Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song

Ofer Tchernichovski,[1]*† Partha P. Mitra,[2]* Thierry Lints,[1]*
Fernando Nottebohm[1]*

Song imitation in birds provides good material for studying the basic biology of vocal learning. Techniques were developed for inducing the rapid onset of song imitation in young zebra finches and for tracking trajectories of vocal change over a 7-week period until a match to a model song was achieved. Exposure to a model song induced the prompt generation of repeated structured sounds (prototypes) followed by a slow transition from repetitive to serial delivery of syllables. Tracking this transition revealed two phenomena: (i) Imitations of dissimilar sounds can emerge from successive renditions of the same prototype, and (ii) developmental trajectories for some sounds followed paths of increasing acoustic mismatch until an abrupt correction occurred by period doubling. These dynamics are likely to reflect underlying neural and articulatory constraints on the production and imitation of sounds.

Vocal imitation is guided by auditory information, requires intact hearing, and is very sensitive to the age or reproductive condition of the individual (1, 2). The brain circuits that govern this skill in songbirds have been described (3). We here report on conditions that bring vocal learning under fine experimental control and provide a detailed acoustic analysis of the sound transformations that underlie the learning process.

Zebra finch (*Taeniopygia guttata*) males develop their song between 35 and 90 days after hatching, a time known as the sensitive period for vocal learning (4). This song consists of complex sounds ("syllables") separated by silent intervals (5). A song motif is composed of dissimilar syllables repeated in a fixed order (5). When a young male zebra finch is reared singly in the company of an adult male, it develops a song that is a close copy of the sounds and temporal order of that male's song (4, 6). Acquisition of the auditory memory of the model song can start as early as 25 days after hatching, but this onset can be delayed by withholding exposure to the model (7, 8). Once acquired, a stored representation of the model song can be converted to a motor imitation. This conversion has been modeled by assuming simple Hebbian and reinforcement learning rules (9). Nevertheless, past technical limitations encountered when studying early song development have left much of the fine-grained structure of the imitation process unexplored.

In many songbirds, as in humans, first acquisition of auditory memories of learned sounds occurs before the onset of vocal learn-

ing (10). Under such conditions, it can be difficult to distinguish between the learned and the innate component of the developing sounds. In the zebra finch, however, the sensory phase of model acquisition overlaps with the period of motor development of learned vocalization (5). We took advantage of this overlap to delay model acquisition so as to obtain a baseline of "untutored" song during the early subsong stage, and we then examined the effect of exposure to a model song during the remainder of the sensitive period for vocal learning. Untutored subsong was recorded, and then birds were trained, starting on day 43 after hatching, to peck at a key to trigger a short song playback from a small speaker housed within a plastic male model (11). Training persisted until the end of the experiment. To enhance the identification of time-frequency structure in subsong, which typically consists of poorly structured sounds, we used multitaper spectral analysis techniques and estimated spectral derivatives that act like "edge detectors" in the time-frequency plane of the spectrogram (12) (Fig. 1). Figure 1B presents spectral derivatives of the emerging song of a bird just before train-

Fig. 1. An example of training. (A) Acclimation to the training apparatus from days 30 to 42 after hatching, in the presence of a plastic model of an adult male (on middle perch). (B) Untutored subsong was recorded on day 43. Spectral derivatives provide a representation of song that is similar but superior to the traditional sound spectrogram. Instead of power spectrum versus time, we present directional derivatives (changes of power) on a gray scale so that the detection of frequency contours is locally optimized. This was particularly useful for the analysis of juvenile song. (C) The keys were then uncovered. The bird learned to peck on either one of the keys to induce a short song playback from the plastic model. (D) Song playback was composed of two renditions of the song motif (the "model") depicted. The overall daily exposure was limited to 28 s. As shown, the bird's song had changed by (E) the second and (F) the third day of training.
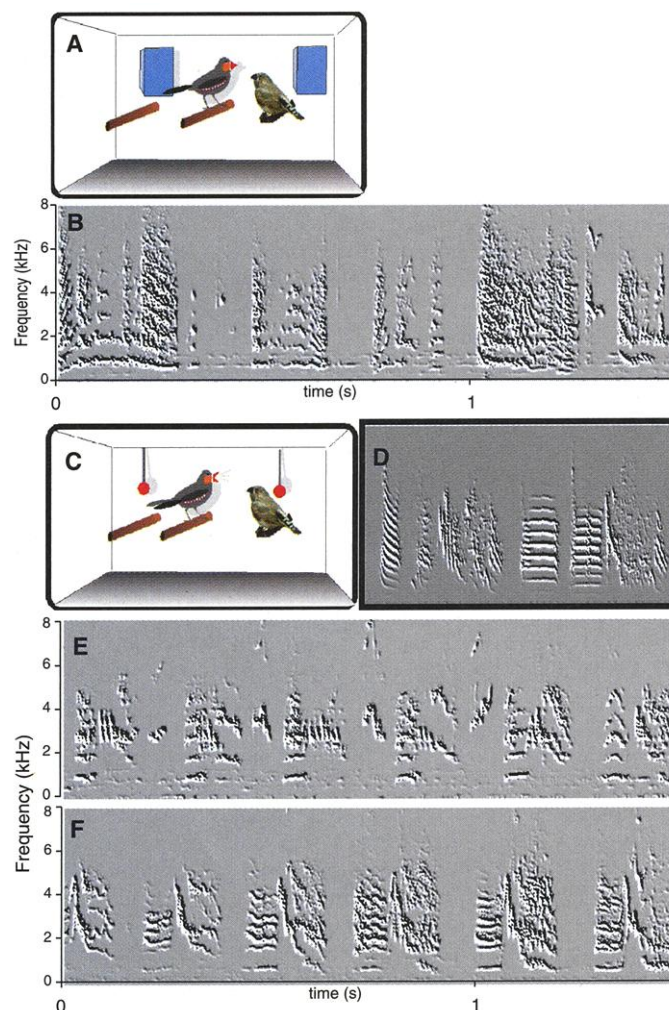
[1]Field Research Center, The Rockefeller University, Millbrook, NY 12545, USA. [2]Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA.

