

Neural processing of auditory feedback during vocal practice in a songbird

Georg B. Keller¹ & Richard H. R. Hahnloser¹

Songbirds are capable of vocal learning and communication^{1,2} and are ideally suited to the study of neural mechanisms of complex sensory and motor processing. Vocal communication in a noisy bird colony and vocal learning of a specific song template both require the ability to monitor auditory feedback^{3,4} to distinguish self-generated vocalizations from external sounds and to identify mismatches between the developing song and a memorized template acquired from a tutor⁵. However, neurons that respond to auditory feedback from vocal output have not been found in song-control areas despite intensive searching^{6–8}. Here we investigate feedback processing outside the traditional song system, in single auditory forebrain neurons of juvenile zebra finches that were in a late developmental stage of song learning. Overall, we found similarity of spike responses during singing and during playback of the bird's own song, with song responses commonly leading by a few milliseconds. However, brief time-locked acoustic perturbations of auditory feedback revealed complex sensitivity that could not be predicted from passive playback responses. Some neurons that responded to playback perturbations did not respond to song perturbations, which is reminiscent of sensory-motor mirror neurons^{8,9}. By contrast, some neurons were highly feedback sensitive in that they responded vigorously to song perturbations, but not to unperturbed songs or perturbed playback. These findings suggest that a computational function of forebrain auditory areas may be to detect errors between actual feedback and mirrored feedback deriving from an internal model of the bird's own song or that of its tutor. Such feedback-sensitive spikes could constitute the key signals that trigger adaptive motor responses to song disruptions^{10,11} or reinforce exploratory motor gestures for vocal learning¹².

The field L region and the caudolateral mesopallium (CLM) are interconnected brain areas not part of the traditional song-control system and are analogous to the auditory cortex in mammals in that they receive the main stream of auditory input from the thalamus, as well as feedback from motor-related areas^{13–17}. Neurons in field L and CLM of awake and anaesthetized animals respond robustly to a large variety of auditory stimuli such as white noise, the bird's own song (BOS), and conspecific songs^{18–21}. These features make field L and CLM potential substrates for the integration of self-generated and external sounds and for monitoring singing-related auditory feedback. To explore this hypothesis, we made extracellular recordings from CLM and field L neurons in juvenile male zebra finches using chronically implanted miniature motorized microdrives (Supplementary Fig. 1). Our strategy was first to probe singing-related firing in these neurons for evidence of motor-specific processing beyond passive auditory responses elicited by playback of the BOS, and then to investigate the feedback sensitivity of singing-related spikes by delivering brief acoustic stimuli during singing.

Singing and playback-related firing was similar in most cells (Fig. 1a), despite the large differences in sound amplitudes in vocal

and playback conditions and despite the variable direction of the playback source relative to the bird's head. Average firing rates in vocal and playback conditions were typically above baseline rates (Fig. 1b) and were highly correlated ($r = 0.77$, $P < 10^{-10}$, $n = 92$; Supplementary Figs 2 and 3). Firing rates were also well matched

