An Associational Model of Birdsong Sensorimotor Learning II. Temporal Hierarchies and the Learning of Song Sequence

TODD W. TROYER1 AND ALLISON J. DOUPE2

1,2Department of Psychiatry, 2Department of Physiology, 1,2W. 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 RA, the premotor nucleus 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 associative 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 (Nordene and Nordene 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
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) → sensory (HVc_AFP) → next motor (HVc_RA) → 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 → 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
solution involves the concerted action of multiple associa-
tional 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 syl-
labing. 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 gen-
erator 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 →
sensory mapping from HVc_RA → HVc_AFP. The other half
of the interaction depends on connections from HVc_AFP →
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 →
sensory → next motor → next sensory, etc. This hypothesis
borrows from classical chaining ideas (James 1983), as well as
more recent computational models (Kleinfeld and Sompoli-
sky 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 to-
ward a desired goal is to bias the pattern of neural activity
toward the desired state. Associational plasticity then strength-
ens 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). Associ-
atational plasticity then changes the pattern of connections be-
 tween 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 → 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 learn-
ing, 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 partic-
lar, by altering connections between HVc and RA, the AFP
changes the pattern of vocal output and hence auditory reaf-
ference. 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 syllab-
le learning (Fig. 2). The connections from HVc_AFP →
HVc_RA are necessary for sequence generation. Without the
d context signals carried by these connections, activity within

---

**Table 1. Functional hypotheses for sequence learning**

<table>
<thead>
<tr>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reciprocal interaction in HVc</td>
</tr>
<tr>
<td>2. Slow sensory → motor “context” signal (HVc_AFP → HVc_RA)</td>
</tr>
<tr>
<td>3. Sensory → motor associations guide development of connections from A FP → RA</td>
</tr>
<tr>
<td>4. AFP guided changes in RA trigger renewed efference copy learning in HVc</td>
</tr>
</tbody>
</table>
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 A 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 $Mot_A$ 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 $Sen_A → (Ctxt_A → Mot_A → Sen_B → . . .)$. 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

![Figure 3](image3.png) 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 $Mot_A$ in RA. Syllable B follows syllable A. $Ctxt_A$ also elicits $Sen_B$, the efference copy corresponding to $Mot_B$. $Sen_B → Ctxt_B → Mot_C$ leads to the production of syllable C, etc.

![Figure 4](image4.png) 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 $Sen_A → Ctxt_A → Mot_B$, resulting in the transition from A to B.
the vocalization that follows syllable A. First, in HVc, due to the slow connections from HVc_AFP → 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 → Ctxt_A → Mot_B.

SENSORY → MOTOR MAPPING FROM THE AFP → 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 → 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 → next motor transition, Sen_A → Ctxt_A → Mot_B, sequence learning is not yet complete. This is because by altering synapses in RA, the AFP has perturbed the motor activity → 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 → Ctxt_A → Sen_D → Ctxt_D. Presumably, the context signal from syllable D, Ctxt_D, is mapped onto Mot_B 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 → 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 → 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 → Sen_D. In this way, the HVc sequence generator is able to track the AFP-induced changes in RA. By combining the appropriate sensory → motor and motor → sensory mappings, the model learns the chain of sensory-motor associations that reproduces the tutor sequence: Sen_A → (Ctxt_A → Mot_B) → Sen_B . . . .

FIG. 6. Renewed efference copy learning. Since AFP-guided sequence learning alters the projection from HVc_RA → 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.
METHO DS

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 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 → 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 → 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 → 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 → 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 → RA. To counteract this tendency, AFP → 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 output 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 → motor associational learning follows, causing AFP assemblies encoding a particular tutor syllable to project most strongly to RA assemblies encoding the same syllable. (Fig. 5). After the sensory → 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 → 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 → RA connections perform a sensory → 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 → 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 → 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 → 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 suberved by mechanisms tied to developmental critical periods. Our model makes predictions regarding the functional information carried by the AFP → 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 matrix, $M_{ij}$, the matrix of co-fluctuations between patterns of RA activity for a given syllable and the patterns of RA activity for the next syllable

$$M_{ij}^{	ext{next}} = \frac{1}{250} \sum_{n=250, n\neq i}^{250} \left[ r_i(n-1) - \bar{r}(n-1) \right] \left[ r_j(n) - \bar{r}(n) \right]$$

