



Published in final edited form as:

Nature. 2013 March 7; 495(7439): 59–64. doi:10.1038/nature11967.

Elemental gesture dynamics are encoded by song premotor cortical neurons

Ana Amador^{1,†}, Yonatan Sanz Perl², Gabriel Mindlin², and Daniel Margoliash¹

¹University of Chicago

²University of Buenos Aires

Abstract

Quantitative biomechanical models can identify control parameters used during movements, and movement parameters encoded by premotor neurons. We fit a mathematical dynamical systems model including subsyringeal pressure, syringeal biomechanics, and upper vocal tract filtering to the songs of zebra finches. This reduced the dimensionality of singing dynamics, described as trajectories in pressure-tension space (motor “gestures”). We assessed model performance by characterizing the auditory response “replay” of song premotor HVC neurons to presentation of song variants in sleeping birds, and by examining HVC activity in singing birds. HVC projection neurons were excited and interneurons were suppressed with near-zero time lag, at times of gesture trajectory extrema. Thus, HVC precisely encodes vocal motor output via the timing of extreme points of movement trajectories. We propose that the sequential activity of HVC neurons represents the sequence of gestures in song as a “forward” model making predictions on expected behavior to evaluate feedback.

For a given set of movements, sets of movement parameters tend to be correlated with each other, so that it is difficult to resolve if motor cortical neurons encode different sets of static parameters (e.g. position, velocity, direction) or even to distinguish between static and time-dependent parameters (e.g. path trajectory)¹. In principle, the motor coding problem can be addressed by developing quantitative models that describe the biomechanics of the movements². To the extent that such models capture the actual control elements used to produce a movement, this permits motor neuron activity to be evaluated in a natural framework. We examined motor control in the bird song system from this perspective, creating a dynamical systems model of the avian vocal organ (syrinx) that captures many of the rich set of vocal behaviors that characterize bird songs³.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence and requests for materials should be addressed to dan@bigbird.uchicago.edu.

[†]Present address: Department of Physics, University of Buenos Aires

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions: AA, GBM and YSP developed the syringeal model, GBM and YSP modeled the songs, AA conducted surgeries, sound recordings and collected the electrophysiological data, AA, GBM, and DM conceived of and designed the experiments and prepared the manuscript, and all four authors participated in data analysis.

There are no competing financial interests.

We assessed predictions of the biomechanical model by taking advantage of a neuronal replay phenomenon⁴⁻⁶. Neurons in the nucleus HVC, a secondary motor/association cortical structure which is the most central structure known to be essential for singing, emit precise premotor activity when a bird sings⁵⁻⁷, and have responses that are very similar in timing and structure⁶ that are highly selective for the bird's own song (BOS) when a bird listens to playback of song^{8,9}. In zebra finches, there is a striking state-dependent neuronal replay phenomenon⁴ associated with song learning¹⁰, so that the strongest and most selective auditory responses are recorded in sleeping birds. We used responses to song in sleeping adult zebra finches as a proxy for evaluating the structure of singing, and then tested emerging hypotheses in singing birds.

Validating a song model: static parameters

The avian vocal organ is a nonlinear device¹¹⁻¹³ capable of generating complex sounds even when driven by simple instructions^{14,15}. We extended a low dimensional model of the avian syrinx and vocal tract that can capture a variety of acoustic features like the precise relationship between fundamental frequency and spectral content of zebra finch song^{16,17}. The model used here is summarized in Fig. 1. A two dimensional set of equations describes the labial dynamics (see Methods) (Fig.1, $x(t)$, red trace). Flow fluctuations are fed into a vocal tract, generating an input sound $P_i(t)$ (green trace). The tract filters the sound and is characterized as a trachea, modeled by a tube, which connects to the oro-esophageal cavity (OEC), here modeled as a Helmholtz resonator¹⁸ (see Methods). The output of the model is a time trace representing the uttered sound ($P_{out}(t)$) (blue trace).

Using this model, we created synthetic versions of the songs our test birds sang. Time dependent parameters of the model describing the labial dynamics were reconstructed to account for the time dependent acoustic properties of the sound (see Methods). Following^{3,16,17}, for each bird's song we used an algorithmic procedure to reconstruct unique functions for the air sac pressure ($\alpha(t)$) and the tension of syringeal labia ($\beta(t)$). The result of the procedure for one song is illustrated in Fig. 2, showing that many features observed in the spectrograph of the recorded song (Fig. 2a) were also apparent in the synthesized song (Fig. 2b). Relatively simple time traces of reconstructed pressure and tension arose from fitting the bird's song (Fig. 2c). These two functions drove the nonlinear equations for the labia to produce a wide range of diverse acoustic features. The parameter space of pressure vs. tension was organized by bifurcation curves (Fig. 2d, black lines), i.e. curves in the parameter space that separated regions where the model presented qualitatively different dynamics (sound patterns). Only one region (Fig. 2d, gray region) corresponded to oscillatory behavior, i.e. labial oscillations resulting in sound pressure fluctuations. Two features of the pressure-tension trajectories resulting in sound output were apparent (Fig. 2d). One, most of the control parameters were maintained close to bifurcation curves, facilitating rapid changes in the quality of sound output with small changes in parameter values. Two, there were many sounds that were characterized principally by movements in pressure or tension but not both.

Song was described by the sequence of these pressure-tension trajectories, which we call gestures, with gesture onsets and offsets defined as discontinuities in either the pressure or

tension functions (Fig. 2c). Gestures include movements that do not result in phonation, such as pressure patterns associated with mini-breaths between syllables¹⁹ but our recordings here were limited to airborne sounds. In a sample of 8 modeled songs, there were 13 ± 4 gestures per motif (largest basic unit of song, a repeated sequence of syllables). The distribution of gesture durations (mode = 22.5 ± 2.5 ms, range 4–142 ms) was non-Gaussian, with 33% of the gestures > 30 ms, and a long tail corresponding to slowly varying sounds such as constant frequency harmonic stacks (Fig. 2e).

This simple model captured essential features of sound production in a framework of labial tension and subsyringeal pressure over which birds have direct motor control^{20–22}. Whereas the actual syrinx has considerable additional complexity, the model provided for substantial dimensionality reduction. This allowed us to capture a wide range of acoustic features in a small set of time dependent parameters.

We tested the model by comparing responses of HVC neurons to the broadcast of the modeled song (mBOS) and BOS in sleeping birds (Fig. 3). Responses to a grid of mBOS stimuli with identical timing but different spectra from BOS identified optimal estimates for two remaining free static parameters (Supplementary Fig. 1). In sleeping birds, song system neurons are exceptionally selective and it was far from trivial to induce a response: for example mBOS generated without the OEC component failed to elicit response. In a case where we mis-estimated the duration of a component of BOS by 5 ms, a neuron responded strongly to BOS but not at all to the synthetic song (Supplementary Fig. 2b). Over a population of 30 neurons, the best mBOS elicited $58\% \pm 8\%$ of the response to BOS (Supplementary Note 1). Both phasic projection neurons ($HVC_{(p)}$) ($N=15$) and tonic interneurons ($HVC_{(i)}$) ($N=15$) responded selectively to mBOS over non-BOS stimuli (Supplementary Note 1). These results show that a low dimensional model representing an approximation of peripheral mechanics is sufficient to capture behaviourally relevant features of song.