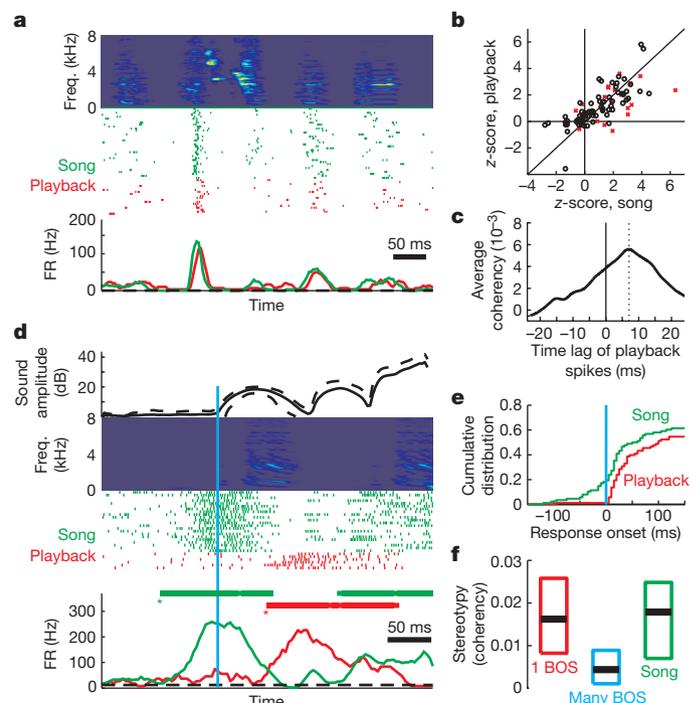


Figure 1 | Comparison of active responses with passive responses. Active responses were similar to passive responses, but were more stereotyped and slightly anticipatory. **a**, Activity during song (green rasters) leads activity during BOS playback (red rasters) in this example neuron. Top: spectrogram of an example song motif (high sound amplitudes in red and low amplitudes in blue). Bottom: average firing rate (FR) curves. **b**, Scatter plot of firing rates (z-scores) in individual cells. Cells were stimulated with one version (black circles) or many versions (red crosses) of the BOS. **c**, The mean coherency function between singing- and playback-related spike trains peaked 6.8 ms (dotted line) after singing-related spikes. **d**, The spike rasters in this neuron are anticipatory to song onset (blue line) and delayed relative to playback onset. Significant deviations from baseline firing are marked by green and red horizontal bars, and onset times are indicated by asterisks. Top: mean sound amplitude (solid line) plus/minus standard deviation (dashed line). **e**, Cumulative distribution of response onset times. **f**, Median firing stereotypy (black bars) during song ($n = 92$) is similar to stereotypy during playback of a single BOS ($n = 24$) but higher than stereotypy during playback of many versions of the BOS ($n = 68$). Second and third quartiles are shown by coloured boxes.

¹Institute of Neuroinformatics, University of Zurich/ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

on a finer timescale. In 69 of 92 cells, the spike-coherency function^{22,23} averaged over all pairings of song and playback trials displayed a significant peak (>2 jackknife standard deviations above zero) within ± 20 ms. The peak of the mean coherency function (averaged over 92 cells) was significant and occurred 6.8 ms after singing-related spikes (Fig. 1c), indicating that, overall, singing-related activity slightly preceded playback-evoked responses (only 0.5 ms of this lag can be explained by the closer proximity of the sound source to the bird's ears during singing). This anticipatory behaviour of singing-related activity suggests that in addition to auditory inputs, cells in field L and CLM also received inputs from a vocal-related, non-auditory source. Consistent with this view, in roughly one-fifth of the cells we observed firing increases before onset of the first introductory note of a song bout (Fig. 1d, e and Supplementary Fig. 4), which demonstrates a source of non-auditory drive in these cells during singing. Hence, some neurons seemed to integrate auditory with non-auditory signals, of which the latter may have reflected information about song-motor activity, for example as part of a motor estimate of auditory feedback.

Motor-specific processing was also evident by analysis of firing stereotypy, which was higher during singing than during playback of different versions of the BOS (Wilcoxon rank-sum (WRS) test, $P < 10^{-8}$; Fig. 1f). This stereotypy difference could be attributed neither to intrinsic differences between song and playback stimuli (because the latter were copies of the former) nor to differences in average firing rates (Supplementary Fig. 3). By contrast, the firing stereotypy during singing did not differ significantly from the stereotypy during playback of one unique BOS stimulus (WRS test,

$P = 0.42$). Thus, singing-related firing stereotypy was higher than predicted by passive responses, but was commensurate with intrinsic synaptic noise, suggesting that auditory responses may be partly subsumed during singing by a motor-specific source of stereotyped synaptic input.

Brief acoustic stimuli delivered during singing provide an effective means of operant conditioning of song features¹¹. Such stimuli are thus ideally suited to probing auditory feedback sensitivity. In 50% of song motifs (randomly selected) we applied a brief perturbing stimulus through a second loudspeaker that was time locked to a given syllable (see Methods, Fig. 2a and Supplementary Figs 5 and 6). In agreement with previous reports on adult birds⁶, we found that feedback perturbations did not induce immediate spectral or temporal changes in vocal output (all analyses were restricted to song motifs with conserved syllable sequences; see Supplementary Figs 7–9 and Methods).

