ANALYSIS OF SONG DEVELOPMENT

O. Tchernichovski · P.P. Mitra

Towards quantification of vocal imitation in the zebra finch

Abstract The transition from an amorphous subsong into mature song requires a series of vocal changes. By tracing song elements during development, we have shown that the imitation trajectory to the target could not be predicted based on monotonic progression of vocal changes, indicating an internal component that imposes constraints on song development. Here we further examine the nature of constraints on song imitation in the zebra finch. We first present techniques for identifying and tracing distinctive vocal changes, and then we examine how sequences of vocal change are expressed and coordinated. Examples suggest two types of constraints on song imitation, based on the nature of the temporal context. *Developmentally diachronic* constraints are imposed by sequential dependencies between vocal changes as a function of developmental time, whereas *developmentally synchronic* constraints are given by the acoustic context of notes within the song. Finally, we show that the tendency of birds to copy certain sounds in the song model before others might be related to such constraints. We suggest that documenting the full range of distinctive vocal changes and the coordination of their expression would be useful for testing mechanisms of vocal imitation.

Keywords Birdsong · Development · Sound analysis · Vocal learning · Zebra finch

Abbreviations *BOS* bird's own song · *DVC* distinctive vocal change

O. Tchernichovski (✉)
Department of Biology,
The City College of the City University of New York,
138 and Convent Ave, New York NY 10031, USA
E-mail: ofer@ccny.cuny.edu
Tel.: +1-212-6508540
Fax: +1-212-6508959

P.P. Mitra
Department of Theoretical Physics,
Bell Laboratories, Lucent Technologies

Introduction

At about 30 days post hatch, the juvenile zebra finch starts producing the soft, squeaky sounds of its subsong. Within several weeks, these ill-formed sounds transform markedly, as the bird masters the complex sounds of an adult male's tutor song. This process exhibits several stages (Immelmann 1969) and requires auditory feedback (Konishi 1965), suggesting that the bird compares the sounds of its own song (bird's own song, BOS) to the auditory memory of a song model it is attempting to imitate (Brainard and Doupe 2000). Some aspects of this sensory-motor process can be studied on a behavioral level: observations of song development show that the bird can approximate a song model via a generative procedure, by transforming the acoustic features of sounds (Thorpe 1958; Immelmann 1969; Clark et al. 1987; Tchernichovski et al. 2001). In addition, the bird can approximate model sounds by narrowing its vocal repertoire via a selective procedure so as to eliminate sounds that do not match a template (Marler and Nelson 1992).

Vocal changes, however, are not fully determined by the acoustic error between the BOS and the model, since there is also an internal component of song development that must be taken into account: First, some vocal changes are hardwired and occur regardless of training (e.g., in socially isolated birds; Price 1979), and second, even vocal changes that are specific to the imitation of a particular model are sometimes counterintuitive, such that the imitation trajectory appears to be indirect (Tchernichovski et al. 2001).

In this study, we take the behavioral analysis one step further, and make a preliminary attempt to atomize the generative processes underlying song learning in a few simple cases. We then explore the elementary units of vocal changes and examine sequences of vocal changes to uncover constraints that shape (and could potentially predict) imitation trajectories. To gain statistical power for detecting interactions between environmental and

internal factors during song imitation, we limit our analysis to the imitation of a single song model across different training regimes. Using this approach we can compare vocal changes across birds that copy the same sounds.

Materials and methods

In this study we combined new data with results obtained and techniques used in several previous studies (Tchernichovski et al. 1999, 2000, 2001; Tchernichovski and Mitra 2002). Most of the methods are already published and are presented here briefly with appropriate references.

Animals

The data presented here are from young male zebra finches that were raised at the Rockefeller University Field Research Center and were kept under constant (12:12) photoperiod and food ad libitum throughout the experiments.

Experimental groups

In Figs. 2, 3, 4, and 5) we present data from eight birds that received the same training regime as in an earlier study (Tchernichovski et al. 2001). Young males were raised by their mothers (no adult male present) until 30 days old. Each juvenile was then placed alone in a soundproof chamber that contained a plastic model of an adult zebra finch male (Fig. 1A). Training started on day 43, when we inserted two keys into the training box (Fig. 1B; Adret 1993). Within 1–36 h birds began to peck at either one of the two keys. Key pecking induced the playback of the same short (1.4 s) model song from a tiny speaker placed inside the plastic bird (Fig. 1C). Each playback consisted of two identical repetitions of a single song motif recorded from an adult bird. Each day consisted of two training sessions. During each session we reinforced the first ten 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 40 song motifs (28 s). Training continued until birds were 90 days old, when song development was complete. In Fig. 6(see Results: Statistical analysis of imitation priorities) we present data from three different training groups.

Brief training

In nine birds the training regime was similar to that described above, except that birds were trained only for 2 consecutive days: keys were inserted to the training box on day 43 post-hatch. Birds were trained for 48 h, starting from the time of the first key peck (which occurred within less than 36 h from the time of inserting the keys). This gave an overall exposure of less than 1 min to song playbacks (four training sessions of 20 song motifs). The keys were then removed and birds were kept socially isolated in the same training boxes until 90 days old, but received no further training.

Moderate training

Six birds were trained continuously from day 30 and until day 100 post-hatch. The training regime was similar to that described above (20 song motifs per session). Compared to other training regimes tested so far, moderate training gave the most accurate imitation. Data from Tchernichovski et al. (1999).
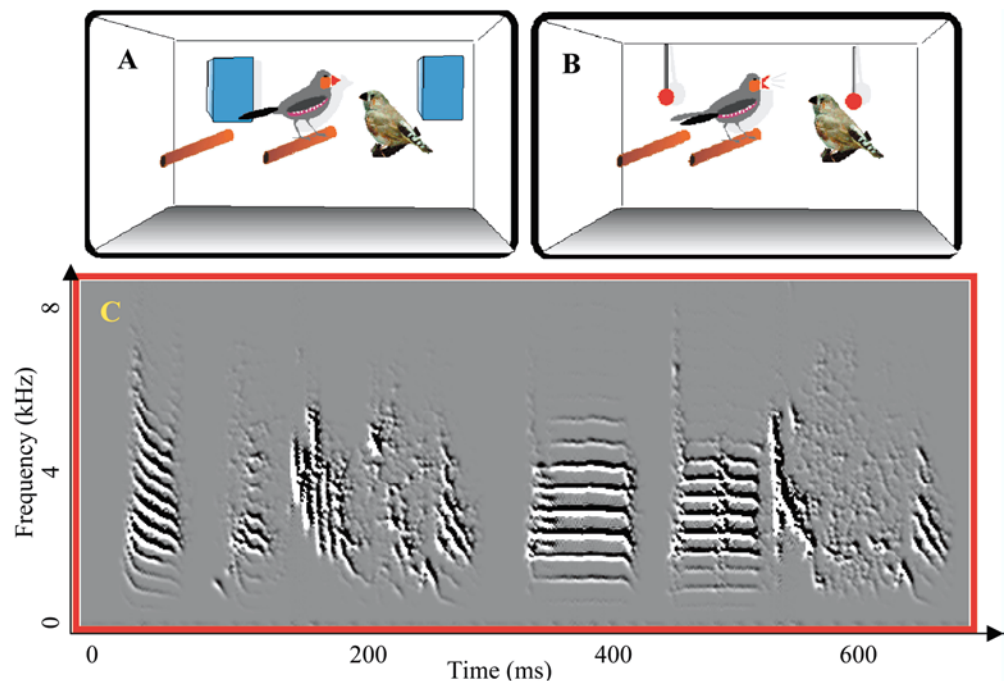
Overtraining

Eight birds were trained continuously from day 30 until day 100 post-hatch, reinforcing an unlimited quota of key-pecks, i.e., every key-peck induced a song playback. For reasons unknown, over-training inhibits song imitation in zebra finches. Data from Tchernichovski et al. 1999).

