RESEARCH ARTICLE

An Adapting Auditory-motor Feedback Loop Can Contribute to Generating Vocal Repetition

Jason D. Wittenbach[1�], Kristofer E. Bouchard[2,3�¤], Michael S. Brainard[2,4], Dezhe Z. Jin[1]*

1 Department of Physics and Center for Neural Engineering, the Pennsylvania State University, University Park, Pennsylvania, United States of America, 2 Department of Physiology and Center for Integrative Neuroscience, University of California at San Francisco, San Francisco, California, United States of America, 3 Department of Neurosurgery and Center for Neural Engineering and Prosthesis, University of California at San Francisco, San Francisco, California, United States of America, 4 Howard Hughes Medical Institute, San Francisco, California, United States of America

� These authors contributed equally to this work.
¤ Current address: Biological and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
* dzj2@psu.edu

Abstract

Consecutive repetition of actions is common in behavioral sequences. Although integration of sensory feedback with internal motor programs is important for sequence generation, if and how feedback contributes to repetitive actions is poorly understood. Here we study how auditory feedback contributes to generating repetitive syllable sequences in songbirds. We propose that auditory signals provide positive feedback to ongoing motor commands, but this influence decays as feedback weakens from response adaptation during syllable repetitions. Computational models show that this mechanism explains repeat distributions observed in Bengalese finch song. We experimentally confirmed two predictions of this mechanism in Bengalese finches: removal of auditory feedback by deafening reduces syllable repetitions; and neural responses to auditory playback of repeated syllable sequences gradually adapt in sensory-motor nucleus HVC. Together, our results implicate a positive auditory-feedback loop with adaptation in generating repetitive vocalizations, and suggest sensory adaptation is important for feedback control of motor sequences.

Author Summary

Repetitions are common in animal vocalizations. Songs of many songbirds contain syllables that repeat a variable number of times, with non-Markovian distributions of repeat counts. The neural mechanism underlying such syllable repetitions is unknown. In this work, we show that auditory feedback plays an important role in sustaining syllable repetitions in the Bengalese finch. Deafening reduces syllable repetitions and skews the repeat number distribution towards short repeats. These effects are explained with our computational model, which suggests that syllable repeats are initially sustained by auditory

Competing Interests: The authors have declared that no competing interests exist.

feedback to the neural networks that drive the syllable production. The feedback strength weakens as the syllable repeats, increasing the likelihood that the syllable repetition stops. Neural recordings confirm such adaptation of auditory feedback to the auditory-motor circuit in the Bengalese finch. Our results suggests that sensory feedback can directly impact repetitions in motor sequences, and may provide insights into neural mechanisms of speech disorders such as stuttering.

Introduction

Many complex behaviors—human speech, playing a piano, or birdsong—consist of a set of discrete actions that can be flexibly organized into variable sequences [1–3]. A feature of many variably sequenced behaviors is the occurrence of repetitive sub-sequences of the same action. Examples include trills in music, repeated syllables in birdsong, and syllable/sound repetitions in stuttered speech. A central issue in understanding how nervous systems generate complex sequences is the role of sensory feedback versus internal motor programs [4] (Fig 1a). At one extreme (the serial chaining framework), the sensory feedback from one action initiates the next action in the sequence; therefore sensory feedback is critical for sequencing the actions [5, 6]. However, because of the delays in both motor and sensory processing in nervous systems, it has been argued that a sequence generation mechanism relying solely on sensory feedback would be too slow to account for the execution of fast sequences such as typing and speech [1]. At the other extreme, sequences are generated by internal motor programs controlling sequence production without the use of sensory feedback [7–9]. However, there is ample evidence that sensory feedback can affect action sequences [10–14]. Despite the ubiquity of sequencing in behavior, the neural mechanisms of how sensory feedback interacts with internal motor programs to influence discrete actions remain largely unexplored.

Here, we study the role of sensory feedback in the production of repetitive vocal sequences using the Bengalese finch as a model system. The Bengalese finch produces songs composed of discrete acoustic events, termed syllables, organized into variable sequences (Fig 1b). However, sequence production is not random [15], as the transition probabilities between syllables are statistically reproducible across time [13, 16]. A prominent feature of the songs of several songbird species, including the Bengalese finch, is syllable repetition [15, 17–21] (e.g. 'b' in Fig 1b). For a given repeated syllable, the number of consecutively produced repeats (the repeat number) varies. The first order Markov process, in which the probability of repeating a syllable is constant, is a simple model for generating syllable repetitions. Such a process produces a monotonically decreasing distribution of repeat numbers, with the most probable repeat number (peak repeat number) being one (Fig 1c, black curve). Indeed, many repeated syllables in the songs of the Bengalese finch do have such distributions [20]. However, there are also repeated syllables that violate the predictions of the Markov process. These syllables are typically long repeated, and their distributions of repeat numbers are peaked, with the most probable repeat number being much greater than one [20–23] (Fig 1c, red curve). In the songs of the Bengalese finch, the transition probabilities between syllables are altered shortly after deafening [24, 25] or in real-time by delayed auditory feedback [13], demonstrating that disturbing auditory feedback can disturb sequence generation.

Songbirds are prominent models for studying the neural basis of complex sequence production. Experimental data from sensory-motor song nucleus HVC (proper name) of singing zebra finches have led to neural network models of the internal motor program for sequence generation that instantiate first-order Markov processes [26]. This suggests that additional
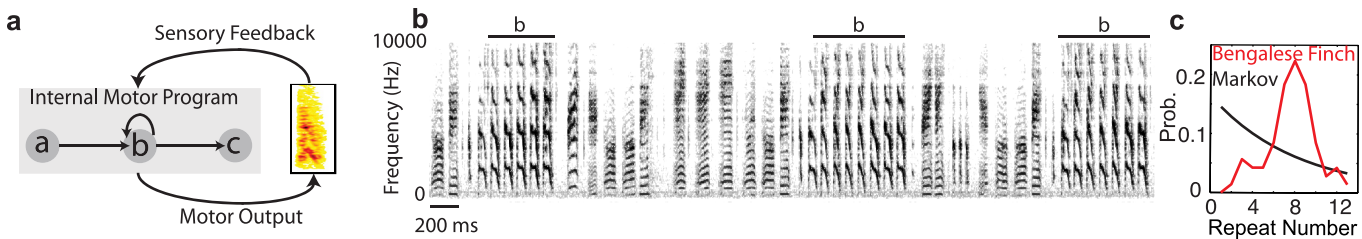
Fig 1. Bengalese finch song and the generation of sequences. a: Diagram of sensory-motor circuit for sequence generation. An internal motor program generates transitions between actions ('a', 'b', 'c', etc) while sensory feedback from the actions (motor outputs) impinges on the motor program. b: Example of Bengalese finch song. Spectrogram (power at frequency vs. time) of an adult Bengalese finch song, which consists of several syllables (denoted with letters) produced in probabilistic sequences. A prominent feature of Bengalese finch songs is the presence of syllable repetitions, some with long repeat sequences (e.g. syllable 'b'). c: Probability distribution of repeat counts for syllable 'b' from an individual Bengalese finch (red), and the predicted probability distribution for a Markov process using first order transition probabilities.

mechanisms contribute to the generation of non-Markovian distributions of repeat numbers [20, 21, 26]. One possibility is that, because of sensory-motor delays, auditory feedback from the previous syllable interacts with the internal motor program to contribute to the transition dynamics for subsequent syllables [13, 14, 27, 28]. For repeated syllables, we hypothesized that the interaction of auditory-feedback and ongoing motor activity forms a positive-feedback loop that contributes to sustaining syllable repetition beyond the predictions of a Markov process (Fig 1a). However, such positive-feedback architectures are inherently unstable, prone to indefinite repetition (i.e. perseveration). Across sensory modalities, a common feature of sensory responses to repeated presentations of identical physical stimuli is a gradual decrease of response magnitude (i.e. response adaptation) [29]. We therefore hypothesized that auditory inputs are subject to response adaptation, which gradually reduces the strength of the positive feedback loop over time. Thus, an auditory-motor feedback loop with response adaptation is predicted to contribute to the generation of non-Makovian repeated syllable sequences by both pushing repeat counts beyond the expectations of a Markov process and simultaneously preventing indefinite repetitions of the syllable. We tested these hypotheses using computational modeling combined with behavioral and electrophysiological experiments.

Results

A network model with adapting auditory feedback

The critical features of our framework for repeat generation are: (1) the population of neurons generating a repeated syllable receives a source of excitatory input in addition to the recurrent excitation from the sequencing network, and (2) the strength of this input adapts over time during repeat generation. For concreteness, we instantiate this framework as a 'branched-chain' network with adapting auditory feedback, and place this network in nucleus HVC. In songbirds, HVC has been proposed to contain an internal motor program for the generation of song sequences [26, 30–35]. HVC sends descending motor commands for song timing to nucleus RA (the robust nucleus of the arcopallium), which in turn projects to brainstem areas controlling the vocal organs [36, 37] (Fig 2a). HVC also receives input through internal feedback loops from the brainstem [38], via Uva (nucleus uvaeformis) and NIf (the interfacial nucleus of the nidopallium) [39]. Experiments in the zebra finch have shown sparse sequential firing of the RA projecting HVC neurons (HVC_{RA}) during singing [30, 31, 35]. This has led to the hypothesis that the motor program for sequence production in HVC includes sequential "chaining" of activity, in which populations of HVC_{RA} neurons responsible for generating a
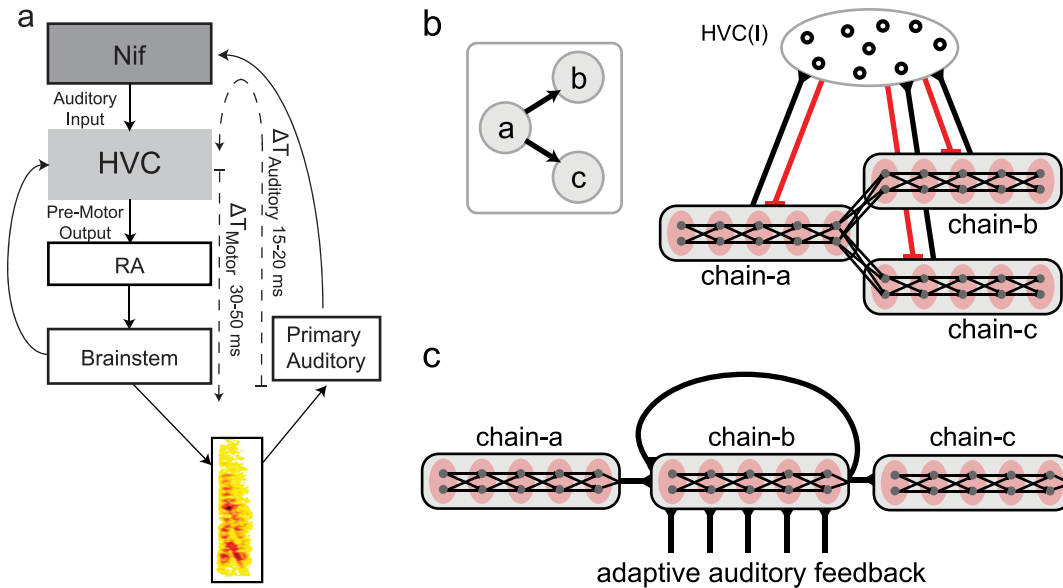
Fig 2. Avian song system and branched chain network with adapting auditory feedback. a: Diagram of the avian song system. HVC is a sensory-motor integration area that receives auditory input from high-level auditory nuclei such as NIf (nucleus interfacealus), and sends temporally precise motor controls signals to nucleus RA (robust nucleus of the arcopallium), which projects to the vocal brainstem areas. There is a pre-motor latency of 30–50 ms (ΔT Motor) between activity in HVC and subsequent vocalization. Additionally, there is a latency of 15–20 ms (ΔT Auditory) for auditory activity to reach HVC. This makes for a total auditory-motor latency between pre-motor activity and resulting auditory feedback of 45–70 ms. b: Example of a branch point in a probabilistic sequence (left). Syllable 'a' can transition to either syllable 'b' or 'c'. Such probabilistic sequences can be produced by a branched chain network (right). Here, each syllable is produced by a syllable-chain, in which groups of HVC_{RA} neurons (grey dots in red ovals, grouped in grey rectangles for a given syllable) are connected unidirectionally in a feed-forward chain (black lines with triangles are excitatory connections). The end of chain-a connects to the beginning of chain-b and chain-c. Spike activity propagates through chain-a and drives downstream neurons in RA to produce syllable a. At the end of chain-a, the activity continues to chain-b or chain-c via the branched connections. Only one syllable chain can be active at a time, as enforced by winner-take-all mechanisms mediated through local feedback inhibition from the HVC_I neurons (red lines are inhibitory connections). c: The branched chain network with adapting auditory feedback for generating repeating sequences of syllable 'b'. The end of chain-b reconnects to its beginning and to chain-c. Auditory feedback from syllable 'b' is applied to chain-b, and biases the repeat probability when the activity propagates to the branching point. The feedback is weakened as syllable 'b' repeats due to use-dependent synaptic depression which models stimulus-specific adaptation of the auditory signal.

doi:10.1371/journal.pcbi.1004471.g002

syllable drive the neuronal populations that generate subsequent syllables either directly within HVC or through the internal feedback loop [31, 34, 35, 40, 41] (Fig 2b).

