

# Vocal Exploration Is Locally Regulated during Song Learning

Primoz Ravbar,<sup>1</sup> Dina Lipkind,<sup>3</sup> Lucas C. Parra,<sup>2</sup> and Ofer Tchernichovski<sup>3</sup>

Departments of <sup>1</sup>Biology and <sup>2</sup>Biomedical Engineering, City College of New York, City University of New York, New York, New York 10031, and

<sup>3</sup>Department of Psychology, Hunter College, City University of New York, New York, New York 10065

Exploratory variability is essential for sensorimotor learning, but it is not known how and at what timescales it is regulated. We manipulated song learning in zebra finches to experimentally control the requirements for vocal exploration in different parts of their song. We first trained birds to perform a one-syllable song, and once they mastered it, we added a new syllable to the song model. Remarkably, when practicing the modified song, birds rapidly alternated between high and low acoustic variability to confine vocal exploration to the newly added syllable. Furthermore, even within syllables, acoustic variability changed independently across song elements that were only milliseconds apart. Analysis of the entire vocal output during learning revealed that the variability of each song element decreased as it approached the target, correlating with momentary local distance from the target and less so with the overall distance within a syllable. We conclude that vocal error is computed locally in subsyllabic timescales and that song elements can be learned and crystallized independently. Songbirds have dedicated brain circuitry for vocal babbling in the anterior forebrain pathway (AFP), which generates exploratory song patterns that drive premotor neurons at the song nucleus RA. We hypothesize that either AFP adjusts the gain of vocal exploration in fine timescales or that the sensitivity of RA premotor neurons to AFP/HVC inputs varies across song elements.

## Introduction

Learning to perform motor actions, such as throwing darts, or continuous actions, such as performing a dance, involves a tradeoff between two motor requirements: exploration to find motor states that can efficiently produce a desired outcome and consolidation of the performance when such outcome is reached. For example, when learning to throw darts, the initial exploratory variability of throws is usually high (Müller and Sternad, 2009). As performance improves and darts hit closer to the target, the throws will stabilize. So, the variability of the action should decrease when average performance is closer to the desired goal (namely, the variance decreases with the bias). However, in learning of continuous actions, such as dancing, in which performance can be evaluated locally, different parts of the action might impose conflicting requirements: although some parts may require exploratory variability, other parts may require stabilization if they are already close to the target (Doya, 2000). In other words, the desired tradeoff between exploration and stabilization might vary across segments of the continuous action. This conflict does not exist in the learning of goal-directed actions in which the outcome of the action can only be evaluated globally. How can conflicting demands for variability and stability in different parts of the action be satisfied during learning of continuous actions? A possible solution would be to segment the action into parts in which variability could be regulated locally so that exploration

can be confined predominantly to those parts of the action that need to change most (i.e., based on the local bias; Doya, 2000).

Here we examine how exploratory variability is regulated during developmental learning of continuous actions, using bird-song as a model. Songbirds are capable of performing complex and stereotyped songs learned by imitation of other birds (Immelmann, 1969; Slater, 1983). In the laboratory, birds can be trained to imitate various song playbacks (Adret, 1993; Lipkind and Tchernichovski, 2011) while recording the entire vocal output (Tchernichovski et al., 2000, 2004). In zebra finches (*Taeniopygia guttata*), song learning starts at approximately day 30 after hatch, and by day 100, the song becomes stereotyped (crystallized). In juvenile birds, variable song patterns are generated by a dedicated brain circuitry, the anterior forebrain pathway (AFP) (Olveczky et al., 2005; Aronov et al., 2008), and variability is used for vocal exploration (Andalman and Fee, 2009; Charlesworth et al., 2011). Lesions of AFP prevent song learning (Bottjer et al., 1984; Scharff and Nottebohm, 1991; Brainard and Doupe, 2000; Haesler et al., 2007) and result in extremely stereotyped performance (Scharff and Nottebohm, 1991; Olveczky et al., 2005). During song development, neural control gradually shifts to a second vocal center, HVC (Aronov et al., 2008), which generates extremely accurate and sparse electrophysiological activity, resulting in highly stereotyped song patterns (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007). Thus, as HVC takes over the control of song production, acoustic variability decreases until the song becomes fully stereotyped, with some residual (but nevertheless functional) variability originating from AFP (Tumer and Brainard, 2007; Aronov et al., 2008; Andalman and Fee, 2009).

We used two approaches to test whether birds can selectively regulate the gain of vocal exploration in different parts of their song. In our first experiment, we manipulated song learning to impose

Received July 18, 2011; revised Jan. 4, 2012; accepted Jan. 7, 2012.

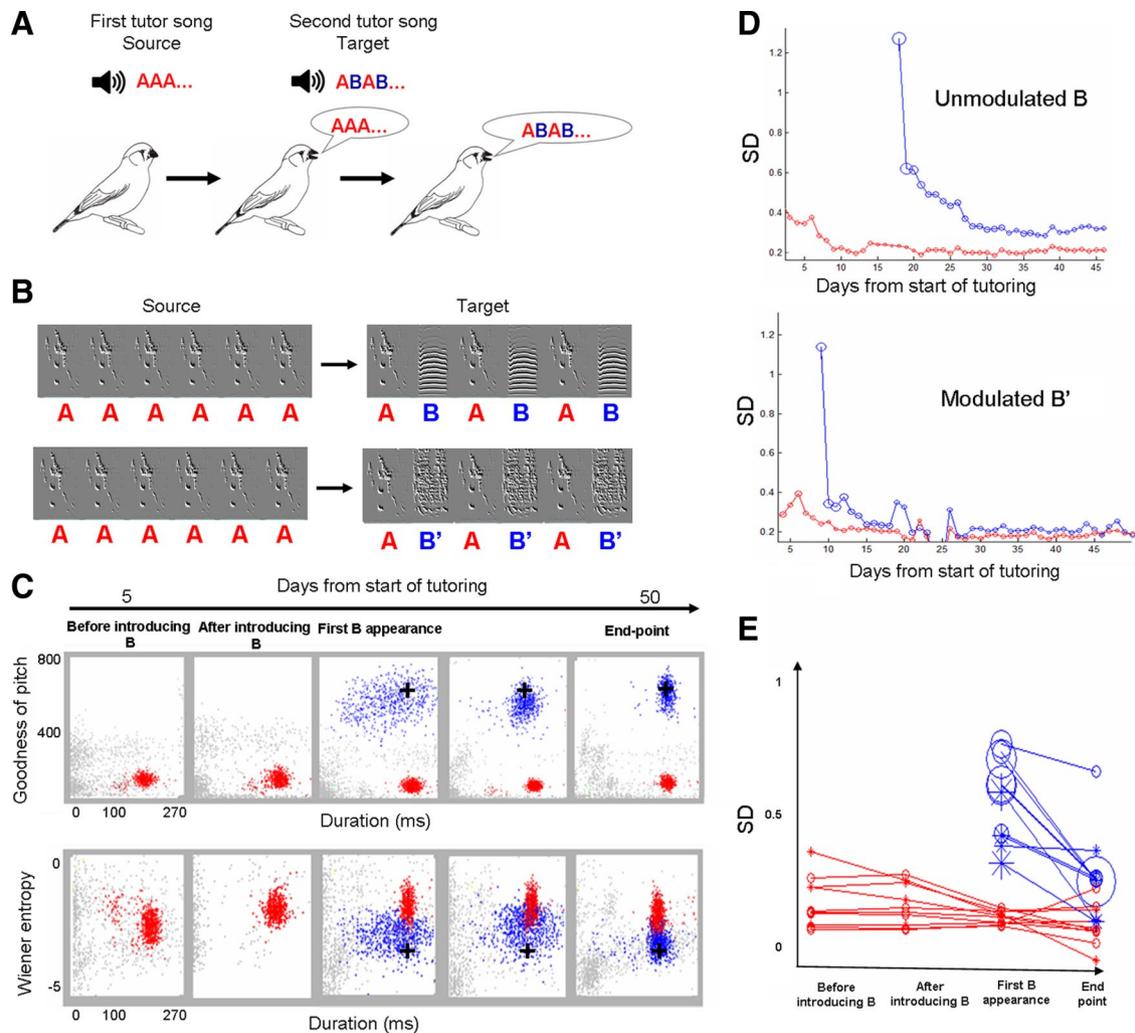
Author contributions: P.R., D.L., and O.T. designed research; P.R. and D.L. performed research; P.R. and L.C.P. analyzed data; P.R., L.C.P., and O.T. wrote the paper.

