Article

Neuron

# Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex

## Highlights

- The population of HVC projection neurons is continuously active during singing
- Analysis shows that HVC activity is not tightly organized around motor gestures
- However, projection neurons and interneurons are rhythmically modulated by the song
- Coherence between projection neurons and song is stronger in juveniles than adults

## Authors

Galen F. Lynch, Tatsuo S. Okubo, Alexander Hanuschkin, Richard H.R. Hahnloser, Michale S. Fee

## Correspondence

fee@mit.edu

## In Brief

Lynch et al. show that HVC is not active only around motor gestures, as previously hypothesized. Instead, the population of HVC projection neurons is continuously active during song, with rhythmic modulations locked to song. Juvenile HVC exhibits stronger song-related modulation.

Lynch et al., 2016, Neuron *90*, 877–892
May 18, 2016 
http://dx.doi.org/10.1016/j.neuron.2016.04.021

CellPress

Neuron
Article

# Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex

Galen F. Lynch,[1,4] Tatsuo S. Okubo,[1,4] Alexander Hanuschkin,[2,3] Richard H.R. Hahnloser,[2,3] and Michale S. Fee[1,*]
[1]Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich 8057, Switzerland
[3]Neuroscience Center Zurich (ZNZ), Zurich 8057, Switzerland
[4]Co-first author
*Correspondence: fee@mit.edu
http://dx.doi.org/10.1016/j.neuron.2016.04.021

## SUMMARY

**Songbirds learn and produce complex sequences of vocal gestures. Adult birdsong requires premotor nucleus HVC, in which projection neurons (PNs) burst sparsely at stereotyped times in the song. It has been hypothesized that PN bursts, as a population, form a continuous sequence, while a different model of HVC function proposes that both HVC PN and interneuron activity is tightly organized around motor gestures. Using a large dataset of PNs and interneurons recorded in singing birds, we test several predictions of these models. We find that PN bursts in adult birds are continuously and nearly uniformly distributed throughout song. However, we also find that PN and interneuron firing rates exhibit significant 10-Hz rhythmicity locked to song syllables, peaking prior to syllable onsets and suppressed prior to offsets—a pattern that predominates PN and interneuron activity in HVC during early stages of vocal learning.**

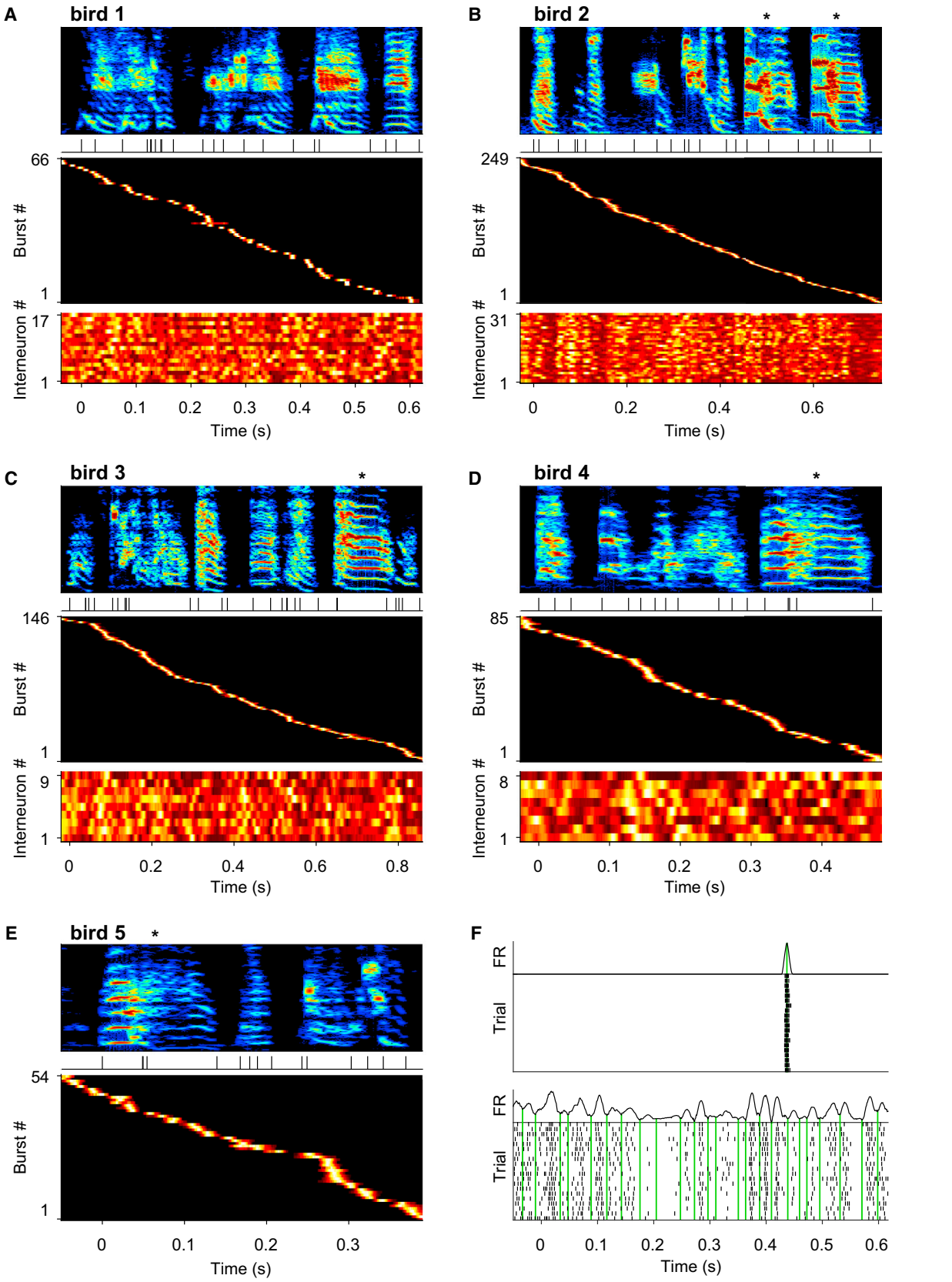
## INTRODUCTION

Learned motor behaviors require animals to acquire and produce precisely timed motor sequences. The origin of temporal control in motor behaviors remains unclear. While cortical motor regions traditionally have been viewed as encoding features of motor gestures (Evarts, 1968; Georgopoulos et al., 1988; Griffin et al., 2015; Hatsopoulos, 2005; Kalaska, 2009; Mussa-Ivaldi, 1988; Paninski et al., 2004; Todorov, 2000), these models do not address how sequences of motor gestures are produced or appropriately timed. Recently, a number of studies have suggested that, beyond just representing features of motor output, motor regions may have intrinsic oscillatory dynamics (Churchland et al., 2012) or sequential dynamics (Mita et al., 2009; Murakami et al., 2014; Shenoy et al., 2013) to act as their own pattern generators.

A similar set of issues has arisen recently in the field of vocal control in the songbird. Birdsong is a complex learned behavior involving a well-characterized set of brain regions (the song system). Both the production and learning of birdsong requires the avian premotor nucleus HVC (used as a proper name) (Aronov et al., 2008; Long and Fee, 2008; Nottebohm et al., 1976; Vu et al., 1998), a likely analog of mammalian motor cortex (Pfenning et al., 2014). HVC contains at least two classes of projection neurons (PNs) (Dutar et al., 1998; Kubota and Taniguchi, 1998; Mooney, 2000)—neurons projecting to the premotor nucleus robustus of the arcopallium (RA), part of a downstream motor pathway that innervates muscles of the vocal organ, and neurons projecting to Area X, part of a basal ganglia-thalamocortical loop necessary for song learning (Ali et al., 2013; Bottjer et al., 1984; Charlesworth et al., 2012; Scharff and Nottebohm, 1991). Both types of PNs generate sparse and brief (5- to 10-ms) bursts of spikes reliably active at one or a few times in the song, with different neurons active at different times (Fujimoto et al., 2011; Kozhevnikov and Fee, 2007; Long et al., 2010; Prather et al., 2008). The activity of HVC PNs has inspired two incompatible views of HVC function.

In one view, the population of HVC PNs bursts continuously throughout the song, essentially forming a clock that governs the processes of song production and learning at every moment in song (continuous-time model) (Fee and Goldberg, 2011; Glaze and Troyer, 2007; Hahnloser et al., 2002; Leonardo and Fee, 2005; Long and Fee, 2008; Long et al., 2010). For example, in one simple mechanistic model of song production, a unique ensemble of RA-projecting HVC ($HVC_{(RA)}$) neurons bursts at each time in the song, driving a continuous and varying pattern of activity in the downstream vocal premotor pathway (Fee et al., 2004; Leonardo and Fee, 2005). Continuous activity in $HVC_{(RA)}$ neurons also figures into recent models of sequence generation in which these neurons form a synaptically connected chain that supports the propagation of bursting activity (Glaze and Troyer, 2007; Hanuschkin et al., 2011; Jin et al., 2007; Li and Greenside, 2006; Long and Fee, 2008; Long et al., 2010). Finally, a recent model of vocal learning (Fee and Goldberg, 2011) incorporates X-projecting HVC ($HVC_{(X)}$) neurons to form a continuous-time code (or clock), allowing temporal specificity in learning (Charlesworth et al., 2011; Fee and Goldberg, 2011; Ravbar et al., 2012).

All of these models require continuous sequential activity in the population of HVC PNs; thus, we refer to them collectively as continuous-time models. While the simplest instantiations of




(legend on next page)

these models generally assume an equal number of neurons participating at each time in the sequence, such a uniform density of bursts is not required for their function.

An alternative hypothesis has been proposed recently (Amador et al., 2013), in which HVC activity is organized around a small number of time points in the song corresponding to discrete events in the trajectory of vocal control parameters, called gesture trajectory extrema (GTE). Amador et al. (2013) reported that bursts of PNs, as well as the minima of interneuron firing rates, were tightly synchronized with GTEs (4- to 7-ms SD). Based on their observations, Amador et al. (2013) went on to conclude that HVC PNs, as a population, must be silent between GTEs. Not only are these findings incompatible with the continuous-time models described above, but the near-zero latency of HVC bursts relative to song features rules out a premotor role for HVC in vocal output. In short, the GTE model would force a dramatic revision of our understanding of the mechanisms of song production.

Here we set out to reexamine the relation between HVC activity and vocal structure in the song by analyzing a large dataset of PNs and interneurons recorded in singing birds. Our analysis initially focused on neurons recorded in birds with mature adult song, testing the continuity and uniformity of HVC burst coverage, as well as investigating the alignment of PN and interneuron activity to GTEs and other song features. In adult birds, we found that HVC PNs, as a population, provide continuous and nearly uniform coverage of song. However, we also found that both $HVC_{(X)}$ neurons and interneurons exhibit a rhythmic modulation in firing rate coherent with an underlying ~10-Hz rhythmicity in song syllable structure, a feature not predicted by either the GTE or uniform models. Further, analysis of HVC neurons recorded in juvenile birds suggested that such rhythmic modulation may be a remnant of a developmental process in which oscillatory HVC activity drives the rhythmic repetition of early prototype syllables (Okubo et al., 2015; Tchernichovski et al., 2001).