We quantified the propensity of cells to respond to perturbations either in the vocal or the playback condition in terms of a response bias b confined to values between $b = -1$ (no perturbation response during song) and $b = 1$ (no perturbation response during playback). Similarly, we quantified the sensitivity of responses to perturbed feedback in terms of a selectivity s that was normalized to $s = 0$ if the firing did not change during feedback perturbations and to $s = 1$ if the firing doubled during perturbations (see Methods). Overall, 66 of 67 cells significantly responded to feedback or playback perturbations (54 of 67 to feedback, and 53 of 67 to playback; see Methods). Many cells responded robustly to perturbations in both conditions (Fig. 2b). However, almost 20% of cells had a strong playback bias

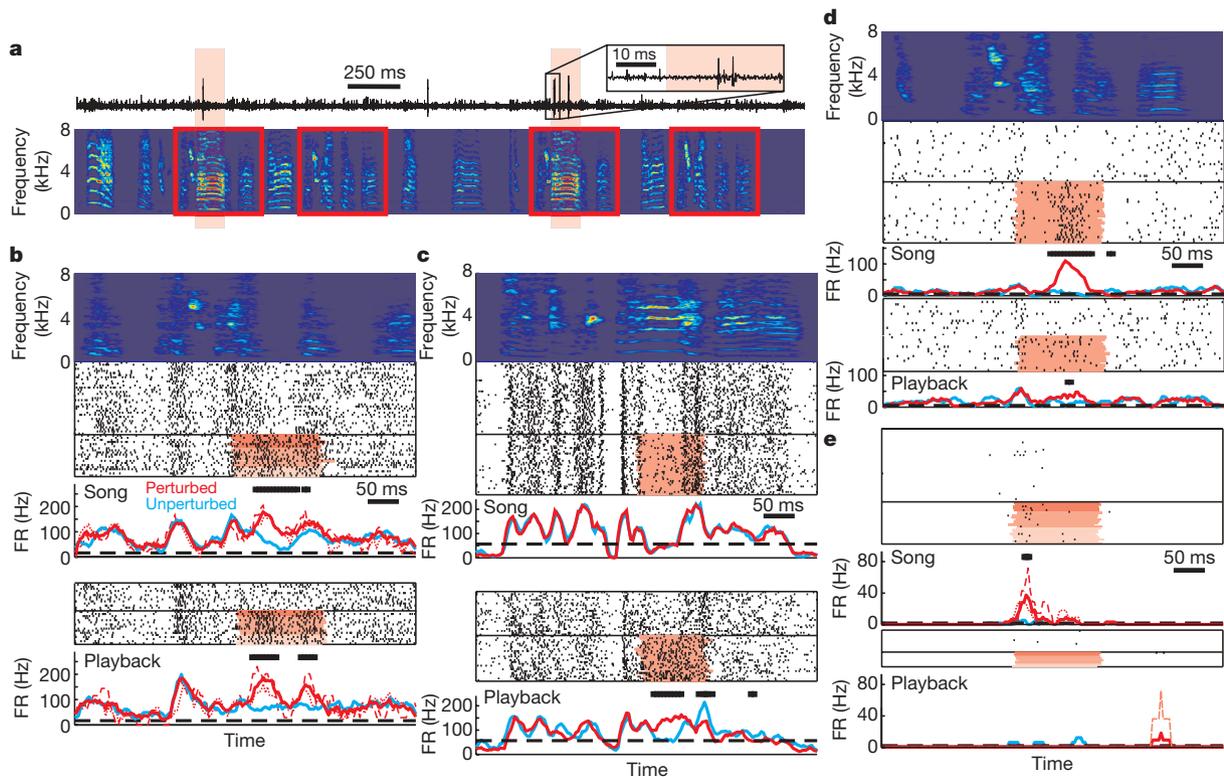


Figure 2 | Example perturbation responses. **a**, Bottom: spectrogram of a song that was twice perturbed by a long-call stimulus (red shading). Song motifs are delimited by red boxes. Top: extracellular voltage trace of a simultaneously recorded neuron (inset, spike burst). **b**, Neuron with increased firing during perturbation of song (top) and of BOS playback (bottom). Perturbing stimuli are indicated by red shaded areas; lighter shading corresponds to lower sound amplitudes. Average firing rate curves are shown for cases without perturbation (blue line) and with perturbation (solid red line; dashed and dotted red lines are averages over trials with high

and low perturbation amplitudes, respectively). Also shown are average spontaneous firing rate (dashed horizontal lines) and the times of significant perturbation responses (black horizontal bars). $b = 0.0$, $s = 0.42$. **c**, As in **b**, but for a neuron that responds to perturbations only in the playback, but not the vocal condition. $b = -0.95$, $s = 0.0$. **d**, **e**, Two feedback-sensitive neurons in the same bird with selective ($b = 0.38$, $s = 1.38$; **d**) and highly selective ($b = 0.38$, $s = 7.11$; **e**) responses for song perturbations (**e** shows the same cell as in **a**).

