

An Associational Model of Birdsong Sensorimotor Learning

II. Temporal Hierarchies and the Learning of Song Sequence

TODD W. TROYER¹ AND ALLISON J. DOUPE²

^{1,2}*Department of Psychiatry, ²Department of Physiology, ^{1,2}W. M. Keck Center for Integrative Neuroscience, and ²Sloan Center for Theoretical Neurobiology at UCSF, University of California, San Francisco, California 94143-0444*

Received 16 February 2000; accepted in final form 2 May 2000

Troyer, Todd W. and Allison J. Doupe. An associational model of birdsong sensorimotor learning. II. Temporal hierarchies and the learning of song sequence. *J Neurophysiol* 84: 1224–1239, 2000. Understanding the neural mechanisms underlying serially ordered behavior is a fundamental problem in motor learning. We present a computational model of sensorimotor learning in songbirds that is constrained by the known functional anatomy of the song circuit. The model subsumes our companion model for learning individual song “syllables” and relies on the same underlying assumptions. The extended model addresses the problem of learning to produce syllables in the correct sequence. Central to our approach is the hypothesis that the Anterior Forebrain Pathway (AFP) produces signals related to the comparison of the bird’s own vocalizations and a previously memorized “template.” This “AFP comparison hypothesis” is challenged by the lack of a direct projection from the AFP to the song nucleus HVC, a candidate site for the generator of song sequence. We propose that sequence *generation* in HVC results from an associative chain of motor and sensory representations (motor → sensory → next motor . . .) encoded within the two known populations of HVC projection neurons. The sensory link in the chain is provided, not by auditory feedback, but by a centrally generated efference copy that serves as an internal prediction of this feedback. The use of efference copy as a substitute for the sensory signal explains the ability of adult birds to produce normal song immediately after deafening. We also predict that the AFP guides sequence *learning* by biasing motor activity in nucleus RA, the premotor nucleus downstream of HVC. Associative learning then remaps the output of the HVC sequence generator. By altering the motor pathway in RA, the AFP alters the correspondence between HVC motor commands and the resulting sensory feedback and triggers renewed efference copy learning in HVC. Thus, auditory feedback-mediated efference copy learning provides an indirect pathway by which the AFP can influence sequence generation in HVC. The model makes predictions concerning the role played by specific neural populations during the sensorimotor phase of song learning and demonstrates how simple rules of associational plasticity can contribute to the learning of a complex behavior on multiple time scales.

INTRODUCTION

Like many complex behaviors, birdsong is arranged in a temporal hierarchy. In zebra finches, song consists of a few short introductory notes, followed by several repetitions of a stereotyped sequence of vocal gestures, or “syllables,” separated by brief periods of silence (Sossinka and Böhner 1980). Song is learned in two phases. First, birds listen to

and memorize a tutor song, or “template” (Konishi 1965; Marler 1964). Later, during *sensorimotor* learning, birds use auditory feedback from their own vocalizations to gradually match their vocal output to the template. In the companion paper (Troyer and Doupe 2000), we focused on one level of the hierarchy for song and showed how simple associational (Hebbian) learning rules could be used to learn the motor representations for individual tutor syllables. The syllable learning model addresses the important problem of feedback delay and demonstrates that associational plasticity naturally leads to the learning of an efference copy, or internal prediction, of the auditory feedback. This internal prediction can then be compared with the memorized tutor song to guide sensorimotor learning.

In this paper, we address a second fundamental problem in motor learning, the question of serial order in behavior (Lashley 1951), by extending our syllable learning model to account for the learning of syllable sequence. As in our companion paper, we use simple rules of associational plasticity and assume that the template comparison signals that guide learning are provided by the Anterior Forebrain Pathway (AFP), a circuit that passes through avian basal ganglia, thalamic, and cortex-like nuclei before projecting back onto the motor pathway (see Troyer and Doupe 2000; Fig. 1). We also assume a functional segregation between the two known populations of projection neurons in song nucleus HVC (Nordeen and Nordeen 1988; HVC used as proper name, Margoliash et al. 1994), with AFP-projecting HVC neurons (HVC_AFP) receiving auditory feedback and encoding signals in sensory coordinates, and HVC neurons projecting to the robust nucleus of the archistriatum (RA; HVC_RA) more closely tied to a motor code (Troyer and Doupe 2000). The main biological constraint addressed in this paper is the hierarchical organization of the motor pathway (Fig. 1): the detailed motor programs for individual syllables are believed to be contained in nucleus RA (Vu et al. 1994; Yu and Margoliash 1996), whereas the central pattern generator for song sequence is likely to be found upstream of RA, perhaps within the song nucleus HVC (Vu et al. 1994). Our sequence learning model addresses two key questions left unanswered by current experimental data: what is the mechanism for sequence generation in HVC, and how can signals from the AFP guide

Present address and address for reprint requests: T. W. Troyer, Dept. of Psychology, University of Maryland, College Park, MD 20742 (E-mail: ttroyer@psyc.umd.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

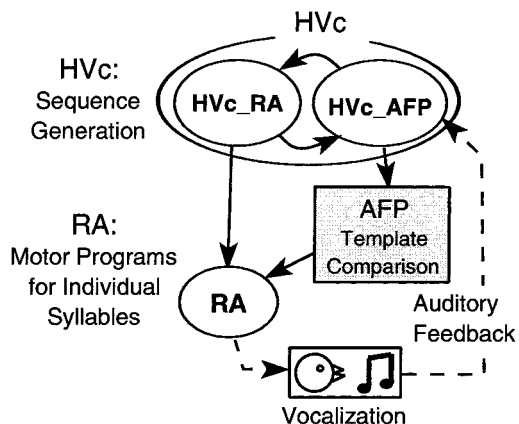


FIG. 1. Encoding of motor hierarchy within the song circuit. The fine temporal structure within individual song syllables is believed to be encoded in RA (Yu and Margoliash 1996). The pattern generator for song sequence is believed to be located upstream of RA, possibly within HVC (Vu et al. 1994). We have shown the two populations of HVC projection neurons (Nordeen and Nordeen 1988) in separate ovals, although these are intermixed and interconnected in HVC. We assume auditory feedback enters the song system via inputs to Hvc_AFP.

sequence learning given that there are no known connections from the AFP to Hvc?

We propose that sequence generation results from a reciprocal chaining of motor and sensory representations [motor (Hvc_RA) \rightarrow sensory (Hvc_AFP) \rightarrow next motor (Hvc_RA) \rightarrow next sensory (Hvc_AFP) . . .] between the two populations of HVC projection neurons. Our model differs from classic “associative chaining” models (James 1983) in that the “sensory” component in this chain is actually an efference copy, a motor signal that serves as a prediction of the expected sensory feedback (Sperry 1950; von Holst and Mittelstaedt 1980). We also propose that AFP-guided teaching signals act to remap the connections from HVC to RA, so that the output of the HVC pattern generator maps onto the sequence of motor features (encoded in RA) that matches the memorized tutor song (cf. Doya and Sejnowski 1998). However, simply remapping HVC outputs cannot explain AFP-guided learning *within* HVC. In our model, auditory feedback-driven efference copy learning provides the crucial link between the AFP and HVC. By altering the HVC outflow tract, the AFP alters the association between Hvc_RA motor activity and the auditory feedback received by Hvc_AFP. The resulting efference copy learning then changes the motor-sensory interaction underlying sequence generation in HVC.

Our model demonstrates that associational learning, distributed throughout the motor pathway, is sufficient for learning both individual syllables and their proper sequence. The model provides a specific hypothesis for how basal ganglia–forebrain loops could contribute to learning a sequential behavior and highlights key computational problems imposed by the functional anatomy of the song circuit. More generally, the model provides a framework that relates the neural mechanisms underlying song learning to fundamental problems in motor learning and speech production.

Model and approach

In this paper, we extend our previous model for learning individual syllables (Troyer and Doupe 2000) to address the

learning of syllable sequence. Our sequence learning model subsumes our syllable model, accomplishing syllable learning as well as the learning of syllable sequence. The structure of this paper mirrors that of the preceding companion paper (Troyer and Doupe 2000) and relies on the same underlying biological assumptions. We present our results in the form of two closely related models: a “conceptual model” containing a self-consistent set of functional hypotheses, and a “computational model” that incorporates these hypotheses into a working computer algorithm. In this section, we describe the functional problems addressed by our sequence learning model and outline the key elements of our proposed solutions. Then, we present our conceptual model, which describes our functional hypotheses in greater detail. Quantitative results from our computational model are presented in the RESULTS section. Because our model is relatively abstract at the level of local circuits, implementation of these hypotheses was governed chiefly by considerations of computational simplicity. Related issues are described in the METHODS but are not crucial for understanding the main functional implications of the model. The details of our computer algorithm are confined to an appendix.

Problems addressed

Our model explores how song learning can result from associational learning, guided by template comparison signals transmitted by the AFP. We do not address learning the detailed temporal structure within each syllable, nor learning the length of syllables and intersyllable gaps. Timing of song syllables is provided by a rhythmically clocked premotor drive arriving in Hvc_RA (Troyer and Doupe 2000). While the *timing* of this drive is fixed, its *pattern* is completely random; the magnitude of each component of the premotor input is generated independently for each vocalization produced by the model. The model’s task is to take this unstructured premotor timing signal and convert it to a sequence of syllables matched to the tutor template.

The learning of motor representations for individual song syllables was addressed in the preceding companion paper (Troyer and Doupe 2000). This model contained three key functional elements. By associating premotor commands in Hvc_RA with auditory feedback arriving in Hvc_AFP, a motor \rightarrow sensory efference copy mapping develops between the two populations of HVC projection neurons (Fig. 2, marked 1). After this mapping develops, Hvc_AFP activity driven by a given Hvc_RA motor command encodes a sensory prediction of the vocal output resulting from that command. This prediction is then compared with the template in the AFP, resulting in a global reinforcement signal that modulates plasticity in all RA neurons (Fig. 2, marked 2). This reinforcement learning leads to a pattern of connectivity in RA in which neurons encoding the same tutor syllable become strongly connected (Fig. 2, marked 3). As a result, RA has a strong tendency to produce coherent patterns of motor activity matched to the syllables in the tutor template.

Given our adoption of the AFP comparison hypothesis, the most difficult problem regarding sequence learning is the following: how can the AFP guide learning given that 1) the only known output from the AFP projects to RA, and 2) the site of sequence generation is likely to be upstream of RA? Our

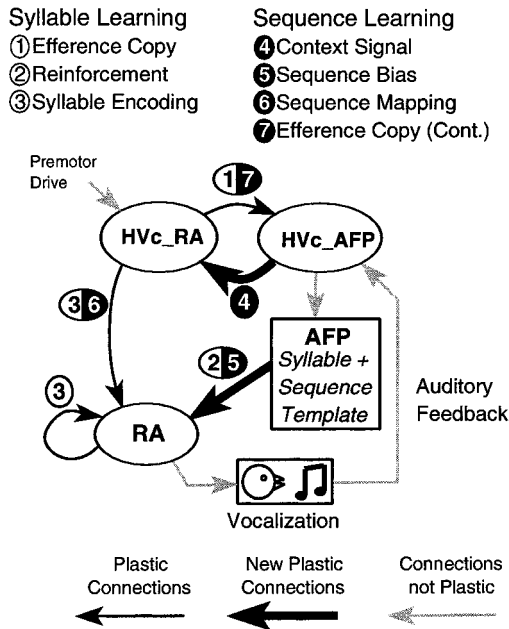


FIG. 2. Network architecture. White circles: functional connections addressed in our syllable learning model (Troyer and Doupe 2000). Black circles: functional connections important for sequence learning (Troyer and Doupe 2000). The association of Hvc_RA premotor activity and auditory feedback input to leads to a motor \rightarrow sensory efference copy mapping between these neural populations (1). Reinforcement signals from the Anterior Forebrain Pathway (AFP) (2) are used to reorganize intrinsic RA connections so that they encode the motor representations for individual tutor syllables (3). Sequence generation results from a reciprocal interaction involving the sensory \rightarrow motor efference copy mapping followed by a slow “context” signal that flows from Hvc_AFP \rightarrow Hvc_RA (4). The AFP uses template information to bias RA toward the appropriate syllable transitions (5). This alters associations in the motor pathway so that the connections from Hvc_RA \rightarrow RA map the output of the Hvc pattern generator onto the correct syllable representations in RA (6). Alterations in the motor pathway lead to renewed efference copy learning (7). Black arrows: plastic connections. Thick arrows: new connections added to address sequence learning. Gray arrows: connections not subject to associational plasticity.

solution involves the concerted action of multiple associational mechanisms acting at different levels of the motor hierarchy. For ease of presentation, we will break this problem into three smaller problems, described below (see *Conceptual model*). However, our choice of solution to each individual problem is affected by the other two, as well as constraints imposed by our solution to the problem of syllable learning. The key to our model is the concept of efference copy, which serves to link all model components into a coherent hypothesis regarding the multiple sensory-motor interactions involved in song learning.