## RESULTS

### HVC Bursting Produces Continuous Coverage of Time in the Song

While the continuous-time models make no specific prediction about the distribution of bursts in the song, they require that there be no periods of inactivity in the population of PNs. To test this prediction, we analyzed a large dataset of 450 neurons recorded during singing, which included 595 burst times generated by 384 PNs. The distribution and continuity of PN bursts in the song motif were quantified by identifying bursts from a histogram of spike times (Figure 1F). Subtle rendition-to-rendition variations in song speed were accounted for by time-warping the spike trains relative to syllable onsets and offsets (Kozhevnikov and Fee, 2007) (see the Experimental Procedures). Visual inspection of the distribution of bursts in time (Figures 1A–1E and S1A–S1E) suggested that the bursts provide nearly complete coverage of the song motif. The fraction of the motif covered by bursts (covered fraction) was quantified by determining the patches in time covered by the bursts of each neuron (see the Experimental Procedures). The covered fraction was defined, over all neurons, as the fraction of the entire song motif covered by at least one patch. Across birds, the median covered fraction was 96%, suggesting a very high degree of burst coverage in most birds (Figure 2A). A significantly lower covered fraction of only 67% was observed in one bird in which the fewest neurons were recorded (bird 1). By simulating a large number ($10^4$) of surrogate datasets, we found that the observed covered fraction for all birds was consistent with random sampling from a uniform underlying distribution of burst times (uniform model; Figure 2A; $p > 0.05$ for all birds). Note that this does not imply that the bursts are truly uniformly distributed, only that the observed coverage metric is consistent with random burst placement.

We next examined how the covered fraction would be expected to vary with the number of neurons recorded. For each bird, we generated surrogate datasets by resampling and shuffling different numbers of bursts from our data (see the Experimental Procedures). The number of bursts was expressed as fold coverage (Figures 2B and S2A), defined as the ratio of total burst duration to motif duration. Across birds, the simulated covered fraction exhibited a rapidly saturating dependence on fold coverage. Indeed, a fold coverage of ten resulted in a very high probability (>95%) of achieving total coverage (100% covered fraction) by bursts in the surrogate datasets (Figure S2B). Finally, the observed covered fraction for the individual PN subtypes was well explained by this saturating dependence (blue curve, Figures 2B and S2A).

### Quantification of HVC Clustering

Although the bursts of PNs in our dataset provided nearly complete coverage of the song, the bursts appeared to exhibit some degree of clumping in time (Figures 1A–1E). We wondered if this apparent clustering is simply due to random sampling of neurons or, alternatively, if it is more consistent with the alignment of many bursts at GTE times, as predicted by the GTE model. To address this question, burst times within the song motif were inferred from the histogram of time-warped spikes (see the Experimental Procedures). We examined the distribution of intervals between all burst times recorded in a single bird (population inter-burst interval, IBI), and we compared it to the uniform model (Figures 3A and S3A), in which burst times are described by a

**Figure 1. A Large Dataset of HVC Neurons Recorded in Singing Birds**

(A–E) Summary of PN and interneuron data for birds 1–5, respectively. Spectrograms shows the song motif for each bird; tick marks below indicate GTE times. Below is a heatmap summary of PN bursting activity. Each row shows the smoothed firing rate (normalized by peak firing rate) of a single burst; multiple bursts from one neuron appear on different rows, sorted by burst time. At the bottom is a heatmap summary of interneuron activity. Each row shows the smoothed firing rate of one recorded interneuron, normalized by its peak firing rate. No interneurons were recorded in bird 5.

(F) Burst and minima times (green vertical lines) are calculated from smoothed time histograms (black traces) of PN and interneuron spike trains (top and bottom, respectively). Vertical axis range is 500–7,500 Hz for all spectrograms.

See also Figure S1.

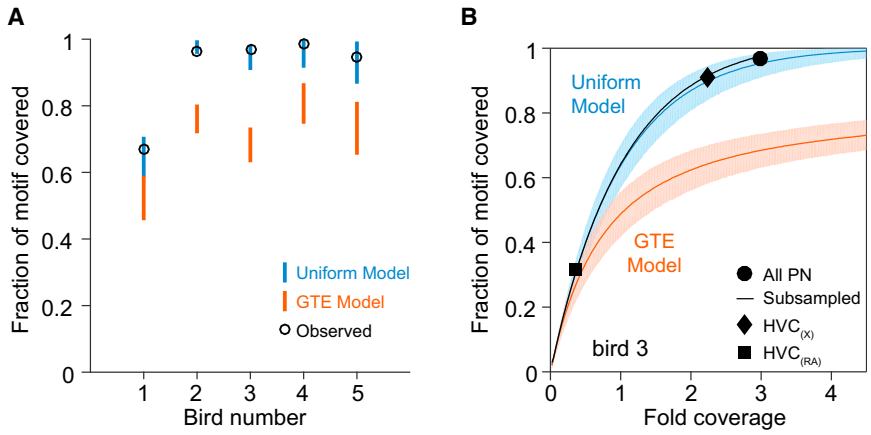



**Figure 2. Fraction of Song Motif Covered by Bursts**

(A) Summary of the fraction of the song motif covered by PN bursts for all five birds (covered fraction; black circles). Also shown is the 95% confidence interval of the covered fraction determined from random shuffling of burst times (vertical blue bars) and the 95% confidence interval of covered fractions predicted by the GTE model (i.e., with bursts placed around GTEs; vertical orange bars).

(B) The covered fraction calculated as a function of the number of bursts in the dataset (black trace, bursts randomly subsampled from observed bursts). Subsampled dataset size is expressed as fold coverage (sum of burst durations divided by motif duration). Data points show observed covered fraction for all bursts (black circle), $HVC_{(X)}$ neurons (black diamond), and $HVC_{(RA)}$ neurons (black square). Also shown is median covered fraction in 10,000 surrogate datasets in which bursts are randomly shuffled (uniform model; blue line) and median covered fraction for surrogate datasets in which bursts are placed around GTEs (GTE model; orange line). Model results are shown as a function of dataset size (fold coverage; see the Experimental Procedures). Shaded bands indicate 95% confidence interval (2.5 to 97.5 percentiles) of Monte Carlo results.

Results for all birds (birds 1–5) and simulations to determine the relation between probability of total coverage and fold coverage are shown in Figure S2.

Poisson process with an exponential distribution of IBIs. For four of five individual birds, the measured IBI distributions were consistent with the uniform model (Kolmogorov-Smirnov test, $p > 0.28$ for four birds), while the remaining bird (bird 2) showed a small deviation from uniformity (K-S test, $p = 0.014$, not significant after Bonferroni correction for five comparisons). A similar pattern of results was found with other measures of the IBI distribution focusing on the longest and shortest intervals (Figures 3B, 3C, S3B, and S3C). Bursts also were analyzed using a sensitive metric of clustering (Ripley's-L), and were found to be consistent with the uniform model in four of five birds (Figure 3D; $p > 0.32$). A small but significant self-clustering of bursts was detected by the Ripley's-L metric in one bird (Figure S3D; bird 2; $p = 4.0e{-}4$; 10% excess burst density in a 20-ms window compared to random burst placement). Thus, while largely consistent with a uniform distribution of burst times, these results reveal a small degree of non-uniformity that will be addressed further.

We next wondered how these measures of burst distribution compare to the predictions of the GTE model. For each bird, a large number ($10^4$) of surrogate datasets was generated, in which bursts were placed around measured GTEs with a Gaussian distribution. For each bird, the width of this Gaussian distribution was chosen to reproduce the burst-to-GTE variance reported by Amador et al. (2013) (average SD of 8 ms). In comparison to the uniform model, the GTE model predicts that bursts would cover a significantly lower fraction of the song (Figures 2B and S2A). It also predicts a larger number of short IBIs in the IBI distribution (Figures 3A, 3B, S3A, and S3B), due to the clustering of many bursts around a small number of GTEs, and a longer maximum IBI, due to the lack of bursts between GTEs (Figures 3C and S3C). The high degree of burst clustering in the GTE model was readily detected by the Ripley's-L metric (Figures 3D and S3D). For all birds, the measured values of each of these metrics were inconsistent with the predictions of the GTE model (covered fraction: $p \leq 2e{-}4$ in all birds; IBI: $p \leq 4e{-}4$ in birds 1, 2, 3, and 5 and $p \leq 0.013$ in bird 4; Ripley's-L: $p \leq 5e{-}3$ in all birds).

The analyses described above, and shown in the main figures, were carried out with GTEs identified using the method described by Amador et al. (2013) (Figures S1F–S1I). These analyses also were repeated using a recently published method for automatic identification of GTEs (Boari et al., 2015). While these methods produced somewhat different GTE times (Figures S1A–S1E, S1J, and S1K), the data were inconsistent with the predictions of the GTE model using either set of GTEs (Figure S3H).

## Coordination of HVC PN and Interneuron Activity

HVC interneurons play a significant role in several models of sequence generation and learning in HVC (Amador et al., 2013; Dutar et al., 1998; Kosche et al., 2015; Long et al., 2010; Markowitz et al., 2015; Marler and Peters, 1982; Mooney and Prather, 2005; Vallentin et al., 2016), and some models make specific predictions about the relative timing of interneuron and PN activity. For example, the GTE model predicts that local minima in the firing rate modulations of HVC interneurons coincide with GTEs (Amador et al., 2013). Interneurons were recorded in four of five adult birds ($HVC_i$ n = 65; birds 1–4), and they were found to spike continuously during the song motif while exhibiting song-locked firing rate modulations (Figures 1A–1F), as previously described (Kosche et al., 2015; Kozhevnikov and Fee, 2007; Markowitz et al., 2015; Yu and Margoliash, 1996). No significant correlation was observed between interneuron minima and burst times (Figures 3E and S3I) ($p = 0.60$, population mean; 1,728 minima; average of 26 ± 5 mean ± SD minima per motif per neuron; see the Experimental Procedures). Instead, interneuron firing rates increased at times of high PN burst density (Figures 3F and 3G); however, this pattern was somewhat different for $HVC_{(X)}$ and $HVC_{(RA)}$ subtypes. The correlation of interneuron firing rates with $HVC_{(X)}$ activity exhibited a strong peak at zero lag and a slow oscillation at longer lags (Figures 3F and S3J; burst correlation peak 4 ± 2 ms, ± SE; positive lag indicates PN leading interneuron; significance of peak, $p \leq 1e{-}4$), while the correlation with $HVC_{(RA)}$ activity revealed a

