Neural mechanisms for learned birdsong

Richard Mooney

Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA

Learning by imitation is essential for transmitting many aspects of human culture, including speech, language, art, and music. How the human brain enables imitation remains a mystery, but the underlying neural mechanisms must harness sensory feedback to adaptively modify performance in reference to the object of imitation. Although examples of imitative learning in nonhuman animals are relatively rare, juvenile songbirds learn to sing by copying the song of an adult tutor. The delineation of neural circuits for birdsong raises the promise that this complex form of vocal learning, which bears strong parallels to human speech learning, can be understood in terms of underlying neural mechanisms. This promise is now being more fully realized, with recent experimental advances leading to better understanding of the central motor codes for song and the central mechanisms by which auditory experience modifies song motor commands to enable vocal learning.

Imitation is the foundation for transmitting much of human culture. Most importantly, the child's ability to vocally mimic the speech of parents and peers forms the foundation for spoken language (Locke 1993). Despite the fundamental importance of speech learning to human societies, and the widespread use of vocal communication by other animals, vocal learning in nonhuman species is quite rare. Oscine songbirds (order: Passeriformes) produce learned courtship songs (Fig. 1) that are culturally transmitted from one generation to the next, affording an experimentally tractable system in which to study the neurobiological underpinnings of vocal imitation.

Songbirds are a highly diverse group, comprising roughly 4000 species, and the song learning process has been studied in detail in only a small fraction (<10%) of these. Even this small sample reveals great diversity in the pace at which song is learned, the range of model songs juvenile birds will imitate, whether one or both sexes sing, and whether song serves primarily to attract mates or also is employed in territorial defense (for more detailed reviews, see Konishi 1985; Marler 1990; Catchpole and Slater 1995; Doupe and Kuhl 1999; Mooney et al. 2008). Despite this behavioral diversity, general aspects of the song-learning process appear to be well conserved across species.

A common theme of song learning is that it occurs during a juvenile sensitive period and involves two distinct phases—sensory learning and sensorimotor learning—that depend on auditory experience (Fig. 2). During sensory learning, the young bird listens to and memorizes one or more tutor songs. In the ensuing phase of sensorimotor learning, the pupil relies on auditory feedback to match its song to the memorized model. Sensorimotor learning ends with song “crystallization,” a process wherein the song becomes highly stereotyped and usually much less dependent on auditory feedback.

During sensory learning, the juvenile songbird listens to and memorizes the song of a tutor, which in nature is oftentimes its father or a neighboring adult male conspecific. Although a human appreciation for song learning probably dates back many centuries, the modern study of song learning emerged from the systematic description of regional intraspecific song dialects (Thorpe 1958). The first experimental evidence of song learning was the finding that juveniles exposed to conspecific songs from other regions would subsequently sing these foreign dialects (Marler and Tamura 1964). Furthermore, juveniles raised in acoustic isolation from other birds develop songs lacking many of the fine variations in pitch and duration that typify wild-type songs (Thorpe 1954, 1958; Immelmann 1969; Marler 1970; Marler et al. 1972; Price 1979). Finally, in many instances, young songbirds can copy the song of another species (Immelmann 1969; Baptista and Petrinovich 1984, 1986; Marler 1991). All of these findings underscore that imitation underlies normal song development.

Sensory learning in songbirds shares important features with forms of sensory imprinting described in other birds and mammals. First, sensory learning is restricted to a sensitive period: In almost all species studied to date, song models heard after the first two or three months after hatching are copied poorly, if at all (Thorpe 1958; Immelmann 1969; Marler and Peters 1987, 1988). Second, sensory learning can be remarkably fast: Accurate imitation may ensue from hearing a tutor song only several hundred times or less, which constitutes only a few minutes of total auditory experience (Hultsch and Todt 1989a,b). Third, because sensory and sensorimotor learning are often separated by many months (Fig. 2; Marler and Peters 1981, 1982a), sensory learning results in a long-lasting and perhaps permanent memory, as is typical of other imprinted stimuli. Finally, the juvenile songbird shows an innate predisposition to copy certain songs but not others. For example, a juvenile swamp sparrow that hears songs of its own and other species during sensory learning displays a strong bias to copy the conspecific songs and subsequently retain these copies in its adult repertoire (Marler and Peters 1987, 1989; Marler 1990). Presumably, selection of an appropriate tutor is aided by an innately specified auditory representation in the juvenile's brain that encodes basic features of conspecific song. Such an innate representation may also explain why acoustic features of isolate songs share common features within species and differ across species (Marler and Sherman 1985).

During the second phase of song learning—sensorimotor learning—the juvenile begins to sing, gradually modifying its song to match the tutor song. Sensorimotor learning often occurs weeks or even months after the juvenile has left the company of its tutor, indicating that vocal matching can be accomplished entirely from memory (Marler and Peters 1981, 1982a). Even though the juvenile need not hear the tutor during sensorimotor learning, hearing still plays a critical role. More than four decades ago, Mark Konishi discovered that juvenile white-crowned sparrows deafened after extensive auditory experience of a tutor song, but prior to sensorimotor learning developed highly noisy and poorly...
structured songs (Konishi 1965b). Interestingly, Konishi also found that isolate white crowned sparrows deafened prior to the onset of singing develop similarly noisy and disorganized songs. Thus, during sensorimotor learning, the juvenile uses auditory feedback to match its song to an internalized song model, regardless of whether it previously heard a tutor song. Normally, the model is an acoustic memory of the tutor song; in isolates, the model must be innately specified. This internal song model—whether learned or innate—is often referred to as a song “template.”

In contrast to sensory learning, sensorimotor learning occurs slowly and involves extensive vocal experimentation. The three major stages to sensorimotor learning—subsong, plastic song, and crystallized song—unfold over many weeks or even months (Immelman 1969; Marler and Peters 1982b; Tchernichovski et al. 2001). The first stage—subsong—consists of soft and rambling vocalizations reminiscent of infant babbling and lacking evidence of imitation. The second stage—plastic song—exhibits phrase-like structure containing identifiable copies of at least parts of the tutor song. Plastic song can be distinguished from crystallized song by higher levels of bout-to-bout variability and by slower adaptive changes (i.e., song plasticity) that are manifested over days to weeks and that render the song increasingly similar to the tutor song. Because plastic song often spans two months or more and the juvenile may sing thousands of times each day, many tens or even hundreds of thousands of song renditions may be necessary to achieve accurate imitation. Sensorimotor learning terminates in a process known as crystallization, wherein levels of acute variability and vocal plasticity are greatly diminished. The adult male’s highly stereotyped crystallized song functions to attract mates and defend territory from other males.

Pioneering studies of song learning relied on human scoring of analog recordings and sonograms, limiting detailed analysis to only a small part of the juvenile’s massive song output (Immelman 1969; Marler 1970; Marler and Peters 1982b). In two seminal studies using automated digital recording and analytical techniques, Ofer Tchernichovski and his colleagues provided a note-by-note account of how a male juvenile zebra finch imitates its tutor (Tchernichovski et al. 2001; Deregnaucourt et al. 2005). Because sensory learning and sensorimotor learning normally overlap in this species, Tchernichovski first raised juveniles in isolation and then employed an operant playback method to precisely control their exposure to tutor song. About 24 h after the initial tutoring session, the acoustic features of the juvenile’s song phrase become more diverse, reflecting its initial attempts at imitation. Bout-to-bout song structure also becomes more variable with the onset of imitation and, quite intriguingly, this variability displays a daily rhythm: Song structure markedly deteriorates following each night of sleep, only to recover during the first several hours of morning singing. During this morning recovery, the juvenile’s song increases in its similarity to that of the tutor and then reaches a plateau, despite continued abundant singing throughout the rest of the day.

The net result of nightly deterioration and daily recovery is adaptive, because the structure of the juvenile’s song becomes more similar to the tutor song from one day to the next (Deregnaucourt et al. 2005). Moreover, juveniles that display greater song deterioration after sleep ultimately mimic the tutor song with greater accuracy, suggesting that song deterioration somehow facilitates imitation. One possibility is that nightly deterioration allows for more complete exploration of vocal-acoustic “space,” helping to push the system out of local minima that may otherwise prevent the bird from achieving a closer match to the tutor song.

