Decrystallization of adult birdsong by perturbation of auditory feedback

Anthony Leonardo & Masakazu Konishi

Computation and Neural Systems Program and Division of Biology, MC 216-76, California Institute of Technology, Pasadena, California 91125, USA

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[1,2]. 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 their song to deteriorate slowly. This 'decrystallization' consisted of a marked loss of the spectral and temporal stereotypy seen in crystallized song, including stuttering, creation, deletion and distortion of song syllables. After normal feedback was restored, these deviations gradually disappeared and the original song was recovered. Thus, adult birds that do not learn new songs never-

theless retain a significant amount of plasticity in the brain.

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[3]. 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. At this stage, the song is referred to as crystallized. Some birds, like the zebra finch, maintain their crystallized song throughout adulthood and are called age-limited learners. Open-ended learners like canaries can, in contrast, learn new songs in adulthood[4].

The stability of song in age-limited learners was previously thought to be maintained without auditory feedback[1,2]. 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[5–7]. Six to eight weeks after deafening, the song of adult zebra finches deteriorates, and shows addition and deletion of syllables, 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
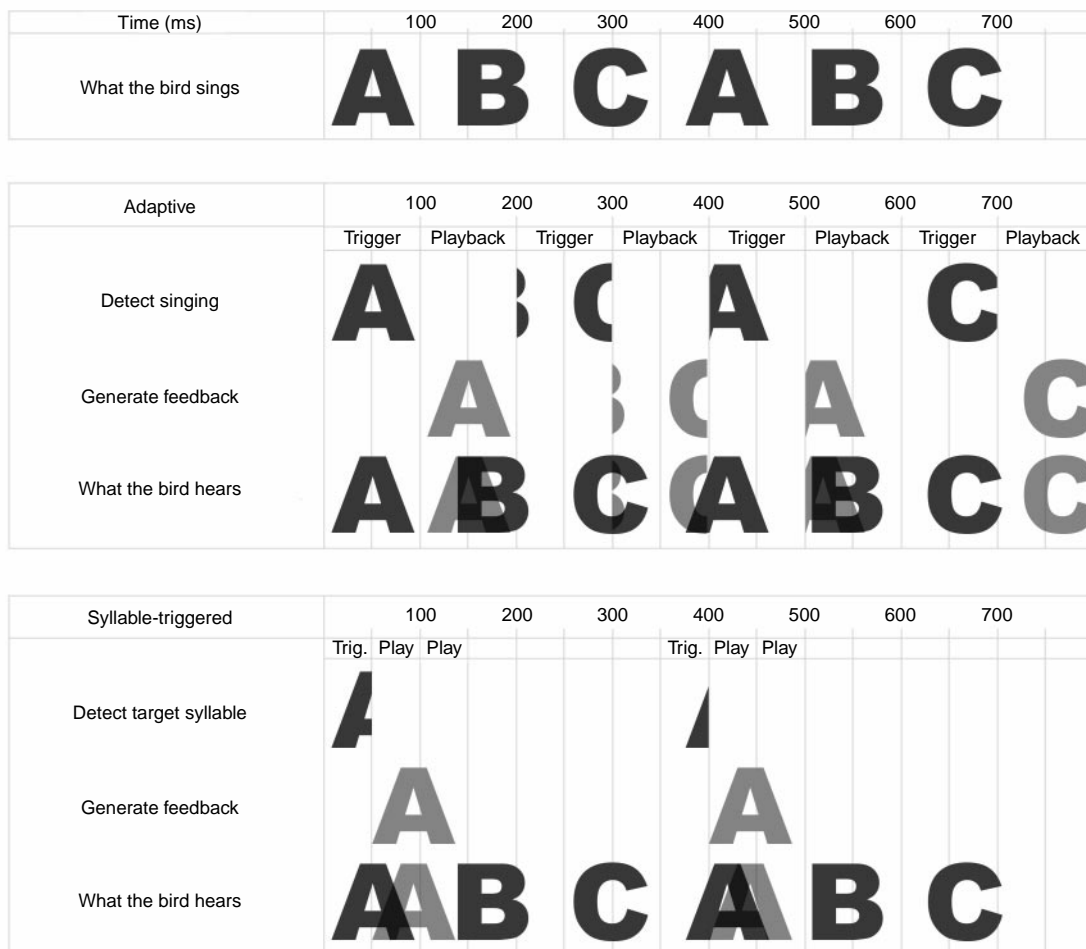

