Chapter 3 : Decrystallization of adult birdsong by perturbation of auditory feedback

Young birds learn to sing by using auditory feedback to compare their own vocalizations to a memorized or innate song pattern; if they are deafened as juveniles they will not develop normal songs (Konishi, 1965; Nottebohm, 1968). The completion of song development is called crystallization. After this stage, song shows little variation in its temporal or spectral properties. However, the mechanisms underlying this stability are largely unknown. Here we present evidence that auditory feedback is actively used in adulthood to maintain the stability of song structure. We found that perturbing auditory feedback during singing in adult zebra finches caused a slow deterioration of song. This "decrystallization" consisted of a dramatic loss of the spectral and temporal stereotypy seen in crystallized song, and included stuttering, creation, deletion and distortion of song syllables. After normal feedback was restored, these deviations gradually disappeared and the original song was recovered. Thus, the brains of adult birds that do not learn new songs nonetheless retain a significant amount of plasticity.

3.1 Introduction

The song of zebra finches consists of three levels of organization: syllables, which are individual sound components of the song separated by silent intervals; motifs, which are sequences of syllables; and bouts, which are sequences of motifs (Sossinka and Bohner, 1980). The spectral structure of the syllables is unstable in the juvenile bird. Similarly, motifs and bouts are organized differently from song to song. However, as the song gradually assumes its adult form, these different levels of organization become highly stereotyped, the variability of the spectral structure of the syllables becomes extremely small, and the bird sings these syllables in a highly predictable order (Marler, 1970). At this stage the song is referred to as crystallized (Figure 3.1). Some birds, like the zebra finch, maintain

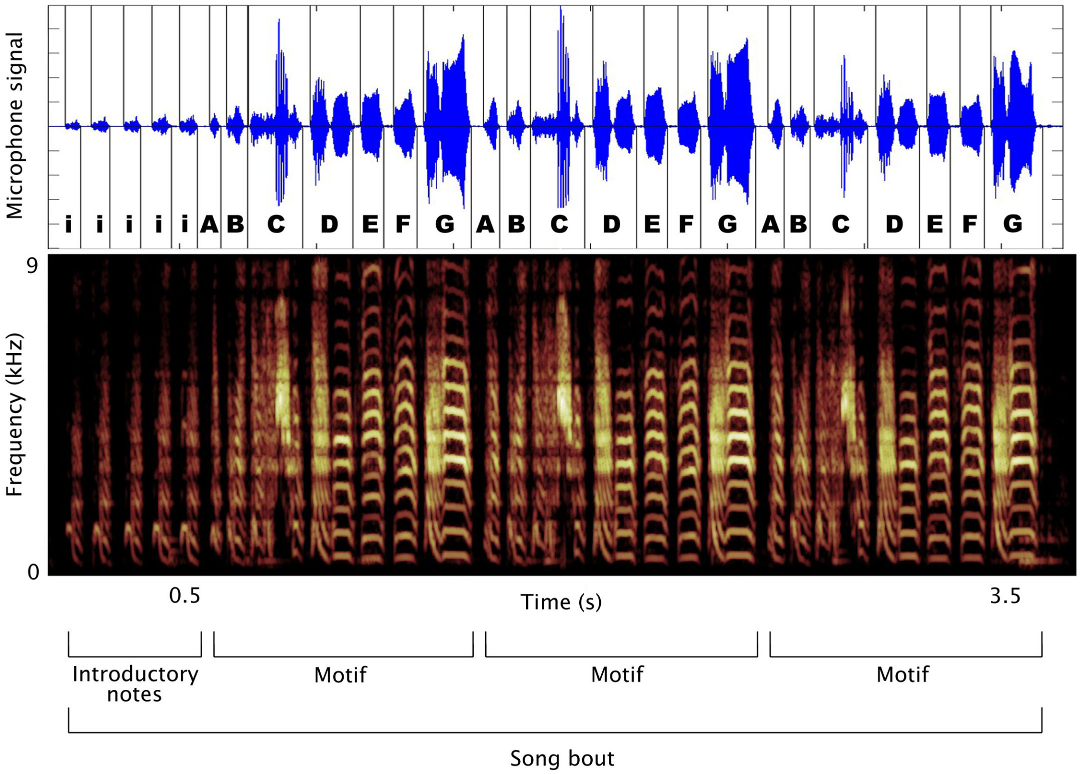


Figure 3.1. The normal structure of adult zebra finch song. Top trace shows the raw microphone signal, parsed into discrete bursts of sound (syllable). Bottom trace shows the time-frequency spectrogram of the song. Syllables are produced in a repeated sequence called a motif. During a bout of singing, the motif is repeated a variable number of times.

their crystallized song throughout adulthood and are called age-limited learners. In contrast, open-ended learners, like canaries, can learn new songs in adulthood (Marler and Peters, 1987).

The stability of song in age-limited learners was previously thought to be maintained without auditory feedback (Konishi, 1965; Nottebohm, 1968). Recent reports, however, show that deafening these birds after crystallization causes changes in song, suggesting that some auditory feedback is important for song maintenance throughout life (Nordeen and Nordeen, 1992; Okanoya and Yamaguchi, 1997; Wooley and Rubel, 1997). Six to eight weeks after deafening, adult zebra finch song undergoes a deterioration involving the addition and deletion of syllables, the abnormal repetition of syllables (stuttering), and modified syllable sequences. By opening the auditory feedback loop, deafening shows how well the song pattern generator can maintain its original output without this signal. However, to learn how the song control system works, it is necessary to manipulate auditory feedback without disabling either the auditory or vocal control system. The ability to modify feedback signals dynamically allows particular spectral and temporal components of the song to be chosen for manipulation. Furthermore, the effects of restoring normal auditory feedback after exposure to abnormal feedback can be observed. We show here how these manipulations cause the adult song to undergo dramatic changes.