These various observations have encouraged speculation that variability is an essential ingredient for song learning, perhaps because it provides the stepping stones over which more gradual, substantial, and persistent changes to song can occur. A causal role for variability in sensorimotor learning seems likely, motivating intense efforts to identify the neural mechanisms that generate song variability. Nonetheless, the causal link between acute variability and sensorimotor learning has yet to be established.

Crystallization, which marks the end of sensorimotor learning, is characterized by increased song stereotypy (i.e., diminished bout-to-bout variability and evening-to-morning changes to song) (Deregnaucourt et al. 2005), and decreased dependence on auditory feedback for stable song maintenance (Konishi 1965b; Price 1979). The extent to which dependence on feedback diminishes following crystallization varies across species, an important point to keep in mind when considering the various attempts to isolate feedback-related signals in the singing bird. At one extreme, the crystallized songs of adult white-crowned sparrows appear largely impervious to deafening (Konishi 1965b). In contrast, adult zebra finches and Society finches maintain their crystallized songs via auditory feedback. In these species, deafening (Nordeen and Nordeen 1992), distorting song (i.e., by damaging the vocal nerve) or delaying or distorting auditory feedback (both referred to here as DAF) (Williams and McKibben 1992; Leonardo and Konishi 1994)

Figure 1. Song structure and nomenclature. A sound spectrograph, or sonogram, of the song of an adult male white-crowned sparrow (Zonotrichia leucophrys). Notes (arrows) are the elementary units of song, and they are grouped to form syllables (letters), a defined sequence of which constitutes the song phrase or motif. In this sonogram, darker traces correspond to louder sounds. This song is ~1 sec in duration.

Figure 2. Developmental timelines of song learning. During sensory acquisition the juvenile bird listens to and memorizes one or more tutor songs. During sensorimotor learning the juvenile matches its own song to the memorized model using auditory feedback. Sensorimotor learning includes an initial stage of babbling-like vocalizations known as subsong, followed by more structured but variable plastic song. Song crystallization marks the end of sensorimotor learning. Different songbird species illustrate similarities and variations in the timing of these two stages. (A) In zebra finches, the sensory and sensorimotor periods overlap extensively, and song crystallization is complete between 90 and 120 d after hatching. (B) In white-crowned sparrows, as with other temperate zone songbird species, sensory acquisition and sensorimotor learning are separated by many months, and song crystallization occurs only at the end of the first year of the bird’s life. In sparrows, seasonal changes in day length can trigger renewed plastic singing and recrystallization.
cause the previously crystallized song to become more plastic. Over time, these plastic changes include spectral distortion of individual syllables and temporal reorganization of the motif. In an adult zebra finch, deafening or exposure to DAF only triggers this plastic process of song “decrystallization” after weeks or even months (the delay correlating with the animal’s age), whereas similar manipulations in the juvenile trigger maladaptive vocal plasticity within several days (Price 1979; Williams and McKibben 1992; Leonardo and Konishi 1999; Lombardino and Nottebohm 2000; Kozhevnikov and Fee 2007). Notably, the Society finch, a semi-domesticated species closely related to the zebra finch, maintains an acute dependence on auditory feedback in adulthood, with deafening or exposure to DAF triggering song deterioration within days (Okanoya and Yamaguchi 1997; Woolley and Rubel 1997).

Although highly stereotyped, a bird’s crystallized song still varies slightly from one bout to the next (Deregnaucourt et al. 2005). This residual variability is quite small: Even adult Society finches, which express a relatively high level of variability in their crystallized songs, still control fundamental frequency of individual syllables to within ~1%—a value much more precise than observed for human speech phones and only slightly less precise than repetitions of a single note sung by a trained human vocalist (Sundberg et al. 1996; Amir et al. 2003; Tumer and Brainard 2007). The functional significance of this residual variability remains unclear. On the one hand, persistent variability in crystallized song may reflect behaviorally irrelevant “noise” that the adult songbird cannot detect or control. Alternatively, the adult may not only sense this residual variability but also be able to exploit it to adaptively modify its song.

To distinguish between these possibilities, Tumer and Brainard disrupted auditory feedback in a contingent-dependent manner: Higher pitched renditions of a target syllable triggered a concomitant noise burst, whereas with lower pitched renditions the bird experienced normal feedback (Tumer and Brainard 2007). Adult male Society finches subjected to this protocol displayed rapid (~1 d or less) adaptive shifts in fundamental frequency of the target syllable, ultimately reducing the number of renditions that triggered noise bursts. Furthermore, vocal changes were restricted to the fundamental frequency of the target syllable, could involve frequency shifts in either direction, depending on the applied contingency, and did not involve other syllables or other acoustic features of the target syllable. These findings indicate that adult male Society finches can detect and control subto variations in their crystallized songs, and suggest that they can harness this variability to enable song plasticity.

These findings also raise the possibility that this adult form of song plasticity employs an error correction mechanism similar to that used for sensorimotor learning. Alternatively, the form of song plasticity studied by Tumer and Brainard may reflect a form of operant learning in which syllables that trigger noise bursts undergo negative reinforcement. Interestingly, when an adult Bengalese finch is outfitted with tiny headphones, allowing the pitch of feedback of a target syllable to be systematically shifted up or down, it adjusts its song by shifting the pitch of the target syllable in the opposite direction, indicative of an error correction process (Sober and Brainard 2009).

Independent of whether song variability enables adult vocal plasticity, it is likely to serve an important communication function. The crystallized songs of male zebra finches display different amounts of acoustic variability depending on social context, with songs directed toward a nearby female exhibiting significantly less variability than “undirected” songs (Sossinka and Bohner 1980; Kao et al. 2005). These different levels of variability are significant in terms of song’s courtship function, because directed songs elicit more potent receptive responses from adult female zebra finches than do undirected songs (Woolley and Doupe 2008). More generally, the effects of social context on song variability support speculation that undirected song constitutes a rehearsal phase that fine tunes directed song “performance,” an idea that might be tested by determining whether a male’s directed song deteriorates, or becomes less attractive to females, when he is prevented from singing undirected songs.

In summary, vocal learning in songbirds and humans share major similarities: both occur during a juvenile sensitive period, both depend on early auditory experience of acoustic models, and both rely on auditory feedback to enable vocal matching of species-appropriate models. Moreover, the auditory periphery of birds and humans share basic organizational features, such as a tonotopically organized sensory epithelium, and birdsong and speech both require precisely integrated vocal and respiratory muscle activity. Therefore, the brains of juvenile songbirds and humans must accomplish highly similar sensorimotor transformations to enable vocal learning. As a result, neurobiological studies of birdsong are likely to reveal mechanisms that are highly pertinent to understanding how humans use audition to learn and maintain speech.

Neural circuits for song learning: General concepts

Prior to delving into songbird neurobiology, it is useful to conceptualize some of the major tasks the brain must accomplish to enable vocal learning. One is to generate motor commands. Another is a means of varying these motor commands to match the song model. Accurate matching in turn depends on the brain’s ability to use auditory feedback to evaluate song performance in reference to a memorized song model. An influential idea is that this evaluation is accomplished by a comparator circuit, the inputs to which are the auditory feedback signal and the memorized song model (Fig. 3A). When these two inputs differ, the comparator emits a permissive or instructive signal that causes song to change, subsequently minimizing the difference between the song and the memorized model.

The nature of the comparator remains hypothetical and how auditory feedback, the memorized song model, and instructive signals are instantiated in the songbird’s brain is not well understood. In contrast, substantial light has been shed on the neural mechanisms for singing and generating acute song variability. Because song variability may be an important ingredient for song learning, an important goal is to identify where variability signals originate in the brain and how these signals might be affected by auditory feedback. Moreover, lesion studies indicate that the same brain pathway necessary for high levels of acute song variability is also essential to slower forms of song plasticity, including the slow changes to song that occur during sensorimotor learning, raising the possibility that this pathway constitutes part of a comparator circuit. Thus, it is useful to first explore the neural machinery that enables singing, song variability, and plasticity, and then consider how this machinery may be harnessed by the auditory system to enable song learning.

