Article

Neuron

Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch

Highlights

- A novel head-fixed singing bird preparation for two-photon imaging of network activity
- Observation of skilled motor behavior in a cortical circuit
- Use of premotor activity during singing to test a range of population models
- Song-related dynamics divorced from ongoing motor movements

Authors

Michel A. Picardo, Josh Merel, Kalman A. Katlowitz, ..., Eftychios A. Pnevmatikakis, Liam Paninski, Michael A. Long

Correspondence

mlong@med.nyu.edu

In Brief

Picardo et al. use two-photon imaging and intracellular electrophysiology in the singing zebra finch to demonstrate that premotor cortical activity does not reflect ongoing song-related kinematics but, instead, appears to form an abstract population sequence during song performance.

Picardo et al., 2016, Neuron *90*, 866–876
May 18, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.neuron.2016.02.016

CellPress

CellPress

Neuron
Article

Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch

Michel A. Picardo,[1,2] Josh Merel,[3,4] Kalman A. Katlowitz,[1,2] Daniela Vallentin,[1,2] Daniel E. Okobi,[1,2] Sam E. Benezra,[1,2] Rachel C. Clary,[1,2] Eftychios A. Pnevmatikakis,[3,4,5] Liam Paninski,[3,4] and Michael A. Long[1,2,*]

[1]New York University Neuroscience Institute and Department of Otolaryngology, New York University Langone Medical Center, New York, NY 10016, USA

[2]Center for Neural Science, New York University, New York, NY 10003, USA

[3]Department of Statistics and Center for Theoretical Neuroscience, Columbia University, New York, NY 10027, USA

[4]Grossman Center for the Statistics of Mind, Columbia University, New York, NY 10027, USA

[5]Simons Center for Data Analysis, Simons Foundation, New York, NY 10010, USA

*Correspondence: mlong@med.nyu.edu

http://dx.doi.org/10.1016/j.neuron.2016.02.016

SUMMARY

The zebra finch brain features a set of clearly defined and hierarchically arranged motor nuclei that are selectively responsible for producing singing behavior. One of these regions, a critical forebrain structure called HVC, contains premotor neurons that are active at precise time points during song production. However, the neural representation of this behavior at a population level remains elusive. We used two-photon microscopy to monitor ensemble activity during singing, integrating across multiple trials by adopting a Bayesian inference approach to more precisely estimate burst timing. Additionally, we examined spiking and motor-related synaptic inputs using intracellular recordings during singing. With both experimental approaches, we find that premotor events do not occur preferentially at the onsets or offsets of song syllables or at specific subsyllabic motor landmarks. These results strongly support the notion that HVC projection neurons collectively exhibit a temporal sequence during singing that is uncoupled from ongoing movements.

INTRODUCTION

Song production in the zebra finch provides an excellent opportunity to examine the processes that shape the representation of a single complex learned behavior as it progresses from dedicated higher-order centers through downstream targets, ultimately leading to the flexion of muscles needed to produce the song (Ashmore et al., 2005; Leonardo and Fee, 2005). Although the motor pathway in the songbird is composed of well identified and spatially segregated neural circuits (Nottebohm et al., 1976), the means by which this singing behavior is represented in these regions is still a matter of debate (Troyer, 2013). One controversy centers on a single forebrain area (called HVC) that plays a central role in song production (Long and Fee, 2008; Vu et al., 1994). In one view, HVC projection neurons reflect motor-related aspects of the ongoing song (Amador et al., 2013; Boari et al., 2015), similar to the coding scheme observed in the mammalian primary motor cortex (M1) (Churchland et al., 2012; Evarts, 1968; Georgopoulos et al., 1986; Todorov, 2000). In the other view, these neurons may represent relative time within the motor act (Fee et al., 2004; Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010) without regard to the kinematics of the vocal apparatus. Similar abstract motor representations have been shown to exist in higher-order cortical sites such as the supplementary motor area (Matsuzaka et al., 2007; Mita et al., 2009; Nachev et al., 2008; Tanji and Shima, 1994). An intermediate hypothesis, in which HVC may represent both movement and elapsed time, may also be valid and has been suggested to exist in M1 (Carpenter et al., 1999; Lu and Ashe, 2005, 2015; Matsuzaka et al., 2007). Distinguishing between these mechanisms of motor control would be a significant step forward in our understanding of how singing behavior is encoded within HVC, potentially extending to skilled behaviors in other forebrain circuits.

Technical challenges using traditional electrophysiological approaches in singing birds have prevented a clear characterization of the HVC premotor network (Amador et al., 2013; Day et al., 2013; Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010; Markowitz et al., 2015). Previously, song-related neural responses were collected one at a time over the span of many weeks and then aligned to singing behavior (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010; Okubo et al., 2015). This process is inefficient, and coding of the behavior at an ensemble level could have shifted considerably during that time (Huber et al., 2012; Okubo et al., 2015; Peters et al., 2014), potentially leading to misinterpretations that would be avoided using a method that could allow for a "snapshot" of large neural populations in a single recording session. In the rodent, in vivo two-photon imaging has enabled the study of large populations of motor cortical neurons during the performance of

CrossMark

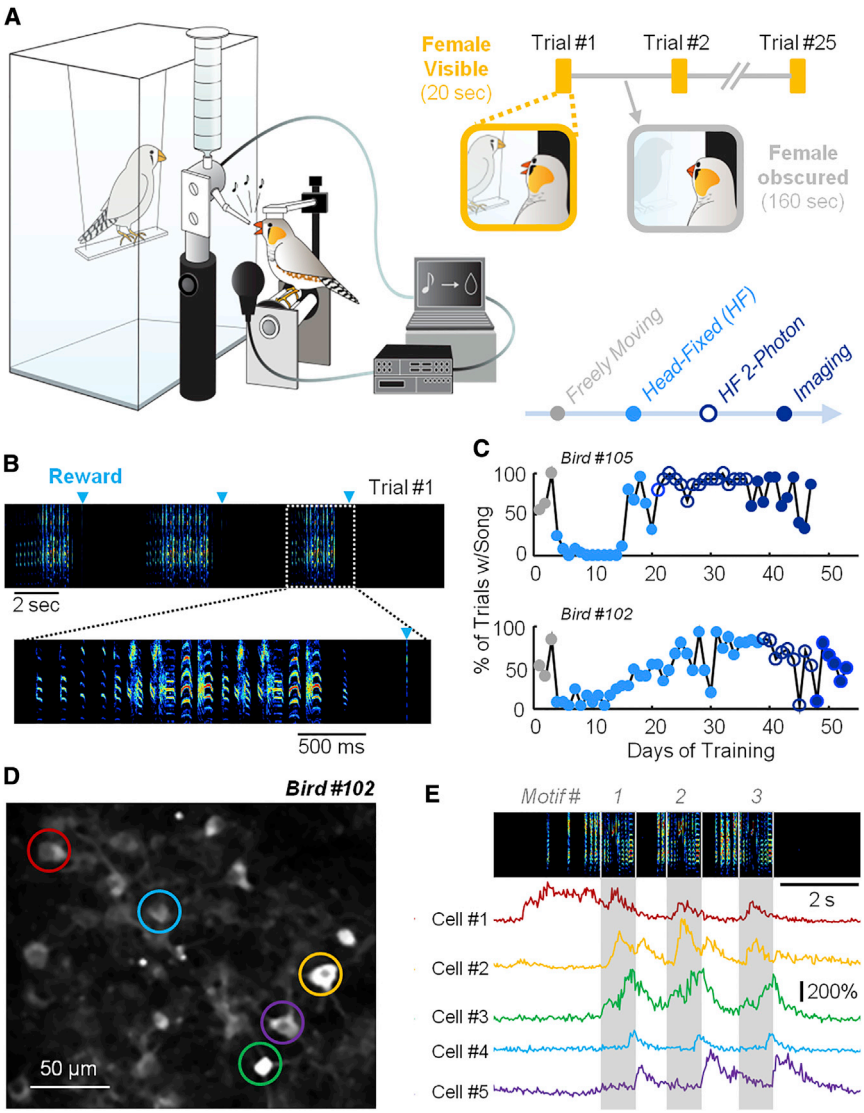

Figure 1. Imaging Neural Activity in HVC of a Head-Fixed Zebra Finch

(A) A schematic of the arena used to train head-fixed zebra finches. In our training paradigm, we used polarized glass to provide visual access to a female.

(B) An example spectrogram (frequency, 0.5–8 kHz) showing singing behavior elicited during a trial.

(C) A timeline displaying the progression of training conditions. The percentage of trials with song for each day is plotted for two birds.

(D) GCaMP6s-expressing neurons within HVC.

(E) Examples of fluorescence measurements in the neurons circled in (D). Individual song motifs are highlighted in gray.

RESULTS

Two-Photon Calcium Imaging in the Head-Fixed Singing Bird

We wanted to test the proposed hypotheses of song representation by examining the neuronal activity at a population level with two-photon microscopy (Denk et al., 1990). Recent experiments have used two-photon imaging of HVC activity in the anesthetized zebra finch to measure sensory responses within that structure (Graber et al., 2013; Peh et al., 2015). To adapt this approach to address motor dynamics, we developed a head-fixed singing bird preparation. Because we were not able to elicit head-fixed song in initial attempts (n = 6 birds), we used operant conditioning to encourage this behavior. Male zebra finches were separated from females by a glass partition whose transparency was under experimental control, and they were given visual access to females at regular intervals (Figure 1A). Males were rewarded for singing during those epochs (Figure 1B; Figures S1A–S1D) and then transitioned to a head-fixed context when song could be evoked reliably (Figure 1C; Figures S1E, S1F, and S1I). The vast majority of these birds (71 of 78) produced at least one head-fixed song motif (Table 1). Although head fixation can alter behavior and the associated neural activity for certain tasks (Ravassard et al., 2013), we found that the quality ($p = 0.27$, $F = 1.47$, one-way ANOVA) and timing ($p = 0.95$, paired t test) of singing behavior for five individual birds were not significantly different compared with song produced during free movement (Figures S1J–S1L).

A subset of the head-fixed singing birds was acclimated to the two-photon microscope (Figures S1G and S1H; Movie S1). In those individuals, we virally expressed a genetically encoded calcium indicator (GCaMP6s or GCaMP6f) (Chen et al., 2013) to report the activity of HVC neurons during singing. To calibrate the calcium indicator in vivo, we performed simultaneous

tractable behaviors (Huber et al., 2012; Li et al., 2015; Peters et al., 2014).

For these experiments, we developed a head-fixed preparation that enabled us to image populations of HVC projection neurons during singing. We then applied a new computational algorithm that enabled high-precision estimates for event timing by integrating across song trials. Additionally, we analyzed a series of intracellular recordings in which spikes and excitatory postsynaptic events were used to reflect motor-related network activity within HVC. First, we demonstrate that the rate of HVC premotor bursts during silent gaps in the song does not differ relative to epochs of active singing. We then found that HVC activity occurs continuously within the context of a syllable rather than concurrent with identified motor components of the song. Our results show that behaviorally relevant temporal sequences within HVC of the zebra finch are uncoupled from the properties of each constitutive movement, akin to higher-order cortical areas in primates (e.g., Tanji and Shima, 1994).

Table 1. Zebra Finch Training

	Freely Moving	Head-Fixed (No 2P)	Head-Fixed 2P	Head-Fixed 2P and Imaging
Number of birds trained	109	78	17	6
Number of sessions	540	2,474	187	85
Number of trials	18,500	61,750	2,834	1,405
Trials with song	7,463	11,811	1,509	756
Number of song motifs	37,112	66,025	6,800	2,868
Motifs per trial	2.0	1.1	2.4	2.0

Shown is the performance of zebra finches in each phase of the training program. Under the last two conditions, only a subpopulation of head-fixed birds was used for this study, and further training was carried out under the two-photon microscope (2P).

juxtacellular recordings and calcium imaging (Figure S2A) of bursting activity in an anesthetized and pharmacologically disinhibited zebra finch (Figures S2B–S2E). The latency from the first spike of the burst to the onset of the calcium transient was 5.9 ± 3.7 ms (n = 3 neurons from 2 birds). The standard deviation of onset timing within individual neurons was 3.5 ± 1.3 ms (e.g., Figure S2E), which supported the notion that GCaMP6 could reliably report neural activity with high temporal precision. During singing, the majority of HVC neurons exhibited one (cells 1, 4, and 5) or a few (cells 2 and 3) distinct calcium transients (Figures 1D and 1E), consistent with previous measurements taken with a head-mounted CMOS camera (Markowitz et al., 2015). For this study, we restricted our analysis to these sparsely active neurons that are likely to represent HVC premotor [$HVC_{(RA)}$] or basal ganglia-projecting [$HVC_{(X)}$] cells (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010), as opposed to other observed neurons that showed sustained fluorescence during singing and probably represent inhibitory interneurons (Hahnloser et al., 2002; Kosche et al., 2015).

Increasing Temporal Resolution by Integrating Data across Behavioral Renditions

We next asked whether our approach could provide sufficient temporal resolution to test hypotheses concerning the representation of singing behavior within the HVC circuit. High temporal resolution is critical for detecting the precise onset of (~10 ms) bursting responses within HVC (Hahnloser et al., 2002; Long et al., 2010). However, the onset time of putative burst events was difficult to estimate precisely in individual trials (e.g., Figures 2A, 2C, and 2E). To address this problem, we took advantage of the stereotyped timing inherent in both the zebra finch song as well as the underlying HVC bursting activity (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007). For a given focal plane, we imaged during several renditions of the song (Figure 2B) and aligned calcium traces from multiple trials to the singing behavior (Figure 2D). Using a Bayesian method (Figures S2D and S2F–S2I), we found that combining fluorescence measurements across trials could greatly sharpen the temporal resolution of our estimates for burst onsets (Figures 2C–2G; see Experimental Procedures for details). For instance, in an example cell that had been imaged for 23 song motifs, the burst events could be detected with significantly more precision when considering all trials (mean σ_{onset} = 3.0 ms; Figure 2F) compared with estimates taken from single trials (mean σ_{onset} = 16.3 ms; Figure 2E). Across our dataset, the uncertainty of our estimate for burst onsets decreased as a function of the number of motifs analyzed (Figure 2G) and as a function of the signal-to-noise ratio of measurements from individual neurons (Figure 2H).