Projection neurons burst at gesture extrema

We then evaluated the activity of HVC neurons relative to model dynamics, analyzing the timing of spike bursting relative to the pressure-tension trajectories used to synthesize mBOS. This identified a compelling relation between the timing of $HVC_{(p)}$ spikes and the pressure-tension trajectories. For example, in Fig. 4a the spiking of two neurons (coded with different colors) is shown relative to the BOS spectrograph, oscillograph and reconstructed pressure and tension time series. One neuron bursts once, at the transition between descending frequency modulations and a constant frequency “high note”. The other neuron bursts twice, once when the pressure during a high note reached a maximum, the other time at the transition between a high frequency chevron and a broadband frequency modulated sound. Similar relations between spike burst timing and gestures were observed for 14 of the 15 $HVC_{(p)}$ (Supplementary Figs 2 and 3). In one case, a neuron emitted bursts in the interval between syllables. We hypothesize this pattern might arise if the bursts are associated with mini-breaths during singing¹⁹. Only the 17 bursts occurring during phonation were considered for further analysis.

Examining the responses of the $HVC_{(p)}$ on pressure vs. tension plots demonstrated that neurons burst preferentially at gesture trajectory extrema (GTE) associated with gestures (Fig. 4b). A gesture has at least two GTE, at its beginning and end, and up to two additional GTE, if the absolute maxima of pressure and/or tension represent unique and distinct time points. No additional GTE result in cases where the absolute maximum is not distinct in time, e.g., multiple local maxima with same magnitude. Of the 17 bursts (14 $HVC_{(p)}$), 11 (65%) were aligned with onsets/offsets, and 6 (35%) were aligned with pressure or tension maxima. In a sample of 5 songs, there were 28 ± 4 GTE per song (165 total GTE). From a total of 60 gestures, 20 (33.4%) had only onset and offset GTEs; 30 (50%) had in addition a unique peak in pressure (3 GTEs per gesture); 5 (8.3%) had in addition a unique peak in tension (3 GTEs per gesture); and 5 (8.3%) had in addition unique peaks in both pressure and tension. The distribution of time intervals between successive GTE (mode = 9 ± 1 ms, range 4 – 116 ms) was non-Gaussian, with 66% of the intervals > 30 ms (Fig. 4c). This is graphically emphasized with tick marks showing all GTEs in Fig. 4a and Supplementary Figs 2,3. Most gestures corresponded to notes (the smallest unit of song organization recognized by ornithologists), yet motor activity at GTE maxima could subdivide notes, for example where a neuron burst and the pressure reached a maximum in the middle of a constant frequency harmonic stack (Supplementary Fig. 2). These examples highlight that for some $HVC_{(p)}$ the patterns of activity would not be interpretable with a purely spectrographic analysis of song⁵. We also observed cases where $HVC_{(p)}$ burst at the onset of relatively pure pressure-only or tension-only trajectories, with a preponderance for pressure-only trajectories (Fig. 2d). If such neurons project to distinct regions of HVC's afferent targets, which are organized based on the syringeal muscles and interactions with respiratory system, such observation could help resolve the long-standing riddle of HVC's topographic organization.

To quantify these observations, we calculated the time between each spike in each burst to the closest GTE for all 17 bursts. The resulting distribution was approximately Gaussian, with bursts on average preceding the closest GTE (mean = -5.6 ± 0.3 ms, $\sigma = 6.7 \pm 0.3$ ms; Fig. 4d). A bootstrap procedure (Supplementary Note 2) confirmed that the correspondence to the closest GTE was statistically significant (F test, $P < 0.045$). This indicates that the timing of $HVC_{(p)}$ bursts is associated with the timing of GTE. Given a minimal delay between activity of $HVC_{(p)}$ and sound production estimated between 25–50 ms²³, the minimal 15 ms delay for auditory feedback to HVC⁸, and that the duration of intervals between GTE varied greatly (Fig. 4c), it is remarkable that the timing of $HVC_{(p)}$ bursting was synchronized with near-zero time lag to a model of actual behavioral output.

Interneurons are suppressed at GTE

We also noted a relation between the minima in the activity of $HVC_{(i)}$ and the timing of GTE. To characterize this, for each interneuron, we binned the spikes in 10 ms windows for each acoustic presentation. The resultant average response traces were smoothed and the minima in the smoothed traces were identified (see Methods). For an example neuron, the average response is shown in green, the superimposed smoothed curve in black, and the minima in red dots (Fig. 5a, bottom panel). Each $HVC_{(i)}$ did not have minima at all GTE, but across all neurons, we observed a close alignment between the times of the minima and

the times of GTE. (A non-significant relation was observed for maxima of $HVC_{(i)}$ activity; Supplementary Fig. 4.) Computing the differences between the time of each minimum that occurred during phonation and the closest GTE resulted in a distribution that was approximately Gaussian (mean = $-0.82 \text{ ms} \pm 0.60 \text{ ms}$, $\sigma = 7.3 \pm 1.4 \text{ ms}$; Fig. 5b). We compare this distribution to the distribution of randomly positioned minima within each motif using the bootstrap procedure and found them to be significantly different (F test, $P < 0.016$, Supplementary Note 2). Additional tests identified marginally significant locking to GTE for one of four birds (Supplementary Note 3). Thus, the precise activity of $HVC_{(i)}$ ⁷ can help shape the timing of $HVC_{(p)}$. This suggests a simple model where bursts of activity of $HVC_{(p)}$ suppress activity in $HVC_{(i)}$, whose ongoing activity helps shape the next $HVC_{(p)}$ burst.

A representation of gestures during singing

Given that our results were obtained by broadcasting songs to sleeping birds, it is natural to inquire if during singing the activity of HVC neurons are also locked to gesture transitions. Previous results have demonstrated tight temporal locking comparing daytime singing activity and auditory-driven responses during sleep of single RA neurons in zebra finches⁴, and HVC neurons in awake swamp sparrows and Bengalese finches that respond to auditory stimulation⁶, but similar observations have yet to be reported for zebra finch HVC neurons. We made recordings from HVC in singing birds ($N = 3$ birds), including 5 phasic neurons bursting during phonation (recorded in two of the three birds, Fig. 6, Supplementary Fig. 5); one neuron had two bursts per motif, and 10 tonic neurons. We confirmed that during singing, all sparse bursts of $HVC_{(p)}$ occurred at gesture transitions (Fig 4e). Following the same analysis as in sleeping birds (but here, since each motif of song could vary, it was independently modeled), we observed for singing birds even more precise timing of $HVC_{(p)}$ than was observed during sleeping (cf. Fig. 4d, e). The Gaussian fit for the population of phasic neurons recorded during singing (mean = $-1.35 \text{ ms} \pm 0.10 \text{ ms}$, $\sigma = 4.0 \pm 0.1 \text{ ms}$; Fig. 4e) was significantly different from the bootstrapped random distribution (F test, $P < 0.025$, see Supplementary Note 2 and Supplementary Fig. 6). The minima activity of tonic neurons recorded during singing also showed precise timing relative to GTEs (Gaussian fit for the minima: mean = $-0.12 \text{ ms} \pm 0.4 \text{ ms}$, $\sigma = 4.0 \pm 0.4 \text{ ms}$; Fig. 5c), and this was significantly different than the bootstrapped random distribution (F test, $P < 0.002$). Additional analyses demonstrated significant locking of minima to GTE in two of three singing birds (Supplementary Note 3). As for sleeping birds, the maxima of tonic neural activity showed no evidence of a significant locking to the GTEs (Supplementary Fig. 4c). Finally, examining the data from a prior study of zebra finches²⁴ we observed that during singing the timing of $HVC_{(RA)}$ bursts were closely associated with the timing of $HVC_{(X)}$ bursts (Supplementary Fig. 7. In light of our results, this supports the hypothesis that all classes of HVC neurons are active in relation to the timing of gestures, although the multiple subtypes of $HVC_{(RA)}$, $HVC_{(X)}$, and $HVC_{(i)}$ have yet to be evaluated.