Our model for generating syllable sequences starts with such a synaptic chain framework. The details of this model have been described previously [26] and are summarized in Materials and Methods. In synaptic chain models, each syllable is encoded in a chain network of HVC_{RA} neurons (Fig 2b). Spike propagation through the chain produces the encoded syllable by driving appropriate RA neurons. To generate variable syllable transitions, the syllable-chains are connected into branching patterns. At a branch point, syllable-chains compete with each other through a winner-take all mechanism mediated by the inhibitory HVC interneurons (HVC_I), allowing only one branch to continue the spike propagation. The selection is probabilistic due to intrinsic neuronal noise, which provides a source of stochasticity in the winner-take-all competition (Fig 2b). In this model, syllable repetition is generated by connecting the syllable-chains to themselves at the branching points [26, 34]. In branched chain networks, the transitions between the syllable-chains are largely Markovian, and for repeating syllables this implies that repeat number distributions should be a decreasing function of the repeat number—in particular, the most probable (or "peak") repeat number will be one [26] (Fig 1c). However, many repeated syllables in Bengalese finch song have repeat distributions that are highly non-

Markovian, with peak repeat numbers much larger than one [20–23]. This implies additional processes beyond synaptic chains contribute to generating non-Markovian repeated sequences.

Here we incorporate auditory feedback into the branching chain network model and show that, when this feedback is strong and adapting, non-Markovian repeat distributions emerge. In HVC, as in many sensory-motor systems, including the human speech system [42, 43], the same neuronal populations that are responsible for the generation of the behavior also respond to the sensory consequences of that behavior, i.e. the bird's own song (BOS) [14, 44–46]. HVC receives much of its auditory input from NIf [47–50], which can provide real-time auditory feedback during singing (Fig 2a) [51]. However, because of the time it takes to propagate motor commands to the periphery (30–50 ms) and process the subsequent auditory signals (15–20 ms) (Fig 2a), auditory feedback is necessarily delayed relative to the motor activity that generated it [1, 13, 14, 28]. This sensory-motor delay for HVC (45–70 ms) is on the order of the duration of a syllable, making it possible for auditory feedback to influence HVC motor programs and the transition dynamics between syllables [13, 14, 27] (Fig 2a).

We first tested the feasibility of this mechanism using biophysically detailed neural network models. To illustrate this model, we focus on generating sequences of the form 'ab^{n}c', where syllable 'a' transitions to syllable 'b', 'b' repeats a variable number of times (n), and transitions to 'c' (e.g. 'abbbbbbbc'). For concreteness, we model the adapting input as an auditory feedback signal to the network, though in principle this adapting input could reflect recurrent circuit-activity that is non-sensory. To incorporate auditory feedback into the previous model, each HVC_{RA} neuron in chain-b is contacted by excitatory synapses carrying auditory inputs triggered by the production of syllable 'b' (Fig 2c). We assume that the auditory synapses are made by axons from NIf, which is a major source of auditory input to HVC [47–50] and is selective to BOS [49]. When auditory feedback is present, the auditory synapses receive spikes from a Poisson process, assumed to be from the population of NIf neurons responding to syllable 'b' (Materials and Methods) (Fig 2c). The auditory synapses are subject to short-term synaptic depression, resulting in gradual adaptation of responses to repeated inputs [52, 53]. Specifically, due to the synaptic depression, the average strength of the auditory inputs to chain-b decreases exponentially during the repeats of syllable 'b' (Materials and Methods).

In Fig 3, we show results from an example network in which the auditory input to chain-b is strong and the spiking dynamics produce repeats of syllable 'b' with large repeat numbers. A spike raster for a standard single run of the network is shown in Fig 3a. Once spiking was initiated in chain-a (through external current injection), spikes propagated through chain-a, and activated chain-b. Chain-b repeated a variable number of times before the spike activity exited to chain-c and stopped once it reached the end of chain-c. As chain-b continued to repeat, the synapses carrying the feedback signal weakened over time due to adaptation (Fig 3b).

Analyzing multiple trials, we find that the probability of chain-b transitioning to itself (repeat probability) also decreases over time, though the repeat probability is only meaningful at the transition times—i.e. when the activity reaches the end of chain-b (Fig 3c). Examining the feedback strength at these transition times across the same trials allowed us to understand how the instantaneous feedback strength affects the repeat probability (Fig 3d). Not surprisingly, we found that the repeat probability increases with the strengths of the auditory synapses. Repeat probability p_r as a function of the feedback strength could be well fit with the sigmoidal function (Fig 3d, red curve)

$$p_r(A) = 1 - \frac{c}{1 + \eta A^v}, \tag{1}$$

where $A > 0$ represents the strength of the auditory synapses, $\eta, v > 0$ are parameters controlling the shape of the curve, and $0 < c < 1$ is a parameter for the repeat probability when there
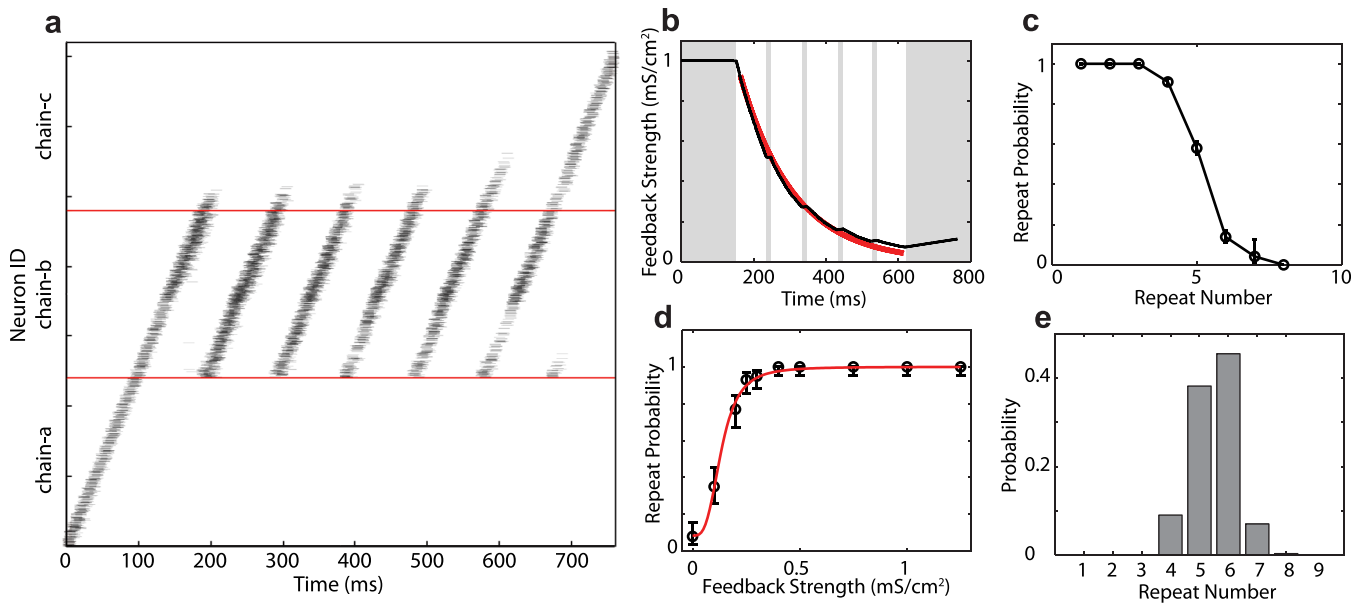
Fig 3. Strong, adapting auditory feedback produces peaked repeat distributions in branching chain neural networks. a. Raster plot of the spikes of the neurons in the network model. Neurons are ordered according their positions in the chains. Each dot represents the spike time of a neuron. Spikes are subsampled and the image is smoothed so that darker areas represent stronger activity at a particular location/time. Spikes propagate from chain-a to chain-b. Chain-b repeats a variable number of times (in this case, 6) before activity exits to chain-c. b. The average strength of auditory synapses decreases once they are activated by auditory feedback input. The red line is the fit to an exponential function with a decay time constant $\tau = 148$ ms. There is auditory feedback from syllable-b to HVC_{RA} neurons in chain-b during the times indicated by white areas. c. The probability of chain-b repeating decreases as the repeat number increases. d. The repeat probability of chain-b as a function of the average synaptic strength of the auditory inputs that the chain receives at the transition time. The bars are 90% confidence intervals (Wilson score with continuity correction). The red line is a fit with a sigmoidal function. e. The probability distribution of the repeat numbers of syllable 'b'. All probabilities computed over 1,000 simulations.

doi:10.1371/journal.pcbi.1004471.g003

is no auditory feedback (i.e. $A = 0$), which is determined by the connection strengths of the network at the branching point. Note that, when the auditory input $A = 0$, the repeat probability is $p_r = 1 - c$, and conversely, as A is large, p_r approaches 1.

Initially, the strong auditory feedback biases the network toward repeating and so the repeat probability is close to 1. If the strong excitatory input resulting from auditory feedback were constant, the network would perseverate on repeating syllable 'b' indefinitely (a result of the positive feedback loop). However, because of the short-term synaptic depression, the auditory input to chain-b when syllable 'b' repeats decreases exponentially over time (Fig 3b, red line; time-constant of $\tau = 148$ ms for this particular network). Even so, the repeat probability stays close to 1 as long as the auditory input is strong enough. Further weakening of the feedback reduces the repeat probability more significantly, making repeat-ending transitions to chain-c more likely. For this network, this process produced a repeat number distribution peaked at 6, as shown in Fig 3e. These results demonstrate that branched-chain networks receiving adapting excitatory inputs can generate repeat distributions that are non-Markovian.

Statistical model for the repeat number distributions

The repeat number distributions from our network model can be described using a simple statistical model with a small number of parameters. In our network model, the gradual reduction of excitatory drive from auditory feedback as a syllable is repeated reduces the probability that the syllable transitions to itself, and thus reduces the repeat probability. Eq (1) describes the

dependence of the repeat probability p_r on the auditory input strength, A. The synaptic depression model tells us how A changes with time. Sampling this at the transition times describes how A changes with the repeat number, n. At the end of the nth repeat of the syllable, A reduces to

$$A(n) = a_0 e^{-nT/\tau}, \tag{2}$$

where a_0 is the initial strength of the auditory feedback, τ is the time constant of the input decay, and T is the duration of the syllable. Combining this with the dependence of the repeat probability on A, shown in Eq (1), we find that the repeat probability after the nth repetition of the syllable is given by

$$p_r(n) = 1 - \frac{c}{1 + \eta a_0^v e^{-nvT/\tau}} = 1 - \frac{c}{1 + ab^n}, \tag{3}$$

where $a = \eta a_0^v$ and $b = e^{-vT/\tau}$. Therefore, there are effectively three parameters (a, b and c) for how p_r depends on n. We call Eq (3) the sigmoidal adaptation model of repeat probability.

The network sequence dynamics can be represented with a state transition model, in which a single state corresponds to the repeating chain. The state can transition to itself with a probability $p_r(n)$ given by Eq (3), or exit the state with probability $1 - p_r(n)$. This single state transition model can accurately fit the repeat number distributions generated by the network simulations with varying parameters, as shown in Fig 4a (all fit errors below their respective benchmark errors, which characterize the fitting errors expected from the finiteness of the data set—see Materials and Methods).

This model contains the Markov model and a previously described 'geometric adaptation' model [20] as special cases (Materials and Methods). Both of these models fail to fit the simulated data, even when a large number of states/parameters are used (Fig 4b and 4c). On the other hand, we have shown that the sigmoidal model provides an accurate fit with a single state and a small number of parameters. Therefore, relative to other statistical models, the single-state transition model with sigmoidal adaptation parsimoniously and accurately replicates the syllable repetition statistics of our network model.

Using the single state transition model with sigmoidal adaptation, we explored how peak repeat numbers depend on the initial feedback strength and the adaptation strength (defined by the related parameter, α, in the synaptic depression model, Materials and Methods) (Fig 4d). Here we see that, for a given adaptation strength, there is a threshold feedback strength at which the peak repeat number is greater than 1, and this threshold increases with increasing adaptation strength. This demarcates the transition between Markovian (peak repeat number = 1) and non-Markovian (peak repeat number > 1) repeat distributions (black-to-red transition in (Fig 4d)). Further increases in the feedback strength result in larger peak repeat numbers. Conversely, for a given feedback strength, increasing the adaptation strength results in a reduction of the peak repeat number. Together, these results demonstrate that a large range of peak repeat numbers can be generated through various combinations of feedback and adaptation strengths, and suggest that there is a threshold feedback strength required to produce non-Markovian repeat distributions.