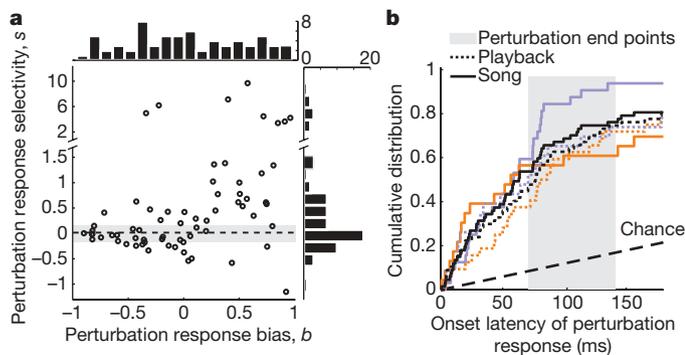


Figure 3 | Summary of perturbation selectivity. **a**, Scatter plot of bias, b , versus selectivity, s ($n = 67$ cells, circles), the marginal distributions (histograms, showing number of cells), and the median chance selectivity (dashed line) with first and third quartiles (grey shading; see Methods). **b**, Cumulative distributions of perturbation-response onset latencies during song (solid lines) and playback (dotted lines). Latencies were widely distributed, both across all birds (black lines) and in two individual birds (orange and purple lines). The estimated contribution caused by random fluctuations is shown by the dashed line (labelled 'Chance'; see Methods), and the range of perturbation end points by the grey shaded area.

and small selectivity for distorted feedback ($b < -0.5$ and $|s| < 0.5$, $n = 12$); these cells responded to playback perturbation, but did not respond to even high-intensity song perturbation (Fig. 2c). About 10% of cells showed very high selectivity for distorted feedback ($s > 3$, $n = 8$), and many of them tended to be quite unresponsive to perturbed playback (Fig. 2d, e and Supplementary Fig. 10a). Feedback perturbations predominantly induced firing increases ($s > 0$), although we observed firing suppression in a few cells (Fig. 3a and Supplementary Fig. 10b, c). In both conditions, sound amplitudes during perturbation responses were significantly lower than average (Student's t -test, $P < 10^{-10}$), in agreement with a monotonic relation between perturbation amplitude and response selectivity s (Supplementary Fig. 11). However, onset times of perturbation responses were widely distributed across cells (Fig. 3b), revealing a cell-specific temporal modulation of perturbation sensitivity.

In six of eight cells with selectivity $s > 3$, peak firing rates during perturbed song were more than three times higher than peak firing rates during unperturbed song (in the remaining two cells, peak firing rates were more than 40% higher). Thus, responses to perturbed feedback could largely exceed all of unperturbed singing-related activity, suggesting that high selectivity for distorted auditory feedback derives from a precisely timed and strongly coherent synaptic drive.

We also explored whether high selectivity was associated with suppression of neural activity during self-initiated vocalizations, because such suppression is a common gain-control mechanism found in the auditory brain areas of animals as diverse as crickets²⁴, bats¹ and marmosets²⁵. The baseline firing in highly selective cells ($s > 3$) was lower than for all other cells (5.7 Hz versus 19.5 Hz, $P = 0.028$; Supplementary Fig. 12a); nevertheless, highly selective cells were relatively more suppressed during singing (WRS test of equal median z -scores, $P = 0.005$; Supplementary Fig. 12b).