Figure 1 Protocols for constructing the feedback signals. In the adaptive protocol, the computer alternated between recording vocalizations and playing the last vocalization back to the bird. In the syllable-triggered protocol, the computer monitored the bird's song in 50-ms bins and detected the production of a single target syllable. In both protocols, the bird heard the superposition of his own vocalizations and the computer-generated feedback.

either the auditory or vocal motor 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 produce dramatic changes in the adult song.

We developed a computer-controlled system to perturb auditory feedback and designed two feedback paradigms (Fig. 1). In both methods, the computer detected singing and then played 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). To investigate 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.

We recorded the songs of five adult male zebra finches in separate, sound-attenuated chambers for several weeks before beginning the experiment. The variability of their songs was well within that found in crystallized songs. The birds were then placed in the feedback system. After 1–4 months in this environment, four of the five birds
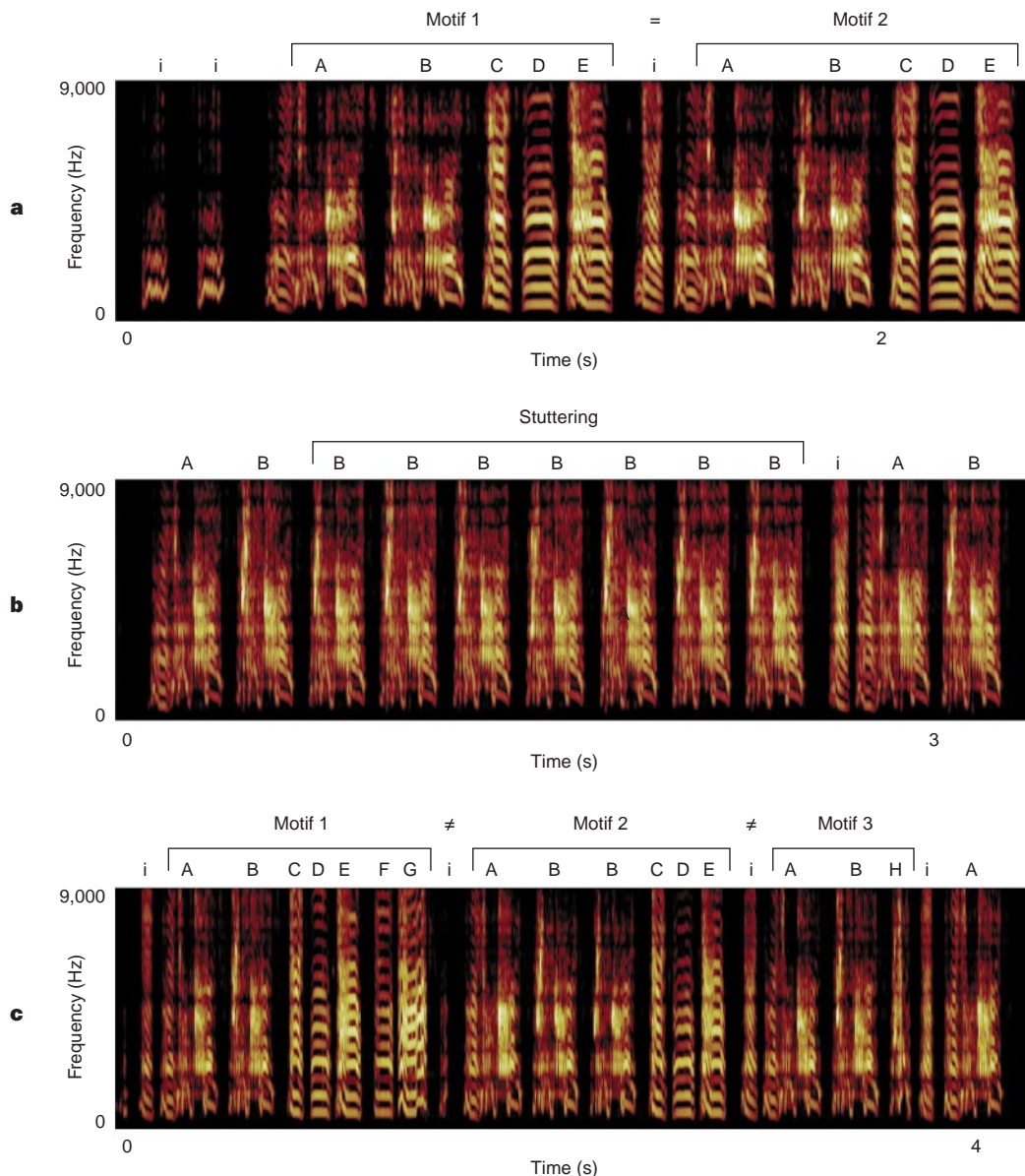