Sigmoidal adaptation model fits diverse repeat number distributions of Bengalese finch songs

To see whether the non-Markovian repeat distributions generated with our network model can accurately describe syllable repeat number distributions of actual Bengalese finch songs, we recorded and analyzed the songs of 32 Bengalese finches. We identified the song syllables and
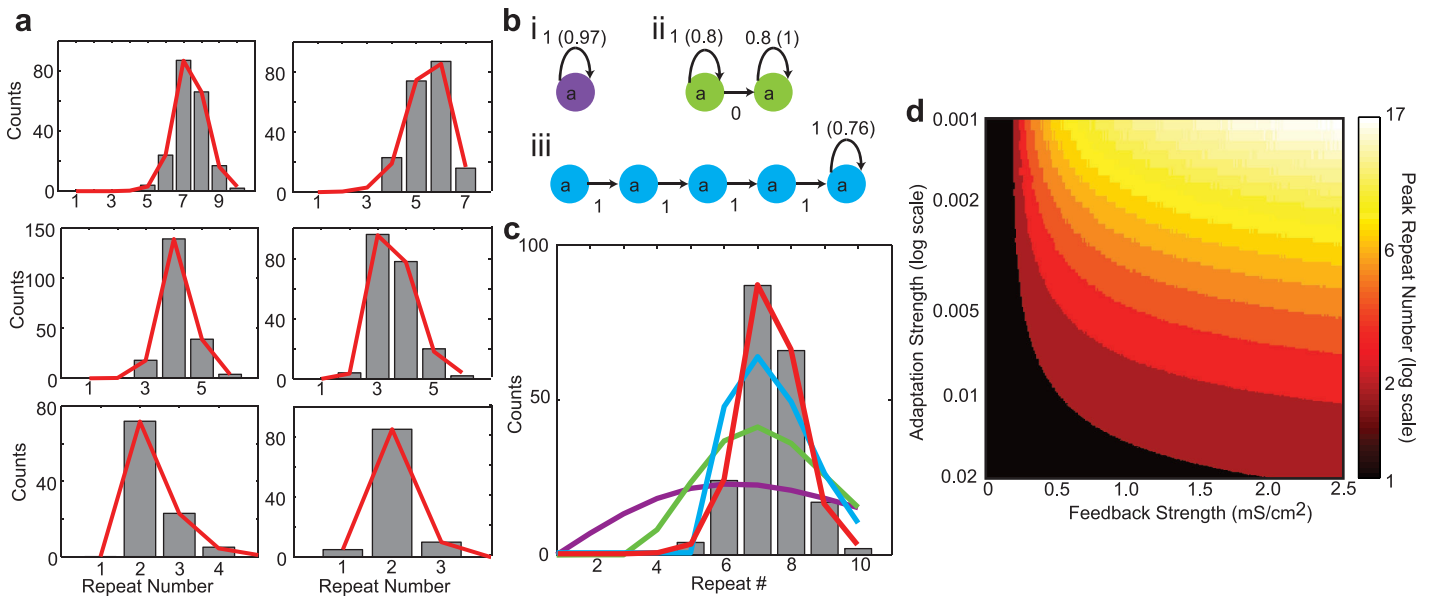
Fig 4. Sigmoidal adaptation model of repeats and model predictions. a: Six example repeat count histograms (black bars) from the neural network simulations with adapting auditory feedback and the best fit distributions from the sigmoidal adaptation model (red lines). The decay constants of the auditory feedback and the syllable lengths are varied to produce different repeat number distributions. Syllable lengths are changed by altering the number of groups per chain. All fit errors are smaller than benchmark errors. b: Best fit geometric adaptation models for the first histogram in a. With geometric adaptation, the probably of a state repeating is decreased by a constant factor with each consecutive repeat (Materials and Methods): (i) single state; (ii) two-states, both repeating; (iii) multiple-states, only final state repeating. In all cases, numbers on arrows are initial transition probabilities while the number in parenthesis is the constant adaptation factor. c: Comparison of model fits. Red is the best fit of the sigmoidal adaptation model with one state. Other colors are best fits of the corresponding models in b. The sigmoidal adaptation model provides a superior fit while only requiring a single state. d: Peak repeat number as a function of the initial auditory feedback strength and the adaptation strength generated using the sigmoidal adaptation model. The peak repeat number increases for increasing initial feedback strength and decreases for increasing adaptation strength. For a given adaptation strength, there is a threshold feedback strength at which repeat distributions become non-Markovian (i.e. peak repeat number > 1).

obtained the syllable sequences (Materials and Methods). Our data set contains more than 82,000 instances of 281 unique syllables, of which 71 are repeating syllables. Since the simulations of the network model are slow, we used the single state transition model with sigmoidal adaptation to fit the repeat number distributions for these syllables. As demonstrated above, the statistical model (Eq (3)) captures the essential features of our network model, and succinctly represents the repeat number distributions produced by the network simulations.

In Fig 5a, we show six examples of Bengalese finch repeat count histograms (grey bars) with different peak repeat counts (peak repeat count increases across plots i-vi.), and the best-fit model distributions (red lines). These examples show a range of distribution peaks and shapes, from small peak numbers with long rightward tails (i), to large peak numbers with tight, symmetric tails. Interestingly, we found that three repeated syllables (out of 71) had clear double-peaked distributions, with a prominent peak at repeat number 1 and another peak far away (two of which are displayed in panels ii and vi). These double peaked distributions cannot be explained with a single state transition model. A simple explanation is that the single peak and the broad peak are generated by two separate states (or neural substrate), as postulated in Jin & Kozhevinov (the "many-to-one mapping" from multiple chains in HVC to the same syllable type) [20]. Here we removed the single peak at repeat number 1 for these three syllables and only analyzed the longer repeat parts. The state transition with sigmoidal adaptation model does an excellent job of fitting the wide variety of peaks and shapes of the repeat distributions found in the Bengalese finches.
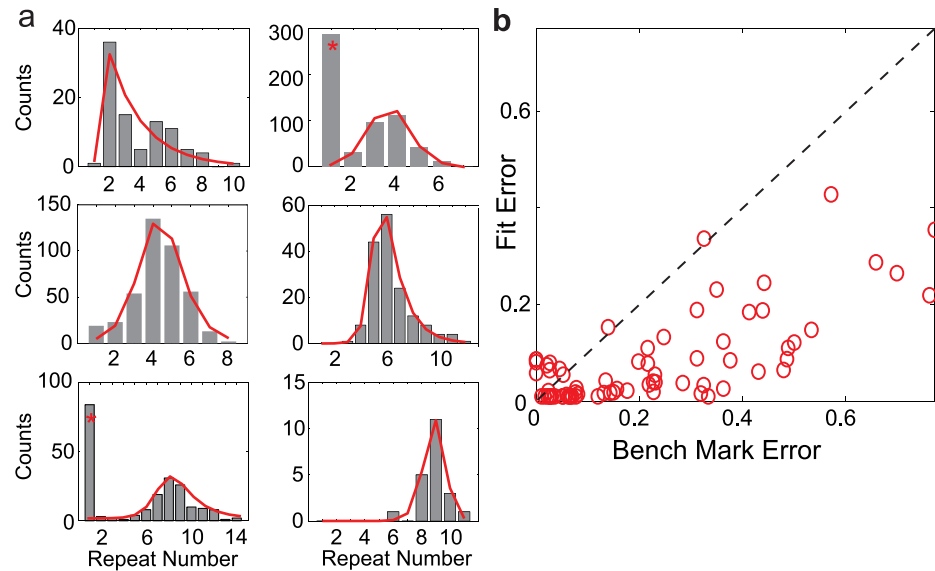
Fig 5. Sigmoidal adaptation model fits diverse repeat number distributions of Bengalese finch songs.
a: Six example Bengalese finch repeat count histograms (grey bars) and the best-fit model distributions (red lines). Peak repeat count increases from left-to-right and down columns. Distribution marked with (*) provide two examples of repeat distributions that have clear double peaks. For these cases, the peaks at repeat number 1 are excluded. b: Scatter plot of fit error vs. benchmark error. Each red circle corresponds to the distribution for one repeated syllable from the song database. The fit errors are smaller than the benchmark errors in the vast majority of cases (86%).

doi:10.1371/journal.pcbi.1004471.g005

The results comparing the fit errors from the sigmoidal adaption model to benchmark errors across all 71 repeating syllables are shown in Fig 5b (Materials and Methods; see also [20]). The vast majority of fit errors from the feedback adaptation model are below their respective benchmark errors (86% of fit errors below the benchmark error), demonstrating that the model does an excellent job of fitting the diverse shapes of Bengalese finch song repeat number distributions. Therefore, the single state transition model with sigmoidal adaptation, and by extension the branched-chain model with adaptive auditory feedback, can successfully describe the syllable repeat number distributions in Bengalese finch songs.

Removal of auditory feedback in Bengalese finches by deafening reduces peak repeat numbers

In our framework, auditory feedback from the previous syllable arrives in HVC at a time appropriate to provide driving excitatory input to HVC neurons that generate the upcoming syllable. For repeated syllables, this creates a positive feedback loop which is responsible for generating peak repeat numbers greater than 1 (adaptation drives the process to extinction). Therefore, a key prediction is that without auditory-feedback driven excitatory input, the peak-repeat number should shift toward 1. To test this prediction, we deafened six Bengalese finches by bilateral removal of the cochlea, and analyzed the songs before and soon after they were deafened (2–4 days) (Materials and Methods).

We found that deafening greatly reduces the peak repeat-counts. For example, in Fig 6a, we display spectrograms and rectified amplitude waveforms of the song from one bird prior to deafening (top) and soon after deafening (2–3 days post-deafening). We see that deafening
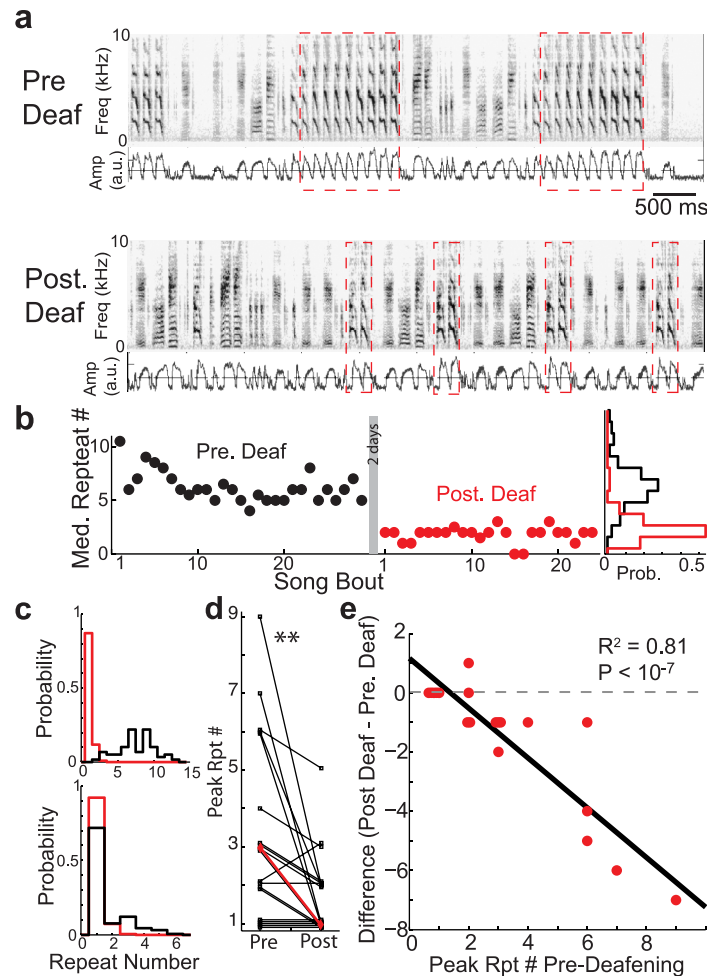
Fig 6. Removal of auditory feedback in Bengalese finches by deafening reduces peak repeat counts.
a: Example spectrograms and rectified amplitude waveforms (blue traces) for the song of one bird before
(top) and after (bottom) deafening. Red dashed boxes demarcate the repeated syllables. b: Median repeat
counts per song of the syllable from before deafening (black) and after deafening (red). Rotated probability
distributions at the right hand side display the repeat counts across all recorded songs before (black) and
after (red) deafening. c: Additional examples of repeat distributions pre- (black) and post- (red) deafening. For
syllables that were repeated many times, deafening caused sharp reductions in repetitions, resulting in
repeat number distributions that are more Markovian (upper graphs). Deafening had less of an effect on
syllables that were repeated fewer times (lower graphs). d: Deafening results in a significant decrease in the
peak repeat numbers. Individual syllables are in black (overlapping points are vertically shifted for visual
clarity), median across syllables is in red. (Wilcoxon sign-rank test, $p < 10^{-2}$, $N = 19$). e: Peak repeat
numbers before deafening vs. the differences in peak repeat numbers before and after deafening. Red dots
correspond to syllables and black line is from linear regression. Larger decreases in peak repeat numbers for
syllables that were repeated many times before deafening ($R^2 = 0.81$, $p < 10^{-7}$, $N = 19$).

doi:10.1371/journal.pcbi.1004471.g006

reduces the number of times that the syllable (red-dashed box) is repeated. The time course of
repeat generation from this bird is examined in more detail in Fig 6b, where we plot the median
repeat counts per song of the syllable from Fig 6a before deafening (black) and after deafening
(red). Here we see that, even in the first songs recorded post-deafening, there is a marked
decrease in the produced number of repeats. This data further exemplifies that repeat counts
per song is generally stable across bouts of singing within a day both before and after deafening.
Across days, repeat counts continued to slowly decline with time since deafening, though the

co-occurrence of acoustic degradation of syllables makes these later effects difficult to interpret [24, 54]. Nonetheless, the rapidity of the effect of deafening underscores the acute function of auditory feedback in the generation of repeated syllables.