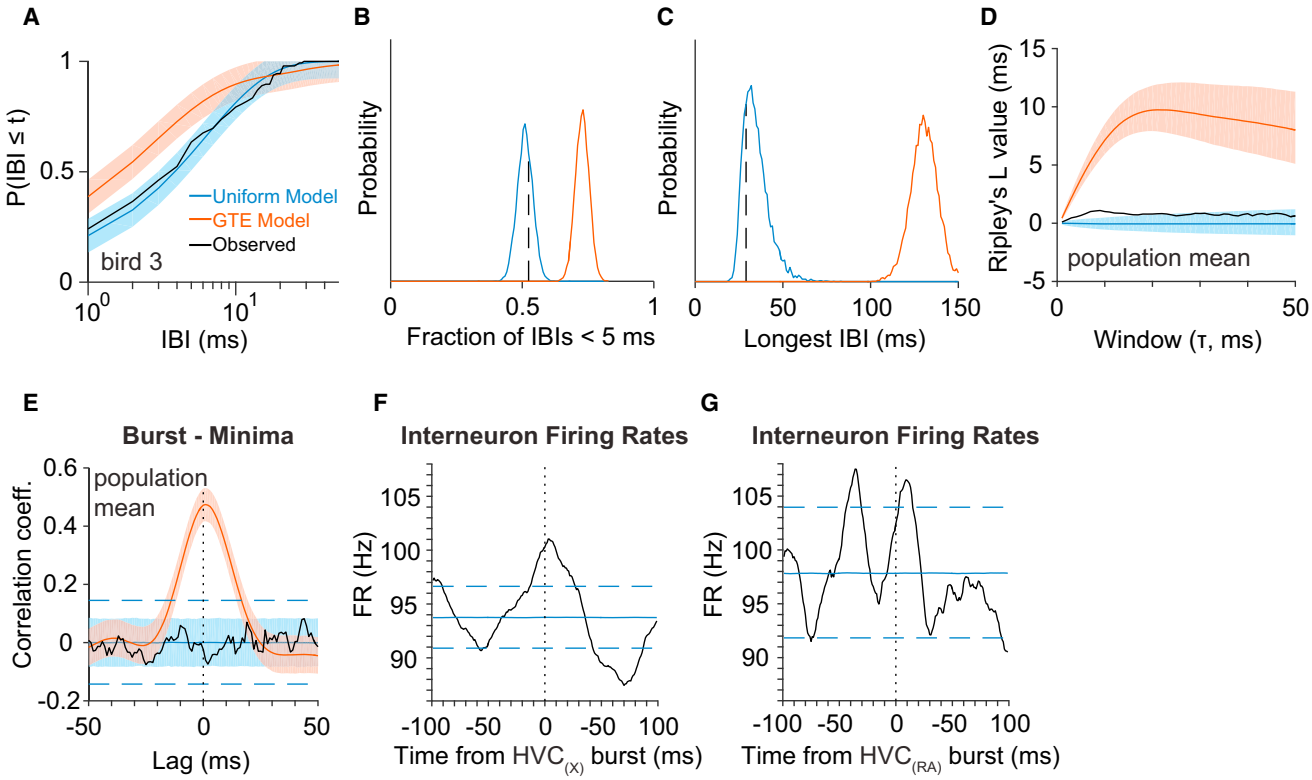



**Figure 3. Quantification of HVC Clustering**

(A) Cumulative distribution of the time intervals between all bursts recorded in bird 3 (IBIs; black trace), is compared to the median prediction of the uniform and GTE models (blue and orange lines, respectively; shaded region indicates 95% confidence interval).

(B) Fraction of IBIs less than 5 ms in bird 3 (vertical dashed black line) and distributions predicted by the uniform and GTE models (blue and orange curves, respectively) are shown.

(C) Longest interval between bursts recorded in bird 3 (black dashed line) is shown.

(D) Ripley's-L metric of clustering for bursts recorded across all birds is plotted as a function of the window size $\tau$ (black trace) and compared to the prediction of the uniform and GTE models (blue and orange solid lines, respectively; shading indicates 95% confidence interval).

(E) Cross-correlation between PN burst times and interneuron minima averaged across birds (black trace, negative lags indicate that minima precede bursts) is compared to the predictions of the uniform and GTE models (solid blue and orange curves, respectively; shading indicates 95% confidence interval in each lag bin, horizontal dashed lines indicate 95% confidence interval for maxima and minima of the uniform model anywhere in this window).

(F and G) Population average interneuron firing rate aligned at (F) $HVC_{(X)}$ burst times and (G) $HVC_{(RA)}$ burst times (black trace; median firing rate within 100 ms of burst, solid blue line). Clustering metrics for $HVC_{(X)}$ and $HVC_{(RA)}$ neurons are shown in Figure S3.

more complex pattern with multiple peaks (Figures 3G and S3K) and with greater bird-to-bird variability, perhaps reflecting the smaller number of $HVC_{(RA)}$ neurons recorded.

## Quantification of Alignment of Bursts and Minima to Temporal Features of Song

We next set out to determine whether PN bursts or interneuron minima are significantly aligned to syllable onsets, offsets, or GTEs. A cross-correlation analysis revealed that bursts are not significantly aligned at times of syllable onsets ($p = 0.64$), syllable offsets ($p = 0.93$), or GTEs ($p = 0.40$) (Figures 4A–4C and S4A). Similarly, interneuron minima exhibited no significant alignment at syllable onsets, offsets, or GTEs (Figures 4D and S4E). This result was confirmed with an alternative metric (Amador et al., 2013) based on the distribution of time differences between neural events and the GTEs closest to them (see the Experimental Procedures). These time differences were found to vary more widely (SD 8.7–17.9 ms for bursts and 9.2–16.8 ms for minima) than previously reported (4.0 ms for bursts and 4.0 ms for minima) (Amador et al., 2013). The measured time differences are inconsistent with the predictions of the GTE model ($p_{GTE} < 2.0e^{-4}$ for bursts and minima), and are instead consistent with a uniform random placement of burst and minima times in all birds (Figures 4E, 4F, S4F, and S4I; bursts: minimum $p = 0.046$, not significant after Bonferroni correction for five comparisons; minima: minimum $p = 0.015$, not significant, four comparisons). A separate analysis of $HVC_{(X)}$ and $HVC_{(RA)}$ neurons produced similar results (Figures S4G and S4H), as did an analysis carried out for GTEs identified by the automated method of Boari et al. (2015) (Figure S4B).

While we found no evidence for clustering of bursts at syllable onsets or offsets, we found that the density of HVC bursts was significantly higher in the first half of song syllables than in the second half, principally due to $HVC_{(X)}$ bursts ($HVC_{(X)}$ burst density 1.5-fold higher in first half than second half; $p = 0.018$,

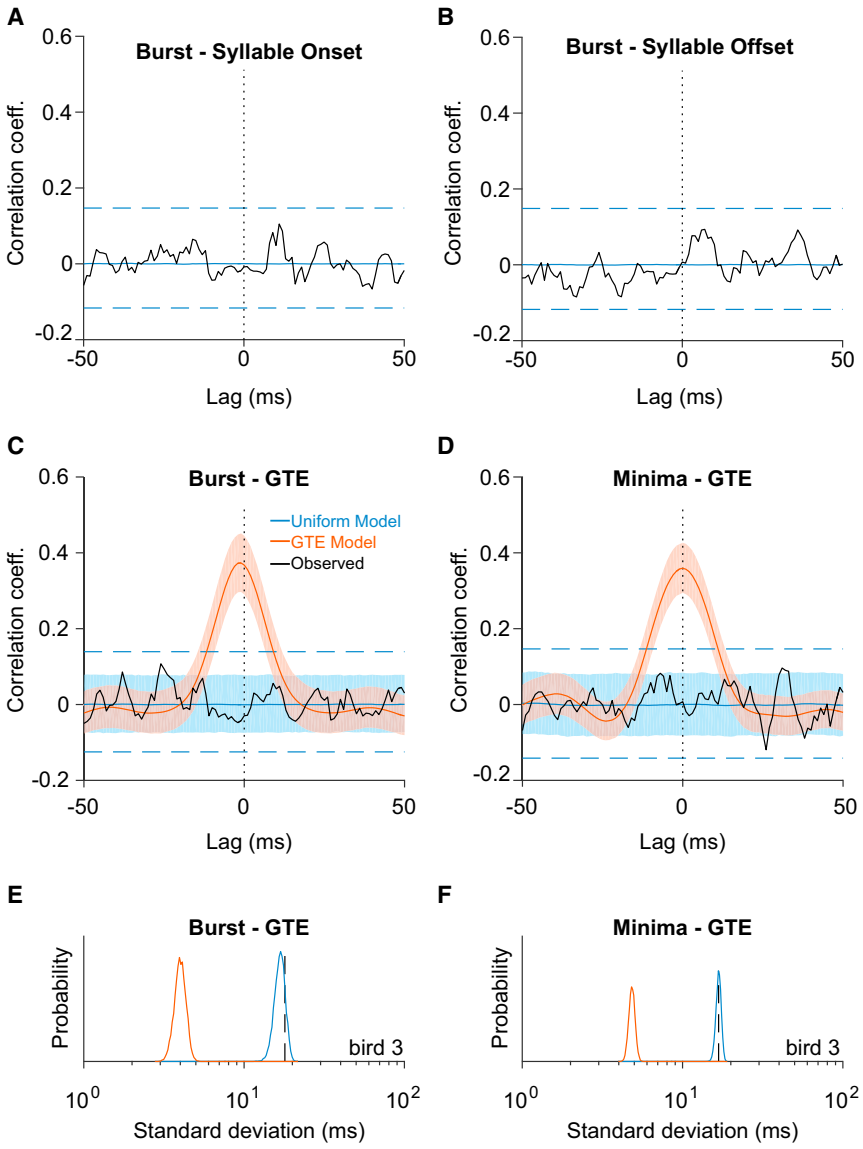



**Figure 4. Variations in Density of HVC Activity around Temporal Features of Song**

(A) Cross-correlation between bursts and syllable onsets, averaged across all birds (black trace, bursts lead onsets for negative lags). Horizontal dashed lines indicate 95% confidence interval for maxima and minima for shuffled data anywhere in this window.

(B) Cross-correlation between bursts and syllable offsets, averaged across all birds, is shown.

(C) Cross-correlation between bursts and GTEs times, averaged across all birds (black line), is compared with the median predicted value of both the uniform and GTE models (solid blue and orange lines, respectively; shading indicates 95% confidence interval in each bin; horizontal dashed lines indicate 95% confidence interval for maxima and minima anywhere in this window, for uniform model).

(D) Cross-correlation between interneuron minima and GTE times, averaged across the four birds in which interneurons were recorded, is shown.

(E) Metric of alignment between bursts and GTEs, based on the distribution of time differences between bursts and the GTEs closest to them. The observed SD for bird 3 is shown (dashed vertical line, blue and orange curves indicate predictions of uniform and GTE models, respectively).

(F) Same analysis as (E), but for interneuron minima. Results for all birds, as well as identified $HVC_{(X)}$ and $HVC_{(RA)}$ neurons, are shown in Figure S4.

Wilcoxon signed-rank). No significant modulation was observed for $HVC_{(RA)}$ bursts (1.24-fold higher; $p = 0.65$). These findings suggest that the density of HVC bursts exhibits a slow syllable-related modulation not apparent in the cross-correlation analyses of Figures 4A and 4B.

## Covariation of HVC Spike Timing and Song Timing

The continuous-time model predicts that HVC bursts exert a causal premotor influence on the timing of song structure, such as syllable onsets or offsets, regardless of variations in burst density. To see if our data are consistent with this view, we adopted a measure how HVC spike timing co-varies with these events (Ali et al., 2013). The cross-trial correlation between song-aligned time-warped spike trains across different song renditions was computed as a function of the delay of the alignment windows. An analysis of simulated spike trains, based on a chain model of HVC sequences, shows that the cross-trial correlation exhibits a peak when the alignment windows are advanced by a delay corresponding to the premotor latency (Figure S5). For measured PN spike trains, we observed a clear and unique maximum in this correlation at a premotor delay of 29 ± 4 ms (Figure 5A, ±SE; see the Experimental Procedures). A similar latency was obtained separately for $HVC_{(X)}$ and $HVC_{(RA)}$ subtypes (Figures 5B and 5C; $HVC_{(RA)}$: 28 ± 16 ms ; $HVC_{(X)}$: 28 ± 5 ms). Interneurons exhibited a broad peak with no clearly defined maximum (Figure 5C).