Song recording

All songs were recorded digitally with a sampling rate of 44.1 kHz and 16-bit accuracy. In the birds represented in Figs. 2, 3, 4, and 5) we recorded a few minutes of song from each bird at least once a day during days 42–52 after hatching and at least once a week

Fig. 1 A The training apparatus included a plastic model, similar in size and coloration to an adult zebra finch, placed in the middle of the box. B Training starts when two keys are inserted to the box. By pecking at either of the two keys, the young bird could induce a song playback from a tiny speaker placed inside the plastic model. C Spectral derivatives (see Materials and methods) of the song model. In subsequent figures specific parts of this model are circumscribed by a red rectangle

thereafter until day 90. In the three groups of birds represented in Fig. 6 only the mature song was recorded on day 90 post hatch.

Data analysis

Spectral derivatives

Spectral derivatives provide a representation of song that is similar but superior to the traditional sound spectrogram (Fig. 1C). Instead of power spectrum versus time, we perform multi-taper spectrum analysis and estimate spectral derivatives (changes of power) in both time and frequency axes. From time and frequency derivatives we compute directional derivatives which are locally optimized so as to detect frequency contours of any arbitrary angle in time-frequency plan. Directional spectral derivatives are plotted on a gray scale, where negative values are dark and positive values are bright (Tchernichovski et al. 2000).

Similarity measurements

We measured similarity between songs using the default parameter setting of Sound Analysis 2 software (Tchernichovski et al. 2000; Tchernichovski and Mitra 2002). The procedure can compare any two song motifs (e.g., song model versus BOS) and detect sections of similarity across the two songs based on four song features: pitch, FM, Wiener entropy, and spectral continuity. The similarity indices used here are all 'percentage of significant similarity', which is an estimate of the proportion of sounds in the song model to which a similar version may be found in the BOS. In the data presented in Fig. 6B, F we present a breakdown of similarity across single notes: to do so, we categorized a note as similar to the model if at least 50% of it was included in a similarity section.

Automatic tracing of imitation trajectories

To automatically trace imitation trajectories we first compared the mature version of the BOS to the song model. Song sections that were similar to the model were then traced backwards by comparing the mature version to an earlier version (e.g., a day earlier) and so on recursively as long as the procedure could keep track of the similarity. This procedure does not require partitioning of the developing song prior to analysis (the similarity measurement is performed on the entire song, and detects sections of similarity, which are then compared to the entire version of the song in a previous day, and so on). Note that the procedure can only trace relatively short sections of similarity within a song – tracing the entire song motif in a single section is rarely possible; in general, the larger the section traced, and the younger the bird, the more likely the procedure to fail (Tchernichovski and Mitra 2002). To quantify the vocal changes from one hour to the next (Fig. 2D) we examined songs at the beginning of each hour and sampled the first 12 renditions of the appropriate syllable.

Results

Distinctive vocal changes

When examining different developmental stages of a song, it is sometimes possible to automatically detect sections of similarity across versions and to align them during an extended developmental time, as illustrated in Fig. 2A. The red rectangle presents a case where the boundaries of a vocal change can be clearly detected. When the scope of a vocal change is limited in both song time and developmental time, we call it a distinctive vocal change (DVC).

Figure 2B presents a very simple vocal change that (perhaps not coincidentally) occurred in a bird that produced an inaccurate imitation of the model song. As shown, a single prototype sound of long duration gave rise to two harmonic stacks, similar to those of the song model (sounds c and d in the Fig. 2B). Examining this imitation trajectory from the bottom to the top of the figure shows that on day 52 post-hatch, a localized attenuation appeared at the middle of the prototype. The attenuation then expanded until it became a stop (day 60), splitting the sound into two. This example indicates that syllable boundaries can change during song development, and other examples will show (Fig. 5) that the acoustic type of sounds can also change during song development. It is therefore very difficult to partition the emerging song into reproducible syllable units of various types. Here we present a complementary approach of partitioning song development ''vertically'' into sequences of vocal changes. Note that identifying units of vocal change requires a smooth transition from the primitive to the mature version of a song interval, but it does not require classification of sounds in the emerging song.

The imitation trajectory described above is an example of a DVC, as it has a *limited scope* in both song time and developmental time: We can identify the boundaries of the change by tracing the drop in the amplitude envelope. To estimate the attenuation we divide the syllable into three equal parts, measure minimum and maximum amplitude for each part and define:
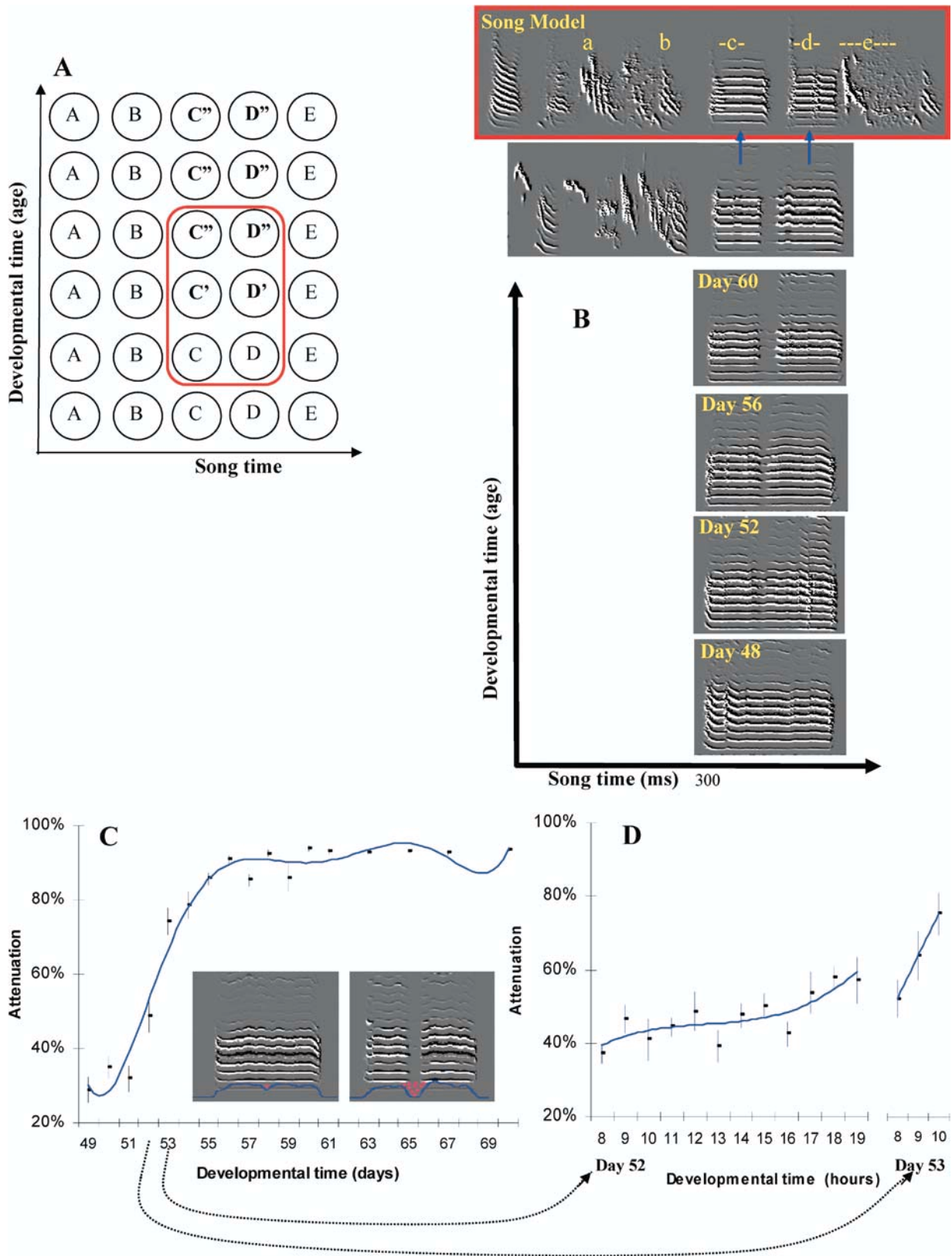
$$\text{Attenuation } (\%) \\ = 1 - \frac{Minimum\,amp\,(middle\,part)}{Average\,\{Max\,amp\,(first\,part),\,Max\,amp\,(last\,part)\}},$$