Similar results were seen across the other repeated syllables. Fig 6c shows the repeat number distributions for two additional birds before (black) and after (red) deafening. In these cases, deafening resulted in repeat number distributions that monotonically decayed. The peak repeat numbers pre and post deafening for all 19 syllables in our data set are presented in Fig 6d. Across the 19 repeated syllables from 6 birds, deafening significantly reduced the number of consecutively produced repeated syllables (Fig 6d, $p < 0.01$, sign-rank test, $N = 19$, medians demarcated in red, overlapping points are vertically shifted), although there was variability in the effect magnitude: the effect of deafening appeared larger for the repeat with larger initial repeat number (compare upper and lower panels of Fig 6c). This suggests that the degree to which deafening reduces peak repeat number depends on the initial repeat number. We examined the change in peak repeat number resulting from deafening as a function of the peak repeat number before deafening (Fig 6e, red dots correspond to data from individual syllables, overlapping points are horizontally offset for visual display). We found that the magnitude of decrease in peak repeat numbers after deafening grows progressively larger for syllables with greater peak repeat numbers before deafening ($R^2 = 0.81$, $p < 10^{-7}$, $N = 19$). This suggests that repeated syllables with larger repeat numbers are progressively more dependent upon auditory feedback for repeat production. Interestingly, after two days of hearing loss, one of the deafened Bengalese finches in our experiments had a repeat that was minimally affected by deafening, and several birds retained peak repeat number around 2, not all the way to 1 as predicted for a Markov process (Fig 6d). None-the-less, these deafening results are consistent with the hypothesis that the generation of repeated syllables is driven, in-part, by a positive-feedback loop caused by excitatory auditory input during singing.

HVC auditory responses to repeated syllables gradually adapt

A key prediction of the adaptive feedback model for repeat generation is that auditory responses of HVC neurons should decline over the course of repeated presentations of the same syllable. To test this hypothesis, we examined the properties of HVC auditory responses to repeated syllables in sedated birds (Materials and Methods). An example recording from an HVC multi-unit site in response to playback of the bird's own song (BOS) stimulus is presented in Fig 7a, which displays the stimulus oscillogram (top), and the average spike rate in response to the stimulus (bottom). Multiple renditions of the repeated syllable are demarcated by red-dashed boxes, and we see that the evoked HVC auditory responses to repeated versions of the same syllable gradually declined.

The example presented above suggests that auditory responses to repeated presentations of the same syllable adapt over time. However, in the context of BOS stimuli, the natural variations that occur in syllable acoustics, inter-syllable gap timing, and in the identity of the preceding sequence, make it difficult to directly compare responses to different syllables in a repeated sequence. Therefore, to examine how responses to repeated syllables are affected by the length and identity of the preceding sequence, for each bird we constructed a stimulus set of long, pseudo-randomly ordered sequences of syllables (10,000 syllables in the stimulus, one prototype per unique syllable, median of all inter-syllable gaps used for each inter-syllable gap, derived from the corpus of each bird's songs, Materials and Methods). This stimulus allows a systematic investigation of how auditory responses to acoustically identical syllables depend on the length and syllabic composition of the preceding sequence [28]. Auditory responses at 18 multi-unit recordings sites in HVC from 6 birds were collected for this data set, which
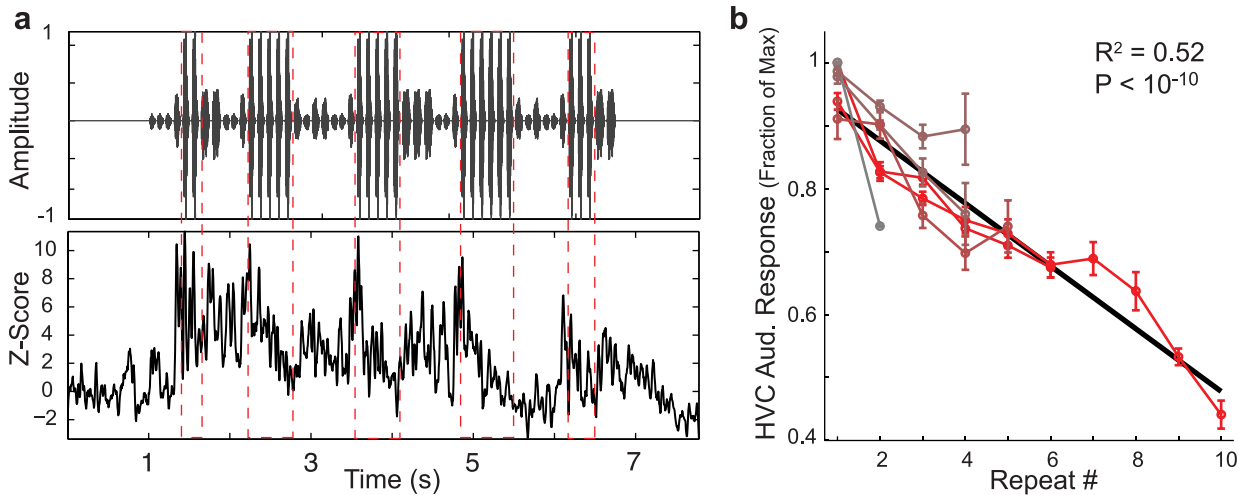
Fig 7. HVC auditory responses to repeated syllables gradually adapt. a: Example auditory recording from an HVC multi-unit site in response to playback of the BOS (bird's own song) stimulus. Top panel is the song oscillogram. Bottom plot is the average response rate across trials. Adaptation of HVC auditory responses to a repeated syllable (demarcated by red-dashed lines) is observed. b: Responses to the last syllable in a repeat as a function of the repeat number. Data are presented as mean ± s.e. of normalized auditory responses across sites for a given repeated syllable (11 sites in 4 birds, 6 repeated syllables). Data are colored from grey-to-red with increasing peak repeat number. Across syllables and sites, the response to the last syllable in a repeat declines with increasing repeat number. Black line is from linear regression ($R^2 = 0.523$, $p < 10^{-10}$, $N = 24$, slope = -5%).

doi:10.1371/journal.pcbi.1004471.g007

contained 40 unique syllables. Of these 40 syllables, 6 syllables in 4 birds (with 11 recording sites) were found to naturally repeat.

We used these stimuli to systematically examine how auditory responses to a repeated syllable depend on the number of preceding repeated syllables. We found that HVC auditory responses gradually declined to repeated presentations of the same syllable. In Fig 7b, for each uniquely repeated syllable (different syllables are colored from grey-to-red with increasing max repeat number), we plot the average normalized auditory response (mean ±s.e. across sites) to that syllable (e.g. 'b') as a function of the repeat number (e.g. repeat number 5 corresponds to the last 'b' in 'bbbbb'). Across HVC recordings sites and repeated syllables, the response to the last syllable declined as the number of preceding repeated syllables increased ($R^2 = 0.523$, $p < 10^{-10}$, $N = 24$, slope = -5%).

Thus, auditory responses to repeated syllables gradually adapt as the number of preceding repeated syllables increases, providing confirmation of a key functional mechanism of the network model.

Non-Markovian repeated syllables are loudest and evoke the largest HVC auditory responses

To generate non-Markovian repeat distributions, we have proposed that the sequence generation circuitry is driven, in part, by auditory feedback that provides excitatory drive to sensory-motor neurons that control sequencing. Specifically, auditory feedback from the previous syllable arrives in HVC at a time appropriate to provide driving excitatory input to neurons that generate the upcoming syllable. This predicts that if HVC auditory responses are positively modulated by sound amplitude, feedback associated with louder syllables should provide stronger drive to the motor units, and thus generate longer strings of repeated syllables for a given rate of adaptation. This logic is supported by the sigmoidal adaptation model, which predicts a
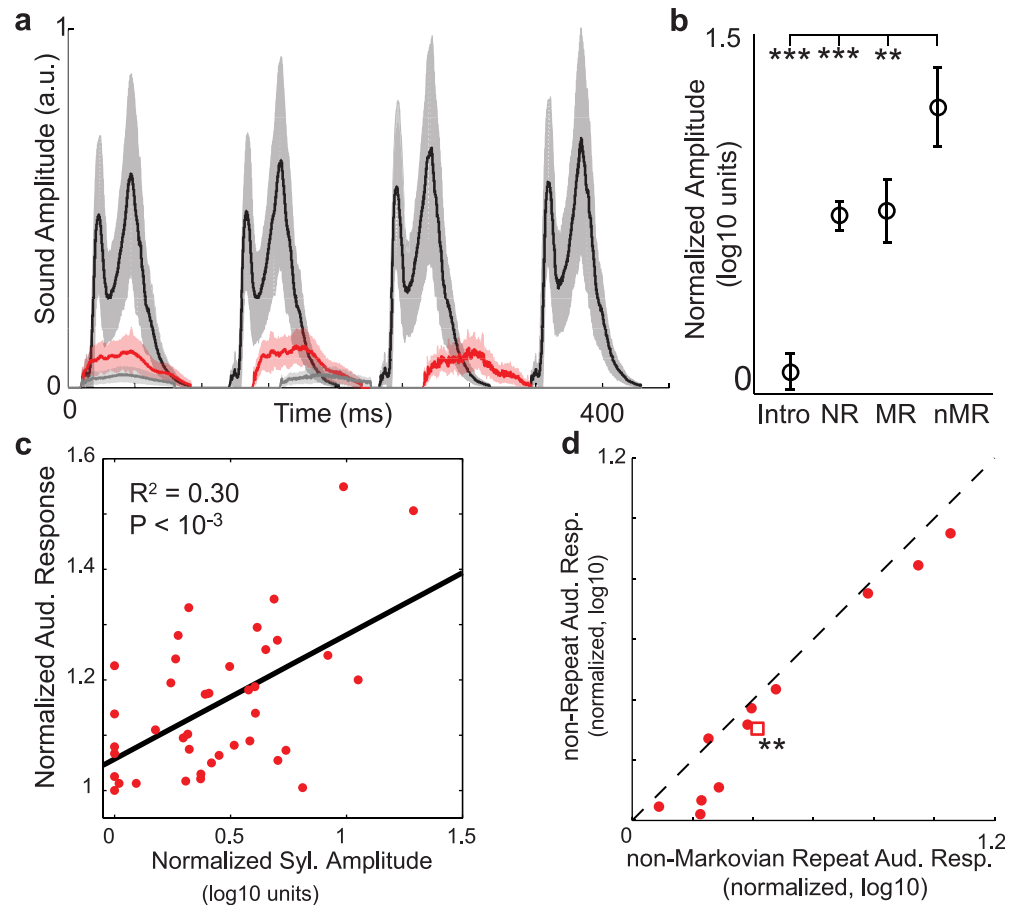
Fig 8. Non-Markovian repeated syllables are loudest and evoke the largest HVC auditory responses.
a: Mean ±s.d. amplitude waveforms for a non-Markovian repeated syllable (black), a Markov repeated syllable (red), and an intro note (grey) from the songs of one bird. b: Mean ±s.e. normalized peak amplitudes of song vocal elements. Intro notes (Intro), non-repeated syllables (NR), Markov-repeated syllables (MR, peak repeat number = 1), and non-Markovian repeated syllables (nMR, peak repeat number > 1). non-Markovian repeated syllables are significantly louder than other vocalizations ($p < 10^{-3}$, $p < 0.01$, Wilcoxon sign-rank test, Bonferroni corrected for $m = 3$ comparisons). c: Scatter plot of normalized auditory responses to a syllable as a function of the normalized amplitude of that syllable. Black line is from regression ($R^2 = 0.30$; $p < 10^{-3}$, $N = 40$ syllables). d: Paired comparison of normalized auditory responses to non-repeated syllables (NR) and non-Markovian repeated syllables (nMR). Repeated syllables illicit larger auditory responses. ($p < 0.01$, Wilcoxon sign-rank test, $N = 11$ sites). Circles: data; square: median.

doi:10.1371/journal.pcbi.1004471.g008

threshold auditory feedback strength at which the peak repeat number becomes greater than one (i.e. non-Markovian, Fig 4b). Behaviorally, this predicts that non-Markovian sequences of repeated syllables should be composed of the loudest syllables in the bird's repertoire.

We tested this behavioral prediction by comparing the amplitudes of Bengalese finch vocalizations based on their repeat structure. Fig 8a plots the rectified amplitude waveforms (mean ±s.d.) of a few consecutively produced repetitions of a non-Markovian repeated syllable (black), a Markovian repeated syllable (red), and 'introductory' note (grey) from one bird. The non-Markovian repeated syllable is qualitatively louder than the other repeated vocalizations in the birds' repertoire. To quantitatively test this prediction, we measured the peak amplitude of the 281 unique syllables in our data set, and normalized this to the minimum peak amplitude

across syllables (Materials and Methods). We categorized each syllable in our data set according to whether it was an introductory note (Intro), a non-repeated syllable (NR: repeats = 0), a Markovian repeated syllable (MR: peak repeat number = 1), or a non-Markovian repeated syllable (nMR: peak repeat number > 1). In Fig 8b, we plot the mean ±s.e. of the normalized peak amplitudes of these syllable groups across the data set. As exemplified by the data in Fig 8a, we found that non-Markovian repeated syllables were significantly louder than the other vocalizations in a bird's repertoire (***: $p < 10^{-3}$, **: $p < 10^{-2}$, sign-rank test, Bonferroni corrected for m = 3 comparisons). Therefore, syllables with non-Markovian repeat distributions are typically the loudest vocalizations produced by a bird.