## Covariation of HVC Activity with Song Spectral Features

Although our findings suggest that HVC PNs are continuously active throughout song, the evidence described above—that bursts are less prevalent in the second half of song syllables—led us to wonder if such variations may be related to song spectral features (Tchernichovski et al., 2000). The burst density and population-average firing rate of PNs were analyzed for correlation with six commonly used spectral features (pitch goodness, Wiener entropy, amplitude, amplitude modulation, frequency modulation, and gravity center). Of these, only pitch goodness, a measure of harmonicity of sound, exhibited significant correlation after Bonferroni correction for multiple comparisons (Figures 6A and S6A; six comparisons). However, further analysis revealed that HVC PN activity was significantly weaker only during

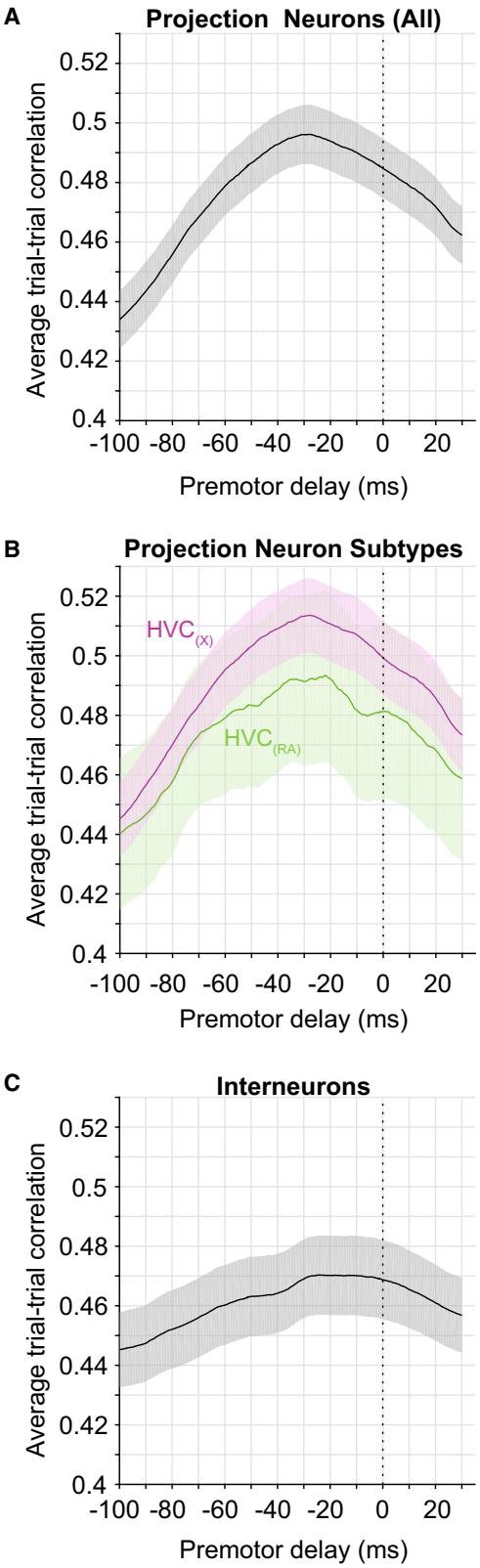



**Figure 5. Covariation between HVC Spike Timing and Song Timing**

Syllable onset and offset times are used to time-warp spike trains across different song renditions. The mean correlation of spike trains from different renditions is plotted as a function of how much the time-warping intervals are shifted relative to syllable onsets and offsets. Note that a maximum correlation, indicating optimal spike train alignment across renditions, occurs when the warping intervals are shifted 29 ms prior to syllable onsets and offsets.

(A) Mean spike-train correlation as a function of shift for HVC PNs is shown (shaded region indicates SEM).

(B) Spike-train correlation is plotted separately for identified $HVC_{(X)}$ and $HVC_{(RA)}$ PNs (purple and green, respectively).

(C) Spike-train correlation for HVC interneurons. Simulations validating this technique are shown in Figure S5.

low-frequency (<750-Hz) harmonic sounds, while the activity during high-frequency (>750-Hz) harmonic sounds was actually higher on average than during non-harmonic sounds (Figure 6B; one-way ANOVA, based on burst density $p = 0.0013$, based on firing rate $p = 4e{-}5$; compared to average burst density, the median burst density during low pitch, high pitch, and non-harmonic sounds was 0.43, 1.27, and 1.08, respectively). Note that, despite the lower burst rate during low-frequency harmonic elements (LHEs; median burst density 57% lower than average density), bursts in our dataset occurred throughout these syllable elements (Figure 6C).

A separate analysis of different PN subtypes revealed a significantly lower $HVC_{(X)}$ burst probability during LHEs than expected for random burst placement (ratio of probabilities 0.56; $p = 0.001$, two-tailed, bionomial test; $n = 269$ bursts). No significant modulation was found for $HVC_{(RA)}$ neurons (ratio 0.78; $p = 0.65$; $n = 47$ bursts). HVC interneurons exhibited a similar pattern of modulation with LHEs as did $HVC_{(X)}$ neurons (Figures S6C and S6D).

The reduced HVC activity during LHEs was most prominent at zero premotor latency, but it reached statistical significance in a range of latencies from −20 ms (burst preceding LHE) to +10 ms (burst following LHE).

Notably, in our dataset, LHEs occurred significantly more often in the second half of song syllables than in the first half (average probability of LHE 2.47-fold higher; $p < 0.01$, Wilcoxon signed-rank), consistent with the lower density of $HVC_{(X)}$ bursts during the second half of song syllables.

## HVC Firing Rates Exhibit Slow Syllable-Related Modulations

These observations led us to carry out a more detailed quantification of syllable-related modulations in the spiking activity of HVC neurons. Syllable onset-aligned and offset-aligned spike histograms of the interneuron population revealed a significant slow modulation peaking prior to syllable onsets (maximum at 18 ± 2 ms, ±SE, see the Experimental Procedures) and dipping prior to syllable offsets (minimum at 23 ± 2 ms) (Figures 7A, 7B, S7A, and S7B; average over all birds, $p \leq 1e{-}4$ for both comparisons). Interneurons also exhibited a weak but significant increase in firing rate prior to GTEs (Figure 7C; 23 ± 2 ms; $p \leq 1e{-}4$). PN firing rates exhibited a pattern of modulation roughly similar to that of interneurons, but this did not reach statistical significance in the windows prior to syllable onset or syllable offset (Figures 7G–7I; average over all PNs). Such modulations in PN firing rates may be explained in part by a slow modulation

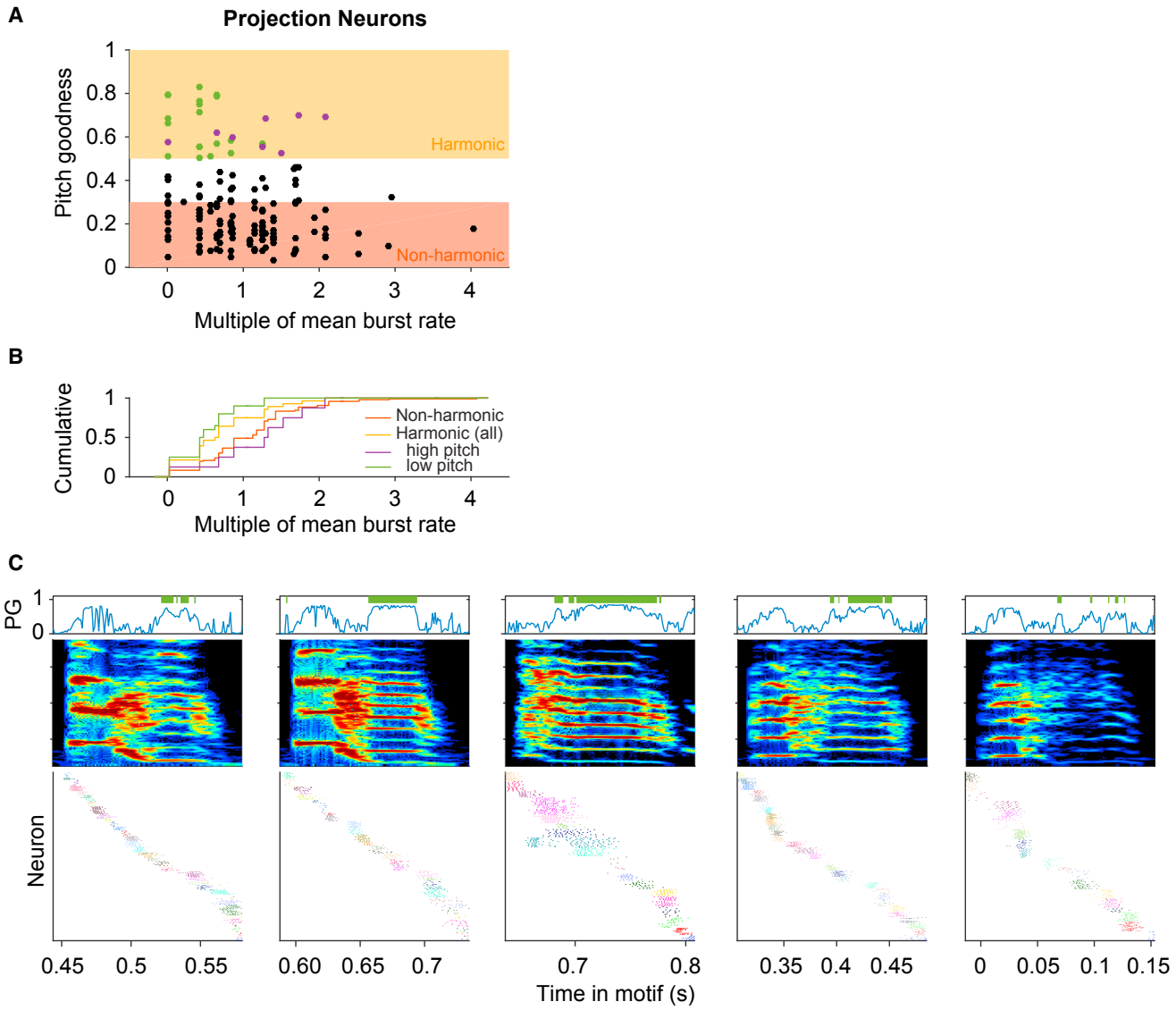



**Figure 6. Covariation of HVC Activity with Song Spectral Features**

(A) Scatterplot of HVC burst density, expressed as a multiple of mean burst rate, versus pitch goodness, averaged in 15-ms intervals. Periods of harmonic sound (pitch goodness > 0.5) are further classified as either low pitch (<750 Hz, green dots) or high pitch (>750 Hz, purple dots).

(B) Cumulative distribution of the burst rate during song times classified as non-harmonic (red), harmonic (yellow), low-pitch harmonic (green), or high-pitch harmonic (purple).

(C) Raster plots of PN spike trains shown for five song syllables with strong harmonic components. Pitch goodness is shown at top (blue trace). Times classified as low-pitch harmonic elements are indicated by green bars at top. Syllables shown here are indicated in Figure 1 by an asterisk above the spectrograms. A similar analysis with firing rates is shown in Figure S6.

in burst density, as described above, and also by a significantly higher spike rate within bursts around syllable onsets compared to bursts around syllable offsets (average 307 Hz in a window −30 ms to +10 ms around onsets; average 256 Hz in the same window around offsets; p = 0.001, rank-sum test).

While most adult zebra finch song is not highly rhythmic, it has been reported that these songs can contain an underlying rhythm in the 10-Hz range (Saar and Mitra, 2008), leading us to wonder if the observed HVC firing rate modulations also may have a rhythmic component. Consistent with previous reports, we found that the song amplitude profile of all of our adult birds exhibited a distinct spectral peak near 10 Hz (Figures 7D and S7C, range 7–10 Hz). The population spike trains of both interneurons and PNs exhibited a broad spectral peak near this same frequency (Figures 7D, 7J, S7C, and S7F). Further analysis revealed a large peak in the cross-spectral density between interneuron spike trains and song amplitude (Figures 7E and S7D, average over all interneurons), as well as a significant coherence at 10 Hz (Figures 7F and S7E; peak coherency at 10 Hz = 0.40 ± 0.03 SE; p < 1e−4; phase = $-0.81\pi$ radians).