(1)

which is simply the ratio between the depth of the drop (minimum amplitude of the middle part) and the mean maximal amplitudes of first and last parts. We can now estimate the rate of the change by measuring the depth of the drop from day to day (Fig. 2C) and from hour to hour (Fig. 2D). As shown, the attenuation progressed slowly during the afternoon of day 52 post-hatch and rapidly during the morning of day 53.

In addition to the *limited scope*, there are two features that add interest and generality to this vocal change: First, this was not an idiosyncratic change, but a step in a learning process, as it eventually led to an approximation of model sounds. Second, the acoustic feature that has changed (amplitude) can be studied in specific mechanistic terms: attenuation can be achieved, for example, by reducing air pressure or by briefly obstructing the upper airways (similar to inserting a 'stop consonant' in human speech).

Developmentally diachronic constraints

The example discussed above is of an isolated vocal change. However, the song-imitation process involves

many vocal changes, some of them related. The imitation trajectory we have just examined (Fig. 2B–D) is in fact more complex than the description given above suggests. As we described, the bird inserted a stop, which generated two indistinguishable harmonic stacks. Subsequently, the bird started to increase the amplitude of the second sound until it became about 50% louder than the first sound, so that a continuous prototype was

◀

Fig. 2 **A** an illustration of a song interval composed of different sounds, which can be automatically aligned during an extended developmental time as long as vocal changes are gradual. The *red rectangle* shows the boundaries of a distinctive vocal change (DVC). Note that in practice, the interval of sound is analyzed continuously and the symbols stand for aligned FFT time windows of sound rather than 'syllables'. **B** An automatically traced imitation trajectory leading to a copy of the model harmonic stacks denoted as *c* and *d*. The bird inserted a stop, splitting a prototype harmonic stack into two sounds. **C** A quantitative tracing of the insertion of stop as it progresses from one day to the next. The *blue curve* represents the change in the depth of the attenuation (see *blue amplitude curve in the spectral image*). Each data point represents the mean value for 12 renditions of the syllable and the error bars represent the standard error (SE). **D** A quantitative tracing of the insertion of stop as it progresses from one hour to the next during the rapid phase of the vocal change (days 52 and 53 post-hatch). Each data point represents the mean value for 12 renditions of the syllable and the error bars represent SE

eventually transformed into two distinguishable sounds. Figure 3A presents a 3-D image of the change in amplitude envelope during those vocal changes, suggesting that the amplitude of the second harmonic stack started increasing only after the stop was completed. To examine the dynamics of this amplitude modulation we use the same measurements as in Eq. 1, namely dividing the syllable into three equal parts, measuring minimum and maximum amplitude for each part and defining:

$$relative\ amplitude = \frac{Max\ amp\ (last\ part)}{Max\ amp\ (first\ part)}, \quad (2)$$

which gives a value of 1 when there is no amplitude difference, and more than 1 when the second sound is louder than the first.

As shown in Fig. 3B, the two vocal changes had very different time-courses: the insertion of stop occurred within a few days, whereas the amplification of the second sound progressed slowly over 20 days or so. Furthermore, the data suggest that the onset of amplification occurs only after the stop was inserted: For example, on day 53, 10:00 a.m. (Fig. 2D) the mean attenuation was 69–81% (95% confidences interval, $n = 12$ renditions) indicating a statistically significant attenuation, and the mean relative amplitude was 0.9–1.14 (95% confidence interval, $n = 12$ renditions) which does not suggest a real amplitude modulation. To further validate that the onset of amplification occurs only after the stop was inserted we plot for each syllable the attenuation (Eq. 1) versus the amplification (relative amplitude, Eq. 2). Figure 3C presents the plot for all data sampled during the vocal change (on days 50–60). As shown, relative amplitude values are distributed evenly around 1 until attenuation levels are above 80% (which is virtually a stop). In other words, a tendency to increase the relative amplitude of the second part is detectable only in syllables where the insertion of stop is at a very advanced stage or complete.