Using Calcium Imaging to Investigate Motor Representation at a Population Level

When the timing of HVC bursts was determined, we examined the relationship between these events and singing behavior. We asked whether bursts occurred preferentially during periods of active singing. Zebra finches produce a repeated series of vocal elements, known as "syllables," that are approximately 75–300 ms in length and are separated by shorter (15- to 100-ms) silent "gaps" during inspiration (Hartley and Suthers, 1989) (Figures 3A and 3B). Although syringeal muscles can be activated during both syllables and gaps, electromyogram recordings during song production show dynamic patterns of muscle activation that correlate with the rapidly changing acoustic features of song (Goller and Cooper, 2004; Riede and Goller, 2010; Suthers et al., 1994; Vicario, 1991a). As a result, the kinematic hypothesis would predict that HVC projection neurons would be significantly more active during syllables than during silent gaps, whereas the temporal sequence hypothesis would not anticipate a difference between these conditions. In total, we considered 294 burst events (250 cells) from five birds producing a total of 45 syllables (Figure 3; Figure S3). We found no relationship between the relative frequency of HVC projection neuron bursts and the presence of a syllable (74.0 bursts/s) or a gap (76.7 bursts/s) ($p > 0.05$, paired t test), supporting the notion that bursts are forming a continuous sequential representation throughout the song motif. Additionally, there was no temporal alignment of burst events to the onsets (Figure 3C) or offsets (Figure 3D) of syllables.

We next wanted to examine the fine structure of syllables to test whether HVC projection neuron activity is best explained by the presence of specific motor movements or by a continuous temporal sequence. Because a direct assessment of muscular activity was not feasible in our preparation, we turned to a previously established tool for estimating vocal muscle gestures from recorded songs (Amador et al., 2013; Boari et al., 2015; Figure S4). These events, called gesture-trajectory extrema (GTEs), are associated with changes in syringeal tension or subsyringeal pressure (Amador et al., 2013) and can be estimated by analyzing song structure (Boari et al., 2015). We first examined the relationship between GTEs and burst onsets (e.g., Figure 3B) and found no significant correlation between these events at any time lag in both individual birds and across the population (Figure 3E). To validate our method, we performed a simulation where burst events were placed at syllabic and subsyllabic time points and adjusted for associated experimental uncertainties as well as the inherent temporal variance suggested previously to exist between these events (Amador et al., 2013). In

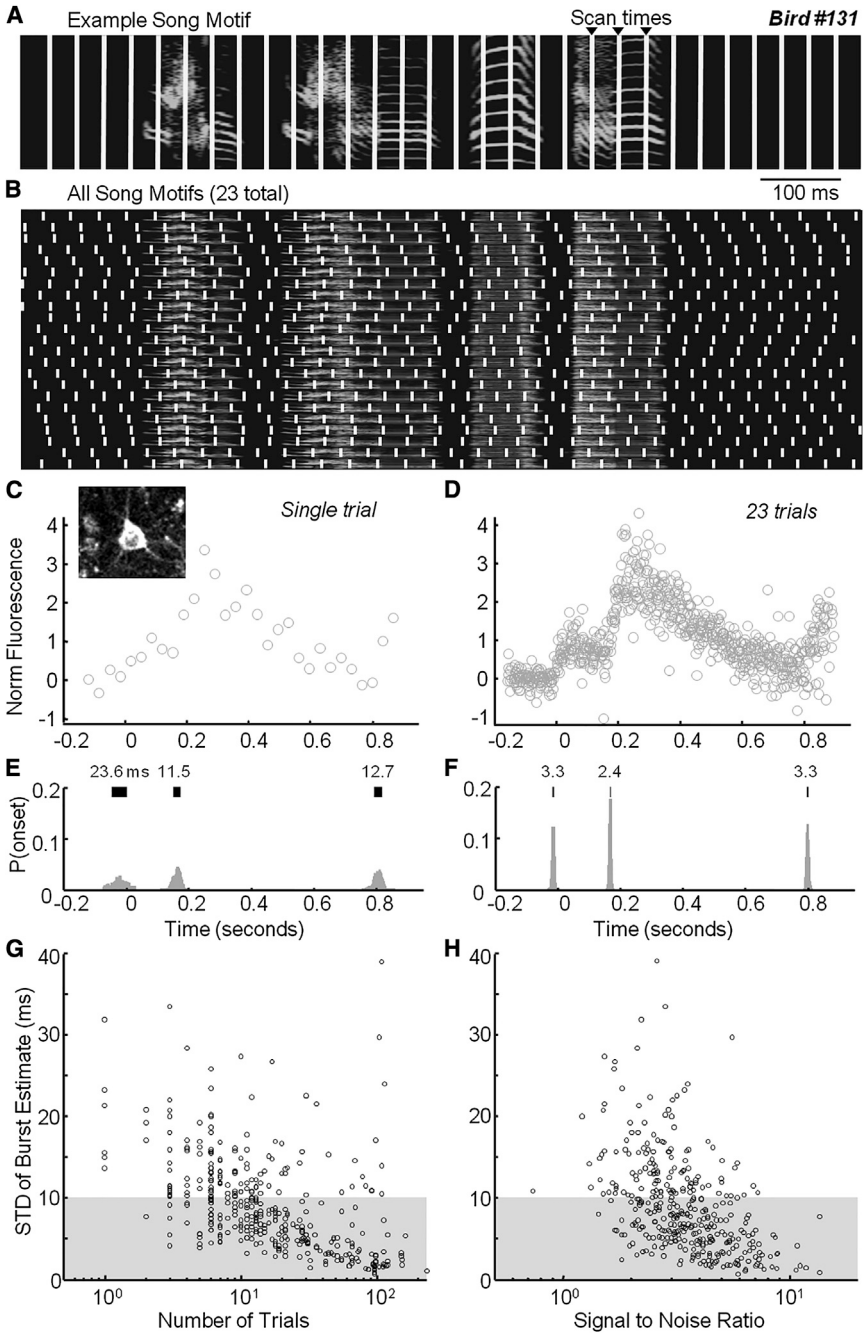

Figure 2. Integrating Imaging Data across Trials

(A and B) Spectrograms with vertical lines representing the scan times for a single motif (A) and across 23 motifs (B).

(C and D) Fluorescence transients for one neuron (inset). We show the results from a single trial (29 points) (C) as well as measurements taken across all trials (667 points) (D).

(E and F) Histograms represent an estimate of the posterior probability of burst onsets in the above panels. SDs (in milliseconds) are provided for each peak (horizontal bar, mean ± 1 SD).

(G and H) The SDs of onset estimates varied as a function of the number of trials (n = 376 bursts) (G) and the signal-to-noise ratio of the fluorescence measurements (n = 361 bursts) (H) for single neurons. The region shaded in gray indicates values less than 10 ms.

these simulated positive controls, we were able to detect a significant correlation between motor events and bursting activity (Figure 3F).

Hypothesis Testing of Network Representation

To formally test hypotheses concerning motor representations within HVC, we defined the predictions of two models: a movement model based on GTE events and a uniform distribution model in which bursts are spread evenly throughout the song (Figure 3G). When we smoothly interpolated between these two models with a likelihood-ratio test to consider a range of mixed coding schemes (see Experimental Procedures for details), we found that our data strongly favored a uniform distribution (Figure 3H), supporting the notion of a song-related temporal sequence within HVC. As before, we confirmed the efficacy of this method by finding a strong preference for the GTE model in our simulated positive control (Figure 3I). One limitation of the likelihood-ratio approach is that it tests the GTE model with no temporal offset, which was a choice that was originally motivated by the previous claim that GTE and bursts co-occur (Amador et al., 2013). Although our decision to focus on a zero latency relationship is consistent with our observation of a lack of strong correlations at a variety of time offsets (e.g., Figure 3E), we repeated the log likelihood analysis at a range of possible offsets (−30 to 30 ms). The repeated testing at various negative time offsets can help to counteract any systematic inaccuracy in our estimate of the lag between the true burst time and the calcium onset resulting from our calibration conditions (Figure 4A). The positive time offsets enable us to consider a "causal" scenario in which the accepted premotor delays (~15–20 ms) existing between HVC activity and song (Kozhevnikov and Fee, 2007) are incorporated into the analysis (Figure 4B). Across these various tests, we continued to find strong support for a uniform distribution compared with a subsyllabic kinematic model.

We considered a number of potential sources of error that may have biased our findings. First, our ability to precisely estimate burst onset times for individual neurons was variable (Figures

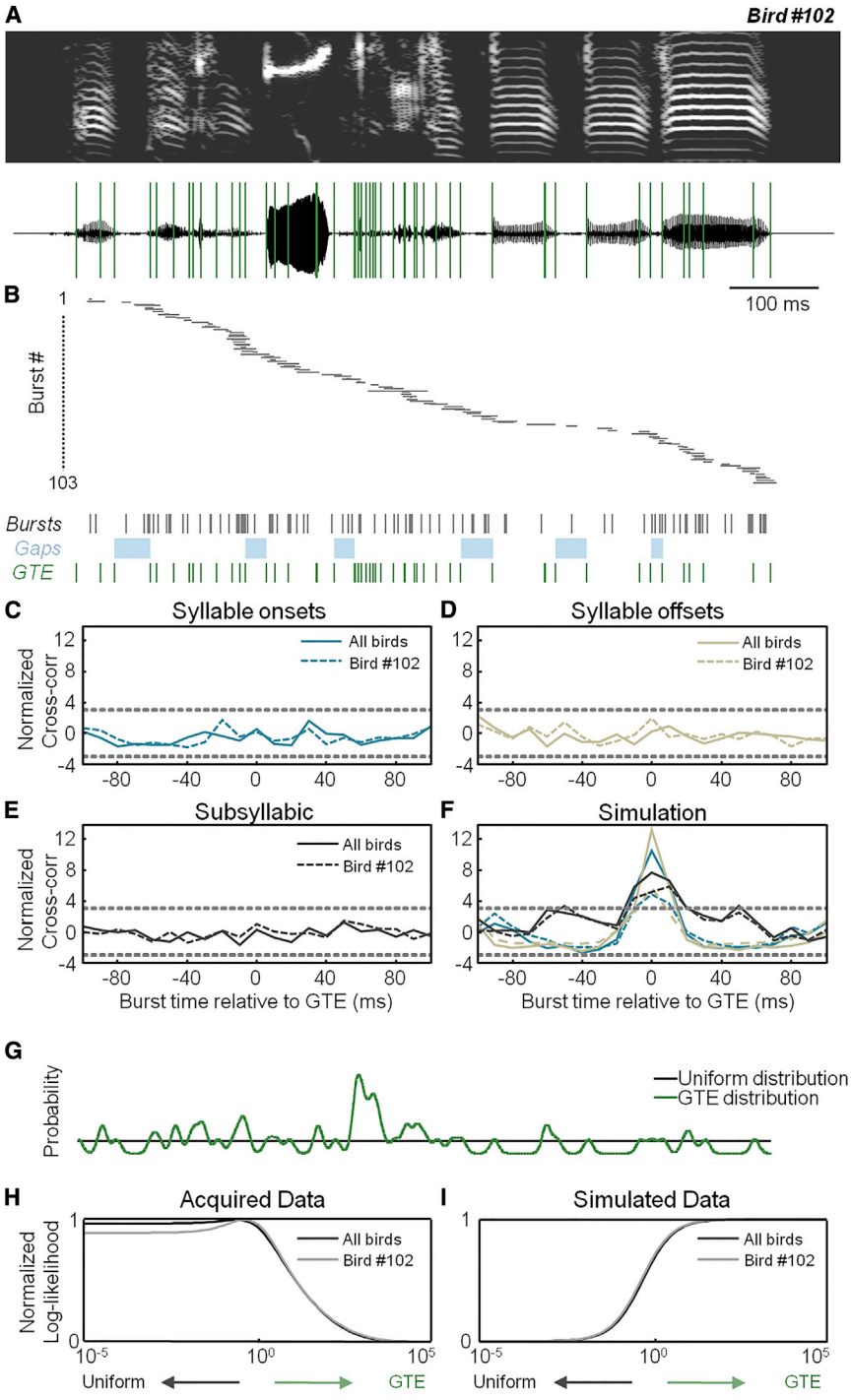

Figure 3. Timing of HVC Network Activity during Singing

(A) The spectrogram (top) and waveform (bottom) of a song motif with 58 GTE overlaid (green lines).

(B) Inferred onset times of 103 burst events from 90 HVC projection cells aligned to the song motif. Horizontal bars represent burst onset times (±1 SD). Below, we show burst onsets (black), silent gaps (light blue), and GTE times (green).

(C–E) Normalized cross-correlation (Experimental Procedures) between burst onset times and syllable onsets (C), syllable offsets (D), and subsyllabic time points (E). Negative offsets mean that the burst precedes the motor event. Dashed lines, ± 3 SD from the null model (Experimental Procedures).

(F) The normalized cross-correlation for simulated data based on syllable onsets, offsets, and subsyllabic time points.

(G) Probabilities of burst times in two alternative models (uniform, black; GTE, green).

(H) The normalized log-likelihood ratio of the predictive model as a function of the relative contributions from the uniform and GTE models, with positive values for R favoring the uniform model. The R value for bird 102 is 23.6, and that for the population of five birds is 84.2, which provides strong support for the uniform model over the GTE model. Note that the value for "c" is plotted on the x axis (Experimental Procedures).

(I) The normalized log-likelihood for simulated data. The large negative values for R (R_{102}, −45.6; R_{All}, −94.7) demonstrate that our positive control simulation strongly supports the GTE model.

2G and 2H). Although our analyses take this variability into account (see Experimental Procedures), it is possible that neurons with the highest amount of uncertainty were obscuring an existing relationship between bursts and GTEs. To address this concern, we restricted our analysis to only bursts with an onset uncertainty of less than 10 ms (186 of 294 events) and again saw no significant relationship between GTE and HVC burst timing (Figures S5A–S5D). Second, zebra finches may generate some muscle movements during silent gaps, such as "mini-breaths" (Hartley and Suthers, 1989), whose relationship with HVC activity is unclear (Andalman et al., 2011). Because we are unable to reconstruct gestures outside of phonation, we repeated the analysis after removing bursts that occurred during gaps and found no significant change in the results (Figures S5E and S5F).

Using Intracellular Recording to Investigate Motor Representation at a Population Level

To provide an additional test of our hypotheses, we next used intracellular recordings to examine the activity of song-related populations of cells within HVC. We first considered the onset of burst responses of individual projection neurons (bird 33, n = 12 cells, 17 bursts; Figures 5A and 5B) to provide an additional validation of our analytical approach in a condition with traditionally less temporal uncertainty than calcium imaging. In our analysis of 28 bursting events in 21 cells from 3 birds, we did not find any evidence of a relationship between burst times and GTE (Figure 5C), despite simulations indicating that we would be able to find a relationship if one existed (Figure 5D).