We developed a computer-controlled system to perturb auditory feedback. We designed two feedback paradigms (see Figure 3.2). In both methods, the computer detected singing and then played back a feedback signal to the bird. Because the playback signals were delivered through an overhead speaker, the birds heard a superposition of natural and artificial sounds. In the first paradigm, the adaptive protocol (n = 3 birds), a computer alternated between recording vocalizations and playing back the last vocalization to the bird. The position of the feedback varied depending on the exact timing of the bird's song, and the feedback changed as the song changed (see Methods). In order to explore the effects of feedback on a more local level, we designed a second syllable-triggered protocol (n = 2 birds) in which we perturbed the feedback of only a single target syllable. The computer recognized and played back a stored copy of this syllable each time the bird produced it. The timing of the feedback signal was fairly constant across different song deliveries.

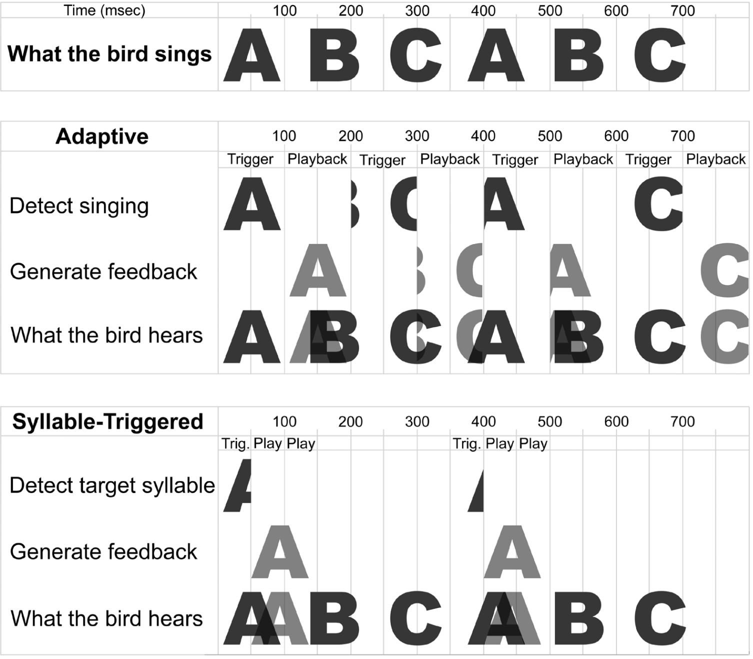


Figure 3.2. Protocols for constructing the feedback signals. In the adaptive protocol, the computer alternated between recording vocalizations and playing the last vocalization back to the bird. In the syllable-triggered protocol, the computer monitored the bird's song in 50 msec bins and detected the production of a single target syllable. In both protocols, the bird heard the superposition of his own vocalizations and the computer generated feedback.

3.2 Results

We recorded the songs of five adult male zebra finches in separate, sound attenuated chambers for several weeks before beginning the experiment. This procedure allowed us to ascertain that the variability of their songs was well within that found in crystallized songs. The birds were then placed in the feedback system. After a period of 1-4 months in this environment, four of the five birds showed dramatic changes from their original songs. To study the progression of these changes, we allowed some of the birds to sing without artificial feedback on a randomly chosen 10-15% of their song deliveries one day per week. All of the analyzed data consisted of recordings of the bird singing by himself, with no artificial feedback being produced at the time. After the feedback was permanently stopped, we tracked all the birds for another 8-16 months to determine if they were able to recover their original songs.

Decrystallization of the song occurred in both global song organization and local spectral structure, and consisted of the emergence of new spectral and temporal properties and the increased occurrence of properties that were rare in the baseline song. We use the term decrystallization to refer collectively to all of the perturbation-induced changes from the original quantitative and statistical structure of the song; this does not necessarily imply a return to a juvenile state of song structure. These changes include stuttering, creation, deletion and distortion of song syllables. As the decrystallization progressed, the proportion of normal songs decreased and that of abnormal ones increased. However, baseline songs were still produced with low probability even during the peak of song degradation.

The changes seen in the three adaptive protocol birds were very similar to those seen in deafened birds (Nordeen and Nordeen, 1992). Stuttering occurred in all three adaptive protocol birds and was the most dramatic change in song organization. Both complex syllables (Figure 3.3 a, b) and modified introductory notes (Figure 3.3 c, d) were stuttered. A secondary effect of stuttering was a substantial increase in the maximum observed song length. Other changes to song organization induced by the feedback were the addition of new syllables to the song and, infrequently, the deletion of old syllables from the song (Figure 3.4). Many zebra finch syllables contain sets of frequencies that are integer multiples of a common fundamental frequency; such sets are called harmonic stacks (Figure 3.1, syllable D). The same three birds also showed spectral distortion in their song syllables

