ARTICLE

doi:10.1038/nature11967

Elemental gesture dynamics are encoded by song premotor cortical neurons

Ana Amador[1]†, Yonatan Sanz Perl[2], Gabriel B. Mindlin[2] & Daniel Margoliash[1]

Quantitative biomechanical models can identify control parameters that are used during movements, and movement parameters that are 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 reduces the dimensionality of singing dynamics, described as trajectories (motor 'gestures') in a space of syringeal pressure and tension. Here we assess model performance by characterizing the auditory response 'replay' of song premotor HVC neurons to the presentation of song variants in sleeping birds, and by examining HVC activity in singing birds. HVC projection neurons were excited and interneurons were suppressed within a few milliseconds of the extreme time points of the gesture trajectories. Thus, the HVC precisely encodes vocal motor output through activity at the times of extreme points of movement trajectories. We propose that the sequential activity of HVC neurons is used as a 'forward' model, representing the sequence of gestures in song to make predictions on expected behaviour and 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 whether motor cortical neurons encode different sets of static parameters (for example, position, velocity and direction), or even to distinguish between static and time-dependent parameters (for example, path trajectory)[1]. In principle, the motor coding problem can be addressed by developing quantitative models that describe the biomechanics of the movements[2]. To the extent that such models capture the actual control elements used to produce a movement, this permits motor cortical neuron activity to be evaluated in a natural framework. We examined motor control in the avian 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 behaviours that characterize bird songs[3].

We assessed predictions of the biomechanical model by taking advantage of a neuronal replay phenomenon[4–6]. Neurons in the nucleus HVC, a secondary motor or association cortex structure (among the structures that are essential for singing, this is the most synaptically distant from the syrinx), emit precise premotor activity when a bird sings[5–7]. Interestingly, when a bird listens to playback of its own song, the same HVC neurons have responses that are very similar in timing and structure to their responses during singing[6], and that are highly selective for the bird's own song (BOS)[8,9]. In zebra finches, there is a notable state-dependent neuronal replay phenomenon[4] associated with song learning[10], so that the strongest and most selective auditory responses are recorded in sleeping birds. We used the responses of HVC neurons in sleeping adult zebra finches to evaluate the responses to BOS and artifical BOS variants, and then tested emerging hypotheses by recording from singing birds.

Validating a song model and estimating static parameters

The avian vocal organ is a nonlinear device[11–13] that is 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 such as 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 (Fig. 1; see Methods). Flow fluctuations are fed into a vocal tract, generating an input sound ($P_{in}(t)$). The tract filters the sound and is characterized as a trachea, modelled by a tube, which connects to the oro-oesophageal cavity (OEC), modelled here as a Helmholtz resonator[18] (see Methods). The output of the model is a time trace representing the uttered sound ($P_{out}(t)$).

Using this model, we created synthetic versions of the songs that 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 procedures described previously[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 present 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 versus tension was organized by bifurcation curves (Fig. 2d); curves in the parameter space that separated regions in which the model presented qualitatively different dynamics (sound patterns). Only one region (Fig. 2d) corresponded to oscillatory behaviour in which labial oscillations resulted in sound-pressure fluctuations. Two features of the pressure–tension trajectories resulting in sound output were apparent (Fig. 2d). First, 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; and second, 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

[1]Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57th Street, Chicago, Ilinois 60637, USA. [2]Department of Physics, FCEN, University of Buenos Aires, Intendente Guiraldes 2160, Pabellon 1, Buenos Aires 1428, Argentina. †Present address: Department of Physics, FCEN, University of Buenos Aires, Intendente Guiraldes 2160, Pabellon 1, Buenos Aires 1428, Argentina.