The first problem we address is the problem of sequence generation, i.e., what is the nature of the central pattern generator for song? We propose that sequence generation results from a reciprocal interaction between the two populations of Hvc projection neurons (Table 1, number 1). The solution naturally incorporates the mechanism of efference copy, which contributes one half of this interaction by providing a motor \rightarrow sensory mapping from Hvc_RA \rightarrow Hvc_AFP. The other half of the interaction depends on connections from Hvc_AFP \rightarrow Hvc_RA. These are hypothesized to provide slow signals carrying information from one syllable to the next (Fig. 2, marked 4). We call such signals “context” signals. Thus, se-

quences are generated as a chain of mappings from motor \rightarrow sensory \rightarrow next motor \rightarrow next sensory, etc. This hypothesis borrows from classical chaining ideas (James 1983), as well as more recent computational models (Kleinfeld and Sompolinsky 1988) of sequence generation.

The second problem we address is the problem of how AFP signals guide sequence learning at the level of RA. The most straightforward method of directing associational learning toward a desired goal is to bias the pattern of neural activity toward the desired state. Associational plasticity then strengthens the connections consistent with this pattern. In our model, we assume that the AFP generates an expectation of the next syllable in the tutor sequence and uses this expectation to bias RA activity (Table 1, number 2; Fig. 2, marked 5). Associational plasticity then changes the pattern of connections between Hvc and RA so that syllables are produced in the proper sequence (Fig. 2, marked 6). Note that this solution gives rise to an additional problem to be solved before the AFP can bias RA activity in the proper direction: template information is stored in sensory coordinates, but the required bias must be in motor coordinates. We propose that a sensory \rightarrow motor mapping is learned between the AFP and RA soon after the initial period of efference copy learning (Table 1, number 3; see *Conceptual model*).

The third problem we address is the problem of sequence learning at the level of Hvc. While the mechanism outlined above is sufficient for a rudimentary form of sequence learning, it fails as a complete model. In particular, it fails to account for any learned changes in the number or sequence of premotor commands formed upstream of RA. In our model, the efference copy provides the key link between learning at the level of RA and learning upstream of RA, in Hvc. In particular, by altering connections between Hvc and RA, the AFP changes the pattern of vocal output and hence auditory reafference. This in turn induces new efference copy learning in Hvc (Table 1, number 4; Fig. 2, marked 7) via the same mechanism described in our syllable learning model (Troyer and Doupe 2000). Since efference copy mapping plays a key role in the Hvc pattern generator, the new efference copy learning alters the sequence of Hvc outputs (see *Conceptual model*). In addition to providing a specific mechanism for how the AFP affects sequence generation in Hvc, the need for ongoing efference copy learning is consistent with experiments demonstrating that auditory feedback is required throughout development (Price 1979).

In addressing the problem of sequence learning, we have added two new sets of connections to our model for syllable learning (Fig. 2). The connections from Hvc_AFP \rightarrow Hvc_RA are necessary for *sequence generation*. Without the context signals carried by these connections, activity within

TABLE 1. *Functional hypotheses for sequence learning*

1. Reciprocal interaction in Hvc
A. Fast motor \rightarrow sensory efference copy signal (Hvc_RA \rightarrow Hvc_AFP)
B. Slow sensory \rightarrow motor “context” signal (Hvc_AFP \rightarrow Hvc_RA)
2. AFP sequence teaching signals bias RA activity toward next tutor syllable
3. Sensory \rightarrow motor associations guide development of connections from AFP \rightarrow RA
4. AFP guided changes in RA trigger renewed efference copy learning in Hvc

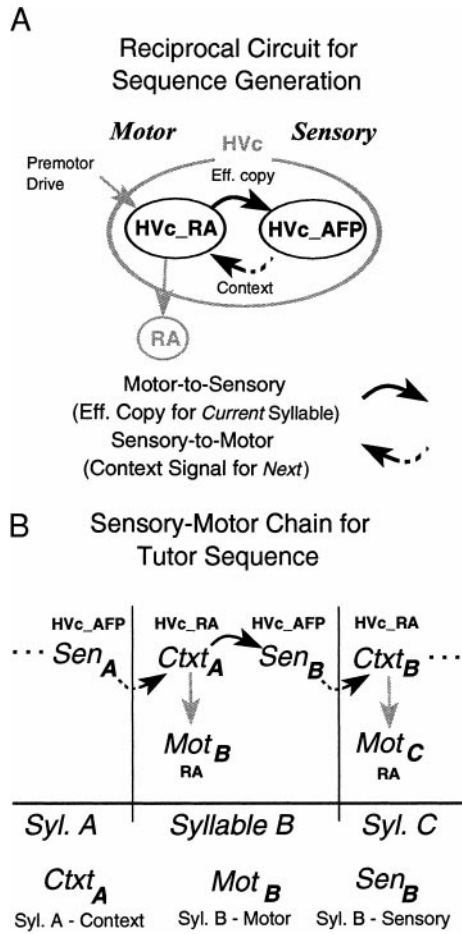


FIG. 3. Sequence generation. A: sequence generation results from a reciprocal interaction between representations in HVC_RA (motor) and HVC_AFP (sensory). The final motor output of the model depends on the mapping from HVC_RA → RA. B: schematic of the mappings necessary for correct reproduction of the tutor sequence ... A → B → C ... Suppose an efference copy, Sen_A , is represented in HVC_AFP (in sensory coordinates). This is followed by the HVC_RA context representation $Ctxt_A$, which is mapped onto Mot_B in RA. Syllable B follows syllable A. $Ctxt_A$ also elicits Sen_B , the efference copy corresponding to Mot_B . $Sen_B \rightarrow Ctxt_B \rightarrow Mot_C$ leads to the production of syllable C, etc.

HVC_RA would not be affected by activity related to the previous syllable and the sequence of HVC outputs would be random (Troyer and Doupe 2000). Patterned connections from the AFP → RA are necessary for *sequence learning* in our model. Without these connections, information stored in the AFP related to the tutor sequence cannot be used to guide learning in the motor pathway.

Conceptual model

PROBLEM 1: SEQUENCE GENERATION. We propose that sequences of song syllables are generated by a reciprocal interaction between motor (HVC_RA) and sensory/efference copy (HVC_AFP) activity within HVC (Table 1, number 1): motor → sensory prediction → next motor → next sensory prediction → ... (Fig. 3A). The motor → sensory component of this interaction is subserved by the efference copy mapping between HVC_RA and HVC_AFP. This mapping is learned early in development by associating HVC_RA motor commands with auditory feedback arriving back in HVC_AFP, as

described in our model for syllable learning (Troyer and Doupe 2000). Figure 3B shows how these mappings result in the reproduction of the tutor song *after learning is complete*, using the transition from syllable A to syllable B as an example. Let Sen_A denote the sensory representation for syllable A in HVC_AFP. This representation is elicited by the efference copy mapping during production of A. Via the connections from HVC_AFP → HVC_RA, Sen_A elicits a context signal $Ctxt_A$ that drives activity in HVC_RA during the syllable following syllable A. $Ctxt_A$ maps onto the motor representation Mot_B in RA, and the model produces syllable B after syllable A. This is the sensory prediction → next motor component of the interaction. With an accurate efference copy mapping, $Ctxt_A$ also elicits an efference copy representation Sen_B in HVC_AFP. This motor → sensory prediction component of the interaction completes the cycle. Thus, correct sequence learning in our model depends on learning the chain of mappings $Sen_A \rightarrow (Ctxt_A \rightarrow Mot_B) \rightarrow Sen_B \rightarrow \dots$. Note that our implementation of this functional circuit is highly simplified: HVC_RA → HVC_AFP connections transmit *only* fast motor → sensory (efference copy) signals, whereas HVC_AFP → HVC_RA connections transmit *only* slow sensory → next motor (context) signals. More realistic circuit models of HVC will be required to explore possible local circuit mechanisms subserving this reciprocal flow of activity.

PROBLEM 2: SEQUENCE LEARNING IN RA. In our model, the AFP uses template information to generate “sequence teaching” signals that bias RA activity toward the proper tutor sequence (Table 1, number 2). The details of how these signals reorganize the motor pathway to produce correct sequence transitions are illustrated in Fig. 4, using the transition from syllable A to syllable B as an example. In our model, the efference copy representation, Sen_A , that is registered in HVC_AFP during the production of syllable A, generates two distinct signals during

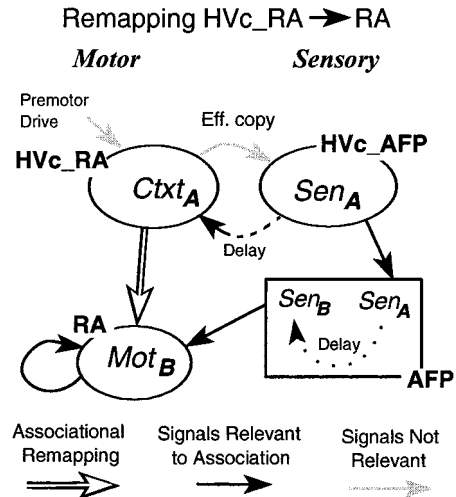


FIG. 4. AFP-guided sequence learning. Schematic showing the learning of the transition from syllable A to B. The efference copy for A, Sen_A , results in a context signal, $Ctxt_A$, that arrives in HVC_RA after a delay. Sen_A is also passed on to the AFP. Using previously stored template information, the AFP generates, after an appropriate delay, the sensory representation for the next syllable in the tutor song, Sen_B . Sen_B biases RA toward the motor pattern Mot_B . Associational learning (white arrow) between the context signal $Ctxt_A$ in HVC_RA and Mot_B in RA ensures that future productions of syllable A will evoke the composite mapping $Sen_A \rightarrow Ctxt_A \rightarrow Mot_B$, resulting in the transition from A to B.

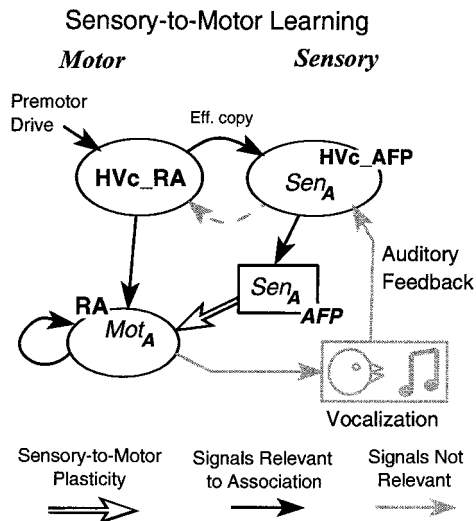


FIG. 5. Learning a sensory \rightarrow motor mapping between AFP and RA. After the initial phase of efference copy learning, the Hvc_RA activity that produces motor activity for syllable A in RA (Mot_A) will also produce a sensory prediction (Sen_A) of that motor activity in the AFP (black arrows). This leads to associational learning between AFP assemblies encoding syllable A in sensory coordinates and the RA assemblies encoding syllable A in motor coordinates (white arrow).

the vocalization that follows syllable A. First, in Hvc, due to the slow connections from Hvc_AFP \rightarrow Hvc_RA, Sen_A results in a context signal, $Ctxt_A$, that is input to Hvc_RA. Second, the AFP receives the efference copy Sen_A from Hvc_AFP and generates the sequence teaching signal for syllable B, after an appropriate delay. This signal is input to RA and biases RA activity toward the next motor representation in the tutor sequence, Mot_B . Since both of these signals exert their effects with a one syllable delay, during the syllable following A, neurons in Hvc_RA that are part of the context representation $Ctxt_A$ tend to be co-active with RA neurons comprising the motor representation Mot_B . Associational learning then strengthens the connections between these sets of neurons (Fig. 4, white arrow). In this way, the context representation $Ctxt_A$ gets mapped onto Mot_B , and the model learns the transition $Sen_A \rightarrow Ctxt_A \rightarrow Mot_B$.