Previously it was concluded that the timing of song syllables was unrelated to the timing of $HVC_{(p)}$ discharge^{5,24} in singing birds. Given the sparse bursting⁵ of these cells this led to the idea that the output of HVC had a time clock-like function with a nearly uniform “tick” size of approximately 10 ms ²³ supported by a “syn-fire” chain of synaptic activity across

HVC_(p)⁵. Instead we find that the bursting of HVC_(p) and modulation of HVC_(i) activity is timed to significant instances of motor gestures. The sequential firing across the population of HVC_(p) unfolds in an ordered fashion⁵ but time is not explicitly represented in HVC. Instead, the statistics of HVC activity are closely tied to syringeal/vocal tract mechanics. Given the broad distribution of times between GTE, if HVC activity is synchronized with GTEs this is inconsistent with a syn-fire network that is active at every moment. The distinction between these two models of HVC has additional broad implications for the functional organization of the song system, for song learning, and for motor coding.

Since gestures vary greatly in duration, and RA only has access to the times of GTE, then downstream components of the motor pathway (RA and presumably brainstem) should generate independent dynamical information to sustain the detailed structure within each gesture (cf.^{23,25}). Previous experimental results, including the effects of electrical stimulation of HVC or RA during singing²⁶ and lesions of nuclei afferent to HVC²⁷ implicate information in HVC encoding larger units of song. This might arise if some gestures or transitions are over-emphasized in HVC relative to others. Finally, gestures are learned, which is consistent with the physiological properties of HVC neurons: integration over hundreds of milliseconds and multiple syllables, non-linear summation over syllables in a sequence preceding the excitatory response, and selective response to BOS^{4,8,9,28,29,30}. The information about groupings of gestures such as syllables can be carried in these integrated signals. This also re-emphasizes synaptic modification in HVC, not just changes at HVC-RA synapses, are associated with feedback mediated sensorimotor learning (cf. ²³). HVC also projects to the cortico-basal ganglia pathway which contributes to learning-mediated synaptic modification in RA by introducing variance into song output^{31,32}. This suggests the hypothesis that the variance is structured not in an auditory framework but around specific features of song motor gestures.

A forward model for vocomotor control

If activity in HVC is in synchrony with little time lag with motor gestures occurring at the periphery this would tend to bring it into temporal register with fixed (circa 15 ms) delayed auditory³³, proprioceptive²⁰, or brainstem³⁴ feedback. This allows movements to be represented in HVC by gestures of greatly varying duration (with dynamics principally generated through internal HVC interactions) while each gesture is referenced to a common time framework for evaluating feedback (with feedback arriving through distinct, extrinsic inputs). This suggests that projection neurons represent a prediction about the actual behavioral output at that moment in time, constituting an unexpected form of a “forward” or predictive model to resolve the problem of the delay in sensorimotor control³⁵. Assuming that behavior is subdivided into gestures, and only the transitions (GTE) are represented by HVC output (HVC_(p)), then the intervals between the transitions could accumulate feedback information by modifying the tonic activity of HVC_(i) and subsequently the spike bursting of HVC_(p). Indeed, HVC receives multiple sources of feedback including input from the primary motor cortex RA³⁶, thalamic input carrying brainstem respiratory, auditory, and proprioceptive information^{21,34,37}, and forebrain auditory input³⁸.

We have described song organization based on gestures, by taking advantage of the dynamical systems modeling framework to go beyond spectrographs. These features of motor systems organization could obtain generally³⁹. Our data support Sherrington's long-standing hypothesis that the motor cortex is a synthetic organ, representing segments of whole movements^{1,40}. In humans the production of speech and the performance of athletes and musicians are an exceptional example of highly precise learned skilled behavior that could share mechanisms to those described here. Developing corresponding models for human speech production should help inform speech and language pathologies where sequential behavior is disrupted.

Methods

Subjects, songs, and surgeries

All procedures were in accordance with a protocol approved by the University of Chicago Institutional Animal Care and Use Committee. Songs were recorded from 12 birds and electrophysiology was conducted on 9 adult male zebra finches (*Taeniopygia guttata*) bred in our colony. Birds were prepared for recordings with surgeries using standard techniques to implant a head pin (for auditory experiments)¹⁰ or motorized microdrive (for singing experiments)⁵. For auditory experiments, adults were maintained on a 16/8 reversed light cycle in sound isolation boxes. Songs were recorded and filtered using custom software (SABER, A.S. Dave) then edited (Praat, P. Boersma and D. Weenink, www.praat.org). Edited songs included two or three repetitions of one motif, and were typically 2 – 4 s in duration. Birds were prepared for recordings with surgeries using standard techniques to implant a head pin (for auditory experiments)¹⁰ or motorized microdrive (for singing experiments)⁵. Birds were allowed to recover for 2 or 3 days before the first of the days of recordings, and rest for at least 2 days between recording sessions.

Electrophysiology, stimulus presentation, and spike analysis

HVC extracellular recordings were performed in head-fixed sleeping or singing tethered birds. Recordings were post-processed with a spike-sorting algorithm (Klusters, L. Hazan, klusters.sourceforge.net and custom software written by C.D. Meliza) to separate the times of spike events for each unit. For experiments in singing birds, all well-isolated neurons are reported. For auditory experiments, only BOS responsive neurons were recorded. The auditory stimuli were presented randomly with an interstimulus interval of 7 ± 1 s. The neural response to each song is quantified in terms of the Z score²⁵:

$$Z = \frac{\mu_S - \mu_{BG}}{\sqrt{\text{Var}(S) - \text{Var}(BG) - 2\text{Covar}(S, BG)}}$$

where μ_S is the mean response during the auditory stimulus (S) and μ_{BG} is the mean response during background activity (BG). The denominator of the equation is the standard deviation of $(S - BG)$. The background was estimated by averaging the firing rate during a 2 sec period. The Z score of the mBOS, CON, and REV were normalized to the BOS Z score, and averages across neurons were reported as mean of normalized responses \pm s.e.m. For interneurons, the strength of the response varied across the motifs⁴². We picked the last

(second or third) motif, which gave the strongest response, to analyze the timing of spikes relative to GTE. This minimized false peaks and troughs in the response profiles. In singing birds, interneurons fired reliably for each motif and all motifs were incorporated into the analysis. The average response of each interneuron (1 ms resolution) was smoothed using a Savitsky Golay filter (polynomial local regression⁴¹) and the minima were identified using a 21-point sliding window.