The song system, a central circuit for singing and song learning

A specialized forebrain circuit essential to singing and song learning distinguishes the songbird’s brain from the brain of birds that do not learn to vocalize (Fig. 4; Nottebohm et al. 1976, 1982; Kroodsma and Konishi 1991; Wild 1997a,b, 2004). This forebrain circuit sits atop an evolutionarily more ancient brainstem circuit important to the production of unlearned vocalizations in birds as well as mammals. The integrated forebrain and brainstem circuit, referred to collectively as the song system, ultimately affects vocalization through the muscles of the avian vocal organ, or
display singing-related activity (McCasland and Konishi 1981; McCasland 1987; Yu and Margoliash 1996; Hessler and Doupe 1999a; Hahnloser et al. 2002; Leonardo 2004; Leonardo and Fee 2005), and many of these neurons also respond to auditory stimulation (Katz and Gurney 1981; McCasland and Konishi 1981; Margoliash 1983, 1986; Doupe and Konishi 1991; Vicario and Yohay 1993), pointing to a functional linkage by which auditory information could influence vocal control.

The telencephalic song nucleus HVC is at the sensorimotor apex

In its total scope, the song system is a highly distributed network spanning the entire brain and conveying descending song motor signals, as well as ascending streams of auditory, proprioceptive, and recurrent motor information (Fig. 4B). Such a highly distributed network, with both feed-forward and recurrent architecture, is unlikely to function in a purely feed-forward manner. Nonetheless, certain anatomical and functional features support the view that the telencephalic nucleus HVC resides at or near the top of this sensorimotor hierarchy, making it a useful point of entry into this complex system.

Conspicuously, HVC is a common component of both the SMP and the AFP, while also serving as a major telencephalic target of ascending auditory and recurrent motor pathways. Indeed, the SMP and AFP arise from two distinct populations of HVC projection neurons (PNs) (Katz and Gurney 1981; Kirn et al. 1991; Fortune and Margoliash 1995; Dutar et al. 1998; Mooney 2000). One HVC PN type (HVC₃₄) innervates the telencephalic song nucleus RA, which contains PNs that send long axons down into the brainstem, where they terminate on syringeal motor neurons and respiratory premotor neurons (Nottebohm et al. 1976; Wild 1993). Thus, the SMP provides the anatomical basis by which the telencephalon exerts control over vocalization, a hallmark of both human speech and birdsong, but not of most other animal vocalizations.

The other HVC PN type (HVC₅₆) innervates area X, which along with the thalamic nucleus DLM and the anterior telencephalic nucleus LMAN, forms a pathway through the anterior forebrain that indirectly links HVC to RA (Nottebohm et al. 1982; Ohkohata and Saito 1987; Foster and Bottjer 1998; Luo et al. 2001). The basic organization of the songbird AFP resembles that of mammalian cortical-basal ganglia pathways: Area X is located in the striatum, DLM is located in the thalamus, and HVC and LMAN are equivalent to cortical structures (Doupe et al. 2005). The axons of LMAN PNs bifurcate as they exit the caudal margin of the nucleus; one branch travels caudally in the telencephalon and forms fast excitatory synapses in RA (Mooney 1992), while the other branch terminates in area X, forming a feedback loop within the AFP (Nixdorf-Bergweiler et al. 1995; Luo et al. 2001). A prominent feature of this organization is that RA serves as the sole target through which the AFP can directly influence song motor commands.

Song motor codes in the SMP

The pivotal location of HVC in the sensorimotor network and the destructive effects of HVC lesions on song motivate intense interest in understanding how HVC neurons function during singing and song learning. Pioneering electrophysiological recordings in the HVC of singing birds detected activity time-locked to individual syllables that increased prior to sound production and persisted in deafened birds, hinting at a motor rather than an auditory nature (McCasland and Konishi 1981). Furthermore, singing-related activity appears to propagate down the SMP, arising earlier in HVC than in RA (McCasland 1987). Subsequent studies found that focal microstimulation in HVC or RA transiently disrupts song, consistent with the idea that singing-related
activity in the SMP conveys motor commands (Vu et al. 1994; Ashmore et al. 2005). Moreover, targeted laser ablation of HVC<sub>RA</sub> but not HVC<sub>X</sub> neurons, severely degrades song structure (Scharff et al. 2000), indicating that song motor commands emanate from HVC directly to RA, and not through the AFP. Although experiments such as these show that the SMP conveys song motor signals, they cannot provide insight into the underlying motor codes.

The most direct way to decipher the song motor code is to record the singing-related activity of identified neurons at various levels in the SMP. Though this may seem straightforward, recording single unit activity in singing birds can be challenging, because singing involves powerful vocal and respiratory movements accompanied by high-amplitude head vibrations. In retrospect, a further challenge to recording the singing-related activity of individual HVC<sub>RA</sub> neurons is that they rarely fire action potentials except when the bird is singing (Yu and Margoliash 1996; Hahnloser et al. 2002). Unfortunately, recordings from interneurons cannot detect the pattern of activity transmitted out of HVC to affect song.

Michale Fee and his co-workers overcame these challenges by using a miniature motorized microdrive and antidromic stimulation methods to record from identified HVC<sub>RA</sub> neurons (Hahnloser et al. 2002). What they found was quite remarkable: Single HVC<sub>RA</sub> neurons fire only one very brief (~10 msec) burst of action potentials at a single time during the entire ~1 sec motif, with different neurons bursting at different times in the motif. These extremely sparse and precise patterns of activity suggest that the entire ensemble of HVC<sub>RA</sub> neurons could function to specify the timing of syllables, notes, and even the intervening silent "gaps." Indeed, some HVC<sub>RA</sub> and RA neurons burst during these silences, consistent with this idea.

If the output of the HVC<sub>RA</sub> ensemble provides a timing signal for song, what and where is the mechanism that determines tempo? One possibility is that bursting activity propagates through a chain of synaptically connected HVC<sub>RA</sub> neurons, much like a series of falling dominoes, creating a timing signal that spans the entire motif. In this case, song tempo will be determined by the biophysical properties of the HVC local network. Alternatively, HVC<sub>RA</sub> neurons may burst in a serial fashion because they read out a timing signal generated outside of HVC. Long and Fee reasoned that if song tempo is determined by the activity of the HVC local network, then song should slow as HVC is cooled (Long
and Fee 2008). Using a small Peltier device to focally manipulate HVC temperature in singing birds they discovered that song tempo slowed, on average, almost 3% per degree Celsius that HVC was cooled. Moreover, song tempo slowed similarly across all timescales, ranging from individual notes (~10 msec) to the entire motif (~1 sec), including the silent gaps between syllables. In contrast, cooling HVC resulted in only small changes to the song's acoustic structure (such as pitch or amplitude) while, surprisingly, cooling RA exerted only minimal song effects of any kind. These findings provide direct evidence that a substantial amount of song timing information is generated by the HVC local network. Nonetheless, these experiments do not exclude a role for upstream structures in contributing to or coordinating timing signals for song. Indeed, pathways from brainstem SMP components to HVC (Fig. 4B) are likely to play an important role in initiating or terminating syllable sequences, because electrical stimulation in the brainstem can cause truncation and restarting of the motif (Ashmore et al. 2005), and because the durations of syllable gaps vary more than the durations of syllables (Glaze and Troyer 2006).

An important remaining question is the nature of the mechanism in HVC that sets song tempo. A leading candidate is a "synfire" chain comprising HVCRA neurons linked together by their local excitatory connections, an idea incorporated into modeling studies of central song patterning networks (Li and Greenside 2006; Jin et al. 2007). In fact, the axons of HVCRA neurons extend local collaterals before exiting HVC, forming excitatory synapses with other HVCRA cells, as well as interneurons and HVCX cells, providing a potential substrate for a synfire chain (Mooney 2000; Mooney and Prather 2005). Moreover, isolated in vitro preparations of HVC can be induced to generate rhythmic synaptic activity in response to high frequency electrical stimulation, suggesting that the HVC network can generate and propagate patterned activity (Solis and Perkel 2005). In the future, a better understanding of how song timing information is generated will rest on identifying how action potential bursts of HVCRA neurons are generated and how they propagate from one cell to the next through the local synaptic network.

