Development of temporal structure in zebra finch song

Christopher M. Glaze and Todd W. Troyer
Program in Neuroscience and Cognitive Science, Department of Psychology, University of Maryland, College Park, Maryland

Submitted 6 July 2012; accepted in final form 20 November 2012

Glaze CM, Troyer TW. Development of temporal structure in zebra finch song. J Neurophysiol 109: 1025–1035, 2013. First published November 21, 2012; doi:10.1152/jn.00578.2012.—Zebra finch song has provided an excellent case study in the neural basis of sequence learning, with a high degree of temporal precision and tight links with precisely timed bursting in forebrain neurons. To examine the development of song timing, we measured the following four aspects of song temporal structure at four age ranges between 65 and 375 days posthatch: the mean durations of song syllables and the silent gaps between them, timing variability linked to song tempo, timing variability expressed independently across syllables and gaps, and transition probabilities between consecutive syllable pairs. We found substantial increases in song tempo between 65 and 85 days posthatch, due almost entirely to a shortening of gaps. We also found a decrease in tempo variability, also specific to gaps. Both the magnitude of the increase in tempo and the decrease in tempo variability were correlated on gap-by-gap basis with increases in the reliability of corresponding syllable transitions. Syllables had no systematic increase in tempo or decrease in tempo variability. In contrast to tempo parameters, both syllables and gaps showed an early sharp reduction in independent variability followed by continued reductions over the first year. The data suggest that links between syllable-based representations are strengthened during the later parts of the traditional period of song learning and that song rhythm continues to become more regular throughout the first year of life. Similar learning patterns have been identified in human sequence learning, suggesting a potentially rich area of comparative research.

SEQUENCE LEARNING is one of the touchstone questions in neuroscience (e.g., Lasley 1951; Hikosaka et al. 2002; Keeler et al. 2003; Rhodes et al. 2004), and song learning in zebra finches has served as an excellent model system for understanding sequence learning and production. Adult songs are highly stereotyped and have a well-defined temporal structure spanning multiple time scales. In the zebra finch, song learning can be roughly divided into two overlapping processes (Immelmann 1969; Marler 1970; Konishi 1985; Brainard and Doupe 2000): “sensory acquisition,” in which a young bird ~20–65 days posthatch (dph) is exposed to the song of one or more tutors and forms an auditory template; and “sensorimotor learning,” in which the juvenile bird ~35–90 dph learns to produce song based on that template. Once learned, songs are composed of several repeats of 500- to 1,000-ms-long “motifs,” consisting of a sequence of three to seven “syllables,” with 50- to 250-ms-long stereotyped vocalizations separated by silent gaps. In adult zebra finch song, the spectral features of individual song syllables are highly stereotyped, as are the timing and sequencing of syllable production.

To date, most studies of song development have focused on how the spectral features of song are learned (e.g., Tchernichovski et al. 2001; Deregnaucourt et al. 2005; Shank and Margoliash 2009; Charlesworth et al. 2011). Further electro-physiological studies have demonstrated that basic spiking characteristics in the robust nucleus of the arcopallium (RA) develop concomitantly with spectral features (Shank and Margoliash 2009; Ölveczky et al. 2011), and both song features and patterns of neural activity stabilize by ~90 dph ( Ölveczky et al. 2011).

However, few studies have focused on how the overall temporal structure of song is learned. There is converging evidence for syllable-based representations in the adult motor code (Cynx 1990; Franz and Goller 2002; Schmidt 2003; Solis and Perkel 2005; Glaze and Troyer 2006; Cooper and Goller 2006; Andalman et al. 2011), begging the question of how they are formed and what role they play in song learning. There are indications that syllable-sequence variability during the later stages of song development is sensitive to lesions of the lateral magnocellular nucleus of the anterior nidopallium (LMAN), the output nucleus of a basal ganglia circuit involved in song learning (Scharff and Nottebohm 1991; Bottjer et al. 1984; Ölveczky et al. 2005). There is also evidence for changes to learned song timing during both early development and young adulthood that are similarly sensitive to perturbations to the basal ganglia circuit (Brainard and Doupe 2001; Goldberg and Fee 2011). The gradual switch from random timing during the early stages of song learning to more stereotyped timing reflects a switch in motor control from the LMAN to the song motor pathway (Aronov et al. 2011). However, it remains unclear whether there are in fact distinct neural processes associated with learning syllable timing and transitions that are different from learning the spectral features of individual syllables (Troyer and Doupe 2000).