This work was supported by National Institutes of Health Grant NIDCD41445 (O.T.). We thank Julia Hyland Bruno for proofreading and Tina Roeske for discussion and advice.

Correspondence should be addressed to Primoz Ravbar, Department of Biology, City College of New York, Marshak Science Building, 138th Street and Convent Avenue, New York, NY 10031. E-mail: primoz.ravbar@gmail.com.

DOI:10.1523/JNEUROSCI.3740-11.2012

Copyright © 2012 the authors 0270-6474/12/323422-11\$15.00/0



**Figure 1.** Vocal exploration is confined to newly added syllables. **A**, The AAAA → ABAB altered-target training procedure. **B**, Spectral derivatives (sonograms) showing the source and target song models. **C**, Scatter plots of syllable features (goodness of pitch vs duration and Wiener entropy vs duration in the same bird). The red cluster corresponds to syllable A and the blue cluster to syllable B (unmodulated version). The + symbol indicates the position of the target syllable B. **D**, Variability ( $SD^{syll}$ ) of syllables A and B across development. Examples from two birds trained with unmodulated B (top) or highly modulated B' (bottom) shows combined variability of four syllable features ( $SD^{syll}$ ; see Methods and Materials). The sizes of circles are proportional to the Euclidian distance from the target (bias). **E**, Variability ( $SD^{syll}$ ) of syllables A and B (or B') across birds ( $n = 10$ ) for the following periods: before introducing syllable B (or B') (days -3 to 0 of target training), just after the onset of training with syllable B (or B') (days 1 to 3 of target training), just after the appearance of syllable B (or B'), and at the endpoint (days 85–95 after hatch). Six birds were trained with unmodulated B (circles), and four birds were trained with highly modulated version of B (B') (asterisks). The sizes of circles and asterisks are proportional to the distance from target (bias). Note that the variability of A cluster does not change significantly after the appearance of B (or B') ( $p > 0.98$ ), whereas variability of B (or B') drops significantly ( $p < 0.003$ , single-tailed  $t$  test).

conflicting demands for exploration and stabilization in different parts of the song. In the second experiment, we analyzed the development of song syllables to investigate the relationship between exploratory variability and vocal error at subsyllabic timescales.

### Materials and Methods

**Animal care.** All experiments were conducted in agreement with National Institutes of Health guidelines and were reviewed and approved by the Institutional Animal Care and Use Committee of City College of New York, City University of New York. Zebra finches were bred in family cages. Fathers were removed when clutch mates were 7–8 d old or younger, and thereafter birds were raised by their mothers and were not exposed to songs. On day 30–32 after hatch, male birds were individually isolated in sound-attenuation chambers.

**Altered training procedure.** All birds were tutored with operant song playbacks from 43 to 90 d after hatch, as described by Tchernichovski et al. (1999). Two birds were trained with playbacks of a naturally occurring song throughout the experiment. All other birds were trained using altered training procedure (Lipkind and Tchernichovski, 2011); birds were trained sequentially with two song models (“source” and “target”; Fig.

1A, B). Source and target song models for training were composed from natural syllables. Twenty-eight birds were trained with playbacks of the source song starting from day 43 after hatch. The source song was composed of a single, complex song syllable (syllable A) repeated in a short bout (AAAA...). Singing behavior was recorded continuously, and songs were analyzed daily to determine whether the source model was imitated. For birds that learned the source before day 63 ( $n = 15$ ), we switched their training to playbacks of the target song (ABAB...), and other birds were excluded because switching training after this age can rarely lead to imitation of the altered song target. Ten birds successfully imitated the target model, six with non-modulated harmonic stack (B) and four with broadband, highly modulated syllable (B') (Fig 1B).

**Analysis groups.** Analysis of altered target training was performed on the six birds that performed a recognizable imitation of the target ABAB model and on the four birds that performed a recognizable imitation of the AB'AB' target (overall  $n = 10$  birds). Analysis of intra-syllabic events was performed using one syllable per bird: we analyzed the imitation of syllable A (the source song) in a sample of six birds trained by the altered-training procedure described above and also in two birds trained with playbacks of a naturally occurring song model, using the one complex

song syllable from each (overall  $n = 8$  birds). Having six repeats of the learning of the same syllable allowed us to compare the learning of similar intra-syllabic song elements across birds. Results from the 11 intra-syllabic song elements in the two birds that were trained with a natural song model were used to further validate the generality of the effect across syllable types.

**Song recording and features calculation.** We audio recorded (16 bit, 44.1 kHz) each bird continuously from day 32 to day 90 after hatch using Sound Analysis Pro 1.4 (Tchernichovski et al., 2000). Recording epochs containing songs were automatically identified and saved, and song features (amplitude, pitch, Wiener entropy, etc.) were computed as in the study by Tchernichovski and Mitra (2004). Multitaper spectral analysis was performed (two tapers were used with bandwidth parameter = 1.5) with time windows of 10 ms, advancing in steps of 1 ms such that song features were computed for every millisecond (Tchernichovski et al., 2000). Syllable boundaries were identified using a stationary threshold of sound amplitude.

**Sampling of song data.** For all daily measures, we sampled the entire singing output from the latest third of the daily singing activity to avoid the often strong oscillations in song structure during the morning singing (Derégnaucourt et al., 2005).

**Detection of syllable types.** Once the raw sound data was segmented to syllables, we summarized the acoustic structure of each syllable by computing syllable features: duration, mean frequency, mean Wiener entropy, etc. Syllable features can be displayed in a scatter plot as in Figure 1C. Some of the data points in the scatter plots represent calls and cage noise, whereas others represent song syllables. In the adult bird, song syllables form distinct clusters that correspond to syllable types. These types often appear during early stages of song learning and can be tracked over song development, as shown for the two syllable types A and B (and B'), in Figure 1C.

We detected syllable types semiautomatically by the Sound Analysis Pro clustering procedure, using a hierarchical nearest-neighbor clustering algorithm (Tchernichovski and Mitra, 2004). Starting from the endpoint (day 100 after hatch), the procedure automatically recognizes the clusters, and those are tracked one by one backward in time, until the cluster is no longer detectable (confirmed by visual inspection). The time of appearance of a new cluster (such as syllable B in this case) is the first time when a syllable type can be recognized by our tracking procedure. For all 10 birds trained with the altered-target training, we visually confirmed the day of appearance of cluster B. The backtracking is done in steps of 5000 syllables continuously over the entire singing records. In each step, the user observes the clusters and confirms that the cluster is clearly visible with no detectable residuals. During early development, clusters are larger and less dense, and therefore thresholds for cluster detection needed occasional adjustments. To confirm that this does not introduce artifacts, scatter plots were observed as movies (DVD maps; Tchernichovski and Mitra, 2004) with each cluster displayed with a different color, to confirm by visual inspection that clusters are robust and consistently identified. In cases when identity of clusters was inconsistent, we repeated the tracking from the day inconsistency started. Clustering and backtracking was relatively easy in the altered-target training because the two syllables A and B were chosen to be very distinct in their features.

**Distance and variability estimates.** In general, performance error has two components: performance variance and performance bias. In Figure 1C, the size of the blue cluster corresponds to the performance variability. The distance between the center of the blue cluster and the target (+ symbol) corresponds to the performance bias, and we will refer to it hereafter as the “distance.” The term “vocal error” was used in the abstract and discussion only as a generic, nontechnical term, when referring to developmental changes in the Euclidean distances from the target.

In this study, we test whether variability is modulated locally in song time and whether bias and variability correlate better (in developmental time) when they are computed relative to the target that is locally defined in song time. In other words, we want to test whether performance variability is associated with the local similarity to the target. A major challenge in variability estimations is that our mea-

asures might be affected by nonstationarities in the bias as a result of the natural course of learning and the diurnal oscillations in song structure. As elaborated in the sections below, we used three approaches to reduce such effects. First, we excluded the morning songs (when most diurnal oscillations occur) from the calculation of daily estimates. Second, we computed our estimates in relatively small windows of 50 consecutive renditions, and only then we averaged those estimates over each day. Third, we removed trends from the data of each window before computing distance and variability.

**Variability across renditions for whole-syllable features ( $SD^{syll}$ ).** We investigated variability at the entire syllable level using four syllable features: duration, mean Wiener entropy, mean goodness of pitch, and mean frequency. The units of syllable features were normalized to units of median absolute deviation (MAD) based on the distribution of those features across a large sample of zebra finch songs (Tchernichovski et al., 2000):

$$MAD = \text{median}(|\mathbf{x} - \text{median}(\mathbf{x})|),$$

where  $\bar{\mathbf{x}}$  represents a vector containing the collection of feature values. Normalization allows us to pull variability estimates across syllable features to obtain a single variability measure.