Chronic recordings in singing birds also reveal that the sparse timing signal emitted by HVC is transformed within RA into a more continuous pattern of activity, because single RA PNs burst at many (~10) precise times during a motif (Yu and Margoliash 1996; Leonardo and Fee 2005). This physiological transformation suggests that an array of HVCRA neurons that burst at different times in the motif provide convergent input to individual RA neurons and that single HVCRA neurons make divergent connections with many different RA neurons; both of these features are consistent with the known connectivity (Kittelberger and Mooney 1999, 2005). A crucial consequence of this network arrangement for sensorimotor learning is that the song's temporal features are largely specified by the pattern of synapses that HVCRA axons make in RA. In fact, modeling simulations show that temporally sparse activity in HVC acting through such a network optimizes learning rates, because error signals that "correct" the connections in an HVCRA neuron only introduce changes at a single time in the motif (Fiete et al. 2004). However, this network view minimizes the role of local connectivity between RA neurons in contributing to or modifying song motor commands from HVC, which seems unlikely given that RA interneurons and projection neurons are highly interconnected (Sapiro et al. 1999; Sizemore and Perkel 2008).

What function might RA serve in generating or transforming song motor commands emanating from HVC? One possibility is that an explicit timing signal emanating from HVC is transformed within RA into a motor signal more specifically correlated with acoustical features of the bird's song. However, the mean population activity of small ensembles of RA neurons and mean spectral features of song are on average uncorrelated, indicating that different ensembles of RA neurons may contribute to similar sounds at different times in the motif (Leonardo and Fee 2005). Nonetheless, a more recent analysis found that trial-to-trial variability in individual RA neurons could predict a significant portion of trial-to-trial acoustic variability in the bird's song (Sober et al. 2008). This is remarkable, given that the "jitter" in an RA neuron's spike timing across motif renditions is in the millisecond range (Chi and Margoliash 2001). Although this correlation does not necessarily imply a causal influence of single RA neurons on song, factors that drive variations in the firing patterns of ensembles of RA neurons could drive variable performance, a major attribute of sensorimotor learning.

The AFP, an engine for song variability

Numerous lines of evidence point to the AFP as the source of acute song variability. In juvenile zebra finches engaged in sensorimotor learning, reversible inactivation of LMAN sharply reduces the trial-to-trial variability of their plastic songs (Olveczky et al. 2005). In adult male zebra finches, singing-related activity in the AFP and acoustic variability are higher during undirected song (Sossinka and Bohner 1980; Jarvis et al. 1998; Hessler and Doupe 1999b), and these context-dependent changes in song variability are abolished by LMAN lesions (Kao and Brainard 2006). This effect results specifically from reduced variability of undirected songs, underscoring that the AFP acts to inject variability into undirected songs, rather than to reduce the variability of directed songs (Kao and Brainard 2006). Notably, LMAN lesions do not abolish other context-dependent aspects of the male's courtship display, such as "dancing," head bowing, and beak wiping, illustrating that the male can still faithfully interpret social cues provided by the female, while also emphasizing the AFP's highly specialized role in vocalization (Kao and Brainard 2006).

Multiple lines of evidence indicate that LMAN neurons actively drive song variability through the fast excitatory synaptic connections they make with RA neurons. First, chronic recordings made in the LMAN of freely behaving songbirds reveal that neuronal activity increases during singing (Hessler and Doupe 1999a), with individual neurons firing in bursts roughly time-locked with song features (Leonardo 2004; Kao et al. 2008). This singing-related activity persists following deafening, suggestive of a motor origin, for which HVC is the most likely source (Hessler and Doupe 1999a; Kozhevnikov and Fee 2007; Prather et al. 2008). Second, singing-related activity in LMAN decreases with social context (Jarvis et al. 1998), with LMAN neurons displaying more variable spike timing and more frequent bursting during undirected singing (Hessler and Doupe 1999b; Kao et al. 2008). Spike timing variability in LMAN is even higher in neurons in juvenile finches engaged in sensorimotor learning, raising the possibility that developmental changes in LMAN firing properties could underlie developmental changes in acoustic variability (Olveczky et al. 2005). Together, these results suggest that variable LMAN firing patterns, perhaps especially variable bursting activity of LMAN neurons, drive song variability.

If natural variations in LMAN activity drive song variability, then artificially modulating LMAN activity during singing should affect song output. Consistent with this prediction, syllable-triggered microstimulation in LMAN can alter song structure, with vocal effects becoming detectable as early as 30 msec after stimulus onset and terminating ~65 msec after stimulus offset (Kao et al. 2005). These rapid effects strongly suggest that LMAN neurons drive song variability through the fast excitatory synapses they make in RA. In fact, LMAN and HVC axon terminals form adjacent synapses on the same dendrites of RA projections neurons, providing a discrete cellular locus where signals from the AFP could
modulate descending motor signals from HVC (Canady et al. 1988; Mooney and Konishi 1991; Mooney 1992). The functional properties of the synapses that LMAN and HVC axons make on RA neurons also provide clues as to how such modulation could be achieved: Near resting potential, LMAN synapses excite RA neurons mostly through NMDA receptors, whereas HVC synapses excite RA neurons predominantly through AMPA receptors (Kubota and Saito 1991; Mooney and Konishi 1991; Mooney 1992; Stark and Perkel 1999). When HVC<sub>RA</sub> and LMAN neurons fire together during singing, voltage-dependent blockade of NMDA receptors could be relieved, permitting LMAN inputs to strongly depolarize RA neurons. This arrangement may enable variably firing LMAN neurons to add “noise” in a conditional manner to the motor signals from HVC, causing the singing-related firing patterns of RA neurons—and thus song—to become more variable. Indeed, pharmacologically silencing LMAN causes an immediate decrease in the variability of RA spike times and blocking NMDA receptors in RA triggers an immediate decline in the juvenile’s bout-to-bout song variability, consistent with this idea (Olveczky et al. 2005, 2006).

Developmental regulation of song variability

The variable singing-related activity of LMAN neurons and the functional heterogeneity of HVC and LMAN synapses on RA neurons could contribute to the developmental regulation of song variability. One idea is that the bursting activity of LMAN neurons, which becomes less pronounced with age (Kao and Brainard 2006), may be especially effective in driving postsynaptic responses in RA, perhaps because bursting increases transmitter release or facilitates temporal summation. Nevertheless, context-dependent changes in song variability disappear entirely in older (>4 yr) male zebra finches, even though their LMAN neurons continue to exhibit context-dependent changes in bursting activity (Kao and Brainard 2006). This divergence in levels of acoustic and neuronal variability hints that developmental changes to LMAN > RA synapses may be especially important in regulating song variability.

In fact, NMDA-EPSCs evoked by LMAN > RA synapses decline in duration over sensorimotor learning (Stark and Perkel 1999; White et al. 1999; Livingston et al. 2000), and shorter currents could be less effective at injecting noise into the RA spike pattern. Decreased efficacy of the LMAN > RA synapse also could be achieved by hyperpolarization of RA neurons, because NMDA receptor-mediated currents are sensitive to voltage-dependent blockade. When the candidate hyperpolarizing agent is noradrenaline (NE), which can suppress spontaneous firing of RA neurons by increasing resting potassium conductance (Solis and Perkel 2006); interestingly, NE also can suppress LMAN input to RA presynaptically (Sizemore and Perkel 2008). Therefore, developmental increases in NE or other factors that diminish the functional efficacy of LMAN synapses in RA, or lead to their elimination, could reduce song variability.

Hence, a prevailing view is that ontogenetic changes in the relative efficacy of HVC and LMAN synapses in RA account for the transition from plastic to crystallized song. During plastic song, interactions between more equally weighted HVC and LMAN synapses on RA neurons drive variable song output. As the HVC inputs onto individual RA neurons strengthen and those from LMAN weaken, vocal sequences would become increasingly stereotyped. When HVC synapses reach maximum efficacy relative to LMAN synapses, song variability reaches the lower bound manifested in crystallized song. This maximum may reflect the point at which HVC inputs onto an RA neuron are sufficiently potent to drive it at its maximum firing rate.