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 $M_{ij}^{	ext{next}}$ to an ideal syllable transition matrix, $M_{ij}^{	ext{seq}}$, $M_{ij}^{	ext{seq}} = 4$, if assembly $j$ forms part of the representation for the syllable following the syllable coded by assembly $i$; $M_{ij}^{	ext{seq}} = -1$, 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 → motor connections (HVc_RA → RA) and HVc_RA → sensory connections (HVc_RA → HVc_AFP). 2) To quantify the development of the sensory → motor mapping (Fig. 5), we computed the CC between the pattern of AFP → 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, $M^{eq}$, and the pattern of intrinsic RA connections as in Troyer and Doupe 2000. $M^{eq}_{ij} = 4$, if assembly $i$ forms part of the representation for same syllable as assembly $j$; $M^{eq}_{ii} = -1$, otherwise. Diagonal entries were excluded. 4) To evaluate sequence-related connectivity, we multiplied the HVc_AFP → HVc_RA and HVc_RA → 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 $M^{eq}$ 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 consists of 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 → 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 → 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-
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 \( i \)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 \( i \)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. 10A. Figure 10B shows which connections are most important during each of the song learning stages traced in Fig. 10A. 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 → motor mapping also develops between the AFP and RA (stage 1a, 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
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 → HVc_RA → RA (thick dashed line) yields correct sensory → next motor syllable transitions. Note that improvement in the sequencing of RA activity happens before the learning of the appropriate connectivity from HVc_AFP → HVc_RA → RA, since AFP-driven sequence transitions are necessary to drive sequence-related learning. The reorganization of the HVc_RA → 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 → 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 → 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 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 → 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 × 5 = 40). AFP circuitry was adjusted for the different template,
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 → 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 → 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-

![Figure 12](image-url) 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.
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 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 → sensory → next motor → 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 CxtA, 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 associaional 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 Doolding 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 → 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 → 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 → 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 → 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 represen-
tations 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).

**Individually 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 (~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 plast-
icity 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 associa-
tional 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
(McCauley et al. 1997), and whisker twitching in rats
(Nicolesis et al. 1995).

**Sequence generation (>100 ms).** Learning temporal structure
on time scales greater than individual syllables poses signifi-
cantly fewer problems than learning structure at fine time
scales. Sensory → 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 Notebohm
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 Notebohm 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
fully 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
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 → 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 reperto-
ire (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 represen-
tations 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 de-
velopmental time courses of syllable morphology and syl-
vable sequence will be crucial for understanding the mech-
nisms 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: i)
adding plastic connections from HVc_AFP → HVc_RA and from the
AFP → 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 $r^{HR}$, $r^{RA}$, $r^{HA-E}$, $r^{HA-M}$, $r^{HA-I}$, $r^{HA-G}$ 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 contributed 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 $\rho^{AFin}$ and $\rho^{AFout}$. We use $\text{post, pre}$ 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 ($r^{HA-E}$ and $r^{HA-M}$). Calculation of $r^{HA-E}$ and $r^{HA-G}$ 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: $\text{aff}_HR(n) = p(n) + \sum [HR, HA] r_i^{EC}(n - 1)$. Output firing rates are determined as in SLA(2).

3. Calculate earlier portions of HVc_AFP activity ($r^{HA-E}$ and $r^{HA-M}$). 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).

5. Calculate activity in AFP input assemblies follows SLA(5):

$$ r_i^{AFout}(n) = 0.15 \times r_i^{AFin}(n) $$

$$ r_i^{AFin}(n) = 0.75 \times e^{[0.15 + 0.85(R_i^{EC}(n))] - I} $$

$$ r_i^{AFout}(n) = 0.10 \times r_i^{AFout}(n - 1) $$

where $r_i^{AFin}$ is the firing rate for assembly $i$ in RA proportional to the total amount of excitatory input received from the AFP. $R_i(n) = \sum [RA, AF] r_i^{AFout}(n)/5[RA, AF]$. The correction factor $5[RA, AF]$ ensures that the magnitude of reinforcement corresponds to that used in SLA.

6. 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):

$$ \text{aff}_HR(n) = \sum [HR, HA] r_i^{HR}(n) + \sum [RA, AF] r_i^{AFin}(n) - (r_i^{AFout}(n)). $$

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 between the final stages of syllable learning (syllables 7850–11,500).

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

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. $\tau_{decay} = 225$ ms and $b^{HR} = 6.133$ 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, $r^{ext}$, 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: $r_i^{ext}(n) = \{35 [HA-E](n) + 20 [HA-M](n) + 35 [HA-G](n)\}/(35 + 20 + 35)$. To reduce instabilities resulting from reciprocal positive feedback in HVc, the context signal for each syllable was normalized to have average value equal to 1, i.e., $r_i^{ext}(n) = r_i^{ext}(n)/\langle r_i^{ext}(n) \rangle$.

9. Calculate running averages of activity. Same as in SLA(9), with $r_i^{AFin} = r_i^{RF}$.

Increased number of tutor syllables

In some simulations, the number of tutor syllables was increased from five to eight. The number of AFP 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 → motor learning in the connections from the AFP → RA, the learning rate was slowed, $k_{RA, RF} = 0.5 \times k_{RA, AF}$, and the LTP/LTD threshold was reduced slightly, $b_{RA, AF} = 3$. To push the model harder, all learning rates were increased by a factor of 5, i.e., $k_{post, pre} \times 5 \times k_{post, pre} = k_{post, pre} \times 5 \times k_{post, pre}$. All other parameters remained fixed.

We thank B. Baird, D. Buonomano, C. Linster, A. Krukowiecki, 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