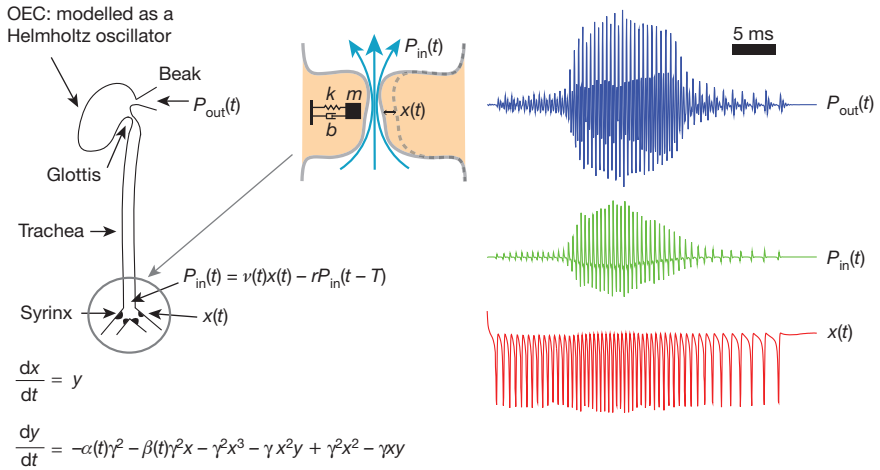

Figure 1 | Schematized view of a dynamical systems model describing syringeal labial dynamics and tracheal vocal-tract filtering. The syringeal membrane was modelled as a mass (m) with damping (b) and a restitution (spring) force (k). Normal form equations for labial position ($x(t)$, red line) were integrated, computing the input pressure at the vocal tract ($P_{in}(t)$, green line), filtered by the oro-oesophageal cavity (OEC), and ultimately the total output pressure ($P_{out}(t)$, blue line). γ, time constant; r , reflection coefficient of the trachea; T, propagation time along trachea; ν, proportional to the mean velocity of the flow; y, velocity (see Methods).

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 syllables[19], but our recordings were limited here to airborne sounds. In a sample of 8 modelled 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 lasting less than 30 ms, and it had a long tail corresponding to slowly varying sounds, such as constant-frequency harmonic stacks (Fig. 2e).

This simple model captured the essential features of sound production in a framework of labial tension and subsyringeal pressure over which birds have direct motor control[20–22]. Although the syrinx has considerable additional complexity, the model described the vocal system in a low-dimensional space, enabling us to capture a wide range of acoustic features using a small set of time-dependent parameters.

We tested the model by comparing responses of HVC neurons to the broadcast of the modelled song (mBOS) with the 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 difficult to induce a response: for example, mBOS generated without the OEC component failed to elicit responses. On one occasion we mis-estimated the duration of a component of BOS by 5 ms, and found one neuron that responded strongly to the BOS but not at all to the synthetic song (Supplementary Fig. 2b). Over a population of 30 neurons, the best mBOS elicited 58% ± 8% of the response to BOS (Supplementary Note 1). Phasic projection neurons (HVC_p; $n = 15$) and tonic interneurons (HVC_i; $n = 15$) both 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 next evaluated the activity of HVC neurons relative to model dynamics, by analysing the timing of spike bursting relative to the pressure–tension trajectories used to synthesize the mBOS. This identified a compelling relationship 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 colours) is shown relative to the BOS spectrograph, oscillograph and reconstructed pressure and tension time series. One of the neurons burst once, at the transition between descending frequency modulations and a constant frequency 'high note'. The other neuron burst twice, once when the pressure during a high note reached a maximum, and once at the transition between a high frequency chevron and a broadband frequency-modulated sound. Similar relationships between spike-burst timing and gestures were seen 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 propose that this pattern might arise if the bursts are associated with mini-breaths during singing[19]. Only the 17 bursts occurring during phonation were considered for further analysis.

Examination of the responses of the HVC_p using pressure versus tension plots showed 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. In cases in which the absolute maximum is not distinct in time, no additional GTE result (for example, if there are multiple local maxima with the same magnitude). Of the 17 bursts (14 HVC_p), 11 (65%) were aligned with onsets or 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 GTE in total). From a total of 60 gestures, 20 (33.4%) had only onset and offset GTE; in addition, 30 (50%) had a unique peak in pressure (3 GTE per gesture), 5 (8.3%) had a unique peak in tension (3 GTE per gesture); and 5 (8.3%) had 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, and 66% of the intervals had a duration of less than or equal to 30 ms (Fig. 4c). Most gestures corresponded to notes (the smallest unit of song organization recognized by ornithologists), but motor activity at GTE maxima could subdivide notes; for example, at the point at which 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 song[5]. We also observed cases in which HVC_p burst at the onset of relatively pure pressure-only or tension-only trajectories, with a tendency towards pressure-only trajectories (Fig. 2d). If such neurons project to distinct regions of the HVC's afferent targets, which are organized based on the syringeal muscles and interactions with respiratory system, such observation could help to resolve the long-standing riddle of the HVC's topographic organization.