If amplitude is a contributing factor to repeat generation, then HVC auditory responses should be positively modulated by syllable amplitude. However, previous work in the avian primary auditory system has found a population of neurons that is insensitive to sound intensity [55], and amplitude normalized auditory responses have been utilized in previous models of sequence encoding in HVC auditory responses [56]. Therefore, we first examined whether auditory responses were positively modulated by syllable amplitude. To make recordings from different sites/birds comparable, we normalized both the syllable amplitudes (relative to mean) and auditory responses (relative to minimum). The scatter plot in Fig 8c plots the normalized syllable amplitudes vs. the normalized auditory responses (averaged across sites within a bird), for the 40 syllables in in our data set [28]. We found a modest but significant positive correlation between auditory responses and syllable amplitude (R^2 = 0.30;$p < 10^{-3}$, N = 40 syllables). We next examined whether the increased amplitude of repeated syllables resulted in increased HVC auditory response to these syllables. We performed a paired comparison of normalized auditory responses to non-repeated syllables (NR) and non-Markovian repeated syllables (nMR) at the 11 sites where auditory responses to repeated syllables were collected (Fig 8d). We found that repeated syllables had significantly larger auditory responses than non-repeated syllables ($p < 0.01$, sign-rank test, N = 11 sites). Thus, HVC auditory responses are sensitive to syllable amplitude, and repeated syllables elicit larger auditory responses than non-repeated syllables, likely due to being the loudest syllables that a bird sings. Therefore, the strong auditory feedback associated with these loud repeated syllables may be a key contributor to their non-Markovian repeat distributions.

Discussion

We have provided converging evidence that adapting auditory feedback directly contributes to the generation of long repetitive vocal sequences with non-Markovian repeat number distributions in the Bengalese finch. A branching chain network model with adapting auditory feedback to the repeating syllable-chains produces repeat number distributions similar to those observed in the Bengalese finch songs. From the network model we derive the sigmoidal adaptation model for repeat probability, and show that it reproduces the repeat distributions of both the branching chain network and Bengalese finch data. Removal of auditory feedback by deafening reduced the peak repeat number, confirming one of the key features of the proposed mechanism. Furthermore, recordings in the Bengalese finch HVC show that auditory responses of HVC adapt to repeated presentations of the same syllable, providing evidence for another key feature of the proposed mechanism. Finally, we found that non-Markovian repeated syllables are louder than other syllables and elicit stronger auditory responses, suggesting that a threshold auditory feedback magnitude is required to generate long strings of repeated syllables, in agreement with modeling results. Together, these results implicate an adapting, positive auditory-feedback loop in the generation of long repeated syllable sequences, and suggest that animals may directly use normal sensory-feedback signals to guide behavioral

sequence generation with sensory adaptation preventing behaviorally deleterious perseveration.

In our framework, a positive feedback loop to a repeating syllable provides strong excitatory drive to that syllable and sustains high repeat probability. The strength of this feedback gradually reduces while the syllable repeats, preventing the network from perseverating on the repeated syllable. The combination of strong, positive feedback and gradual adaptation allows the production of non-Markovian repeat number distributions in the branching chain networks. It should be emphasized that this feedback mechanism is not necessary for repeat syllables with Markovian repeat number distributions [20]. Such Markovian repeats are short, and can be simply generated with self connections in the branching chain network model without auditory feedback [26].

We have conceptualized the adapting feedback as short-term synaptic depression of the NIf to HVC synapses resulting from auditory feedback. However, neither the exact source of the feedback nor the mechanism generating the adaptation is critical for our model. Indeed, the adaptation of auditory responses could arise from a variety of pre- and/or post-synaptic mechanisms anywhere in the auditory pathway, such as in the auditory forebrain [57], the auditory responses of NIf [47–50] or other auditory inputs to HVC such CM (caudal mesopallium) [58], or within HVC itself. The biophysical origin of the auditory adaptation in HVC observed in our experiments remains to be determined. Our experiments showing the adaptation of auditory feedback for the repeated syllables were performed on passively listening birds. Future experiments on singing birds are required to see whether such adaptation occurs in the singing state.

Previous experiments that deafened Bengalese finches showed that removal of auditory feedback has immediate impact on the song syntax of the Bengalese finch [24, 25, 54]. The main effect reported was the increased randomness in the syllable sequences. However, the impacts on syllable repeats was not analyzed. Our own deafening experiments showed that long repeated syllables are particularly vulnerable to loss of hearing, and their repeat number distributions shift close to Markov distributions two days after deafening. The Markovian repeats, on the other hand, were not affected as much. These new observations supports the idea that non-Markovian repeats rely more on auditory feedback than Markovian ones, as suggested by our computational model. However, it should be noted that the deafening results are consistent with our model but do not prove it. There could be alternative explanations, including possible systematic changes in the stress level, the arousal states, the neural circuits in the auditory and motor areas during the recovery from deafening. Future experiments that directly manipulate auditory feedback online in intact brain will help to further test our model.

After two days of hearing loss, one of the deafened Bengalese finches in our experiments maintained peaked repeat number distributions, and several birds retained peaked repeat numbers around 2, not all the way to 1 as predicted for a Markov process. One possible explanation is the existence of multiple chains that produce syllables with similar acoustic features [20]. Such a "many-to-one mapping" could produce residual non-Markovian features in the repeat number distribution after deafening. Another possibility is that there are several internal feedback loops to HVC within the song system that could contribute to repeating syllables. For example, there are direct anatomical projections from RA back to HVC [59] as well as through the medial portion of MAN (mMAN) [60]. Furthermore, there are connections from vocal brainstem nuclei to HVC through Uva and NIf [38, 61]. Although the signals transmitted through these internal feedback loops are poorly understood, they are likely to contribute to the temporal/sequential structure of song [62]. These internal feedback loops may also contribute to, or even be the main routes of connecting the syllable encoding chains in HVC, rather than the direct connections between the chains within HVC as assumed in our network model. Furthermore, such internal feedback loops could be one site of adapting excitatory drive that

contributes to the generation of non-Markovian repeats. However, our deafening results suggest that auditory feedback is a primary source of excitatory drive for repeat generation. Our modeling results will not change if such internal feedback loops are used instead of the direct connections for sequence generation, or instead of auditory feedback as the route of adapting positive feedback.

The feedback delay time plays an important role in our model, as the feedback signal must return to HVC in time to exert an influence on the selection of the next syllable. We have hypothesized a simple scenario where these feedback signals are auditory in nature. Each is tuned to a specific syllable in the bird's repertoire and targets entire chains within HVC. In this case, there is a simple constraint on the delay time for the auditory feedback to exert its influence on the song sequence: the total delay time must be less than the duration of the syllable under examination. Different delay times conforming to this constraint would lead to slight changes in the repeat distribution due to small differences in the initial amount of adaptation experienced on the first repetition, but with no qualitative differences. This constraint could be pushed beyond its limit by very short syllables that terminate before the auditory feedback would return to HVC, precluding the ability of auditory feedback to influence the subsequent transition. If non-auditory internal feedback loops were to carry such a signal, the delay time—and thus the corresponding constraint—could be significantly shorter than predicted for the auditory case. Another possibility is that the delay makes the auditory feedback effective only after the syllable has repeated once or twice. The initial repeats could be sustained by the intrinsic self-connections of the chain network encoding the repeated syllable ([Fig 2c](#)). The auditory feedback can then arrive to sustain a long repetition. If the self-connections are weak, the syllable tends to stop at one or two repetitions; but once it repeats more than once or twice, the arriving auditory feedback can take over and sustain a long repetition. This could be another mechanism for the double peaked repeat number distributions we have observed ([Fig 5a](#)), in addition to the possibility of a "many-to-one" mapping from HVC to the syllable types. It will be interesting to distinguish these possibilities in future studies.

We observed that non-Markovian repeated syllables are typically the loudest syllables in a bird's repertoire. Furthermore, HVC responses to repeated syllables were significantly greater than responses to non-repeated syllables. Together, these results suggest that louder syllables provide stronger auditory feedback to HVC. This is consistent with our model, in which non-Markovian repeats are strongly influenced by auditory feedback to HVC, though by no means does our model predict such a result. The relationship between the syllable amplitude and repeat length can be further tested with experiments that manipulate syllable amplitudes online with realtime auditory feedback [22]. It should be noted, however, we are not suggesting that a syllable is loud because of a strong auditory input to HVC. The control of syllable amplitude could depend on multiple neural mechanisms. It remains to be investigated why the non-Markovian repeated syllables are louder than other syllables.

Our framework can be extended to allow auditory feedback to influence transition probabilities beyond repeated syllables. In general, because the auditory-motor delay in HVC due to neural processing is on the order of a syllable duration ([Fig 2a](#)), auditory feedback from the previous syllable arrives in HVC at a time to contribute to the motor activity for the current syllable [13, 14, 28]. For a diverging transition of syllable 'a' to either 'b' or to 'c', as shown in [Fig 2b](#), auditory feedback from syllable 'a' can be applied to chain-b and chain-c. Depending on the amount of feedback on each chain, the transition probability to 'b' or 'c' can be enhanced or reduced by the feedback. Our model for repeating syllables ([Fig 2c](#)) can be thought of as a special case of this general scenario, in which the repeating syllable-chain receives much stronger auditory input than the competing chain. The strong auditory feedback for repeated syllables may in part reflect synaptic weights that have been facilitated by Hebbian mechanisms

operating on the repeated coincidence of auditory feedback with motor activity [28]. This framework is consistent with the observations that manipulating auditory feedback experimentally can change the transition probabilities [13]. Auditory feedback plays a secondary role in determining the song syntax in our proposed mechanism. The allowed syllable transitions are encoded by the branching patterns of the chain networks. Auditory feedback biases the transition probabilities, to varying degrees for different syllable transitions. The secondary role of auditory feedback on the syntax could be the reason for the individual variations seen in a previous deafening experiment [24]. Indeed, it was observed that one Bengalese finch maintained its song syntax 30 days after deafening [24]. The secondary role of feedback in our model is in contrast to the model of Hanuschkin et al, who relied entirely on auditory feedback for determining syllable transitions [27]. However, as in the Hanuschkin model, our model emphasizes the role of auditory feedback in shaping song syntax.

We have theorized that auditory feedback provides direct inputs to HVC_{RA} neurons in controlling syllable repetitions in the Bengalese finch. Whether auditory feedback can reach HVC_{RA} neurons in the Bengalese finch is not yet known. Recent experiments that recorded projection neurons intracellularly in HVC of the zebra finch, whose song consists of fixed sequences of syllables, demonstrated that auditory feedback is gated off and does not provide inputs to the projection neurons during singing [63, 64]. On the other hand, it was shown that the firing rates in HVC of the Bengalese finch changed during singing when the auditory feedback was manipulated [14], suggesting that auditory feedback can influence HVC during singing in this species. It is possible that the differences in sequence complexity between these species may in part be due to different online sensitivities to auditory feedback [24]. Syllable repetitions are common in many other songbird species, including the canary [19]. It remains to be seen whether auditory feedback plays an important role in syllable repetitions in species other than the Bengalese finch. The differences of sensory-motor integration during singing in different species of songbirds need further investigations.

Probabilistic state transition models have been used for describing variable birdsong syntax with high accuracy [20]. Multiple states for a single syllable are often required for the state transition model to capture the statistical properties of the syllable sequences, resulting in the partially observable Markov model with adaptation (POMMA) [20]. Such many-to-one mapping manifests as multiple peaks in the repeat number distributions in our data (Fig 5a). However, some of the multiple states in POMMA could also be due to the inaccurate description of history-dependence of the transition probabilities. The geometric adaptation model for the repeat probability, used in the previous work [20], often leads to multiple states to accurately capture the non-Markovian repeat number distributions, as shown in Fig 4b and 4c. In contrast, the sigmoidal adaptation model for the repeat probability, derived from our network model, enables accurate description of such distributions using a single state. Thus the sigmoidal adaptation model should reduce the complexity of POMMA for the Bengalese finch song syntax.

For motor control with continuous trajectories, such as reaching movements or articulation of single speech phonemes, it has been proposed that internal models estimate sensory consequences of motor commands, compare these estimates to actual sensory feedback, and use the difference as error signals to correct ongoing motor commands [65–68]. Along these lines, recent recordings in the auditory areas Field-L and CLM (caudolateral medopallium) of the zebra finch showed that, during singing, a subset of neurons exhibit activity that is similar to, but precedes, the activity induced by playback of the birds own song [69]. These data have led to the hypothesis that the songbird auditory system encodes a prediction of the expected auditory feedback ("forward model") used to cancel expected incoming auditory feedback signals [39, 65, 69, 70]. According to such a forward model interpretation, as long as feedback matches expectation, auditory feedback does not reach HVC and therefore does not contribute to song

generation during singing [64]. At the surface, this seems at odds with our framework in which auditory feedback has a direct role in song generation, in particular for repeats. One possible resolution is that due to the probabilistic syllable transitions, auditory feedback cannot be fully predicted and canceled by the forward model since the motor actions themselves are not entirely predictable. Such imperfect cancelation allows direct influence of auditory feedback on syllable sequences. Another possibility is that due to the increased loudness of non-Markovian repeated syllables, residual auditory input reaches HVC and contributes to song generation.