To assess variability of syllable features across renditions of each syllable type (as defined above), we computed SDs in small windows of 50 consecutive renditions ( $N = 50$ ). This window size was determined empirically as large enough for computing robust estimates of central tendencies and variability but small enough to minimize noise attributable to daily oscillations and other nonstationarities.

In each window, we computed mean syllable features for each syllable captured by the feature value  $x_{wsf}$ ; of feature  $f$ , for syllable type  $s$ , at sample  $j$ , within the window  $w$ . Because the measurements of variability could be affected by daily shifting of the means, we removed the linear trend using the Matlab `detrend(...)` function:

$$\bar{\mathbf{x}}_{wsf} = \text{detrend}(\mathbf{x}_{wsf}),$$

where  $\bar{\mathbf{x}}_{wsf}$  is a detrended version of vector  $[X_1 \dots X_N]$  ( $\mathbf{x}_{wsf}$ ) for syllable  $s$  in window  $w$  for the syllabic feature  $f$ .

For each syllable type  $s$  and feature  $f$ , we then obtained a daily estimate of variability across 50 repeats ( $N$ ) and averaged across all windows taken from the last third of a day ( $W$  ranges between approximately 50 to 500 windows):

$$SD_{sf} = \sqrt{\frac{1}{W} \frac{1}{N} \sum_{w=1}^W \sum_{j=1}^N (x_{wsfj} - \text{mean}(\bar{\mathbf{x}}_{wsf}))^2}.$$

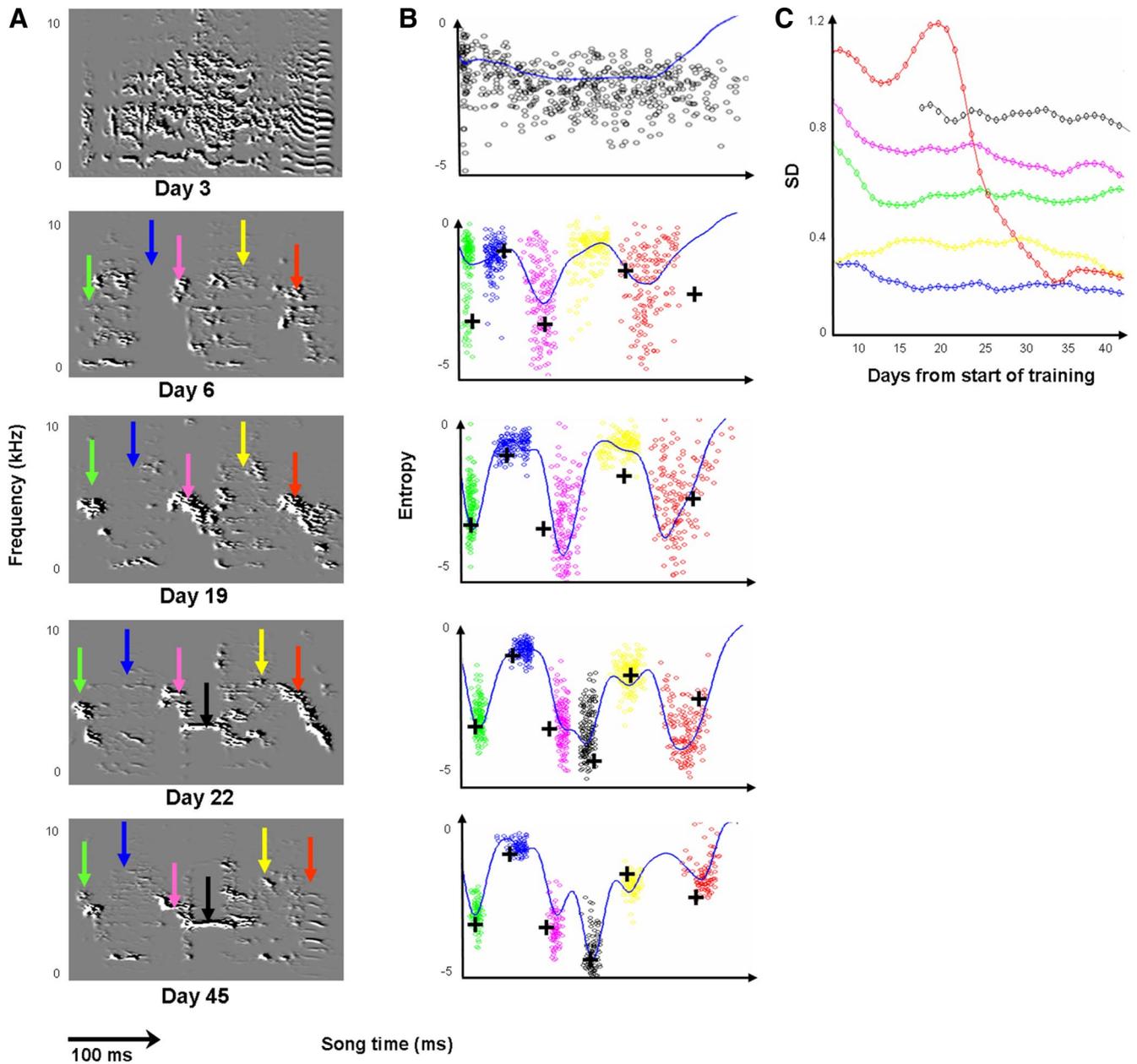
Across the four syllable features, our daily estimate of variability for each syllable type is then

$$SD^{syll} = \sum_{f=1}^4 SD_{sf}.$$

**Distance of syllable-features from the target ( $Dist^{syll}$ ).** For the analysis at the whole-syllable level, we were only concerned with mean features across the entire syllable. Thus, the mean syllable distance from the target (song model syllable) was computed as the Euclidian distance of scaled syllable features from the model (Tchernichovski et al., 2000), using daily estimates by syllable type. For each syllable type  $s$ , syllable feature  $f$ , and a target syllable  $T_{sf}$  (the syllable-feature value in the corresponding song model syllable), we then obtained a daily estimate of the Euclidean distance:

$$Dist_{sf} = \sqrt{\sum_{w=1}^W (\text{mean}(\bar{\mathbf{x}}_{wsf}) - T_{sf})^2},$$

where  $W$  is the number of windows per day. Across the four syllable features, our daily estimate of the Euclidean distance for the target is



**Figure 2.** Time courses of changes in variability within a syllable. **A**, Sonogram of a developing syllable from its first appearance (Day 3) until its consolidation (Day 45). Arrows point to approximate locations of the intra-syllabic events identified. Note that, on day 3, we could not yet identify any intra-syllabic events. **B**, Identified intra-syllabic events can be traced across trials. The blue curve represents Wiener entropy traces averaged across 50 syllables. Intra-syllabic events are shown as colored clusters. Each point in the clusters represents a Wiener entropy minimum. In two of the clusters (blue and yellow), the entropy minima were so subtle that they are not always visible in the averaged curve. Overall, the intra-syllabic events become less variable as the clusters get smaller. (The variability of the red cluster, however, seems to increase on day 22.) Note that, at day 3, we can identify syllabic types but Wiener entropy minima (black circles) do not form intra-syllabic events (clusters) until day 6. The + symbol represents the position of corresponding intra-syllabic events in the target syllable. **C**, Developmental time courses of variability ( $SD_{eT}$ ) for all identified intra-syllabic events. Variability decreases across all intra-syllabic events, but this decrease is asynchronous.

$$Dist^{syll} = \frac{1}{4} \sum_{f=1}^4 Dist_{fT}$$

Note that  $Dist^{syll}$  is an estimate of the overall bias from the target.

**Detection of intra-syllabic events.** Much of our analysis was based on detecting significant time events within syllables during song development. This allowed us to track variability and Euclidean distances from the song model at the subsyllabic level. Empirically, we found that, among several features, Wiener entropy is the first feature to show consistently identifiable transitions within a syllable. We therefore looked at the distribution of local minima in Wiener entropy curves within a syllable throughout song development. Wiener entropy is a measure of the

width of the power spectrum and ranges between zero for white noise and negative values, usually between  $-3$  to  $-8$ , for pure tones and harmonic stacks (Tchernichovski et al., 2000). A Wiener entropy local minimum represents a moment within the syllable at which the concentration of spectral energy reaches a local peak—this corresponds to moments in which harmonic stacks or pure tone events are most clearly defined (Fig. 2 A, B).