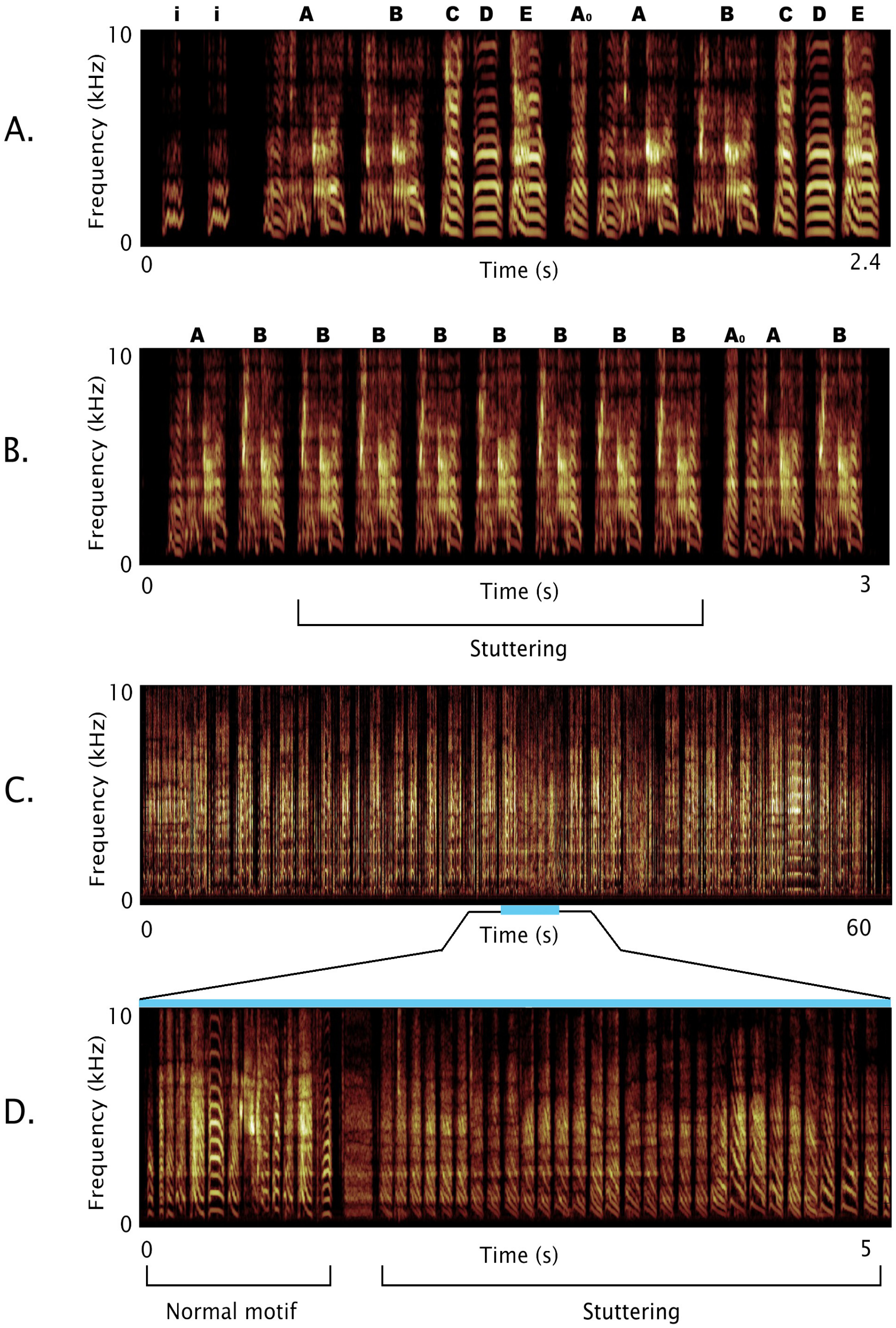
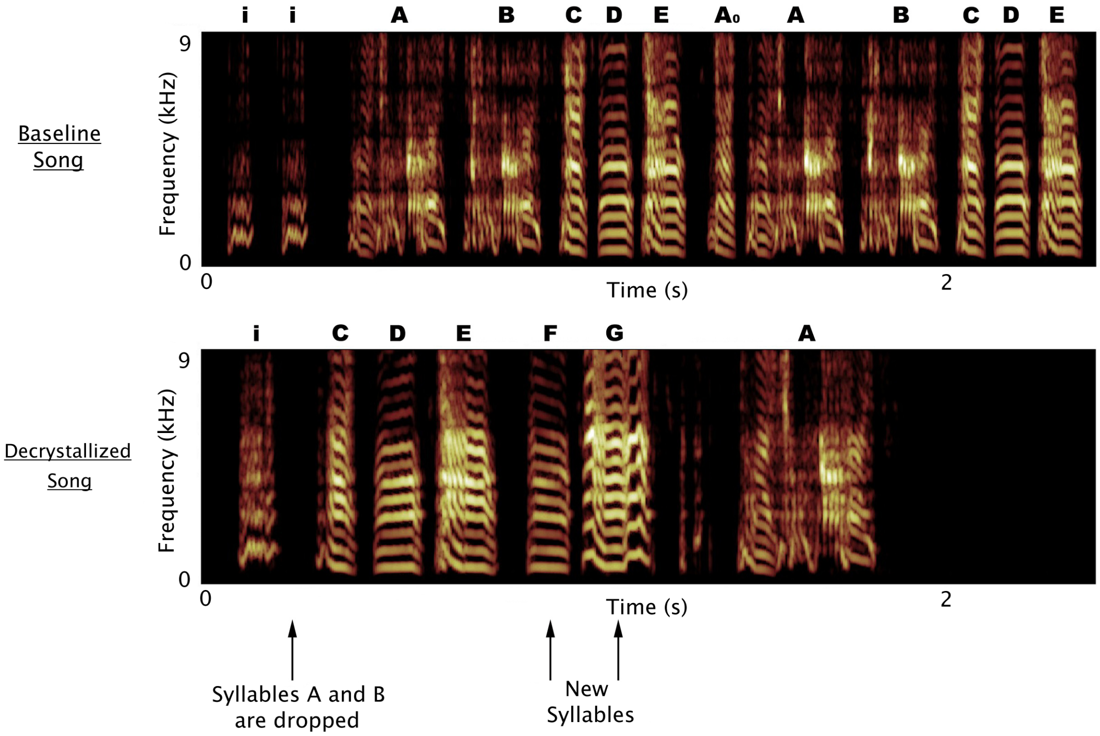