SENSORY \rightarrow MOTOR MAPPING FROM THE AFP \rightarrow RA. If the sequence teaching signal for syllable B, which we assume to be encoded in sensory coordinates in the AFP, is to bias RA motor activity toward syllable B, a sensory \rightarrow motor mapping between the AFP and RA is required (Table 1, number 3). In our sequence learning model, the required map develops soon after the initial period of efference copy learning, and before syllable learning is complete. With an accurate efference copy, Hvc_RA excites a sensory representation in the output neurons of the AFP (via Hvc_AFP) that corresponds to the motor activity in RA. For example, if Hvc_RA drives motor activity in RA that is relatively well matched to tutor syllable A, it will also drive an efference copy within Hvc_AFP that leads to excitation within the AFP output neurons encoding tutor syllable A (Fig. 5). Associative learning then strengthens connections between the AFP neurons encoding syllable A in sensory coordinates and the RA neurons encoding A in motor coordinates. Note that to *develop* the appropriate mapping between the AFP and RA, the output neurons in the AFP must encode a sensory representation of the *current* syllable. To *use* the map

to bias RA activity toward the tutor sequence, these same AFP output neurons must encode a representation of the *next* syllable. Our model simply assumes that AFP efferents contain a combination of these signals. Possible explanations for how the components of this mixed signal could exert distinct functional influences in RA are described in the METHODS.

PROBLEM 3: SEQUENCE LEARNING IN HVC. Even though the model has learned the correct efference copy \rightarrow next motor transition, $Sen_A \rightarrow Ctxt_A \rightarrow Mot_B$, sequence learning is not yet complete. This is because by altering synapses in RA, the AFP has perturbed the motor \rightarrow sensory matching necessary for an accurate efference copy in Hvc. In particular, Hvc_RA neurons belonging to the representation for $Ctxt_A$ originally mapped onto some particular combination of motor representations in RA. For example, perhaps $Ctxt_A$ originally mapped most strongly onto syllable D. With an accurate efference copy, these same Hvc_RA neurons were mapped onto the corresponding combination of sensory representations in Hvc_AFP, Sen_D . Remapping $Ctxt_A$ onto Mot_B in RA alters this correspondence, and the Hvc sequence generator produces the following set of mappings: $Sen_A \rightarrow Ctxt_A \rightarrow Sen_D \rightarrow Ctxt_D$. Presumably, the context signal from syllable D, $Ctxt_D$, is mapped onto Mot_E in RA. Therefore, syllable B (produced by $Ctxt_A$) will be followed, not by C, but by E. However, such errors in the efference copy component of the Hvc sequence generator are continually corrected by renewed auditory feedback-driven learning in the Hvc_RA \rightarrow Hvc_AFP connections (Table 1, number 4): $Ctxt_A$ excites Mot_B in RA, leading to an auditory feedback signal Sen_B arriving in Hvc_AFP (Fig. 6). Therefore, Hvc_RA \rightarrow Hvc_AFP connections between Hvc_RA neurons belonging to $Ctxt_A$ and Hvc_AFP neurons belonging to Sen_B are strengthened (Fig. 6, white arrow), supplanting the “old” connections from $Ctxt_A \rightarrow Sen_D$. In this way, the Hvc sequence generator is able to track the AFP-induced changes in RA. By combining the appropriate sensory \rightarrow motor and motor \rightarrow sensory mappings, the model learns the chain of sensory-motor associations that reproduces the tutor sequence: $Sen_A \rightarrow (Ctxt_A \rightarrow Mot_B) \rightarrow Sen_B \dots$

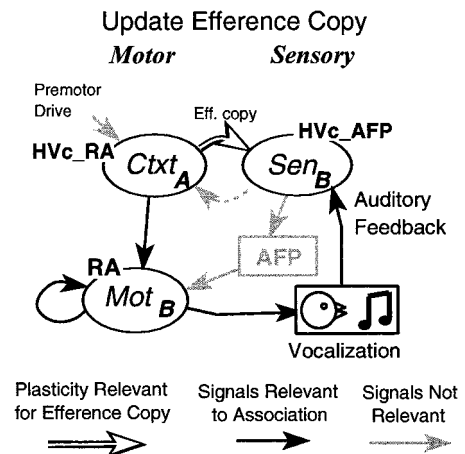


FIG. 6. Renewed efference copy learning. Since AFP-guided sequence learning alters the projection from Hvc_RA \rightarrow RA, renewed efference copy learning (white arrow) is required so that $Ctxt_A$ projects onto Sen_B in Hvc_AFP (cf. Troyer and Doupe 2000, Fig. 4A). Thus, auditory feedback is necessary throughout development to maintain an accurate efference copy.

METHODS

The model presented in this paper is an extension of the syllable learning model described in the preceding companion paper (Troyer and Doupe 2000). To account for the generation and learning of song sequence, we added two new sets of synaptic connections to this model (Fig. 2B). Because our model is relatively abstract at the level of local circuits, the choice of how these connections were embedded in our computer algorithm was governed chiefly by considerations of computational simplicity (a variety of biological mechanisms could contribute to their functionality). An understanding of the theoretical issues related to our implementation is not necessary to understand our simulation results. Most features of the model are described in detail in Troyer and Doupe (2000). We discuss here only new additions to the model. The final subsection in the METHODS describes the method we used for quantifying the time course of model development.

Most simulations of the complete model contained 25,000 syllables, over 5,000 more than were typically needed for model output to become stereotyped (see APPENDIX). Computer simulations were written using the MATLAB simulation environment (version 5.3; The Mathworks, Natick, MA). Typical simulations took ≈ 3 h when run using a 400-MHz Pentium II processor. Details regarding simulations and parameters are contained in the APPENDIX.

HVc_AFP \rightarrow HVc_RA connections

To account for sequence generation, connections from HVc_AFP to HVc_RA were added (Fig. 2B). These connections are assumed to be functionally “slow synapses” that carry information from one syllable to the next (cf. Kleinfeld and Sompolinsky 1988). For computational simplicity, the functional separation of HVc connections was strict: HVc_RA \rightarrow HVc_AFP connections carried *only* efference copy information related to the current syllable, and the HVc_AFP \rightarrow HVc_RA connections broadcast signals that affected only the subsequent syllable. However, our general approach requires only a functional imbalance between the two populations of HVc projection neurons. A strict separation is not crucial. To match the functional delay in the HVc_AFP \rightarrow HVc_RA pathway (≈ 50 ms), a corresponding delay was introduced in the time window for synaptic plasticity in these connections (see APPENDIX). In general, we followed the principle that the time window for synaptic plasticity should be roughly proportional to the time scale of encoding for the information passed over that synapse. RA connections, which encode the detailed motor programs within each syllable, had the shortest plasticity window, and the HVc_AFP \rightarrow HVc_RA context synapses had the longest.

Since it relies on reciprocal excitatory connections, the pattern generator within HVc tended to be unstable. To help control this positive feedback, we 1) normalized the size of the context signal during each syllable (see APPENDIX), and 2) included “adaptation” in the HVc_RA assemblies. HVc_RA adaptation was of the same form as the HVc_AFP adaptation included to cancel the delayed auditory feedback (Troyer and Doupe 2000). However, because HVc_RA adaptation was included to counteract an overall build up of HVc activity, its decay time (225 ms) was considerably longer than the decay time of HVc_AFP adaptation (115 ms).

AFP \rightarrow RA connections and signals

The circuitry within the three song nuclei that make up the AFP could, in principle, subserve a variety of complex processing tasks. Our model treats the entire AFP as a “black box” performing the necessary calculations related to template comparison (see APPENDIX for details). Our algorithm was governed chiefly by computational

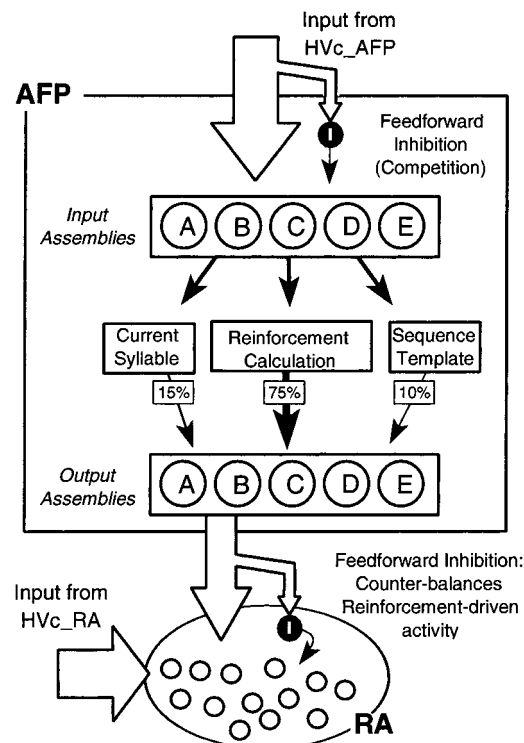


FIG. 7. Processing in the AFP (see METHODS for details). Input from HVc_AFP excites “feedforward” inhibition (filled circle, top) that implements a competition between AFP input assemblies (only those assemblies receiving significantly more than the average amount of input will be active). Three different calculations are performed on the results of this competition. 1) The match between this efference copy activity and the tutor song is determined (see Troyer and Doupe 2000). The results in a single “reinforcement” value that is strongly broadcast to all AFP output assemblies, accounting for 75% of their activity. 2) Patterned activity related to the current syllable is passed on unchanged, accounting for 15% of AFP output assembly activity. 3) Patterned activity related to the current syllable is delayed for the duration of one syllable and then delivered to AFP output assemblies in a pattern that is shifted forward one syllable in the tutor sequence. This shifting mechanism is how tutor sequence is stored in the AFP, and the shifted signal accounts for 10% of AFP output assembly activity. Feedforward inhibition in RA (filled circle, bottom) counteracts the strong reinforcement signal, leaving the patterned signal to affect RA activity via the pattern of AFP \rightarrow RA connections. After the initial period of sensory-motor matching (see RESULTS), signal (2) is redundant with the strong motor input from HVc_RA, leaving signal (3) to be the main contribution to altering activity in RA.

simplicity, but most calculations could be implemented relatively easily by a variety of biologically plausible circuits.

Processing within the AFP is shown in Fig. 7. Each AFP “input assembly” receives input from the HVc_AFP assemblies encoding sensory features related to the corresponding tutor syllable (the nature of the encoding scheme used in our model is described in Troyer and Doupe 2000; Fig. 6). Input is also received by a single inhibitory unit that broadcasts its output to all input assemblies. This “feedforward inhibition” implements a form of competition in which the only active AFP assemblies are those that receive significantly more input than average.

The main difficulty for our model is that the AFP is assumed to simultaneously broadcast three distinct signals that are important for separate aspects of sensorimotor learning. Each of these calculations is represented by a separate box in the middle of Fig. 7: 1) to guide syllable learning, the AFP transmits a nonspecific reinforcement signal that modulates plasticity in RA; 2) to organize a sensory \rightarrow motor mapping between the AFP and RA, the AFP forms a sensory representation related to the current syllable; 3) to guide sequence learning, the AFP must generate, with a one

syllable delay, a sequence teaching signal that biases RA activity toward the next syllable in the tutor sequence. A possible neural substrate for this delayed sequence teaching signal is the axon collaterals that transmit information from the lateral portion of the magnocellular nucleus of the anterior neostriatum (LMAN), the output nucleus of the AFP, to area X, the input nucleus of the AFP (see Fig. 1C, Troyer and Doupe 2000). The appropriate delay is roughly 75 ms, the length of a typical song syllable (≈ 115 ms) minus the processing delay contributed by the AFP (≈ 40 ms). Note that signals 1 and 2 are used to guide plasticity in RA but are not required to influence RA activity. In contrast, the purpose of signal 3 is to guide activity, but in principle, could disrupt learning in the AFP \rightarrow RA pathway.

In our implementation, the three signals are not segregated at the level of AFP outputs: the activity within the AFP output assemblies is just a summation of signals 1–3. The input to each RA assembly is then calculated as a sum of AFP outputs, weighted by the pattern of synaptic strengths from the AFP \rightarrow RA. This input serves both as a source of additive external input summed with RA input coming from HVC, and as a modulatory term in the RA plasticity rule (see APPENDIX). The modulation of RA plasticity in our model is completely phenomenological. Candidate mechanisms include release of trophic factors by AFP efferents (Johnson et al. 1997) or downstream effects of calcium entering through AFP glutamatergic synapses, which are dominated by NMDA receptors (Mooney and Konishi 1991).