Usually, within 2–4 d after a syllable type emerges (i.e., after the emergence of a visible cluster in the space of syllable features; Fig. 1C), minima of Wiener entropy become consistent across renditions of the syllable type. For each bird, we sampled one complex syllable type for analysis, with at least three Wiener entropy local minima (corresponding to distinct “notes” visible in the spectrogram of the syllable) at the endpoint of song development.

We measured Wiener entropy minima for all occurrences of a syllable type for each day of development (~600–10,000 renditions per day). For each rendition of the syllable type, we smoothed the Wiener entropy time course using 30 ms Hanning window using the Matlab `hann(...)` function and then detected local minima (usually three to six per syllable). We then plotted histograms of Wiener entropy local minima time points and detected clusters of these time points based on visual inspection. We called these clusters intra-syllabic events (as in Fig. 2B). Thus, intra-syllabic events are clusters of Wiener entropy minima that consistently appeared in a recognizable timing within a syllable. In practice, once an intra-syllabic event could be recognized, we were usually able to track it continuously thereafter (except in a few cases of bifurcations). We analyzed each intra-syllabic event separately during its entire developmental time course. We used Wiener entropy as the defining feature of intra-syllabic events and determined time of event and mean frequency at the time of event as additional features of intra-syllabic events.

**Variability across renditions of intra-syllabic events ( $SD_{ef}$ ).** Once Wiener entropy minima formed distinct clusters (intra-syllabic events; Fig. 2B), it became possible to estimate the variability of feature values across renditions of intra-syllabic events. As for the analysis of whole-syllable measures, mean values and SDs were computed over windows of 50 consecutive renditions of events in each cluster (intra-syllabic event). Here too, we used the Matlab `detrend(...)` function to remove trends that might result from the natural progression of learning. To obtain daily estimates of variability for each day, we pooled the SDs computed across windows as described above for whole-syllable features.

Variability was computed for each feature and for each intra-syllabic event separately—events are defined as well-identified clusters in time with a local minima in Wiener entropy. We denote with  $x_{wefj}$  the values of feature  $f$  sampled at the time of such intra-syllabic events for  $j = 1 \dots N$  consecutive renditions ( $N = 50$ ) in a running window  $w$  to obtain a vector  $\mathbf{x}_{wef} = [x_{wef1} \dots x_{wefN}]$ . The window of 50 samples was chosen to eliminate the effect of shifting mean over 1 d. Note that, once the time points of Wiener entropy minima were determined, other song features were sampled at those same time points.

We then again removed the trend from each window,  $\bar{\mathbf{x}}_{wef} = \text{detrend}(\mathbf{x}_{wef})$ , obtain a detrended vector of consecutive events in window  $w$ . This vector was computed for each significant intra-syllabic event across the syllable.

For each intra-syllabic events  $e$  and feature  $f$ , we obtained a daily estimate of variability for the last third of the day:

$$SD_{ef} = \sqrt{\frac{1}{W} \frac{1}{N} \sum_{w=1}^W \sum_{j=1}^N (x_{wefj} - \text{mean}(\bar{\mathbf{x}}_{wef}))^2}$$

**Local distances of intra-syllabic events from song model ( $Dist_{ef}$ ).** Euclidean distances for each intra-syllabic event were computed relative to the corresponding events in the target (as noted above, Wiener entropy was used to determine events). Distances were computed for each day using the daily means of acoustic features:

$$Dist_{ef} = \sqrt{\sum_{w=1}^W (\text{mean}(\bar{\mathbf{x}}_{wef}) - T_{ef})^2}$$

where  $Dist_{ef}$  is the daily local Euclidean distance from the target of a significant intra-syllabic event  $e$  for the feature  $f$ , averaged across all windows of the last third of the day  $W$ . These daily measures provide a time course of the local distances over development. Again, these distances represent a bias of the birds in reproducing the target features.

To test how the developmental time courses of daily estimates of variability ( $SD_{ef}$ ) and of distance from target ( $Dist_{ef}$ ) might be related, we computed the correlation between them,  $r_{ef}^{\text{local}} = \text{corr}(SD_{ef}, Dist_{ef})$ , where correlation is determined across days of development. We refer to this correlation as the local correlation.

**Global distance of intra-syllabic events from target.** To estimate Euclidean distance from the template over the entire syllable, but using the same statistical terms, we simply calculated the mean Euclidean distance

over all intra-syllabic events within a syllable, to obtain daily estimates of the distance for the entire syllable (global distance). The course of global distances across development was then correlated with the time course of variability of each intra-syllabic event:  $r_{ef}^{\text{global}} = \text{corr}(SD_{ef}, \frac{1}{E} \sum_{e=1}^E Dist_{ef})$ ,

where  $\frac{1}{E} \sum_{e=1}^E Dist_{ef}$  is the mean of the local distances from the target across all intra-syllabic events within a syllable (typically  $E = 3-6$  events per syllable). We computed  $r_{ef}^{\text{global}}$  for each song feature  $f$  and for each intra-syllabic event  $e$  separately.

**Diurnal oscillations in variability.** To assess diurnal effect, we repeated the computation of variability as above, except that we segregated the data into the first and last thirds of the day (corresponding approximately to morning vs evening renditions). Note that estimates of variability obtained by our method do not depend on the alignment of syllables, nor do they depend on the modulation of syllabic features. Highly structured syllables tend to be more modulated, whereas in less structured syllables, acoustic features are more “flat,” which could result in biased estimates of variability because higher diversity of vocal states (typical of highly developed syllables) may produce higher estimates of differences between syllabic renditions, which in turn may result in higher estimates of variability (across renditions). Identifying intra-syllabic events circumvents this problem.

## Results

### Vocal exploration is confined to newly learned syllables

We first manipulated song learning so that only one part of the song would require vocal exploration. We used an altered-target training approach (Lipkind and Tchernichovski, 2011) and included in the analysis only birds that produced a recognizable imitation of the song models (see Materials and Methods). Juvenile zebra finches were first trained with a source song model (AAAA) consisting of a bout of a single syllable, and, once imitation was detected, we altered the training to a target song model (ABAB), in which the new syllable was either a harmonic stack ( $n = 6$  birds) or a highly modulated syllable ( $B'$ ,  $n = 4$  birds).

We recorded and analyzed the entire vocal output of each bird during the transition AAAA → ABAB, automatically segmenting the songs into syllables. The structure of each syllable was summarized by four features: duration, mean Wiener entropy, mean frequency, and goodness of pitch (Tchernichovski et al., 2000). We then performed cluster analysis of mean syllable features to identify the A and B syllable types (see Materials and Methods). For both types of syllables A and B, we started the cluster analysis on the last day of song development and then traced the clusters back in time. The earliest day when we could still detect the cluster of syllable B we call “appearance of B.”

Figure 1C presents an example of one bird, showing scatter plots of three syllable features (duration vs mean goodness of pitch or Wiener entropy) in different stages of song learning. By the time we altered the tutoring, the cluster that corresponded to syllable A (red) was already small and dense. In contrast, the cluster that corresponded to the newly learned syllable B (blue) was initially much larger and highly scattered. The variability across renditions ( $SD^{\text{syll}}$ ) of syllable B (Fig. 1C, blue cluster) and then gradually decreased and approached the variability of syllable A (Fig. 1D). Similar results were obtained with the addition of syllable B' (bottom panel).

Interestingly, there was no apparent increase in the variability of syllable A when syllable B first appeared (Fig. 1C), indicating that, when the bird sang ABAB (or AB'AB'), it rapidly altered between performing a highly stereotyped and a highly variable syllable. As the features of syllable B approached those of the target (the model syllable; indicated by the + symbol in Fig. 1C), the size of the B cluster (indicating its variability) decreased. In

other words, the distance between the centroid of the B cluster and the target (bias) tends to decrease with the variability of the performance (Fig. 1E).

To test across birds ( $n = 10$ ) whether variability of syllable A was affected by the appearance of new, highly variable, syllable B, we calculated the variability ( $SD^{syll}$ ) of both clusters during four time periods: the last 3 d of source model (AAAA) training, the first 3 d of target (ABAB) training, just following the appearance of syllable B, and at the endpoint (days 90–93 after hatch). As shown in Figure 1E, the variability across renditions ( $SD^{syll}$ ) of syllable A (red) did not increase after the onset of target training, nor during the 3 d period after the new, highly variable cluster B (blue circles) or B' (blue \* symbol) emerged ( $p = 0.98$ , paired  $t$  test) and did not differ from  $SD^{syll}$  at the end of development ( $p > 0.98$ , paired  $t$  test). At the same time,  $SD^{syll}$  for syllable B (or B') decreased significantly across birds ( $p < 0.001$ , paired  $t$  test). These results indicate that, when practicing the modified song, birds can rapidly alternate between high and low acoustic variability, confining variability to the newly added syllable.