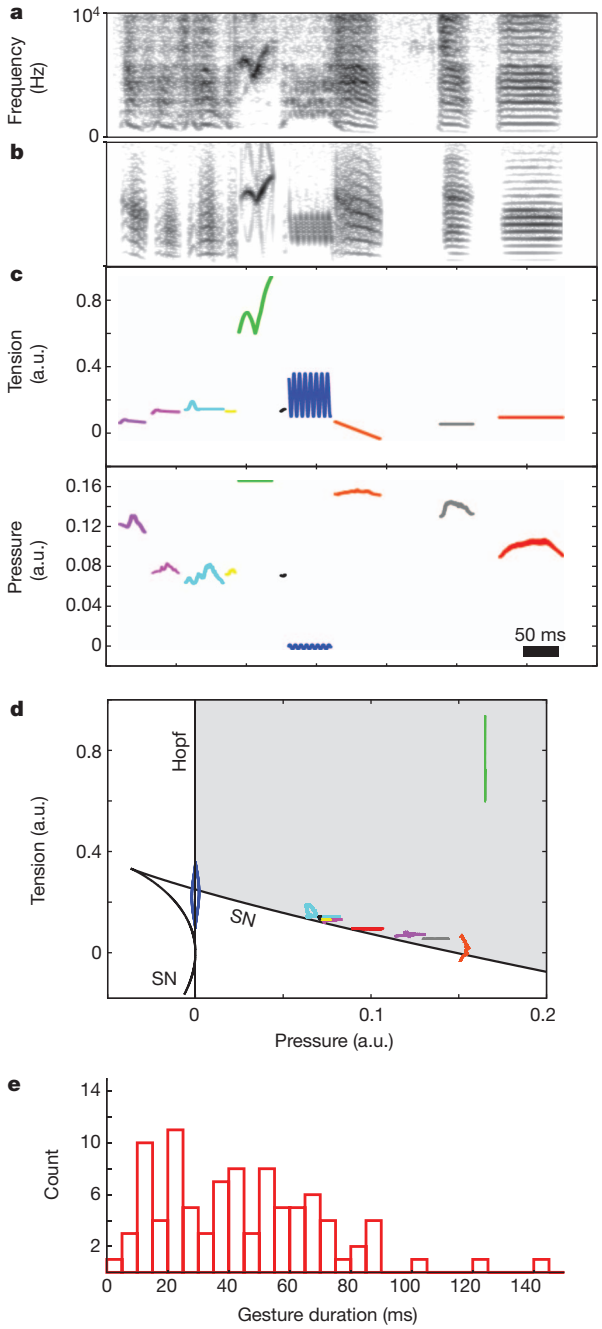

Figure 2 | A low-dimensional model for reconstructing gestures. **a**–**e**, Spectrographs of a bird's song (**a**) and a model synthetic song (**b**). Song is described by fitted parameters $\alpha(t)$ and $\beta(t)$, proportional to air-sac pressure and labial tension, respectively (**c**; each distinct sound is colour coded). Each sound (same colour code as in **c**) is generated by a continuous curve in the parameter space of the model, that we define as a 'gesture' (**d**). Oscillations in the vicinity of a Hopf bifurcation are almost tonal, whereas oscillations near a saddle-node (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 grey area indicates the region of phonation. The distribution of gesture durations for five birds is displayed in **e**. a.u., arbitrary units.

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 and, on average, the bursts preceded the closest GTE (mean $= -5.6 \pm 0.3$ 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 ms[23], the minimal 15-ms delay for auditory feedback to HVC[8], and that the duration of intervals between GTE varied greatly (Fig. 4c), it is notable that the timing of HVC_p bursting was synchronized with a near-zero time lag to a model of actual behavioural output.

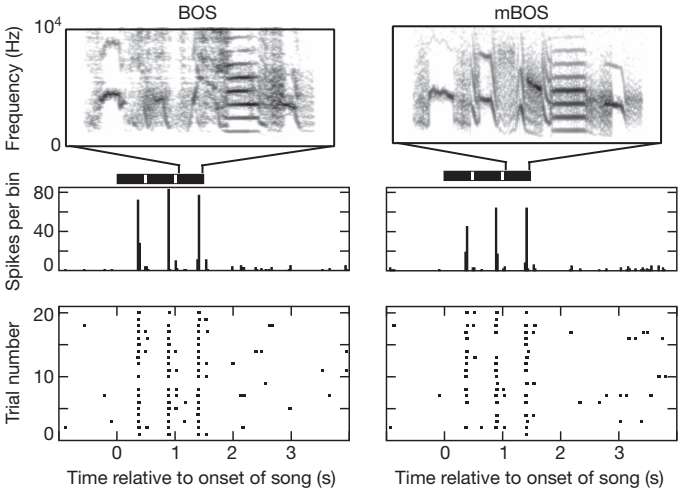

Figure 3 | Testing the low-dimensional model. The activity of HVC-selective neurons of sleeping birds in response to the presentation of BOS and mBOS was similar. The timing of the three repeated motifs that were presented is indicated by the bold horizontal lines.