How does the superposition of signals 1–3 in AFP output neurons exert separate effects in RA? The nonspecific reinforcement component of the AFP activity (signal 1) is separated from the two patterned components by its magnitude: we assume that the reinforcement signal contributes 75% of the input to AFP output assemblies. AFP output is then dominated by this reinforcement signal, and the resulting modulation of RA plasticity can be used to guide syllable learning. To allow the two patterned signals to play their role in song learning, we assume that the AFP also excites a population of inhibitory interneurons local to RA (Fig. 7, filled circle, *bottom*). This feedforward inhibition counteracts the nonspecific (reinforcement) component of the AFP input to RA, causing this nonspecific input to have little effect on spiking activity in RA. However, inhibition would not be expected to cancel trophic effects of AFP inputs and hence would not block reinforcement mediated by neurotrophins. In an alternative scenario, inhibition that is proximal to the cell body might eliminate spiking but not prevent the depolarization within distal dendrites by inputs from HVC_RA or other RA neurons. Thus, calcium entry through NMDA receptors at AFP synapses could still be used to modulate plasticity within the dendritic tree, even though the currents flowing through these receptors are counteracted by inhibition arriving at the soma.

In addition to explaining how the nonspecific reinforcement component of the AFP activity is prevented from disrupting patterns of RA activity, we must explain how to prevent it from disrupting the learning in the AFP \rightarrow RA pathway. By definition, a large reinforcement signal that is expressed as high activity in all AFP output assemblies will also lead to increased plasticity within all RA assemblies. This correlation between nonspecific presynaptic firing in the AFP and nonspecific modulation of plasticity in RA tends to strengthen all synapses from the AFP \rightarrow RA. To counteract this tendency, AFP \rightarrow RA synapses were assigned a higher plasticity threshold (see APPENDIX).

The action of the AFP activity related to the current efference copy (signal 2) is straightforward: after the efference copy mapping from HVC_RA to HVC_AFP gives an accurate prediction of the motor input from HVC_RA to RA (Troyer and Doupe 2000), the AFP assembly corresponding to the current syllable will be most active when RA assemblies corresponding to that syllable are also active. Sensory \rightarrow motor associational learning follows, causing AFP assemblies encoding a particular tutor syllable to project most strongly to RA assem-

blies encoding the same syllable. (Fig. 5). After the sensory \rightarrow motor matching is accomplished, the input from the AFP activity related to signal 2 will be redundant with the (stronger) input to RA from HVC.

Our functional requirements for the sequence teaching signal (signal 3) are that it biases RA activity toward the next syllable in the tutor sequence, but does not disrupt the learning in the AFP \rightarrow RA pathway driven by signal 2. To implement the proper bias, the processing box marked “Sequence Template” in Fig. 7 accepts a pattern of input, waits for one syllable, and then excites AFP output assemblies in a pattern that is shifted one syllable forward in the tutor sequence. Since the AFP \rightarrow RA connections perform a sensory \rightarrow motor mapping, this signal will bias RA toward the next motor command in the tutor sequence (Fig. 4). The reason that this signal does not disrupt the associations necessary to develop a sensory \rightarrow motor mapping to RA is that, before sequence learning is accomplished, the inputs from HVC_RA to RA are strong and their sequence is random. Therefore, AFP activity for the subsequent syllable (signal 3) will not be strongly correlated with RA activity and hence will not contribute significantly to plasticity in the AFP \rightarrow RA connections. After the model begins to produce the proper sequence, the motor patterns in RA driven by HVC_RA will be matched to the sequence teaching signal syllable (signal 3). Hence, the associational plasticity related to signal 3 will simply reinforce the sensory \rightarrow motor mapping originally organized by signal 2.

Our implementation represents only one of many plausible ways in which different signals could exert different effects in RA. A conceptually simple solution to the problem of segregation would be to have different functional signals carried by distinct classes of AFP projection neurons. However, developing such a separation could be difficult. Another alternative is for different signals to be encoded in different temporal patterns of AFP activity (e.g., bursting versus tonic). These could preferentially excite separate receptors in RA and/or trigger different plasticity mechanisms in RA. Finally, since the three signals make crucial contributions to learning at different times during song learning (see Fig. 11 in RESULTS), their functions could be subserved by mechanisms tied to developmental critical periods. Our model makes predictions regarding the functional information carried by the AFP \rightarrow RA pathway. Further experiments will be required to determine the possible neural substrate for these signals.

Quantifying learning time course

To obtain quantitative results regarding the time course of learning in the model, we measured how closely the statistics of RA motor output matched the statistics of the tutor song, as well as measuring how closely important patterns of connectivity matched the properties of an “ideal” model that would accurately reproduce the tutor song. The measure used to compute these matches was the correlation coefficient (CC) applied to the elements of the relevant connection matrices (see METHODS in Troyer and Doupe 2000). Syllable-related activity was quantified as in Troyer and Doupe 2000. Sequence-related activity was quantified by dividing the model output into 250 syllable epochs and constructing M^{next} , the matrix of co-fluctuations between patterns of RA activity for a given syllable and the patterns of RA activity for the next syllable

$$M_{ij}^{\text{next}} = \frac{1}{250} \sum_{n=250(m-1)+1}^{250m} [r_i(n-1) - \bar{r}(n-1)][r_j(n) - \bar{r}(n)]$$

where $r_i(n)$ is the activity level in the i th RA assembly, and $\bar{r}(n)$ is the average activity across assemblies during syllable n . We used the CC to compare M^{next} to an ideal syllable transition matrix, M_{ij}^{seq} : $M_{ij}^{\text{seq}} = 4$, if assembly j forms part of the representation for the syllable following the syllable coded by assembly i ; $M_{ij}^{\text{seq}} = -1$, otherwise. Diagonal entries were included.

In addition to monitoring patterns of RA activity, we monitored development in four sets of connections. 1) The accuracy of the efference copy map was quantified by calculating the correlation coefficient between the pattern of $HVc_RA \rightarrow$ motor connections ($HVc_RA \rightarrow$ RA) and $HVc_RA \rightarrow$ sensory connections ($HVc_RA \rightarrow HVc_AFP$). 2) To quantify the development of the sensory \rightarrow motor mapping (Fig. 5), we computed the CC between the pattern of $AFP \rightarrow$ RA connection strengths and the ideal pattern of connectivity, in which the AFP assembly representing a given tutor syllable would have connections only onto RA assemblies encoding the motor features belonging to that syllable. 3) To quantify the progress of syllable learning, we computed the CC between the ideal syllable correlation matrix, M^{sy} , and the pattern of intrinsic RA connections as in Troyer and Doupe 2000. $M_{ij}^{seq} = 4$, if assembly j forms part of the representation for same syllable as assembly i ; $M_{ij}^{seq} = -1$, otherwise. Diagonal entries were excluded. 4) To evaluate sequence-related connectivity, we multiplied the $HVc_AFP \rightarrow HVc_RA$ and $HVc_RA \rightarrow$ RA connection matrices. The resulting matrix represents the influence of each HVc_AFP assembly on each RA assembly via the context signal in HVc (Fig. 3). The correlation coefficient between this matrix and M^{seq} was used to measure the development of sequence-related connectivity.

RESULTS

Our model explores how song learning can result from associational plasticity, guided by template comparison signals transmitted by the AFP. The representation of the sensory and motor aspects of song in our model is described in detail in our companion paper (Fig. 6 in Troyer and Doupe 2000). Briefly, the information encoded within each neural population (HVc_RA , HVc_AFP , RA, and the AFP) is represented by the activation value of a number of processing units, each meant to capture the average level of activity within a connected set of neurons or “cell assembly” (Hebb 1949). For most simulations, the tutor song contains five syllables, with each syllable composed of eight abstract vocal features. The features encoding different syllables are assumed to be unique, so we number the features according to tutor syllable (syllable A, features 1–8; syllable B, features 9–16; etc.). Each of 40 RA assemblies encodes the motor aspect of one vocal feature, and each of 40 HVc_AFP assemblies encodes the sensory aspect of one feature. The template for syllables is stored in the connections from $HVc_AFP \rightarrow$ AFP, and the template for tutor sequence is stored by circuitry internal to the AFP (see METHODS).

Sensorimotor learning is accomplished in three stages. The first two stages were explored in our companion paper (Troyer and Doupe 2000). At the beginning of the simulation, all connections in the motor pathway are unstructured, and the premotor drive initiating each syllable drives unorganized patterns of RA activity (Fig. 8A). During the initial, efference copy learning stage, associations between the HVc_RA motor activity and the resulting auditory feedback input to HVc_AFP cause a motor \rightarrow sensory efference copy mapping to develop between these two populations (stage 1; Figs. 4A, 8 in Troyer and Doupe 2000). In the second, syllable learning stage, the AFP evaluates the efference copy signals and broadcasts template matching “reinforcement” signals that reorganize synaptic strengths in RA so that assemblies corresponding to individual tutor syllables are co-active (stage 2; Fig. 8B; Figs. 4A, 10 in Troyer and Doupe 2000). In this paper, we focus on the final, sequence learning stage, in which “sequence teaching” signals from the AFP act in concert with the sequence gener-

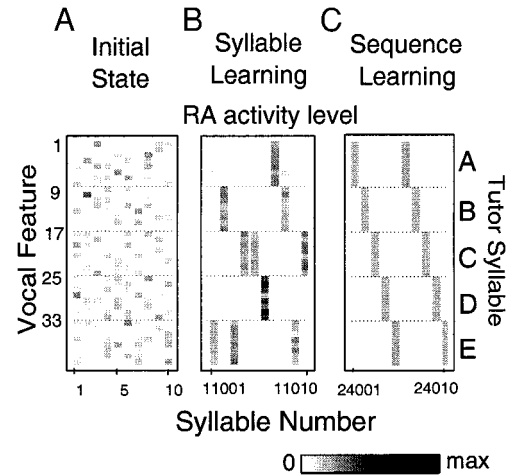


FIG. 8. Overview of model behavior. RA assemblies (40 total) are grouped along the vertical axis according to the tutor syllable to which they correspond (labeled A–E). Bar shows color scale for this and subsequent figures. A: RA activity during the first 10 simulated syllables (numbered from start of simulation). RA activity is unorganized and random. B: syllable learning. RA activity during each syllable is well-matched to one of the tutor syllables, but syllables are produced in a nearly random order. C: sequence learning. By syllable 25,000, activity is matched to the tutor representation, with syllables produced in the proper sequence.

ation mechanism in HVc so that syllable representations are produced in the correct order, $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A \dots$ (stage 3; Fig. 8C). It is important to note that a segregation between developmental stages is not embedded within our learning rule or network architecture. Rather, all synapses in HVc and RA are plastic, and this plasticity lasts throughout the simulation. Thus, development is driven by interdependent patterns of association that emerge during song learning.

Sequence learning

The key to sequence learning in the model is the ability of signals from the AFP to bias RA activity toward the proper syllable transitions (Fig. 9A, arrows). Acting over multiple syllables, this in turn biases the association between HVc_RA and RA activity. The resulting change in connections from $HVc_RA \rightarrow$ RA connectivity leads to the production of appropriate syllable transitions (Fig. 4). Auditory feedback ensures that an accurate efference copy mapping is maintained (Fig. 6). The gradual improvement of syllable transitions is shown in Fig. 9B.