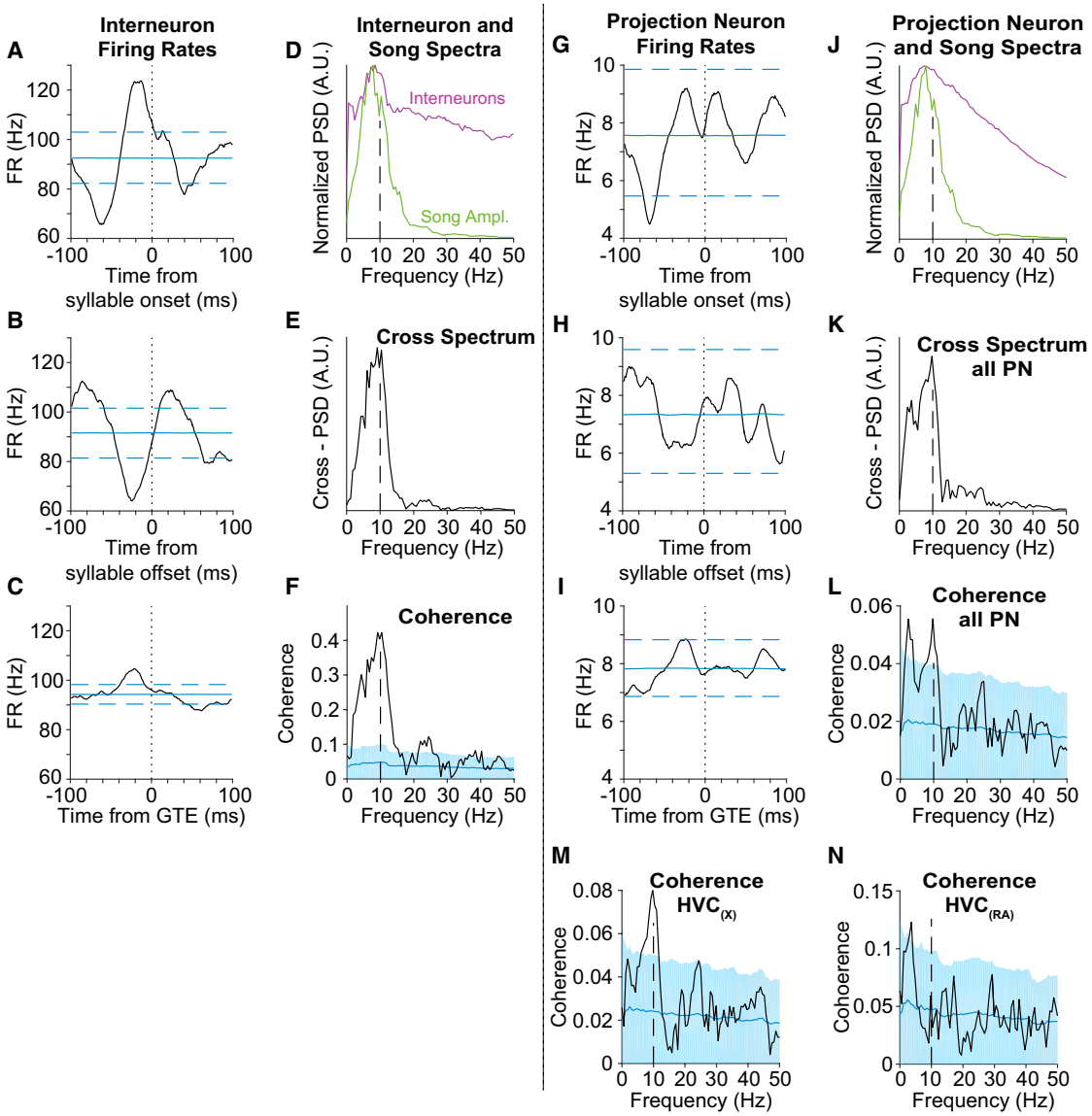



**Figure 7. Rhythmic Modulation of HVC Firing Rates Locked to Syllable Structure**

(A) Peri-event time histogram (PETH) of HVC interneuron firing rates aligned at syllable onsets, averaged across all birds, is shown (black trace, solid blue line indicates median interneuron firing rate; horizontal dashed lines indicate 95% confidence interval for maxima and minima in this window for shuffled data).

(B) Same as (A) is shown, but for interneuron firing rates aligned at syllable offsets.

(C) Histogram of interneuron firing rates aligned at GTE times is shown.

(D) Normalized spectrum of song amplitude (rhythm spectrum, green curve) and interneuron spike trains (purple curve), averaged over all birds. Vertical dashed line indicates 10 Hz.

(E) Cross-spectral density of the song amplitude and interneuron spike trains, averaged across birds, is shown.

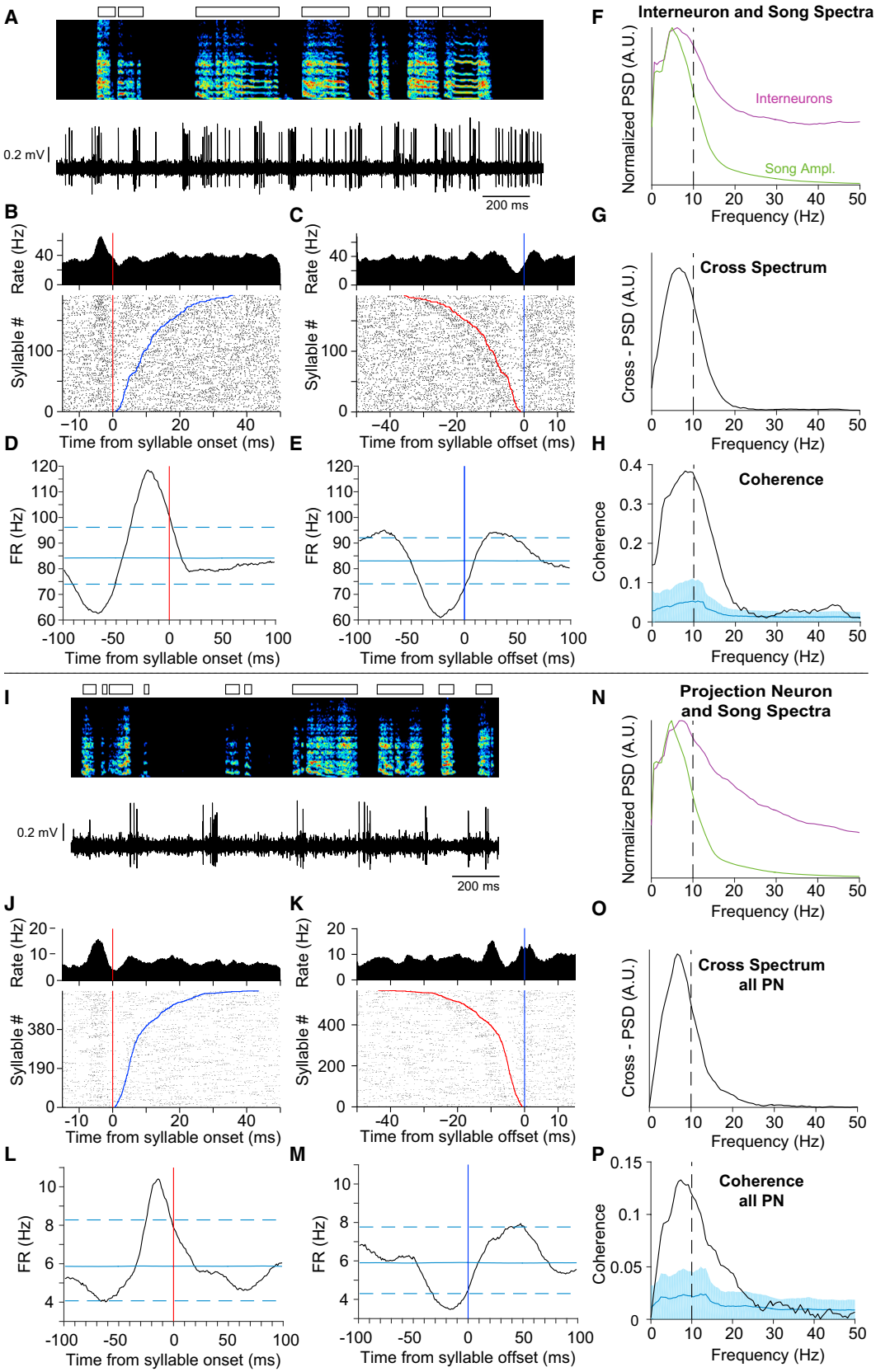
(F) Magnitude of coherency between song amplitude and interneuron spike trains, averaged across birds, is shown (black trace; solid blue curve, median; shading, 95% range of shuffled datasets).

(G–L) Same as in (A)–(F) are shown, but for all PN spike trains.

(M and N) Magnitude of coherency between song amplitude and PN spike trains, plotted separately for (M) $HVC_{(X)}$ and (N) $HVC_{(RA)}$ neurons.

Results for individual birds are shown in Figure S7.

A similar coherence analysis for PNs showed a weaker but significant peak in the cross-spectral density and coherence near 10 Hz (Figures 7K, 7L, S7G, and S7H; average over all PNs; peak coherency at 10 Hz = 0.056 ± 0.015; p = 0.013; phase = $-0.84\pi$ radians). Separate analysis of PN subtypes revealed no significant modulation in the population of $HVC_{(RA)}$ neurons (coherency at 10 Hz = 0.06 ± 0.04; p = 0.69; n = 57 neurons), but it did reveal significant rhythmic modulation of $HVC_{(X)}$ firing rates (coherency at 10 Hz = 0.08 ± 0.02; p = 0.0012; n = 238 neurons). The phase of the $HVC_{(X)}$ coherence was close to that of




*(legend on next page)*

interneurons ($-0.74\pi$ and $-0.81\pi$ radians, respectively). A similar coherence analysis using burst times rather than spike times yielded a qualitatively similar pattern of findings, suggesting that $HVC_{(X)}$ neurons exhibit a rhythmic modulation of burst density. While we cannot rule out rhythmic modulation of $HVC_{(RA)}$ activity, no significant modulation was observed in any of our analyses.

### HVC Activity in Juvenile Birds

Recordings in HVC of juvenile birds suggest that the rhythmic modulation in adult HVC may be explained in a developmental context. HVC PNs in young birds exhibit strongly rhythmic syllable-locked bursting that emerges at an early stage of song learning (Okubo et al., 2015), in which birds produce rhythmic protosyllables at roughly 10 Hz (Liu et al., 2004; Saar and Mitra, 2008; Tchernichovski et al., 2001). We analyzed a dataset of 116 HVC neurons recorded in young birds before they developed a mature song motif (n = 20 birds; 43–54 days post-hatching (dph); spanning subsong and protosyllable stages). We found that interneurons (n = 29 neurons; 15 birds) exhibited firing patterns qualitatively similar to those observed in adult birds (Figures 8A, S8A, and S8B) (Kosche et al., 2015; Kozhevnikov and Fee, 2007; Markowitz et al., 2015; Yu and Margoliash, 1996). They generated continuous spiking and bursting at an average firing rate of 81 ± 27 Hz during singing and 14 ± 7 Hz outside of singing (mean ± SD) (Figure S8C). The majority of individual interneurons exhibited a peak in firing rate prior to syllable onsets (Figure 8B; 27 of 29; threshold $p < 0.005$) and a dip prior to syllable offsets (Figure 8C; 24 of 29; $p < 0.005$). A similar pattern of modulation was observed in the population average interneuron firing rate (Figures 8D and 8E; peak 18 ± 1 ms prior to onset, ±SE; dip 22 ± 2 ms prior to offset; $p < 1e{-}4$). Spectral analysis revealed a significant peak in the spike spectrum for a large fraction (17/29) of individual interneurons in our juvenile dataset (between 4 and 15 Hz; Figures S8A and S8B), as well as in the population average spike spectrum (Figure 8F). A corresponding peak at ~10 Hz was seen in the cross-spectral density and in the coherence between song amplitude and the population of interneuron spikes (Figures 8G and 8H; peak coherency at 10 Hz = 0.38 ± 0.05, SE; $p < 1e{-}4$).