*All authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: tcherno@mail.rockefeller.edu

ing (on the morning of training day 1), and Fig. 1, E and F, shows the spectral derivatives on training days 2 and 3. As shown, song had changed remarkably by 2 days after the onset of training: Sounds became more structured and appeared in a more predictable temporal order. On training day 3, some sounds were already similar to those of the model song.

Indirect Imitation Trajectories

Viewed most simply, an imitation trajectory could be represented by a path leading directly from the acoustic features of sounds produced before exposure, to those of a target sound present in the model song (13). Alternatively, an imitation trajectory might deviate from a direct path to negotiate constraints imposed, for example, by propensities of brain function and/or the physics of sound generation by the vocal organ. For example, in the bird's vocal organ, as in some musical instruments, pitch might become unstable across certain ranges of airflow (14). An indirect trajectory of sound changes could either avoid those ranges or take advantage of them. An example of an automatically traced (15) imitation trajectory of a simple harmonic stack (16) is presented in Fig. 2A. A raw imitation of the model's harmonic stack was apparent on training day 5, although the pitch at that time was slightly higher than the model's. We measured the pitch of harmonic stacks produced by this bird every day until the pitch matched that of the model. As shown, the pitch error between the developing sound and the model sound increased slowly and consistently from training days 5

to 13, and then when the pitch reached the frequency of the model's first harmonic, it was corrected by an abrupt period doubling. Assessed on the basis of pitch alone, model approximation in this case was indirect, but in other cases, it was direct (Fig. 2B). A period-doubling trajectory makes sense only if the initial pitch is higher than that of the model (17). To test for a possible effect of initial pitch on the trajectory taken, we examined a sample of 10 period-doubling trajectories and found that the initial pitch ranged between 61 and 582 Hz above that of the target. Tracking 10 non–period-doubling trajectories gave a significantly lower initial pitch (range, 156 Hz below the target to 132 Hz above it; t test, $P < 0.01$). We conclude that the initial pitch predicts most, though not all, period-doubling trajectories.

The above findings do not contradict the model-approximation theory, which does not specify how the approximation is achieved. For example, the "indirect" trajectory may be operationally short if it is easier to increase pitch and then take advantage of the nonlinear dynamics of sound production by the syrinx to reduce pitch by half (14). Our results show, however, that the imitation trajectory of even a simple sound cannot be explained by just invoking a gradual reduction of acoustic error. We now describe the song imitation process more generally.

The Early Generative Phase

The song development process is complex, and to make progress in quantification, we reduced the song to a set of four features:

Wiener entropy, spectral continuity, pitch, and frequency modulation (12, 18) (Fig. 3A). On any 1 day, we characterized the song by the distribution of these features, computed on a frame-by-frame basis over a 10-s time period (12). This obviated the need for partitioning or classifying sounds (12, 15). Thus, in this section, study of song development is reduced to studying the development of feature distributions. We studied the changes in both the mean values and the SDs of the features. In addition, we examined how the feature distribution approached that of the model using the Kolmogorov-Smirnov (KS) statistic (the changes in mean, SD, and KS statistic do not necessarily mirror each other).

We used the SDs of features to define a measure called "feature diversity" (19), which is an estimate of the range of different sounds produced by a bird during a specific stage of learning. As shown in Fig. 3B, training birds with a model song induced an abrupt increase of song feature diversity by the second day of training (paired t test, $n = 42$, $P < 0.0001$; in 29 out of 42 birds trained, feature diversity increased above the upper 0.05% confidence interval, and in 2 birds, it decreased below the lower 0.05% confidence interval). The changes in the mean feature values for Wiener entropy and spectral continuity are shown in Fig. 3C. On average, the sounds produced had higher spectral continuity and lower Wiener entropy, signifying higher temporal stability and higher tonality respectively; that is, sounds became more structured. Moreover, not only were pitch, spectral continuity, and Wiener entropy val-