Figure 3.3. Stuttering of syllables in decrystallized birdsong. Normal (A) and decrystallized (B) songs of bird d6, who frequently stuttered a complex song syllable after two weeks of exposure to perturbed auditory feedback. Bird d5 (C, D) showed a dramatic increase in average song bout length due to the stuttering of an introductory note. Note that these abnormal songs were produced by the birds after the removal of feedback and were not simply an instantaneous effect of auditory feedback perturbation.



Syllables A and B are dropped

New Syllables

Figure 3.4. Addition and deletion of syllables in decrystallized birdsong. Top panel: normal song of bird d6. Bottom panel: decrystallized song of bird d6, showing the deletion of syllables A and B, and the addition of new syllables F and G.

(Figure 3.5) including wobbles in the harmonic structure of simple notes and the production of two superimposed harmonic stacks, suggesting a loss of precise control over the vocal organ (the syrinx; Suthers, 1990; Fee et al., 1998). There was considerable variability in the magnitude and time course of the changes between different birds, but significant changes were generally seen within six weeks. Finally, all the changes in song structure described above could occur within different motifs in the same bout (Figure 3.6). This is significant because one of the hallmarks of crystallized song is its robust temporal stereotypy - an identical syllable sequence is maintained within all the motifs of a bout. Decrystallized song, in contrast, lacks this stereotypy.

Two birds received feedback in the syllable-triggered protocol. In one of the birds, significant changes in the spectrum of the target syllable appeared in less than a week and increased in magnitude for the remainder of the feedback period. No changes were seen in the spectrum or the ordering of any of the other syllables in the song. The changes to the

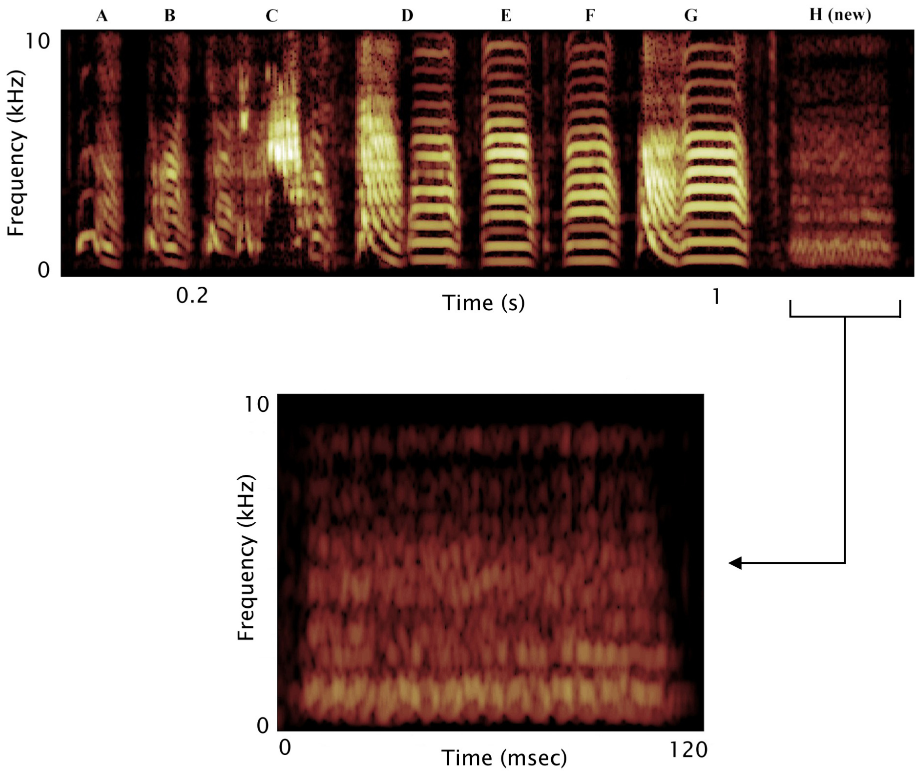


Figure 3.5. Spectral distortion in decrystallized song syllables. Syllable H appeared in the song of bird d4 after decrystallization. It has a highly abnormal sound spectrum, containing no discernable harmonics or well controlled spectral structure. Juvenile birds produce sounds very similar to syllable H before they have learned to control the syrinx properly.

target syllable consisted of the appearance of harmonic frequencies around previously single frequency portions of the syllable's spectrum (see Figure 3.7). The presence of these additional harmonics grew more frequent with time, until eventually this initially tonal syllable was sometimes produced as a distorted harmonic stack. The ability to alter the spectral structure of a sound in response to altered auditory feedback has also been observed in humans (Houde and Jordan, 1998). The second bird used in this protocol showed no changes in his song after 11 weeks of feedback perturbation.