In conclusion, field L and CLM responses equally reflect processing of self-generated and external auditory inputs, as made evident by the similarity of average firing rates in active and passive conditions (Fig. 1b and Supplementary Fig. 3), by the uniformly distributed bias index (Fig. 3a) and by similarity of onset-latency curves (Fig. 3b). Hence, the auditory forebrain seems to form an invariant representation of actively and passively perceived songs for integrating and comparing auditory feedback with the songs of other birds. Dissimilarities between active and passive sound processing were evident in terms of a motor-specific drive. Consequently, some neurons showed singing-related activity that resembled playback-evoked

activity but was insensitive to perturbed feedback. Such behaviour is reminiscent of auditory-vocal mirroring reported in HVC⁸ neurons and could arise from corollary discharges elicited by an efference copy of motor commands. On the other hand, neurons that were largely quiescent during singing, except when the auditory signal was perturbed, are reminiscent of some neurons in primate auditory cortex that strongly respond to frequency-shifted auditory feedback²⁶. Vocal-mirror spikes could contribute to the generation of highly perturbation-selective responses, provided that such spikes are able to precisely counterbalance the excitatory drive elicited by sensory feedback. We find some evidence for such counterbalancing in perturbation-selective neurons in terms of their relatively strong firing suppression during song, even though suppression was rare overall, unlike in monkey auditory cortex²⁵ and in an auditory ganglion of crickets²⁴ (note that in crickets, responses are not perturbation selective, despite this suppression).

Relatively few cells specialized into highly selective perturbation detectors, yet their mere existence suggests that auditory feedback is analysed in the auditory forebrain with reference to an internal model (Supplementary Fig. 13). For example, vocal mirror responses could represent predicted auditory feedback, which helps the bird to generate a stable perception of its song in the midst of a noisy colony. Accordingly, highly feedback-sensitive responses would reflect prediction errors of auditory feedback; such errors could signal song disruptions or simplify vocal learning, according to some forward-model theories²⁷. Alternatively, given that the birds in our study were in the process of learning a tutor song, vocal mirroring could constitute online replay of the tutor memory, as evidenced by similar firing stereotypies during song and during playback of a single song template (Fig. 1f). Similarity of vocal and playback-related firing could thus be a reflection of a good match between the actual song and the memorized auditory template, of which the latter may feed into CLM and field L through the caudomedial nidopallium^{28,29}. According to such a template-replay interpretation, responses in highly perturbation-selective neurons represent performance errors signalling the dissimilarity between the perturbed song and the tutor memory, a property with obvious benefits for song learning^{12,27}. The ability of birds to correlate perturbations with subtle motor variability¹¹ suggests a functional connection between perturbation-selective neurons and premotor neurons, although a direct link between perturbation selectivity and song learning remains to be observed.

METHODS SUMMARY

Subjects and electrophysiology. All experiments were carried out in accordance with protocols approved by the Veterinary Office of the Canton of Zurich, Switzerland. Data were collected from six juvenile male zebra finches (60–92 days old). Microdrives with three independent electrodes were implanted 1.1–2.1 mm anterior and 1.5–2.0 mm lateral of the midsagittal sinus (under a fixed head angle of 70° or 90°) using methods previously described³⁰. After each experiment, the brain was removed for histological examination of unstained slices to verify the location of reference lesions. Neurons were classified as putative CLM or putative field L based on anatomical location¹⁵.

Data analysis. Differences between average firing rates in unperturbed and perturbed trials were assessed using a WRS test ($P = 0.05$) on the number of spikes in 30-ms time windows. Windows were shifted in 5-ms steps and only sequences of at least two subsequent windows with $P < 0.05$ were considered significantly different.

We defined the response bias as $b = (|d_f| - |d_p|) / (|d_f| + |d_p|)$, where d_f is the feedback perturbation response and d_p is the playback perturbation response, each defined as the difference in average firing rates between perturbed and unperturbed conditions in a time window extending from the time of onset of perturbation up to 50 ms after perturbation ends. We defined the response selectivity as $s = d_f / r_s$, where r_s is the average firing rate during unperturbed song motifs.

The stereotypies of singing and playback-related spike patterns were assessed using the average coherency of spike rasters at zero time lag (sound traces were aligned by the amplitude onset of the detected syllable). Differences between singing and playback-related firing stereotypies were detected using a WRS test (same BOS, $P = 0.3$; different BOS, $P < 10^{-7}$).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 May; accepted 29 September 2008.

Published online 12 November 2008.