Interneurons are suppressed at GTE

HVC_i show local minima and maxima in their tonic activity throughout song, and we noted that there was a relationship between the minima 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; an example neuron is shown in 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 relationship 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 $= -0.82$ ms $\pm$ 0.60 ms, $\sigma = 7.3 \pm 1.4$ ms; Fig. 5b). We compared 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 timing of minima relative to GTE for one of four birds (Supplementary Note 3). Thus, the precise activity of HVC_i (ref. 7) can help to shape the timing of HVC_p. This suggests a simple model in which bursts of activity of HVC_p suppress activity in HVC_i, whose ongoing activity helps to 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 enquire whether the activity of HVC neurons are also locked to gesture transitions during singing. Previous results have shown similar patterns of spike bursts between the daytime singing activity and the auditory-driven responses during sleep of single neurons of the robust nucleus of the arcopallium (RA) in zebra finches[4]. HVC neurons in awake swamp sparrows and Bengalese finches, which respond to auditory stimulation, also show auditory–motor correspondance in their activity patterns[6]. However, similar

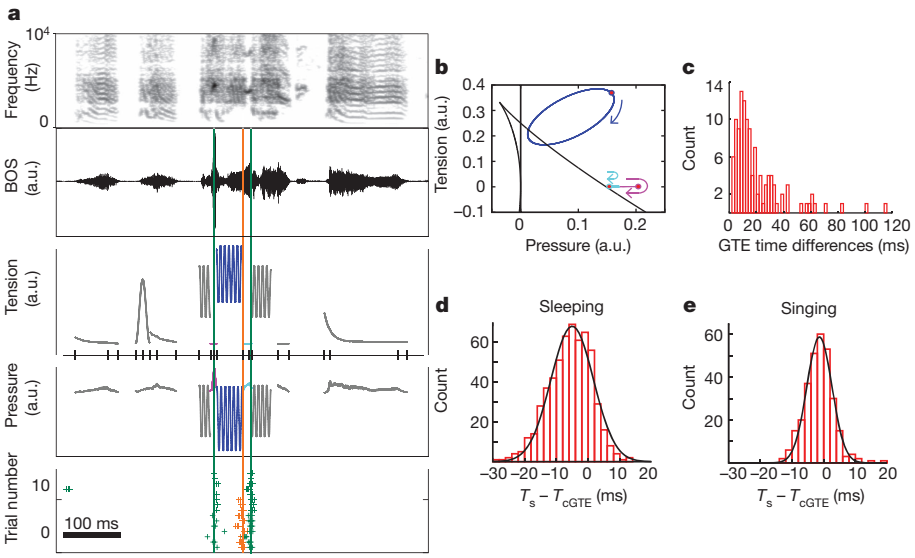

Figure 4 | Timing of gestures relative to bursting of projection neurons. **a**, Song spectrograph and oscillograph (top two panels); reconstructed parameters of pressure and tension (third and fourth panels; the gestures of interest are coloured magenta, blue and cyan), with tick marks indicating the times of all GTE; and raster plots of the responses of two neurons (bottom panel, coloured dark green and orange), together with their closest GTE, indicated with lines of the same colours. **b**, The trajectories in pressure–tension parameter space, with a point indicating the mean position of a burst, and arrows indicating the trajectory direction (the movement may follow and return along the same pressure–tension path, hence the curved arrows). Colours as in part **a**. **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 (T_{cGTE}) in sleeping birds ($n = 14$ HVC_p, 5 birds). **e**, The same analysis of **d** but for singing birds ($n = 5$ HVC_p, 2 birds).

observations have yet to be reported for zebra finch HVC neurons. We made recordings from HVC in singing birds ($n = 3$ birds), including ten tonic neurons, and five phasic neurons that burst during phonation, recorded in two of the three birds (Fig. 6 and Supplementary Fig. 5); one phasic neuron had two bursts per motif. These recordings confirmed that during singing, all sparse bursts of HVC_p occurred at gesture transitions (Fig. 4e). The same analysis that was used in sleeping birds was then used for singing birds (except that as each motif of song could vary, it was independently modelled here), and this showed that the timing of HVC_p for singing birds was even more precise than during sleeping (see Fig. 4d, e). The Gaussian fit for the population of phasic neurons recorded during singing (mean = -1.35 ms $\pm$ 0.10 ms, $\sigma = 4.0 \pm 0.1$ 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 GTE (Gaussian fit for the minima: mean = -0.12 ms $\pm$ 0.4 ms, $\sigma = 4.0 \pm 0.4$ ms; Fig. 5c), and this was significantly different to the bootstrapped random distribution (F-test, $P < 0.002$). Additional analyses showed significant locking of minima to GTE in two out 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, examination of data from a previous study of zebra finches[24] showed that during singing the timing of HVC_{RA} bursts were closely associated with the timing of the bursts of the other class of HVC projection neurons, which project to Area X (HVC_X), the basal ganglia component of the song system (Supplementary Fig. 7). Our results support 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.