Fig. 2. Indirect and direct approaches to the imitation of harmonic stacks. **(A)** Spectral derivatives of a developing harmonic stack in reference to a syllable from a model song (left). The pitch of the harmonic stack is given at the top of each panel (16). A quantitative examination of the pitch error between this developing harmonic and the model harmonic stack shows a gradual increase of error, followed by an abrupt period doubling that reduced the error in a single step (right). The graph presents the mean pitch values of harmonic stacks produced by this bird across 30-s samples of subsong recorded on each training day. **(B)** An example of harmonic stack imitation where pitch error gradually decreased until a match to the model syllable was reached.
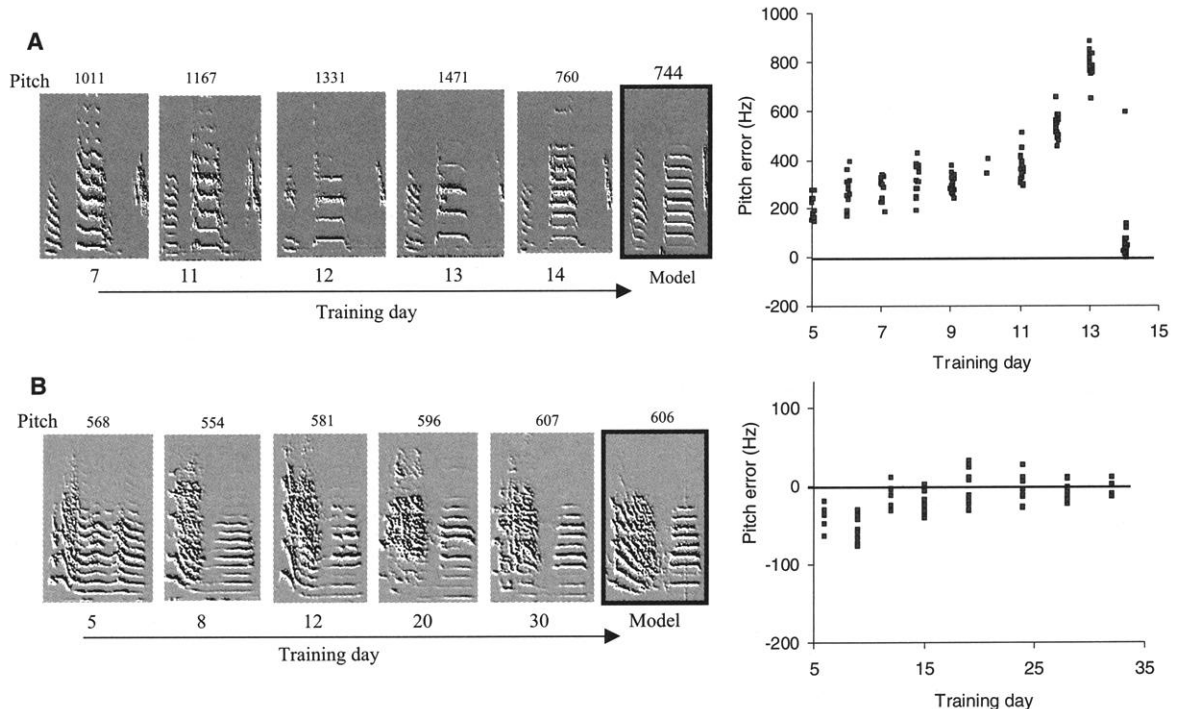
ues on the second day of training significantly different from those on day 1 (paired t test, $n = 42$, $P < 0.01$), they also differed from those of age-paired males kept under similar conditions but without exposure to a model song (t test, $n1 = 42$, $n2 = 12$, $P < 0.05$). We do not have statistical data on moment-to-moment changes across birds, but in one bird that was recorded continuously during the second day of training, we observed abrupt acoustic changes over a period of 3 hours (Fig. 3D). The changes of feature diversity for this bird from day 1 to day 2 were within the high end of the typical range (percentile = 26).
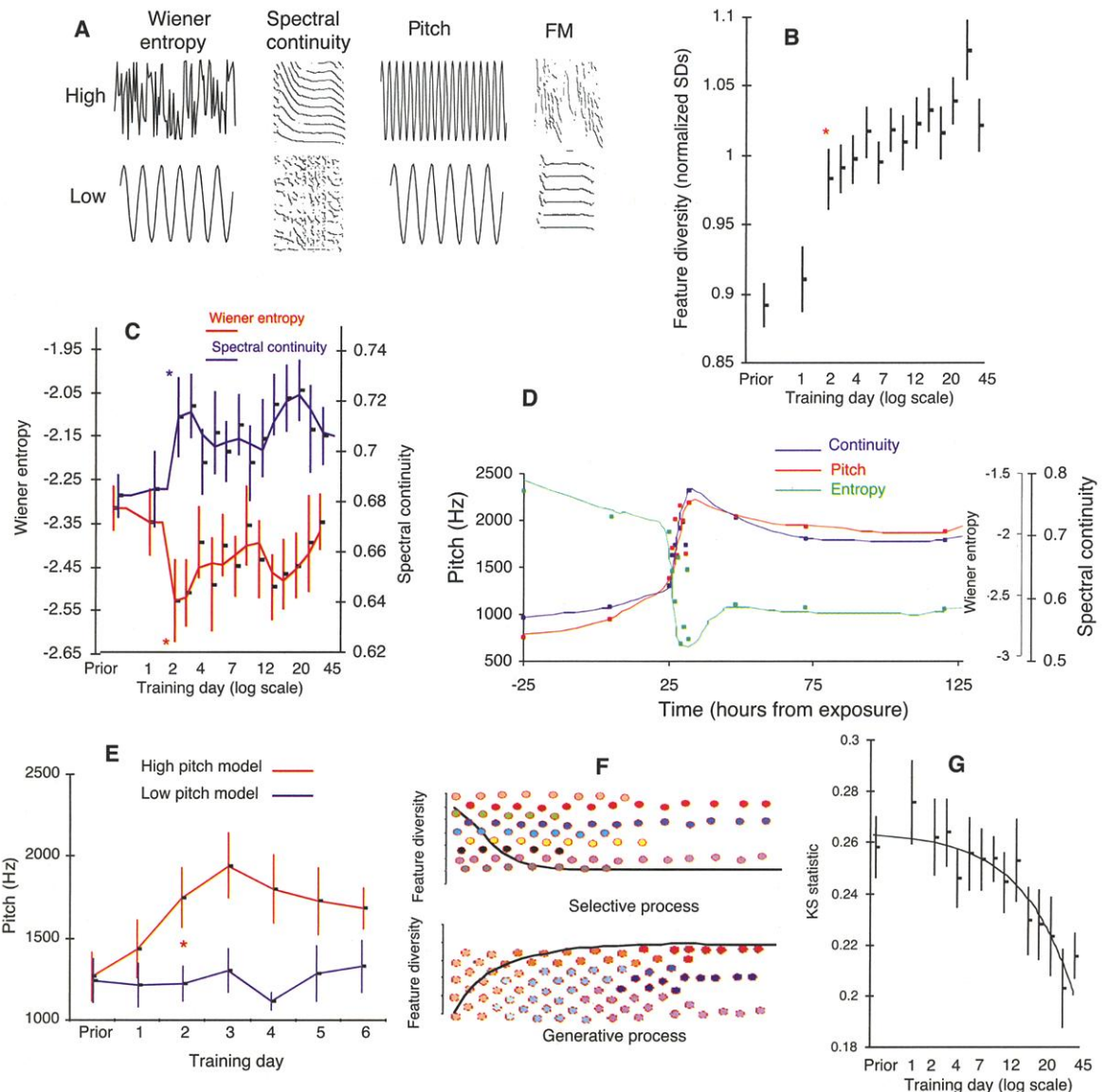
It could be that the transitions that occurred during the second day of training resulted from excitement caused by song stimulation, rather than from learning. To test for a model-specific effect, we trained 10 birds with a high-pitch (up