It seems likely that the outcome of the first vocal change is relevant (and perhaps even necessary) for
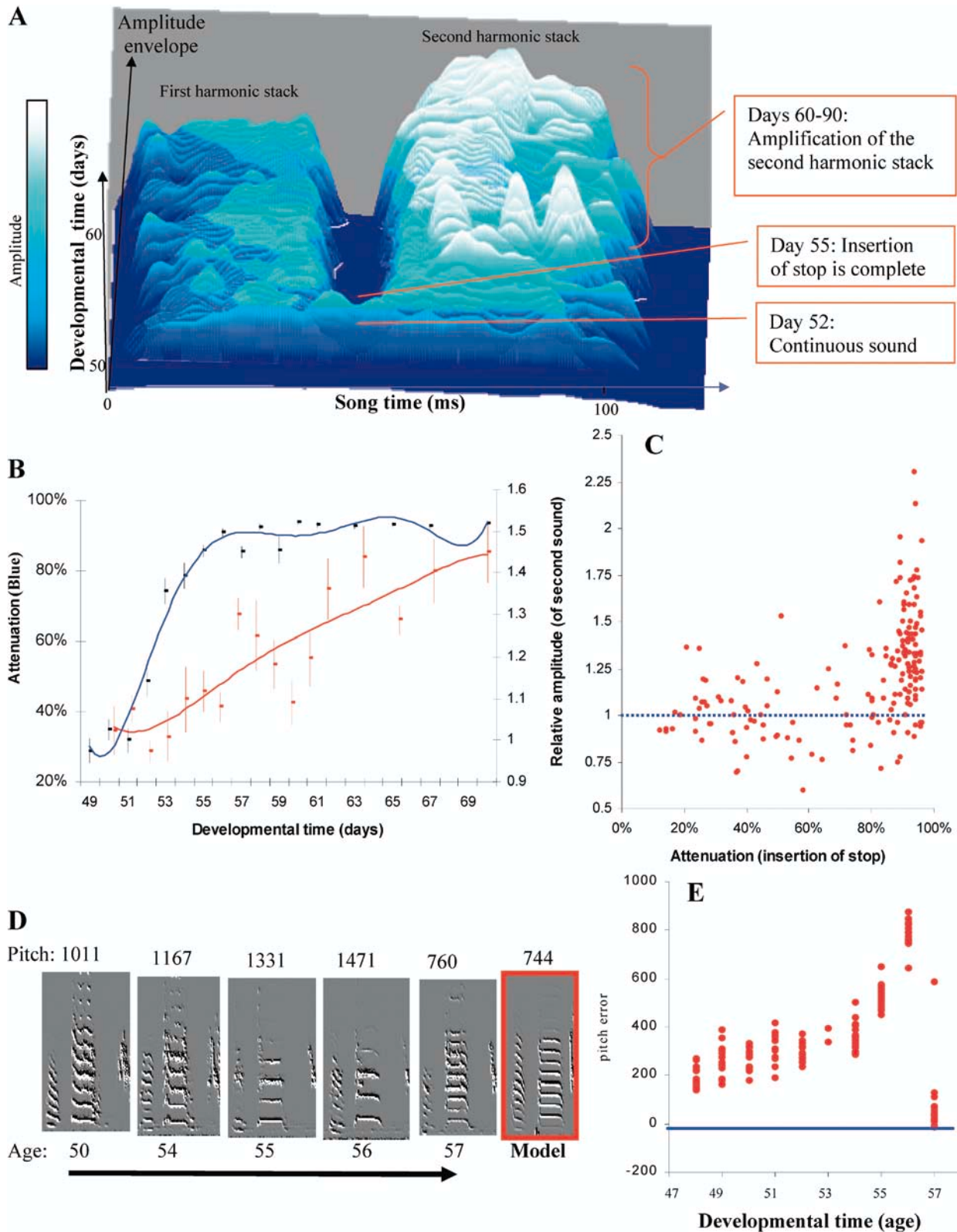
activating the next vocal change. When a vocal change appears to be contingent on the completion of a different vocal change, we call it a *developmentally diachronic constraint*. Such constraints are manifested in the imitation trajectory as a function of developmental time (the *y*-axis of Fig. 4A). For example, when one vocal change (blue rectangle, c, d→C, D) produces the input of the next vocal change (red trapezoid, D→D"), we say that the first vocal change is a developmentally diachronic constraint on the second one (Fig. 4A).

Overall, song becomes more structured during development and indeed, each of the vocal changes described above (insertion of stop and amplitude modulation) added some structure to the song: the first generated two sounds out of one, and the second vocal change made them distinguishable. The second vocal change, however, did not approximate the model: in the song model, the amplitude of the first harmonic stack was higher than that of the second and hence, by amplifying the second harmonic stack the bird reduced the similarity to the song model. A similar phenomenon is also apparent in the next example:

The correction of pitch error via the period-doubling trajectory presented in Fig. 3D, E is another example of developmentally diachronic constraint (from Tchernichovski et al. 2001). In this case, the first vocal change gradually increased the pitch, drifting it away from that of the model sound until it was twice the pitch of the model sound. At that point, a second vocal change halved the pitch in a single step (period doubling), matching it to that of the model. Here also, the gradual increase of pitch is likely to be a developmentally diachronic constraint required for the expression of period doubling. If the period doubling event must wait for the conclusion of the pitch increase, it might be possible to prevent the period doubling by delaying the learning. For example, after the age of 90 days song structure does not change much, if at all. The question is: had the first vocal change (the gradual increase of pitch) failed to reach the first harmonic of the song model by day 90, would the pitch 'freeze' on that level? This has not been investigated yet, but we have seen a few cases of slow-learning birds in which period doubling failed to occur, and the pitch remained almost twice as high as that of the model.

Developmentally synchronic constraints

The challenge of song imitation is not only of changing sounds appropriately, but also of incorporating those changes into the emerging song structure. Consider, for example, a vocal change that expands the duration of a sound: D→D" (red trapezoid, Fig. 4A). This vocal change could be constrained by the need to shift the neighboring sounds (and of increasing the duration of the song motif). Constraints on vocal changes in *song time* (*x*-axis) are defined as developmentally synchronic constraints, as opposed to developmentally diachronic

A

Amplitude envelope

First harmonic stack

Second harmonic stack

Days 60-90: Amplification of the second harmonic stack

Day 55: Insertion of stop is complete

Day 52: Continuous sound

Amplitude

Developmental time (days)

Song time (ms)

B

Attenuation (Blue)

Developmental time (days)

C

Relative amplitude (of second sound)

Attenuation (insertion of stop)

D

Pitch: 1011   1167   1331   1471   760   744

Age: 50   54   55   56   57   Model

E

pitch error

Developmental time (age)

constraints we discussed earlier, which act in *developmental time* (*y*-axis).

We will illustrate this by presenting an interesting case (Fig. 4B): Note that the model song is shown twice, aligned with the start of the pupil's song on day 50 (bottom of figure) and with the end of the pupil's song on day 95 (top of figure). As indicated by the arrows, by day 50 post-hatch a bird had already copied a raw version of the beginning and end of the song model but has not yet copied the harmonic stacks c. In order to copy

◄

Fig. 3 A A 3-D image of the change in the amplitude envelope during and after inserting a stop. To eliminate the effect of proximity to the microphone, we normalized each amplitude curve by the mean amplitude during the first 100 ms of the sound. As shown, the relative amplitude of the second call increases after the stop until it became 50% higher than that of the first call. **B** A quantitative tracing of the amplitude modulation (*red line*) in reference to the insertion of stop (*blue line*, both curves are 5th order polynominal smoothing). **C** A plot of attenuation (insertion of stop) versus amplitude modulation for each syllable sampled throughout the period of vocal change (days 50–60). **D, E** Period-doubling trajectory: the pitch error (the difference between bird's pitch and model pitch) increases smoothly until it reach the first harmonic of the song model, and is then corrected abruptly by period doubling (from Tchernichovski et al. 2001)

sound c in its proper temporal context, the bird must open a gap (insert a time slot) and generate an additional sound in the middle of its emerging song; otherwise, it would overwrite an existing match. This requirement is imposed by the song structure present on day 50, and is therefore a *developmentally synchronic constraint* on imitating that harmonic stack: As shown in Fig. 4B, between days 50 and 95, the bird 'stretched' the duration of the chirp denoted as b, and transformed it to the harmonic stack denoted as c (b→c). That is, the bird has transformed a copy of model-sound b to a copy of model-sound c. Note that the alignment of the sound with the song model had changed while it was transformed, as indicated by its alignment in reference to the end of the song model (top of figure) as opposed to its alignment in reference to the beginning of the song model in the earlier version (bottom of figure). During the transformation, the sound was not only stretched (time warped) but was also separated from the syllable just prior to it, as all the sounds to the left of the vocal change were shifted farther to the left, opening a 50-ms time slot that gave room for the time warping.