Reconstruction of motor gestures

We assumed flow induced oscillations of opposing labia as a sound source model for birdsong production¹⁴. This model assumes that for high enough airflow values, the labia start to oscillate with a wavelike motion. Assuming two basic modes active (a flapping like motion and a lateral displacement of the tissues, appropriately out of phase), a system of equations describe the dynamics of the medial position $x(t)$ of one of the opposing labia, at one of the sound sources. These read

$$\frac{dx}{dt} = y \quad \frac{dy}{dt} = (1/m)(-k(x)x - (b(y)+cx^2)y + a_{lab}p_{av})$$

where the first term in the second equation is the restitution in the labium, the second term accounts for the dissipation, and the last term for the force due to the interlabial pressure. The average pressure p_{av} can be written in terms of the displacement and its velocity³. These equations describe a set of qualitatively different dynamical regimes. To gain independence from the details of any particular model presenting these regimes, we worked with a normal form that unfolds into a Saddle node in limit cycle bifurcation and a Hopf bifurcation. The normal form, which is analytically derived⁴³, constitutes the simplest set of equations for any model in which oscillations arise in either of these two bifurcations. Once this reduction is performed, the selection of parameters that allow obtaining a sound with specific acoustic features gives rise to unique values. The normal form equations are shown in Fig. 1, and display the same set of dynamical regimes³ as the physical model, with scaling through a time constant γ . Once $x(t)$ is computed, the pressure at the input of the tract is computed as $P_i(t) = \alpha(t)x(t) - rP_i(t-T)$ where T is the time for a sound wave to reach the end of the tube and return, and $\alpha(t)$ is proportional to the average mean velocity of the flow. The transmitted pressure fluctuation $P_i(t) = (1-r)P_i(t-0.5T)$ forces the air in the glottis, which is approximated by the neck of a Helmholtz resonator (used to model the OEC^{3,44}), i.e., a large container with a hole, such that the air in its vicinity oscillates due to the springiness of the air in the cavity. A linear set of three ordinary differential equations accounts for the dynamics of the air flow and pressure in this linear acoustic device³, resulting in the final output pressure $P_{out}(t)$ (Fig. 1).

We reconstructed the parameters driving the equations of the normal form ($\alpha(t)$ and $\beta(t)$), as well as the parameters describing the tracheal length and the OEC cavity in such a way that the synthesized sounds presented the same fundamental frequencies and spectral content as natural song. Reconstructions over sequential sound segments gave estimates of the time dependence of physiological parameters used during song production. A linear integrator ($\tau = 2.5$ ms) was used to compute the envelope of the sound signal. A threshold was used to

identify phonating segments. For those longer than 20 ms, we decomposed the recorded songs into successive 20 ms segments (time between consecutive segments $\Delta t = 1/20000$ s). These were short enough to avoid large variation of the physiological gestures, and long enough to compute spectral content. For each segment, we computed the spectral content index (SCI)¹⁶ and the fundamental frequency. A search in the parameter space $(\alpha(t), \beta(t))$ was performed over a grid so that the synthetic sounds produced would match the fundamental frequencies of the song segment being fitted. Over the set of $(\alpha(t), \beta(t))$ values selected, a search was performed so that SCI of the synthetic sounds matched the value of the song segment³. For sound segments shorter than 20 ms, the fundamental frequency was computed as follows. First, we selected the relative maxima of the sound signal that reached the sound envelope. Then, the fundamental frequency was computed as the inverse of the time difference between the next two consecutive selected maxima. The SCI at that time was estimated as the average value among all the possible SCI values, corresponding to that frequency in the framework of the model¹⁶. With those estimations of fundamental frequency and SCI, $(\alpha(t), \beta(t))$ were computed. Brief segments were typically fast trills. We modeled those as rapid oscillations of pressure and tension, with the amplitude of the pressure oscillations such that the maxima fall in the phonating region, and amplitude of the tension oscillations such that the frequency range of the vocalization was reproduced. We found that most of the parameters could be well approximated by either fractions of sine functions, exponential decays, constants, or combinations of those.

Using these analytic functions as parameters of the model to generate a synthetic copy of the recorded song resulted in a noiseless surrogate song (e.g., Supplementary Fig. 1, Noise=0). The addition of noise allowed the gradual recovery of realistic timbric features. In the text, the dimensionless variable Noise varied between 0 and 40, with Noise=5 corresponding to a fluctuation size equal to 2.5 percent of the maximum range of the $\beta(t)$ parameter. Notice that the timbric effect will be more important for low frequency sounds, which explore a small range of $\beta(t)$.

For each bird, the length of the trachea was chosen so that the frequencies close to 2.5 kHz and 7 kHz in the bird's song were the first and second resonances of a tube closed at one end. This corresponds to a length of 3.5 cm⁴⁵. Typically, zebra finch songs present a third important resonance around 4 kHz. The parameters of the Helmholtz resonator were adjusted so that its resonant frequency would account for this resonance³. The synthetic songs for sleeping birds were generated before doing the electrophysiological experiments. For singing birds all song reconstructions were also performed blind to the spike data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Richard H. R. Hahnloser for help with the microdrives and techniques used to record from singing birds. We thank Henry D. I. Abarbanel, Timothy Q. Gentner, Howard C. Nusbaum, and Stephanie E. Palmer for valuable comments on the manuscript. Supported by a Human Frontiers Science Program cross-disciplinary fellowship award to AA, NIDCD006876, CONICET and UBA awards to GBM and YSP, and NIDCD and NSF/CRCNS awards to DM.