Figure 2 Normal and decrystallized birdsong spectrograms. **a**, Two motifs of a normal, crystallized zebra finch song. Each syllable has a stable and well defined spectral structure. The ordering of the song syllables is identical in the two motifs. The syllable sequence 'A B C D E' represents the baseline motif for this bird. The 'i's represent introductory notes, which occur in variable numbers at the beginning of motifs. **b**, The decrystallized song of the bird shown in **a**, with a motif containing stuttering of the song syllable 'B'. **c**, Another example of the

decrystallized song of the bird shown in **a**. The syllable sequences now vary from motif to motif, indicating a loss of the fundamental temporal stereotypy which characterizes crystallized zebra finch song. Motifs one and three contain new song syllables (F, G, H) which were not present in the baseline motif shown in **a**. Motif two contains a repetition of syllable 'B'. After the feedback was removed, the sequences shown in **b** and **c** gradually disappeared and only the baseline song was produced.

showed dramatic changes in their 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 analysed data consisted of recordings of the bird singing by himself, with no artificial feedback. After the feedback was permanently stopped, we tracked all the birds for another 8–16 months to determine whether they could 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 increased occurren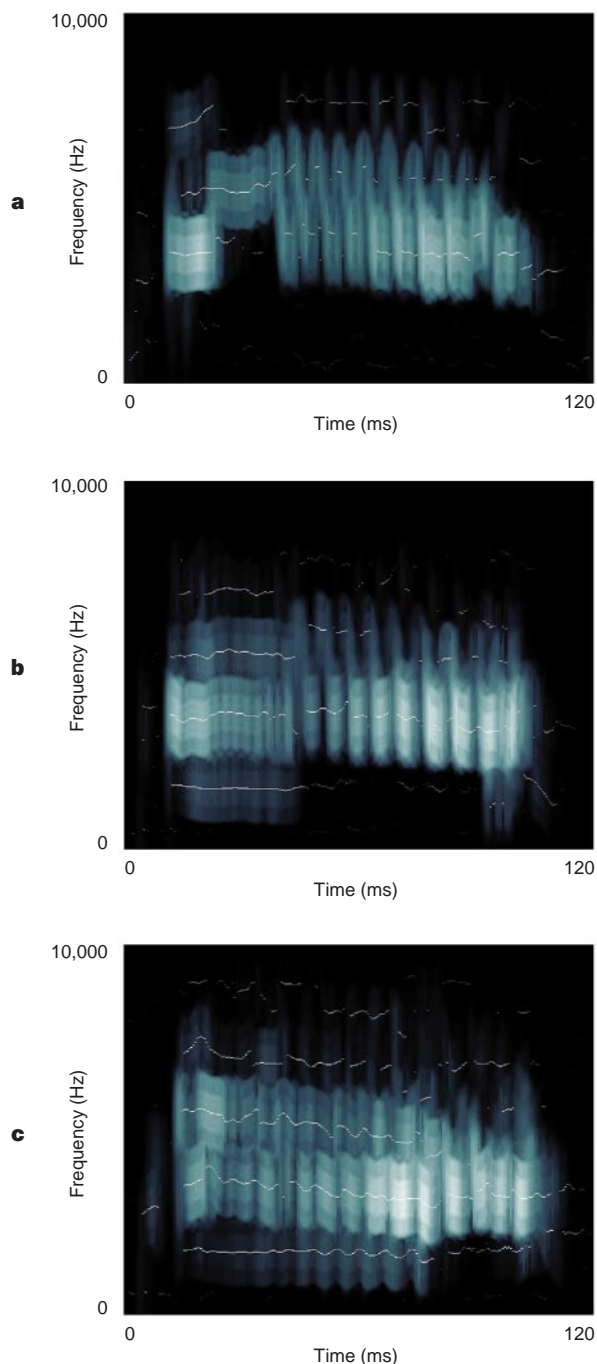ce of properties that were rare in the baseline song. We use the term decrystallization to refer collectively to all of the perturbation-induced changes from the original quantitative and statistical structure of the song; this does not necessarily imply a return to a juvenile state of song structure. 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 at the peak of song degradation.