A remaining issue is how these two synaptic inputs interact during subsong, the earliest stage of sensorimotor learning. Interestingly, when bilateral HVC lesions are made in very young (<45 d) zebra finches, they still produce vocalizations that resemble subsong, and these persistent vocalizations can be blocked by inactivating either RA or LMAN (Aronov et al. 2008; see also Nottebohm et al. 1976; Simpson and Vicario 1990 for an account of vocalizations produced following HVC lesions). Apparently, the juvenile bird’s earliest attempts to sing reflect RA activity under the influence of LMAN, a view further supported by the observation that some LMAN neurons fire action potential bursts preceding subsong utterances (Aronov et al. 2008). Thus, the transition from subsong to plastic song reflects a functional switch wherein singing-related activity in RA neurons is initially driven solely by LMAN and subsequently driven by HVC and LMAN. Notably, this functional switch closely parallels the staggered anatomical development of these two inputs, with LMAN axons innervating RA prior to subsong, and HVC axons densely innervating RA only near the transition to plastic song (Herrmann and Arnold 1991; Akutagawa and Konishi 1994; Mooney and Rao 1994). Viewed broadly, then, the several stages of sensorimotor learning appear to reflect reciprocal shifts in the weights of LMAN and HVC synapses on RA neurons.

The AFP may provide an instructive signal important to adaptive song plasticity

Beyond its role in generating high levels of bout-to-bout variability, the AFP is also essential to slower changes to song (i.e., song plasticity) in juvenile and adult birds. Following LMAN lesions, the songs of juvenile zebra finches exhibit abnormally high levels of stereotypy and change little or not at all, as if they had undergone premature crystallization (Bottjer et al. 1984; Scharff and Nottebohm 1991). Furthermore, LMAN lesions in adult zebra finches prevent song deterioration normally triggered by deafening (Brainard and Doupe 2000) or by other forms of vocal plasticity that can slowly arise following damage to the vocal nerve (Williams and Mehta 1999). More broadly, the finding that LMAN lesions prevent various forms of juvenile and adult song plasticity, including those triggered by disrupting auditory feedback, is consistent with the idea that LMAN transmits an instructive signal to RA that drives adaptive vocal plasticity.

Of course, lesion experiments alone cannot distinguish whether LMAN provides an instructive signal or simply acts permissively to allow such vocal plasticity. Notably, Kao et al. (2005) found that stimulation at a single locus in LMAN tended to alter the target syllable in a consistent manner, for example, increasing its pitch or amplitude, rather than triggering degradation or increased variability. These systematic effects suggest that, in addition to driving song variability, LMAN activity could provide instructive signals that systematically bias vocal performance toward a certain goal. Another clue that LMAN transmits such biased instructive signals during singing is that the temporal pattern of fluctuations in the mean firing rates of LMAN neurons is similar during directed and undirected song (Hessler and Doupe 1999a; Kao et al. 2008).

More recent evidence also suggests that LMAN provides instructive information that acts to bias song away from vocal errors (Andalman and Fee 2009). As described previously, a finch exposed to noise bursts whenever it sings higher pitch variants of a target syllable will slowly (approximately hours) shift the syllable’s pitch downward, subsequently reducing the probability that singing will trigger noise playback (Tumer and Brainard 2007). The initial expression of this adaptive shift depends on the AFP, because pharmacologically silencing LMAN causes the target syllable’s pitch to immediately revert toward its original value (Andalman and Fee 2009). Although these findings cannot rule out that LMAN functions permissively to gate a bias signal arising from another source (i.e., HVC), LMAN neurons exhibit
song premotor activity and electrical stimulation in LMAN during singing can alter the targeted syllable in a consistent way, suggesting LMAN is the source of an instructive signal. One caveat is that it is unknown whether the mechanisms that enable adaptive pitch shifts in the late juvenile and adult birds studied by Andalman and Fee (2009) (as well as by Tumer and Brainard 2007) are the same as those that enable juvenile sensorimotor learning. Therefore, an important goal will be to determine whether LMAN normally functions during sensorimotor learning to bias song in such a way to minimize differences between auditory feedback and the memorized tutor song.

The finding that LMAN lesions abolish both acute variability and slower forms of song plasticity also strengthens the idea that these two processes are causally related. Specifically, by driving acute song variability, variably bursting LMAN neurons might enable small vocal changes that could accumulate over days and weeks, resulting in more substantial and persistent adaptive alterations in song structure. Though this idea is both tantalizing and probable, the link between song variability and adaptive plasticity awaits experimental elucidation. As a thought experiment, at least, one could imagine examining whether birds can adaptively modify perfectly stereotyped syllables in response to feedback perturbation.

Although an explicit connection between song variability and adaptive vocal plasticity is yet to be established, current evidence provides some insights into how these processes could be linked at a synaptic or even molecular level. One possibility is NMDA receptors at LMAN–RA synapses operate over fast timescales (i.e., hundreds of milliseconds) to drive variability and over longer timescales (i.e., hours to days) to enable long-lasting forms of synaptic plasticity that are ultimately manifested as persistent changes to song. Additionally or alternately, such long-lasting changes could reflect a trophic influence of LMAN on RA, mediated most likely by brain-derived neuronotrophic factor (BDNF) (Johnson et al. 1997; Li et al. 2000; Kittelberger and Mooney 2005). A mark of this influence is that LMAN lesions made during sensorimotor learning (but not those made following crystallization) trigger relatively rapid (~24 h) consolidation of the synapses that HVC axons make with RA neurons (Kittelberger and Mooney 1999). Intriguingly, this rapid consolidation resembles the synaptic consolidation that occurs more slowly over the course of sensorimotor learning, suggesting that changes in the efficacy of HVC to RA synapses could play an important role in sensorimotor learning. An important future direction will be to explore whether LMAN axons on RA neurons can sustain hetero- and/or homosynaptic forms of plasticity likely to underlie learning-related changes in song motor commands.

Identifying the source of song variability

Illuminating how song variability is controlled centrally during sensorimotor learning requires identifying where the variability signal first originates. A first step in answering this question is to determine where context-dependent changes in singing-related activity first emerge in the song system. Although these comparisons have not yet been made in HVC, HVCX cells show an extremely high degree of temporal precision in their singing-related activity (Kozhevnikov and Fee 2007; Prather et al. 2008), whereas area X neurons, like those in LMAN, exhibit context-dependent changes in their singing-related activity (Hessler and Doupe 1999a). These observations suggest that more variable firing patterns of singing-related activity first arise within the AFP. However, whether these changes arise first in area X or originate later in the AFP, only to be fed back into area X from LMAN, remains uncertain. This could be resolved by testing whether inactivating LMAN abolishes context-dependent changes in area X activity.

The current view that context-dependent changes in neuronal activity arise in the AFP motivates strong interest in identifying factors that modulate AFP activity. One leading candidate is dopamine, because both area X and LMAN receive dopaminergic and/or noradrenergic input from midbrain VTA neurons (Lewis et al. 1981; Bottjer et al. 1989; Soha et al. 1996; Gale and Perkel 2006), and dopamine can modulate the intrinsic and synaptic properties of area X neurons (Ding and Perkel 2002, 2004; Ding et al. 2003). Furthermore, VTA neurons exhibit singing-related activity, and levels of this activity, along with dopamine levels in area X, are higher in male zebra finches when they sing directed songs (Sasaki et al. 2006; Yanagihara and Hessler 2006). These findings hint that dopamine could serve as an important regulator of AFP activity and song variability.

The potential involvement of dopamine in song performance is also tantalizing because dopamine signaling has been implicated in certain forms of reinforcement learning. Recordings made in monkeys previously trained in an operant task detect some dopaminergic VTA neurons that encode discrepancies between a predicted outcome, such as a juice reward normally delivered following a conditioned stimulus, and the reward’s actual occurrence (Schultz 1986; Romo and Schultz 1990; Schultz et al. 1993). The ability of a subset of VTA neurons to encode prediction errors, as well their ability to predict accurately when the expected outcome should occur, are essential features of error signals in models of reinforcement learning (Schultz et al. 1997; Schultz 2002).

In the songbird’s brain, a predicted outcome could be computed by a motor-based signal trained through experience to estimate associated auditory feedback (Troyer and Doupe 2000), while the actual outcome would be conveyed by the real feedback signal. Prediction errors could be generated by neurons that detect when these two signals differ (Fig. 3B), as might occur when the bird makes a vocal “mistake” or experiences distorted feedback. If certain songbird VTA neurons behave like those studied in the primate to encode prediction errors, vocal errors or distorted feedback would be expected to modulate their singing-related activity. This change in VTA activity could, in turn, alter dopamine release in the AFP to affect song output, possibly by directly or indirectly affecting levels of song variability. This scenario is highly speculative, however, and dopamine in the songbird could serve other functions exclusive of any role in reinforcement learning. Therefore, experimentalists will need to resolve several important issues before a link between dopamine and reinforcement learning can be made for songbirds.