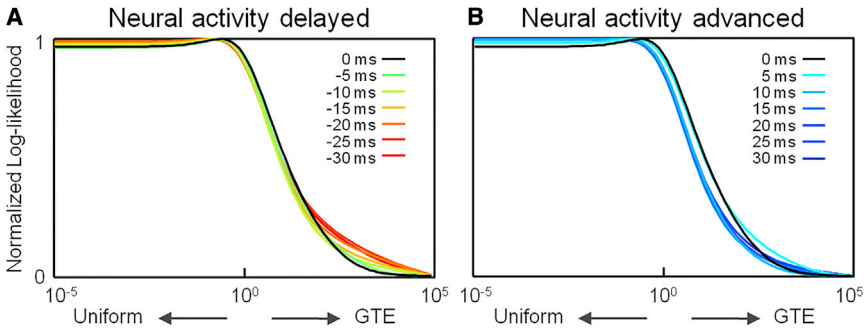

Figure 4. Testing the Temporal Relationship between Neural Activity and Movement

(A and B) Normalized log-likelihood of data acquired across all five imaged birds as a function of the relative strength of the GTE model presented with a series of additional time offsets in which burst times are either delayed (A) or advanced (B) relative to gestural time points. The values for R_{All} across all delays range from 77.4 to 118.1.

We next turned our attention to the subthreshold membrane potential of these HVC neurons. We have recently demonstrated that the series of stereotyped synaptic events onto identified HVC projection neurons is driven by motor-related inputs (Vallentin and Long, 2015). Previous paired recordings in slices of HVC have shown extensive interconnectivity within that nucleus, including links between excitatory neurons (Kosche et al., 2015; Mooney and Prather, 2005). As a result, we reasoned that identified excitatory events onto individual HVC projection neurons (Figure 6B, see example traces) could be used to simultaneously monitor the activity of multiple premotor neurons. We estimated onset times for the most consistently detected postsynaptic potentials (Figures 6A–6C; Experimental Procedures). We then repeated this analysis for all 15 HVC projection neurons within one bird (n = 187 events; Figure 6D) and across a group of three birds (n = 333 events from 26 cells; Figures S6A–S6C). Additionally, there was no connection between the incidence of synaptic onset times and the presence of either a syllable (165.7 events/s) or a gap (164.1 events/s) ($p > 0.05$, paired t test). When considering syllabic and subsyllabic structure, we also found no significant relationship between synaptic events and behavior (Figures 6E–6K; Figures S6D and S6E), including separate analyses where gaps were not included (Figures S6F and S6G) and a range of delays were introduced (Figure S6H). The possibility existed that GTE-related events may be observed in one cell class but not the other (n = 19 $HVC_{(RA)}$ and 7 $HVC_{(X)}$; Figure 6D), but we found no additional relationship when considering each cell type separately (Figure S7). Because of potential errors in our ability to detect synaptic events, we also analyzed the average membrane potential at syllable onsets and offsets as well as GTE times, and we did not find any significant deviation from zero across all neurons tested (n = 33 cells).

DISCUSSION

In our investigation, we focused on the forebrain nucleus HVC of the zebra finch because these neurons represent a well-defined skilled behavior using a sparse and reliable code. We developed a method for investigating the neural mechanisms underlying song production in a head-fixed context, which represents a rare example of a restrained social behavior (Lenschow and Brecht, 2015; Oomura et al., 1983). Using this approach, we provided population-level support for the notion that HVC projection neurons exhibit an abstract representation of elapsed time, suggesting a scheme for encoding behavior that is divorced from the kinematics of ongoing movements (Lu and Ashe, 2015; Matsuzaka et al., 2007; Tanji and Shima, 1994).

Our first step was to determine whether the population data we collected from the imaging (n = 5 birds) and electrophysiology (n = 3 birds) experiments were modulated temporally. Despite observing a large number of motor time points in each bird, obvious gaps in activity were still evident (e.g., Figures S3A and S6C). These nonuniformities may arise from a number of sources. For instance, because our recordings were spatially restricted within HVC, a sampling bias may exist when burst events from nearby cells exhibit similar timing (Graber et al., 2013; Markowitz et al., 2015; Peh et al., 2015). We next tested whether any nonuniformities could be explained by the coincident activity of neurons with motor movements. We greatly expanded on previous efforts to address this issue (Amador et al., 2013; Kozhevnikov and Fee, 2007) by quantitatively examining a number of possible relationships between HVC neuronal activity and singing. First, we find that HVC projection neurons are not preferentially active during syllables compared with silent gaps despite the clear increase in the number and complexity of motor commands during these times (Goller and Cooper, 2004; Riede and Goller, 2010; Suthers et al., 1994; Vicario, 1991a). Second, we added to this analysis by further showing that no structured neural activity existed when considering syllable onsets or offsets separately. Third, we tested whether the timing of HVC projection neuron activity is modulated by subsyllabic motor gestures, as measured by GTE (Amador et al., 2013; Boari et al., 2015).

To analyze the relationship between GTE and neural activity, we used a quantitative approach to demonstrate that the data are better described by a simple uniform model than by a gestural hypothesis. Moreover, this analysis allowed us to test for the possibility of a mixed coding scheme in which a population of neurons represents both kinematics and some other feature independent of the ongoing movement. By examining these intermediate models in our likelihood analysis, we did not see support for the idea of partial coding of GTEs by the population. It should be stated that the poor fit of our data with the GTE model does not necessarily mean that the underlying event times are distributed uniformly but, rather, that any nonuniformities cannot be explained by aspects of motor behavior tested as part of this study. The method for identifying these GTE time points was derived from models of subsyringeal pressure and syringeal muscle activity (Goller and Cooper, 2004; Méndez et al., 2010; Perl et al., 2011). However, future work should directly measure these factors in conjunction with population activity to further confirm our findings.

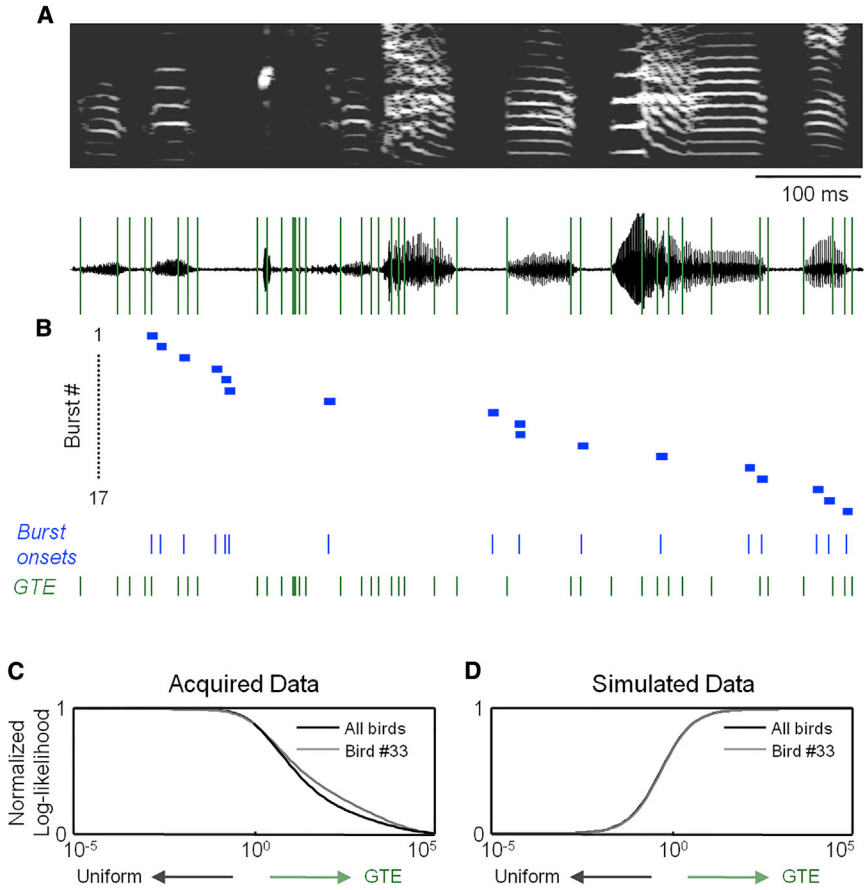

Figure 5. Timing of HVC Bursts during Singing

(A) The spectrogram (top) and waveform (bottom) of a song motif with 45 GTE overlaid (green lines).

(B) Onset times of 17 burst events measured electrophysiologically from 12 HVC projection cells aligned to the song motif. Horizontal bars, depicted with a standard width of 10 ms, are centered on burst onset times. Below, we show burst onsets (blue) and GTE times (green).

(C) The normalized log-likelihood of the acquired data as a function of the relative strength of the GTE model (R_{33}, 32; R_{All}, 49.7).

(D) The normalized log-likelihood for simulated data (R_{33}, −9.0; R_{All}, −17.7).

An important future direction is to understand the processes that enable a transformation between the abstract temporal sequence within HVC and the production of singing behavior. A candidate site that is likely to be central to this conversion is the robust nucleus of the archopallium (RA), which sits between HVC and the motor neurons that coordinate syringeal and respiratory activity (Vicario, 1991b). Each RA neuron receives synaptic input from multiple HVC neurons (Fee et al., 2004; Garst-Orozco et al., 2014), and spiking within this structure covaries with certain song features (Sober et al., 2008). Despite a wealth of single-unit recordings in this region during singing (Leonardo and Fee, 2005; Yu and Margoliash, 1996), an analysis linking RA spiking activity to GTE has not yet been performed.

Sequences similar to those described in this study have been shown to exist across a variety of brain regions, such as the hippocampus (Pastalkova et al., 2008), parietal cortex (Harvey et al., 2012), and basal ganglia (Mello et al., 2015). Within HVC, computational models (Fiete et al., 2010; Jun and Jin, 2007) and recent experimental findings (Okubo et al., 2015) suggest that this kind of population sequence may arise as part of a developmental process. Previous electrophysiological results have shown that the circuitry within HVC (Kosche et al., 2015; Mooney and Prather, 2005; Solis and Perkel, 2005) is likely to play an important role in generating these sequences (Long and Fee, 2008; Vu et al., 1994). Specifically, strong excitatory collaterals within HVC appear to underlie the propagation of network activity through a feedforward synaptic chain (Abeles, 1991; Long et al., 2010). A combination of our high temporal resolution imaging technique with emerging anatomical approaches for circuit reconstruction (Briggman et al., 2011) or electrophysiological methods (Ko et al., 2011) could help to identify mechanisms underlying sequence generation within HVC.

EXPERIMENTAL PROCEDURES

Animals

We used adult (>90 days post-hatching) male and female zebra finches that were obtained from an outside breeder and maintained in a temperature- and humidity-controlled environment with a 12 hr:12 hr light:dark schedule. All animal maintenance and experimental procedures were performed according to the guidelines established by the Institutional Animal Care and Use Committee at the New York University Langone Medical Center.

Song Detection and Reward

We recorded singing behavior with an omni-directional lavalier condenser microphone (AT803, Audio-Technica) and amplified the signal with a solid-state preamplifier (Ultragain Pro MIC2200, Behringer). Song was detected using a digital signal processor (RX8, Tucker-Davis Technologies), and custom software was written using the RPvdsEx interface (Tucker-Davis Technologies) and MATLAB. Because the zebra finch produces a song containing short gaps between syllables and motifs, singing behavior was defined as time periods in which the ratio of high-frequency power to low-frequency power (0–1 kHz and 1–7 kHz, respectively) was greater than 3 for more than 50% of a 1-s sliding window. Singing behavior was only evaluated during 20-s trial periods when the female zebra finch was visible (e.g., Figure 1A). Liquid rewards were administered using a gravity-fed line though a solenoid (NResearch) that was opened for 200 ms to dispense approximately 20 μl of water. To evaluate the detection algorithm, we manually noted false negatives and false positives, enabling us to calculate the sensitivity [true positives/(true positives + false negatives)] and positive predictive value [true positives/(true positives + false positives)].

Surgical Procedures

All surgical procedures were performed under isoflurane anesthesia (1%–3% in oxygen) following established guidelines. Prior to the onset of training, a small (3.9 × 4.55 × 1.3 mm) stainless steel head plate with two threaded holes was affixed with dental acrylic over the inner leaflet of the skull, with the

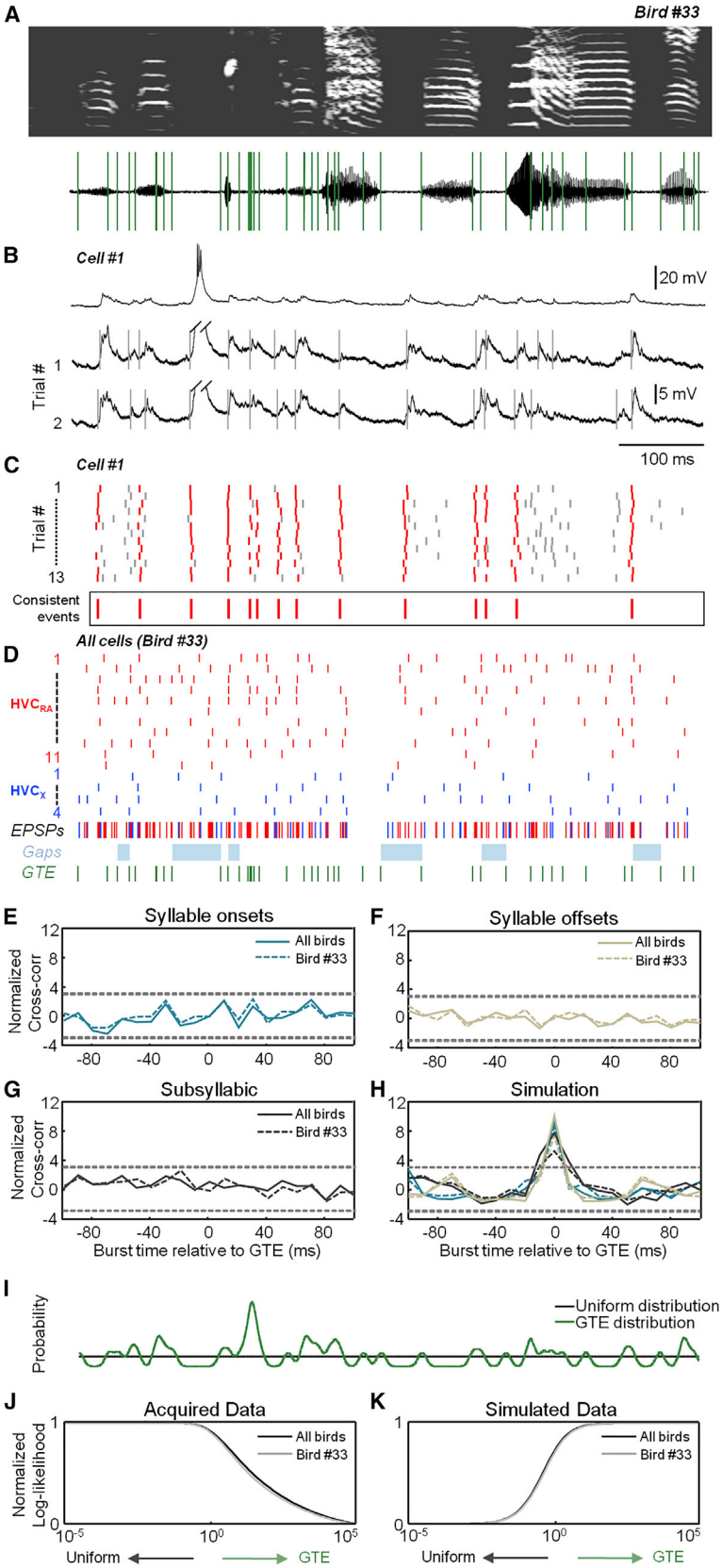

Figure 6. Timing of HVC Synaptic Inputs during Singing

(A and B) A sonogram and GTE time points (A) for the bird shown in Figure 5 and song-related intracellular recordings from an $HVC_{(RA)}$ neuron aligned to the song motif (B). The traces below have been truncated (oblique lines). Vertical gray lines indicate identified synaptic events.