### Assessing vocal exploration at subsyllabic timescales

The results above suggest that exploratory variability is not uniformly distributed over the song but depends on the stage of learning for individual syllables. However, song syllables are natural units of song production, and we wondered whether a similar effect might exist even during the continuous singing action within a syllable. With simple syllables such as harmonic stacks, this question is difficult to assess, but many song syllables (including syllable A and B') are of complex type, with several notes and rapid transitions alternating in timescales of a few milliseconds. Can the bird learn such intra-syllabic structures by adaptively changing variability at subsyllabic timescale? In other words, what are the natural timescales of vocal exploration in song time?

To answer this question, we examined song learning at the subsyllabic level, using song development data from randomly selected six birds trained with the altered-target training procedure but focusing on vocal changes within syllable A and also in two birds trained with a single song model. Having six imitation trajectories of syllable A allowed us to test for possible interaction between specific intra-syllabic structures and time courses of learning and variability across birds. The two additional birds were used to confirm qualitatively (over 11 intra-syllabic events in those two birds) whether findings generalize to other complex syllable types. Within each syllable, we used the time course of Wiener entropy to identify local minima of Wiener entropy values, which indicate time points at which the syllable is most tonal. Across renditions, those minima often formed robust clusters, which we call intra-syllabic events (see Materials and Methods).

Figure 2 demonstrates the tracking of intra-syllabic events over development. On the third day after the onset of training, syllable A first appeared as a distinct type (identifiable cluster in feature space as in Fig. 1C). Three to 7 d later, intra-syllabic events gradually appeared (Fig. 2A, B, Day 6). This pattern was observed in all eight birds: syllable types appeared  $3.2 \pm 2.1$  d (means and SD) after the onset of training, whereas intra-syllabic events appeared  $7.6 \pm 3.1$  d after the training onset. This delay suggests a hierarchical process, in which coarse structures (at the syllabic level) consolidate before the fine structures (at the subsyllabic level). As illustrated in Figure 2B, once identified, intra-syllabic events can be tracked continuously over song development (except for bifurcations).

To visualize an example of the detailed relationship between variability and distance from target, we plotted them against each other as shown for one syllable in Figure 3, A and B. Distances of Wiener entropy and mean frequency were computed in reference to the endpoint (Fig. 3A) and also in reference to the song model (Fig. 3B). As shown, in all intra-syllabic events, variability decreased during development but at different rates. For most events, there was an association between the distance from the song model (or from the endpoint) and the level of variability (Fig. 3C). For example, looking at Figure 3B, for events 1 and 6, the initial variability was high in both, but in event 1, variability and distance from target decreased substantially in  $\sim 5$  d, whereas in event 6, variability and distances both remained at high levels for  $\sim 20$  d. Another interesting case is event 3, which bifurcated after 10 d into events 4 and 5. Note that event 4 was initially closer to the target and also of lower variability than event 5. Figure 3C shows the same data on a scatter plot, in which colors indicate individual intra-syllabic events.

### Vocal exploration correlates with local distance from the target on a subsyllabic timescale

We wondered whether the magnitude of vocal exploration differs across intra-syllabic events and, if so, whether different time courses of variability might mirror differences in the learning pace of different intra-syllabic events, i.e., whether decrease in distance from the target drives adaptive decrease in variability separately for each event. To test for this, we computed the Euclidean distances of centroids of intra-syllabic events (i.e., centroids of Wiener entropy minima clusters as in Fig. 2B) from the corresponding events in the target (the position of these events in the target is indicated in Fig. 2B by the + symbol). We call these distance estimates local distances and computed them for each feature separately ( $Dist_{ef}$ ; see Materials and Methods). We then examined whether the local distance can explain the local variability ( $SD_{ef}$ ) better than the global distance, which is the average Euclidean distance of intra-syllabic events within the syllable. Note that the global distance should provide a more stable estimate of distance compared with individual local distances of which it is composed and therefore, by default, should provide slightly better correlations. We then computed the time course of global distance across development and correlated it with the time course of variability for each intra-syllabic event, for each feature ( $SD_{ef}$ ). This estimate,  $r_{ef}^{global}$ , was then compared with the correlation between the local distance and variability of intra-syllabic events,  $r_{ef}^{local}$ . In summary, we test whether the variability in each intra-syllabic event correlates better with its own distance from the target than with the overall distance from the target (calculated over all the subsyllabic events within that syllable).

Figure 4A shows plots of  $r_{ef}^{local}$  versus  $r_{ef}^{global}$ , for Wiener entropy, mean frequency, and time position. The data include all identified intra-syllabic events across all birds ( $n = 41$  intra-syllabic events from 8 birds). Measures of intra-syllabic events within a syllable cannot be considered as independent samples. Therefore, for the purpose of statistical testing, we computed median correlations across intra-syllabic events for each bird (across three to six events constituting one syllable), to obtain a single statistical estimate per bird, i.e., median  $r_{ef}^{local}$  and median  $r_{ef}^{global}$  per bird. Statistical analysis was restricted to differences from the song model, and differences from the endpoint are presented only for qualitative inspection. Variability was significantly more correlated with local distance than with global

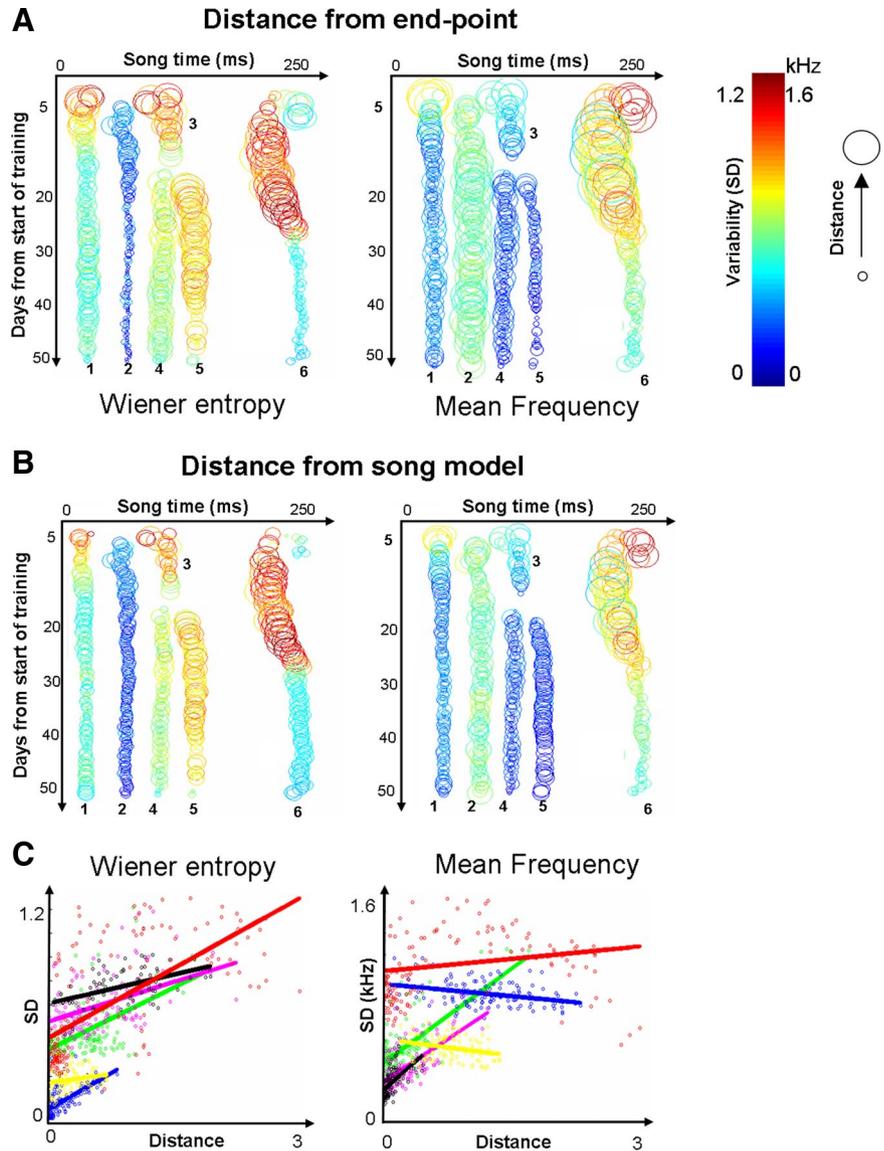
distance in all song features tested: Wiener entropy ( $p = 0.015$ ), mean frequency ( $p = 0.04$ ), and time points ( $p = 0.044$ , paired, two-tailed  $t$  test;  $n = 8$  birds).