Decrystallization essentially consisted of a large increase in the variability of the song. In order to quantify the changes associated with the arrangement and variability of syllable sequences, we developed an automatic method to sort the data acquired on a given day into syllable types by classifying the different spectral and temporal features of each

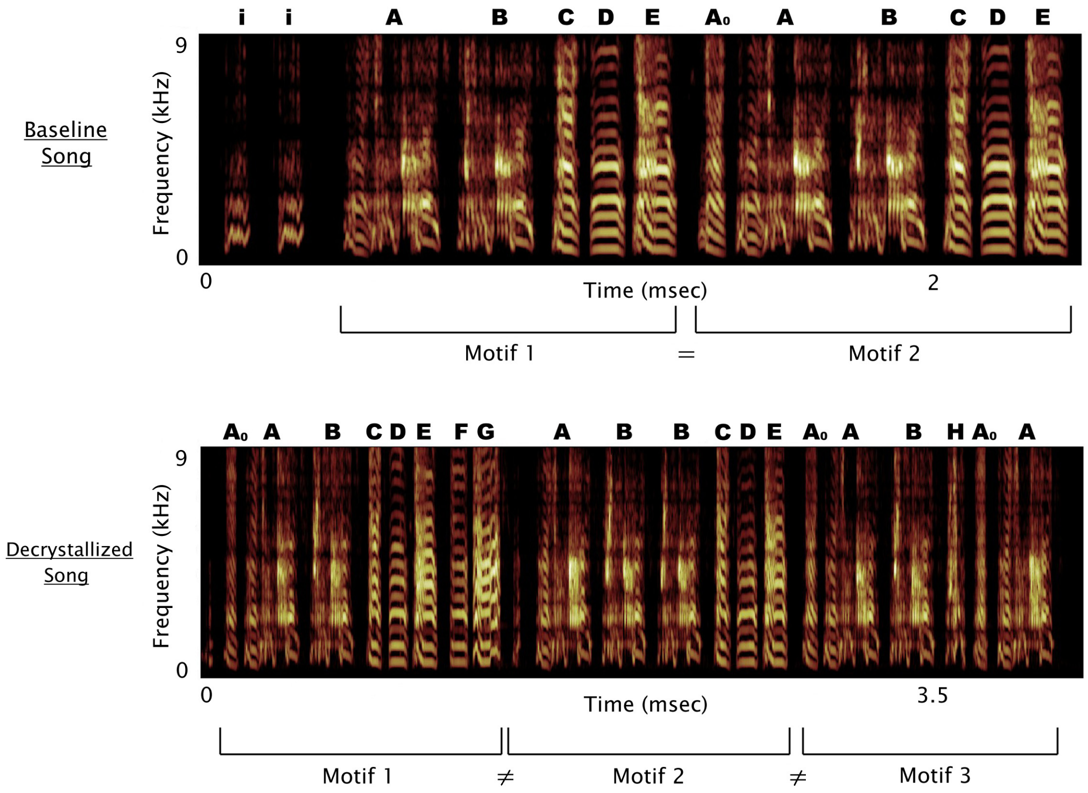


Figure 3.6. Decrystallized songs contained increased variability in syllable sequencing. A hallmark of normal zebra finch song is that song bouts are comprised of a single repeated motif. After decrystallization, in contrast, birds frequently produced song bouts comprised of different motifs.

syllable (see Methods). This transforms the raw voltage waveforms recorded from the microphone into a series of syllable strings, such as "A B C D E", from which we then calculated the probability of different syllable sequences. We defined the baseline motif as the most probable sequence of syllables that the bird repeated in a bout before the feedback period began. Figure 3.8a shows the probability of singing the baseline motif as a function of time for one bird. After receiving one month of perturbed feedback, the probability of the baseline motif for this bird decreased by a factor of six.

The probabilistic sequencing of syllables by the bird on a particular day can be fully characterized as a Markov chain (Feller, 1968), in which the likelihood of singing a particular syllable depended only on the occurrence of the last syllable produced, and not on any prior syllables. The bird's song over the course of the experiment can then be described as a