References

1. Hatsopoulos NG, Xu Q, Amit Y. Encoding of movement fragments in the motor cortex. *J Neurosci.* 2007; 27:5105–5114. [PubMed: 17494696]
2. Nishikawa K, et al. Neuromechanics: an integrative approach for understanding motor control. *Integr Comp Biol.* 2007; 47:16–54. [PubMed: 21672819]
3. Perl YS, Arneodo EM, Amador A, Goller F, Mindlin GB. Reconstruction of physiological instructions from Zebra finch song. *Phys Rev E.* 2011; 84
4. Dave AS, Margoliash D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science.* 2000; 290:812–816. [PubMed: 11052946]
5. Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature.* 2002; 419:65–70. [PubMed: 12214232]
6. Prather JF, Peters S, Nowicki S, Mooney R. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature.* 2008; 451:305–310. [PubMed: 18202651]
7. Yu AC, Margoliash D. Temporal hierarchical control of singing in birds. *Science.* 1996; 273:1871–1875. [PubMed: 8791594]
8. Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci.* 1983; 3:1039–1057. [PubMed: 6842281]
9. Margoliash D. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci.* 1986; 6:1643–1661. [PubMed: 3712002]
10. Shank SS, Margoliash D. Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature.* 2009; 458:73–77. [PubMed: 19079238]
11. Amador A, Goller F, Mindlin GB. Frequency modulation during song in a suboscine does not require vocal muscles. *J Neurophysiol.* 2008; 99:2383–2389. [PubMed: 18287554]
12. Elemans CPH, Laje R, Mindlin GB, Goller F. Smooth operator: avoidance of subharmonic bifurcations through mechanical mechanisms simplifies song motor control in adult zebra finches. *J Neurosci.* 2010; 30:13246–13253. [PubMed: 20926650]
13. Fee MS, Shraiman B, Pesaran B, Mitra PP. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature.* 1998; 395:67–71. [PubMed: 12071206]
14. Mindlin, GB.; Laje, R. *The Physics of Birdsong.* Berlin: Springer Verlag; 2005.
15. Laje R, Gardner TJ, Mindlin GB. Neuromuscular control of vocalizations in birdsong: A model. *Phys Rev E.* 2002; 65:051921:1–051921:8.
16. Sitt JD, Amador A, Goller F, Mindlin GB. Dynamical origin of spectrally rich vocalizations in birdsong. *Phys Rev E.* 2008; 78:011905:1–011905:6.
17. Amador A, Mindlin GB. Beyond harmonic sounds in a simple model for birdsong production. *Chaos.* 2008; 18:041023:1–041023:6.
18. Riede T, Suthers RA, Fletcher NH, Blevins WE. Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc Natl Acad Sci U S A.* 2006; 103:5543–5548. [PubMed: 16567614]
19. Hartley RS, Suthers RA. Air-flow and pressure during canary song - direct evidence for mini-breaths. *J Comp Physiol A.* 1989; 165:15–26.
20. Suthers RA, Goller F, Wild JM. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *P Natl Acad Sci USA.* 2002; 99:5680–5685.
21. Wild JM. Functional neuroanatomy of the sensorimotor control of singing. *Ann NY Acad Sci.* 2004; 1016:438–462. [PubMed: 15313789]
22. Suthers RA, Goller F, Pytte C. The neuromuscular control of birdsong. *Philos T Roy Soc B.* 1999; 354:927–939.
23. Fee MS, Kozhevnikov AA, Hahnloser RH. Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci.* 2004; 1016:153–170. [PubMed: 15313774]
24. Kozhevnikov AA, Fee MS. Singing-related activity of identified HVC neurons in the zebra finch. *J Neurophysiol.* 2007; 97:4271–4283. [PubMed: 17182906]
25. Fiete IR, Hahnloser RHR, Fee MS, Seung HS. Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *J Neurophysiol.* 2004; 92

26. Vu ET, Mazurek ME, Kuo YC. Identification of a Forebrain Motor Programming Network for the Learned Song of Zebra Finches. *J Neurosci*. 1994; 14:6924–6934. [PubMed: 7965088]
27. Williams H, Vicario DS. Temporal patterning of song production: Participation of nucleus uvaeformis of the thalamus. *J Neurobiol*. 1993; 24:903–912. [PubMed: 8228968]
28. Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci*. 1992; 12:4309–4326. [PubMed: 1432096]
29. Nick TA, Konishi M. Neural auditory selectivity develops in parallel with song. *J Neurobiol*. 2005; 62:469–481. [PubMed: 15616963]
30. Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. Neural correlates of categorical perception in learned vocal communication. *Nat Neurosci*. 2009; 12:221–228. [PubMed: 19136972]
31. Brainard MS, Doupe AJ. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature*. 2000; 404:762–766. [PubMed: 10783889]
32. Olveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol*. 2005; 3:e153. [PubMed: 15826219]
33. Konishi M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol*. 1965; 22:770–783. [PubMed: 5874921]
34. Ashmore RC, Wild JM, Schmidt MF. Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J Neurosci*. 2005; 25:8543–8554. [PubMed: 16162936]
35. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. *Science*. 1995; 269:1880–1882. [PubMed: 7569931]
36. Roberts TF, Klein ME, Kubke MF, Wild JM, Mooney R. Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *J Neurosci*. 2008; 28:3479–3489. [PubMed: 18367614]
37. Coleman MJ, Roy A, Wild JM, Mooney R. Thalamic gating of auditory responses in telencephalic song control nuclei. *J Neurosci*. 2007; 27:10024–10036. [PubMed: 17855617]
38. Bauer EE, et al. A synaptic basis for auditory-vocal integration in the songbird. *J Neurosci*. 2008; 28:1509–1522. [PubMed: 18256272]
39. Mulliken GH, Musallam S, Andersen RA. Forward estimation of movement state in posterior parietal cortex. *Proc Natl Acad Sci U S A*. 2008; 105:8170–8177. [PubMed: 18499800]
40. Leyton SS, Sherrington CS. Observations on the excitable cortex of the chimpanzee, orangutan and gorilla. *Q J Exp Physiol*. 1917; 11:135–222.
41. Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. *Numerical Recipes: The Art of Scientific Computing*. Third edn. Cambridge: Cambridge University Press; 2007.
42. Sutter ML, Margoliash D. Global synchronous response to autogenous song in zebra finch HVC. *J Neurophysiol*. 1994; 72:2105–2123. [PubMed: 7884447]
43. Guckenheimer, J.; Holmes, P. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Berlin: Springer Verlag; 1997.
44. Fletcher NH, Riede T, Suthers RA. Model for vocalization by a bird with distensible vocal cavity and open beak. *J Acoust Soc Am*. 2006; 119:1005–1011. [PubMed: 16521762]
45. Daley M, Goller F. Tracheal length changes during zebra finch song and their possible role in upper vocal tract filtering. *J Neurobiol*. 2004; 59:319–330. [PubMed: 15146548]