The changes seen in the three adaptive protocol birds were very similar to those seen in deafened birds[5–7]. A series of spectrograms representative of these results is shown in Fig. 2. Stuttering occurred in all three adaptive-protocol birds and was the most dramatic change in song organization. Both complex syllables and modified introductory notes were stuttered. A secondary effect of stuttering was a substantial increase in the maximum 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. Many zebra finch syllables contain sets of frequencies that are integer multiples of a common fundamental frequency; such sets are called harmonic stacks (Fig. 2a, syllable D). The same three birds also showed spectral distortion in their song syllables, including wobbles in the harmonic structure of simple notes and the production of two superimposed harmonic stacks, indicating a loss of precise control over the vocal organ (the syrinx)[8,9]. 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. This is significant because one of the hallmarks of crystallized song is its robust temporal sterotypy—an identical syllable sequence is maintained within all the motifs of a bout. Decrystallized song, in contrast, lacks this stereotypy (Fig. 2c).

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 target syllable consisted of the appearance of harmonic frequencies around previously single-frequency portions of the syllable's spectrum (Fig. 3). 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. Humans can also alter the spectral structure of a sound in response to altered auditory feedback[10]. 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. 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 using the different spectral and temporal features of each syllable. This transformed the raw voltage waveforms recorded from the microphone into a series of syllable strings, such as 'A B C D E', from which we calculated the probability of different syllable sequences. We define the baseline motif as the most probable sequence of syllables that the bird repeated in a bout before the feedback period began. Figure 4a 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[11]. Thus, the likelihood of singing a particular syllable depends 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 Markov chain which evolves in time (Fig. 4a), and
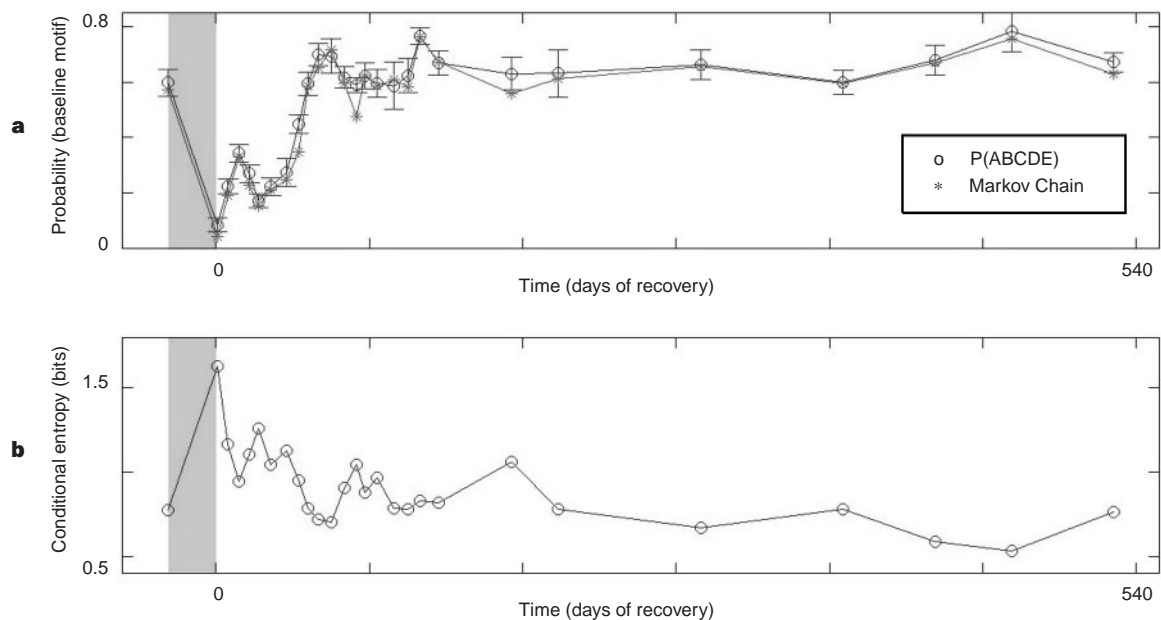