The effect was similar across the 11 intra-syllabic events analyzed for the two birds trained with a natural song model and across the 31 events analyzed for syllable A in the other six birds: in both groups of birds, the average  $r^{\text{local}}$  was higher than the average  $r^{\text{global}}$ . For Wiener entropy, in the group trained with the natural model average  $r^{\text{local}}$  was 0.38 and average  $r^{\text{global}}$  was 0.26 ( $n = 11$  events); in the group trained with syllable A, average  $r^{\text{local}}$  was 0.53 and average  $r^{\text{global}}$  was 0.34 ( $n = 30$  events).

Comparing results in reference with song model and in reference with endpoints (Fig. 4A, top vs bottom) shows similar effects, but negative correlations were more frequent when variability is compared with the distance from the song model. Interestingly, in those cases, the bird had first reached the model values but then continued with vocal changes farther away from the model, perhaps actively diverging from the model (Tchernichovski and Nottebohm, 1998).

We next examined whether changes in variability ( $SD_{ef}$ ) and local distance occurred simultaneously within the 1 d time resolution of our study. To test for this, we computed correlations introducing lags between distance and variability (cross-correlations; Fig. 4B). We found that there was no significant deviation of correlations from lag = 0 ( $p > 0.5$ , paired  $t$  test), indicating that changes are indeed simultaneous within the 1 d time resolution analyzed.

An alternative explanation to our findings would be that perhaps certain song elements in a particular syllable are both easier to learn and faster to stabilize (regardless of learning). If this was the case, then the order in which intra-syllabic events are learned would be conserved for a particular syllable. Six of eight birds used for analysis of intra-syllabic events were trained with the same syllable A. Therefore, to test for this possibility, we compared the learning of four equivalent intra-syllabic events in six birds that were trained with the same target syllable (A). We assessed the rate of learning of each intra-syllabic event by computing the overall change in distance from the song model (local distance) and calculating the developmental time when the event reached half of this distance (learning duration). ANOVA of these learning durations showed no consistent order of learning durations among intra-syllabic events ( $p = 0.124$ ;  $F = 2.25$ ; across  $n = 4$  events in 6 birds). This analysis does not have the power to exclude the possibility that the rate of learning might still depend on the features of specific intra-syllabic events, but the lack of evidence to such an effect makes it unlikely that it is still strong enough to explain our results, and therefore, the only correlation that holds is the one between the local distance and variability.

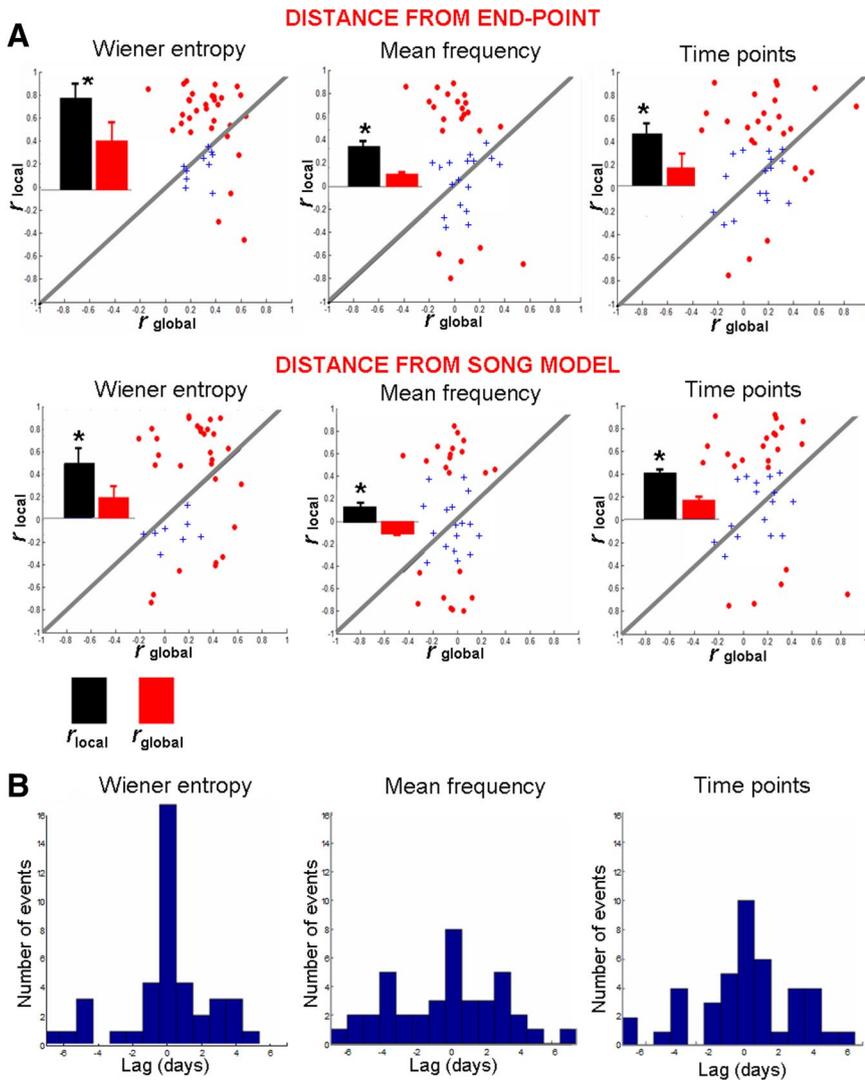


**Figure 3.** Variability decreases when intra-syllabic events reach their targets. **A**, Across development, variability of significant events decreases near their local endpoints. Color of circles corresponds to  $SD_{ef}$ , and size represents the Euclidian distance of intra-syllabic events from their endpoints. **B**, The same variability data as in **A** is presented with the Euclidian distance from the corresponding intra-syllabic event in the song model. **C**, Scatter plots of  $SD_{ef}$  versus distance from song model for Wiener entropy (left) and mean frequency (right). Different colors represent individual intra-syllabic events.

Together, these results indicate that the changes in local variability are best explained by the local distance from the model, suggesting that birds can evaluate the distance locally, at short timescales of no more than 20–50 ms, and maintain high exploratory variability primary in those intra-syllabic events in which local distance is high. It is particularly interesting that variability not only of spectral features but also of time points was better correlated with local distance (of timing) than with global timing distance, which suggests that time jitter of intra-syllabic events is locally gated within a developing syllable.

#### Morning song is more variable than afternoon song

In the analyses thus far, with daily units taken from the afternoon and evening songs, developmental time courses of distance and variability were often monotonic. However, previous investigations have shown that, during periods of rapid learning, there are



**Figure 4.** Variability of intra-syllabic events is better correlated with local than with global distance. **A**, We correlated time courses of variability ( $SD_{ef}$ ) of intra-syllabic events, for each feature separately, with their local distance to obtain  $r_{ef}^{local}$  and with global distance to obtain  $r_{ef}^{global}$ . The errors were computed for Euclidian distances from both endpoint (top) and the song model (bottom). Local and global correlates ( $r_{ef}^{local}$  and  $r_{ef}^{global}$ ) were computed for all intra-syllabic events, across eight birds, separately for each feature. Local correlates are significantly greater than global correlates for all three song features (see Results). For visualization, we denoted nonsignificant correlations by blue + symbols. The bars present mean and SEM for correlations with local (black) and global (red) correlations. **B**, Histograms of lags of intra-syllabic events. Exploratory variability changes simultaneously with local distance. We computed cross-correlations of local distance and exploratory variability ( $SD_{ef}$ ) for all intra-syllabic events across all eight birds studied. Histograms of lags were computed for each of the three song features (Wiener entropy, mean frequency, and time positions). In every feature, the highest number of intra-syllabic events had lag of zero. There was no significant deviation ( $p > 0.5$ , paired  $t$  test) from lag = 0 in any song feature.

strong diurnal oscillations in song structure, such that the morning song is less structured and less similar to the song model (Derégnaucourt et al., 2005). We would predict that, if a bird can adaptively change the magnitude of variability, the morning song, which is less similar to the model, should be more variable than the afternoon song. However, a previous study (Miller et al., 2010) showed that syllables become more, rather than less, variable after daily practice (comparing morning with afternoon songs). We therefore performed a similar analysis, but, instead of examining variability at the syllable level, we computed variability of intra-syllabic events ( $SD_{ef}$ ). As shown in Figure 5, **A** and **B**, variability of intra-syllabic events (Wiener entropy and their timing) is significantly higher in the mornings than late afternoons, which is consistent with the hypothesis that young birds can in-

crease or decrease variability adaptively. Note that both spectral noise within syllables and variability between syllabic renditions of intra-syllabic events ( $SD_{ef}$ ) are higher during morning singing (Fig. 5B). Therefore, the higher variability in the afternoon song reported by Miller et al. (2010) is probably an outcome of higher diversity in the mean values of intra-syllabic events within a syllable (an effect that we see in our data). At the low level, however, the morning song appears to be more variable than the afternoon song.