For PNs, a majority of both cell types (total of 87 neurons; 18 birds) exhibited a peak in firing rate prior to syllable onset and a dip prior to syllable offset (Figures 8I–8K; significant peak in 55/87 neurons, $p < 0.005$; significant dip prior to offsets in 27/87 individual PNs, $p < 0.005$). Notably, the population average of both $HVC_{(X)}$ and $HVC_{(RA)}$ neuron firing rates exhibited a pattern of modulations around syllable onsets and offsets similar to that of interneurons (Figures S8D and S8E; peak 14 ± 4 ms prior to onset, $p < 1e{-}4$; dip 17 ± 4 ms prior to offset, $p < 1e{-}4$). Spectral analysis revealed a significant peak in coherence between the population of PN spikes and song amplitude (Figure 8P; peak coherency at 10 Hz = 0.12 ± 0.03; $p < 1e{-}4$); a separate analysis of PN subtypes revealed a robust and significant peak in coherence at ~10 Hz for both $HVC_{(X)}$ neurons (Figure S8F; coherency magnitude at 10 Hz = 0.13 ± 0.03; $p = 1e{-}4$; n = 57 antidromically identified neurons) and $HVC_{(RA)}$ neurons (coherency magnitude at 10 Hz = 0.12 ± 0.08; $p = 0.04$; n = 16 antidromically identified neurons). Notably, the phase relation between song amplitude and HVC spiking was similar for both PN subtypes and interneurons at the peak frequency (10 Hz; $HVC_{(X)}$ phase = $-0.74\pi$; $HVC_{(RA)}$ phase = $-0.68\pi$; interneuron phase = $-0.84\pi$), suggesting that the rhythmic modulations in these neuron types are nearly synchronized.

In summary, in juvenile birds, interneurons and PNs of both subtypes exhibit significant syllable-related rhythmic modulation at ~10 Hz. For PNs, the size of this modulation was substantially larger in juvenile birds than in adults (by a factor of 2), and the difference was marginally significant in a statistical test of coherency differences ($p = 0.06$, one-tail permutation test) (Maris et al., 2007). For interneurons the modulation was not significantly different in juvenile and adult birds ($p = 0.65$).

## DISCUSSION

In this study, we set out to examine neural coding in HVC and understand its relation to song structure and timing. We initially

**Figure 8. HVC Activity in Juvenile Birds**

(A) Example of a putative HVC interneuron recorded during juvenile song (47 dph) is shown.

(B) Syllable onset-aligned raster and histogram for the neuron shown in (A), sorted by syllable duration. Red line indicates the syllable onset and blue line indicates syllable offset.

(C) Syllable offset-aligned raster and histogram for the neuron in (A) are shown.

(D) Mean syllable onset-aligned activity over all the putative HVC interneurons recorded from 43 to 54 dph (n = 40 neurons from 17 birds). Horizontal blue line indicates the median; dashed lines indicate 95% confidence interval for maxima and minima in this window for shuffled data.

(E) Same as in (J) is shown, but for syllable offset-aligned activity.

(F) Normalized spectrum for song amplitude and interneuron spike trains (green and purple curves, respectively), averaged over all birds. Vertical dashed line indicates 10 Hz.

(G) Cross-spectrum between song amplitude and interneuron spike trains is shown.

(H) Magnitude of coherency between song amplitude and interneuron spikes is shown (black curve; blue curve, median; shaded region, 95% confidence interval in shuffled data).

(I) Example of an HVC PN recorded during juvenile subsong ($HVC_{(X)}$, 45 dph) is shown.

(J) Syllable onset-aligned raster and histogram for the neuron in (I) are shown.

(K) Syllable offset-aligned raster and histogram for the neuron in (I) are shown.

(L) Mean syllable onset-aligned activity over all HVC PNs recorded between 40 and 55 dph (n = 194 neurons from 20 birds). Blue horizontal line indicates the median, and dashed lines indicate 95% range of maxima and minima in this window.

(M) Same as in (L) is shown, but for syllable offset-aligned activity.

(N–P) Same as (F)–(H), but for PNs.

See also Figure S8.

focused on the following two models of HVC coding: (1) continuous-time models in which the bursts of HVC PNs are continuously, and perhaps uniformly, distributed throughout the song; and (2) the GTE model, in which PN bursts and interneuron minima are clustered at specific time points in the song corresponding to the transitions and extrema of inferred motor trajectories (Amador et al., 2013). We examined a large dataset of HVC PNs and interneurons recorded in five birds with mature songs and found that our data are largely consistent with the predictions of the continuous-time model and are inconsistent with those of the GTE model. However, further analysis revealed aspects of HVC coding that are not captured by either of these models. These include a pronounced rhythmic modulation of interneuron firing rates related to syllable boundaries, a weaker rhythmic modulation of HVC PNs synchronized to that of interneurons, as well as a significant reduction in HVC activity during LHEs in the song.

An essential feature of the continuous-time models is that bursting activity extends throughout the song motif without interruption. In principle, song could be tiled with a sequence of bursts placed edge-to-edge, in which case only 100 neurons, each generating one burst of 10-ms duration, could encode a song motif of 1-s duration (fold coverage of one). Alternatively, a large number of bursts might be placed randomly throughout the song. We have estimated that with a fold coverage of only ten, such random placement would yield complete coverage in more than 95% of instances. In contrast, the large number of PNs in HVC (>16,000 of each class) (Wang et al., 2002; Wild et al., 2005) corresponds to a fold coverage of over 160, which would guarantee complete coverage even if burst density were strongly non-uniform—for example, even if parts of the song had a factor of ten lower burst density than average. While we did find evidence of non-uniformity in burst density (57% lower during LHEs and 33% lower during the latter half of song syllables), such variations were far weaker than would be required to produce gaps in burst coverage. This conclusion is further supported by the nearly complete coverage (median > 96%) directly measured in our dataset. Finally, we found no evidence for the dramatic non-uniformities in burst density predicted by the GTE model; a direct test of burst alignment to discrete temporal events in song, such as syllable onsets, offsets, or GTEs, revealed no evidence for clustering around these events.

One of the most surprising aspects of the GTE model, as formulated by Amador et al. (2013), is the proposed alignment of $HVC_{(RA)}$ bursts to GTEs with near-zero latency. This claimed alignment would preclude any premotor role for HVC in controlling vocal gestures, and it stands in stark contradiction with a large number of reports documenting a premotor function for HVC. These include lesion studies demonstrating the necessity of HVC for adult song (Aronov et al., 2008; Nottebohm et al., 1976; Simpson and Vicario, 1990) and recordings of HVC neurons showing premotor song-related activity (Ali et al., 2013; McCasland, 1987; McCasland and Konishi, 1981; Schmidt, 2003). Other studies revealed that disruption of HVC activity via electrical stimulation affects adult song production (Ashmore et al., 2005; Vu et al., 1994; Wang et al., 2008) and that acute manipulation of dynamics in HVC with temperature affects the temporal structure of the song (Andalman et al., 2011; Goldin et al., 2013; Long and Fee, 2008).

In contrast, the premotor role of HVC is a key aspect of the continuous-time model, in which temporal features of song, such as onsets, offsets, or acoustic transitions within a syllable, are each causally driven by a small population of $HVC_{(RA)}$ neurons active at some premotor latency prior to that feature. Based on an analysis of the covariation of spike timing and song timing, we found that HVC likely exerts a premotor influence on song after a delay of ~29 ms. This is consistent with the 35-ms latency determined from a similar analysis of multiunit data (Ali et al., 2013), and it is reasonably consistent with the sum of a 15-ms estimated premotor latency from RA to vocal output (Fee et al., 2004) and the ~10-ms antidromic latency from HVC to RA (Hahnloser et al., 2002).

There are a number of possible explanations for the discrepancy between our findings and those of Amador et al. (2013). Perhaps the most important difference is the size of the dataset. The large number of PNs we recorded in each bird (40–136 neurons) reveals a more complete picture of the HVC population activity, showing near continuous coverage of bursts, even during LHEs. Indeed, the ~50% lower burst density during LHEs may have been a contributing factor to the conclusion of Amador et al. (2013) that bursts occur so rarely during harmonic sounds. An additional difference between these studies may be the fact that the majority of neurons in our dataset were recorded in young adult birds that had recently completed learning their songs. $HVC_{(X)}$ neurons, which comprise the majority of neurons in these datasets, are involved in learning and may exhibit different patterns of bursting in young birds than in older birds. Recent findings also indicate a degree of heterogeneity in the connectivity and function of medial and lateral portions of HVC (Basista et al., 2014). Thus, some differences in the observed firing patterns could be explained if the Amador et al. (2013) dataset and our dataset included neurons recorded in different regions of HVC.

Causal manipulations of HVC PNs would provide a more direct test of the continuous-time and GTE models than measurements of burst density. For example, the chain model of sequence generation in HVC predicts that transiently silencing all $HVC_{(RA)}$ neurons at any time in the chain would terminate the song sequence. Also, the continuous-time model of $HVC_{(X)}$ neurons in learning predicts that transiently silencing these neurons at any time in the song would selectively render the bird incapable of modifying its song at the corresponding time. In contrast, the GTE model would predict no effect of silencing HVC PNs at times between GTEs.

We discovered several distinct, yet related, deviations from uniformity in the activity of $HVC_{(X)}$ neurons that may have an interesting relation to vocal learning. It is notable that the weaker $HVC_{(X)}$ activity at the ends of song syllables in our dataset is accompanied by an increased probability of finding LHEs. Such variation in song structure is a relatively common feature of zebra finch song syllables, which frequently begin with complex modulations and end with a downsweep in pitch or an LHE (Price, 1979). Thus, the increased density of $HVC_{(X)}$ bursts at times of high song complexity in the early part of song syllables may reflect the need for a denser representation in time in the Area X learning circuit. This may be related to the non-uniform time basis seen in a sensorimotor learning circuit in weakly electric fish (Kennedy et al., 2014).

Notably, for $HVC_{(RA)}$ neurons, we found no evidence of syllable-related modulation of burst density (related to LHEs, to early or late parts of the syllable, or to rhythmic modulation coherent with song structure). These observations suggest that $HVC_{(RA)}$ neurons may have a more uniform distribution in the song compared to $HVC_{(X)}$ neurons, further confirmation of which would require recordings from larger numbers of $HVC_{(RA)}$ neurons in individual birds.

The rhythmic modulation of HVC firing rates at ~10 Hz in adult birds may have a developmental origin. The adult song motif, consisting of three to seven syllables repeated every 0.5 to 1 s, emerges during learning from an earlier stage of song development in which primitive prototype syllables are rhythmically repeated at 10 Hz (Aronov et al., 2008; Liu et al., 2004; Okubo et al., 2015; Saar and Mitra, 2008; Tchernichovski et al., 2001). During this protosyllable stage, individual $HVC_{(X)}$ and $HVC_{(RA)}$ neurons generate rhythmic bursts locked to protosyllables (Okubo et al., 2015). Early on, these bursts occur at latencies predominantly clustered ~20 ms prior to syllable onsets. Such clustering, combined with a rhythmic repetition of protosyllables at ~10 Hz, would explain the observed coherence at ~10 Hz between HVC PN firing rates and song amplitude.