It was thought previously that the timing of song syllables was unrelated to the timing of HVC_p discharge in singing birds[5,24]. Given the sparse bursting of these cells, this led to the idea that the output of HVC had a clock-like function with a nearly uniform 'tick' size of approximately 10 ms (ref. 23) supported by a 'syn-fire' chain of synaptic activity across HVC_p (ref. 5). Instead, we find that the

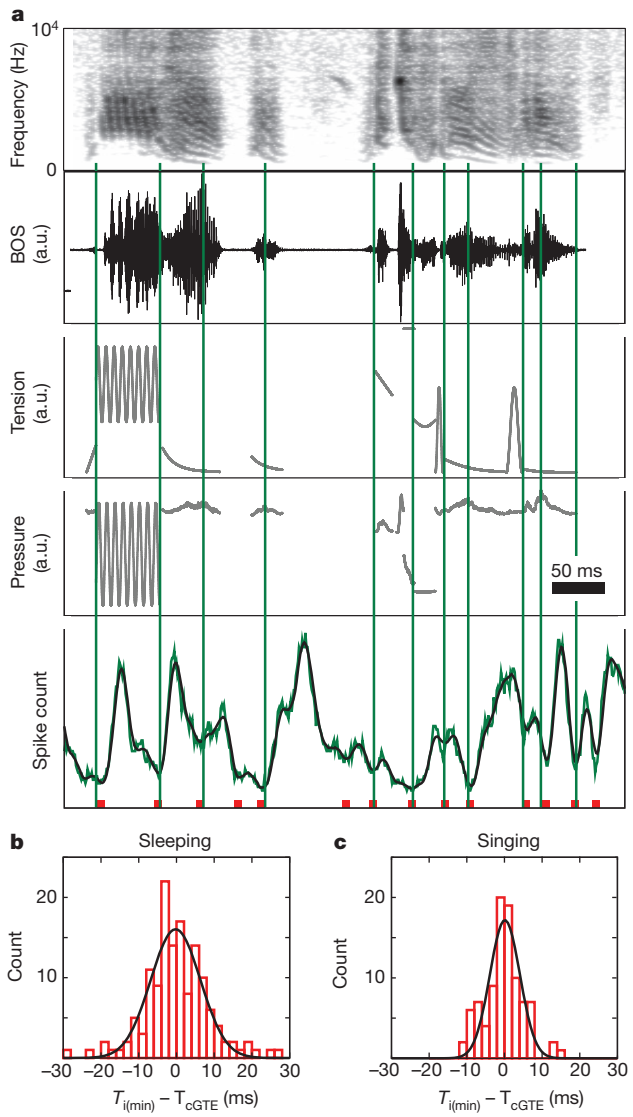

Figure 5 | Suppressed interneuron activity is associated with GTE. **a**, Song spectrograph and oscillograph, reconstructed parameters of pressure and tension, and raster plots, organized as in Fig. 4a, but with spike count response to the song (10-ms bin, 20 repetitions; green line) for one HVC_i, and 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 indicate the position of the closest GTE to each minima. **b**, distribution of time differences between spike response minima ($T_{i(min)}$) and their closest GTE (T_{cGTE}) in sleeping birds ($n = 15$ HVC_i, 5 birds). **c**, Same analysis as **b** but for singing birds ($n = 10$ HVC_i, 3 birds).

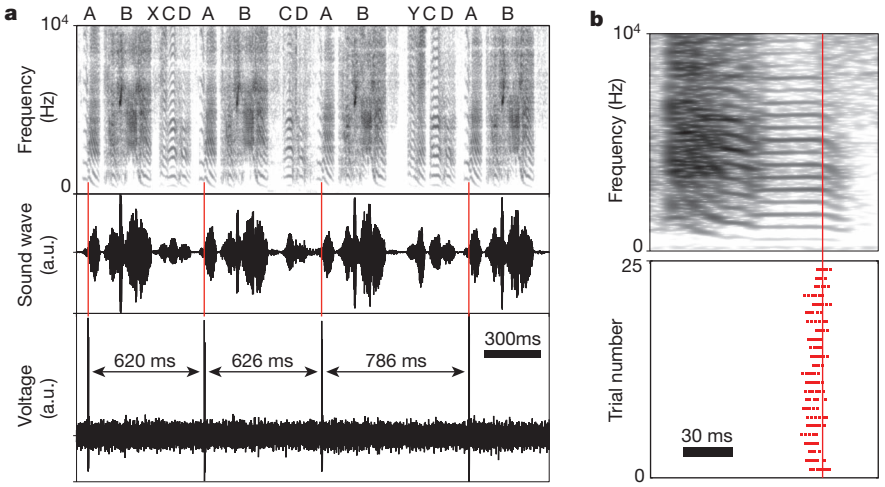

Figure 6 | During singing, HVC projection neurons fired in the vicinity of GTE. a, An HVC_p neuron with spike bursting that is time-locked to the vicinity of a GTE, even as the syllable sequence and time-interval varies. Letters over the sonograph identify the syllable types. **b**, For another bird, the bursts of a HVC_p neuron are locked to a GTE in the vicinity of a subtle acoustic transition.

bursting of HVC_p and modulation of HVC_i activity is timed to particular time points of motor gestures. The sequential firing across the population of HVC_p unfolds in an ordered fashion[5], 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 GTE this is inconsistent with a syn-fire network that is active at every moment. The differences between these two models of HVC have additional broad implications for the functional organization of the song system, for song learning and for motor coding.