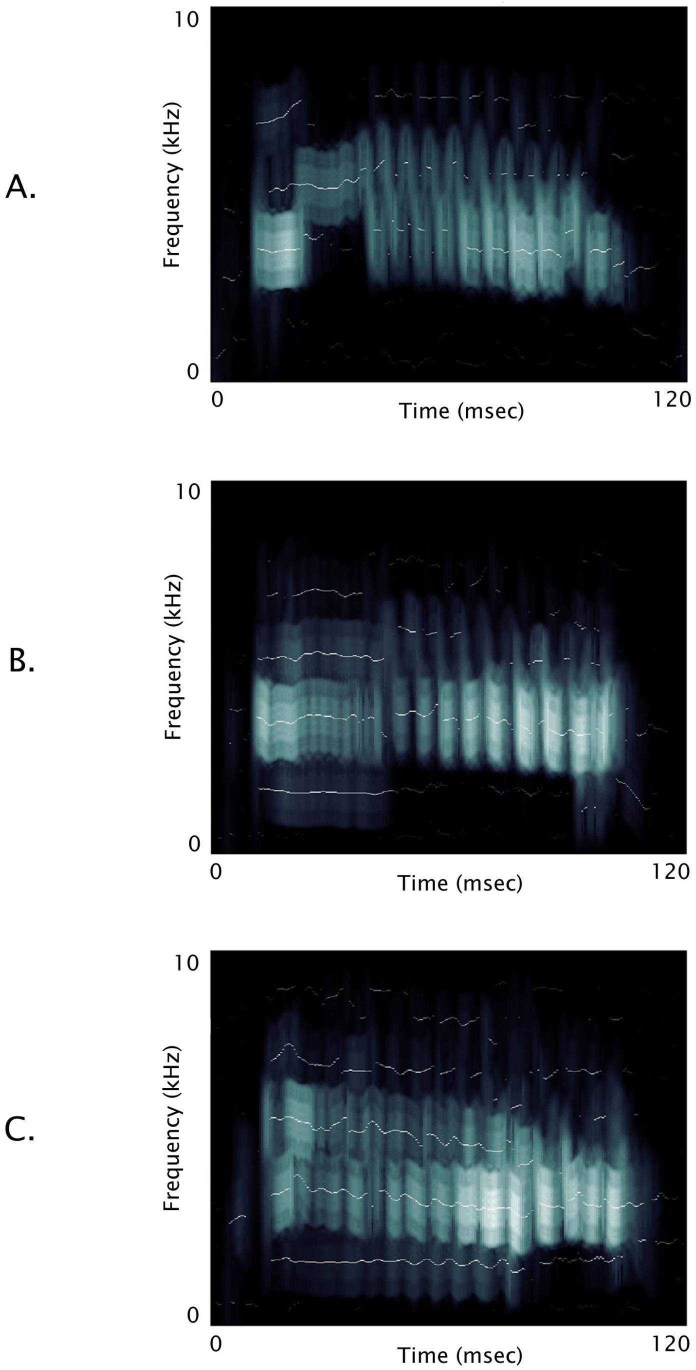


Figure 3.7. Decrystallization of a single syllable. A: baseline (pre-feedback) version of the syllable. B: After one week of syllable-triggered feedback. Harmonic frequencies (white lines) have appeared around the single frequency in the early portion of the syllable (at t = ~ 30 msec). C: After one month of feedback, additional harmonic frequencies now stretch throughout the duration of the syllable.

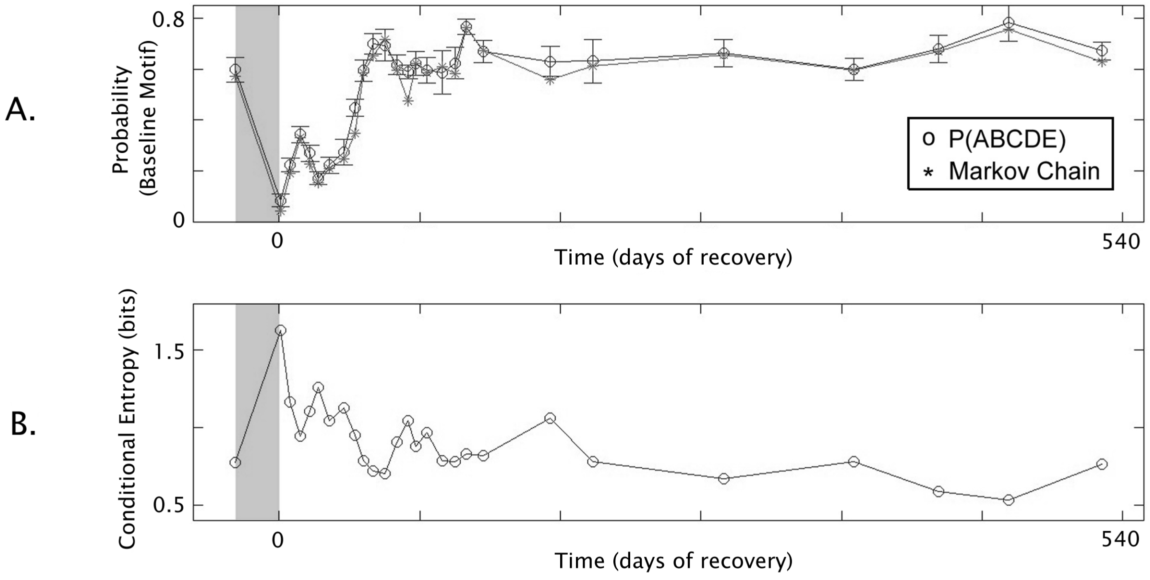


Figure 3.8. Time course of decrystallization and recovery. A) The probability of one bird singing his baseline motif as a function of time. The shaded area represents the feedback interval. Circles represent the baseline motif probability calculated from the entire sequence of syllables (Prob {A B C D E}), while asterisks represent the baseline motif probability calculated from the syllable to syllable transition probabilities of the Markov chain for the song (Prob{A} * Prob{B|A} * Prob{C|B} * Prob{D|C} * Prob{E|D}). The two curves are identical to within 4% error. B) The conditional entropy of the song, which is a measure of song variability, for the same bird.

Markov chain which evolves in time (see Figure 3.8a), and the conditional entropy of the Markov chain can be used as an estimate of the variability of the song on a particular day (Figure 3.8b; Cover and Thomas, 1991). The conditional entropy measures the uncertainty in observing syllable B, given that syllable A was just produced. If each syllable is followed only by a single other syllable (only one type of syllable sequence is produced), then the conditional entropy will be low, whereas if any syllable can follow any other syllable (many different sequences are produced), the conditional entropy will be high. We examined the data for all three adaptive protocol birds using this method and found that birds in a decrystallized state had a significantly higher conditional entropy than they did in their baseline state. Taken together, the time course of the baseline motif probability and the conditional entropy shows that the bird in Figure 3.8 went from being a low variance singer with essentially a single motif, to being a high variance singer who produced a large number of different motifs.