to 6086 Hz) model song and 10 birds with a low-pitch (up to 2567 Hz) model song. As shown in Fig. 3E, the high-pitch model induced the generation of higher pitch sounds than the low-pitch model as early as training day 2 (t test, $P < 0.05$), indicating that the early training-induced changes in acoustic structure were, at least partly, due to vocal learning. The first and largest changes in Fig. 3, B through D, occurred during the second day of training, after a night's sleep (20). Thereafter, feature diversity increased moderately but significantly (correlation coefficient $r = 0.65$, $P < 0.05$). We conclude that training birds with a model song induced the production of structured sounds. This effect is in line with a generative process (Fig. 3F), where structure emerges concurrently with a steep increase in the diversity of sounds produced. We refer to this period of song learning as the early generative phase.

Although the untutored subsong of any one individual exhibited low feature diversity, the feature distribution across individuals spanned a wide range of feature values [e.g., the subsong of five randomly selected untutored birds had a low mean feature diversity of 0.87; when these samples were pooled, they gave a high feature diversity of 0.97 (compare to Fig. 3B)]. We therefore wondered whether birds that (by chance) started with song features similar to those of the model imitated better than other birds. This was not so: There was no correlation between the KS statistic before training and the KS statistic when imitation was complete ($n = 42$, $r^2 = 0.06$). In other words, similarity to the song model before training did not predict the quality of final imitation.

Although the feature distributions showed marked changes during the first days of train-

Fig. 3. The effect of training on song features. (**A**) The four song features: Wiener entropy (a measure of "tonality" from pure tone to white noise), spectral continuity (a measure of the continuity of frequency contours), pitch, and frequency modulation (FM) (the change of pitch over time, e.g., frequency downsweep). (**B** through **D**) The effect of training with a model song across birds ($n = 42$); error bars represent SE uniformly throughout the panels. There are significant changes of feature diversity (**B**) and in Wiener entropy and spectral continuity (**C**) during the second day of training. An example of moment-to-moment changes in song features in an individual bird is shown in (D). (**E**) Exposure to a high-pitch model song versus exposure to a low-pitch model song. The high-pitch model induced higher pitch sounds in the bird's song starting on training day 2. (**F**) Processes of selection and generation can lead to similar outcomes. To distinguish between them, one needs longitudinal data. In the case of selection, feature diversity (represented here by the range of colors) decreases as imitation proceeds, whereas in the case of generation feature, diversity increases [see (B)]. (**G**) We used the KS statistic to trace the approximation of the developing song to the model song in terms of distances between the distributions of features. The KS statistic is 0.0 when a perfect match is achieved (e.g., when a song is compared to itself).

ing, they nevertheless did not move substantially closer to the corresponding distributions of the model song during that time (Fig. 3G). The following example clarifies this: The mean pitch of the subsong shown in Fig. 1B was 952 Hz, whereas the mean pitch of the model (Fig. 1D) was 1265 Hz. On training day 2, however, the mean pitch of the bird's song was 1933 Hz. The mean pitch on day 2, though higher than its initial value, was less similar to the model because high-pitch sounds were now disproportionately more frequent than in the model song. The slow changes in the KS statistic indicate that although the generative phase is intense, achieving accurate similarity to the entire model song is a slow process, to which we now turn.

Transition from Repetitive to Serial Production

Immelmann and others (4, 5, 21) found that syllables imitated from a model song appeared in subsong before the appropriate serial order of syllables was apparent. We observed that young zebra finches tend to produce back-to-back repetitions of similar sounds. As imitation proceeded, the birds produced a greater diversity of syllables delivered in a serial order, as in the model song. These events are reminiscent of the emergence in human infants of reduplicated (canonical) babbling (22–24) and the transition to variegated babbling (10, 25). It is therefore of interest to measure this transition and to examine how it might be affected by training the bird with a specific model song.