As learning progresses, HVC firing patterns change in a number of ways, as does the song. The distribution of bursts of both PN subtypes increasingly spreads out to occupy a greater range of latencies and form a more uniform sequence of bursts throughout developing song syllables (Okubo et al., 2015). Individual PNs burst less often and more selectively as new syllables emerge, until, in adult birds, few individual neurons generate rhythmic bursts at 10 Hz and the majority of PNs generate a single burst during the song motif. Our findings suggest that, ultimately, the population of $HVC_{(X)}$ neurons retains a significant rhythmic modulation locked to the underlying song rhythm, while the $HVC_{(RA)}$ population may not. It is unknown whether these modulations serve a functional role in song learning that persists into adulthood or are simply a vestige of an earlier highly rhythmic stage of song development.

Our findings provide insight into the possible role of HVC interneurons during singing. We found that the firing rates of $HVC_{(X)}$ neurons and interneurons are modulated synchronously, as revealed by the peak at zero lag in the burst-interneuron cross-correlations (Figure 3F) and the nearly identical phase of $HVC_{(X)}$ and interneuron coherence with song amplitude. These findings may argue against a view that $HVC_{(X)}$ bursts occur selectively during synchronized gaps in inhibition (Gibb et al., 2009; Kosche et al., 2015; Markowitz et al., 2015), and instead they support the view that inhibition may serve to balance increased activity within the excitatory network (Jin et al., 2007; Long et al., 2010). While the relation between $HVC_{(RA)}$ bursts and interneuron firing rates in our data is less clear, we find no evidence for increased $HVC_{(RA)}$ bursting at times of reduced interneuron spiking.

The rhythmic modulation of HVC activity may be related to observations of oscillatory activity in other motor systems, including physiological tremor (McAuley and Marsden, 2000), and rhythmic vocal/motor behaviors, such as lip-smacking in primates (Ghazanfar et al., 2013) and babbling in human infants (MacNeilage, 2008). While natural human speech is not strictly rhythmic, continuous natural speech is characterized by quasi-periodic modulation in sound amplitude and orofacial movements in the 3- to 8-Hz range, a modulation most easily detected in the coherence between speech amplitude and neural measures of speech control (Alexandrou et al., 2016; Ruspantini et al., 2012).

An analysis of rhythmic activity in primate motor cortex during reaching movements recently has been used to argue that rhythmic firing patterns in motor cortex at ~3 Hz result from an oscillatory dynamical system in cortex that generates and controls movements (Churchland et al., 2012). Although we found significant rhythmic modulation of firing rates in $HVC_{(X)}$ neurons, which likely correspond to corticostriatal neurons in mammalian neocortex, we found no evidence of rhythmic modulation of premotor $HVC_{(RA)}$ neurons. Selective lesions of $HVC_{(X)}$ neurons impair vocal learning but leave song production intact (Scharff et al., 2000), suggesting that these neurons do not participate in the on-line control of song vocalizations and consistent with the hypothesis that that $HVC_{(X)}$ neurons transmit timing information to basal ganglia circuits underlying vocal learning (Fee and Goldberg, 2011). Thus, our findings support the view that rhythmic activity in HVC may play a role in vocal learning, but they do not support a role for these modulations in adult vocal motor production.

## EXPERIMENTAL PROCEDURES

Please refer to the Supplemental Experimental Procedures for additional details.

### Subjects and Neural Recordings

Electrophysiological recordings were obtained from song control nucleus HVC in five zebra finches (*Taeniopygia guttata*) with mature song motifs (age range: 59 to 200 dph). HVC neurons also were recorded in juvenile zebra finches at earlier stages of song development (age range: 43 to 54 dph; n = 20 birds), comprising birds in the subsong and protosyllable stages as assessed by song structure (Okubo et al., 2015). Single-unit recordings of antidromically identified HVC neurons during singing were carried out using a motorized microdrive, and PNs were identified by antidromic stimulation from nucleus RA and Area X, as described elsewhere (Fee and Leonardo, 2001; Okubo et al., 2014). The care and experimental manipulation of subjects was carried out in accordance with NIH guidelines and has been reviewed and approved by the MIT Institutional Animal Care and Use Committee.

Bursts times were calculated from the peri-event time histogram (PETH) of spikes aligned to the song motif. Spike times were first time-warped by identifying the onset and offset of each syllable in the motif, calculating the amount by which each syllable and silent gap must be stretched or compressed to match a single fiducial rendition of the song motif, and then linearly warping spike times within each onset-offset and offset-onset window by this amount. Candidate burst windows were identified as periods when the average firing rate exceeded a low fixed threshold of 10 Hz. Candidate bursts were accepted if spikes occurred in the candidate burst window on at least 50% of song renditions. The time of the burst within the song motif was determined as the average of the spike times in the burst accumulated over all song renditions on which the neuron was recorded.

Local minima in the firing rate of HVC interneurons were extracted as described in the original GTE study (Amador et al., 2013). We generated a PETH of interneuron activity, which was then smoothed using a Savitzky-Golay filter with window size 10 ms and order 3. The minima of the smoothed PETH were found with a 20-ms sliding window.

### GTE Analysis

GTE times were extracted from the songs using procedures described in Amador et al. (2013) (Figures S1F–S1I). The approach is to use a dynamical model of the vocal organ (the syrinx) to infer the trajectory of two control parameters

—air sac pressure and labial tension (Amador and Mindlin, 2008; Perl et al., 2011). Continuous segments of control parameters are called gestures, and local maxima in either of the two control parameters within a gesture are called extrema. These, together with the beginning and end of the gesture, are identified as GTEs. The GTE analyses were carried out using custom MATLAB code by authors who did not have access to the neural data. A second set of GTEs was extracted using an automated method (Boari et al., 2015).

### Coverage Analysis

Each HVC spike in the burst window was assumed to exert a postsynaptic effect lasting 5 ms (Garst-Orozco et al., 2014) and was replaced with a 5-ms postsynaptic square pulse. For each burst, a region of the song was considered covered if, across recorded renditions, at least three post-synaptic pulses overlapped in time. This procedure yielded a small patch of time covered by the burst. The covered fraction was computed as the total covered duration divided by the duration of the motif (including flanking regions). The covered fraction predicted under the uniform and GTE models was determined as described in the Supplemental Experimental Procedures.

### Cross-Correlation Analysis

Cross-correlations between two series of event times, such as GTE times and burst times, were carried out using point-process Pearson normalized cross-covariances (Gabbiani and Cox, 2010), with slight modifications to allow for the possibility of bursts occurring simultaneously over the population of neurons. For display and testing purposes, this point process cross-correlation was binned in a sliding 5-ms bin with 1-ms steps. All p values reported are two sided, unless otherwise specified, and do not assume a symmetric distribution of the test statistic under each model.

For analyses showing modulation of firing rates around discrete events (burst times in Figures 3F and 3G or syllable onsets/offsets in Figures 7 and 8), the relative time between individual spikes and each discrete event was determined and binned with 1-ms bins. The resulting PETHs were smoothed with a 20-ms sliding window and normalized to display average firing rate in each bin. The precise temporal locations of peaks and minima in PETHs were extracted by fitting a quadratic function to a 30-ms window centered at the global maximum/minimum. SE was assessed by bootstrap methods. Significance of PETH maxima (or minima) was determined using Monte Carlo methods to estimate the 95% confidence interval of maxima (or minima) over the entire 100-ms time window for circularly shuffled spike trains. To combine results across birds, all neurons were given equal weight.

### Ripley's Clustering Analyses

To determine whether there is a clustering of bursts, minima, and GTE times, we used Ripley's-L analysis to detect significantly non-uniform clustering of points (Dixon, 2002). In brief, Ripley's-L finds the difference between the expected number of points observed within a time interval, $\tau$, around each point, and the expected number for randomly arranged points (Poisson distributed). To determine what window size ($\tau$) resulted in optimal discriminability between the uniform and GTE models, we computed the d' discriminability metric between the distributions of the L-statistic in these two models. The resulting curve of d' as a function of $\tau$ had a peak at 8 ms, and this is the value of $\tau$ we used for all statistical comparisons of the data with these models.

### Spectral Analysis of Song Amplitude and Spike Trains

Song amplitude was calculated as previously described (Aronov et al., 2011). Briefly, the microphone signal was bandpass filtered, rectified, and smoothed. The spectral analysis of the song amplitude and spike times was carried out using code from the Chronux package (Mitra and Bokil, 2007). Quantities calculated include power spectral density, cross-power spectral density, and coherency. For adult data, spectral quantities were computed for time series extending over the entire duration of the motif. For juvenile data, song bouts were divided into non-overlapping segments of 750-ms duration. Cross-spectrum was computed by adding coherently across neurons and birds. Significance of coherency was assessed by shuffling spike times over 10,000 Monte Carlo trials. Rhythmicity of interneurons recorded in juvenile birds was assessed by determining if there was a significant peak in the spike spectrum in the range of 4–15 Hz and further requiring that the global maximum of spike spectrum occur within this range. To combine spectral analysis results across birds, all neurons were averaged with equal weight.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and eight figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.04.021.

## AUTHOR CONTRIBUTIONS

G.F.L. and M.S.F. conceived and designed the study. T.S.O. collected experimental data for birds 2–5 and juvenile birds. G.F.L. and M.S.F. analyzed the neural data. A.H. and R.H.R.H. performed syringeal modeling to find GTEs. R.H.R.H. and M.S.F. provided funding and support. All authors contributed to writing the manuscript.

## ACKNOWLEDGMENTS

Funding to M.S.F. was provided by the NIH (grant R01DC009183) and to T.S.O. by the Nakajima Foundation and Schoemaker Fellowship. R.H.R.H. acknowledges support by the European Research Council (ERC-Advanced Grant 268911) and the Swiss National Science Foundation (grant 31003A_127024). We thank Michael Stetner and Emily Mackevicius for critical feedback and Michael Lynn for preliminary analyses of data from bird 1.

Received: June 19, 2015
Revised: February 17, 2016
Accepted: April 11, 2016
Published: May 18, 2016

## REFERENCES

Alexandrou, A.M., Saarinen, T., Kujala, J., and Salmelin, R. (2016). A multimodal spectral approach to characterize rhythm in natural speech. J. Acoust. Soc. Am. *139*, 215–226.

Ali, F., Otchy, T.M., Pehlevan, C., Fantana, A.L., Burak, Y., and Ölveczky, B.P. (2013). The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. Neuron *80*, 494–506.

Amador, A., and Mindlin, G.B. (2008). Beyond harmonic sounds in a simple model for birdsong production. Chaos *18*, 043123.

Amador, A., Perl, Y.S., Mindlin, G.B., and Margoliash, D. (2013). Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature *495*, 59–64.

Andalman, A.S., Foerster, J.N., and Fee, M.S. (2011). Control of vocal and respiratory patterns in birdsong: dissection of forebrain and brainstem mechanisms using temperature. PLoS ONE *6*, e25461.

Aronov, D., Andalman, A.S., and Fee, M.S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science *320*, 630–634.

Aronov, D., Veit, L., Goldberg, J.H., and Fee, M.S. (2011). Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. J. Neurosci. *31*, 16353–16368.

Ashmore, R.C., Wild, J.M., and Schmidt, M.F. (2005). Brainstem and forebrain contributions to the generation of learned motor behaviors for song. J. Neurosci. *25*, 8543–8554.

Basista, M.J., Elliott, K.C., Wu, W., Hyson, R.L., Bertram, R., and Johnson, F. (2014). Independent premotor encoding of the sequence and structure of birdsong in avian cortex. J. Neurosci. *34*, 16821–16834.