After the removal of the artificial feedback, the temporal and spectral variability in the songs of all three adaptive protocol birds slowly decreased over the following weeks and months. The songs eventually recrystallized and were again stable. Both the probability of the baseline motif and the conditional entropy of the song returned to their baseline levels. The occurrences of stuttering, abnormal sequencing of syllables, and modified spectral organization gradually became infrequent and they were replaced by the temporal and spectral organizations characteristic of the original song. A complete recovery took about two to four months. The syllable-triggered bird made a partial recovery by 8 months after the cessation of feedback. The slow progression of these changes suggests that auditory feedback does not exert a great deal of instantaneous control over the production of song, but instead has a cumulative effect on song maintenance. The recrystallized songs appeared to remain stable indefinitely. We tracked one bird for a year after his recovery and saw no departures from the baseline song structure.

3.3 Discussion

Our results demonstrate that zebra finches need auditory feedback to maintain their songs in adulthood. This species, which does not modify its song or learn new songs after crystallization, appears to retain a great deal of plasticity in its auditory-vocal control system. This plasticity is sufficient to produce modifications in both the temporal pattern of song organization and the spectral structure of individual syllables. This finding is not consistent with the classical depiction of song development in which a dynamic learning period in youth ends in a static maintenance period in adulthood. Thus, the distinction between age-limited and open-ended learners may not be as sharp as these names imply. Furthermore, the recrystallization of the bird's original song shows clearly that, despite the destabilization of the behavior, a memory of the original song remains which may be recovered upon restoration of normal auditory feedback. This indicates that the song is maintained by an active control process, and therefore, somewhere in the song control system, we would expect to find neurons which are sensitive to perturbations of auditory feedback.

3.4 Methods

Birds were housed in custom designed plexiglass cages and were paired with a female who lived in a separate partition of the cage. The playback speaker's output was calibrated to be approximately the same as the sound level of the bird's vocalizations (80-90 dB SPL in the bird's ear). At the start of the baseline song recordings, the birds ranged from 130 to 300 days old (mean age was 200 days; zebra finches reach adulthood at 90 days). The singing rates of the different birds ranged from tens to thousands of motifs per night. However, there was no apparent correlation of age or singing volume with the magnitude of the effects we observed. For the two birds who did not sing frequently, we collected no baseline data during the feedback period.

Adaptive Feedback Perturbation. Microphone data were sampled at 40 kHz, after being low-pass filtered (10 kHz cutoff, 7-pole anti-aliasing filter). As is shown in Figure 3.2, the computer continuously acquired data from the microphone in segments of 100 msec. Each of these "bins" of data was passed through a software-based infinite-impulse response (IIR) filter_(the "trigger filter"), that was used to detect song vocalizations while avoiding noise artifacts (such as pecking, wing flaps, and low amplitude calls). The playback of sound occurred when the output of the trigger filter exceeded an RMS threshold. The artificial feedback was thus produced with a 100 msec delay, and coincided with a silent interval in the bird's song or with the following syllable, depending on the timing of the song with respect to the bin borders created by the computer. The alternation between recording vocalizations and playing sounds prevented a positive feedback loop from developing. The playback stimulus was constructed from the 100 msec sound that had caused the trigger event, and was narrowband filtered before being played back to the bird. A narrowband filter was chosen to make the playback sounds difficult to localize. For one bird a wideband IIR filter was used. No significant differences in results were observed between this bird and the two narrowband birds. Delayed feedback was played to the birds, rather than other interfering stimuli, in order to replicate the spectral structure of auditory feedback that these birds normally hear. The effects of white noise will be examined in future work.

The birds exposed to the adaptive protocol showed changes in their songs similar to those of deafened birds. Because deafening can be caused by prolonged exposure to excessively loud sounds (Wooley and Rubel, 1997), we used a behavioral test to demonstrate that the conditions used in our experiment did not cause deafness. Zebra finches respond to sounds by vocalizing. At the beginning and the end of the feedback period, we examined differences in the calling probability of two of the adaptive protocol birds to the playback of quiet sounds versus no sounds. The sounds were a variety of natural stimuli including conspecific calls and songs. Both birds produced significantly more calls during the presentation of quiet sounds ($p < 0.001$, generalized-likelihood-ratio test for Bernoulli random variables), which indicates that the birds could hear the sounds.

Syllable-Triggered Perturbation. Recognition of the target syllable was achieved by using a series of IIR filters in conjunction with each other to perform a logical operation (e.g., power in band X and not in band Y). The triggering was based on a small segment of the time-varying spectrum of the target syllable, which was unique to that syllable. An original copy of the crystallized trigger syllable was used as the playback stimulus for the duration of the experiment. Typical zebra finch syllables are 80-150 msec in length. The triggering resolution was 50 msec, which was short enough to ensure that the feedback always overlapped the trigger syllable itself (and partially on the following syllable). The other syllables of the song received no feedback.

Spectral Analysis. We calculated the time-frequency spectrogram for each song with a sliding window (5-8 msec), in which each time point consisted of the direct multitaper estimate of the power spectrum (with a time-bandwidth product $NW = 3$ or 4; Ho et al., 1998). The spectrograms shown in all of the figures were analyzed in this manner. For the analysis associated with Figure 3.7, a harmonic analysis was then used to determine the location and magnitude of the jumps in the discrete spectrum of each syllable (Thomson, 1982) based on the multitapered spectral estimates. This analysis revealed additional statistically significant harmonic frequencies in the target syllable after the feedback was presented to the bird ($p < 0.01$). An example of these results is shown in Figure 3.7.