(C) All excitatory postsynaptic potentials (EPSPs) detected across 13 song motifs. Consistently detected events (Experimental Procedures) are indicated with red lines (mean values below).

(D) The consensus synaptic event times for 11 $HVC_{(RA)}$ and 4 $HVC_{(X)}$ neurons (red and blue, respectively) recorded in the same bird. Below, we show EPSP onsets (red or blue, as above), silent gaps (light blue), and GTE times (green).

(E–G) Normalized cross-correlation (Experimental Procedures) between burst onset times and syllable onsets (E), syllable offsets (F), and subsyllabic time points (G). Dashed lines, ±3 SD from the null model.

(H) Normalized cross-correlation for simulated data based on syllable onsets, offsets, and subsyllabic time points.

(I) Probabilities of burst times in two alternative models (uniform, black; GTE, green).

(J) The normalized log-likelihood of the acquired data as a function of the relative strength of the GTE model (R_{33}, 240.9; R_{All}, 493.9).

(K) The normalized log-likelihood for simulated data (R_{33}, −97.8; R_{All}, −157.7).

posterior edge approximately 5 mm rostral to the bifurcation of the sagittal sinus. Cranial window surgery and viral injection were performed when the zebra finch demonstrated the ability to sing reliably under the head-fixed condition (Figure 1; Figure S1). HVC was identified electrophysiologically during surgery using a bipolar stimulating electrode placed into RA (Long et al., 2010). In most birds, we injected AAV9.Syn.GCaMP6s.WPRE.SV40 (Penn Vector Core) into HVC using a beveled pipette (opening diameter, 30 μm; length of bevel, 100 μm). In two birds, we injected a 1:1 mix of AAV9.CamKII0.4.Cre.SV40 and either AAV9.CAG.Flex.GCaMP6f.WPRE.SV40 (bird 192) or AAV9.CAG.Flex.GCaMP6s.WPRE.SV40 (bird 193). We performed three to six injections (30–100 nl/site) separated by approximately 300 μm using an oil-based pressure injection system (Nanoject II, Drummond Scientific). At the end of the injection procedure, we sealed a 3-mm-diameter circular coverglass (#1 thickness, Warner Instruments) with Kwik-Sil adhesive (WPI) and fixed the edge of the glass with cyanoacrylate. We also cemented a black plastic ring (inner diameter, 5 mm; outer diameter, 7.5 mm) around the cranial window to prevent light contamination during imaging.

Two-Photon Imaging

We used a customized movable objective microscope (Sutter Instrument) to scan our field of view (frame rate, 28.8 Hz unless noted otherwise) with a resonant system (Thorlabs) and ScanImage 4.2 software (Pologruto et al., 2003). All imaging was done using a 16× water immersion objective (Nikon) with a numerical aperture of 0.8 and a working distance of 3 mm. To protect the microscope from external light contamination, we wrapped it with a light-attenuating material. Additionally, we fit a black balloon to the tip of the objective on one end and the black ring surrounding the optical window on the other (Dombeck et al., 2010). The excitation source was a mode-locked Ti:sapphire laser (Chameleon, Coherent) tuned at 920 nm and controlled by a Pockels cell (Conoptics 302RM). Fluorescent light was detected using a GaAsP photomultiplier tube (H10770PA-40 PMT module, Hamamatsu) and a wide detection path (2-in collection lens).

Two-Photon Targeted Electrophysiological Recordings

In vivo electrophysiological recordings of HVC neurons expressing GCaMP6s were performed in isoflurane-anesthetized zebra finches. For these birds, we did not affix the cranial window to the brain with Kwik-Sil. Rather, the glass coverslip was sealed at the edges with cyanoacrylate, and a small opening (~300 μm in diameter) was produced near the center of the glass using a carbide burr. Borosilicate glass pipettes were fabricated on a horizontal puller (Sutter Instrument) with impedance values in the range of 4–5 MΩ and filled with an solution of K-gluconate (150 mM) and 5 μM of Red Alexa 594 (Molecular Probes, Invitrogen) to visualize the pipette under the microscope. Because HVC projection neurons produce only infrequent bursts outside of the context of singing (Hahnloser et al., 2002; Long et al., 2010), we applied 0.1 mM of the $GABA_A$ receptor antagonist gabazine (Sigma) to the surface of the craniotomy to increase the rate of spontaneous bursting activity (Mooney and Prather, 2005). Signals were recorded using a Neurodata IR183A single-channel amplifier (Cygnus Technology) and custom MATLAB acquisition software. Data were low pass-filtered at 5 kHz and digitized with a National Instruments digital-to-analog converter (acquisition rate, 40 kHz). For these experiments, the soma was scanned at 52 Hz, and these data were aligned with spiking activity.

Intracellular Recordings

Intracellular recordings during singing were carried out as described previously (Vallentin and Long, 2015). Briefly, a motorized intracellular microdrive was installed on the head of the zebra finch. For antidromic identification of projection neurons, we implanted a bipolar stimulating electrode into the RA and/or area X. Sharp electrodes with an impedance of 70–130 MΩ were backfilled with 3 M potassium acetate and inserted into HVC. Acceptable recordings were defined as having a spike height of more than 40 mV, a resting membrane potential more hyperpolarized than −50 mV, and a total recording duration of more than 3 min. When stable recordings were achieved, a female bird zebra finch was presented to elicit directed singing. Neurons with fewer than two song motifs were excluded from our analysis.

Additional experimental procedures related to analytical methods can be found in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.02.016.

AUTHOR CONTRIBUTIONS

M.P., D.O., and M.L. designed the experiments. M.L., M.P., J.M., L.P., and K.K. wrote the paper. M.P., D.V., S.B., R.C., and K.K. collected the data. M.P., E.P., J.M., and K.K. analyzed the data. J.M., E.P., L.P., and K.K. created relevant computational tools for analysis.

ACKNOWLEDGMENTS

This research was supported by the NIH (R01NS075044), the New York Stem Cell Foundation, the Rita Allen Foundation, the Esther A. and Joseph Klingenstein Foundation, the Simons Foundation (Global Brain Initiative), DARPA N66001-15-C-4032 (SIMPLEX), and EMBO (ALTF 1608-2013). We thank Aimee Chow, Celine Cammarata, and Brandon Robinson for technical assistance and help with further development of our training protocol. We thank Florin Albeanu and Simon Peron for assistance with two-photon imaging and Jeff Gauthier and Arnaud Malvache for help with analysis. We thank Florin Albeanu, Dmitriy Aronov, Brenton Cooper, Kishore Kuchibhotla, and John Long for comments on earlier versions of this manuscript. We thank Gabriel Mindlin and Marcos Trevisan for assistance with identifying GTE time points. We acknowledge the GENIE Program and the Janelia Farm Research Campus; specifically, Vivek Jayaraman, Ph.D.; Rex A. Kerr, Ph.D.; Douglas S. Kim, Ph.D.; Loren L. Looger, Ph.D.; and Karel Svoboda, Ph.D. from the GENIE Project, Janelia Farm Research Campus, Howard Hughes Medical Institute.

Received: December 5, 2015
Revised: January 14, 2016
Accepted: February 4, 2016
Published: May 18, 2016

REFERENCES

Abeles, M. (1991). Corticonics: Neural Circuits of the Cerebral Cortex (Cambridge University Press).

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.

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.

Boari, S., Sanz Perl, Y., 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.

Briggman, K.L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. Nature *471*, 183–188.

Carpenter, A.F., Georgopoulos, A.P., and Pellizzer, G. (1999). Motor cortical encoding of serial order in a context-recall task. Science *283*, 1752–1757.

Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

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.

Day, N.F., Terleski, K.L., Nykamp, D.Q., and Nick, T.A. (2013). Directed functional connectivity matures with motor learning in a cortical pattern generator. J. Neurophysiol. *109*, 913–923.

Denk, W., Strickler, J.H., and Webb, W.W. (1990). Two-photon laser scanning fluorescence microscopy. Science *248*, 73–76.

Dombeck, D.A., Harvey, C.D., Tian, L., Looger, L.L., and Tank, D.W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nat. Neurosci. *13*, 1433–1440.

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

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

Fiete, I.R., Senn, W., Wang, C.Z., and Hahnloser, R.H. (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron *65*, 563–576.

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., Schwartz, A.B., and Kettner, R.E. (1986). Neuronal population coding of movement direction. Science *233*, 1416–1419.

Goller, F., and Cooper, B.G. (2004). Peripheral motor dynamics of song production in the zebra finch. Ann. N Y Acad. Sci. *1016*, 130–152.

Graber, M.H., Helmchen, F., and Hahnloser, R.H. (2013). Activity in a premotor cortical nucleus of zebra finches is locally organized and exhibits auditory selectivity in neurons but not in glia. PLoS ONE *8*, e81177.

Hahnloser, R.H., 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.

Hartley, R.S., and Suthers, R.A. (1989). Airflow and pressure during canary song: direct evidence for mini-breaths. J. Comp. Physiol. A *165*, 15–26.

Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature *484*, 62–68.

Huber, D., Gutnisky, D.A., Peron, S., O'Connor, D.H., Wiegert, J.S., Tian, L., Oertner, T.G., Looger, L.L., and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. Nature *484*, 473–478.

Jun, J.K., and Jin, D.Z. (2007). Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. PLoS ONE *2*, e723.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjöström, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87–91.

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.

Lenschow, C., and Brecht, M. (2015). Barrel cortex membrane potential dynamics in social touch. Neuron *85*, 718–725.

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

Li, N., Chen, T.W., Guo, Z.V., Gerfen, C.R., and Svoboda, K. (2015). A motor cortex circuit for motor planning and movement. Nature *519*, 51–56.

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.

Lu, X., and Ashe, J. (2005). Anticipatory activity in primary motor cortex codes memorized movement sequences. Neuron *45*, 967–973.

Lu, X., and Ashe, J. (2015). Dynamic reorganization of neural activity in motor cortex during new sequence production. Eur. J. Neurosci. *42*, 2172–2178.

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.

Matsuzaka, Y., Picard, N., and Strick, P.L. (2007). Skill representation in the primary motor cortex after long-term practice. J. Neurophysiol. *97*, 1819–1832.

Mello, G.B., Soares, S., and Paton, J.J. (2015). A scalable population code for time in the striatum. Curr. Biol. *25*, 1113–1122.

Méndez, J.M., Dall'Asén, A.G., Cooper, B.G., and Goller, F. (2010). Acquisition of an acoustic template leads to refinement of song motor gestures. J. Neurophysiol. *104*, 984–993.

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.

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.

Nachev, P., Coulthard, E., Jäger, H.R., Kennard, C., and Husain, M. (2008). Enantiomorphic normalization of focally lesioned brains. Neuroimage *39*, 1215–1226.

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., 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.

Oomura, Y., Yoshimatsu, H., and Aou, S. (1983). Medial preoptic and hypothalamic neuronal activity during sexual behavior of the male monkey. Brain Res. *266*, 340–343.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsáki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. Science *321*, 1322–1327.

Peh, W.Y., Roberts, T.F., and Mooney, R. (2015). Imaging auditory representations of song and syllables in populations of sensorimotor neurons essential to vocal communication. J. Neurosci. *35*, 5589–5605.

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.

Peters, A.J., Chen, S.X., and Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity during motor learning. Nature *510*, 263–267.

Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online *2*, 13.

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D., Cushman, J., Aghajan, Z.M., and Mehta, M.R. (2013). Multisensory control of hippocampal spatiotemporal selectivity. Science *340*, 1342–1346.

Riede, T., and Goller, F. (2010). Functional morphology of the sound-generating labia in the syrinx of two songbird species. J. Anat. *216*, 23–36.

Sober, S.J., Wohlgemuth, M.J., and Brainard, M.S. (2008). Central contributions to acoustic variation in birdsong. J. Neurosci. *28*, 10370–10379.

Solis, M.M., and Perkel, D.J. (2005). Rhythmic activity in a forebrain vocal control nucleus in vitro. J. Neurosci. *25*, 2811–2822.

Suthers, R.A., Goller, F., and Hartley, R.S. (1994). Motor dynamics of song production by mimic thrushes. J. Neurobiol. *25*, 917–936.

Tanji, J., and Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. Nature *371*, 413–416.

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

Troyer, T.W. (2013). Neuroscience: The units of a song. Nature *495*, 56–57.

Vallentin, D., and Long, M.A. (2015). Motor origin of precise synaptic inputs onto forebrain neurons driving a skilled behavior. J. Neurosci. *35*, 299–307.

Vicario, D.S. (1991a). Contributions of syringeal muscles to respiration and vocalization in the zebra finch. J. Neurobiol. *22*, 63–73.

Vicario, D.S. (1991b). Organization of the zebra finch song control system: II. Functional organization of outputs from nucleus Robustus archistriatalis. J. Comp. Neurol. *309*, 486–494.

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.