Boari, S., Perl, Y.S., Amador, A., Margoliash, D., and Mindlin, G.B. (2015). Automatic reconstruction of physiological gestures used in a model of birdsong production. J. Neurophysiol. *114*, 2912–2922.

Bottjer, S.W., Miesner, E.A., and Arnold, A.P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science *224*, 901–903.

Charlesworth, J.D., Tumer, E.C., Warren, T.L., and Brainard, M.S. (2011). Learning the microstructure of successful behavior. Nat. Neurosci. *14*, 373–380.

Charlesworth, J.D., Warren, T.L., and Brainard, M.S. (2012). Covert skill learning in a cortical-basal ganglia circuit. Nature *486*, 251–255.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Dixon, P. (2002). Ripley's K function. Encycl. Environmetrics *3*, 1796–1803.

Dutar, P., Vu, H.M., and Perkel, D.J. (1998). Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVc of the adult zebra finch. J. Neurophysiol. *80*, 1828–1838.

Evarts, E.V. (1968). Relation of pyramidal tract activity to force exerted during voluntary movement. J. Neurophysiol. *31*, 14–27.

Fee, M.S., and Goldberg, J.H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience *198*, 152–170.

Fee, M.S., and Leonardo, A. (2001). Miniature motorized microdrive and commutator system for chronic neural recording in small animals. J. Neurosci. Methods *112*, 83–94.

Fee, M.S., Kozhevnikov, A.A., and Hahnloser, R.H.R. (2004). Neural mechanisms of vocal sequence generation in the songbird. Ann. N Y Acad. Sci. *1016*, 153–170.

Fujimoto, H., Hasegawa, T., and Watanabe, D. (2011). Neural coding of syntactic structure in learned vocalizations in the songbird. J. Neurosci. *31*, 10023–10033.

Gabbiani, F., and Cox, S.J. (2010). Mathematics for Neuroscientists (Burlington, MA: Academic Press).

Garst-Orozco, J., Babadi, B., and Ölveczky, B.P. (2014). A neural circuit mechanism for regulating vocal variability during song learning in zebra finches. eLife *3*, e03697.

Georgopoulos, A.P., Kettner, R.E., and Schwartz, A.B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. J. Neurosci. *8*, 2928–2937.

Ghazanfar, A.A., Morrill, R.J., and Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. Proc. Natl. Acad. Sci. USA *110*, 1959–1963.

Gibb, L., Gentner, T.Q., and Abarbanel, H.D.I. (2009). Inhibition and recurrent excitation in a computational model of sparse bursting in song nucleus HVC. J. Neurophysiol. *102*, 1748–1762.

Glaze, C.M., and Troyer, T.W. (2007). Behavioral measurements of a temporally precise motor code for birdsong. J. Neurosci. *27*, 7631–7639.

Goldin, M.A., Alonso, L.M., Alliende, J.A., Goller, F., and Mindlin, G.B. (2013). Temperature induced syllable breaking unveils nonlinearly interacting timescales in birdsong motor pathway. PLoS ONE *8*, e67814.

Griffin, D.M., Hoffman, D.S., and Strick, P.L. (2015). Corticomotoneuronal cells are "functionally tuned". Science *350*, 667–670.

Hahnloser, R.H.R., Kozhevnikov, A.A., and Fee, M.S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature *419*, 65–70.

Hanuschkin, A., Diesmann, M., and Morrison, A. (2011). A reafferent and feedforward model of song syntax generation in the Bengalese finch. J. Comput. Neurosci. *31*, 509–532.

Hatsopoulos, N.G. (2005). Encoding in the motor cortex: was Evarts right after all? Focus on "motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks". J. Neurophysiol. *94*, 2261–2262.

Jin, D.Z., Ramazanoğlu, F.M., and Seung, H.S. (2007). Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. J. Comput. Neurosci. *23*, 283–299.

Kalaska, J.F. (2009). From intention to action: motor cortex and the control of reaching movements. Adv. Exp. Med. Biol. *629*, 139–178.

Kennedy, A., Wayne, G., Kaifosh, P., Alviña, K., Abbott, L.F., and Sawtell, N.B. (2014). A temporal basis for predicting the sensory consequences of motor commands in an electric fish. Nat. Neurosci. *17*, 416–422.

Kosche, G., Vallentin, D., and Long, M.A. (2015). Interplay of inhibition and excitation shapes a premotor neural sequence. J. Neurosci. *35*, 1217–1227.

Kozhevnikov, A.A., and Fee, M.S. (2007). Singing-related activity of identified HVC neurons in the zebra finch. J. Neurophysiol. *97*, 4271–4283.

Kubota, M., and Taniguchi, I. (1998). Electrophysiological characteristics of classes of neuron in the HVc of the zebra finch. J. Neurophysiol. *80*, 914–923.

Leonardo, A., and Fee, M.S. (2005). Ensemble coding of vocal control in birdsong. J. Neurosci. *25*, 652–661.

Li, M., and Greenside, H. (2006). Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *74*, 011918.

Liu, W.C., Gardner, T.J., and Nottebohm, F. (2004). Juvenile zebra finches can use multiple strategies to learn the same song. Proc. Natl. Acad. Sci. USA *101*, 18177–18182.

Long, M.A., and Fee, M.S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature *456*, 189–194.

Long, M.A., Jin, D.Z., and Fee, M.S. (2010). Support for a synaptic chain model of neuronal sequence generation. Nature *468*, 394–399.

MacNeilage, P. (2008). The origin of speech (OUP Oxford).

Maris, E., Schoffelen, J.-M., and Fries, P. (2007). Nonparametric statistical testing of coherence differences. J. Neurosci. Methods *163*, 161–175.

Markowitz, J.E., Liberti, W.A., 3rd, Guitchounts, G., Velho, T., Lois, C., and Gardner, T.J. (2015). Mesoscopic patterns of neural activity support songbird cortical sequences. PLoS Biol. *13*, e1002158.

Marler, P., and Peters, S. (1982). Long-term storage of learned birdsongs prior to production. Anim. Behav. *30*, 479–482.

McAuley, J.H., and Marsden, C.D. (2000). Physiological and pathological tremors and rhythmic central motor control. Brain *123*, 1545–1567.

McCasland, J.S. (1987). Neuronal control of bird song production. J. Neurosci. *7*, 23–39.

McCasland, J.S., and Konishi, M. (1981). Interaction between auditory and motor activities in an avian song control nucleus. Proc. Natl. Acad. Sci. USA *78*, 7815–7819.

Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y., and Tanji, J. (2009). Interval time coding by neurons in the presupplementary and supplementary motor areas. Nat. Neurosci. *12*, 502–507.

Mitra, P., and Bokil, H. (2007). Observed Brain Dynamics (Oxford University Press).

Mooney, R. (2000). Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. J. Neurosci. *20*, 5420–5436.

Mooney, R., and Prather, J.F. (2005). The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. J. Neurosci. *25*, 1952–1964.

Murakami, M., Vicente, M.I., Costa, G.M., and Mainen, Z.F. (2014). Neural antecedents of self-initiated actions in secondary motor cortex. Nat. Neurosci. *17*, 1574–1582.

Mussa-Ivaldi, F.A. (1988). Do neurons in the motor cortex encode movement direction? An alternative hypothesis. Neurosci. Lett. *91*, 106–111.

Nottebohm, F., Stokes, T.M., and Leonard, C.M. (1976). Central control of song in the canary, Serinus canarius. J. Comp. Neurol. *165*, 457–486.

Okubo, T.S., Mackevicius, E.L., and Fee, M.S. (2014). In vivo recording of single-unit activity during singing in zebra finches. Cold Spring Harb. Protoc. *2014*, 1273–1283.

Okubo, T.S., Mackevicius, E.L., Payne, H.L., Lynch, G.F., and Fee, M.S. (2015). Growth and splitting of neural sequences in songbird vocal development. Nature *528*, 352–357.

Paninski, L., Fellows, M.R., Hatsopoulos, N.G., and Donoghue, J.P. (2004). Spatiotemporal tuning of motor cortical neurons for hand position and velocity. J. Neurophysiol. *91*, 515–532.

Perl, Y.S., Arneodo, E.M., Amador, A., Goller, F., and Mindlin, G.B. (2011). Reconstruction of physiological instructions from Zebra finch song. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *84*, 051909.

Pfenning, A.R., Hara, E., Whitney, O., Rivas, M.V., Wang, R., Roulhac, P.L., Howard, J.T., Wirthlin, M., Lovell, P.V., Ganapathy, G., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. Science *346*, 1256846.

Prather, J.F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature *451*, 305–310.

Price, P.H. (1979). Developmental determinants of structure in zebra finch song. J. Comp. Physiol. Psychol. *93*, 260–277.

Ravbar, P., Lipkind, D., Parra, L.C., and Tchernichovski, O. (2012). Vocal exploration is locally regulated during song learning. J. Neurosci. *32*, 3422–3432.

Ruspantini, I., Saarinen, T., Belardinelli, P., Jalava, A., Parviainen, T., Kujala, J., and Salmelin, R. (2012). Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2-3 Hz. J. Neurosci. *32*, 3786–3790.

Saar, S., and Mitra, P.P. (2008). A technique for characterizing the development of rhythms in bird song. PLoS ONE *3*, e1461.

Scharff, C., and Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. J. Neurosci. *11*, 2896–2913.

Scharff, C., Kirn, J.R., Grossman, M., Macklis, J.D., and Nottebohm, F. (2000). Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. Neuron *25*, 481–492.

Schmidt, M.F. (2003). Pattern of interhemispheric synchronization in HVc during singing correlates with key transitions in the song pattern. J. Neurophysiol. *90*, 3931–3949.

Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical control of arm movements: a dynamical systems perspective. Annu. Rev. Neurosci. *36*, 337–359.

Simpson, H.B., and Vicario, D.S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. J. Neurosci. *10*, 1541–1556.

Tchernichovski, O., Nottebohm, F., Ho, C.E., Pesaran, B., and Mitra, P.P. (2000). A procedure for an automated measurement of song similarity. Anim. Behav. *59*, 1167–1176.

Tchernichovski, O., Mitra, P.P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. Science *291*, 2564–2569.

Todorov, E. (2000). Direct cortical control of muscle activation in voluntary arm movements: a model. Nat. Neurosci. *3*, 391–398.

Vallentin, D., Kosche, G., Lipkind, D., and Long, M.A. (2016). Neural circuits. Inhibition protects acquired song segments during vocal learning in zebra finches. Science *351*, 267–271.

Vu, E.T., Mazurek, M.E., and Kuo, Y.C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. J. Neurosci. *14*, 6924–6934.

Vu, E.T., Schmidt, M.F., and Mazurek, M.E. (1998). Interhemispheric coordination of premotor neural activity during singing in adult zebra finches. J. Neurosci. *18*, 9088–9098.

Wang, N., Hurley, P., Pytte, C., and Kirn, J.R. (2002). Vocal control neuron incorporation decreases with age in the adult zebra finch. J. Neurosci. *22*, 10864–10870.

Wang, C.Z.H., Herbst, J.A., Keller, G.B., and Hahnloser, R.H.R. (2008). Rapid interhemispheric switching during vocal production in a songbird. PLoS Biol. *6*, e250.

Wild, J.M., Williams, M.N., Howie, G.J., and Mooney, R. (2005). Calcium-binding proteins define interneurons in HVC of the zebra finch (Taeniopygia guttata). J. Comp. Neurol. *483*, 76–90.

Yu, A.C., and Margoliash, D. (1996). Temporal hierarchical control of singing in birds. Science *273*, 1871–1875.