Some similarities between non-Markovian syllable repetitions in birdsong and sound/syllable repetitions in stuttered speech have been observed in the past [71–73]. In persons who stutter, repeating syllables within words ('to-to-to-today', for example) is a prominent type of speech disfluency [74–76]. Auditory feedback plays an important, but poorly understood, role in stuttered speech. For example, altering auditory feedback, including deafening [74], noise masking [77, 78], changing frequency [79], and delaying auditory feedback reduces stuttering [80]. Conversely, delayed feedback can induce stuttering in people with normal speech [10, 11]. Auditory processing may be abnormal both in zebra finches with abnormal syllable repetitions and in persons who stutter [71]. Our observation that deafening reduces syllable repetitions in Bengalese finch songs echoes the reduction of stuttering after deafening in persons who stutter [74]. In general agreement with our proposed role of auditory feedback in repeat generation, some theories suggest that persons who stutter have weak feed-forward control and overly rely on auditory feedback for speech production [67]. It will be interesting to see whether further quantitative analysis of the statistics of stuttered speech would reveal additional behavioral similarities, such as non-Markovian distributions and increased amplitude; to our knowledge no such examination exists. Such similarities could point to shared neural mechanisms with syllable repetition in birdsong, especially the possibility that auditory feedback plays a key role. However, our study also provides a cautionary note to the interpretation of repeated syllables in birdsong as 'stutters'. Our analysis shows that syllables with non-Markovian repeat distributions are loud and require strong auditory feedback. In contrast, syllables with Markovian repeat distributions are quieter and are less reliant on auditory feedback for their generation. We propose that it is the former type of syllable repetition that shares similarity to stuttering in humans.

Materials and Methods

Ethics Statement

All procedures involving animals were performed in accordance with established animal care protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee (IACUC).

Model neurons

The model neurons for the network simulations are a reproduction of those in previous works [26, 35]. Below, we summarize the key aspects of these models. The reader is referred to these papers for exact details on the equations and constants. Since detailed information about the ion channels of HVC neurons is unavailable, we model both HVC_{RA} and HVC_I neurons as simple Hodgkin-Huxley type neurons, adding extra features to match available electrophysiological data. HVC_I neurons exhibit prolonged tonic spiking during [35]. To simulate this we use a single-compartment model with the standard sodium-potassium mechanism for action potential generation along with an additional high-threshold potassium current that allows for rapid spike generation.

A distinctive feature of HVC_{RA} neurons is that their activity comes in the form of precise bursts during song production [30, 35]. This bursting activity increases the robustness of signal propagation along chains of these neurons [32, 35]. A study of the subthreshold dynamics of HVC_{RA} neurons during singing suggests that this bursting is an intrinsic property of these cells [35]. We generate this intrinsic bursting behavior with a two-compartment model [26, 32, 35]. A dendritic compartment contains a calcium current as well as a calcium-gated potassium current. When driven above threshold, these currents produce a stereotyped calcium spike in the form of a sustained (roughly 5 ms) depolarization of the dendritic compartment. A somatic compartment contains the standard sodium-potassium currents for generating action potentials. These compartments are ohmically coupled so that a calcium spike in the dendrite drives a burst of spikes in the soma.

All compartments also contain excitatory and inhibitory synaptic currents. Action potentials obey kick-and-decay dynamics. All synaptic conductances start at 0. When an excitatory or inhibitory action potential is delivered to a compartment, the corresponding synaptic conductance is immediately augmented by an amount equal to the strength of the synapse. In between spikes, the synaptic conductances decay exponentially toward zero.

Branching synfire chain model

The network topology underlying all of the more advanced models below is the branching synfire chain network for HVC [26]. HVC_{RA} neurons are grouped into pools of 60 neurons. 20 pools are then sequentially ordered to form a chain. Except for the final pool, all neurons in a pool make an excitatory connection to every neuron in the next pool (Fig 2b). The strengths of these synapses are randomly generated from a uniform random distribution between 0 and $G_{EE, max}$ = 0.09 mS/cm^2. Because of this setup, activating the neurons in the first group sets off a chain reaction where each group activates the subsequent group, leading to a signal of neural activity propagating down the chain with a precise timing. There is one chain for every syllable in the repertoire of the bird. Activity flowing down a given chain drives production of the corresponding syllable through the precise temporal activation of different connections from the HVC_{RA} neurons to RA (not explicitly modeled). To begin to impose a syntax on the song, the neurons in the final pool of one chain make connections to the initial pool of any chain whose syllable could follow its own. This branching pattern encodes the basic syllable transitions that are possible.

When the activity in an active chain reaches a branching point, all subsequent chains are activated, however only one should stay active—the syllable chosen next. This selection is achieved through lateral inhibition between the chains intermediated by HVC_I neurons. There is a group of 1,000 HVC_I neurons. Each HVC_{RA} neuron has a chance of making an excitatory connection to each HVC_I neuron with a probability p_{EI} = 0.05. Each of these connections has a strength randomly drawn from a uniform distribution between 0 and $G_{EI, max}$ = 0.5 mS/cm^2. In turn, each HVC_I neuron has a chance of making an inhibitory connection to each HVC_{RA} neuron with a probability p_{IE} = 0.1. The strengths of these connections are randomly drawn from a uniform distribution between 0 and $G_{IE, max}$ = 0.7 mS/cm^2. This setup gives a rough approximation of global inhibition on the HVC_{RA} neurons which is what leads to the lateral inhibition between the chains that they comprise.

Noise is added to the network to make switching between chains a stochastic process. This noise is modeled as a Poisson process of spikes incident on each compartment of every neuron. The strength of each spike is randomly selected from a uniform distribution from 0 to G_{noise} and every spike has an equal chance of being excitatory or inhibitory. Both compartments of HVC_{RA} neurons receive noise at a frequency of 500 Hz; at the soma G_{noise} = 0.045 mS/cm^2,

while at the dendrite G_{noise} = 0.035 mS/cm². The single compartment of the HVC_I neurons receive noise at a frequency of 500 Hz with G_{noise} = 0.45 mS/cm². In HVC_{RA} neurons, this leads to subthreshold membrane fluctuations of ~ 3 mV; in the HVC_I neurons, the results is a baseline firing rate of ~ 10 Hz.

Each HVC_{RA} neuron also receives an external drive that facilitates robust propagation of signals through the chains. This takes the form of a purely excitatory spike train modeled by a Poisson process with frequency 1,000 Hz. The strength of each spike is chosen from a uniform random distribution from 0 to 0.05 mS/cm².

Auditory feedback model

We incorporate auditory feedback into the branching synfire chain model in a manner similar to the external drive used in [26]. When a syllable is being produced and heard by the bird, some amount of auditory feedback can be delivered to any of the chains in the network in the form of external drives. The relative strength of this feedback drive between chains then biases transition probabilities so that auditory feedback plays an important role in determining song syntax.

The first piece in our model for auditory feedback is determining when auditory feedback from a specific syllable is active. We assume that the first few pools in every chain encode for the silence between syllables. Furthermore, once a syllable is being produced, there is a delay before auditory feedback begins that represents how long it takes for the bird to hear the syllable and process the auditory information. In our simulations, the activity of the 4th pool of every chain is monitored (by keeping track of the number of spikes in the previous 5 ms), with syllable production onset determined by when the population rate crosses a threshold of 43 Hz/neuron. After a delay of 40 ms, auditory feedback from that chain's syllable begins.

The auditory feedback takes the form of an external drive to all of the HVC_{RA} neurons in a chain. Every chain can provide auditory feedback to every other chain, including itself. Thus, if there are N chains, then there are N^2 auditory feedback pathways. Denote the strength of the auditory feedback from chain i to chain j as G_{ij}. Every neuron in a chain will have N synapses, each one carrying the auditory feedback from one of the N chains in the network. The synapses carrying the auditory feedback from chain i to chain j have strengths drawn from a uniform random distribution between 0 and G_{ij}. Setting G_{ij} = 0 implies that there is no auditory feedback from chain i to chain j. When auditory feedback from a chain is active, the corresponding synapses are driven with Poisson processes at a frequency f_{fdbk}.

The model that each neuron receives only one synapse for each auditory feedback source is unrealistic. However, for computational simplicity, we model the feedback this way and consider each high-frequency synapse to be carrying spike trains from multiple sources. Since the kick-and-decay synapse model does not separate sources, this induces no real approximation. Auditory feedback parameters for Fig 2 were tuned to f_{fdbk} = 1,340 Hz and G_{bb} = 1.9 mS/cm².

Synaptic depression model

To implement synaptic depression, we follow a simple phenomenological model used in Abbott et al. [53]. Whenever a synapse is used to transmit a spike, its strength g is decreased by a constant fraction α, so that $g \rightarrow (1 - \alpha)g$. The parameter α is referred to as the depression strength. In between spikes, the synaptic strength recovers toward its base strength g_0 with first order dynamics:

$$\tau_R \frac{dg}{dt} = -(g - g_0). \tag{4}$$

The parameter τ_R is called the synaptic depression recovery time constant. If such a depressing synapse carries a spike train with a constant frequency f, the large-scale effect is an exponential decay to a steady-state strength where recovery and depression are balanced. The time constant of this decay as well as the steady-state strength can be expressed as functions of the model parameters: $\tau(\tau_R, \alpha, f)$ and $g_\infty(\tau_R, \alpha, f)$. See below for a derivation of the exact forms.

In our simulations with synaptic depression on the synapses carrying auditory feedback (in particular Fig 2), we use $\tau_R = 3.25$ s and $\alpha = 0.006$. It should be noted that, since these synapses actually represent the combined effect of multiple synapses (see above), these model parameters should not be taken as biologically representative. However, by matching the large-scale dynamics (τ and g_∞) of the lower-frequency constituent synapses to that of the model synapse, one can find the more biologically relevant underlying depression parameters. Assume that each auditory feedback synapse represents the combined input of N constituent synapses, each carrying a spike train with a frequency f/N so that the model synapse carries a spike train with frequency f. Matching the large-scale dynamics is then expressed as (primes representing biologically relevant parameters)

$$\tau(\tau'_R, \alpha', f/N) = \tau(\tau_R, \alpha, f), \tag{5}$$

$$g_\infty(\tau'_R, \alpha', f/N) = g_\infty(\tau_R, \alpha, f). \tag{6}$$

Since α, τ_R, and f are known from the model, we can solve for α' and τ'_R. With $N = 50$ this gives $\alpha' \approx 0.26$ and $\tau'_R \approx 3.75$ s—reasonable values for short-term depression in cortex [53].

Computational implementation

Both the neural and synaptic depression models take the form of a large system of differential equations. A fourth-order Runge-Kutta scheme is used to numerically integrate these equations with custom code written in C++. When action potentials are generated during a time-step, synaptic conductances and synaptic depression dynamics are immediately updated before the next time-step is taken. All analysis is done with custom code in the MATLAB (The Mathworks, Natick, MA) environment.

Statistical model

To systematically examine how the repeat number distribution depends on the strength a_0 of the auditory feedback and the adaptation strength α, we used the sigmoidal adaptation model, Eqs (2) and (3), to generate repeat number distributions with combinations of these parameters. The decay time constant of the auditory feedback due to synaptic adaptation was set to

$$\tau = \frac{\tau_R}{1 - \tau_R f \log(1 - \alpha)}, \tag{7}$$

where τ_R is the recovery time constant and f is the firing rate of NIf neurons during auditory feedback (see below). Besides a_0 and α, all other parameters are set using those from the network simulations shown in Fig 3 with $T = 100$ ms (approximately the length observed in the simulations). To simulate a repeat bout, we sequentially generate random numbers x_k from a uniform distribution between 0 and 1 and compare each number to $p_r(k)$. The first time that $x_k > p_r(k)$ signifies that a further repeat does not occur, so the bout contains k repeats. A distribution of repeats for a given (a_0, α) combination is produced by simulating the repeat bouts 10,000 times, and the results are shown in Fig 4b, where we plot the peak repeat numbers for the distributions. Because the peak repeat number can go to infinity as adaptation strength goes to 0, for numerical stability we use a minimum adaptation strength of 0.001.

Special cases of the sigmoidal adaptation model

The sigmoidal adaptation contains two interesting special cases: (1) If we set the adaptation constant $\tau \to \infty$, which is equivalent to no adaptation of the auditory synapses, we have $b \to 1$ and the repeat probability becomes a constant, a hallmark of the Markov model for repeats. (2) If $c = 1$, which means the repeat probability is zero when the repeat number is large, and the initial auditory input is small such that when $ab \ll 1$, we have $p_r(n) \approx ab^n$, i.e. the repeat probability decreases by a constant factor with the repeat number. This the geometric adaptation of repeat probability. It was used to describe the repeating syllables in a previous work on the song syntax of the Bengalese finch [20].