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

Neuron, Volume *90*

Supplemental Information

Population-Level Representation of a Temporal Sequence Underlying Song Production in the Zebra Finch

Michel A. Picardo, Josh Merel, Kalman A. Katlowitz, Daniela Vallentin, Daniel E. Okobi, Sam E. Benezra, Rachel C. Clary, Eftychios A. Pnevmatikakis, Liam Paninski, and Michael A. Long

Figure S1

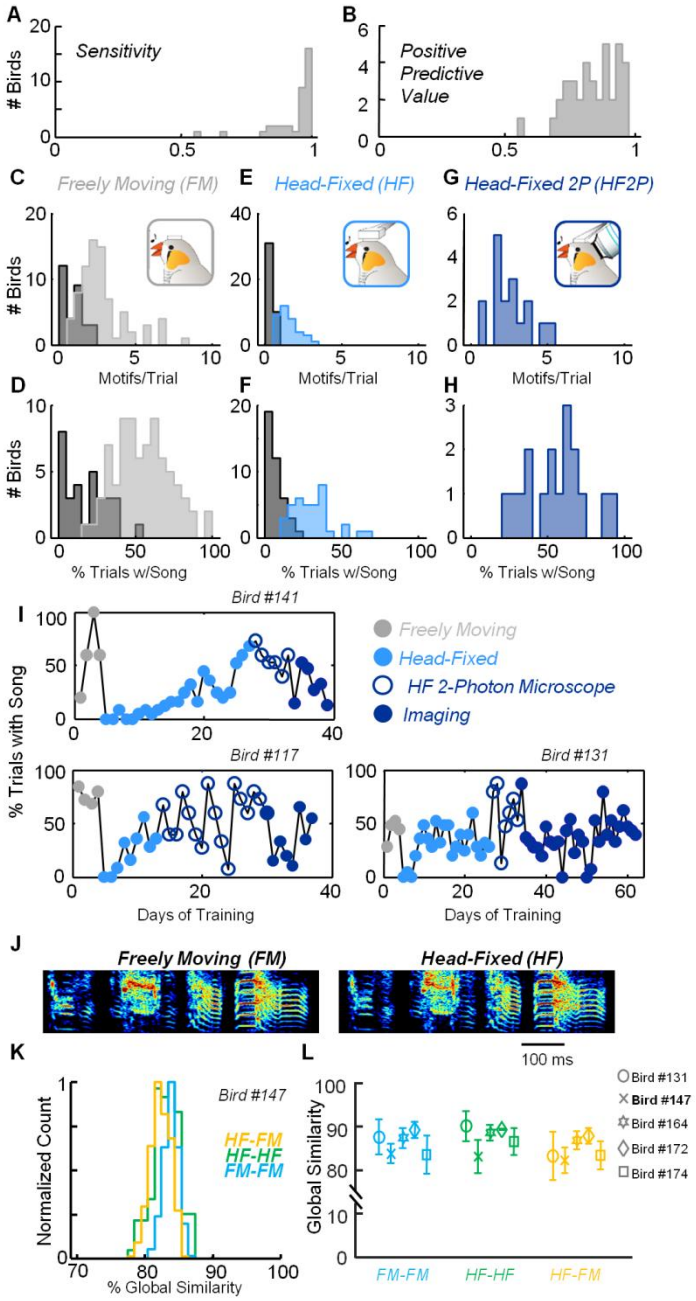

Figure S1. Singing performance in a multi-stage training protocol (Related to Figure 1)
(A,B) Histogram of the sensitivity (proportion of song rewarded: 0.94±0.09) is shown in (A) and positive predictive value (proportion of rewards triggered by song: 0.85±0.09) is shown in (B) for a population of 37 birds that had received a minimum of 100 rewards in the head-fixed configuration. During training, birds progressed through three different phases as they were being shaped to sing under the 2-photon microscope (see Supplementary Table).
(C,D) In the freely moving (FM) condition, the number of motifs per trial in (C) and the percentage of trials with song in (D) for those that passed this stage (light gray) and those who failed (dark gray).

(E,F) In the head-fixed (HF) condition, the number of motifs per trial in (E) and the percentage of trials with song in (F) for those that passed this stage (light blue) and those who failed (dark gray).
(G,H) In the head-fixed 2-photon (HF2P) condition, the number of motifs per trial in (G) and the percentage of trials with song in (H) (dark blue).
(I) The percentage of trials with song for each day is provided for three birds (in addition to those shown in Figure 1C) that completed all phases of training.
(J) Sonograms from Bird #147, which produced similar directed song motifs in the freely moving (FM) and head-fixed (HF) conditions.
(K) Histograms displaying the Global Similarity (see Experimental Procedures) of all song motifs from Bird #147 within and across conditions.
(L) A similar comparison of Global Similarity as shown in (K) for five birds. Birds produced an average of 27.6 and 23.8 motifs in the FM and HF conditions respectively. In our population, Global Similarity measurements (HF-FM: 84.9 ± 2.5; FM-FM: 86.4 ± 2.5; HF-HF: 87.7 ± 2.8) were similar across conditions. All values represent mean +/- SD.

Figure S2

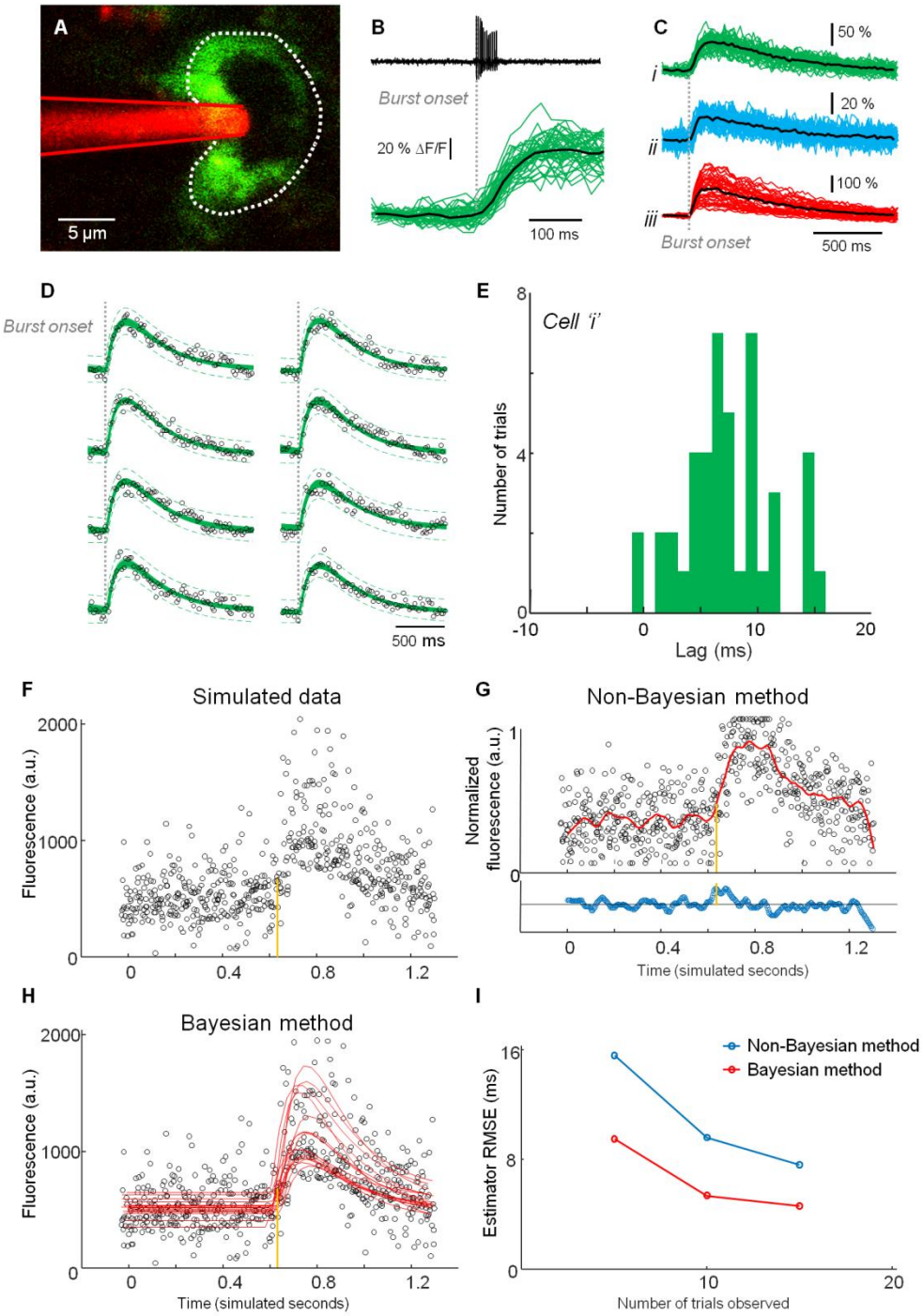

Figure S2. Using calcium indicators and Markov Chain Monte Carlo sampling to estimate burst onset times (Related to Figures 1-4)

(A) An *in vivo* image of the soma of a GCaMP6s-expressing HVC neuron (in green) and a glass pipette (in red).

(B) Fluorescence traces (n = 44 sweeps) from one neuron aligned to the onset of burst events (example shown above). The black line represents an average trace.
(C) The burst-related calcium transients for three HVC projection neurons are shown (aligned to the burst onset). Cell '*i*' is the same neuron featured in (B).
(D) Plot of representative samples from the posterior distribution for 8 individual traces (cell "*i*"). Circles represent the fluorescence measurements for each trial, and the green traces (n = 250) indicate the modeled curve with dashed error curves representing the predictive interval. For further explanation, see Experimental Procedures.
(E) Histogram representing the lag between the first spike in the burst and the inferred onset of 44 bursts for cell *i*.
(F) Simulated fluorescence data of 15 trials with a 30 Hz sampling rate. A joint onset was used in all cases (represented by vertical line).
(G) In a non-Bayesian approach, range normalized data are pooled, and then smoothed (red line). Below, events are detected as increases in the derivative of smoothed trace above a 6SD threshold (red dotted line).
(H) In the Bayesian approach, per trial inference is performed jointly, and onset is inferred by integrating information across trials.
(I) A comparison between the root mean square error of burst onset estimates of the two methods relative to the true onset. While both methods show improvement with more trials, the Bayesian estimator dominates the simpler approach. Moreover, the Bayesian method quantifies the uncertainty in the estimated burst onset time.

Figure S3

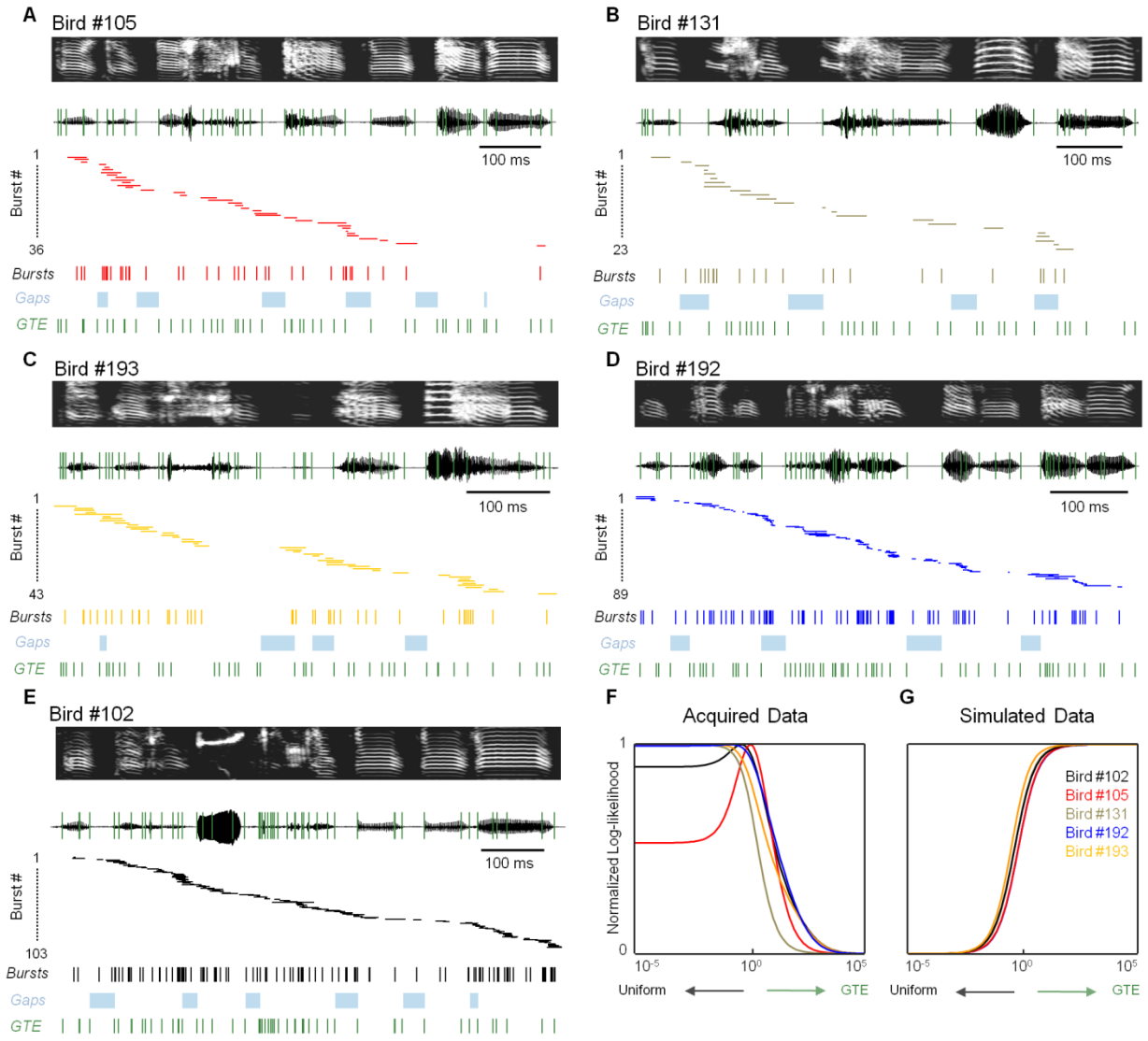

Figure S3. Timing of HVC bursts and GTE during singing for individual birds (Related to Figure 3)
(A-E) The spectrogram (top) and waveform (below) of the song motif from Bird #105 in (A), Bird #131 in (B), Bird #193 in (C), Bird #192 in (D), and Bird #102 in (E) with GTE overlaid as green vertical lines. The middle panel shows the inferred burst onset times for each bird. Each row has a single burst, represented by a horizontal line whose width indicates the uncertainty in the onset estimation. Below, we show burst onsets, silent gaps (light blue), and GTE times (green).
(F,G) The normalized log-likelihood of the acquired data in (F) as well as the simulated GTE-based distribution in (G) as a function of the relative strength of the GTE distribution. The magnitude of the effect can be quantified by the value R, which is equal to LL(c=0)-LL(c=inf). Acquired: R_{102} = 23.6, R_{105} = 1.3, R_{131} = 4.0, R_{192} = 42.7, R_{193} = 12.5. Simulated: R_{102} = -45.6, R_{105} = -9.3, R_{131} = -5.8, R_{192} = -27.6, R_{193} = -6.2.