- Marler, P. & Tamura, M. Culturally transmitted patterns of vocal behavior in sparrows. *Science* **146**, 1483–1486 (1964).
- Thorpe, W. H. The biological significance of duetting and antiphonal song. *Acta Neurobiol. Exp. (Wars.)* **35**, 517–528 (1975).
- Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
- Brainard, M. S. & Doupe, A. J. Auditory feedback in learning and maintenance of vocal behaviour. *Nature Rev. Neurosci.* **1**, 31–40 (2000).
- Immelman, K. in *Bird Vocalizations* (ed. Hinde, R. A.) 61–74 (Cambridge Univ. Press, 1969).
- Leonardo, A. Experimental test of the birdsong error-correction model. *Proc. Natl Acad. Sci. USA* **101**, 16935–16940 (2004).
- Kozhevnikov, A. A. & Fee, M. S. Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97**, 4271–4283 (2007).
- Prather, J. F., Peters, S., Nowicki, S. & Mooney, R. Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature* **451**, 305–310 (2008).
- Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004).
- Leonardo, A. & Konishi, M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399**, 466–470 (1999).
- Tumer, E. C. & Brainard, M. S. Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* **450**, 1240–1244 (2007).
- Fiete, I. R., Fee, M. S. & Seung, H. S. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J. Neurophysiol.* **98**, 2038–2057 (2007).
- Karten, H. J. The ascending auditory pathway in the pigeon (*Columba livia*). II. Telencephalic projections of the nucleus ovoidalis thalami. *Brain Res.* **11**, 134–153 (1968).
- Nottebohm, F., Kelley, D. B. & Paton, J. A. Connections of vocal control nuclei in the canary telencephalon. *J. Comp. Neurol.* **207**, 344–357 (1982).
- Vates, G. E., Broome, B. M., Mello, C. V. & Nottebohm, F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J. Comp. Neurol.* **366**, 613–642 (1996).
- Foster, E. F. & Bottjer, S. W. Axonal connections of the high vocal center and surrounding cortical regions in juvenile and adult male zebra finches. *J. Comp. Neurol.* **397**, 118–138 (1998).
- Zeng, S., Zhang, X., Peng, W. & Zuo, M. Immunohistochemistry and neural connectivity of the Ov shell in the songbird and their evolutionary implications. *J. Comp. Neurol.* **470**, 192–209 (2004).
- Cardin, J. A. & Schmidt, M. F. Song system auditory responses are stable and highly tuned during sedation, rapidly modulated and unselective during wakefulness, and suppressed by arousal. *J. Neurophysiol.* **90**, 2884–2899 (2003).
- Bauer, E. E. *et al.* A synaptic basis for auditory-vocal integration in the songbird. *J. Neurosci.* **28**, 1509–1522 (2008).
- Janata, P. & Margoliash, D. Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *J. Neurosci.* **19**, 5108–5118 (1999).
- Amin, N., Grace, J. A. & Theunissen, F. E. Neural response to bird’s own song and tutor song in the zebra finch field L and caudal mesopallium. *J. Comp. Physiol. A* **190**, 469–489 (2004).
- Kimpo, R. R., Theunissen, F. E. & Doupe, A. J. Propagation of correlated activity through multiple stages of a neural circuit. *J. Neurosci.* **23**, 5750–5761 (2003).
- Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. Sleep-related neural activity in a premotor and a basal-ganglia pathway of the songbird. *J. Neurophysiol.* **96**, 794–812 (2006).
- Poulet, J. F. & Hedwig, B. The cellular basis of a corollary discharge. *Science* **311**, 518–522 (2006).
- Eliades, S. J. & Wang, X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* **89**, 2194–2207 (2003).
- Eliades, S. J. & Wang, X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**, 1102–1106 (2008).
- Jordan, M. I. & Rumelhart, D. E. Forward models: Supervised learning with a distal teacher. *Cogn. Sci.* **16**, 307–354 (1992).
- Phan, M. L., Pytte, C. L. & Vicario, D. S. Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc. Natl Acad. Sci. USA* **103**, 1088–1093 (2006).
- London, S. E. & Clayton, D. F. Functional identification of sensory mechanisms required for developmental song learning. *Nature Neurosci.* **11**, 579–586 (2008).
- Fee, M. S. & Leonardo, A. Miniature motorized microdrive and commutator system for chronic neural recording in small animals. *J. Neurosci. Methods* **112**, 83–94 (2001).

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

Acknowledgements This work was supported by a Swiss National Science Foundation Professorship Grant to R.H.R.H. We thank K. Martin, R. Mooney and R. Douglas for comments on the manuscript.

Author Contributions G.B.K. performed the experiments and analysed the data. Both authors wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.H.R.H. (rich@ini.phys.ethz.ch).

