

A fundamentally different possibility is that HVC may not be involved in auditory processing, but is fundamentally involved in generating and transmitting timing information about the ongoing song. We have previously proposed that activity of RA-projecting HVC neurons may constitute a representation of temporal order in the vocal motor sequence (Hahnloser et al. 2002). Given that HVC\(_{(X)}\) neurons also generate precisely timed sparse sequences of bursts, similar to those observed in
HVC\textsubscript{(RA)} neurons, it is compelling to imagine that HVC\textsubscript{(X)} neurons perform a similar function of transmitting timing information to the basal ganglia. Further support of this idea is the lack of auditory responsiveness of HVC\textsubscript{(X)} neurons during singing, as well as the lack of correlation between the firing of these neurons and spectral structure of the song. Although the precise function of the AFP in song learning is unknown, it seems likely that information about timing of the ongoing motor sequence during learning will play an essential role in that function.

Acknowledgments

The experimental work described here was partially carried out at, and supported by, Bell Laboratories, Lucent Technologies.

Grants

Part of this work was supported by the Whitehall Foundation and by National Institute of Mental Health Grant R01 MH-067105.

References