### Discussion

Using a measurement strategy that brings the timescale of behavioral measurement as close as possible to the timescale of singing-related neural activity during song learning, we presented behavioral evidence that the forebrain circuitry, underlying the production of learned song syllables, can rapidly switch functional states between “exploratory” and “performance” modes.

We studied variability across renditions of song elements (syllables and intra-syllabic events) and found that it is associated with the local requirements for vocal exploration.

During late song development, when variability is low, adding a new syllable to the model triggered high variability confined to the new syllable. A similar effect was observed at the continuous singing action within a syllable. Therefore, the transitions from a highly variable (juvenile-like) to stereotyped (mature-like) song are, in general, asynchronous in song time, even across song elements only several milliseconds apart.

For each intra-syllabic event, the change in variability observed over development could be explained by local deviation (local distance) from the model, with global distance (across all intra-syllabic events within a syllable) showing significantly weaker correlation. This effect was observed across such relatively independent song features as timing, mean frequency, and Wiener entropy. We interpret these results as evidence that, at least in some respects, the juvenile

bird segments its continuous singing behavior into discrete parts that can be learned separately by computing local vocal errors and thus regulating local exploratory variability.

We focus the discussion on three questions raised by our findings. How might the local control of vocal exploration relate to the sensitive period for song learning? What implication might the putative computation of local vocal error have on song learning theory? What neuronal mechanisms could explain the different levels of variability across song elements?

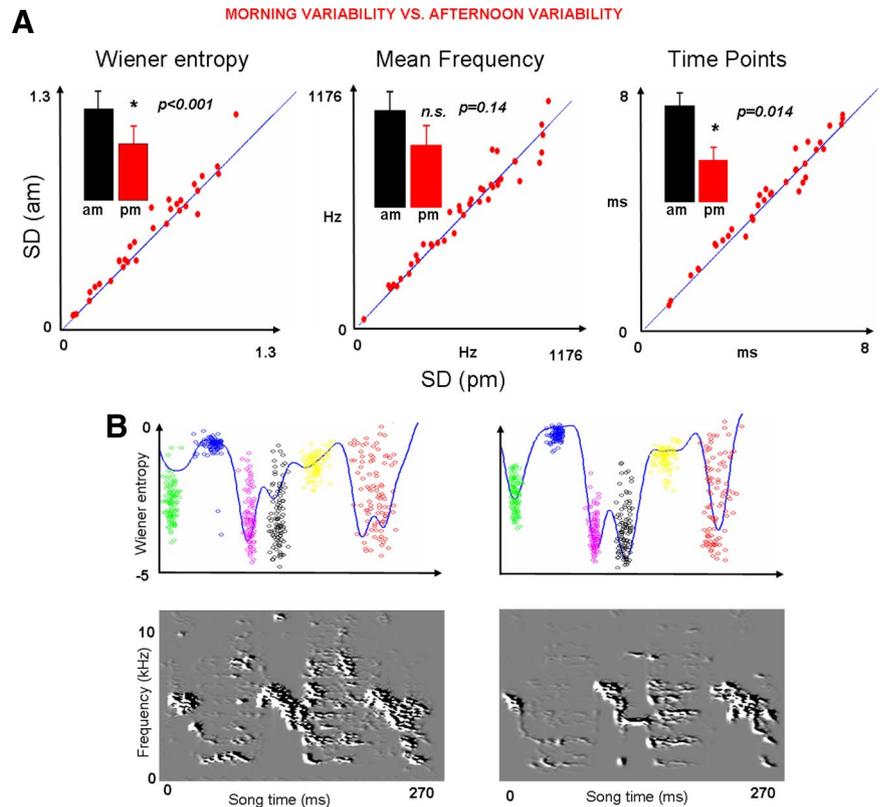
### Piecemeal transition from variable to stereotyped song

Zebra finches rapidly lose their song learning abilities with age and learning as their song turns from variable to stereotyped

(Morrison and Nottebohm, 1993; Boettger and Doupe, 2001), but they nevertheless retain some plasticity into adulthood. The minor residual variability in song features across renditions, which persists even in stereotyped song syllables, is still accessible to negative reinforcement learning (avoidance learning) and can be used to train birds to shift the fundamental frequency of a targeted syllable up or down (Tumer and Brainard, 2007; Andalman and Fee, 2009). Furthermore, microstimulating the AFP (which generates vocal exploration) can trigger highly localized and specific vocal changes in the song of adult birds (Kao et al., 2005). However, even prolonged training of adult birds that succeeded to shift their fundamental frequency did not induce any increments in variability in the targeted syllables (Tumer and Brainard, 2007). Therefore, it is possible to train adult birds to change song features using their residual exploratory variability but only in small steps, suggesting that the developmental transition from high to low variability cannot be easily reversed.

These findings are reminiscent of the results by Knudsen and Knudsen (1989) with adaptive adjustment of auditory orienting behavior in response to displacing prisms in barn owl. Juvenile owls can calibrate their auditory map to adjust for large angular error in their visual field, but adult owls can only adapt to small errors and thus only learn in small steps (Knudsen and Knudsen, 1990). In the adult owl, the ability to adaptively rotate the auditory map is constrained by a sensitive period, possibly at the level of neuronal anatomy or synaptic weights, in auditory neurons in optic tectum (Knudsen, 1985). Analogously, in songbirds, it is the range of active vocal exploration that constrains learning (Tumer and Brainard, 2007). Vocal exploration in juvenile birds provides a broader range of usable song elements than the adult song as a result of stronger variability within and across renditions and stronger diurnal oscillation in song structure (Derégnaucourt et al., 2005).

One interpretation of our results could be that different parts of the song consolidate independently based, at least in part, on local vocal error. Thus, regulation of exploratory variability could be a passive rather than active process. We do not know which of the two hypotheses is correct, but an independent consolidation of song parts based on local error seems sufficient to explain the rapid transitions in variability we observed. Imagine walking through a corridor while looking outside via windows with variably sized openings: each time we cross a narrow window, we are forced to look at the same image, but while crossing a wide-open window, we have a range of images from which to choose. Consequently, variability in the position of our eyes changes quickly as we walk by the windows, although the opening width of each window remains unchanged. By analogy, the rapid transitions we observed between variable and stereotyped song elements do not imply that the range of variability within each song element can



**Figure 5.** Variability of intra-syllabic events tends to be higher in the morning than later in the day. **A**, Developmental time courses of variability ( $SD_{eff}$ ) were computed for each intra-syllabic event from samples taken in the morning and were compared with the time courses of variability ( $SD_{eff}$ ) of the same intra-syllabic event but sampled during late afternoons. In two song features (Wiener entropy and time points), variability was significantly higher in the morning than in the afternoon ( $p < 0.001$  for Wiener entropy and  $p = 0.02$  for time points). For mean frequency, the difference between morning and afternoon variability was not significant ( $p = 0.14$ ). **B**, Sonograms of the same syllable type recorded in the morning (left) and late afternoon of the same day (right). Note that there is more spectral noise in the morning sonogram (am) compared with the late afternoon sonogram (pm). Above the sonograms plots of intra-syllabic events are presented (as in Fig. 2B). Note that clusters (intra-syllabic events) are bigger in the morning (left) than in the afternoon (right). This indicates that both noise within a syllable and variability between syllabic renditions are higher during morning singing.

be changed actively (for example, by injecting more noise to song elements in which performance is far from the target; see below, Hypotheses about neuronal mechanisms).

### Implications for song learning theory

Computing global vocal error might suffice, in principle, to allow for song learning (Fiete et al., 2007), but this would entail that the parts of the song that a bird has already mastered receive the same amount of exploration as do other parts in which the difference from the target is greater (in either acoustic features or timing). The current results suggest instead that vocal error is computed locally in time. We hypothesize that the bird partitions its performance into several discrete events and change the range of exploratory variability independently in each time slice. The Fiete et al. model should work either way, but one would expect that information from local vocal error, if available, should accelerate learning.