Thus, in order to transform b→c the bird had shifted neighboring sounds, overcoming a developmentally synchronic constraint on the time warping. With regards to overwriting existing sounds, the interpretation is not straightforward, but we would like to make two comments. The location of a mismatch between the emerging song (on day 50) and the song model depends on how they are aligned: Aligning from the beginning gives good match between the two songs until we reach sound e. Aligning from the end, however, gives a good match until we reach model sound c. Therefore, overwriting c (as opposed to e), suggests that the bird detected a mismatch by aligning its song in reference to the end of the song model. If this hypothesis is true, than the bird could not have detected the mismatch in real time (Margoliash 2002). Second, it is not clear if the overwriting reflects a failure to insert a sufficiently prolonged time slot, a failure to generate an additional sound, or a failure to detect that it is overwriting.

Note that the order of mastering sounds (in developmental time) can determine the nature of developmentally synchronic constraints in song time. For example, if instead of copying the beginning and the end

of the song model first and then attempting to copy the middle (AC→ABC) the bird had copied the beginning and the middle first and then attempted to copy the end (AB→ABC) there would have been no need for opening a time slot and for shifting sounds. The sensitivity of developmentally synchronic constraints to the order of copying model sounds is the subject of the next sections.

Analysis of imitation trajectories across birds

We have seen in one bird (Fig. 4B) that changing the position of sounds in song time might be subject to developmentally synchronic constraints. What developmental factors could dictate that an initial position of a sound differs from its target position in song time? Figure 5A presents a case of a bird (pupil 1) that copied all the model sounds in good approximation to their position in the song model, whereas Fig. 5B shows another case (pupil 5) where a copy of model sound *a* (vibrato, red arrow) appears in a position that differs from that of the model. Why is the position of the vibrato correct in one bird and incorrect in another bird?

Tracing imitation trajectories of model sounds d and e showed similarities in trajectories of 10 out of 12 birds that were trained with the same song model (Tchernichovski et al. 2001). Figure 5C presents five examples of such trajectories. As shown, the prototype sounds (broadband downsweeps) are similar across birds. In each case, a broadband noisy sound was transformed into a copy of the model harmonic stack d. The fate of the high-pitch sounds, however, varied across birds: pupil 3 deleted the first high pitched sound very late (red arrow). Pupils 1 and 2 deleted the first high-pitched sound earlier, and pupils 4 and 5 did not delete the first high-pitch sound at all, but transformed it into a vibrato. Note that the final copy of sounds d and e in birds that retained the first high pitched sound (pupils 4 and 5) are similar to immature versions of those sounds in birds that deleted that high-pitched sound (e.g., compare to pupil 2, day 55).

We propose that retaining this high-pitch sound and transforming it into a vibrato imposes developmentally synchronic constraints that cannot be fully resolved, and which can explain why pupils 1, 2 and 3 chose to delete that high-pitch prototype. In other words, we suggest that the decision of how to change one sound imposes constraints on the imitation of other sounds later on, and that taking those constraints into account may help the bird achieve a good imitation.

Statistical analysis of imitation priorities

The cases presented in Fig. 5C suggest that imitation trajectories of different birds share common features. As we suggested, similarity in vocal imitation across birds could stem from developmental constraints, but it could also stem from (or be modulated by) perceptual biases,
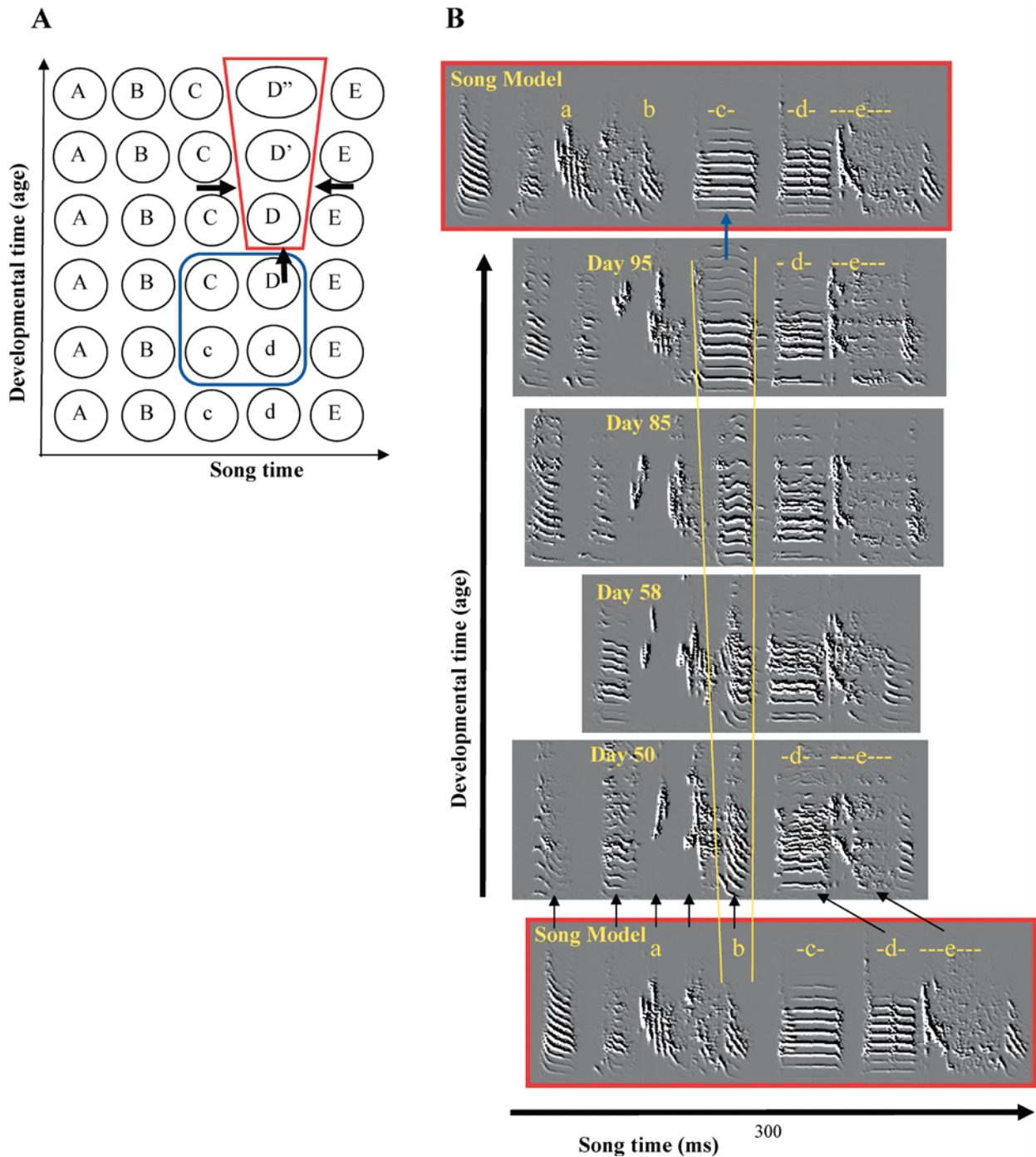
Fig. 4 **A** An illustration of developmental synchronic constraints (*horizontal arrow*) versus developmentally diachronic constraints (*vertical arrow*) on a vocal changes. **B** Examining the emerging song motif of a bird on day 50 show imitation of the beginning and end of the song model (*black arrows*). The imitation trajectory between days 50 and 95 show that a copy of the chirp denoted by *b* was transformed into a copy of the harmonic stack *c*

or priorities (ten Cate 1994) to copy certain 'species-specific sounds'. This idea was introduced by Marler and Nelson (1992) to explain why distant populations of birds retain species-specific characteristics in their songs despite the enormous range of different sounds that they