To measure the transition from repetitive to serial delivery, we estimated the median duration between two repetitions of sound, termed "period" (26). As shown in Fig. 4, this period increased during training from ~300 to 525 ms. We then partitioned the 10-s sample of song

Fig. 4. The transition from repetitive to serial production of sounds. The period of sound repetition (blue) increased during song development. The syllable duration (green) did not change consistently, but the number of syllables per period (red) increased throughout song development. Means ± SE (error bars); $n = 42$.

into syllables (27). As shown in Fig. 4, the mean number of syllables per period increased as well, from 1.5 to 3 syllables per period. Nevertheless, syllable duration did not change, indicating that the increase of period was not due to an increase in syllable duration, nor to the assembly of short syllables into long syllables. Rather, the increase in period reflected the emergence of longer sequences of different syllables. Despite the marked increase in song structure on training day 2, there was only a moderate (nonsignificant) increase in the number of syllables per period during the first week of training. This result is in line with our observation that the increase in song structure was initially confined to one or two types of syllables that the bird sang repeatedly. No significant increase in period was observed in control birds that were kept under similar conditions but that were not trained (28). Moreover, the song of some of our control birds (isolates) consisted of abnormal back-to-back repetitions of a same syllable type (29). We conclude that although sequential patterns of structured sounds appeared in some socially isolated birds, a transition from repetitive to sequential sound production was encouraged by exposure to a model song.

Sound Differentiation in Situ

We now turn to questions of how local transitions of acoustic features are linked to changes in the temporal organization of sounds and whether sounds of different types, such as harmonic stacks, high-pitch notes, and frequency downsweeps, emerge from primitive versions of each type. To assess syllable origins, we used an automated procedure to trace the development of specific imitations backward in time (15, 18) to reconstruct a likely trajectory of sound alterations that could explain the final outcome. Imitation trajectories of four birds are shown in Fig. 5. In bird A, we traced the imitation of two different syllables of a model song: The first syllable includes a high-pitch note, a vibrato, and a slightly modulated harmonic sound. The second syllable, similar to a "male long call" (5), starts with a harmonic downsweep that is followed by a nonmodulated harmonic stack. As shown, the imitation trajectories of these two syllables originate from repetitions of the same early "prototype" sound. The prototype sound was, in this case, more similar to the first syllable and was transformed markedly to give rise to the second one. Transformations occurred, apparently, while the relative positions of the sounds involved remained unchanged. We call this effect "sound differentiation in situ," to convey the notion that antecedents of the sounds of the adult song differentiated in their final temporal relation, with no translocations. Visual inspec-

tion of the sound spectrograms supported this interpretation.

In bird B (Fig. 5), we present an imitation trajectory of a different song syllable. The first part of this syllable consists of a nonmodulated harmonic stack; the second part consists of a sharp high-pitch sweep and a broadband sound. There is a short high-pitch note present in the bird's song on training day 35. There was no similar note in the model song, but the origin of the high note becomes clear when tracking it back in time. Although the first and second parts of the imitated syllable were quite different, they emerged by transforming two back-to-back renditions of the same prototype sound. Once again, sounds were differentiated in situ. Because the prototype was similar to the second part of the model syllable, developing the harmonic stack in the first part required major transitions. This result is counterintuitive, because untutored harmonic sounds could have provided excellent raw material to develop the imitation of the harmonic stack. This latter strategy appears to have been the one first adopted by bird D (Fig. 5) during training day 5. However, the harmonic stack present in that bird's song on day 5 was not generated in situ and was eventually abandoned. On training day 21, bird D started to generate another (rather inaccurate) version of the harmonic stack, but this time in situ, and it persisted into adulthood. This observation suggests that the laborious in situ differentiation of the harmonic sound was necessitated by constraints that hinder sound translocation and did not arise from a lack of "appropriate raw material" for generating this sound.

We examined in 10 birds the development of the syllable shown in Fig. 5, B through D. In all cases the harmonic stack developed in situ and the high-pitch sweep developed before the harmonic stack started to emerge. Therefore, the three examples (Fig. 5, B through D) are representative of how this syllable developed. We infer that some aspects of an imitation trajectory are stereotyped across birds. The in situ differentiation of two back-to-back renditions of a same prototype sound provides a mechanism of transition from the primitive, repetitive state to a mature state where sequences of dissimilar sounds are produced in a fixed order. During this transition, ensembles of sounds, either within or across syllables, differentiate in fixed sequential relations and in a chronological order that is idiosyncratic to that ensemble. We do not know if all types of sounds found in zebra finch song differentiate in situ, but clearly this is a common mechanism. In the current data set, we found at least one instance of in situ differentiation in 9 out of 10 birds (two groups of 5, each group tutored with a different song composed of three or four syllables). In five of these birds, there were
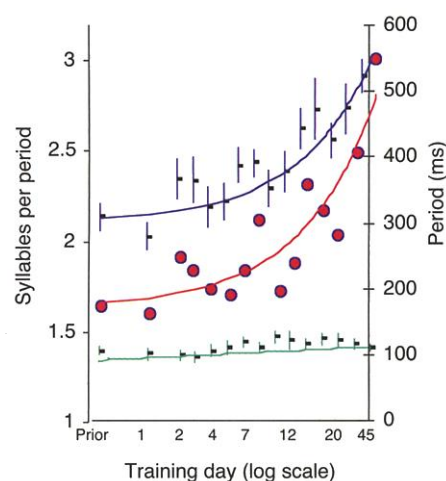
two instances of in situ differentiation. In addition, our observations suggest that syllables of different types need not emerge from primitive versions of each of these types (Fig. 5E).