Putting articulatory dynamics aside, partitioning the continuous action could reduce the complexity of the song learning task. Although partitioning the learning of a continuous action into discrete tasks is sometimes considered unrealistic (Doya, 2000), the overall number of distinct song elements (syllables and notes) in a zebra finch song ranges only from 5 to 20. Consider for example the learning of 10 song elements, each with 10 possible

acoustic states. Assuming that the bird learns notes one by one and notes do not interfere with each other, then the bird needs to learn 10 tasks of 10 states (100 possible states). However, if the error estimate is only available globally, then the bird would have to select among  $10^{10}$  possible states (regardless of the precision of the error computation). Analytic results for a number of stochastic and deterministic algorithms confirm that the learning speed improves with lower dimensionality (Werfel et al., 2005).

Taking articulatory dynamics into account, however, can complicate matters, because manipulating one intra-syllabic event could affect the acoustic states of neighboring events and undo the learning. In this respect, learning coarse song structure before the learning of fine structure could potentially decrease such interactions and facilitate learning. There is evidence that coarse articulatory states are learned before fine states (Méndez et al., 2010). The current results further support this view, because syllable types (clusters) became detectable in our data a few days before the appearance of robust intra-syllabic events. This raises the possibility that, during song development, the bird shifts not only from high to low variability within each vocal state but also from computing coarse errors over long time windows to computing fine errors in narrower windows.

### Hypotheses about neuronal mechanisms

Previous studies have shown that the AFP plays a critical role in song learning (Bottjer et al., 1984; Scharff and Nottebohm, 1991; Brainard and Doupe, 2000) and in the maintenance of song plasticity in adult animals (Brainard and Doupe, 2001). During early song production (Aronov et al., 2008), AFP is the primary source of variable song patterns (subsong). Later on, control gradually shifts to HVC and the primary role of AFP becomes that of “injecting” noise into stereotyped song patterns produced by HVC (Olviczky et al., 2005), enabling the animal to explore different acoustic states during the period of song learning. Our findings suggest that, during song development, the magnitude of vocal exploration is regulated locally, in part based on local error. This would imply that either AFP can adjust the gain of vocal exploration in fine timescales or that the responsiveness to either AFP or HVC input in the downstream premotor neurons (RA) is dependent on song state. Interestingly, in juvenile birds, an increase in HVC activity is associated with vocal stability, whereas decrements in HVC activity are associated with vocal variability (Day et al., 2008). Therefore, an alternative scenario to modulation in AFP activity is that HVC activity increases with learning independently for each vocal element.

For AFP to be able to adjust the gain of its output properly during singing, it needs to “know” the exact song time. This timing information could be provided to AFP via the HVC area X-projecting neurons. Although previous studies showed that lesions to area X, which should block this information, have no obvious immediate effect on song structure or on the magnitude of noise (Goldberg and Fee, 2011), it is possible nevertheless that fine modulation in vocal exploration might be gated by area X (Kojima and Doupe, 2009) or that AFP has other means of generating a song-time dependent signal.

Alternatively, our results could be explained by differential sensitivity of RA premotor neurons to AFP input across song elements. Each premotor RA neuron receives input from several HVC and AFP neurons (Mooney and Konishi, 1991). Variability in weights across those connections could, in theory, result in song-state-dependent sensitivity of RA neurons to AFP input.

Disambiguating between these hypotheses could be attempted by training a bird with our AAAA → ABAB paradigm

and testing the effect of microstimulation from AFP during singing of either syllable A or B, once the highly variable B syllable appears. Such a stimulation should cause brief vocal changes that are song-time specific (Kao et al., 2005). The prediction is that, if the sensitivity of RA neurons to AFP input is higher for the new syllable (B), such stimulation would have a stronger effect if delivered during performance of B. Alternatively, recording from LMAN (lateral magnocellular nucleus of the anterior nidopallidum) should show increased activity or less inter-hemispheric synchrony (Wang et al., 2008) when singing B.

### References

- Adret P (1993) Vocal learning induced with operant techniques: an overview. *Netherlands J Zool* 43:125–142.
- Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci U S A* 106:12518–12523.
- Aronov D, Andalman AS, Fee MS (2008) A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* 320:630–634.
- Boettiger CA, Doupe AJ (2001) Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* 31:809–818.
- Bottjer SW, Miesner EA, Arnold AP (1984) Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* 224:901–903.
- Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404:762–766.
- Brainard MS, Doupe AJ (2001) Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions. *J Neurosci* 21:2501–2517.
- Charlesworth JD, Tumer EC, Warren TL, Brainard MS (2011) Learning the microstructure of successful behavior. *Nat Neurosci* 14:373–380.
- Day NE, Kinnischtzke AK, Adam M, Nick TA (2008) Top-down regulation of plasticity in the birdsong system: “premotor” activity in the nucleus HVC predicts song variability better than it predicts song features. *J Neurophysiol* 100:2956–2965.
- Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O (2005) How sleep affects the developmental learning of bird song. *Nature* 433:710–716.
- Doya K (2000) Reinforcement learning in continuous time and space. *Neural Comput* 12:219–245.
- Fiete IR, Fee MS, Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J Neurophysiol* 98:2038–2057.
- Goldberg JH, Fee MS (2011) Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. *J Neurophysiol* 105:2729–2739.
- Haesler S, Rochefort C, Georgi B, Licznernski P, Osten P, Scharff C (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol* 5:e321.
- Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65–70.
- Immelmann K (1969) Song development in the zebra finch and other estrilid finches. In: *Bird vocalizations* (Hinde RA, ed), pp 61–74. New York: Cambridge UP.
- Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433:638–643.
- Kojima S, Doupe AJ (2009) Activity propagation in an avian basal ganglia-thalamocortical circuit essential for vocal learning. *J Neurosci* 29:4782–4793.
- Knudsen EI (1985) Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl. *J Neurosci* 5:3094–3109.
- Knudsen EI, Knudsen PF (1989) Vision calibrates sound localization in developing barn owls. *J Neurosci* 9:3306–3313.
- Knudsen EI, Knudsen PF (1990) Sensitive and critical periods localization by barn owls for visual calibration of sound. *J Neurosci* 10:222–232.
- Kozhevnikov AA, Fee MS (2007) Singing-related activity of identified HVC neurons in the zebra finch. *J Neurophysiol* 97:4271–4283.
- Lipkind D, Tchernichovski O (2011) Quantification of developmental bird-

- song learning from the subsyllabic scale to cultural evolution. *Proc Natl Acad Sci U S A* 108 [Suppl 3]:15572–15579.
- Méndez JM, Dall'Asén AG, Cooper BG, Goller F (2010) Acquisition of an acoustic template leads to refinement of song motor gestures. *J Neurophysiol* 104:984–993.
- Miller JE, Hilliard AT, White SA (2010) Song practice promotes acute vocal variability at a key stage of sensorimotor learning. *PLoS One* 5:e8592.
- Mooney R, Konishi M (1991) Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proc Natl Acad Sci U S A* 88:4075–4079.
- Morrison RG, Nottebohm F (1993) Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *J Neurobiol* 24:1045–1064.
- Müller H, Sternad D (2009) Motor learning: changes in the structure of variability in a redundant task. In: *Progress in motor control: a multidisciplinary perspective* (Sternad D, ed), pp 439–456. New York: Springer.
- Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* 3:e153.
- Scharff C, Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J Neurosci* 11:2896–2913.
- Slater, PJB (1983) Bird song learning: theme and variations. In: *Perspectives in ornithology* (Brush AH, Clark GA, eds), pp 475–499. Cambridge, UK: Cambridge UP.
- Tchernichovski O, Mitra PP (2004) *Sound Analysis Pro user manual*. Available at <http://ofer.sci.cuny.cuny.edu>.
- Tchernichovski O, Nottebohm F (1998) Social inhibition of song imitation among sibling male. *Proc Natl Acad Sci U S A* 95:8951–8956.
- Tchernichovski O, Lints T, Mitra PP, Nottebohm F (1999) Vocal imitation in zebra finches is inversely related to model abundance. *Proc Natl Acad Sci U S A* 96:12901–12904.
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. *Anim Behav* 59:1167–1176.
- Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450:1240–1244.
- Wang CZ, Herbst JA, Keller GB, Hahnloser RH (2008) Rapid interhemispheric switching during vocal production in a songbird. *PLoS Biol* 6:e250.
- Werfel J, Xie X, Seung HS (2005) Learning curves for stochastic gradient descent in linear feedforward networks. *Neural Comput* 17:2699–2718.