As gestures vary greatly in duration, and as the RA only has access to the times of GTE, downstream components of the motor pathway (the RA and presumably the brainstem) should generate independent dynamical information to sustain the detailed structure within each gesture (compare with refs 23, 25). Previous experiments—including examinations of the effects of electrical stimulation of HVC or RA during singing[26], and lesions of nuclei afferent to the HVC[27]—implicate information in the HVC in the encoding of larger units of song. This might be explained, for example, if gestures at the start of syllables are overemphasized in HVC relative to gestures encoding mini-breaths preceding syllables. Finally, gestures are learned, and this is consistent with the physiological properties of HVC neurons: integration over hundreds of milliseconds and multiple syllables, nonlinear summation over syllables in a sequence preceding the excitatory response, and selective response to BOS[4,8,9,28–30]. The information on groupings of gestures, such as syllables, can be carried in these integrated signals. This also re-emphasizes that synaptic modification in HVC, not just changes at HVC–RA synapses, are associated with feedback-mediated sensorimotor learning (compare with ref. 23). The 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 vocal motor control

If activity in the HVC is synchronized, 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 auditory[33], proprioceptive[20] feedback. This allows movements to be represented in the HVC by gestures of greatly varying duration (with dynamics generated principally through internal HVC interactions), and with each gesture referenced to a common time framework for evaluating feedback (with feedback arriving through distinct, extrinsic inputs). This suggests that projection neurons represent a prediction of the actual behavioural 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 control[35]. Assuming that behaviour is subdivided into gestures, and that only the transitions (GTE) are represented by HVC output (HVC_p), feedback information could accumulate in the intervals between the transitions 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 RA[36], thalamic input carrying brainstem respiratory, auditory and proprioceptive information[21,34,37], and forebrain auditory input[38].

We have described song organization based on gestures, using the dynamical systems modelling framework to replace analysis of songs based on spectrographs. These features of motor systems organization may be represented in other systems and for other behaviors[39]. 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 behaviour that could have similar mechanisms to those described here. The development of corresponding models for human speech production should help to provide insight into speech and language pathologies in which sequential behaviour is disrupted.

METHODS SUMMARY

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 HVC extracellular recordings using standard techniques; by implanting a head pin (auditory experiments)[10] or motorized microdrive (singing experiments)[5]. Recordings were post-processed with a spike-sorting algorithm (Klusters, http://klusters.sourceforge.net) to separate the times of spike events for each unit. In sleeping birds, we picked the last (second or third) motif, which gave the strongest response, to analyse the timing of spikes relative to GTE. This minimized false peaks and troughs in the response profiles. The average response of each interneuron (1-ms resolution) was smoothed using a Savitsky Golay filter (polynomial local regression) and the minima were identified using a 21-point sliding window.

Full Methods and any associated references are available in the online version of the paper.

Received 28 February 2012; accepted 31 January 2013.
Published online 27 February 2013.

1. Hatsopoulos, N. G., Xu, Q. & Amit, Y. Encoding of movement fragments in the motor cortex. *J. Neurosci.* **27,** 5105–5114 (2007).