Figure S4

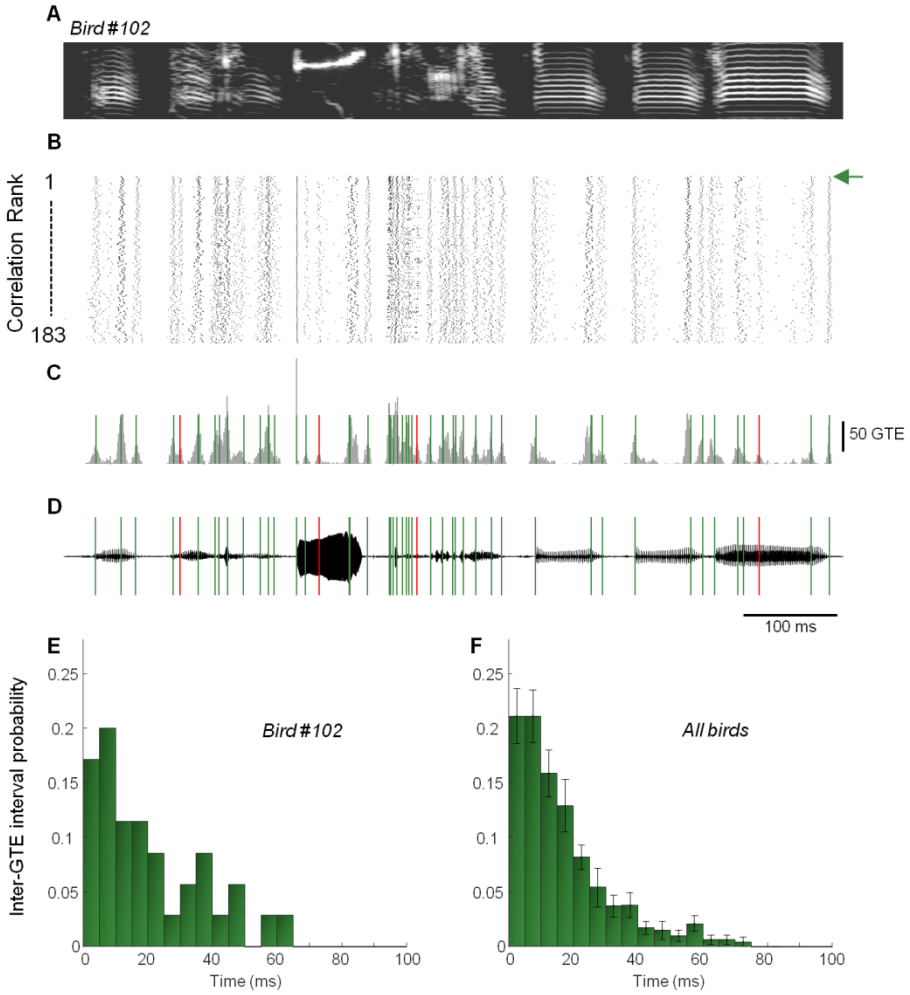

Figure S4. GTE detection algorithm (Related to Figures 3-6)
(A) Example sonogram from Bird #102.
(B) Dot raster of detected GTE for 183 individual motifs. Each row represents all GTE detected on a single motif. Motifs are ranked according to their correlation with the aggregate data in (C) (see Experimental Procedures).
(C) Histogram representing timing of GTE across all trials (bin size: 1ms). Green lines represent the GTE of the top ranked motif (n=54 events) and red lines indicate 4 additional GTE added to that trial based on the underlying distribution (see Experimental Procedures).
(D) Waveform of song motifs form bird #102 with GTE overlaid.
(E) Probability density function of inter-GTE intervals for Bird #102.
(F) Mean distribution of GTE intervals across 8 birds (Error bars = SEM).

Figure S5

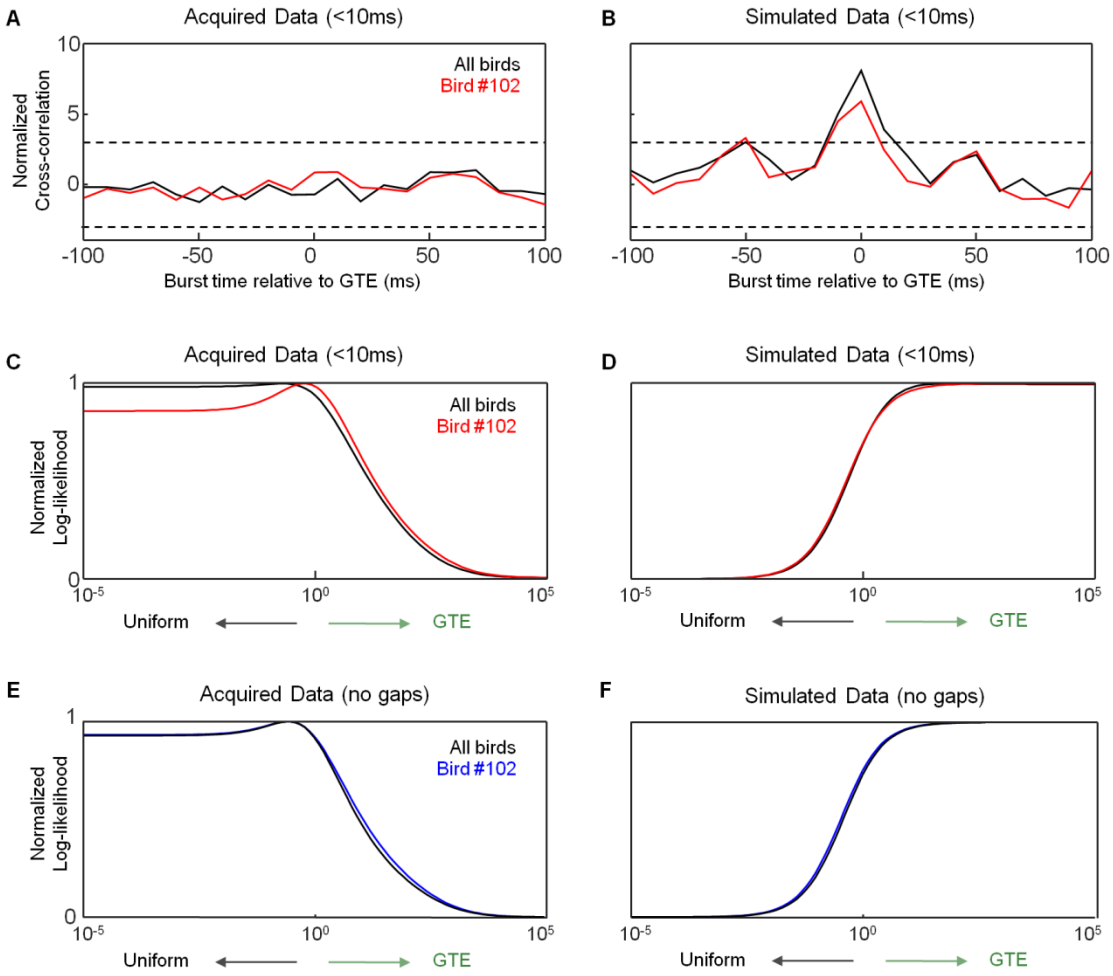

Figure S5. Further analysis of the relationship between HVC neural events and GTE (Related to Figure 3)

(A-D) We restrict our analyses to neurons with a burst onset uncertainty of less than 10 ms (63.2% of the bursts, Figures 2G,and 2H). Cross-correlation analysis of GTE with the acquired data in (A) as well as the simulated data in (B) for Bird #102 (red line) and for our entire population (black line). The normalized log-likelihood of the acquired data in (C) as well as the simulated data in (D) as a function of the relative strength of the GTE distribution. Acquired data: R_{102} = 18.7, R_{All} = 78.8. Simulated data: R_{102} = -34.2, R_{All} = -67.7.

(E-F) We restricted the log-likelihood analysis to events happening only during syllable production (after gaps have been excluded from our analysis). Shown here are normalized log-likelihood analyses of the acquired data in (E) as well as the data simulated from GTE in (F). Acquired: R_{102} = 21.9, , R_{All} = 37.9. Simulated (GTE): R_{102} = -30.9, R_{All} = -60.3. We also quantified maxima points where c>0 as $\hat{c}$ = $\text{argmax}_c(LL)$; M = LL($\hat{c}$)-LL(0). Acquired: $M(\hat{c})_{102}$ = 1.6(0.2), $M(\hat{c})_{All}$ = 2.9 (0.3).

Figure S6

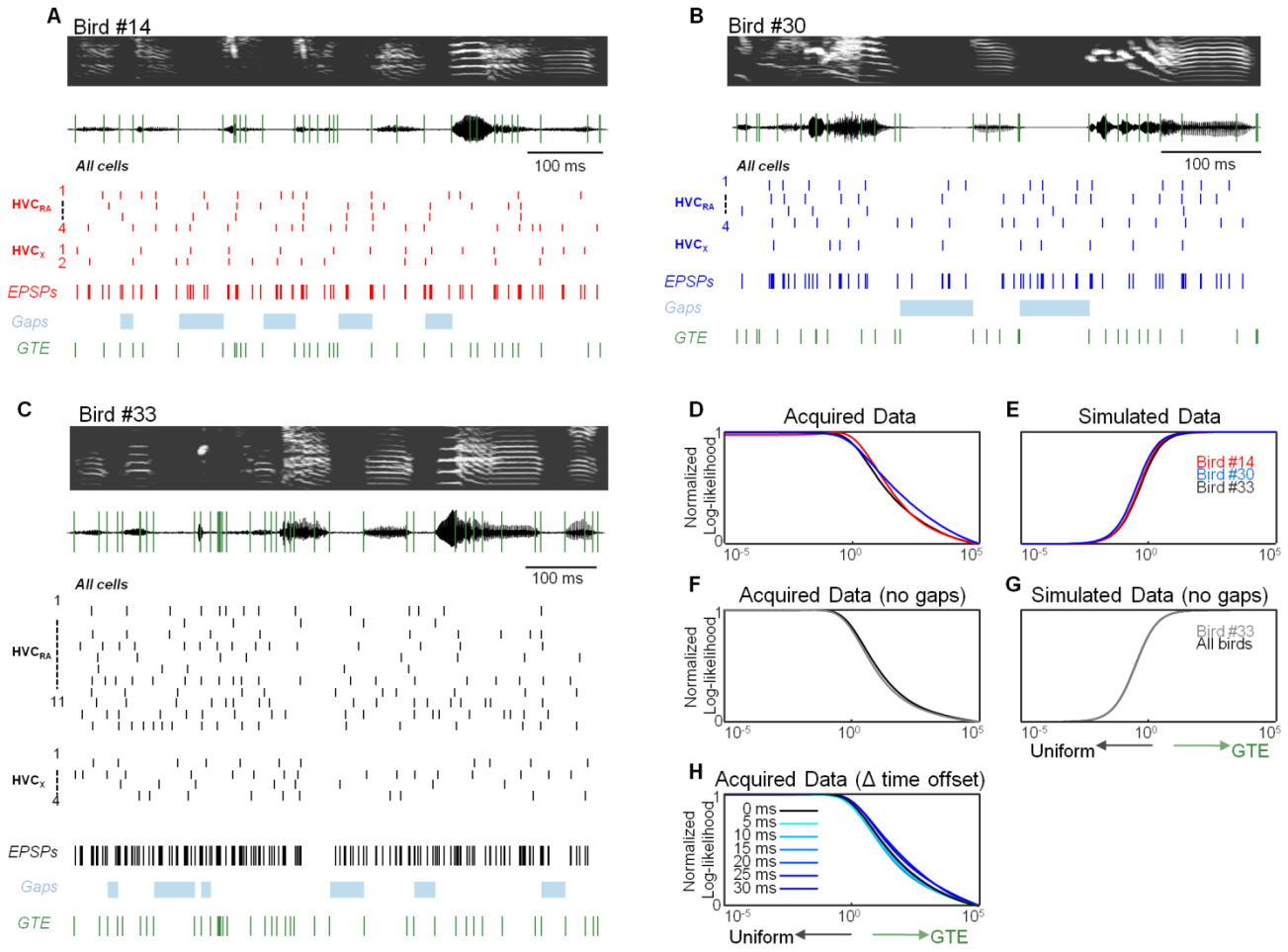

Figure S6. Timing of HVC synaptic inputs and GTE during singing (Related to Figure 6)
(A-C) The spectrogram (top) and waveform (below) of an example motif from Bird #14 in (A), Bird #30 in (B) and Bird #33 in (C) with GTE overlaid as green vertical lines. The middle panel represents the inferred EPSPs onset times for each bird with each row showing all consistently identified EPSPs from individual HVC projection neurons. Below, we show EPSP onsets, silent gaps (light blue), and GTE times (green).
(D-E) The normalized log-likelihood of the acquired data in (D) as well as the simulated GTE-based distribution in (E) as a function of the relative strength of the GTE distribution. Acquired: R_{14} = 109.5, R_{30} = 143.5, R_{33} = 240.9. Simulated: R_{14} = -33.9, R_{30} = -26.1, R_{33} = -97.8.
(F-G) We restricted the log-likelihood analysis to events happening only during syllable production (following Figure S5). Shown here are normalized log-likelihood analyses of the acquired data in (F) as well as the data simulated from GTE in (G). Acquired: R_{33} = 125.3, R_{All} = 216.3. Simulated (GTE): R_{33} = -67.2, R_{All} = -104.6. Acquired: $M(\hat{c})_{33}$ = 0.0(0.0), $M(\hat{c})_{All}$ = 0.4(0.0).
(H) Normalized log-likelihood of data acquired across three birds measured electrophysiologically as a function of the relative strength of the GTE model presented with a series of additional time offsets in which burst times are advanced (0-30 ms) relative to gestural timepoints. The values for R_{All} across all delays range from 449.5 to 546.7.

Figure S7

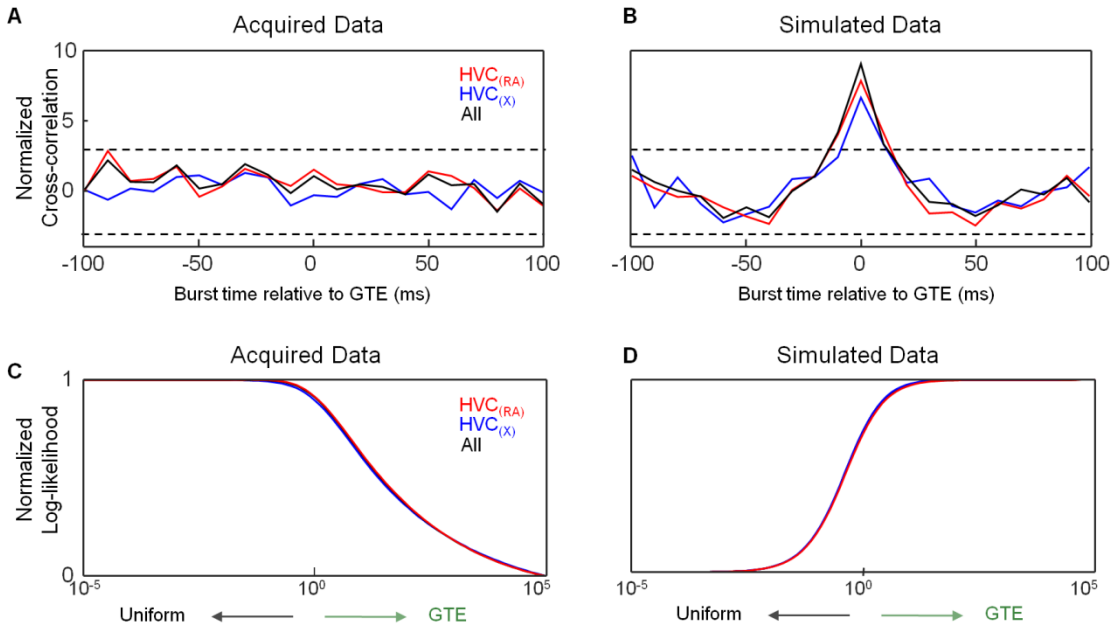

Figure S7. Timing of synaptic inputs relative to GTE in identified HVC projection neurons (Related to Figure 6)
(A,B), Cross-correlations between GTE and EPSP onset times on the acquired data in A as well as simulated data in B for intracellular traces (HVC_{RA} in red; HVC_X in blue; aggregate data in black). (C,D) The normalized log-likelihood of the acquired data (C) as well as the simulated GTE-based distribution (D) as a function of the relative strength of the GTE distribution. Acquired: R_{RA} = 364.6, R_X = 129.3, R_{All} = 493.9. Simulated: R_{RA} = -120.1, R_X = -50.8, R_{All} = -156.5.