Any of these models can be extended to provide better fits to data by allowing multiple states. In these extended models, a repeated syllable is represented by multiple repeating states that all produce that syllable and are connected in series (Fig 4b).

Fitting repeat number distributions

The probability of the syllable repeating N times is given by

$$P(N) = (1 - p_r(N))\Pi_{n=1}^{N-1}p_r(n). \tag{8}$$

The observed repeat number probability $P_o(N)$ is computed by normalizing the histogram of the repeat numbers. The parameters a, b, c are determined by minimizing the sum of the errors

$$E = \sum_N (P(N) - P_o(N))^2, \tag{9}$$

while constraining the parameters ranges $0 < a < 10^8$, $0 < b < 1$, and $0 < c < 1$, using the nonlinear least square fitting function 'lsqcurvefit' in MATLAB. To avoid local minima in the search, 20 random sets of the initial values of the parameters were used for the minimization, and the best solution with the minimal square error was chosen.

Comparing two probability distributions

The difference between two probability distributions $p_1(n)$ and $p_2(n)$ is defined as

$$d = \frac{\max_n |p_1(n) - p_2(n)|}{\max_n(p_1(n), p_2(n))}, \tag{10}$$

i.e. the maximum absolute differences between the two distributions normalized by the maximum of the two distributions [20].

When fitting a functional form to a probability distribution, the difference between the empirical distribution and the fit is compared to a benchmark difference that represents the amount of error expected from the finiteness of data. For a given empirical distribution, the benchmark difference is computed by first randomly splitting the full data set into two groups of equal size and then computing the difference between the distributions resulting from each group. This process is repeated 1,000 times to produce a distribution of differences from simple resampling. The benchmark error is set at the 80th percentile of this bootstrapped distribution. The method of benchmark error was explained in detail previously [20].

Slow-scale depression dynamics

Our model of synaptic depression characterizes the temporal dynamics of synaptic strength, g. Each synapse has a base strength, g_0. The depression model has two parameters: (1) depression strength, α: fraction of strength lost at each spike; (2) recovery time constant, τ_R: rate of

exponential recovery toward g_0. Mathematically it can be described by two rules: 1. at a spike: $g \rightarrow (1 - \alpha)g$; 2. between spikes: $\tau_R dg/dt = - (g - g_0)$. We would like to characterize how this synapse will behave when transmitting a spike train that takes the form of a Poisson process. The analysis is simpler if we consider a regular spike train with frequency f as an approximation. Fortunately, this should still give the average behavior for the Poisson process case. We begin by deriving an iterative map that takes the strength right before one spike and gives the strength right before the next.

Let the strength of the synapse right before a spike be g. Immediately after the spike, the strength will then be $(1 - \alpha)g$. Integrating the equation for recovery (from an initial condition (t_i, g_i) to (t_f, g_f)) yields:

$$g_f = g_0 + (g_i - g_0)e^{-(t_f - t_i)/\tau_R}. \tag{11}$$

Since the spike train is assumed to be regular, we have $t_f - t_i = 1/f$. And since the recovery starts from $g_i = (1 - \alpha)g$, the complete spike-to-spike iterative map is

$$g \rightarrow g_0 + ((1 - \alpha)g - g_0)e^{-1/(\tau_R f)}. \tag{12}$$

Using synaptic strength relative to g_0, i.e. $g = Ag_0$ gives:

$$A \rightarrow 1 - e^{-1/(\tau_R f)} + (1 - \alpha)e^{-1/(\tau_R f)}A. \tag{13}$$

This iterative map has the form $A \rightarrow a + bA$, with $a = 1 - e^{-1/(\tau_R f)}$ and $b = (1 - \alpha)e^{-1/(\tau_R f)}$. If we start with $A = 1$, then this map has a closed-form solution:

$$A_n = \frac{a}{1 - b} + \frac{1 - a - b}{1 - b}b^{n-1}. \tag{14}$$

This is a geometric decrease toward a steady-state value of $a/(1 - b)$ with a ratio of $r = b$. In terms of our model parameters, this is

$$g_\infty = g_0 \frac{1 - e^{-1/(\tau_R f)}}{1 - (1 - \alpha)e^{-1/(\tau_R f)}}, \tag{15}$$

$$r = (1 - \alpha)e^{-1/(\tau_R f)}. \tag{16}$$

This discrete geometric decrease should be well-approximated by continuous exponential decay. The number of spikes needed to produce a fractional decrease of e^{-1} is given by $r^n = e^{-1}$, so that $n = -1/\log r$. Since the inter-spike interval is $1/f$, the time constant of the continuous decay will thus be given by

$$\tau = \frac{n}{f} = \frac{1}{f \log \left[(1 - \alpha)e^{-1/(\tau_R f)}\right]} = \frac{\tau_R}{1 - \tau_R f \log(1 - \alpha)}. \tag{17}$$

While this derivation is for a regular spike train, simulations (not shown) verify that it is also fits the large-scale dynamics of a Poisson spike train with the same mean frequency.

Animals

32 birds were used in this study. All 32 birds contributed to the behavioral analysis (Fig 5). Of these 32 birds, six birds were used in the deafening studies. A different subset of six birds were used in the electrophysiology experiments. During the experiments, birds were housed individually in sound-attenuating chambers (Acoustic Systems, Austin, TX), and food and water were provided ad libitum. 14:10 light:dark photo-cycles were maintained during development and throughout all experiments.

Song collection and analysis

All behavioral analyses, as well as stimulus creation, were done using custom code written in MATLAB. Individual adult male Bengalese finchs were placed in a sound-attenuating chamber (Acoustic Systems, Austin, Tx) to collect audio recordings. An automated triggering procedure was used to record and digitize (44,100 Hz) several hours of the bird's singing. These recordings were then scanned to ensure that more than 50 bouts were obtained. Bouts were defined as continuous periods of singing separated by at least 2 seconds of silence. Bengalese finch songs typically consist of between 5–12 distinct acoustic events, termed syllables, organized into probabilistic sequences. Each bout of singing consists of several renditions of sequences, with each sequence containing between 1 and approximately 40 examples of a particular syllable. The syllables from 15–50 bouts were hand labeled for subsequent analysis.

Deafening

Birds were deafened by bilateral cochlear removal [81, 82]. Complete removal of the cochlea, including the distal end of the auditory nerve, was visually confirmed using a dissecting microscope. After cochlear removal, some birds showed signs of vestibular disturbance that usually resolved in the first few days after surgery. Extra care was taken to ensure that such birds had easy access to seed and maintained full crops. Birds did not exhibit difficulty in perching, feeding, or interacting with other birds after returning to their home cages.

Electrophysiology

The electrophysiological results presented in this study were collected as part of a larger study investigating how sequences and syllable features are encoded in HVC auditory responses. The data used in this study and the associated methods have been described previously [28]. Briefly, for neural recordings, birds were placed in a large sound-attenuating chamber (Acoustic Systems, Austin, TX) and stereotaxically fixed via a previously implanted pin. During electrophysiological recordings, birds were sedated by titrating various concentrations of isoflurane in O2 using a non-rebreathing anesthesia machine (VetEquip, Pleasanton, CA). Throughout the experiment, the state of the bird was gauged by visually monitoring the eyes and respiration rate using an IR camera. Sites within HVC were at least 100 μm apart and were identified based on stereotaxic coordinates, baseline neural activity, and auditory response properties. Experiments were controlled and neural data were amplified with an AM Systems amplifier (x1000), filtered (300–10,000 Hz), and digitized at 32,000 Hz.

Playback of auditory stimuli

Stimuli were band-pass filtered between 300-8,000 Hz and normalized such that BOS playback through a speaker placed 90 cm from the head had an average sound pressure level of 80 dB at the head (A scale). Each stimulus was preceded and followed by 0.5-1 s of silence and a cosine modulated ramp was used to transition from silence to sounds. The power spectrum varied less than 5 dB across 300-8,000 Hz for white-noise stimuli. All stimuli were presented pseudo-randomly.

Creation of pseudo-random stimuli

To probe how repeated syllables are encoded in the population of HVC neurons, we used a stimulus set that consisted of 10 strings of 1000 pseudo-randomly ordered syllables was constructed. The details of this stimulus are described previously [28]. Briefly, for each bird, natural sequences (i.e. sequences produced by a given bird) and non-natural sequences (i.e.

sequences that were never produced by a bird) of length 1 through 10 were concatenated with equal probability into 10 strings of 1000 syllables. For each syllable in the birds repertoire occurring in these stimuli, a single 'prototype' syllable was used based on the distributions of acoustic features of that syllable. The median of all inter-syllable gaps was used for each gap. BOS stimuli created with these elements (synthesized BOS, prototype syllables and median gaps) elicit HVC auditory responses of comparable magnitude to normal BOS stimuli. Additionally, responses to single syllables preceded by the same long sequences in the pseudo-random stimuli are not significantly different from responses in synthesized BOS. Thus, these stimuli isolate sequence variability from other sources of variability in song, and allow investigating how HVC auditory responses to individual syllables are modulated by the preceding sequence.

Spike sorting, calculation of instantaneous firing rates, and responses to individual syllables

Single units were identified events exceeding 6 standard deviations from the mean and/or were spike sorted using in house software based on a Bayesian inference algorithm. Multi-unit neural data were thresholded to detect spikes more than 3 standard deviations away from the mean. Both single and multi-unit spike times were binned into 5 ms compartments and then smoothed using a truncated Gaussian kernel with a standard deviation of 2.5 ms and total width of 5 ms. To characterize the responses to individual target syllables, we defined a response window, which started 15 ms after the onset of the syllable and extended 15 ms after the offset of that same syllable.

Statistics

All statistical tests were performed using either paired sign-rank tests or unpaired rank-sum tests. Throughout the paper, results were considered significant if the probability of Type I errors was $\alpha < 0.05$. Bonferroni corrections were used to adjust α-values when multiple comparisons were performed.

Author Contributions

Conceived and designed the experiments: DZJ MSB JDW KEB. Performed the experiments: JDW KEB. Analyzed the data: KEB. Wrote the paper: DZJ KEB JDW MSB. Computational model: JDW DZJ. Experiments: KEB MSB.

References

1. Lashley KS. The problem of serial order in behavior. In: Jeffres LA, editor. Cerebral Mechanisms in Behaviour. Oxford, England: Wiley; 1951. p. 112–136.

2. Doupe aJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annual review of neuroscience. 1999; 22:567–631. doi: 10.1146/annurev.neuro.22.1.567 PMID: 10202549

3. Rhodes BJ, Bullock D, Verwey WB, Averbeck BB, Page MPa. Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. Human Movement Science. 2004; 23:699–746. doi: 10.1016/j.humov.2004.10.008 PMID: 15589629

4. Grossberg S. The adaptive self-organization of serial order in behavior: speech, language, and motor control. In: Schwab EC, Nusbaum HC, editors. Pattern Recognition by Humans and Machines: Speech Perception. Academic Press; 1986. p. 187.

5. James W. The Principles of Psychology. New York: Dover; 1890.

6. Hull CL. Knowledge and purpose as habit mechanisms. Psychological Review. 1930; 37(6):511. doi: 10.1037/h0072212

7. Grossberg S. Some networks that can learn, remember, and reproduce any number of complicated space-time patterns. I. Journal of Mathematics and Mechanics. 1969; 19(1):53.

8. Rosenbaum DA, Kenny SB, Derr MA. Hierarchical control of rapid movement sequences. Journal of Experimental Psychology: Human Perception and Performance. 1983; 9(1):86. PMID: 6220126

9. Summers JJ, Anson JG. Current status of the motor program: Revisited. Human Movement Science. 2009; 28(5):566–577. doi: 10.1016/j.humov.2009.01.002 PMID: 19230995

10. Lee BS. Effects of delayed speech feedback. The Journal of the Acoustical Society of America. 1950; 22(6):824–826. doi: 10.1121/1.1906696

11. MacKay DG. Self-inhibition and the disruptive effects of internal and external feedback in skilled behavior. Generation and modulation of action patterns. 1986;p. 174–186.

12. Pfordresher PQ, Palmer C. Effects of hearing the past, present, or future during music performance. Perception & Psychophysics. 2006; 68(3):362–376. doi: 10.3758/BF03193683

13. Sakata JT, Brainard MS. Real-time contributions of auditory feedback to avian vocal motor control. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2006 Sep; 26(38):9619–28. doi: 10.1523/JNEUROSCI.2027-06.2006

14. Sakata JT, Brainard MS. Online contributions of auditory feedback to neural activity in avian song control circuitry. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28 (44):11378–11390. doi: 10.1523/JNEUROSCI.3254-08.2008

15. Okanoya K. The Bengalese finch: A window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of Sciences. 2004; 1016:724–735. doi: 10.1196/annals.1298.026 PMID: 15313802

16. Warren TL, Charlesworth JD, Tumer EC, Brainard MS. Variable Sequencing Is Actively Maintained in a Well Learned Motor Skill. Journal of Neuroscience. 2012; 32(44):15414–15425. doi: 10.1523/JNEUROSCI.1254-12.2012 PMID: 23115179

17. Marler P, Peters S. Selective Vocal Learning in a Sparrow. Science. 1977; 198(4316):519–521. doi: 10.1126/science.198.4316.519 PMID: 17842140

18. Boughey MJ, Thompson NS. Song variety in the brown thrasher (Toxostoma rufum). Zeitschrift fur Tierpsychologie. 1981; 56(1):47–58.