Time course of learning

To examine the time course of learning, we considered the properties of an “ideal” solution, in which patterns of connectivity were set so that this ideal model would accurately reproduce the tutor song (see METHODS for detailed definitions). We then quantified how closely important sets of connections matched the ideal model. The match was calculated using the correlation coefficient, a method that gives a value of one for identical connection patterns and values near zero for connection patterns that are uncorrelated. We measured four sets of connections, the efference copy map from $HVc_RA \rightarrow HVc_AFP$, the sensory \rightarrow motor map from the $AFP \rightarrow$ RA, syllable storage in the $RA \rightarrow$ RA connections, and the sensory \rightarrow next motor pathway from $HVc_AFP \rightarrow HVc_RA \rightarrow$

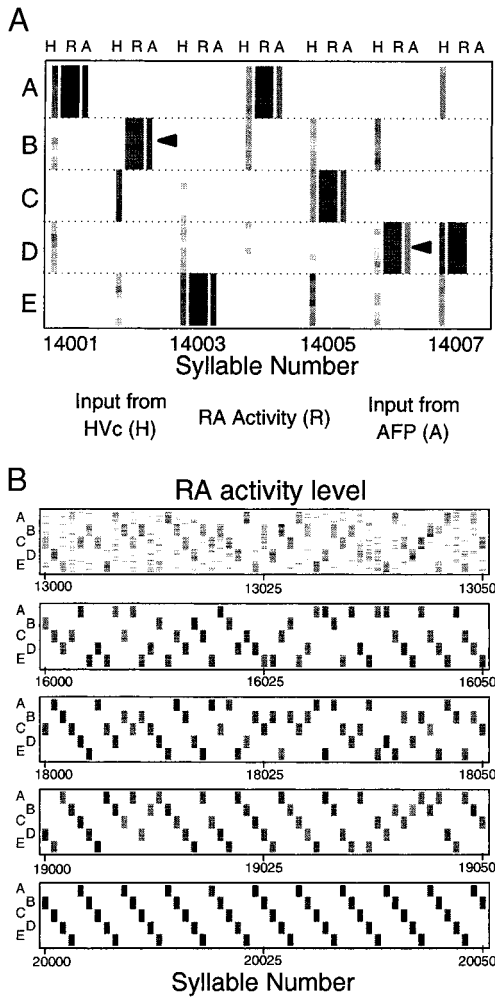


FIG. 9. Sequence learning. *A*: AFP-guided syllable transitions. Hvc input to RA (H), AFP input to RA (A), and RA activity (R) for syllables 14,001–14,007. For syllables 14,002 and 14,006, the input from the AFP ensures proper syllable transitions, overriding “incorrect” input from Hvc (arrows). To emphasize differences in the input to various RA assemblies, the density of shading for H and A represents the amount of input that exceeds the mean for that pathway; inputs weaker than the mean are not shown. *B*: convergence toward proper sequence. Model output for 51 consecutive syllables is shown at 5 different developmental time points. Syllable transitions are initially random but eventually begin to be produced in small strings matching the tutor song. Eventually the entire sequence is learned.

RA. We also measured how closely the motor output from RA matched the tutor song. These calculations were performed for “epochs” consisting of 250 consecutive syllables produced by the model. To quantify the development of tutor syllables, we calculated the matrix of co-fluctuations, whose ij th entry indicates whether assembly i and assembly j have similar patterns of activity. To quantify the development of tutor sequence, we calculated a similar matrix, except that the ij th entry indicates whether activity in RA assembly i during syllable n co-fluctuated with the activity in assembly j during syllable $n + 1$. These matrices were matched to the corresponding matrices computed from the tutor song, again using the correlation coefficient (see METHODS).

The developmental time courses of the multiple, interacting associations underlying model development are summarized in Fig. 10*A*. Figure 10*B* shows which connections are most important during each of the song learning stages traced in Fig.

10*A*. Initially, the only consistent pattern of association in the network is between motor activity and delayed auditory feedback, and the corresponding efference copy mapping develops rapidly (stage 1, dotted line). As accurate efference copies are passed onto the AFP, a sensory \rightarrow motor mapping also develops between the AFP and RA (stage 1*a*, dashed-dotted line; see Fig. 5). An accurate efference copy also causes the AFP to produce consistent reinforcement signals, which reorganize intrinsic RA connections so that RA assemblies corresponding to the same tutor syllable begin to receive common patterns of synaptic input (stage 2, thin solid line). As this happens, the model begins to produce RA activity patterns matched to the tutor syllables (thin dashed line). As syllables are learned, efference copy activity in Hvc_AFP becomes increasingly

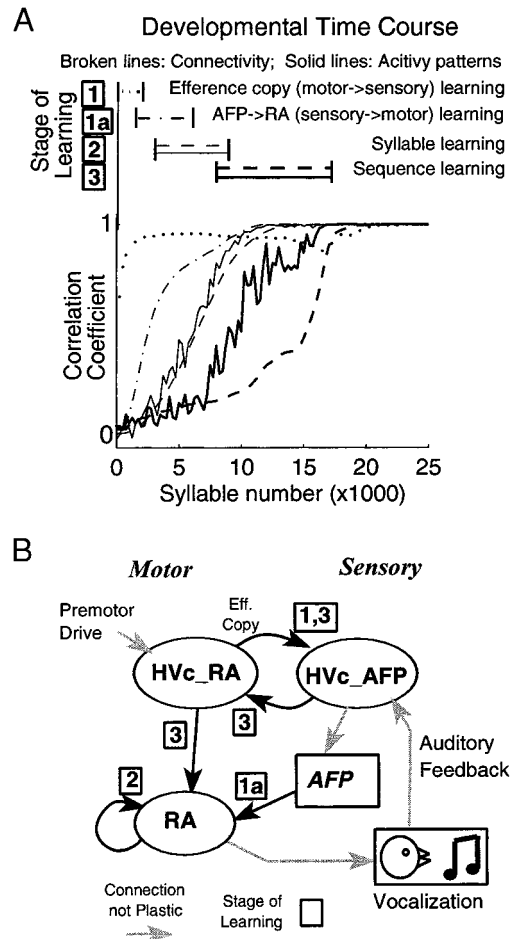


FIG. 10. Summary of developmental time course. *A*: 3 basic stages of development. The initial stage of efference copy learning is nearly complete by syllable 1000 (stage 1, dotted line). As accurate efference copy signals are passed on to the AFP, a sensory \rightarrow motor mapping is learned in the connections from the AFP \rightarrow RA (stage 1*a*, dashed-dotted line; see Fig. 5). Accurate efference copy signals also allow the onset of syllable learning (stage 2). The development of motor activity matched to the tutor song (thin solid line) mirrors the development of appropriate connectivity intrinsic to RA (thin dashed line). Because sequence learning (stage 3) is driven by correct transitions guided by the AFP (Fig. 4), correct sequence activity (thick solid line) occurs before the development of the appropriate composite mapping (Hvc_AFP \rightarrow Hvc_RA \rightarrow RA) in the motor pathway (thick dashed line). Note that reorganization in the Hvc_RA \rightarrow RA pathway that underlies sequence learning disrupts the efference copy match during syllables 8,000–17,000. The correlation coefficients computed are defined in the RESULTS. *B*: involvement of connections in the different stages of learning shown in *A*.

confined to patterns matched to the relatively small number of tutor syllables. These aspects of the model (with the exception of stage 1a) were described in detail in our companion paper (Troyer and Doupe 2000). As syllable learning proceeds, clearly defined sequence teaching signals begin to be produced by the AFP. These begin to bias RA activity toward the tutor sequence (stage 3, thick solid line; see Fig. 9A). This altered activity then remaps the connections from HVC_RA to RA, so that the polysynaptic pathway from HVC_AFP \rightarrow HVC_RA \rightarrow RA (thick dashed line) yields correct sensory \rightarrow next motor syllable transitions. Note that improvement in the sequencing of RA activity happens *before* the learning of the appropriate connectivity from HVC_AFP \rightarrow HVC_RA \rightarrow RA, since AFP-driven sequence transitions are necessary to drive sequence related learning. The reorganization of the HVC_RA \rightarrow RA pathway disrupts the efference copy mapping, which begins to degrade slightly during the period of sequence learning (dotted line, syllables 8,000–17,000). This tension between AFP-guided changes in the motor pathway and renewed efference copy learning continues until both are in rough agreement. This agreement causes a transient decline in the efference copy match (near syllable 16,000), since the HVC \rightarrow RA connection races ahead to the final solution. The efference copy makes a final recovery, and the model produces a stereotyped sequence of song syllables.

Range of model behavior

By presenting results from a single representative simulation, we have demonstrated the plausibility of our core hypothesis that associational learning, distributed widely throughout the song system, is sufficient for sensorimotor matching to a previously memorized template stored in the AFP. Because each stage of the learning is dependent on previously developed associations, a complete assessment of the reaction of our model to changes in model parameters is beyond the scope of this paper (see Troyer and Doupe 2000 for some important manipulations).

Overall, sequence learning was significantly less robust than syllable learning, since it results from continual interplay between the changes in the HVC to RA projection and the efference copy mapping in HVC. The robustness of model behavior at the default set of parameters was assessed by running 10 simulations, each with different random seeds determining the initial pattern of synaptic connectivity and the sequence of premotor drives. All simulations eventually learned the tutor song perfectly. Nine of these simulations followed a similar time course, completing sequence learning near syllable 17,000 (Fig. 11A). However, in one of the simulations, correct learning took significantly longer and was not complete until syllable 25,000 (Fig. 11B). Examination of the output of this simulation reveals that during the period between syllable 15,000 and 20,000 when the other simulations were stringing together series of transitions to match the tutor song, this simulation began to repeat the subsequence A–D, omitting syllable E (Fig. 11C). Since the strong homeostatic mechanisms in the model prevent any RA assemblies from becoming permanently inactive, the model compromised, occasionally inserting a strong version of syllable E in place of syllable D. However, by syllable 23,000, the model began to insert syllable E in its proper place in the sequence, but sometimes

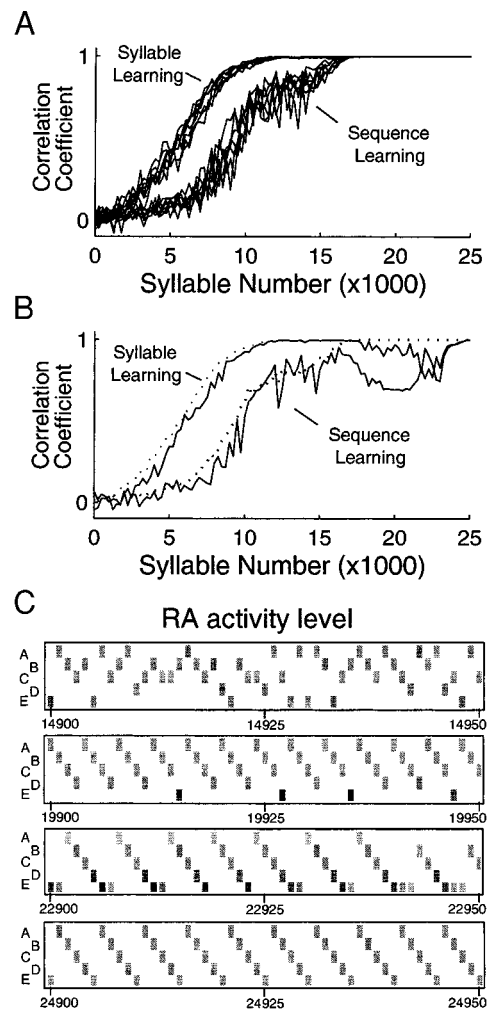


FIG. 11. Variability of learning time course. *A*: Development of syllable-related and sequence-related activity for 9 of 10 repeated simulations. Parameters were fixed at their default values and simulations were run using different random seeds to determine the initial connectivity and the sequence of premotor drives. Output was quantified as in Fig. 10A. *B*: convergence in 1 of the 10 simulations was not complete until syllable 25,000 (solid lines). The average time course of the 9 simulations shown in *A* is plotted for comparison (dotted lines). *C*: model output for simulation plotted in *B*. The model first converged on a suboptimal solution by repeating syllables A–D and occasionally substituting syllable E for D. Due to homeostatic mechanisms that act to keep average activity in all assemblies constant, syllable E had large activity (black rectangles). By syllable 23,000, E was inserted in the proper position but was often repeated 2–3 times. Syllable A was sometimes dropped. Repetitions eventually ceased and by syllable 25,000 the model produced the proper sequence.

syllable E was repeated and sometimes syllable A was dropped. By syllable 25,000, the model had converged on the correct sequence. Personal observation of many simulations revealed that such temporary “compromise” solutions to the competing requirements of associational change in the HVC_RA \rightarrow RA projection and the maintenance of an accurate efference copy mapping within HVC were not uncommon.