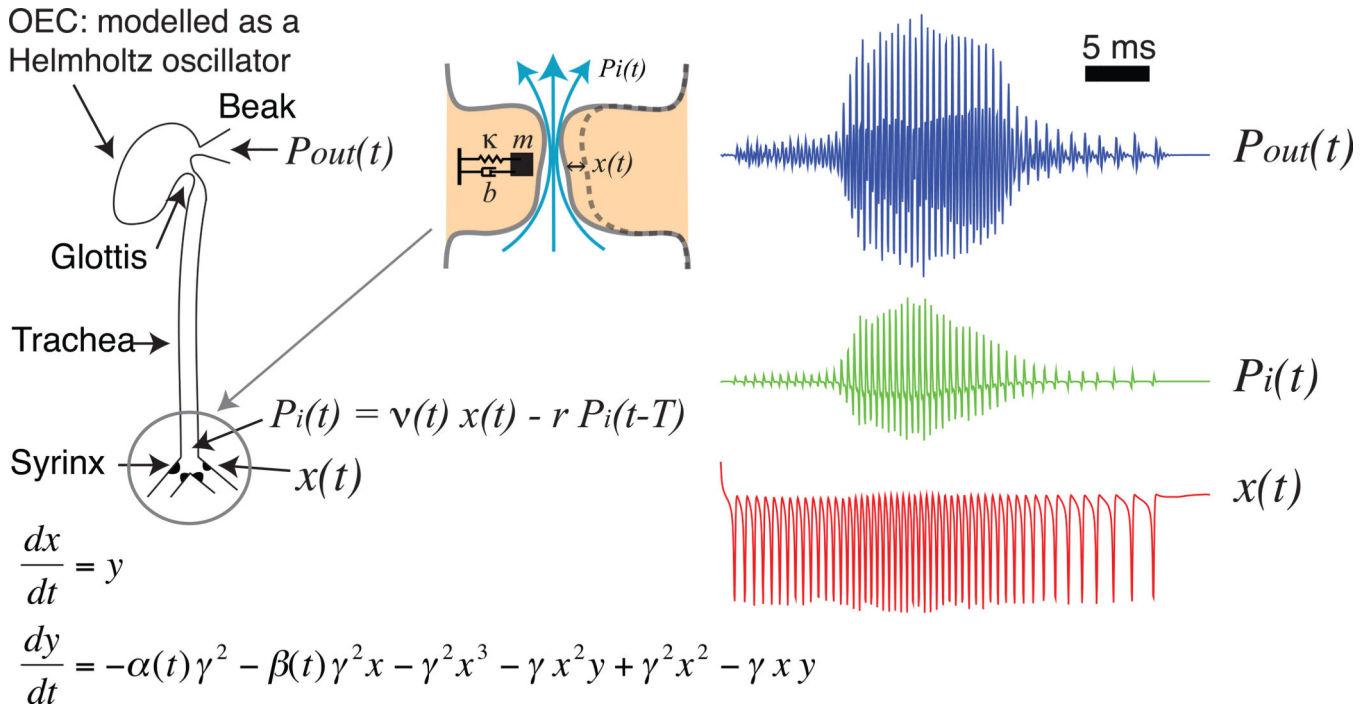


Figure 1. Schematized view of a dynamical systems model describing labial dynamics and vocal tract filtering (trachea and oro-esophageal cavity, OEC)

The syringeal membrane was modeled as a mass (m) with damping (b) and a restitution (spring) force (K). Normal form equations for labial position ($x(t)$, red line) were integrated, computing the input pressure at the vocal tract ($P_i(t)$, green line) and ultimately the total output pressure ($P_{out}(t)$, blue line). v , sound velocity; T , propagation time along trachea; γ , time constant (see Methods).

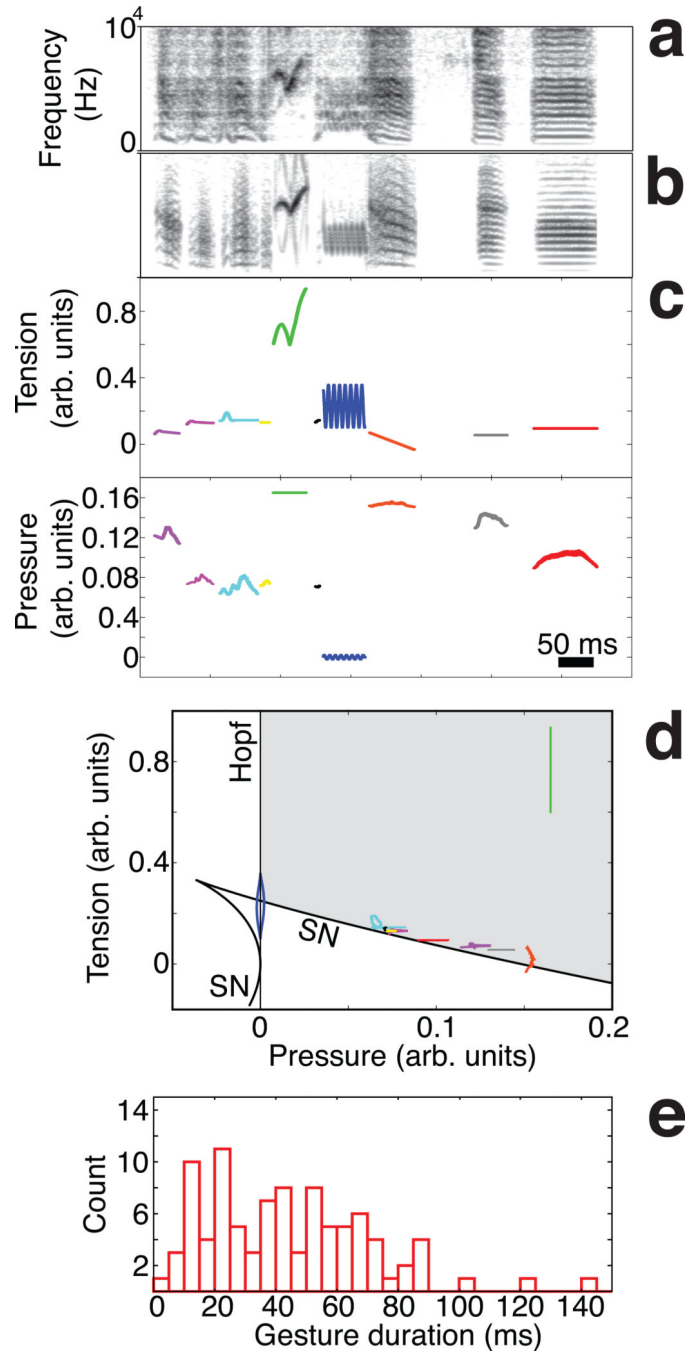


Figure 2. A low dimensional model: reconstructing gestures

Spectrographs of a bird's song (a) and model synthetic song (b). Song is described by fitted parameters $\alpha(t)$ and $\beta(t)$, proportional to air sac pressure and labial tension, respectively (c). Each sound is generated by a continuous curve in the parameter space of the model, a "gesture" (d). Oscillations in the vicinity of a SN bifurcation present rich spectra, typical of zebra finch song. Note that the spectrally poor "high note" (green) is distant from the SN bifurcation. The gray area indicates the region of phonation. The distribution of gesture durations for five birds is displayed in (e).