Syllable Classification. For each syllable, we extracted the length and a number of time-varying parameters based on the spectral analysis described above. A modified K-means clustering algorithm (Selim and Ismail, 1982) was then used to partition the syllables produced on a given day into subsets. These subsets were labeled by the experimenter (syllable A, syllable B, etc.; labeling was done blind to the day on which the data were acquired). For each day of data, approximately 1000 syllables were analyzed. The standard deviations that are shown as error bars in Figure 3.8 were obtained by bootstrapping the probability estimates from the data (Bradley, 1993). The steps underlying this process are described in detail in the following two paragraphs.

For each day of data, an automatic algorithm was used to extract the individual syllables from all of the song files. Syllables were identified in the time-domain as events whose power crossed an RMS threshold (with respect to background activity). For each syllable, we then calculated the length and a number of time-varying parameters (envelope, peak frequency, pitch, goodness-of-pitch, and Wiener entropy) based on the spectral analysis techniques described above. Many previous classification methods for birdsong syllables and songs have been based on direct comparisons (e.g., cross-correlations) of the time-frequency spectrograms of the signals in question. However, the entire spectrum contains so much information that it is difficult to make accurate comparisons of two signals in a reasonable amount of computation time. Furthermore, it is highly unlikely that the bird is actively controlling every single feature contained in the song. Recent experimental work (Fee et al., 1998) has shown that by modulating only a few control parameters, the syrinx can produce an enormous range of vocalizations. This result forms the basis of our analytical methods. We represented each syllable of the bird's song with a few time-varying parameters, which were intended to be analogous to features of the song that were under direct control by the bird. These features are briefly described in the next paragraph.

It is well known the dorsal muscles of the syrinx can rapidly close and open the syrinx during song, producing an amplitude modulation of the signal. This is reflected in the envelope of each syllable, which we calculated via the Hilbert Transform, and then lowpass filtered to 250 Hz (an upper bound on the maximum rate at which the bird could add AM modulation to the signal). The ventral muscles of the syrinx have a role in frequency modulation and in controlling the fundamental frequency of the sound being produced by the syrinx. We estimated the pitch of each syllable using the peak of the cepstrum. Cepstral analysis is a standard technique in the speech community for extracting the pitch of a signal (Kadambe and Boudreauxbartels, 1992). The cepstrum is defined at the spectrum of the log spectrum of a signal. In brief, if a signal has a lot of pitch (like a harmonic stack), frequencies that are spaced at integer multiples of the fundamental frequency will contain more power than the rest of the signal. The cepstrum of this signal will have a peak at the fundamental frequency of the harmonic stack. The magnitude of this peak can be used as an estimate of the goodness-of-pitch, that is, how "pitchy" the sound really was. The Wiener entropy is a measure of the predictability of the signal (Ho et al., 1998). Syllables that are very tonal will have a very low Wiener entropy (highly predictable) whereas those that are white will have a very high Wiener entropy (highly unpredictable). The Wiener entropy most likely reflects a number of parameters that are under control of the bird. Like the other time-varying features discussed, it has a very characteristic waveform for each syllable that the bird produces.

Based on these features, a modified K-means clustering algorithm was then used to partition the syllables produced on a given day into subsets by iteratively minimizing the distance between all syllables and their respective cluster centers. K cluster centers were chosen randomly from the entire set of syllables. Each syllable was then assigned to the nearest cluster based on its distance from the cluster centers. Because the lengths of the syllables we observed typically varied from 30 to 250 msec, the lengths of the different feature vectors also varied. Because of this, the Euclidean distance was not a suitable metric for comparing two syllables. Instead, we used a symmetric dynamic time-warping (DTW) algorithm to compute the distance between two syllables (Casacuberta et al., 1987). The DTW algorithm is another standard tool in the speech community and finds the optimal nonlinear alignment between the two vocalizations (Myers et al., 1980). This compensated for differences in length between different syllables and differences in phase between different iterations of the same syllable (which would wreak havoc with the Euclidean distance). Once each syllable was assigned to a cluster, the cluster centroids were recomputed (based on the new cluster memberships) and the process was started over. Iteration continued until the system of clusters converged and the centroids became stable. The resultant syllable clusters were then labeled by the experimenter, and converted into text strings from which the various probabilities were be computed (e.g., the probability of the bird singing his baseline syllable sequence).

Markov Chain and Conditional Entropy. A Markov chain for the bird's song was constructed for each day of data from the K-means classified syllables. Each state of the chain represented a single syllable in the song, and the bird's song over the course of the experiment was described as a Markov chain which evolved in time. In this framework, the likelihood of singing a particular syllable depended only on the occurrence of the last syllable produced, and not on any prior syllables. We have found that many zebra finch songs can be modeled in this manner using only first-order transition probabilities between syllables (i.e., there is no significant difference between the raw empirical estimate of the probability for some string of syllables and the Markov estimate using only the first-order transition probabilities). This demonstrates that the temporal patterning in the bird's song has a fairly simple statistical structure. In addition, the conditional entropy of the Markov chain can be used as an estimate of the variability of the song on a particular day. Estimating the variability of a particular animals vocalizations has in the past been a problematic issue for the birdsong community. The conditional entropy provides a nice solution to this problem. The conditional entropy of the Markov chain is defined as

$$H(Y \mid X) = -\sum_{i} p(x_i)\, p(y_i \mid x_i) \log(p(x_i \mid y_i)$$

where the summation is taken over all y and x (i.e., the entire set of syllable transition probabilities). P(x) is the probability of seeing syllable x at any point in the song. H(Y|X) measures the uncertainty in observing syllable Y, given that syllable X was just produced. If each syllable was followed only by a single other syllable (i.e., a single type of syllable sequence is produced), then the conditional entropy will be low, whereas if any syllable was followed any other syllable (many different sequences are produced), the conditional entropy will be high.