The first issue to resolve is the cause and effect relationship between activity in the VTA and the AFP. A potential confound is that the songbird VTA receives indirect input from area X, leaving unclear whether context-dependent changes in VTA activity are a cause or consequence of changes in AFP activity (Gale et al. 2008). Reversible inactivation of the VTA, as well as manipulations that interfere with DA signaling in the AFP, could help determine whether activity in dopaminergic VTA neurons triggers context-dependent changes in activity within the AFP.

A second issue is whether dopamine signaling in the songbird involves the fast signals associated with reward prediction error in the primate. In vivo voltammetry measurements targeted to the AFP may be necessary to answer this question, because microdialysis methods are too slow to resolve such fast signals and chronic recordings of VTA neurons cannot assess the time course of dopamine release and reuptake in the AFP. However, even phasic VTA neurons in the primate may respond too slowly and with too much variability to signal reward prediction errors on the millisecond timescale required for song learning (Fiorillo et al. 2008).

A third issue is whether dopamine can act quickly enough to account for rapid context-dependent changes in song and with
sufficient precision to account for temporally specific effects of DAF on song variability. More generally, the slow time course of neuromodulatory signaling seems poorly suited to precisely regulate motor variability on either a trial-by-trial or syllable-by-syllable basis. Perhaps dopamine acts only slowly to reinforce certain motor patterns, perhaps by regulating synaptic plasticity in the AFP (Ding and Perkel 2004).

Finally, assuming a role for the VTA in modulating AFP activity, what types of signals impinge on VTA neurons to affect their activity? In particular, it will be important to identify whether songbird VTA neurons have access to auditory feedback signals, motor-based predictive signals of auditory feedback, and/or resultant prediction errors. Answers to these and other related questions will provide greater insight into how central mechanisms for song variability are dynamically regulated by experience to affect sensorimotor learning.

The search for central representations of auditory feedback

The importance of the AFP to song variability and plasticity raises the exciting possibility that it functions as part of a comparator circuit. Indeed, the ability of LMAN lesions to prevent deafening-induced song deterioration in adult zebra finches is consistent with the idea that LMAN functions as the output of a comparator circuit. In this light, an important goal is to identify where auditory feedback signals are thought to propagate from HVC to the AFP, HVCX neurons are the appropriate cells to target in such tests. Nonetheless, chronic recordings reveal that DAF fails to alter the singing-related activity of HVCX cells either in juvenile zebra finches engaged in sensorimotor learning, or in adult Song variability and plasticity raises the question of the capacity to respond to auditory stimulation in nonsinging states, and their ability to acquire selectivity for spectrally distorted songs prior to the expression of vocal plasticity. Instead, singing-related activity in LMAN appears to be driven by motor signals emanating from HVC, at least in birds producing plastic or crystallized songs (Kozhevnikov and Fee 2007; Prather et al. 2008).

The study by Leonardo (2004) intentionally used adult zebra finches singing directed songs to provide the most stable motor background on which to detect auditory feedback signals. These are also the conditions in which adult zebra finches are most insensitive to feedback perturbations, with many days or even weeks of chronic exposure to DAF required to trigger vocal plasticity (Leonardo and Konishi 1999). Additionally, LMAN is the output stage of the AFP, leaving open the possibility that real-time feedback signals are transmitted into the AFP and then transformed in a way that only exerts effects on LMAN activity after a substantial delay. Thus, a necessary additional step is to test whether perturbations of feedback affect singing-related activity of inputs to the AFP during sensorimotor learning, when those perturbations most rapidly modify song. Because song motor and auditory activity are thought to propagate from HVC to the AFP, HVCX neurons are the appropriate cells to target in such a test. Nonetheless, chronic recordings reveal that DAF fails to alter the singing-related activity of HVCX cells either in juvenile zebra finches engaged in sensorimotor learning or in adult Soci finches, both of which rely acutely on auditory feedback (Kozhevnikov and Fee 2007; Prather et al. 2008). Therefore, neither the input nor output of the AFP appears to have access to a real-time auditory feedback signal.

The absence of singing-related feedback signals in the AFP does not rule out a role for this pathway in error detection, however. As previously discussed, error detection could be accomplished, in part, by estimating the auditory feedback predicted to accompany the motor signal. In its most essential form, a neuron conveying such a prediction should display a motor-derived corollary discharge signal that bears a systematic correspondence to the sensory signal evoked by the associated vocalization. A recent study provides evidence that HVCX cells transmit such a corollary discharge signal to the AFP (Prather et al. 2008). Using a lightweight microdrive (Fee and Leonardo 2001), Prather et al. recorded from antidromically identified HVC cells in the swamp sparrow as it sang and listened to its own songs played through a speaker. When the sparrow was passively listening, an HVCX cell displayed almost identical patterns of activity when the bird sang the same song type, firing at exactly the same time in the effective syllable. Interestingly, these shifts in selectivity can be detected before chronic exposure to distorted feedback has triggered vocal plasticity. Therefore, the attractive idea is that the AFP has access to auditory feedback information, enabling LMAN to modulate song motor commands in an instructive fashion.

Although this idea is attractive, the detection of song-selective auditory activity in the AFP of nonsinging birds does not establish whether AFP neurons respond in real-time to singing-related auditory feedback. One indication that LMAN neurons in the adult zebra finch do not act as real-time sensors of feedback is that the AFP, HVCX neurons are the appropriate cells to target in such tests. Nonetheless, chronic recordings reveal that DAF fails to alter the singing-related activity of HVCX cells either in juvenile zebra finches engaged in sensorimotor learning or in adult Soci finches, both of which rely acutely on auditory feedback (Kozhevnikov and Fee 2007; Prather et al. 2008). Therefore, neither the input nor output of the AFP appears to have access to a real-time auditory feedback signal.

The absence of singing-related feedback signals in the AFP does not rule out a role for this pathway in error detection, however. As previously discussed, error detection could be accomplished, in part, by estimating the auditory feedback predicted to accompany the motor signal. In its most essential form, a neuron conveying such a prediction should display a motor-derived corollary discharge signal that bears a systematic correspondence to the sensory signal evoked by the associated vocalization. A recent study provides evidence that HVCX cells transmit such a corollary discharge signal to the AFP (Prather et al. 2008). Using a lightweight microdrive (Fee and Leonardo 2001), Prather et al. recorded from antidromically identified HVC cells in the swamp sparrow as it sang and listened to its own songs played through a speaker. When the sparrow was passively listening, an HVCX cell displayed almost identical patterns of activity when the bird sang the same song type, firing at exactly the same time in the effective syllable. These neurons displayed almost identical patterns of activity when the bird sang the same song type, firing at exactly the same time in the effective syllable as when the bird was quietly listening to the song played through a speaker. However, this singing-related activity was unaffected by acutely disrupting auditory feedback, pointing to a motor origin, the likely source being the HVCRA cells that directly and indirectly connect to HVCX cells through HVCs local synaptic network (Mooney and Prather 2005). Evidently, HVCX neurons receive a corollary discharge of the song motor signal that closely resembles the auditory signal evoked by the associated vocalization.

These observations support a model in which the singing-related activity of HVCX cells provides a motor-based prediction of auditory feedback, which could be compared to the actual feedback signal to compute an error signal (Fig. 3B). If this model obtains in the songbird, and auditory feedback insensitivity is a general characteristic of the entire HVCX cell population, the
search for comparator neurons can be narrowed to efferent targets of HVCx cells, which include the AFP and other HVC neurons. Because real-time feedback signals have not been detected in the AFP, the search for a comparator may be further narrowed to HVCx cells and HVC interneurons, both of which receive excitatory synaptic input from HVCx cells (Mooney and Prather 2005).