EXPERIMENTAL PROCEDURES

Data Analysis

Data analysis: Song Similarity. We mounted a miniature microphone (FG-23629-P16, Knowles) onto the headplate (Figure S1J-S1L) to determine the extent to which head fixation affected song performance. Signals were amplified using a single-channel audio amplifier (Studio Channel, PreSonus Audio Electronics). The length of the song was calculated by setting an amplitude threshold. Similarity measurements were calculated using Sound Analysis Pro (SAP2011) based on five acoustic features: Pitch, goodness of pitch, amplitude modulation, frequency modulation, and Wiener entropy ($p = 0.05$, interval = 70 ms; and min. duration = 15 ms; asymmetric; applied to the entire motif) (Tchernichovski et al., 2000). Using one motif as the reference, three scores (on a scale of 0-100) were obtained for each pair of motifs: % similarity (large time scale), accuracy (small time scale), and sequential match (metric of the corresponding subsections in reference to the song model) (Tchernichovski et al., 2000). The Global Similarity Score (GSS) was defined as the product of the percent similarity, accuracy, and sequential match of the asymmetric measurement. The final score for each pair of sound files was calculated by averaging the two values generated by pairwise comparisons.

Data analysis: GTE Detection. GTE were algorithmically calculated for each song based on a previous approach (Amador et al., 2013; Boari et al., 2015) with the helpful guidance of Gabriel Mindlin and Marcos Trevisan. We first generated a distribution of these times for each bird by combining GTE across motifs (bin size of 1 ms). This joint distribution consisted of actual GTE times as well as a number of other false positive points that were detected by the algorithm. Additionally, any misalignment of motifs would contribute uncertainty to our estimate if we used the data from the entire distribution. To solve this problem, we settled on an iterative process to find a canonical GTE representation. The GTE for individual trials were first convolved with a smoothing filter whose shape was derived from the width of the GTE population responses (Gaussian fit: σ=2.6ms). We then compared all GTE for single motifs against the population of GTE times, and all trials were ordered based on their correlation to the population distribution (e.g. Figure S4). The canonical example was then manually corrected based on the underlying distribution by taking the median of any missed peaks or removing time points that were not suspected to indicate an important song landmark. Syllable onsets and offsets were defined as the first and last GTE in each syllable.

Data analysis: Two-photon Imaging. For 2-photon imaging experiments in the singing bird, we used a multichannel analog-to-digital converter (Digidata 1550, Molecular Devices) to synchronize the microphone signal with the onset/offset of each acquired frame. We segmented and labeled all syllables using custom MATLAB software. We next captured all song related imaging frames and defined the scan time of each frame as the moment that its central pixel was recorded. Given the slight variability in song timing across trials (Glaze and Troyer, 2006), it was necessary to linearly warp all song motifs to a uniform duration. For each motif, we applied an identical warp factor to the corresponding scan times. We then motion-corrected the acquired images within and across motifs by finding the center of mass of the correlations across frames relative to a set of reference frames (Miri et al., 2011). ROIs (regions of interest) were manually drawn and the fluorescence of each neuron was estimated using freely available software (ImageJ). Burst onsets were detected using a Markov Chain Monte Carlo (MCMC) inference method described below. Briefly, four MCMC chains were run in parallel, each for 4000 sweeps over all parameters (with a burn-in of 1,500 sweeps). The burst onset time was defined as the mean of the median burst time computed within each parallel chain. The uncertainty was defined as the median absolute deviations of each parallel chain, multiplied by the standard 1.48 correction factor (Leys et al., 2013) to enable comparisons with the standard deviation estimates considered below.

Data analysis: Simultaneous Extracellular Recording and Imaging. All juxtacellular data were high-pass filtered by removing the baseline (smoothed, boxcar window of 12.5 ms). Bursts were defined as events with at least three spikes occurring at a high frequency (>100 Hz). In order to avoid contamination from previous calcium transients, we discarded bursts happening more than 1 second after or 1.5 seconds before another burst. In order to determine the latency between the electrophysiological burst activity and the fluorescence response, we calculated the delay between the peak of the first spike in the burst and the corresponding onset in the fluorescent signal, as measured using a Markov Chain Monte Carlo (MCMC) inference method described below.

Data analysis: Burst Detection. For the electrophysiological experiments, song motifs were aligned in a similar manner to the 2-photon imaging data analysis. Bursts were defined as events with at least two spikes occurring at a high frequency (>200 Hz), and onsets were defined as the peak of the first spike in the burst. Given that only single trials were acquired for some neurons, and therefore we were unable to calculate a variance measure in those cases, burst time uncertainty was universally set to 3ms across all conditions, which represents an upper bound of burst time variance in HVC (Kozhevnikov and Fee, 2007).

Data analysis: EPSP Detection. Action potentials were removed from the intracellular traces prior to analysis (cutoff 20 mV above baseline). EPSP peaks were detected by finding the maximum of sets of points 2 mV above a moving baseline (1.025 ms Savitzky-Golay filter). Onsets were assigned to each peak, defined as the minimum within 8 ms before the peak. Algorithmic errors of onset detection were manually corrected. Matching sets of EPSPs across motifs were found for each cell by computing the peaks of the distribution of the data aggregated across all recordings (bin size of 0.1 ms, convolved with a 2.5 ms Gaussian window), and then including the EPSP detected within 5ms in each motif. Inclusion or exclusion of EPSP onsets to a set was manually corrected via inspection of the traces. EPSPs were determined to be consistent if they were found in at least half (minimum 2) of the motifs recorded. Final time point and uncertainty of the EPSPs for each cell was taken to be the mean and standard error of the mean of onsets within each set.

Data analysis: Description of methods for burst onset inference. In this section, we describe our approach for pooling information across trials of stereotyped neural activity, observed via calcium imaging, to infer burst onset times. The approach presented here is an augmentation of our continuous-time Bayesian deconvolution method for single calcium traces (Pnevmatikakis et al., 2013), which is similar to a general approach to deconvolution described elsewhere (Tan and Goyal, 2008). For background on the general Bayesian approach to data analysis consult (Gelman et al., 2013). Burst onsets are modeled as instantaneous events that occur at approximately the same time across trials. Each event is marked with an amplitude which corresponds to burst intensity.

Formally, we must state a model that describes how the parameters of interest relate to the observed calcium traces. We specify the generative model as follows:

$$t_{ij} \sim s_j + \sigma_j \varepsilon_{ij}^s \qquad \varepsilon_{ij}^s \sim \mathcal{N}(0,1) \qquad (1)$$

$$y_i(t) = \sum_j a_{ij} k\left(t - t_{ij}\right) + b_i + \gamma_i \, \varepsilon_i^c(t) \qquad \varepsilon_i^c(t) \sim \mathcal{N}(0,1) \qquad (2)$$

For a given cell, we consider a set of N burst onset times $\{s_j\}$ that are the same across trials. t_{ij} is the time of the jth burst in the ith trial, which is perturbed by Gaussian noise from the shared time s_j. a_{ij} is the amplitude of the burst at time t_{ij} and b_i is the baseline calcium for each trial. The observed calcium signal (Y) for each trial and time is $y_i(t)$. The predicted calcium signal generated from the model is C (i.e. $c_i(t)$ for each trial and time). We will use y_i and c_i to denote the discrete-time traces for the ith trial (with time implicitly indexed). $k(t)$ is the causal difference-of-exponentials kernel, $\left(e^{-t/\tau_f} - e^{-t/\tau_r}\right)1(t \geq 0)$, with τ_f and τ_r the time constants for fall and rise respectively, and with the peak kernel amplitude normalized

to one. In this analysis, we only observe the noisy calcium trace (Y) and we infer the posterior distribution over all other parameters. That is, we infer $p(\{t_{ij}\},\{a_{ij}\},\{b_i\},\{\gamma_i\},\{s_j\},\{\sigma_j\},\tau_r,\tau_f|\,Y)$.
Calcium traces may be acquired at different sampling rates and starting at different times relative to some trigger (e.g. trial-aligned). We can infer the burst onset times in a normalized time and then transform the single trial burst times to the relative time of the calcium trace for that trial. That is, if t_1 is the burst time for trial 1 (reference trial) and t_i is the burst time for trial i, we can compute an offset for trial i, $o_i = u_i - u_1$, and a rescaling coefficient $r_i = (\frac{\Delta_i}{\Delta_1})^{-1}$. Whenever we use a burst time to predict the misaligned, discretized calcium traces, we consider events to occur at $(t_{ij} - o_i) \times r_i$ rather than t_{ij}. By construction, trial 1 is the reference trial, so its offset is 0 and its rescaling coefficient is 1. For clarity of notation, we omit these temporal warping terms, but they are implemented in our analysis routine.

Data analysis: Markov chain Monte Carlo (MCMC) inference algorithm. MCMC methods are used to generate samples from the posterior distribution of parameters of interest and thereby approximate the posterior distribution of the parameters. To ensure that the parameter initialization did not affect our final answer, we include a "burn in" by ignoring the first 1500 sweeps. We initialize inference for each cell by manually specifying burst onset times, then adding a random offset to each (range +/-3 bins, bounded at the trial edges). We found that the results were rather independent of the initialization of the event times, and the algorithm can find good solutions across a wide range of initializations – indeed, we inferred the number of bursts from no initial bursts (i.e. uninitialized) - but pre-determining the number of bursts and very crudely initializing them decreased the burn-in time and allowed us to better standardize the number of post-burn-in samples across neurons (data not shown). The baseline parameter was initialized at zero and the initial amplitudes of the bursts were set to the observed calcium trace value of the time bin nearest the burst onset time. We empirically determined that the values of the exponential filter's rise and decay times had a significant impact on burst onset estimation, so these were also randomized at the beginning of each inference at 100 and 300 ms respectively, with a random offset (range +/- 3 bins, minimum of 1 bin.)

Next we describe the details of the sampling algorithm. We use a Metropolis-within-Gibbs approach in which we hold all parameters except one fixed at a time and sample that parameter probabilistically (i.e. update that parameter). In Gibbs sampling, the parameters are sampled directly from conditional distributions. Metropolis sampling involves proposing moves and probabilistically accepting or rejecting them. A sweep of the sampler consists of five types of "moves", with the option for some to be performed multiple times within a sweep. First we list the move types, and then describe how they are performed:

1. Sample single trial burst times - t_{ij}
2. Sample shared burst onset parameters - s_j, σ_j
3. Sample calcium noise parameter - γ_i
4. Sample amplitudes of the single trial bursts and baseline of each trial - a_{ij} and b_i
5. Sample the time constants (separately) - τ_r, τ_f

1-Sample single trial burst times

To sample burst onset times of single trials, we use random-walk Metropolis, following the approach of the calcium deconvolution sampler (Pnevmatikakis et al., 2013). That is, we propose moves of each burst onset time by drawing from a Gaussian density, centered at the current time of the burst. We accept or reject these moves based on both the Gaussian likelihood $p(Y|t_{ij},\dots)$, which indicates how the timing of the burst affects the data, and the Gaussian group prior $p(t|s)$, which encourages the burst times to be shared across trials:

$$p(t_{ij}|s_j,\sigma_j,\gamma_i,a_{ij},b_i,y_i,\tau_r,\tau_f) \propto p(y_i|t_{ij},\dots)p(t_{ij}|s_j,\dots) \qquad (3)$$

$$\propto \mathcal{N}\left(t_{ij} - s_j, \sigma_j^2\right) \prod_{t=1}^{T} \mathcal{N}(y_i(t) - c_i(t), {\gamma_i}^2); \quad (4)$$

Note that c_i above depends implicitly on the t_{ij} times and on the parameters $a_{ij}, b_i, \tau_r, \tau_f$. We also enforce an exclusion prior, a constraint that moves cannot bring two bursts too close together (i.e. for multibursting cells, the single trial burst onset times are never permitted to be sampled less than 1 bin from each other). This prevents bursts from merging by complete overlap.

2-Sample shared burst onset parameters

To sample estimates of the shared burst times and the parameter corresponding to the jitter of the individual trial burst times relative to the shared burst time, we compute the mean and standard deviation of the burst times across trials given the current sample of the set of t_{ij}:

$$\hat{s}_j = mean_i(t_{ij}) \quad (5)$$

$$\hat{\sigma}_j = std_i(t_{ij} - s_j) \quad (6)$$

We then sample the value of σ_j from the appropriate scaled-inverse chi-squared distribution (implemented via an appropriate inverse gamma distribution) which serves as a standard conjugate prior over variance - this distribution depends on the estimate of the variance $\hat{\sigma}_j^2$. We subsequently sample the shared burst times from a normal centered at their current mean and with standard deviation equal to the standard error of the sample mean, $\hat{\sigma}_j/\sqrt{D}$, where D is the number of trials.

$$s_j \sim \mathcal{N}(\hat{s}_j, \hat{\sigma}_j/\sqrt{D}) \quad (7)$$

$$\sigma_j^2 \sim \textit{Inv-Gamma}(\tfrac{D}{2}, \tfrac{D\hat{\sigma}_j^2}{2}). \quad (8)$$

For increased generality, we permit σ_j to be different for each burst j (but note that this parameter is fixed across trials i). It is also straightforward to decrease the variance of the samples of σ_j by having the conjugate prior be informative, but we found this to be unnecessary here. It is also possible to add a move type which samples $\{t_{ij}\}$ and s_j simultaneously, though again we found this unnecessary.

3-Sample calcium noise parameter

Similar to the procedure for sampling σ_j , to sample γ_i we simply estimate the sample variance and sample from the appropriate scaled-inverse chi-squared uninformative prior.

$$\hat{\gamma}_i \sim std(y_i - c_i) \quad (9)$$

$$\gamma_i^2 \sim \textit{Inv-Gamma}(\tfrac{T}{2}, \tfrac{T\hat{\gamma}_i^2}{2}). \quad (10)$$

Similar to the case for sampling σ_j, for increased generality, we can permit γ_i to be different for each trial.