can master. These two issues are related, as within the context of copying a specific song model, a tendency to copy some sounds before others may impose developmentally synchronic constraints on the learning process. In other words, when a bird copies sound A and then sound B, it could be either because it prefers sound A or because it is easier for the bird to copy A before B. Note that the first explanation is purely perceptual: the bird gives priority to a subset of the song model acoustic structure. The second explanation has an ultimate motor cause, which comes about as primitive sounds are altered as the bird strives to match the model. We will now
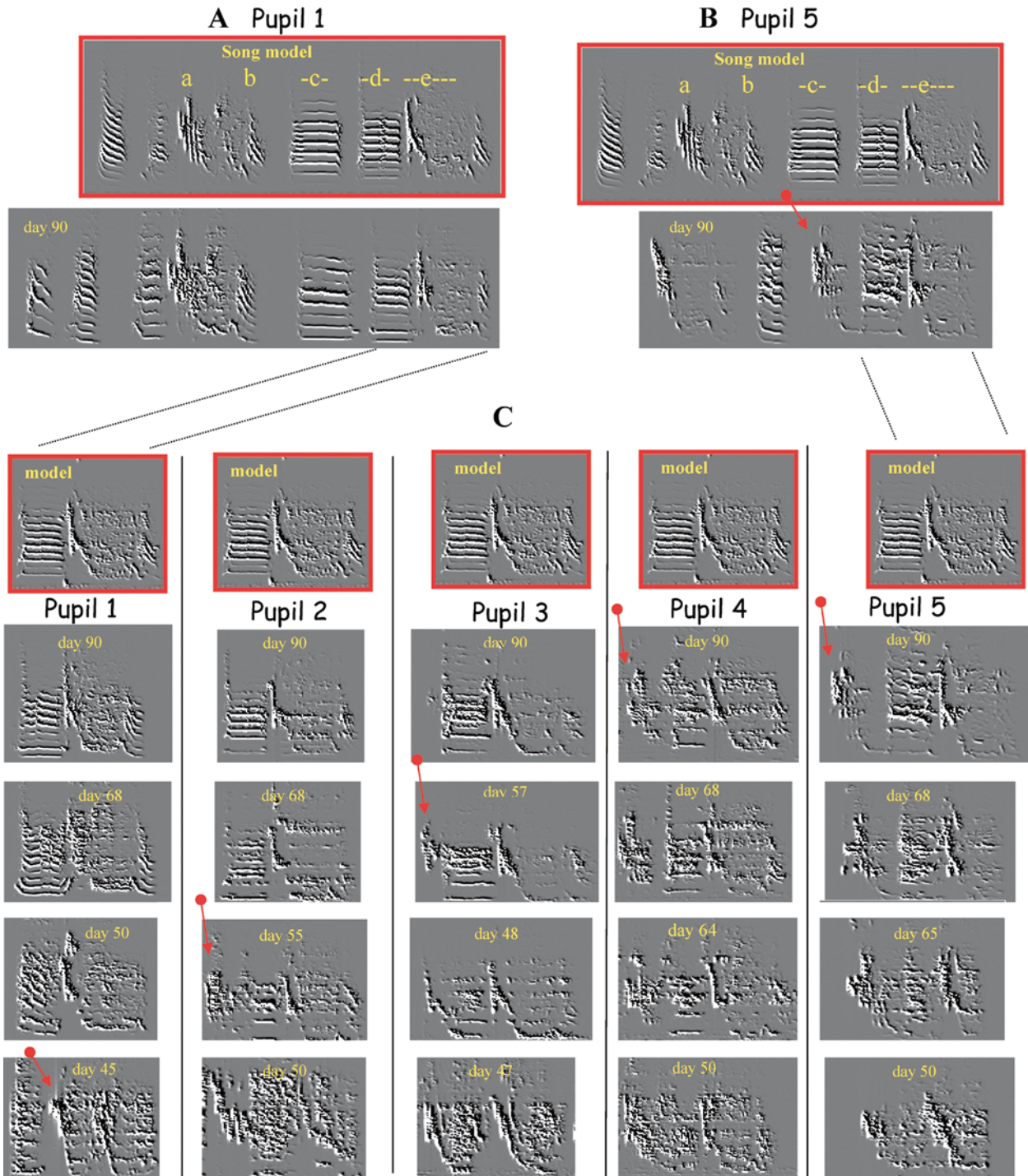
Fig. 5 **A** The mature song of *pupil 1* compared to the model song. **B** The mature song of *pupil 5* compared to the model song. **C** Automatically traced imitation trajectories of sounds *d* and *e* in five birds (pupils 1 and 5 are the same as those in Fig. 5A and B, respectively)

examine if imitation priorities are indeed similar across birds and if those priorities imposes similar developmentally synchronic constraints.

To uncover the priority of copying different parts of a song model we trained nine birds very briefly (for only