Another possibility is that HVCx cells are comparators, but the corollary discharge signal they receive is so strong that auditory feedback does not acutely alter their firing patterns, and instead is only detectable at a subthreshold synaptic level. This scenario seems plausible in adult swamp sparrows, which are likely to resemble other sparrow species that can maintain their crystallized songs without acute auditory feedback (Konishi 1965b). However, the singing-related activity of sensorimotor HVCx neurons in adult Society finches was also found to be insensitive to DAF (Prather et al. 2008), even though their songs change rapidly (over days) in response to auditory feedback perturbations (Okanoya and Yamaguchi 1997; Woolley and Rubel 1997). These and other observations sustain the notion that HVCx cells are gated to exist in purely motor or sensory states, and do not transmit a “blended” motor and auditory feedback signal during singing. Actually, several gating mechanisms have been advanced to explain arousal-mediated suppression of HVC auditory responses seen in sedated or anesthetized zebra finches (Shea and Margoliash 2003; Cardin and Schmidt 2004; Coleman et al. 2007), one or more of which could account for the motor suppression of HVCx cell auditory responses seen in the sparrow and Society finch. Ultimately, even though the prevailing view is that HVCx neurons (as well those in the AFP) do not convey auditory feedback signals, one important step will be to fully characterize the effects of DAF on singing-related HVCx and AFP cell activity at very early stages of sensorimotor learning, when such signals would be expected to be most prominent. Another much more challenging step will be to record intracranially from HVCx cells in singing birds to determine whether they display synaptic responses to feedback perturbations not reflected in their firing patterns.

Because HVC is thought to be a bottleneck through which auditory information reaches the AFP (Bauer et al. 2008; Roy and Mooney 2009), and because HVCx neurons appear to lack auditory responses altogether in the awake bird (Prather et al. 2008), neurons whose action potential activity is sensitive to real-time feedback perturbations must be presynaptic to HVCx cells. Two recent studies provide evidence that neurons directly and indirectly presynaptic to HVCx cells can change their spiking patterns in response to real-time feedback perturbations (Sakata and Brainard 2008; Keller and Hahnloser 2009). One recent study reported that DAF could slightly and transiently suppress singing-related neural activity in HVC of adult Society finches (Sakata and Brainard 2008). This suppression reached a peak ~20 msec after noise onset, similar to the auditory response latencies for HVC neurons. Multiunit recordings in HVC, as used in this study, are thought to represent interneuron activity, which may reconcile these positive findings with the negative effects of DAF on singing-related activity of identified HVCx cells (Kozhevnikov and Fec 2007; Prather et al. 2008). A potential confound to this study is that DAF applied to adult Society finches can trigger immediate changes in the vocal pattern, including a slight slowing of song tempo (Sakata and Brainard 2006), and these changes may be unrelated to those that occur during sensorimotor learning. Nonetheless, short latency suppression could be detected in HVC in the absence of overt song effects, while additional longer latency suppression of HVC activity occurred in those DAF trials with overt motor effects. Ultimately, these findings may provide the first clue that HVC receives real-time feedback signals, even if feedback perturbations only affect song after hours, days, or even weeks.

The possible localization of a real-time auditory feedback signal to HVC interneurons has several potential implications for how feedback could affect song. First, some HVCx cells receive inhibitory synaptic input from interneurons and excitatory input from HVCx cells (Mooney and Prather 2005), an arrangement that could enable HVCx cells to compare a real-time feedback signal with motor-related signals. This interaction may be especially important for “training” HVCx cells to generate a motor estimate of auditory feedback (Troyer and Doupe 2000), or for enabling HVCx cells to function as comparators. Alternatively or additionally, interneurons may function to compare real and predicted feedback signals, because they receive excitatory input from both extrinsic auditory sources and from HVCx cells (Coleman and Mooney 2004; Mooney and Prather 2005; Rosen and Mooney 2006). Finally, some HVC interneurons make synaptic connections onto HVCRA cells, providing a route via which feedback could directly influence song motor commands (Mooney and Prather 2005).

The detection of a putative feedback signal in HVC also raises the question of its source. Regions of the auditory telencephalon provide the major source of auditory information to HVC (Fortune and Margoliash 1995; Yates et al. 1996a; Bauer et al. 2008), making these regions key sources of a real-time feedback signal (Fig. 4B). Actually, a recent study detected neurons sensitive to real-time feedback perturbations in Field L and CLM (Keller and Hahnloser 2009), the primary and secondary auditory regions in the avian telencephalon (Field L and CLM are analogous to mammalian primary and secondary auditory cortices, respectively). The majority of Field L and CLM neurons showed similar patterns of activity whether the bird was singing or listening to its song played through a nearby speaker, and their singing-related activity could be disrupted by DAF. A second cell type also showed similar patterns of activity whether the bird was singing or listening to its song, but its singing-related activity was unaffected by DAF. The third and most rarely encountered cell type was not very active during either singing or listening, but was active when singing-related feedback was disrupted. The discovery of these three distinct cell types raises the possibility that auditory telencephalic neurons do more than simply encode feedback; they may combine a feedback representation with a motor-based predictive signal to generate signals for error correction. Three major lines of inquiry could help to make this determination. The first is to identify whether these three cell types are interconnected with each other in a manner that enables computation of the putative error signal and whether the third cell type provides information to the song system, which could allow its activity to affect song motor commands. The second should seek the nature and source of the putative predictive feedback signal, which presumably depends on input from the song system. The third line of inquiry will need to determine whether the activity of these various auditory telencephalic neuron types is actually harnessed for sensorimotor learning.

Although DAF can be used to identify neurons acutely sensitive to feedback perturbations, a limitation to this method is that it cannot determine whether the resulting changes in activity are harnessed to shape song. For example, only some neurons acutely sensitive to singing-related auditory feedback may play a role in modifying vocalization. Ultimately, a necessary step will be to test whether the activity of central neurons acutely sensitive to feedback perturbations play a causal role in song plasticity. This causal link could be established by using microstimulation methods to test whether altering the singing-related activity of feedback-sensitive neurons affects song in a manner similar to peripheral disruption of auditory feedback.
In search of the template

Notwithstanding the potential importance of predictive motor signals to song learning and maintenance, juvenile songbirds learn to sing by comparing auditory feedback to a memorized model of a tutor song, or song template. The prevailing view is that the song template is an auditory memory, because a juvenile need only hear a tutor to subsequently imitate it (Marler and Peters 1981, 1982a; Adret 1993), and because vocal matching is accomplished largely, if not exclusively, through the use of auditory feedback (Konishi 1965a,b; Funabiki and Konishi 2003). Although the neural manifestation of the template is a matter of ongoing debate, recent research has shed a brighter light on how tutor song experience affects the juvenile songbird’s brain.

Scientists have taken three major experimental approaches to resolve how and where the memory of the tutor song is encoded in the brain. These include pharmacological inactivation of candidate brain areas during sensory learning (Basham et al. 1996; London and Clayton 2008), using immediate early gene (IEG) expression to map brain regions activated by playback of tutor song (Bolhuis et al. 2001; Bolhuis and Gahr 2006), and using electrophysiological methods to search for neurons that display selective auditory responses for the tutor song (Volman 1993; Solis and Doupe 1997, 1999; Solis et al. 2000; Theunissen et al. 2004). Though useful, these methods have not been able to convincingly localize the template to a specific brain region, or to confirm exactly how single neurons may contribute to encoding of tutor song memories.

One fundamental debate that still awaits resolution is whether the template resides within or outside the song system. Historically, the demonstrated importance of the AFP to sensorimotor learning, and the detection of song-selective neurons in this pathway, encouraged the view that this pathway may also be important to sensory learning. This view gained support from the finding that infusing NMDA receptor blockers into the LMAN of juvenile zebra finches just prior to tutoring sessions impaired copying (Basham et al. 1996). Although such treatment might be expected to interfere with initial attempts at imitation, birds tutored on alternate days, and treated with the blocker on “off” days, copied normally (Basham et al. 1996). More recent evidence implicating the AFP in either song memorization or matching comes from the finding that imitation is impaired when expression of the forkhead transcription factor FOXP2 in area X of juveniles zebra finches is suppressed during tutor exposure and sensorimotor learning (Haesler et al. 2007).