To further assess the range of model behavior, we increased the number of syllables to eight, thereby increasing the range of possible sequence transitions. The number of vocal features in each syllable was reduced to five, so that the simulations contained the same number of RA assemblies as before ($8 \times 5 = 40$). AFP circuitry was adjusted for the different template,

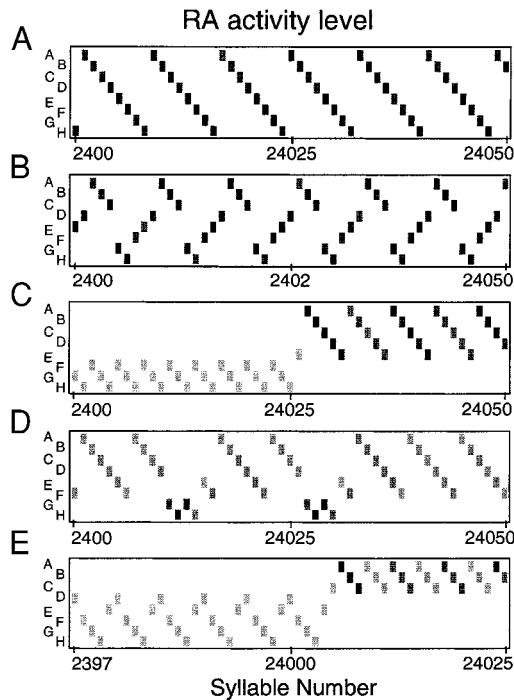


FIG. 12. Imperfect sequence learning. A–E: outcomes of 10 simulations with increased numbers of syllables. Perfect learning (A) occurred in 6 simulations. In one simulation (B), a full sequence of 8 syllables was produced, but the sequence was broken into three “chunks” of syllables. In 2 of these chunks (syllables A–C and G–H), syllables were sung in the proper order. In the 3 other simulations (C–E), the sequence was broken into two subsequences, with subsequences sung in alternation. Transition times between subsequences are determined by the interaction of learning with slow homeostatic mechanisms.

and AFP \rightarrow RA learning was slightly adjusted to ensure that an accurate sensory \rightarrow motor mapping was learned (see APPENDIX). To push the model to make mistakes, all learning rate parameters were increased by a factor of 5. No other parameters were readjusted. The range of RA output for a set of 10 simulations is shown in Fig. 12. Perfect learning occurred in six of the ten simulations. An example is shown in Fig. 12A. In one simulation, the model produced a stereotyped sequence of eight syllables, but this motif consisted of two “chunks” of appropriately copied song, separated by a string of three syllables sung in reverse order (Fig. 12B). In the three other simulations, the full sequence was broken into two repeated subsequences (Fig. 12, C–E). These were sung in alternation, with the rate of alternation controlled by the interaction between associational learning and homeostatic mechanisms that prevent the elimination of either subsequence. In versions of the model with weaker homeostatic mechanisms, syllables outside of the most commonly sung subsequence were simply dropped (not shown).

DISCUSSION

Principal findings and predictions

By constructing a computational model, we have demonstrated that simple rules of associational plasticity, operating throughout the song system, are sufficient to support sensorimotor learning at multiple levels of the temporal hierarchy for song. Learning proceeds in a series of stages, with efference copy learning followed by syllable learning and then sequence

learning. These developmental stages are not predetermined by our learning rule, but follow a cascade of interrelated associations that are guided by template matching signals from the AFP.

In this paper, we focused on the problem of learning song sequence. We propose that sequence generation results from a reciprocal sensory-motor interaction between the two populations of HVC projection neurons: the motor component is encoded primarily in RA-projecting HVC neurons, whereas the sensory component is encoded primarily in AFP-projecting neurons (Katz and Gurney 1981; Kimpo and Doupe 1997; Lewicki 1996; Saito and Maekawa 1993). This mechanism predicts that the participation of neurons in both populations is required for normal sequence generation. We also predict that the slow “context” signals linking one syllable to the next flow primarily from AFP-projecting to RA-projecting neurons. While we have not explored possible neural substrates for this functionally slow connection, Kubota and Taniguchi (1998) have reported that RA-projecting neurons possess an ionic current that delays the initiation of action potentials.

The absence of a direct projection from the AFP to nuclei upstream of RA, the likely site of sequence generation (Vu et al. 1994), poses a significant challenge to the hypothesis that the AFP guides learning of song sequence. One strategy for overcoming this challenge is for the AFP to guide learning within the connections from HVC to RA, so that the outputs from the pattern generator are mapped onto the appropriate sequence of syllable representations in RA (Doya and Sejnowski 1998). Viewed in isolation, this hypothesis predicts the existence of an autonomous pattern generator that is unaffected by outputs from the AFP. In our model, however, a motor \rightarrow sensory efference copy mapping within HVC plays a crucial role in sequence generation. Therefore, we predict that the AFP does affect the pattern generator, although indirectly: AFP-induced changes in RA change the relation between HVC premotor activity and the resulting auditory feedback, triggering renewed learning in HVC and altering the sequence of its premotor outputs (Fig. 6).

Our model predicts that neural activity recorded within the AFP should contain a mixture of three signals. First, to guide syllable learning, the output from the AFP should carry a reinforcement signal that modulates plasticity widely within RA. This reinforcement signal should have a component operating on the time scale of individual syllables. Second, the AFP should carry efference copy information related to the current syllable. This is necessary for associational learning of the appropriate sensory \rightarrow motor mapping from the AFP to RA and should be particularly prominent in the early stages of sensorimotor learning. Finally, to guide sequence learning, the AFP should be able to bias RA activity toward syllable transitions contained within the tutor song. Given our proposed developmental time course of learning (Fig. 10A), we predict that the ability of the AFP to bias RA motor activity should be maximal during the peak period of sequence learning. Early in learning, the AFP to RA connections are expected to be relatively unorganized, and, after sequence learning, the highly organized connections from HVC to RA are expected to dominate the input to RA. This prediction could be tested using cross-

correlation analyses and/or electrically stimulating the output nucleus of the AFP during singing.

Weaknesses of the model

Like our syllable learning model (Troyer and Doupe 2000), the main weakness of the sequence learning model is the simplified representation of the problem. In particular, we have treated the motor hierarchy as having two distinct levels: syllables and sequences of syllables. Questions regarding the mechanisms for starting and stopping song have not been considered, nor have we addressed the possibility that sub-syllabic “notes” might be the true units of song (Cynx 1990). Quantitative data regarding these issues are scant, and more extensive analysis of developing song will be required to constrain more realistic models of learning at multiple levels of the song hierarchy.

Sequences by associative chaining

Our model assumes that sequences are generated as an “associative chain” of sensory and motor representations (motor \rightarrow sensory \rightarrow next motor \rightarrow next sensory . . . ; James 1983; Adams 1984). One important difference in our model is that the sensory components of the chain are internally generated, efference copy representations. Use of an efference copy addresses two of the three main challenges to associative chaining (Rosenbaum 1991). First, efference copy addresses the limitations placed on chaining by feedback delay. Second, our version of chaining depends only on signals generated within the brain and is therefore consistent with retention of motor skills even when sensory feedback has been removed (reviewed in Sanes et al. 1985; Jeannerod 1988). A third challenge for associative chaining models is their inability to account for the errors commonly produced during some sequential behaviors such as speech (Lashley 1951; MacKay 1970; reviewed in Houghton and Hartley 1996). Although a thorough analysis of the variability in zebra finch song sequence has yet to be undertaken, the limited data available suggest that song is sometimes learned in short sequences or “chunks” of song syllables (Williams and Staples 1992). Associative chaining can naturally account for such learning by viewing chunk boundaries as errors in learning appropriate syllable transitions (Fig. 12).

Recent technical advances raise the possibility of testing our chaining hypothesis by selectively photo-ablating neurons within a single population of HVC projection neurons (Scharff et al. 2000). Early results suggest that song is insensitive to disruptions of HVC_AFP, while lesioning HVC_RA neurons can disrupt song. However, the effects of HVC_RA lesions were variable, with <50% of birds showing deterioration of song. More complete lesions and/or more detailed analysis may yield greater insight into the relative contribution of HVC_RA and HVC_AFP neurons to song production.

ASSOCIATIVE CHAINS AND SENSORY SELECTIVITY. The same reciprocal circuit underlying song production may underlie the selectivity of HVC neurons to auditory stimuli (Lewicki and Konishi 1995; Margoliash 1983; Margoliash and Fortune 1992) and may contribute to song perception (Nottebohm et al. 1990; Scharff et al. 1998). In particular, vigorous sensory responses may require that the sequence of incoming auditory

signals be matched to the sequence of sensory expectations that would be elicited by recruiting the motor circuit. In our circuit, auditory stimulation using syllable A of the bird’s own song should excite the sensory representation of A in HVC_AFP. This in turn would excite, with a delay, the HVC_RA context signal Cxt_A , and this should produce efference copy input for syllable B. A match between this internally generated expectation and the auditory signal may lead to an enhanced response. Because the efference copy mapping is learned from associations generated when the bird vocalizes, this mechanism may explain why auditory responses in HVC become tuned to the bird’s own song during the course of sensorimotor learning (Volman 1993). Our model also predicts that neurons within both populations of HVC projection neurons should show sensory-related as well as motor-related activity. Furthermore, since the presentation of multiple syllables may be necessary to fully recruit the motor circuit, this mechanism may underlie the selectivity of some HVC neurons to aspects of the auditory stimuli occurring several hundred milliseconds before the recorded neural response (Lewicki and Arthur 1996; Lewicki and Konishi 1995; Margoliash 1983; Margoliash and Fortune 1992).

Temporal hierarchies and song learning

Our model demonstrates how associational learning, distributed widely throughout the song circuit, can be used to address general problems in sensorimotor learning. Moreover, the model points to specific problems raised by song system anatomy for learning multiple levels of the temporal hierarchy for song (Fig. 1A). The functional roles we propose for the AFP during song learning share similarities with hypotheses regarding the importance of basal ganglia/forebrain loops for reinforcement and sequence learning in mammals (Aldridge and Berridge 1998; Contreras-Vidal and Schultz 1999; Hikosaka et al. 1999; Houk et al. 1995; Matsumoto et al. 1999; Montague et al. 1996).

FINE TEMPORAL STRUCTURE (1–10 MS). Birds are able to produce vocal output that changes on the scale of milliseconds (Fee et al. 1998; Suthers et al. 1994), and it is known that such fine changes affect neural responses in the song system (Theunissen and Doupe 1998) and influence avian behavior (Lohr and Dooling 1998). The possibility that birds learn such fine motor control poses a significant challenge to any model of motor learning. In addition to the fact that sensory \rightarrow motor “inverse” mappings often are not well-defined (Jordan 1995), learning such mappings may be extremely difficult at the finest time scales. First, feedback delay is an order of magnitude longer than the temporal precision of the sensory \rightarrow motor matching. Second, there is likely to be a complex relationship between motor neuron activity and behavioral output due to the physics of the muscles and tissues that produce the behavior (Fee et al. 1998; Goller and Larsen 1997).

Our model relies on reinforcement learning to guide RA connectivity toward patterns encoding individual song syllables. Even though we do not explicitly model the temporal precision of RA motor activity (Yu and Margoliash 1996), our approach satisfies general constraints imposed by the problem of precise motor learning, as well as particular constraints imposed by song system physiology. Most importantly, the use of reinforcement learning avoids the difficult problem of learn-

ing a sensory \rightarrow motor matching at fine time scales, since the evaluation of sensory input yields a single number that is broadcast equally to all RA assemblies. Fine structure within the motor representation for each syllable is assumed to be learned by using the reinforcement signal to guide an initially random exploration of motor space to the appropriate goal. The development of a fine time scale sensory \rightarrow motor mapping between the AFP and RA is particularly unlikely given that the AFP input to RA is almost exclusively mediated by NMDA receptors that have decay times on the order of 40–200 ms (Mooney 1992; Stark and Perkel 1999; White et al. 1999).

Our model also predicts that circuits intrinsic to RA play an important role in encoding the motor programs for individual song syllables (cf. Spiro et al. 1999). Thus, syllable representations can remain stable even during the sequence-related remapping of HVC efferents. Moreover, our model does not require that the precise patterns of RA motor activity be driven by input from HVC, where neural activity has been shown to be temporally less precise (Yu and Margoliash 1996).