4-Sample amplitudes and baseline parameters

We wish to sample the amplitude and baseline parameters from their appropriate conditional distributions. We constrain both of these parameters to bounded intervals (i.e. positive and less than some large value).

$$p\left(a_{ij} \middle| Y, t_{ij}, b_i, \dots\right) \propto p\left(Y \middle| a_{ij}, t_{ij}, b_i, \dots\right) 1(a_{max} \geq a_{ij} \geq a_{min}) \quad (11)$$

$$p\left(b_i \middle| Y, t_{ij}, a_{ij}, \dots\right) \propto p\left(Y \middle| a_{ij}, t_{ij}, b_i, \dots\right) 1(b_{max} \geq b_i \geq b_{min}) \quad (12)$$

Both of these posteriors are truncated normal distributions and can be sampled various ways (e.g. using random-walk Metropolis steps, or jointly from a truncated multivariate normal distribution, as described previously (Pakman et al., 2014)). We found that a random-walk Metropolis approach sufficed here.

5-Sample the time constants

We also wish to sample the time constants from their appropriate conditional distributions. We constrain both of these parameters to bounded intervals, (i.e. both positive and with the rise time smaller/faster than the fall time).

$$p\left(\tau_r \middle| Y, \tau_f, \dots\right) \propto p\left(Y \middle| \tau_r, \tau_f, \dots\right) 1(\tau_f \geq \tau_r \geq \tau_{min}) \quad (13)$$

$$p(\tau_f|Y, \tau_r, \dots) \propto p(Y|\tau_r, \tau_f, \dots)1(\tau_{max} \geq \tau_f \geq \tau_r) \quad (14)$$

It is also straightforward to incorporate a more informative prior if it is believed (e.g. from other research on the calcium indicator properties) that the time constants are distributed according to a more sharply peaked distribution – in this setting, we did not wish to impose a strong prior.

Algorithm: Full inference scheme

The whole procedure of updating each parameter in sequence is run many times to generate many distinct samples that collectively constitute the estimate of the posterior distribution. Matlab code implementing this algorithm will be made available via the website of one of the authors (https://sites.google.com/site/jsmerel/).

Algorithm 1: MCMC inference routine (describes a sweep)

Input: observed calcium signal y_{ij}, initial values for various parameters
Output: Posterior sample of all parameters.

for $j = 1$ **to** *N* bursts **do**
    for i = 1 **to** *D* trials **do**
        Sample t_{ij} using random-walk Metropolis-within-Gibbs (eq. 4)
    end
for j = 1 **to** *N* bursts **do**
    $\hat{\sigma}_j = std_i(t_{ij})$
    $\hat{s}_j = mean_i(t_{ij})$
    $s_j \sim \mathcal{N}(\hat{s}_j, \hat{\sigma}_j/\sqrt{D})$ (eq. 7)
    $\sigma_j^2 \sim$ *Inv-Gamma*$(\frac{D}{2}, \frac{D\hat{\sigma}_j^2}{2})$ (eq. 8)
End
for j = 1 **to** *N* bursts **do**
    for i = 1 **to** *D* trials **do**
        Sample a_{ij} using random-walk Metropolis-within-Gibbs (eq. 11)
    end
end
for i = 1 **to** *D* trials **do**
    Sample b_i using random-walk Metropolis-within-Gibbs (eq. 12)
End
Sample τ_r, τ_f using random-walk Metropolis-within-Gibbs (eq. 13,14)
for i = 1 **to** *D* trials **do**
    $\hat{\gamma}_i \sim std(y_i - c_i)$ (eq. 9)
    $\gamma_i^2 \sim$ *Inv-Gamma*$(\frac{T}{2}, \frac{T\hat{\gamma}_i^2}{2})$ (eq. 10)
end
return a posterior sample of $\{t_{ij}\}, \{a_{ij}\}, \{b_i\}, \{\gamma_i\}, \{s_j\}, \{\sigma_j\}, \tau_r, \tau_f$

For all sampling steps which use a random walk proposal, we determined the proposal variance parameters using adaptive tuning of the random walk proposal variance – when proposals were accepted or rejected, proposal variance was slightly increased or decreased respectively, with the magnitude of this modification decreasing with each sweep, such that the acceptance ratio ended up being reasonable for each parameter and distinct across neurons (Rosenthal, 2011). Additionally, in practice, mixing of samples across sweeps can be faster if certain moves are performed multiple times per sweep.

In addition to the model presented here, we note that some of the trials had components of the calcium response carry over from a previous trial. To handle this, we simply assumed that time zero of all trials had some initial concentration that decayed back to baseline with a time constant corresponding to the fall-time of the difference-of-exponentials kernel. This added one more parameter that we could sample over each sweep in a manner similar to the burst amplitude updates.

Combining data across sweeps and renditions

To aggregate data for each burst, each distribution was automatically clustered and corrected manually into isolated burst times. To ensure consistency, all sweeps included in the distribution were required to contain the same number of bursts for at least 25 sweeps on either side (to allow for stabilization). Rarely, chains that did not converge were removed from further analysis.

Data analysis: Comparison of MCMC inference algorithm with non-Bayesian approach. We simulated data of a single burst measured on multiple trials (n=5, 10, and 15) sampled at 33.3 Hz, assuming no jitter in burst onsets across trials. In order to capture the complexity of real data, each trial had a different sampling offset, baseline, event amplitude, and noise level. The Bayesian approach followed the protocols described above. For the non-Bayesian approach we first normalized all trials between their extrema. This data was then pooled by re-binning (333.3 Hz) and smoothed with a Loess filter (set at 10% of observations). The derivative of this smoothed curve was taken. Events were determined as crossing of a threshold set at 6 SD. We quantified performance as the RMSE of estimates relative to ground truth as a function of number of trials, based on ~50 repeats (failed instances of the method were manually removed) of all trials at each number of observations.

Data analysis: Cross-correlation on the acquired and simulated data. As a first approach to assess the distribution of burst onsets, we performed an intuitive analysis that does not make many assumptions about the data. For this analysis, we binned the events into 10ms bins and examined the cross-correlations between the GTE, syllable onsets, or syllable offsets and the estimated burst onsets. If the behavioral time-points were informative, we would expect to see a peak in the cross correlation near 0, indicating that they are strongly correlated with the burst onsets.

In this analysis, the null model is determined by randomly reshuffling the inter-burst-onset intervals for a given bird to generate a new sequence of bursts with the same interval statistics (we also tried randomly drawing N burst onsets uniformly, where N is the number of bursts observed for that bird; this produces very similar results). We can determine the expected cross-correlation under this model by correlating the randomly sampled bursts with the GTE. We then compute the mean and standard deviation of these at each lag, providing a null model for the cross-correlation.

Simulated comparisons

We can simulate data from a GTE model where burst onsets occur at the times of the GTE plus some random temporal jitter -- this jitter consists of the variability of bursts around the GTE (a parameter of the GTE model which we denote as σ_{GTE}^2 below) combined with the burst onset inference uncertainty. We draw N bursts by randomly sampling the GTE times (with replacement) and adding normally distributed random jitter with variance = $\sigma_{GTE}^2 + \sigma_i^2$ where 'σ_i' is sampled randomly for each simulated burst from the distribution of burst uncertainties for this bird. As expected, as long as the jitter is not unreasonably large,

a significant peak is evident. When the jitter is too large (e.g. > 20ms), the peak is no longer unambiguous (Supplementary Figs. 9 and 12). This defines the resolution we need to unambiguously test the GTE hypothesis with this method. The burst jitter around the GTE has been reported to exhibit ~5ms standard deviation for singing birds (Amador et al., 2013). This simulation analysis suggests that, given our levels of burst onset uncertainty, we would expect to see a peak in these cross-correlations if the bursts actually clustered around GTE.
For these analyses, we only consider burst onsets which fall within the range of the first and last GTE and we correspondingly consider these events to be the bounds of the null model.

Data analysis: Log-likelihood of data under GTE model with baseline. In addition to the cross-correlation analysis, we also performed a more powerful, model-based hypothesis test. The simplest version of the GTE model is that there is an increased probability of finding a burst onset around a GTE and no probability of finding a burst onset elsewhere. The only parameter in this model is the size of the local region around a GTE that has an increased probability of a burst onset. We model this local region with a Gaussian-shaped bump around the GTE (Amador et al., 2013), so the size of this region is given by the standard deviation of the Gaussian, σ_{GTE}. If the standard deviation gets too large, the GTE model becomes equivalent to uniformly sampling the burst times, because the burst times around each GTE blend together and the model predicts that events are equally likely everywhere within the song. When the standard deviation parameter is not too large (e.g. ~5ms, consistent with previous work (Amador et al., 2013)), there are regions with close to zero probability between GTE events.

We then focused on a more general model, of which an exclusive pairing of GTE and burst times is an extreme case. In the following parameterization of the GTE model, we smoothly interpolate between the model in which burst onsets occur uniformly throughout the song (i.e. the uniform model) and the model where burst onsets are localized around GTE (i.e the strong GTE model). This enables us to evaluate the likelihood of the observed data along the continuum of models ranging from uniformly-distributed bursts all the way to the strong GTE model where bursts completely cluster around GTE. If the baseline term dominates, this is support against the relevance of the GTE; conversely, if the GTE term dominates, then the GTE model is relevant.

In this interpolated model the probability of a burst at a given time t is proportional to $\lambda(t)$, with

$$\lambda(t) = b + a(t) \tag{15}$$

$$a(t) = c \times \frac{1}{J}\sum_{j=1}^{J} N\left(t \middle| g_j, \sigma_{GTE}\right) \tag{16}$$

Here b is the baseline probability of finding a burst onset, c is the magnitude of the GTE event locally, J is the number of GTE events, and the GTE contribute a Gaussian shaped bump (with standard deviation σ_{GTE}) of increased probability of burst onsets centered at each GTE (g_j) .

We also need to account for the uncertainty in the burst onset times, so for the i^{th} burst onset, which we estimated to be at time t_i, we must convolve $\lambda(t)$ with our uncertainty in the inference of that burst (s_i, using $s_i \approx 1.4\, MAD_i$ for the imaging data and $s_i \approx \sigma_i$ for the EPSP data) and then appropriately normalize $\lambda(t)$. Let

$$g_i(t|b, c, \sigma_{GTE}) = \lambda(t|b, c, \sigma_{GTE}) * N(t\,|0, s_i), \tag{17}$$

with * denoting convolution, and

$$f_i(t \mid b, c, \sigma_{GTE}) = \frac{g_i(t \mid b, c, \sigma_{GTE})}{\int g_i(t \mid b, c, \sigma_{GTE})\, dt}. \tag{18}$$

The log-likelihood is then:

$$LL(b, c, \sigma_{GTE}) = \sum_{i=1}^{N} ln\, f_i\,(t_i \mid b, c, \sigma_{GTE}). \tag{19}$$

We can plot the log-likelihood of the observed burst times for many values of c and a few values of σ_{GTE} to characterize the optimal c (without loss of generality, we can hold the baseline b fixed and vary c so as

to change the strength of the GTE contributions relative to the uniform baseline for burst times, since the corresponding distributions are normalized). Notice that as $c \rightarrow 0$, the GTE model becomes the model where burst onsets are drawn uniformly; as c gets larger, there is a stronger localization of bursts around the GTE, and this model becomes the strong GTE model. To indicate the extent of this effect on normalized plots, we provide the difference of the log-likelihood values for the uniform (c=0) and GTE (c=Inf) distributions, and we designate this quantity as R.

We also perform this likelihood analysis with the further restriction that only bursts recorded during syllable production are analyzed (Supplementary Figs. 9 and 11). This means that the normalizing integral in the denominator of eq (18) is computed only over intervals for which syllables are occurring, and the sum in eq (19) is restricted to bursts that occurred during a singing interval. For these plots, we also report the locations of the peaks of the LL plots ($\hat{c} = \text{argmax}_c\{LL\}$), as well as their values with respect to the uniform distribution ($M = LL(\hat{c})-LL(0)$), and we designate these quantities as $M(\hat{c})$.

Simulated comparison

To validate our intuitions about the GTE model and verify that we have enough data for this test to be sufficiently powerful, we can simulate burst onset times from a version of the GTE model (e.g. zero baseline and various levels of jitter) and see how the log-likelihood depends on c (for a number of burst onsets equivalent to what we have observed). We sampled N bursts from the GTE times, where N was the number of bursts used in the analysis. We also simulated data from uniform distributions given the bursts restricted to those recorded during singing. Analysis was performed similar to the GTE simulation, except the simulation was run 100 times and the mean and standard deviation was calculated.

SUPPLEMENTAL REFERENCES

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.

Boari, S., Sanz Perl, Y., Amador, A., Margoliash, D., and Mindlin, G.B. (2015). Automatic reconstruction of physiological gestures used in a model of birdsong production. J Neurophysiol, jn 00385 02015.

Dombeck, D.A., Harvey, C.D., Tian, L., Looger, L.L., and Tank, D.W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nat Neurosci *13*, 1433-1440.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). Bayseian Data Statistics, Third Edition edn (CRC Press).

Glaze, C.M., and Troyer, T.W. (2006). Temporal structure in zebra finch song: implications for motor coding. J Neurosci *26*, 991-1005.

Hahnloser, R.H., 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.

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

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the mean. Journal of Experimental Social Psychology *49*, 764-766.

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.

Miri, A., Daie, K., Arrenberg, A.B., Baier, H., Aksay, E., and Tank, D.W. (2011). Spatial gradients and multidimensional dynamics in a neural integrator circuit. Nat Neurosci *14*, 1150-1159.

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.

Pakman, A., Huggins, J., Smith, C., and Paninski, L. (2014). Fast state-space methods for inferring dendritic synaptic connectivity. J Comput Neurosci *36*, 415-443.

Pnevmatikakis, E.A., Merel, J., Pakman, A., and Paninski, L. (2013). Bayesian spike inference from calcium imaging data. Paper presented at: Asilomar Conference on Signals, Systems and Computers.

Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed Eng Online *2*, 13.

Rosenthal, J.S. (2011). Optimal proposal distributions and adaptive MCMC. In Handbook of Markov Chain Monte Carlo, pp. 93-112.

Tan, V.Y.F., and Goyal, V.K. (2008). Estimating signals with finite rate of innovation from noisy samples: A stochastic algorithm. IEEE Trans Signal Process *56*, 5135-5146.

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.

Vallentin, D., and Long, M.A. (2015). Motor origin of precise synaptic inputs onto forebrain neurons driving a skilled behavior. J Neurosci *35*, 299-307.

SUPPLEMENTARY MOVIE LEGEND

Movie S1. Singing behavior in the head-fixed 2-photon configuration (Related to Figure 1)

A movie showing a head-fixed male zebra finch whose HVC was being imaged with a 2-photon microscope while he was singing to two female zebra finches. The visual access was enabled though the clearing of a polarized glass separator. Here the male produced three separate bouts of his courtship song, each of which was rewarded with a drop of water (audible as a click).