Figure 3 Decrystallization of a single syllable. **a**, Baseline (pre-feedback) version of the syllable. **b**, After one week of syllable-triggered feedback. Harmonic frequencies (white lines) have appeared around the single frequency in the early portion of the syllable (at $t = \sim 30$ ms). **c**, After one month of feedback, additional harmonic frequencies now stretch throughout the duration of the syllable.

Figure 4 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 feedback interval. Circles represent the probability calculated from the entire sequence of syllables (Prob {A B C D E}); asterisks represent the 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.

the conditional entropy of the Markov chain can be used as an estimate of the variability of the song on a particular day (Fig. 4b)[12]. 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 Fig. 4 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 became stable again. Both the probability of the baseline motif and the conditional entropy of the song returned to their baseline levels. Stuttering, abnormal sequencing of syllables and modified spectral organization gradually became infrequent and were replaced by the temporal and spectral organizations characteristic of the original song. A complete recovery took about 2–4 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.

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 would indicate.

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; zebrafinches 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 Fig. 1, the computer continuously acquired data from the microphone in segments of 100 ms. 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 a root mean square threshold. The artificial feedback was thus produced with a 100-ms 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 ms of 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 used instead of other interfering stimuli such as white noise so as to replicate the structure of syllables 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[7], we used a behavioural 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 sounds. 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 (for example, 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 ms in length. The triggering resolution was 50 ms, which was short enough to ensure that the feedback always overlapped the trigger syllable itself (and, partially, 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 ms) in which each time point consisted of the direct multitaper estimate of the power spectrum (with a time-bandwidth product NW of 3 or 4) (ref. 13). The data shown in Fig. 2 were analysed in this manner. A harmonic analysis was then used to determine the location and magnitude of the jumps in the discrete spectrum of each syllable by calculating the F-spectrum[14] of each of the multitapered spectral estimates. This analysis revealed additional statistically significant harmonic frequencies in the target syllable after the feedback was presented to the bird ($P < 0.01$) (Fig. 3).

Syllable classification. For each syllable, we extracted the length and a number of time-varying parameters (envelope, peak frequency, pitch, goodness-of-pitch and Wiener entropy[13]) based on the spectral analysis described above. A modified K-means clustering algorithm[15] was then used to partition the syllables produced on a given day into subsets. These subsets were labelled by the experimenter (syllable A, syllable B and so on; labelling was done blind to the day on which the data were acquired). For each day of data, approximately 1,000 syllables were analysed. The standard deviations, which are shown as error bars in Fig. 4, were obtained by bootstrapping the probability estimates from the data[16].