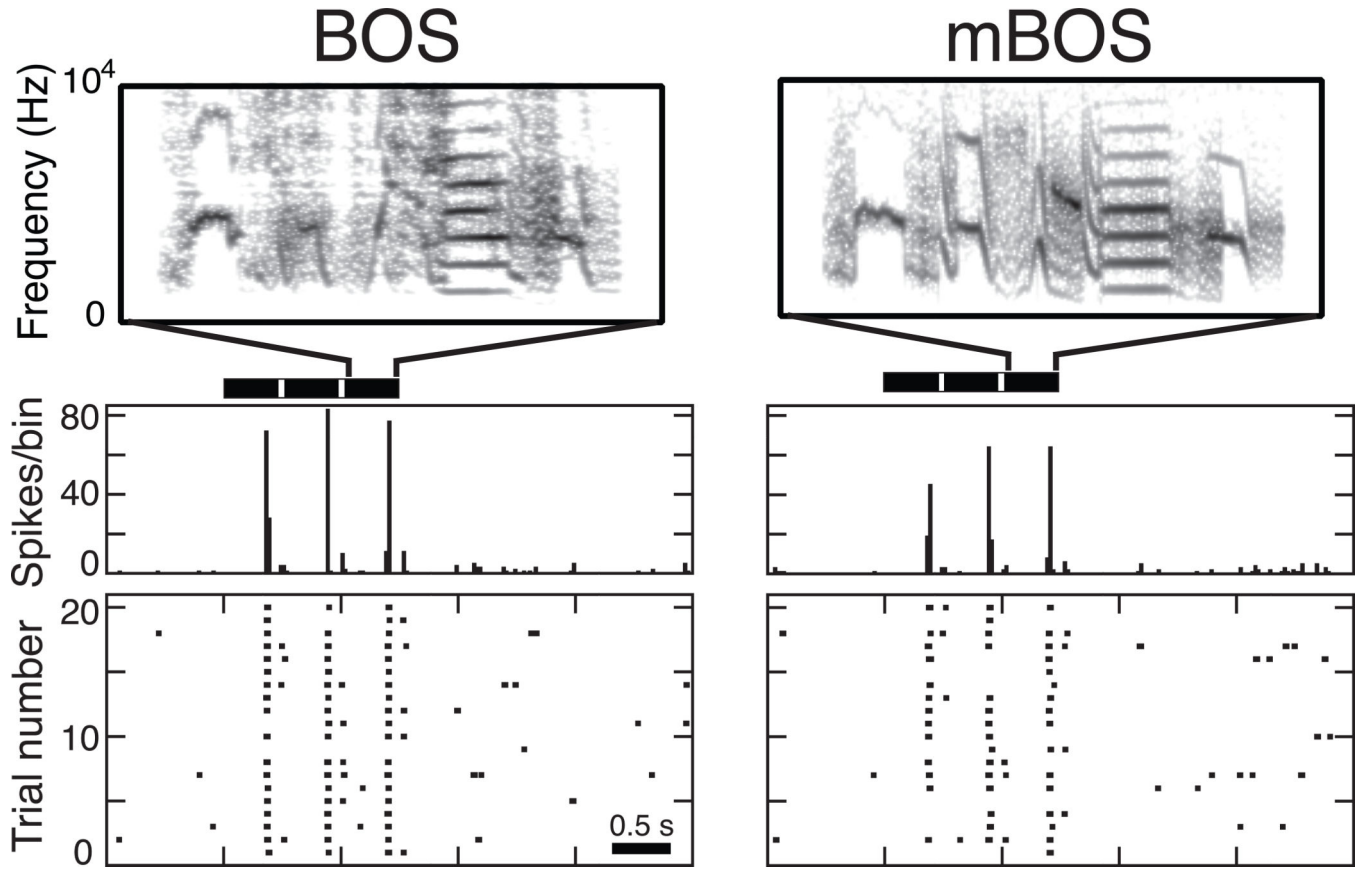


Figure 3. Testing the low dimensional model
The activity of HVC selective neurons of sleeping birds in response to the presentation of BOS and modeled BOS (mBOS) was similar. The timing of the three repeated motifs that were presented is indicated by the bold horizontal lines.

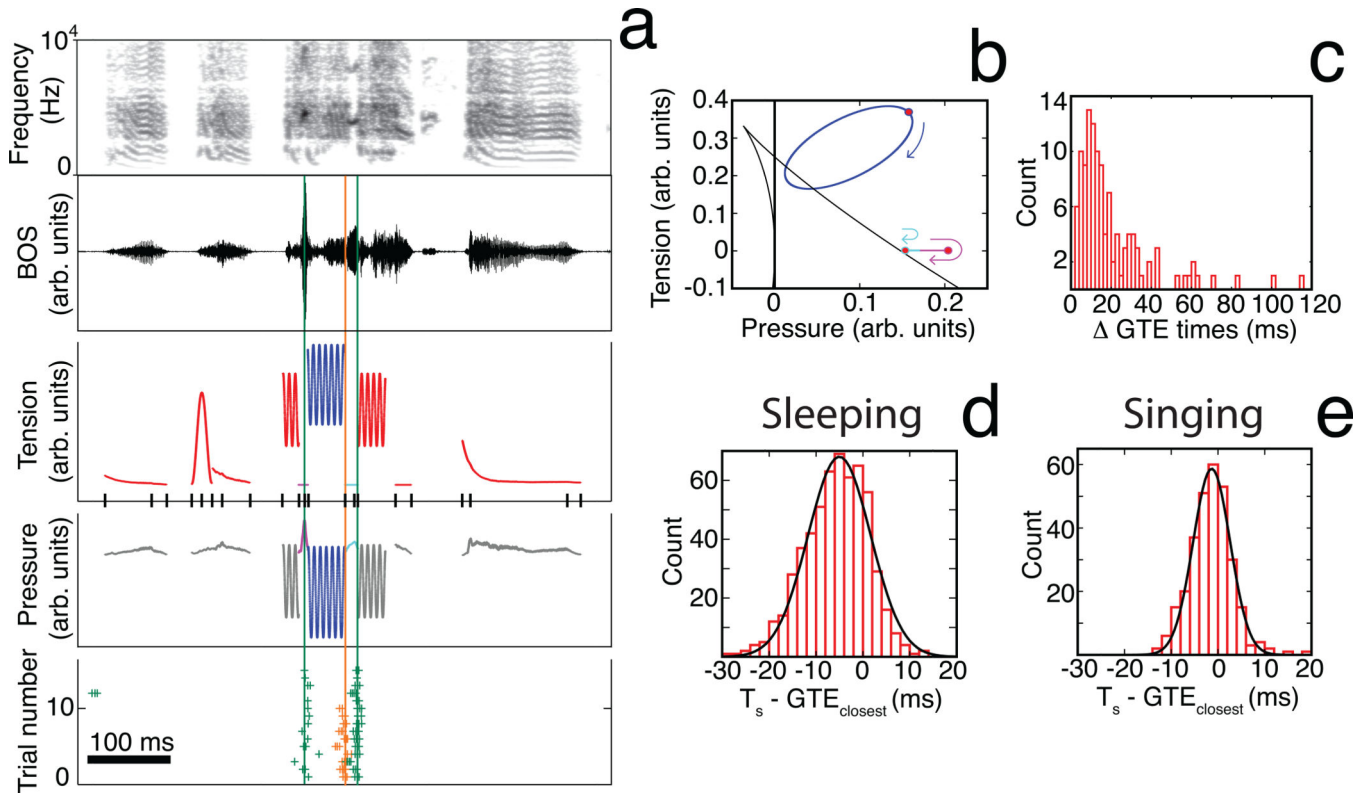


Figure 4. Timing of gestures relative to bursting of projection neurons

a, song spectrograph and oscillograph (top panels); reconstructed parameters pressure and tension (middle panels), with tick marks indicating the times of all GTEs. Bottom panel, raster plots of the responses of two neurons (color coded green and orange), together with their closest GTE, indicated with lines of the same colors. The trajectories (same color coding) in parameter space are displayed in **(b)**, with a point indicating the mean position of a burst, and arrows indicating the trajectory direction. **c**, distribution of time differences between consecutive GTE occurrences ($N = 5$ birds). **d**, distribution of time differences between the time of each spike (T_s) and the time of the closest GTE in sleeping birds ($N = 14$ $HVC_{(p)}$, 5 birds). **e**, The same analysis of **d** on singing birds ($N = 5$ $HVC_{(p)}$, 2 birds).