19. Gardner TJ, Naef F, Nottebohm F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. Science (New York, NY). 2005; 308(2005):1046–1049. doi: 10.1126/science.1108214

20. Jin DZ, Kozhevnikov AA. A compact statistical model of the song syntax in Bengalese finch. PLoS Computational Biology. 2011; 7(3). doi: 10.1371/journal.pcbi.1001108 PMID: 21445230

21. Markowitz JE, Ivie E, Kligler L, Gardner TJ. Long-range Order in Canary Song. PLoS Computational Biology. 2013; 9(5). doi: 10.1371/journal.pcbi.1003052 PMID: 23658509

22. Hampton CM, Sakata JT, Brainard MS. An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. Journal of neurophysiology. 2009 Jun; 101(6):3235–45. doi: 10.1152/jn.91089.2008 PMID: 19357331

23. Fujimoto H, Hasegawa T, Watanabe D. Neural coding of syntactic structure in learned vocalizations in the songbird. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2011; 31 (27):10023–10033. doi: 10.1523/JNEUROSCI.1606-11.2011

24. Okanoya K, Yamaguchi A. Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. Journal of Neurobiology. 1997; 33:343–356. doi: 10.1002/(SICI)1097-4695(199710)33:4%3C343::AID-NEU1%3E3.0.CO;2-A PMID: 9322153

25. Woolley SM, Rubel EW. Bengalese finches Lonchura Striata domestica depend upon auditory feedback for the maintenance of adult song. The Journal of neuroscience: the official journal of the Society for Neuroscience. 1997; 17(16):6380–6390.

26. Jin DZ. Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 2009; 80:1–13.

27. Hanuschkin A, Diesmann M, Morrison A. A reafferent and feed-forward model of song syntax generation in the Bengalese finch. Journal of Computational Neuroscience. 2011; 31:509–532. doi: 10.1007/s10827-011-0318-z PMID: 21404048

28. Bouchard KE, Brainard MS. Neural encoding and integration of learned probabilistic sequences in avian sensorimotor circuitry. Journal of Neuroscience. 2013; 33(45):17710–23. doi: 10.1523/JNEUROSCI.2181-13.2013

29. Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. The Journal of Neuroscience. 2004; 24(46):10440–10453. doi: 10.1523/JNEUROSCI.1905-04.2004 PMID: 15548659

30. Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature. 2002; 419(1989):65–70. doi: 10.1038/nature00974 PMID: 12214232

31. Fee MS, Kozhevnikov Aa, Hahnloser RHR. Neural mechanisms of vocal sequence: Generation in the songbird. Annals of the New York Academy of Sciences. 2004; 1016:153–170. doi: 10.1196/annals.1298.022 PMID: 15313774

32. Jin DZ, Ramazanoglu FM, Seung HS. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. Journal of Computational Neuroscience. 2007; 23:283–299. doi: 10.1007/s10827-007-0032-z PMID: 17440800

33. Long MA, Fee MS. Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature. 2008; 456(November):189–194. doi: 10.1038/nature07448 PMID: 19005546

34. Chang W, Jin DZ. Spike propagation in driven chain networks with dominant global inhibition. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 2009; 79:1–5.

35. Long MA, Jin DZ, Fee MS. Support for a synaptic chain model of neuronal sequence generation. Nature. 2010; 468(7322):394–399. Available from: http://dx.doi.org/10.1038/nature09514 doi: 10.1038/nature09514 PMID: 20972420

36. Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, Serinus canarius. The Journal of comparative neurology. 1976; 165:457–486. doi: 10.1002/cne.901650405 PMID: 1262540

37. Nottebohm F, Paton JA, Kelley DB. Connections of vocal control nuclei in the canary telencephalon. Journal of Comparative Neurology. 1982; 207(4):344–357. doi: 10.1002/cne.902070406 PMID: 7119147

38. Schmidt MF. Pattern of interhemispheric synchronization in HVc during singing correlates with key transitions in the song pattern. Journal of neurophysiology. 2003; 90(August 2003):3931–3949. doi: 10.1152/jn.00003.2003 PMID: 12944542

39. Lewandowski B, Vyssotski A, Hahnloser RH, Schmidt M. At the interface of the auditory and vocal motor systems: NIf and its role in vocal processing, production and learning. Journal of Physiology-Paris. 2013; 107(3):178–192. doi: 10.1016/j.jphysparis.2013.04.001

40. Abeles M. Corticonics: Neural circuits of the cerebral cortex. Cambridge University Press; 1991.

41. Gibb L, Gentner TQ, Abarbanel HDI. Brain stem feedback in a computational model of birdsong sequencing. Journal of neurophysiology. 2009; 102(June 2009):1763–1778. doi: 10.1152/jn.91154.2008 PMID: 19553477

42. Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nature neuroscience. 2004; 7(7):701–702. doi: 10.1038/nn1263 PMID: 15184903

43. Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage. 2010; 50(1):291–301. doi: 10.1016/j.neuroimage.2009.12.035 PMID: 20026224

44. Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. The Journal of Neuroscience. 1992; 12(November):4309–4326. PMID: 1432096

45. Lewicki MS, Konishi M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. Proceedings of the National Academy of Sciences of the United States of America. 1995; 92(June):5582–5586. doi: 10.1073/pnas.92.12.5582 PMID: 7777552

46. Prather JF, Peters S, Nowicki S, Mooney R. Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature. 2008; 451(January):305–310. doi: 10.1038/nature06492 PMID: 18202651

47. Fortune ES, Margoliash D. Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata). Journal of Comparative Neurology. 1995; 360:413–441. doi: 10.1002/cne.903600305 PMID: 8543649

48. Vates GE, Broome BM, Mello CV, Nottebohm F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (Taenopygia guttata). Journal of Comparative Neurology. 1996; 366:613–642. doi: 10.1002/(SICI)1096-9861(19960318)366:4%3C613::AID-CNE5%3E3.0.CO;2-7 PMID: 8833113

49. Coleman MJ, Mooney R. Synaptic transformations underlying highly selective auditory representations of learned birdsong. The Journal of neuroscience. 2004; 24(33):7251–7265. doi: 10.1523/JNEUROSCI.0947-04.2004 PMID: 15317851

50. Cardin JA, Raksin JN, Schmidt MF. Sensorimotor nucleus NIf is necessary for auditory processing but not vocal motor output in the avian song system. Journal of neurophysiology. 2005 Apr; 93(4):2157–66. doi: 10.1152/jn.01001.2004 PMID: 15590726

51. Lei H, Mooney R. Manipulation of a Central Auditory Representation Shapes Learned Vocal Output. Neuron. 2010; 65(1):122–134. doi: 10.1016/j.neuron.2009.12.008 PMID: 20152118

52. Markram H, Tsodyks M, et al. Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature. 1996; 382(6594):807–810. doi: 10.1038/382807a0 PMID: 8752273

53. Abbott L, Varela J, Sen K, Nelson S. Synaptic depression and cortical gain control. Science. 1997; 275 (5297):221–224. doi: 10.1126/science.275.5297.221

54. Woolley SMN, Rubel EW. Vocal memory and learning in adult Bengalese Finches with regenerated hair cells. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2002; 22 (17):7774–7787.

55. Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K. Invariance and sensitivity to intensity in neural discrimination of natural sounds. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28(25):6304–6308. doi: 10.1523/JNEUROSCI.0961-08.2008

56. Drew PJ, Abbott LF. Model of song selectivity and sequence generation in area HVc of the songbird. Journal of neurophysiology. 2003; 89:2697–2706. doi: 10.1152/jn.00801.2002 PMID: 12612042

57. Beckers GJL, Gahr M. Neural processing of short-term recurrence in songbird vocal communication. PLoS ONE. 2010; 5(6). doi: 10.1371/journal.pone.0011129

58. Roy A, Mooney R. Song decrystallization in adult zebra finches does not require the song nucleus NIf. Journal of neurophysiology. 2009; 102(June 2009):979–991. doi: 10.1152/jn.00293.2009 PMID: 19515953

59. Roberts TF, Klein ME, Kubke MF, Wild JM, Mooney R. Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2008; 28(13):3479–3489. doi: 10.1523/JNEUROSCI.0177-08.2008

60. Foster EF, Mehta RP, Bottjer SW. Axonal connections of the medial magnocellular nucleus of the anterior neostriatum in zebra finches. The Journal of comparative neurology. 1997; 382(December 1996):364–381. doi: 10.1002/cne.903820305 PMID: 9183699

61. Wild JM. Neural pathways for the control of birdsong production. Journal of Neurobiology. 1997; 33 (June):653–670. doi: 10.1002/(SICI)1097-4695(19971105)33:5%3C653::AID-NEU11%3E3.0.CO;2-A PMID: 9369465

62. Ashmore RC, Wild JM, Schmidt MF. Brainstem and forebrain contributions to the generation of learned motor behaviors for song. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2005; 25(37):8543–8554. doi: 10.1523/JNEUROSCI.1668-05.2005

63. Hamaguchi K, Tschida Ka, Yoon I, Donald BR, Mooney R. Auditory synapses to song premotor neurons are gated off during vocalization in zebra finches. eLife. 2014; 2014:1–21.

64. Vallentin D, Long MA. Motor Origin of Precise Synaptic Inputs onto Forebrain Neurons Driving a Skilled Behavior. Journal of Neuroscience. 2015; 35(1):299–307. doi: 10.1523/JNEUROSCI.3698-14.2015 PMID: 25568122

65. Jordan MI. Computational aspects of motor control and motor learning. Handbook of perception and action: motor skills. 1996; 2:71–118. doi: 10.1016/S1874-5822(06)80005-8

66. Desmurget M, Grafton S. Forward modeling allows feedback control for fast reaching movements. Trends in cognitive sciences. 2000; 4(11):423–431. doi: 10.1016/S1364-6613(00)01537-0 PMID: 11058820

67. Civier O, Tasko SM, Guenther FH. Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. Journal of fluency disorders. 2010; 35(3):246–279. doi: 10.1016/j.jfludis.2010.05.002 PMID: 20831971

68. Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in human neuroscience. 2011; 5. doi: 10.3389/fnhum.2011.00082 PMID: 22046152

69. Keller GB, Hahnloser RHR. Neural processing of auditory feedback during vocal practice in a songbird. Nature. 2009; 457(January):187–190. doi: 10.1038/nature07467 PMID: 19005471

70. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. Science. 1995; 269(5232):1880–1882. doi: 10.1126/science.7569931 PMID: 7569931

71. Helekar SA, Espino GG, Botas A, Rosenfield DB. Development and adult phase plasticity of syllable repetitions in the birdsong of captive zebra finches (Taeniopygia guttata). Behavioral neuroscience. 2003; 117(5):939–951. doi: 10.1037/0735-7044.117.5.939 PMID: 14570544

72. Kent RD. Research on speech motor control and its disorders: A review and prospective. Journal of Communication disorders. 2000; 33(5):391–428. doi: 10.1016/S0021-9924(00)00023-X PMID: 11081787

73. Voss HU, Salgado-Commissariat D, Helekar Sa. Altered auditory BOLD response to conspecific birdsong in zebra finches with stuttered syllables. PLoS ONE. 2010; 5(12):e4415. doi: 10.1371/journal.pone.0014415

74. Van Riper C. The nature of stuttering. Prentice Hall; 1982.

75. Zebrowski PM. Duration of Sound Prolongation and Sound/Syllable Repetition in Children Who Stutter-Preliminary Observations. Journal of Speech, Language, and Hearing Research. 1994; 37(2):254–263. doi: 10.1044/jshr.3702.254

76. Boey RA, Wuyts FL, Van de Heyning PH, De Bodt MS, Heylen L. Characteristics of stuttering-like disfluencies in Dutch-speaking children. Journal of fluency disorders. 2007; 32(4):310–329. doi: 10.1016/j.jfludis.2007.07.003 PMID: 17963939

77. Martin R, Haroldson SK. Effects of five experimental treatments on stuttering. Journal of Speech, Language, and Hearing Research. 1979; 22(1):132–146. doi: 10.1044/jshr.2201.132

78. Postma A, Kolk H. The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. Journal of Speech, Language, and Hearing Research. 1992; 35(3):537–544. doi: 10.1044/jshr.3503.537

79. Stuart A, Frazier CL, Kalinowski J, Vos PW. The effect of frequency altered feedback on stuttering duration and type. Journal of Speech, Language, and Hearing Research. 2008; 51(4):889–897. doi: 10.1044/1092-4388(2008/065) PMID: 18658059

80. Lane H, Tranel B. The Lombard sign and the role of hearing in speech. Journal of Speech, Language, and Hearing Research. 1971; 14(4):677–709. doi: 10.1044/jshr.1404.677

81. Konishi M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Zeitschrift fur Tierpsychologie. 1965; 22:770–783. PMID: 5874921

82. Brainard MS, Doupe AJ. Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2001; 21(7):2501–2517.