A Revised View of Vocal Imitation

Zebra finches can imitate the song syllables that they first hear during the sensitive period for vocal learning (4, 5). We looked at how imitation proceeded when a model song was first presented at 43 days of age, when juvenile subsong is already in place. The phenom-

enology we describe might have been different had model songs been presented earlier or later. This caveat aside, it is clear that the trajectories for vocal imitation we encountered were neither arbitrary nor straightforwardly determined by acoustic differences between the model syllables and the subsong of a juvenile before training.

The song imitation process may face central and peripheral constraints. Constraints may emerge, for example, from the nonlinear peripheral dynamics of sound production, placing demands on subtle control of the

vocal apparatus if a complex sound is to be imitated. The period-doubling trajectories that we observed represent examples of how the song imitation process may take into account such nontrivial peripheral dynamics. Constraints shaping the imitation trajectory may also emerge from the need to integrate successive vocal gestures delivered with only brief or no intervening silent gaps—a contextual effect that perhaps favors in situ differentiation of sounds. In addition, central constraints may arise from the way in which forebrain song nuclei modulate the activity of

Fig. 5. Sound differentiation in situ. Imitation of specific sounds was tracked back from parts of the mature version of a bird's song (recorded on day 90 after hatching) until the automated procedure could no longer find a suitable match among earlier sounds. The automated tracking of a bird's imitation trajectory backward through time proceeds from the top to the bottom of the figure, and only samples of this tracking are shown here; the emergence of the bird's song is followed from bottom to top. The training day is indicated on each spectrograph. (**A**) Bird A generated two different syllables from successive renditions of a common prototype. (**B** through **D**) Three birds trained with a different model song than bird A generated two parts of a syllable from successive renditions of the same prototype. The imitation trajectories across birds are similar. The red arrow (bird B) indicates a remnant of the prototype that generated this sound. The blue arrow (bird D) indicates harmonic stacks that emerged during early stages of training that were eventually omitted. These harmonic stacks were not identified by the automated procedure, but by visual inspection. (**E**) Two models of song imitation from prototypes. For birds A through D, sounds of different types emerged from a common prototype, as schematized in the right panel.
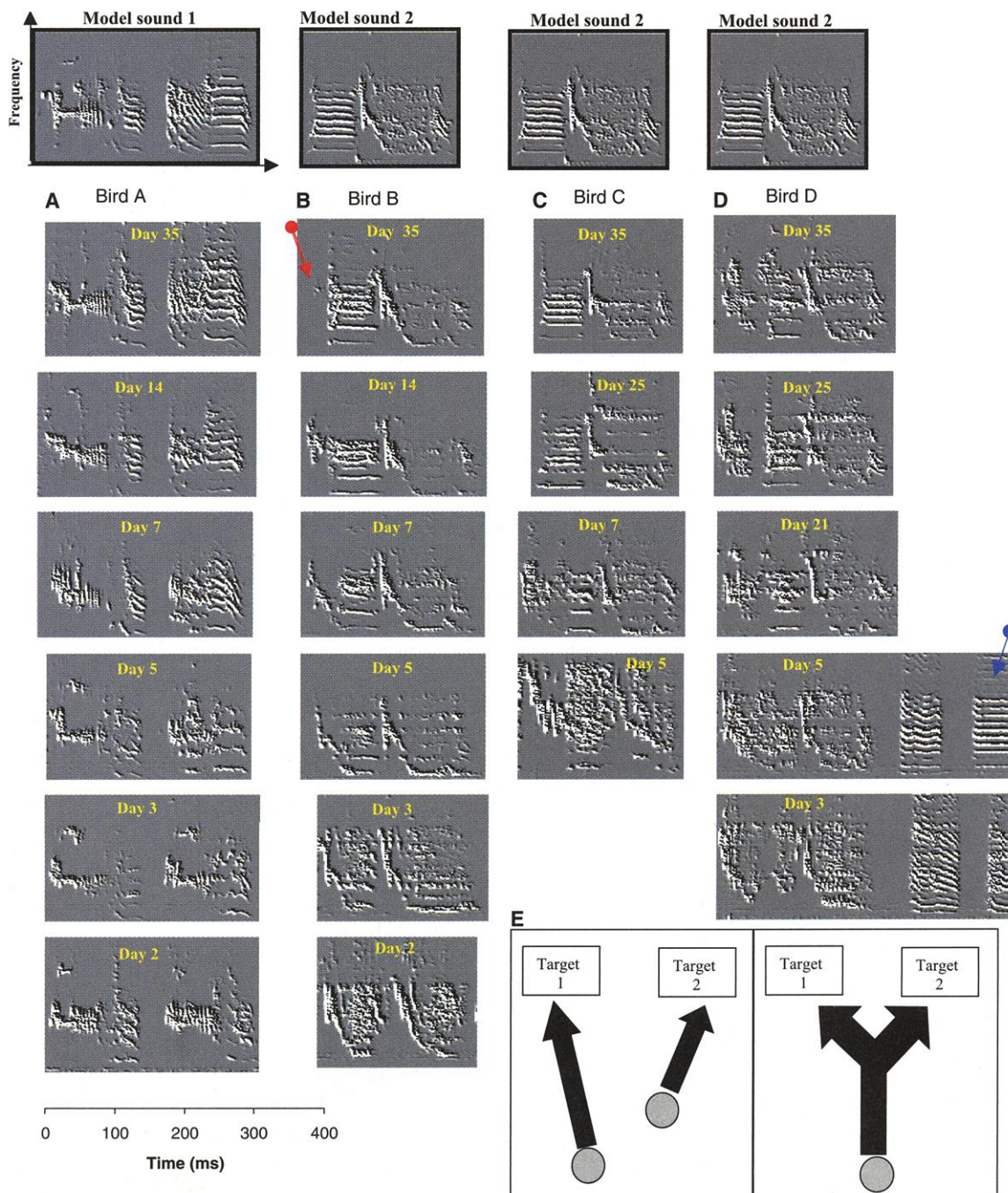
brainstem circuits that first evolved to produce simple, unlearned sounds or rhythmic behaviors.