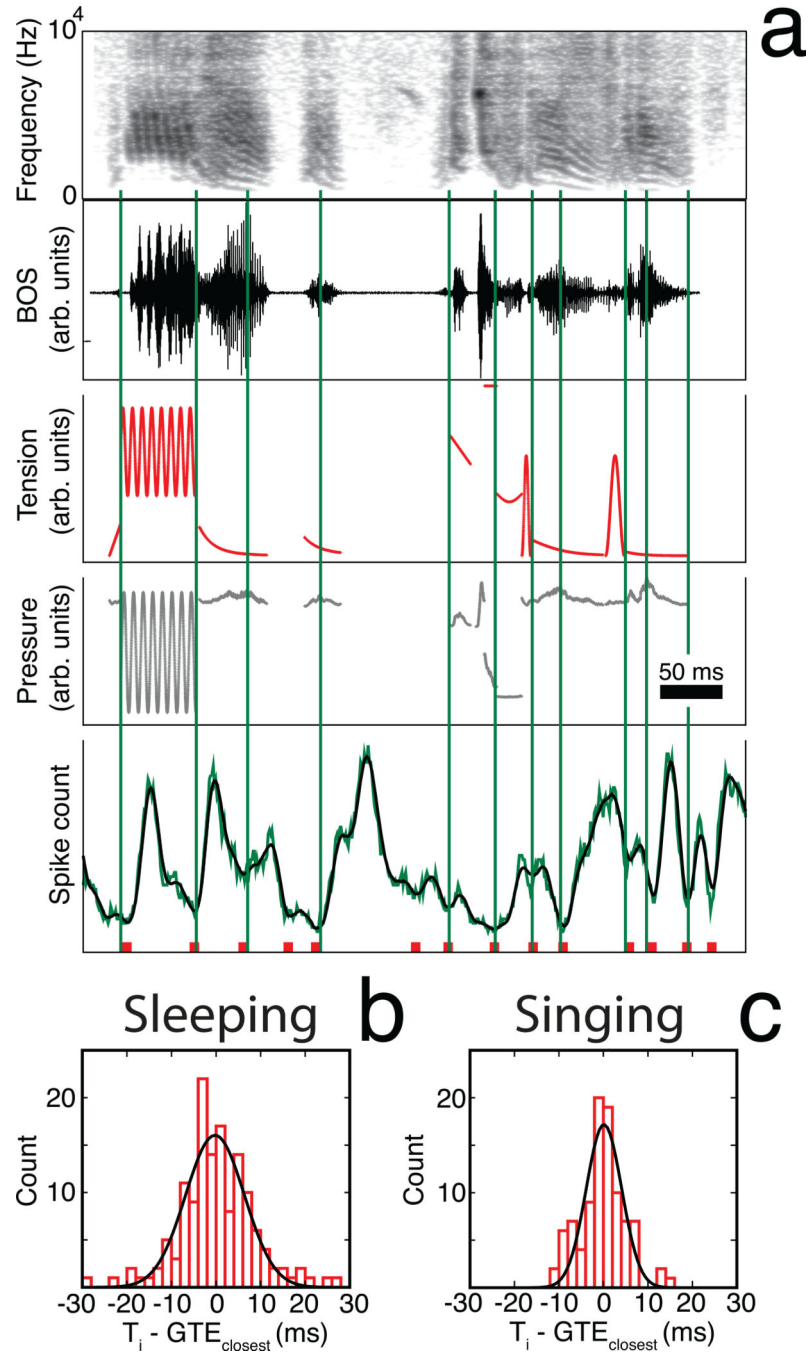


Figure 5. Suppressed interneuron activity is associated with GTEs

a, Organized as in Fig. 4a, but with spike count response to the song (10 ms bin, 20 repetitions; green line) for one $\text{HVC}_{(i)}$, and a smoothed measure of the response (black line; see Methods). Red squares indicate the time of the minima in the smoothed measure, and the vertical lines indicated the position of the closest GTE to each minima. **b**, distribution of time differences between spike response minima and their closest GTE in sleeping birds (15 $\text{HVC}_{(i)}$, 5 birds). **c**, Same analysis in singing birds (10 $\text{HVC}_{(i)}$, 3 birds).

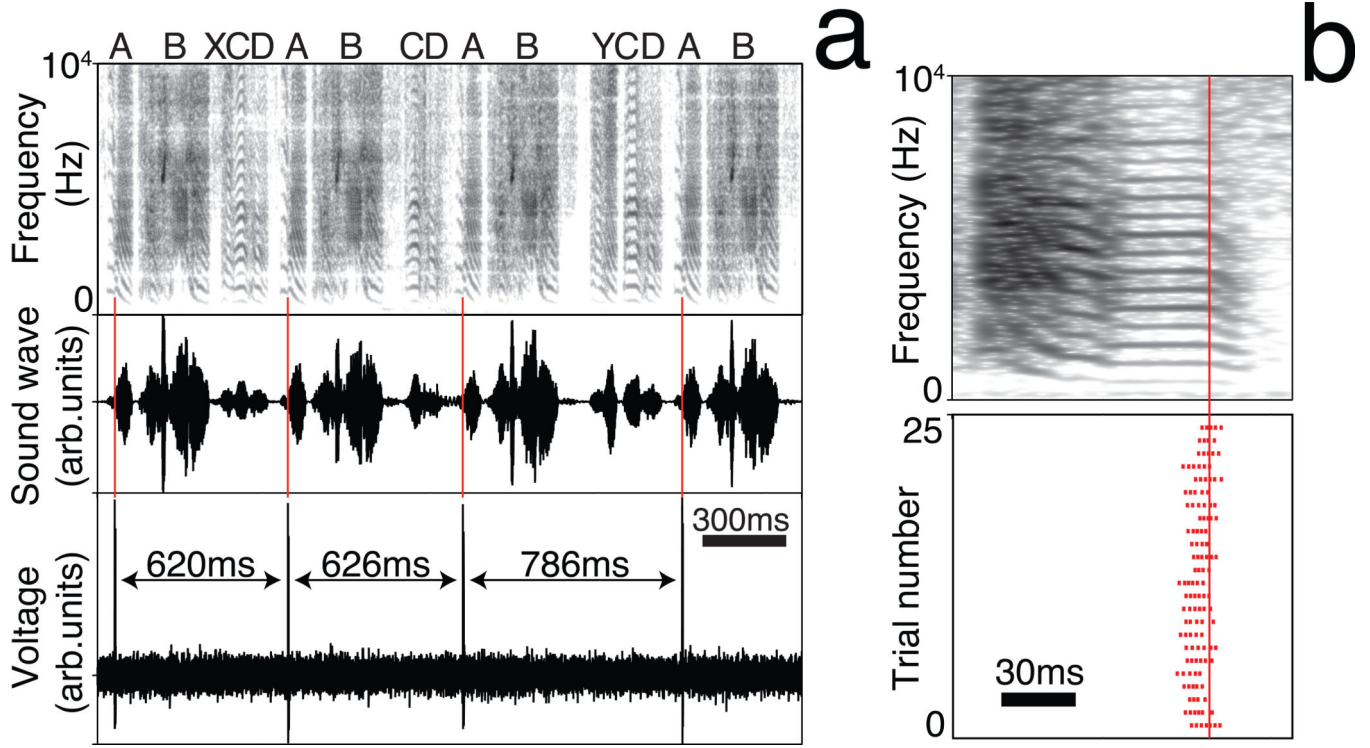


Figure 6. During singing, $HVC_{(p)}$ fired in the vicinity of GTE
a, A $HVC_{(p)}$ neuron bursts locked to the vicinity of a GTE even as the syllable sequence and time interval varies. **b**, For another bird, the burst of a $HVC_{(p)}$ neuron is locked to a GTE in the vicinity of a subtle acoustic transition.