INDIVIDUAL VOCAL GESTURES (≈ 100 MS). Our model uses a Hebbian plasticity rule roughly matched to the time scale of NMDA receptor-mediated currents (40–200 ms). The duration of these currents is of similar duration to both the length of sensory feedback delay and evaluation ($\approx 65 + 40$ ms) and the duration of the individual elements of song (≈ 115 ms). Human speech is disrupted by delayed playback using delays within a similar range (Lee 1950). The similarity between the time scales of internal processing and sensory feedback is important for the workings of our model. A relatively broad window for associational plasticity in HVC is sufficient to span the sensory feedback delay, and the temporal asymmetry of Hebbian plasticity naturally leads to an efference copy mapping between motor and sensory representations within HVC. The use of temporally imprecise, syllable-based neural representations allows for reliable associations even if the window for associational plasticity is relatively broad, eliminating the need for associational learning tightly tuned to the relevant delays in the system. We suggest that feedback delay may set a preferred time scale for sensorimotor learning and may relate to the prevalence of ≈ 4 – 10 Hz rhythms in many motor behaviors, including active touch (Morley et al. 1983), motor tremor (McCaughey et al. 1997), and whisker twitching in rats (Nicolelis et al. 1995).

SEQUENCE GENERATION (>100 MS). Learning temporal structure on time scales greater than individual syllables poses significantly fewer problems than learning structure at fine temporal scales. Sensory \rightarrow motor matching is readily accomplished at the level of syllable-based features, and, as a result, template information encoded in sensory coordinates in the AFP is available to actively influence syllable transitions (Figs. 4 and 10). AFP lesion data are consistent with the active role in sequence generation predicted by our model. Lesions of LMAN, the output nucleus of the AFP, reduce the range of sequence transitions in juvenile birds (Scharff and Nottebohm 1991), as would be expected if AFP outputs were important for generating sequence transitions during sensorimotor learning. In contrast, lesioning the input nucleus of the AFP, area X, increases sequence variability (Scharff and Nottebohm 1991; Sohrabji et al. 1990). Increased variability could result if area X damage led to inconsistent output from LMAN. The most

direct evidence for an active role of the AFP in sequence generation comes from Bengalese finches, an estrildid finch closely related to zebra finches: lesions in the AFP of adult Bengalese finches appear to have an immediate effect on song sequence (Okanoya and Kobayashi 1998).

Learning at longer time scales is also less constrained by the problem of feedback delay. Our estimates suggest that auditory feedback should reach the song system before the onset of the subsequent syllable, raising the possibility that it may play a role in sequence generation. In zebra finches, these reafferent signals appear not to contribute acutely to vocal production since altering the auditory feedback pathway does not have immediate effects on the temporal structure of song (Leonardo and Konishi 1999; Nordeen and Nordeen 1988; Price 1979). In contrast, Bengalese finches that are deafened as adults show a rapid disruption of song sequence (M. Brainard, personal communication; Okanoya and Yamaguchi 1997; Woolley and Rubel 1997), suggesting that auditory reafference may play an important role in this species. However, auditory feedback in Bengalese finches does not appear to contribute to singing at finer time scales, since the degradation of individual syllables parallels the slower postdeafening song degradation seen in zebra finches. In our model based on zebra finches, auditory signals are canceled within HVC (Troyer and Doupe 2000). However, the model could be generalized so that auditory feedback contributes significantly to the context signals that bridge the sensory \rightarrow next motor link in the chain underlying song sequence (Fig. 3).

Motor hierarchies and selective attrition

Juvenile birds in many species produce a large number of syllables that are winnowed down to the final adult repertoire (e.g., Marler and Peters 1982; Nelson and Marler 1994). While we have not yet explored these issues directly, errors made by our model (Fig. 12) suggest how the number of song elements may be influenced by circuitry at both the syllable and sequence levels of the motor hierarchy: RA circuitry may influence the total number of syllable representations encoded, whereas the pattern generator in HVC determines which syllables get incorporated into the final song. This two-level picture may explain the re-emergence of white crown sparrow syllables that were learned during development but dropped from the original adult repertoire (Benton et al. 1998). Quantitative data concerning the developmental time courses of syllable morphology and syllable sequence will be crucial for understanding the mechanisms for learning on multiple time scales.

APPENDIX

Our algorithm for sequence learning extends our previous syllable learning algorithm (SLA), described in detail in the appendix to our companion paper (Troyer and Doupe 2000). We present here only differences and additions to SLA. The main differences were: 1) adding plastic connections from HVC_AFP \rightarrow HVC_RA and from the AFP \rightarrow RA (Fig. 2B); 2) having more complex calculations in the AFP (Fig. 7); 3) adding adaptation to HVC_RA. The additional connections were initialized using the “uniform strategy,” i.e., all synapses were initially set to have equal strength and then perturbed by zero mean Gaussian noise with standard deviation equal to 10% of the unperturbed strength (see Methods in Troyer and Doupe 2000).

As before, we abbreviate HVc_RA as HR, HVc_AFP as HA, and AFP as AF, and let $r^{\text{HR}}, r^{\text{RA}}, r^{\text{HA-E}}, r^{\text{HA-M}}, r^{\text{HA-L}}, r^{\text{HA-G}}$, where E, M, L, G refer to the early, middle, late, and gap portions of HVc_AFP activity. r^{EC} is defined as in SLA and denotes the HVc_AFP activity representing the efference copy passed to the AFP. r^{ctx} denotes the HVc_AFP activity contributing to the context signal and was determined as the average activity during the middle, late, and gap portions of the syllable (see step 8d below). For sequence learning, the AFP has both input and output assemblies (Fig. 7), with rates r^{AFin} and r^{AFout} . We use $[post, pre]$ to denote a matrix of synaptic strengths between a presynaptic and postsynaptic population of assemblies.

Simulations

Running simulations for 25,000 syllables was found to be adequate to guarantee the convergence of sequence learning (Fig. 11). The steps in the sequence learning algorithm are slightly reordered relative to SLA. Since input from the AFP alters RA activity, calculation of AFP activity had to *precede* the calculation of activity in RA. This in turn required calculation of the HVc_AFP activity contributing to the efference copy signal ($r^{\text{HA-E}}$ and $r^{\text{HA-M}}$). Calculation of $r^{\text{HA-L}}$ and $r^{\text{HA-G}}$ had to *follow* the calculation of RA activity, since these depended on the auditory feedback from the current syllable. We let $SLA(n)$ refer to step n in our syllable learning algorithm.

1. Premotor drive. Same as SLA(1), except that p_{drive} was reduced to 16 to compensate for the addition of context input from HVc_AFP.
2. Calculate HVc_RA activity. The afferent input to HVc_RA is calculated as the sum of premotor drive and HVc_AFP context signals: $\text{aff}_i^{\text{HR}}(n) = pi(n) + \sum_j [HR, HA]_{ij} r_j^{\text{ctx}}(n-1)$. Output firing rates are determined as in SLA(2).
3. Calculate earlier portions of HVc_AFP activity ($r^{\text{HA-E}}$ and $r^{\text{HA-M}}$). Same as in SLA(4).
4. Calculate AFP activity and reinforcement. The calculation of AFP activity was more complex than in SLA (see METHODS; Fig. 7). The calculation of activity in AFP input assemblies follows SLA(5): $r_k^{\text{AFin}}(n) = |\text{aff}_k^{\text{AF}}(n) - G_k^{\text{AF}} I - \theta|^+$, with $I = |\langle \text{aff}^{\text{AF}}(n) \rangle - \theta_I^{\text{AF}}|^+$, and $\text{aff}_k^{\text{AF}}(n) = \sum_j T_{kj} \sqrt{r_j^{\text{EC}}(n)}$. The activity r_k^{AFout} in AFP output assembly k was calculated as the sum of three terms (Fig. 7)

$$\begin{aligned} r_k^{\text{AFout}_1}(n) &= 0.15 \times r_k^{\text{AFin}}(n) \\ r_k^{\text{AFout}_2}(n) &= 0.75 \times c^{\text{R}} [0.15 + 0.85 \langle R^{\text{sy}}(n) \rangle] \\ r_k^{\text{AFout}_3}(n) &= 0.10 \times r_{k-1}^{\text{AFin}}(n-1) \end{aligned}$$

where $c^{\text{R}} = 20$ and $R_k^{\text{sy}}(n) = |\hat{R}_k^{\text{sy}} r_k^{\text{AFin}}(n) - \phi_k|^+$ as in SLA(6). Note that $r_k^{\text{AFout}_2}$ is identical for all k . The magnitude of the reinforcement signal for assembly i in RA is proportional to the total amount of excitatory input received from the AFP: $R_i(n) = \sum_k [RA, AF]_{ik} r_k^{\text{AFout}}(n) / (5[RA, AF])$. The correction factor $5[RA, AF]$ ensures that the magnitude of reinforcement corresponds to that used in SLA.

5. Calculate RA activity. The afferent input to RA is calculated as the sum of inputs from HVc_RA and the AFP, with the mean of the AFP input subtracted off due to feedforward inhibition (Fig. 7): $\text{aff}_i^{\text{RA}}(n) = \sum_j [RA, HR]_{ij} r_j^{\text{HR}}(n) + \sum_k [RA, AF]_{ik} [r_k^{\text{AFout}}(n) - \langle r^{\text{AFout}}(n) \rangle]$. Calculation of RA activity is the same as in SLA(3). To monitor convergence, once every 250 syllables, the simulations were continued over the interval $[0, 10]$. As in SLA, the root-mean-square (RMS) difference between short and long simulations [run every 250 syllables to monitor convergence of RA dynamics, see SLA(3)] was less than 0.1 except during the final stages of syllable learning (syllables 8750–11,500).

6. Calculate later portions of HVc_AFP activity ($r^{\text{HA-L}}$ and $r^{\text{HA-G}}$). Same as in SLA(4).

7. Update synaptic strengths. Calculation of plasticity followed the same rule described in SLA(7). The postsynaptic plasticity signal in

RA, $\rho_i^{\text{RA}}(n) = R_i(n) r_i^{\text{RA}}(n)$. The time window for HVc_AFP \rightarrow HVc_RA context learning was given as a difference of exponentials beginning after a 50-ms delay: $\alpha(\tau) = (e^{-(\tau-50)/\tau_{\text{fall}}} - e^{-(\tau-50)/\tau_{\text{rise}}})/a$ for $\tau > 50$ ms. $\tau_{\text{rise}} = 50$ ms, $\tau_{\text{fall}} = 150$ ms, and a is a normalizing constant that ensures that $\alpha(\tau)$ has a maximum value of 1. Learning rate parameters for new plastic connections are $k^{\text{HR,HA}} = 1 \times 10^{-7} \text{ ms}^{-2}$, $k^{\text{RA,RF}} = 3 \times 10^{-11} \text{ ms}^{-2}$. The threshold for long-term potentiation (LTP) and long-term depression (LTD) in HVc_RA was determined using $b^{\text{HR}} = 0.4$. Connections from the AFP \rightarrow RA used a separate LTP/LTD threshold (see METHODS), $b^{\text{RA,RF}} = 3.5$.

8. Update and apply homeostatic mechanisms.

8a. Normalize synaptic strengths. Normalization follows SLA(8a). The total input received by each population remained the same as in SLA; in RA, synaptic strengths were reduced by 20% to accommodate the new connections from the AFP. So $[RA, RA] = 0.15 = 0.4 \times 15/40$; $[RA, HR] = 0.3 = 0.4 \times 15/200$; and $[RA, AF] = 0.6 = 0.2 \times 15/5$. Context inputs from HVc_AFP contributed 20% of the input to HVc_AFP: $[HR, HA] = 0.1 = 0.2 \times 20/40$.

8b. Update inhibitory strengths. Same as in SLA(8b).

8c. Update adaptation. HVc_AFP adaptation follows SLA(8c). HVc_RA adaptation was of the same form but was updated only once at the end of each syllable. $\tau_{\text{decay}}^{\text{HA}} = 225$ ms and $h^{\text{HR}} = 6.133 \text{ ms}^{-1}$. Assuming a constant activity level of 1 during periods of HVc_RA activity, adaptation would have a strength of 12 input units, 60% of the total excitatory input to HVc_RA.

8d. Compute context signal. The context activity, r^{ctx} , was determined from the average HVc_AFP activity during the middle (35-ms long), late (20-ms long), and gap (35-ms long) portions of the syllable: $\hat{r}_j^{\text{ctx}}(n) = [35 r^{\text{HA-M}}(n) + 20 r^{\text{HA-L}}(n) + 35 r^{\text{HA-G}}(n)] / (35 + 20 + 35)$. To reduce instability resulting from reciprocal positive feedback in HVc, the context signal for each syllable was normalized to have average value equal to 1, i.e., $r_j^{\text{ctx}}(n) = \hat{r}_j^{\text{ctx}}(n) / \langle \hat{r}^{\text{ctx}}(n) \rangle$.