We found that different final sounds in the song do not necessarily emerge from a primitive version of each of those types but may be generated from a same prototype. Much experimental work will be necessary to identify the variables that shape these imitation trajectories. In some instances, these trajectories can be indirect, leading us to propose that, in such cases, direct imitation trajectories might be costly in terms of the overall control effort required to arrive at the final sound ensemble. A systematic description of the diversity of prototype sounds and of the operations that birds perform to achieve a wide range of mature syllables may help clarify the contribution of innate and external factors shaping the imitation trajectory.

Although vocal learning remains a highly complex phenomenon, the tools that we used simplify the objective study of its dynamics. A fully automated recording system is now available (30) that can capture the entirety of vocal ontogeny and analyze changes in real time, paving the way for identifying the molecular, cellular, and circuit events that must underlie the moment-to-moment progression toward vocal imitation.

References and Notes

1. W. H. Thorpe, *Ibis* **100**, 535 (1958).
2. M. Konishi, *Z. Tierpsychol.* **22**, 770 (1965).
3. F. Nottebohm, in *The Design of Animal Communication*, M. Hauser, M. Konishi, Eds. (MIT Press, Cambridge, MA, 1999), pp. 63–110.
4. K. Immelmann, in *Bird Vocalization*, R. Hinde, Ed. (Cambridge Univ. Press, Cambridge, 1969), pp. 61–74.
5. R. E. Zann, *The Zebra Finch: Synthesis of Field and Laboratory Studies* (Oxford Univ. Press, Oxford, 1996).
6. O. Tchernichovski, F. Nottebohm, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8951 (1998).
7. L. A. Eales, *Anim. Behav.* **33**, 1293 (1985).
8. A. E. Jones, C. ten Cate, P. J. B. Slater, *J. Comp. Psychol.* **110**, 354 (1996).
9. T. W. Troyer, A. J. Doupe, *J. Neurophysiol.* **84**, 1204 (2000).
10. A. J. Doupe, P. K. Kuhl, *Annu. Rev. Neurosci.* **22**, 567 (1999).
11. Young males were raised by their mothers (no adult male present) until they were 30 days old. Each juvenile was then placed singly in a soundproof chamber that contained a plastic model of an adult zebra finch male (Fig. 1A). At 35 to 40 days after hatching, birds started to produce soft subsong, recorded for each bird on days 42 and 43. Training started on day 43. Within 36 hours, most birds (42 out of 50) began to peck at either one of two keys (31, 32); the 8 birds that failed to peck by that time were not included in the data. We provided two keys to encourage pecking activity (e.g., to overcome preferences to one side of the cage). Pecking either one of the keys induced the playback of the same short (1.4 s) span of song from a tiny speaker placed inside the plastic bird (32). Each playback consisted of two identical repetitions of a single song motif recorded from an adult bird. Training day 1 was taken as the first day the bird activated the song control keys. Each day consisted of two training sessions (32). In brief, during each session, we reinforced the first 10 key pecks with a song playback. Additional key pecks

were allowed but were not reinforced, so that the overall daily quota of model song that a bird could trigger was, at most, 28 s. We trained a total of 42 birds with one of four different model songs, and 12 additional birds were kept in the training apparatus as controls and not trained. A few minutes of song were recorded digitally (16 bits, 44.1 kHz) from each bird at least once a day during days 42 to 52 after hatching and at least once a week thereafter, until day 90 (by which time they produced stable song typical of adults).
12. O. Tchernichovski, F. Nottebohm, C. E. Ho, B. Pesaran, P. P. Mitra, *Anim. Behav.* **59**, 1167 (2000).
13. C. W. Clark, P. Marler, K. Beeman, *Ethology* **76**, 101 (1987).
14. M. S. Fee, B. Shraiman, B. Pesaran, P. P. Mitra, *Nature* **395**, 67 (1998).
15. Similarity measurements were performed as described in (12) and (18). Briefly, two sounds were considered similar if feature analysis for the two sounds yielded at least 90% similarity. In tracing imitation trajectories, we took a conservative approach and also verified the trajectories a posteriori by visual inspection. For example, a syllable from the mature song was compared to a random sample of 10 s of song recorded on an earlier day (e.g., day 83). If the similarity criterion was met and an earlier version of the sound was identified, then this earlier version was used as a point of entry for the next comparison with a still earlier sample of song, and so on recursively. This approach circumvents the need to partition juvenile sounds, because the similarity section automatically provides boundaries. Although the procedure could trace similarity of one sound to two or more earlier fragments (e.g., to test for sound translocation), robust implementation of this capability would require continuous recording, which was not done in the present study. To ensure that imitation trajectories were reproducible, we repeated the procedure using different 10-s samples of sound from each day that song was recorded. We confirmed that an imitation trajectory was reproducible by calculating the similarity across the two independently identified antecedents. We maintained confidence in the trajectory as long as the significant similarity between antecedent versions on a same day was above 80% [(18), pp. 43–45].
16. Harmonic stacks are composed of nonmodulated harmonic frequencies, where the fundamental frequency defines the pitch. Harmonic stacks were traced automatically as described (12, 15, 18). Even though period doubling causes a discontinuous change of pitch, tracing is still possible with the other features. Pitch was measured in the middle of the traced sound.
17. When the initial pitch (at the beginning of the trajectory) was lower than the target, reaching the first harmonic would require that the fundamental be crossed first, matching the pitch to the model sound before period doubling could occur, which then would seem to make period doubling unnecessary.
18. O. Tchernichovski, P. P. Mitra, *Sound Analysis User Manual*, version 1.01 (ed. 2, 2000) (available at http://talkbank.org/animal/sa.pdf).
19. The feature diversity $d(t)$ is a scaled average (across birds and features) of the standard deviations $\sigma_\alpha(t, b)$ of song features $\alpha = 1, \ldots, 4$, computed for a given bird b and for a given time epoch t (t denotes the day of the 10-s sample) with the samples of the features involved during that 10-s epoch. The SDs for different features have different scales; this is taken care of by defining a scale for each feature by pooling all observations for that feature across birds b and measurement windows t