2 days, overall about 1 min of playbacks; see Materials and methods) so as to induce partial imitation of the model song. In particular, we wanted to examine if the brief training would cause impairment specific to one part of the model song. The results of *brief* training are judged in reference to those of *moderate* training with the same song model: As shown in Fig. 6A, brief training was sufficient to induce some imitation, with the similarity to the last part of the song model being high across birds and the similarity to the middle part of the
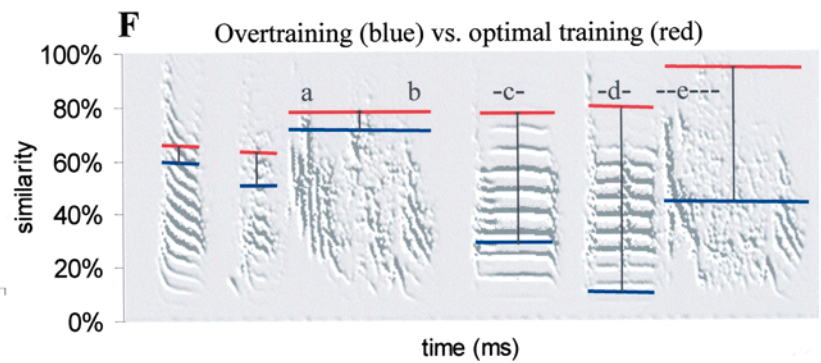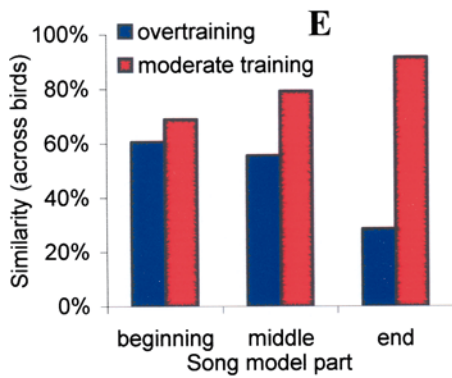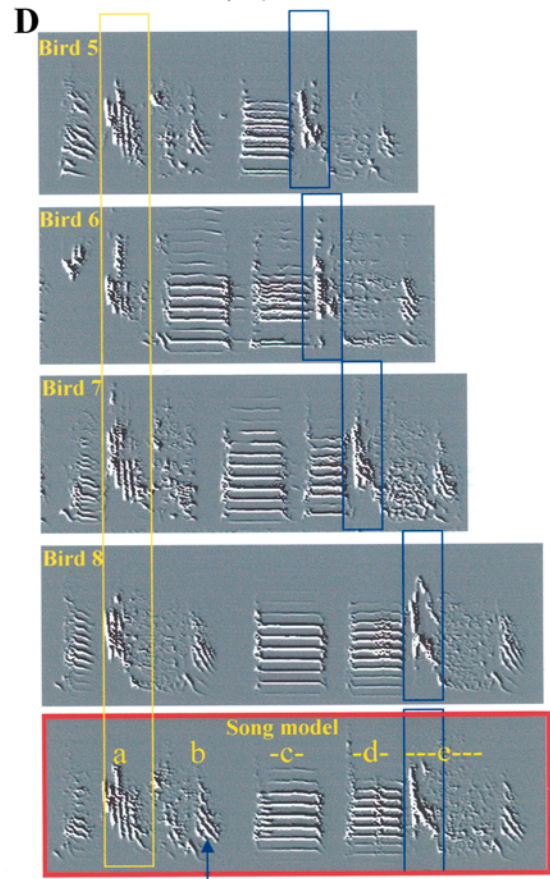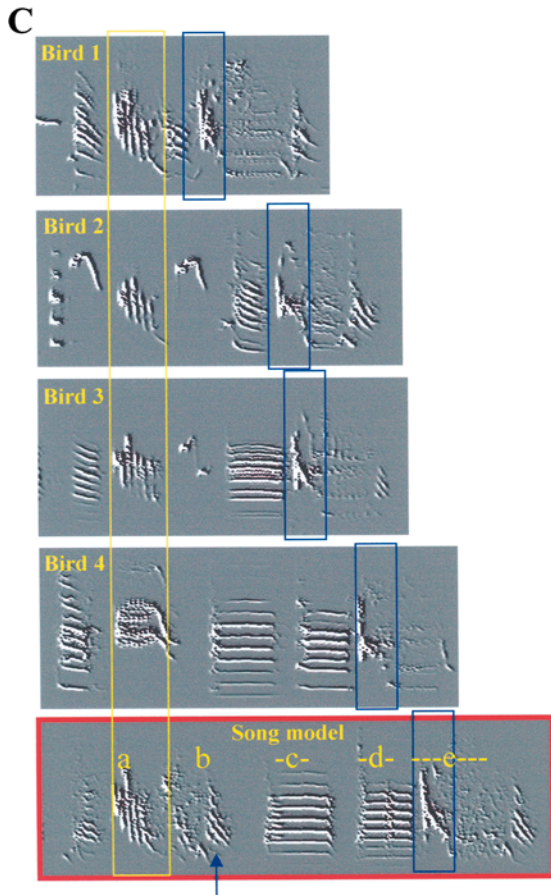
A

B

Brief training (blue) vs. moderate training (red)

C

Bird 1

Bird 2

Bird 3

Bird 4

Song model

a b -c- -d- ---e---

D

Bird 5

Bird 6

Bird 7

Bird 8

Song model

a b -c- -d- ---e---

E

F

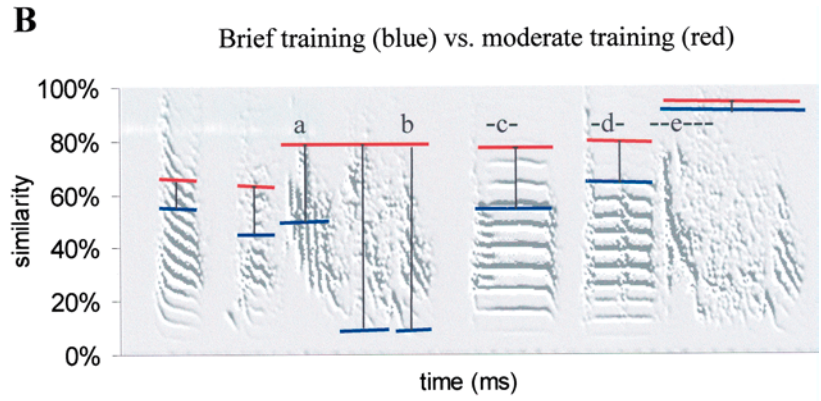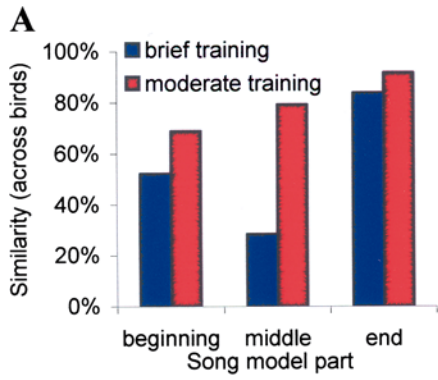Overtraining (blue) vs. optimal training (red)

a b -c- -d- ---e---

◄

Fig. 6 **A** We divided the song model to three equal parts (beginning, middle, and end) and calculated the mean similarity (% significant imitation) of each part for songs of the *brief training* group (*blue bars*) and for songs of the *moderate training* group (*red bars*). **B** A breakdown of similarity for each syllable separately. The *red lines* represent the proportion of moderate training birds that copied each syllable and the *blue lines* represent the proportion of brief training birds that copied each syllable. *Vertical bars* represent the difference between groups. **C, D** Examples of mature song motifs in four birds that received brief training (**C**) versus four birds that received moderate training (**D**). The *yellow rectangles* point to copies of the model vibrato, whereas the *blue rectangles* point to copies of the glissando. Note that in all birds the duration between the two rectangles was shorter than that of the model song. **E** A comparison of similarity to beginning, middle, and end parts of the song model across moderate training (*red bars*) versus over training (*blue bars*) groups. **F** A breakdown of similarity for each syllable separately. The *red lines* represent the proportion of *moderate training* birds that copied each syllable and the *blue lines* represent the proportion of *overtraining* birds that copied each syllable

model being low across birds. Comparing similarity values to first, middle, and last part of the model song across birds showed statistically significant differences across groups (Kruskal-Wallis ANOVA, $H = 6.8$, $P = 0.034$). This bias in the brief training group was particularly strong in sound b (Fig. 6B). Interestingly, this is the same sound that was overwritten by the bird shown in Fig. 4B.

We will now examine a posteriori the variability in song structure within and across groups. Figure 6C presents four examples of songs of the brief training group where we could unequivocally identify copies of model sounds a (vibrato, blue rectangle) and e (glissando, yellow rectangle). We observed two effects: first, the time between sounds a and e was much shorter in the pupil's songs than in the song model (184 ± 66 ms in birds of the brief training group versus 405 ms in the song model, mean $\pm$ SD throughout). Second, the sounds in the neighborhood of sound a (vibrato) varied greatly across birds (similarity values across birds in a 100-ms interval around the vibrato was only $9 \pm 14\%$), whereas the sounds in the neighborhood of the glissando (near the boundaries of the blue rectangle) were much more similar across birds ($63 \pm 19\%$). Interestingly, similar effects (albeit weaker) were observed also in birds that received moderate training. As shown in Fig. 6D, the time between the two landmarks in birds of the moderate training group was shorter than that of the song model (290 ± 84 ms versus 405 ms in the song model, judged in five out of six birds of this group where copies of sounds a and e were identified). The sounds in the neighborhood of the vibrato were variable across birds ($33 \pm 32\%$ similarity), whereas the sounds in the neighborhood of the glissando were very similar across birds ($70 \pm 29\%$ similarity).