Motivated by the detection of BOS-selective neurons throughout the song system, a second approach has been to search for tutor song-selective neurons using electrophysiological recordings in anesthetized birds. In adult songbirds, neurons in LMAN (Doupe and Konishi 1991; Rosen and Mooney 2000), as well as other in other song nuclei (Margoliash 1983, 1986; Volman 1996), respond most strongly to playback of the BOS, and recordings made in juvenile birds reveal that these BOS-selective responses develop in parallel with sensorimotor learning (Volman 1993; Doupe 1997; Solis and Doupe 1997; Theunissen et al. 2004). These observations indicate that the expression of song-selective responses is not triggered simply by hearing the tutor song, but instead requires vocal rehearsal and auditory feedback. However, during sensorimotor learning, many LMAN neurons respond equally to the BOS and the tutor song, and a minority responds most vigorously to the tutor (Solis and Doupe 1997). Notably, tutor song-selective LMAN neurons can be detected even in juvenile birds in which the vocal nerve has been cut (Solis and Doupe 1999), a manipulation that makes BOS very different from tutor song. These results suggest that the some LMAN neurons may respond to the tutor song not simply because of its acoustic similarity to the BOS, but perhaps because they encode a representation of the tutor song.

Another possibility is that auditory neurons in the song system exclusively encode aspects of the bird’s vocal performance, rather than encoding a tutor song representation. Consistent with this idea, a majority of LMAN neurons in both juvenile and adult zebra finches subjected to vocal nerve injury exhibit selectivity for the spectrally distorted BOS, even though these distorted songs are very different from the tutor song (Solis and Doupe 1999; Roy and Mooney 2007). These results are hard to reconcile with a model in which a primary function of LMAN neurons is to encode a lasting memory of the tutor song. However, even if LMAN encodes tutor song memories, they can be overwritten: LMAN neurons in juvenile finches that sequentially copy two different tutors transiently acquire responses to the first tutor’s song, but these responses cannot be detected after LMAN neurons gain responses to the second tutor song (Yazaki-Sugiyama and Mooney 2004). More generally, an overwriting process could explain why song system neurons in adult birds overwhelmingly display selectivity for the BOS over the tutor song.

In the end, the failure to detect a robust tutor song representation in the song system has led to looking for such a representation in the auditory telencephalon, particularly the caudomedial nidopallium (NCM) (Bolhuis and Gahr 2006), which makes reciprocal connections with caudal mesopallium (CM) and thus provides indirect input to the song system (Fig. 48; Yates et al. 1996b). Both immediate early gene (IEG) expression patterns and electrophysiological recordings show that NCM neurons habituate to repeated playback of the same song (Chew et al. 1995; Mello et al. 1995), consistent with NCM being a site of auditory plasticity. Moreover, IEG expression levels in NCM correlate with how well the bird copied the tutor song (Bolhuis et al. 2000, 2001; Terpstra et al. 2004) and NCM neuronal firing rates in adult finches habituate more slowly to playback of the tutor song than to novel songs, with the slowest habituation rates for those tutor songs copied most accurately (Phan et al. 2006). One potential confound is that the most closely copied tutor songs are by definition also those most similar to the BOS, and thus it is unclear which song memory—self or tutor—is reflected in NCM activity. This issue could be resolved by using vocal nerve section to increase the acoustical “distance” between the BOS and the tutor song, as has been done in studies probing tutor song representations in the song system.

Two more recent findings lend additional support to a role for the auditory telencephalon in the formation of tutor song memories. First, reversibly blocking mitogen-activated protein kinase/ extracellular signal–regulated kinase (MEK) signaling in NCA in juvenile zebra finches during, but not after tutoring, has been found to impair subsequent copying (London and Clayton 2008). Although the electrophysiological consequences of such treatment were not well characterized, other studies in songbirds suggest that MEK signaling is important to inducing IEG expression in NCM (Cheng and Clayton 2004), and studies in other systems hint that MEK may play an important role in memory formation (Bozon et al. 2003; Sweatt 2004). Second, Keller and Hahnloser (2009) raise the interesting idea that singing-related activity of putative sensorimotor neurons in Field I and CLM could constitute “online” replay of the tutor song memory. In this view, similar patterns of activity during singing and BOS playback reflect a close match between the bird’s song and the template, while cells that respond only when feedback is perturbed could signal that the bird’s song diverged from the template, providing an error signal for imitation.

The debate is far from over, however, and even more recent evidence rekindles interest in the song system as a site where tutor song experience exerts rapid functional effects, consistent with
the formation of an auditory memory (Shank and Margoliash 2008). Chronic recordings of sleep-related activity in the RA of juvenile finches revealed a sharp increase in bursting activity the first night after their exposure to a tutor song. Intriguingly, bursting in RA increased prior to detectable changes in the entropy variance of the juvenile’s song, a feature that provides the earliest behavioral evidence of imitation (Tchernichovski et al. 2001). Furthermore, the exact pattern of bursting behavior in RA, as measured by interspike interval (ISI) distributions, was influenced by the specific tutor song that the juvenile heard, with more similar ISI distributions characterizing birds that heard the same tutor. However, the emergence of bursting activity in RA also required normal auditory feedback experience, because it could be delayed by using muting or noise-masking to prevent the bird from hearing its song. Therefore, although these effects emerged before demonstrable evidence of imitation, and the pattern of these effects correlated with the specific tutor song the bird had heard, the requirement of auditory feedback suggests they may reflect a rapid instructive interaction between an auditory memory of the tutor and the song motor network. These experiments also raise the possibility that this interaction is consolidated by spontaneous activity generated during sleep, although future experiments will need to verify how patterns of spontaneous and singing-related activity change in the song system during the first minutes of tutor song experience. Finally, although these effects were detected in RA, bursting activity in this song motor area is thought to be driven by inputs from HVC (Dave and Margoliash 2000), suggesting that the proximal site where auditory activity of the tutor affects the song system is in HVC. Much as emerging evidence suggests HVC is where auditory feedback integrates with song motor commands to affect vocal plasticity, HVC may also directly link auditory memories of the vocal model to the machinery of imitation.

Conclusions and future directions

Sensorimotor learning in songbirds is characterized by high levels of bout-to-bout variability and by slower learning-related changes in song. These observations have raised support for the idea that variability provides the stepping stones over which more substantial, gradual, and persistent learning-related changes can occur. An important goal will be to test whether acute song variability is necessary for slower, learning-related changes.

Important advances have been made in understanding the neural mechanisms for generating variability and acute song variation. Current evidence suggests that acute variations in the song pattern arise when a precise timing signal from HVC interacts with variable signals from LMAN. This fast synaptic interaction occurs in the song motor nucleus RA, where HVC and LMAN axons form convergent excitatory synapses on individual song premotor neurons. Beyond these fast synaptic interactions, an important goal will be to identify how these and other synapses in the song system enable the long-lasting forms of synaptic modification that are likely to underlie song motor learning.

Performance evaluation in song learning depends on auditory feedback. Although the central motor codes for song are in the process of being “cracked,” our understanding of central auditory feedback representations for learned vocalizations remains more rudimentary. Recent experiments have identified central neurons that are active during singing and sensitive to auditory feedback perturbations, indicating they encode a real-time representation of auditory feedback. An important next step will be to determine whether and how feedback-related changes in the activity of these neurons are harnessed for modifying vocalization.

Songbirds learn to sing by imitating other birds’ songs. Therefore, evaluation of song performance must involve a comparison of the performance-related feedback with an internal song model. This song model is initially formed by auditory experience of the tutor song. Recent results highlight how a juvenile songbird’s brain changes when a tutor song is heard. Other recent findings provide evidence of a motor-based signal that could function as an estimate of singing-related auditory feedback, reminiscent of the predictive signals that characterize reinforcement learning algorithms. An important future direction will be to determine whether and how these putative song models are combined with performance-related auditory feedback signals to affect vocal learning.

An attractive idea is that a neural comparator detects differences between song-related auditory feedback and the memorized tutor song, emitting a signal that adaptively modifies song to minimize this difference. The nature of the comparator and the output signal that it may generate remain poorly understood. One hint that LMAN may function as the output of the comparator is that it is necessary to both acute song variability and to song plasticity. More recent evidence also suggests that LMAN provides instructive signals that enable correction of vocal errors, at least in late juvenile and adult birds. Two important goals for future research are to establish whether LMAN functions during sensorimotor learning to correct vocal errors in reference to the tutor song model and to define how auditory evaluation of vocal performance ultimately influences this instructive process.

References


Woolley SM, Rubel EW. 1997. Bengalese finches...
Neural mechanisms for learned birdsong

Richard Mooney

Learn. Mem. 2009, 16:
Access the most recent version at doi:10.1101/lm.1065209

References

This article cites 170 articles, 86 of which can be accessed free at: http://learnmem.cshlp.org/content/16/11/655.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.