9. Calculate running averages of activity. Same as in SLA(9), with $r^{\text{AFin}} = r^{\text{AF}}$.

Increased number of tutor syllables

In some simulations, the number of tutor syllables was increased from five to eight. The number of input assemblies was increased accordingly, and the connections from HVc_AFP were increased by a factor of 8/5 so that the total connection strength onto each AFP input assembly remained at 15. To ensure proper sensory \rightarrow motor learning in the connections from the AFP \rightarrow RA, the learning rate was slowed, $k^{\text{RA,AF}} \rightarrow 0.5 \times k^{\text{RA,AF}}$, and the LTP/LTD threshold was reduced slightly, $b^{\text{RA,AF}} = 3$. To push the model harder, all learning rates were increased by a factor of 5, i.e., $k^{\text{post,pre}} \rightarrow 5 \times k^{\text{post,pre}}$ and $k^{\text{inh}} \rightarrow 5 \times k^{\text{inh}}$. All other parameters remained fixed.

We thank B. Baird, D. Buonomano, C. Linster, A. Krukowski, K. Miller, and members of the Doupe lab for many helpful comments. Special thanks to K. Miller for input and support throughout the project.

This work was supported by the McDonnell-Pew Program in Cognitive Neuroscience (T. W. Troyer), and National Institutes of Health Grants MH-12372 (T. W. Troyer) and MH-55987 and NS-34835 (A. J. Doupe).

REFERENCES

- ADAMS JA. Learning of movement sequences. *Psychol Bull* 96: 3–28, 1984.
- ALDRIDGE JW AND BERRIDGE KC. Coding of serial order by neostriatal neurons: a “natural action” approach to movement sequence. *J Neurosci* 18: 2777–2787, 1998.
- BENTON S, MARLER P, NELSON DA, AND DEVOOGD TJ. Anterior forebrain pathway is needed for stable song expression in adult male white-crowned sparrows (*Zonotrichia leucophrys*). *Behav Brain Res* 96: 135–150, 1998.
- CONTRERAS-VIDAL JL AND SCHULTZ W. A predictive reinforcement model of dopamine neurons for learning approach behavior. *J Comput Neurosci* 6: 191–214, 1999.

- CYXN J. Experimental determination of a unit of song production in the zebra finch. *J Comp Psychol* 104: 3–10, 1990.
- DOYA K AND SEJNOWSKI TJ. A computational model of birdsong learning by auditory experience and auditory feedback. In: *Central Auditory Processing and Neural Modeling*, edited by Poon PWF and Brugge JF. New York: Plenum, 1998, p. 77–88.
- FEE M, SHRAIMAN B, PESARAN B, AND MITRA PP. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* 395: 67–71, 1998.
- GOLLER F AND LARSEN ON. A new mechanism of sound generation in songbirds. *Proc Natl Acad Sci USA* 94: 14787–14791, 1997.
- HEBB DO. *The Organization of Behavior*. New York: Wiley, 1949.
- HIKOSAKA O, NAKAHARA H, RAND MK, SAKAI K, LU X, NAKAMURA K, MIYACHI S, AND DOYA K. Parallel neural networks for learning sequential procedures. *Trends Neurosci* 22: 464–471, 1999.
- HOUGHTON G AND HARTLEY T. Parallel models of serial behavior: Lashley revisited. *Psyche* 2, 1996. [Retrieved February 15, 1999, from the World Wide Web: <http://www.cs.monash.edu.au/volume2-1/psyche-95-2-25-lashley-1-houghton.html>]
- HOUK JC, ADAMS JL, AND BARTO AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, and Beiser DG. Cambridge, MA: MIT Press, 1995, p. 248–270.
- JAMES W. *The Principles of Psychology*. Cambridge, MA: Harvard Univ. Press, 1983. [Original work published in 1890.]
- JEANNEROD M. *The Neural and Behavioral Organization of Goal-Directed Movements*. Oxford, UK: Clarendon Press, 1988.
- JOHNSON F, HOHMANN SE, DI STEFANO PS, AND BOTTJER SW. Neurotrophins suppress apoptosis induced by deafferentation of an avian motor-cortical region. *J Neurosci* 17: 2101–2111, 1997.
- JORDAN MI. Computational motor control. In: *The Cognitive Neurosciences*, edited by Gazzaniga MS. Cambridge, MA: MIT Press, 1995, p. 597–610.
- KATZ LC AND GURNEY ME. Auditory responses in the zebra finch's motor system for song. *Brain Res* 211: 192–197, 1981.
- KIMPO RR AND DOUPE AJ. FOS is induced by singing in distinct neuronal populations in a motor network. *Neuron* 18: 315–325, 1997.
- KLEINFELD D AND SOMPOLINSKY H. Associative neural network model for the generation of temporal patterns: Theory and application to central pattern generators. *Biophys J* 54: 1039–1057, 1988.
- KONISHI M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z Tierpsychol* 22: 770–783, 1965.
- KUBOTA M AND TANIGUCHI I. Electrophysiological characteristics of classes of neuron in the HVC of the zebra finch. *J Neurophysiol* 80: 914–923, 1998.
- LASHLEY KS. The problem of serial order in behavior. In: *Cerebral Mechanisms in Behavior (the Hixon Symposium)*, edited by Jeffress LA. New York: Wiley, 1951, p. 112–136.
- LEE BS. Effects of delayed speech feedback. *J Acoust Soc Am* 22: 824–826, 1950.
- LEONARDO A AND KONISHI M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399: 466–470, 1999.
- LEWICKI MS. Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. *J Neurosci* 16: 5855–5863, 1996.
- LEWICKI MS AND ARTHUR BJ. Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16: 6987–6998, 1996.
- LEWICKI MS AND KONISHI M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci USA* 92: 5582–5586, 1995.
- LOHR B AND DOOLING RJ. Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J Comp Psychol* 112: 36–47, 1998.
- MACKEY DG. Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia* 8: 323–350, 1970.
- MARGOLIASH D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3: 1039–1057, 1983.
- MARGOLIASH D AND FORTUNE ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12: 4309–4326, 1992.
- MARGOLIASH D, FORTUNE ES, SUTTER ML, YU AC, WREN-HARDIN BD, AND DAVE A. Distributed representation in the song system of oscines: evolutionary implications and functional consequences. *Brain Behav Evol* 44: 247–264, 1994.
- MARLER P. Inheritance and learning in the development of animal vocalizations. In: *Acoustic Behavior of Animals*, edited by Busnel RG. Amsterdam: Elsevier, 1964.
- MARLER P AND PETERS S. Developmental overproduction and selective attrition: new processes in the epigenesis of birdsong. *Dev Psychobiol* 15: 369–378, 1982.
- MATSUMOTO N, HANAKAWA T, MAKI S, GRAYBIEL AM, AND KIMURA M. Nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. *J Neurophysiol* 82: 978–998, 1999.
- MCCAULEY JH, ROTHWELL JC, AND MARSDEN CD. Frequency peaks of tremor, muscle vibration and electromyographic activity at 10 Hz, 20 Hz and 40 Hz during human finger muscle contraction may reflect rhythmicities in central neural firing. *Exp Brain Res* 114: 525–541, 1997.
- MONTAGUE PR, DAYAN P, AND SEJNOWSKI TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936–1947, 1996.
- MOONEY R. Synaptic basis for developmental plasticity in a birdsong nucleus. *J Neurosci* 12: 2464–2477, 1992.
- MOONEY R AND KONISHI M. Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proc Natl Acad Sci USA* 88: 4075–4079, 1991.
- MORLEY JW, GOODWIN AW, AND DARIAN-SMITH I. Tactile discrimination of gratings. *Exp Brain Res* 49: 291–299, 1983.
- NELSON DA AND MARLER P. Selection-based learning in bird song development. *Proc Natl Acad Sci USA* 91: 10498–10501, 1994.
- NICOLELLIS MA, BACCALA LA, LIN RC, AND CHAPIN JK. Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268: 1353–1358, 1995.
- NORDEEN KW AND NORDEEN EJ. Projection neurons within a vocal motor pathway are born during song learning in zebra finches. *Nature* 334: 149–151, 1988.
- NOTTEBOHM F, ALVAREZ-BUYLLA A, CYXN J, LING C, NOTTEBOHM M, SUTER R, TOLLES A, AND WILLIAMS H. Song learning in birds: the relation between perception and production. *Philos Trans R Soc Lond B Biol Sci* 329: 115–124, 1990.
- OKANOYA K AND KOBAYASHI K. Lesioning a nucleus (area X) of the anterior forebrain pathway affects the production of learned song in adult Bengalese finches (*Lonchura striata* var. *domestica*) (Abstract). *5th Int Soc Neuroethology* 1998, No. 303.
- OKANOYA K AND YAMAGUCHI A. Adult Bengalese finches (*Lonchura striata* var. *domestica*) require real-time auditory feedback to produce normal song syntax. *J Neurobiol* 33: 343–356, 1997.
- PRICE PH. Developmental determinants of structure in zebra finch song. *J Comp Physiol Psychol* 93: 268–277, 1979.
- ROSENBAUM DA. *Human Motor Control*. San Diego, CA: Academic Press, 1991.
- SAITO N AND MAEKAWA M. Birdsong, the interface with human language. *Brain Dev* 15: 31–39, 1993.
- SANES JN, MAURITZ KH, DALAKAS MC, AND EVARTS EV. Motor control in humans with large-fiber sensory neuropathy. *Human Neurobiol* 4: 101–114, 1985.
- SCHARFF C, KIRN JR, GROSSMAN M, MACKLIS JD, AND NOTTEBOHM F. Targeted neuronal death affects neuronal replacement and vocal behavior in adult songbirds. *Neuron* 25: 481–492, 2000.
- SCHARFF C, KIRN JR, MACKLIS JD, AND NOTTEBOHM F. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J Neurobiol* 36: 81–90, 1998.
- SCHARFF C AND NOTTEBOHM F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J Neurosci* 11: 2896–2913, 1991.
- SOHRABJI F, NORDEEN EJ, AND NORDEEN KW. Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behav Neural Biol* 53: 51–63, 1990.
- SOSSINKA R AND BÖHNER J. Song types in the zebra finch (*Poephila guttata castanotis*). *Z Tierpsychol* 17: 123–132, 1980.
- SPERRY RW. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 43: 482–489, 1950.
- SPIRO JE, DALVA MB, AND MOONEY R. Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. *J Neurophysiol* 81: 3007–3020, 1999.
- STARK LL AND PERKEL DJ. Two-stage, input-specific synaptic maturation in a nucleus essential for vocal production in the zebra finch. *J Neurosci* 19: 9107–9116, 1999.
- SUTHERS RA, GOLLER F, AND HARTLEY RS. Motor dynamics of song production. *J Neurobiol* 25: 917–936, 1994.

- THEUNISSEN FE AND DOUPE AJ. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J Neurosci* 18: 3786–3802, 1998.
- TROYER TW AND DOUPE AJ. An associational model of birdsong sensorimotor learning. I. Efference copy and the learning of song syllables. *J Neurophysiol* 84: 1204–1223, 2000.
- VOLMAN SF. Development of neural selectivity for birdsong during vocal learning. *J Neurosci Meth* 13: 4737–4747, 1993.
- VON HOLST E AND MITTELSTAEDT H. The reafference principle: interaction between the central nervous system and the periphery. In: *The Organization of Action*, edited by Gallistel CR. Hillsdale, NJ: Lawrence Erlbaum, 1980, p. 176–209.
- VU ET, MAZUREK ME, AND KUO Y. Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci* 14: 6924–6934, 1994.
- WHITE SA, LIVINGSTON FS, AND MOONEY R. Androgens modulate NMDA receptor-mediated EPSCs in the zebra finch song system. *J Neurophysiol* 82: 2221–2234, 1999.
- WILLIAMS H AND STAPLES K. Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J Comp Psychol* 106: 278–286, 1992.
- WOOLLEY SM AND RUBEL EW. Bengalese finches *Lonchura striata domestica* depend upon auditory feedback for the maintenance of adult song. *J Neurosci* 17: 6380–6390, 1997.
- YU AC AND MARGOLIASH D. Temporal hierarchical control of singing in birds. *Science* 273: 1871–1875, 1996.