2. Nishikawa, K. *et al.* Neuromechanics: an integrative approach for understanding motor control. *Integr. Comp. Biol.* **47,** 16–54 (2007).
3. Perl, Y. S., Arneodo, E. M., Amador, A., Goller, F. & Mindlin, G. B. Reconstruction of physiological instructions from Zebra finch song. *Phys. Rev. E* **84,** 051909 (2011).
4. Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290,** 812–816 (2000).
5. Hahnloser, R. H. R., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419,** 65–70 (2002).
6. Prather, J. F., Peters, S., Nowicki, S. & Mooney, R. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* **451,** 305–310 (2008).
7. Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273,** 1871–1875 (1996).
8. Margoliash, D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* **3,** 1039–1057 (1983).
9. Margoliash, D. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.* **6,** 1643–1661 (1986).
10. Shank, S. S. & Margoliash, D. Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature* **458,** 73–77 (2009).
11. Amador, A., Goller, F. & Mindlin, G. B. Frequency modulation during song in a suboscine does not require vocal muscles. *J. Neurophysiol.* **99,** 2383–2389 (2008).
12. Elemans, C. P. H., Laje, R., Mindlin, G. B. & Goller, F. Smooth operator: avoidance of subharmonic bifurcations through mechanical mechanisms simplifies song motor control in adult zebra finches. *J. Neurosci.* **30,** 13246–13253 (2010).
13. Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* **395,** 67–71 (1998).
14. Mindlin, G. B. & Laje, R. *The Physics of Birdsong*. (Springer Verlag, 2005).
15. Laje, R., Gardner, T. J. & Mindlin, G. B. Neuromuscular control of vocalizations in birdsong: A model. *Phys. Rev. E* **65,** 05192 (2002).
16. Sitt, J. D., Amador, A., Goller, F. & Mindlin, G. B. Dynamical origin of spectrally rich vocalizations in birdsong. *Phys. Rev. E* **78,** 011905 (2008).
17. Amador, A. & Mindlin, G. B. Beyond harmonic sounds in a simple model for birdsong production. *Chaos* **18,** 043123 (2008).
18. Riede, T., Suthers, R. A., Fletcher, N. H. & Blevins, W. E. Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl Acad. Sci. USA* **103,** 5543–5548 (2006).
19. Hartley, R. S. & Suthers, R. A. Air-flow and pressure during canary song: direct evidence for mini-breaths. *J. Comp. Physiol. A* **165,** 15–26 (1989).
20. Suthers, R. A., Goller, F. & Wild, J. M. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proc. Natl Acad. Sci. USA* **99,** 5680–5685 (2002).
21. Wild, J. M. Functional neuroanatomy of the sensorimotor control of singing. *Ann. NY Acad. Sci.* **1016,** 438–462 (2004).
22. Suthers, R. A., Goller, F. & Pytte, C. The neuromuscular control of birdsong. *Phil. Trans. R. Soc. B* **354,** 927–939 (1999).
23. Fee, M. S., Kozhevnikov, A. A. & Hahnloser, R. H. Neural mechanisms of vocal sequence generation in the songbird. *Ann. NY Acad. Sci.* **1016,** 153–170 (2004).
24. Kozhevnikov, A. A. & Fee, M. S. Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97,** 4271–4283 (2007).
25. Fiete, I. R., Hahnloser, R. H. R., Fee, M. S. & Seung, H. S. Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *J. Neurophysiol.* **92,** 2274–2282 (2004).
26. Vu, E. T., Mazurek, M. E. & Kuo, Y. C. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14,** 6924–6934 (1994).
27. Williams, H. & Vicario, D. S. Temporal patterning of song production: Participation of nucleus uvaeformis of the thalamus. *J. Neurobiol.* **24,** 903–912 (1993).
28. Margoliash, D. & Fortune, E. S. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci.* **12,** 4309–4326 (1992).
29. Nick, T. A. & Konishi, M. Neural auditory selectivity develops in parallel with song. *J. Neurobiol.* **62,** 469–481 (2005).
30. Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S. & Mooney, R. Neural correlates of categorical perception in learned vocal communication. *Nature Neurosci.* **12,** 221–228 (2009).
31. Brainard, M. S. & Doupe, A. J. Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404,** 762–766 (2000).
32. Ölveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3,** e153 (2005).
33. Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol.* **22,** 770–783 (1965).
34. Ashmore, R. C., Wild, J. M. & Schmidt, M. F. Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J. Neurosci.* **25,** 8543–8554 (2005).
35. Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. An internal model for sensorimotor integration. *Science* **269,** 1880–1882 (1995).
36. Roberts, T. F., Klein, M. E., Kubke, M. F., Wild, J. M. & Mooney, R. Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *J. Neurosci.* **28,** 3479–3489 (2008).
37. Coleman, M. J., Roy, A., Wild, J. M. & Mooney, R. Thalamic gating of auditory responses in telencephalic song control nuclei. *J. Neurosci.* **27,** 10024–10036 (2007).
38. Bauer, E. E. *et al.* A synaptic basis for auditory-vocal integration in the songbird. *J. Neurosci.* **28,** 1509–1522 (2008).
39. Mulliken, G. H., Musallam, S. & Andersen, R. A. Forward estimation of movement state in posterior parietal cortex. *Proc. Natl Acad. Sci. USA* **105,** 8170–8177 (2008).
40. Leyton, S. S. & Sherrington, C. S. Observations on the excitable cortex of the chimpanzee, orangutan and gorilla. *Q. J. Exp. Physiol.* **11,** 135–222 (1917).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. H. R. Hahnloser for help with the microdrives and techniques used to record from singing birds. We thank H. D. I. Abarbanel, T. Q. Gentner, H. C. Nusbaum and S. E. Palmer for valuable comments on the manuscript. This work was supported by a Human Frontiers Science Program cross-disciplinary fellowship award to A.A., NIDCD006876, ANCyT, CONICET and UBA awards to G.B.M. and Y.S.P., and NIDCD and NSF/CRCNS awards to D.M.

Author Contributions A.A., G.B.M. and Y.S.P. developed the syringeal model, G.B.M. and Y.S.P. modelled the songs, A.A. conducted surgeries, sound recordings and collected the electrophysiological data, A.A., G.B.M. and D.M. conceived and designed the experiments, and prepared the manuscript. All four authors participated in data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.M. (dan@bigbird.uchicago.edu).