Although we did not record song development in the brief training and moderate training groups presented above, the results are consistent with those presented in Fig. 5, i.e., we observed a tendency to copy sound a (vibrato) in a position which is too close to that of sound e.

This tendency, however, was strong in birds that received brief training, and weak in birds that received moderate training between days 30 and 90 post-hatch, where imitation is most accurate. Therefore, the priority of briefly trained birds to copy the beginning and end of the song model could be a consequence of different choices made during early stages of song development, as illustrated in Fig. 5C.

The results of the overtrained groups were very different from those documented above: although the overtrained birds were exposed to the same song model during the same developmental time, they showed different imitation priorities. As shown in Fig. 6E, imitation of the last part of the song model was less accurate (similarity value of 23%) in the overtrained birds, whereas the first and middle parts of the song model were copied much more accurately. Note that this result is very different from that shown in Fig. 6A. As shown in Fig. 6F, the vibrato and the neighboring sounds were copied accurately, whereas most birds did not copy the harmonic stacks and the glissando. Since we did not record song development in these birds we do not know what constraints could have shaped their song development. Nevertheless, it suggests that imitation priorities are not fully determined by the acoustic structure of the model sounds.

Discussion

The purpose of this investigation was to suggest an approach for identifying units of vocal change and to uncover constraints that govern their expression during song imitation in the zebra finch. So far we have only examined a few simple cases and we do not know how much of the vocal learning process could be captured by the approach described here.

Nevertheless, documenting the variety of DVCs across birds and across song models can would the 'toolbox' of vocal changes available to the bird for achieving vocal learning. Since vocal imitation is a one-way process, it would be interesting to search for symmetry across vocal changes: for example, can a bird accomplish the inverse of 'inserting a stop', i.e., 'merging two sounds' with equal facility? It will also be useful to examine what DVCs are available during different developmental stages. Finally, it would be interesting to examine individual variability in the 'toolbox' of available vocal changes, e.g., across good and poor learners. Inherited phenotypes of birds that differ in vocal learning have been identified in Canaries (Mundinger 1995) and it would be useful to have similar preparations in the zebra finch, where both sound analysis and brain measurements are easier to perform.

We suggest that some understanding of the vocal learning procedure and its underlying mechanisms can be achieved by careful analysis of developmentally synchronic and developmentally diachronic constraints on vocal changes. In particular, identifying

conservative and tightly coupled sequences of vocal changes leading to imitation would provide insight into the nature of the computational steps involved in sensory motor transformation. For example, period-doubling trajectories have the appearance of a two-stage approximation mechanism. A systematic analysis of many vocal changes could perhaps tell us, more generally, how many steps are involved in approximating specific sounds, and what level of control system is at play (Powers 1973).

Units of vocal change and constraints on their expression can be explored across behavioral, perceptual (ten Cate 1994), articulatory (Suthers 1999; Goller and Larsen 2002; Fee 2002; Vicario 1991) and central (Jarvis et al. 2002; Lucas et al. 2002; Solis and Doupe 1999; Yu and Margoliash 1996) levels. For example, relating vocal changes to specific articulatory gestures could potentially generalize the descriptive model since the same vocal change could be triggered by more than one gesture and the same gesture could trigger more than one vocal change (Fee et al. 1998) or trigger different vocal changes in different sounds. Similar relations might exist between central and peripheral dynamics (Bernstein 1967; Brezina et al. 2000) of song learning. Overall, the song system provides a unique opportunity to study the information flow across system levels from moment to moment, under tight experimental control, and detailed behavioral analysis of song development would be necessary for understanding vocal learning mechanisms across those levels.

References

Adret P (1993) Operant conditioning, song learning and imprinting to taped song in the zebra finch. Anim Behav 46:149–159

Bernstein N (1967) The co-ordination and regulation of movement. Pergamon Press, London

Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. Nature 404:6779

Brezina V, Orekhova IV, Weiss K (2000) The neuromuscular transform: the dynamic, nonlinear link between motor neuron firing patterns and muscle contraction in rhythmic behaviors. J Neurophysiol 83:207–231

Cate C ten (1994) Perceptual mechanisms in imprinting and song learning. In: Hogan JA, Bolhuis JJ (eds) Causal mechanisms of behavioural development. Cambridge University Press, Cambridge, UK, pp 116–146

Clark CW, Marler P, Beeman K (1987) Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. Ethology 76:101

Goller F, Larsen ON (2002) New perspectives on mechanisms of sound generation in songbirds. J Comp Physiol (in press). DOI 10.1007/s00359-002-0350-6

Fee MS (2002) Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implications for function. J Comp Physiol (in press). DOI 10.1007/s00359-002-0349-z

Fee MS, Shraiman B, Pesaran B, Mitra PP (1998) The role of nonlinear dynamics of the syrinx in the vocalizations of the songbird. Nature 395:67

Immelmann K (1969) Song development in the zebra finch and in other estrildid finches. In: Hinde R (ed) Bird vocalization. Cambridge University Press, Cambridge, pp 61–74

Jarvis ED, Smith VA, Rivas1 MV, Wada K, McElroy M, Smulders TV, Carnici P, Hayashisaki Y, Dietrich F, Wu X, McConnell P, Wang P, Lin S (2002) Integrating the songbird brain. J Comp Physiol (in press)

Konishi M (1965) The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Z Tierpsychol 22:770

Lucas JR, Freeberg TM, Krishnan A, Long GR (2002) A comparative study of avian auditory brainstem responses: correlations with phylogeny and vocal complexity, and seasonal effects. J Comp Physiol (in press). DOI 10.1007/s00359-002-0359-x

Marler P, Nelson D (1992) Neuroselection and song learning in birds: species universals in culturally transmitted behavior. In: Marler P (ed) Seminars on the neurosciences, vol 4. Communication: behavior and neurobiology. Saunders, London, pp 415–423

Margoliash D (2002) Evaluating theories of bird song learning: implications for future directions. J Comp Physiol (in press)

Mundinger PC (1995) Behaviour-genetic analysis of canary song: inter-strain differences in sensory learning, and epigenetic rules. Anim Behav 50:1491–1511

Powers TP (1973) Behavior: the control of perception. Aldine de Gruyter Press, New York

Price P (1979) Developmental determinants of structure in zebra finch song. J Comp Physiol Psychol 93:260

Solis MM, Doupe AJ (1999) Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. J Neurosci 19:4559–4584

Suthers R (1999) The motor basis for vocal performance in songbirds. In: Hauser M, Konishi M (eds) The design of animal communication. MIT Press, Cambridge MA

Tchernichovski O, Mitra PP (2002) Sound Analysis 2 software, user manual and song development database. Source: http://ofer.sci.ccny.cuny.edu

Tchernichovski O, Lints T, Mitra PP, Nottebohm F (1999) Vocal imitation in zebra finches is inversely related to model abundance. Proc Natl Acad Sci. USA 96:12901

Tchernichovski O, Nottebohm F, Ho CE, Bijan P, Mitra PP (2000) A procedure for an automated measurement of song similarity. Anim Behav 59:1167–1176

Tchernichovski O, Mitra, PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. Science 291:2564–2569

Thorpe WH (1958) The learning of song patterns by birds, with especial reference to the song of the chaffinch, *Fringella coelebs*. Ibis 100:535

Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871–1875

Vicario DS (1991) Contribution of syringeal muscles to respiration and vocalization in the zebra finch. J Neurobiol 22:1