1. Konishi, M. The role of auditory feedback in the control of vocalization in the White-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
2. Nottebohm, F. Auditory experience and song development in the chaffinch, *Fringilla coelebs. Ibis* **110**, 549–569 (1968).
3. Sossinka, R. & Bohner, J. Song types in the zebra finch. *Z. Tierpsychol.* **53**, 123–132 (1980).
4. Marler, P. & Peters, S. A sensitive period for song acquisition in the song sparrow, *Melospize melodia*—a case of age limited learning. *Ethology* **76**, 89–100 (1987).
5. Nordeen, K. & Nordeen, E. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57**, 58–66 (1992).
6. Okanoya, K. & Yamaguchi, A. Adult bengalese finches require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.* **33**, 343–356 (1997).
7. Wooley, S. & Rubel, E. Bengalese finches Lonchura-striata-domestica depend on auditory feedback for the maintenance of song. *J. Neurosci.* **17**, 6380–6390 (1997).
8. Suthers, R. Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* **347**, 473–477 (1990).
9. Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. The role of nonlinear dynamics of the syrinx in birdsong production. *Nature* **395**, 67–71 (1998).
10. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
11. Feller, W. *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1968).
12. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, New York, 1991).
13. Ho, C. E., Pesaran, B., Fee, M. S. & Mitra, P. P. Characterization of the structure and variability of zebra finch song elements. *Proc. 5th Joint Symp. on Neural Computation* 76–83 (1998).
14. Thomson, D. J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
15. Selim, S. Z. & Ismail, M. A. K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 81–87 (1984).
16. Bradley, E. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

Neuronal correlates of parametric working memory in the prefrontal cortex

Ranulfo Romo, Carlos D. Brody, Adrián Hernández & Luis Lemus

Instituto de Fisiología Celular, Universidad Nacional Autonóma de México, México D.F. 04510, México

Humans and monkeys have similar abilities to discriminate the difference in frequency between two mechanical vibrations applied sequentially to the fingertips[1–3]. A key component of this sensory task is that the second stimulus is compared with the trace left by the first (base) stimulus, which must involve working memory. Where and how is this trace held in the brain? This question was investigated by recording from single neurons in the prefrontal cortex of monkeys while they performed the somatosensory discrimination task. Here we describe neurons in the inferior convexity of the prefrontal cortex whose discharge rates varied, during the delay period between the two stimuli, as a monotonic function of the base stimulus frequency. We describe this as 'monotonic stimulus encoding', and we suggest that the result may generalize: monotonic stimulus encoding may be the
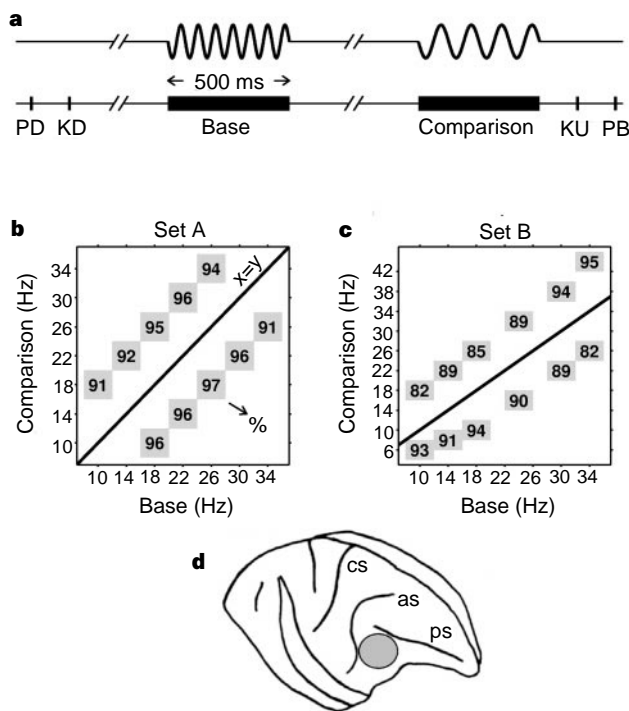
Figure 1 Discrimination task. **a**, Sequence of events during discrimination trials. The mechanical probe is lowered, indenting the glabrous skin of one digit of the hand (PD); the monkey places his free hand on an immovable key (KD); the probe oscillates vertically, at the base frequency; after a delay, a second mechanical vibration is delivered at the comparison frequency; the monkey releases the key (KU) and presses one of two push-buttons (PB) to indicate whether the comparison frequency was higher or lower than the base. **b**, **c**, Stimulus sets used during recordings. Each grey box indicates a base frequency/comparison frequency stimulus pair used; the number inside the box indicates overall per cent correct trials for that base/comparison pair. **d**, Location of recording sites that gave somatosensory working-memory-related responses; CS, central sulcus; AS, arcuate sulcus; PS, principal sulcus.