$$\bar{\sigma}_\alpha = \frac{1}{TN_b} \sum_t \sum_b \sigma_\alpha(t, b)$$

where T is the total number of epochs and N_b is the number of birds. For each training day we compute

$$d(t) = \frac{1}{4N_b} \sum_{\alpha=1}^{4} \sum_{\beta=1}^{N_b} \frac{\sigma_\alpha(t, b)}{\bar{\sigma}_\alpha}$$

The SE of feature diversity is computed for a given epoch by treating the birds as different statistical samples. The error bars on Fig. 3C are SEs of the mean feature values. To obtain these, we computed the mean feature values for individual birds for a given epoch; the grand mean of these individual means is then plotted. The SEs are computed as before by treating the birds as different statistical samples. This measure should not be confused with the feature diversity, which denotes the spread of feature values within individual birds. The two are not, in principle, related.
20. A. S. Dave, A. C. Yu, D. Margoliash, *Science* **282**, 2250 (1998).
21. H. Hultsch, D. Todt, *Anim. Behav.* **44**, 590 (1992).
22. R. E. Stark, in *Child Phonology*, G. H. Yeni-Komshian, J. F. Kavanagh, C. A. Ferguson, Eds. (Academic Press, New York, 1980), pp. 73–92.
23. D. K. Oller, in *Child Phonology*, G. H. Yeni-Komshian, J. F. Kavanagh, C. A. Ferguson, Eds. (Academic Press, New York, 1980), pp. 93–112.
24. M. M. Vihman, in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, B. de Boysson-Bardies *et al.*, Eds. (Kluwer, Dordrecht, Netherlands, 1993), pp. 411–419.
25. L. Roug, I. Landberg, J. Lundberg, *J. Child Lang.* **16**, 19 (1989).
26. Period was calculated by implementing the similarity measurement (15) as follows. Starting from a randomly selected frame within a 10-s epoch, the time frame is first moved forward from the point of origin until it encounters a frame that is less than 75% similar to the sound at the selected frame, so as to depart from the original sound. The time window is moved further forward, until it encounters a sound that is at least 90% similar to the original frame. This gives an instantiation of the time elapsed between two similar sounds. This procedure is repeated 50 times, each time starting from a different randomly selected frame within the 10-s epoch, and the median of the durations thus identified is a measure of periodic structure in the song. We call this the period.
27. To characterize the overall temporal structure of the juvenile song with reproducible statistical results, we partition sounds to syllables (continuous sounds) based on "silent intervals," where no frequency contours are detected (12). Although the statistical estimate of the number of syllables is reproducible, the units themselves (syllables) are not equally well defined in subsong and song and would not be suitable for automated categorization.
28. To test for a possible effect of training on the transition from repetitive to serial delivery of syllables, we raised 12 control birds without exposure to the model song under housing conditions otherwise identical to those encountered by the birds that were trained. We found that the period of the mature song in the control birds was significantly shorter than the final period in the trained birds (366 ± 66 ms versus 524 ± 26 ms, $P < 0.01$) and was within the range of the period soon after the onset of training.
29. P. Price, *J. Comp. Physiol. Psychol.* **93**, 260 (1979).
30. Our software for automated recording, song recognition, and real-time analysis of the imitation process, as well as the raw sound data, are available (at no charge) at http://talkbank.org/animal.
31. P. Adret, *Anim. Behav.* **46**, 149 (1993).
32. O. Tchernichovski, T. Lints, P. P. Mitra, F. Nottebohm, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12901 (1999).
33. We thank M. Konishi, M. Schmidt, P. Marler, and the reviewers for their useful comments; the staff of The Rockefeller University Field Research Center for their technical support; and the TalkBank project for providing Internet storage. Supported by U.S. Public Health Service (PHS) grant DC04722-01 (O.T.), PHS grant MH18343 (F.N.), the Mary Flagler Cary Charitable Trust, the Herbert and Nell Singer Foundation, and the Phipps Family Foundation.