METHODS

Electrophysiology. To maximize the experimental yield, microdrives were equipped with a lateral positioner that allowed for electrode displacement parallel to the brain surface. We recorded from 92 cells during song and playback of the BOS at physiological sound levels at which birds hear each other's songs; in 67 of these we also performed perturbation experiments.

Experimental design. During recording sessions, birds were housed in a sound isolation chamber equipped with a microphone and two loudspeakers (for playback of the BOS and perturbing stimuli) in a triangular arrangement around the centre of the experimental cage (distance, 25 cm). For each bird, the sound amplitudes of BOS stimuli as recorded with the microphone were matched to those of songs.

Sound amplitudes of perturbations were 6 dB above the maximum sound amplitude of BOS stimuli (both measured at the microphone). For each cell, we first determined the perturbing stimulus that elicited the largest response when presented in isolation and then used this stimulus for the rest of the experiment. If none of the stimuli triggered any noticeable response, we used the long-call stimulus. Perturbing stimuli were white noise (3 out of 67 cells), an introductory note (5 out of 67 cells) or a long call (59 out of 67 cells) (Supplementary Fig. 5). The microphone signal was fed to a custom-made, real-time song recognizer that detected the first stereotypic syllable of song motifs using a two-layer neural network trained on spectrotemporal song data. In 50% of detections, a perturbation was delivered. Offline analysis revealed that perturbation onsets had a standard deviation of 4.3 ms relative to syllable onset (aligned by threshold crossing of sound amplitude). Because our subjects were juvenile birds with developing songs, we recorded BOS stimuli less than one day before each experimental session. A set of 10–20 songs was used for BOS playback experiments. Perturbations were confined to recording sessions only (lasting roughly 0.5–2 hours each) to minimize possible effects of song deterioration. BOS playbacks were interspersed with songs as much as possible; we ignored all data in which birds countersang with song playback.

Data analysis. Extracellular voltage traces were digitized at 33 kHz and recorded for offline spike sorting. Firing rates were calculated in 5-ms bins and averaged over multiple renditions or playbacks of the motif. In addition, firing rate curves in Figs 1a, d, 2b–e and Supplementary Figs 4 and 10 were smoothed with a 10-ms Gaussian kernel.

In the plot of cumulative distributions of response-onset times in Fig. 3b, we depicted the estimated chance-related distribution with a dashed black line. The slope of this distribution corresponded to the fraction of neurons with significant firing rate differences between perturbed and unperturbed trials in intervals 50–150 ms prior to perturbation.

To examine dependences of the response bias, b , and selectivity, s , on perturbation (sound) amplitudes, we occasionally presented the perturbing stimulus at sound pressure levels differing by ± 6 dB from normal levels ($n = 40$ cells; Fig. 2a and Supplementary Fig. 11). Excluding either the high- or low-intensity trials led to only small changes in b and s (median absolute change in b , 0.16; median absolute change in s , 0.16).

To estimate the fraction of the response selectivity that was caused by small sample sizes but not by perturbations, we computed a chance response selectivity $s_c = d_c / r_s$, where d_c is the difference in average firing rates between perturbed and unperturbed conditions in a time window preceding the perturbation window. The median, as well as the second and third quartiles of s_c are shown in Fig. 3a by the dashed line and the grey shaded surface, respectively.

To assess changes from background firing, we calculated the z -score of song/playback responses as

$$z = \frac{\mu_s - \mu_b}{\sqrt{\sigma_s^2 + \sigma_b^2 - 2\text{cov}(S, B)}}$$

where μ_s is the average firing rate during song/playback, μ_b is the average baseline firing rate assessed in 1–3-s silent intervals interleaved between playbacks and songs, and σ_s and σ_b are the respective standard deviations. Owing to a lack of pairing of vocal and playback conditions, we set the covariance term $\text{cov}(S, B)$ to 0. We considered cells with a z -score higher than 0.75 to be significantly excited, and cells with a z -score less than -0.75 to be significantly inhibited (the criterion $|z\text{-score}| > 0.75$ corresponds to $P < 0.05$ if the number of trials is larger than five, which was the case for all our cells).

The dashed lines in the scatter plot of average firing rates in song and playback conditions (Supplementary Fig. 3) are linear regressions (passing through the origin) for two sets of neurons: (1) neurons recorded during multiple playbacks of one unique song bout consisting of two motifs (red); (2) neurons recorded during playback of a set of 20 different song bouts (black).