METHODS

Subjects, songs and surgeries. All procedures were carried out 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 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 h–8 h reversed light cycle in sound-isolation boxes. Songs were recorded and filtered using custom software) and then these were edited using Praat (http://www.praat.org). Edited songs included two or three repetitions of one motif, and were typically 2–4 s in duration. Birds were allowed to recover for 2 or 3 days before the first day of recordings, and were rested for at least 2 days between recording sessions.

Electrophysiology, stimulus presentation and spike analysis. HVC extracellular recordings were carried out using head-fixed sleeping or singing tethered birds. Recordings were post-processed with a spike-sorting algorithm (http://klusters.sourceforge.net) 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 score[25]:

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

where μ_S is the mean response during the auditory stimulus (S) and μ_{BG} is the mean response during background activity (BG) (Covar is covariance, Var is variance). The denominator of the equation is the standard deviation of (S – BG). The background was estimated by averaging the firing rate during a 2-s period. The Z scores of the mBOS (model BOS), CON (conspecific song) and REV (reversed BOS) were normalized to the BOS Z score, and averages across neurons were reported as means of normalized responses ± s.e.m. For interneurons, the strength of the response varied across the motifs[41]. To analyse the timing of spikes relative to GTE, we picked the last (second or third) motif, which gave the strongest response. 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 regression[42]) 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 bird song production[14]. This model assumes that for airflow values above a certain threshold, the labia start to oscillate with a wavelike motion. Assuming that two basic modes are 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:

$$\frac{dx}{dt} = y$$

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

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, where a_{lab} is the labial area. The average pressure, p_{av}, can be written in terms of the displacement and its velocity[3]. 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 derived[43], constitutes the simplest set of equations for any model in which oscillations arise in either of these two bifurcations. Once this reduction is carried out, the selection of parameters that enables a sound with specific acoustic features to be obtained gives rise to unique values. The normal form equations are shown in Fig. 1, and display the same set of dynamical regimes[3] 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 $P_{in}(t) = v(t)x(t) - rP_{in}(t - T)$ where T is the time for a sound wave to reach the end of the tube and return, and $v(t)$ is proportional to the average mean velocity of the flow. The transmitted pressure fluctuation $P_{tr}(t) = (1 - r)P_{in}(t - 0.5T)$ forces the air in the glottis, which is approximated by the neck of a Helmholtz resonator (used to model the OEC[3,44]); that is, a large container with a hole, such that the air in its vicinity oscillates owing 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 device[3], 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 $\Delta t = 1/20{,}000$ 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 carried out so that the SCI of the synthetic sounds matched the value of the song segment[3]. 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; next, the fundamental frequency was computed as the inverse of the time difference between the next two consecutive selected maxima; after this, 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 model[16]. With those estimations of fundamental frequency and SCI, $\alpha(t)$ and $\beta(t)$ were computed. In general, brief segments were found to be fast trills. We modelled these as rapid oscillations of pressure and tension, with the amplitude of the pressure oscillations such that the maxima fell in the phonating region, and with the amplitude of the tension oscillations such that the frequency range of the vocalization was reproduced. We found that most of the parameters could be approximated well by fractions of sine functions, exponential decays, constants or a combinations of these.

Use of these analytic functions as parameters of the model to generate a synthetic copy of the recorded song resulted in a noiseless surrogate song (for example, Supplementary Fig. 1 ('Noise = 0')). The addition of noise allowed the gradual recovery of realistic timbre features. The dimensionless variable 'Noise' varied between 0 and 40, with Noise = 5 corresponding to a fluctuation size equal to 2.5% of the maximum range of the $\beta(t)$ parameter. Note that the effect of timbre is 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 cm (ref. 45). Typically, zebra finch songs have a third important resonance at approximately 4 kHz. The parameters of the Helmholtz resonator were adjusted so that its resonant frequency would account for this resonance[3]. The synthetic songs for sleeping birds were generated before the electrophysiological experiments were carried out. For singing birds, all song reconstructions were also performed blind to the spike data.

41. Sutter, M. L. & Margoliash, D. Global synchronous response to autogenous song in zebra finch HVC. *J. Neurophysiol.* **72,** 2105–2123 (1994).
42. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* 3rd edn (Cambridge Univ. Press, 2007).
43. Guckenheimer, J. & Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* (Springer Verlag, 1997).
44. Fletcher, N. H., Riede, T. & Suthers, R. A. Model for vocalization by a bird with distensible vocal cavity and open beak. *J. Acoust. Soc. Am.* **119,** 1005–1011 (2006).
45. Daley, M. & Goller, F. Tracheal length changes during zebra finch song and their possible role in upper vocal tract filtering. *J. Neurobiol.* **59,** 319–330 (2004).