We previously investigated rendition-to-rendition variability in adult song timing and found several patterns that suggested how song representations are organized (Glaze and Troyer 2006, 2007, 2012): first, song intervals share a common source of length variance that we term “tempo variability,” with syllables proportionally less sensitive to this variability than gaps. We also quantified an “independent” component of timing variability that was uncorrelated across syllables and gaps and may be linked with synaptic or neural noise in the pattern generator for song. We have recently developed a statistical “timing variability model” that allows us to separate out these variability components in a principled way (Glaze and Troyer 2012).

To probe the development of the motor code for song timing, we applied the timing variability model to analyze changes in the temporal structure of zebra finch song from 65 dph through 1 yr of age. We found that between 65 and 85 dph, silent gaps decreased in both average duration and their sen-
sitivity to tempo changes, whereas syllables did not. Over that same age range, there were increases in the probability of syllable transitions that were consistent with the adult motif, and those increases were correlated with the decreases in timing parameters gap by gap. In contrast to tempo, independent variability declined for both syllables and gaps and was the only component of timing variability that consistently declined in magnitude across the entire age range examined, decreasing by a factor of two to three through the first year of life. Overall, the data suggest that the final phases of song learning involve the chunking of syllable-based representations into a reliable sequence, with the sequence becoming progressively more rhythmic throughout the first year of life. Similar learning patterns have been hypothesized for mammalian sequence learning, suggesting a potentially rich area of comparative research in birdsong (Graybiel 1998; Sakai et al. 2004; Yin et al. 2009; Costa 2011).

MATERIALS AND METHODS

Analysis was based on songs from seven male birds, recorded between 65 and 375 dph. All care and housing was approved by the Institutional Animal Care and Use Committee of the University of Maryland (College Park, MD). All analysis was performed with Matlab (Mathworks, Natick, MA), with all template matching and dynamic time-warping (DTW) algorithms written as C-MEX routines.

Song collection. Each bird was raised with his parents and clutch mates until 25–30 dph. Individual juveniles were then housed with a single adult tutor (either the father or another adult male) in separate cages within a common sound isolation chamber (Industrial Acoustics, Bronx, NY). At 150 dph, they were released back into the aviary with other birds. Subsequent adult (365+dph) recordings were gathered by returning birds to a chamber, paired with one other adult male. We gathered songs during four age periods: 65–70, 85–90, 125–135, and 365–375 dph.

Chambers were equipped with two directional microphones (Pro 45, Audio-Technica, Stow, OH). Data were recorded at 24,414.1 Hz and consisted of a series of “clips,” periods of sound separated by at least 10 ms of silence, as determined by a sliding window amplitude algorithm. Sound clips separated by <200 ms were included in the same “recording,” and clip onset times were indicated by filling the gaps between clips with zeros. Recordings were attributed to a target bird if the total power was greatest for the microphone directed toward the side of the recording chamber.

We gathered songs during four age periods: 65–70, 85–90, 125–135, and 365–375 dph. We term the collection of songs from a given bird and a given age range as the “bird-by-age sample.” Since songs continue to develop in subtle ways into young adulthood (e.g., Brainard and Doupe 2001), all song analysis treated each time period independently, e.g., song matching was based on templates that were constructed for that period only.

Song selection and template matching. The main thrust of our analysis used our previously published timing variability model (Glaze and Troyer 2012) to separate distinct components of temporal variability and track these over development. The required input to the model is a series of repeated renditions of a single stereotyped sequence of syllables. As a result, our data selection methodology was geared toward identifying and selecting song syllables belonging to each bird’s main song motif. Furthermore, previous studies have shown systematic timing slowing of motifs within a bout of adult song (Chi and Margoliash 2001; Pytte et al. 2007). Given the difficulty of accounting for such within-song timing variability, particularly at the earliest ages, we restricted our timing analysis the first complete motif in each song.

Recordings were initially analyzed using the log amplitude of the fast Fourier transform (FFT) windowed with a 256-point (10.49 ms) Hamming window advanced in 128-point steps. Frequency bins outside the 0.5- to 8.6-kHz range were excluded because song structure is less reliable at the highest and lowest frequencies. To determine if the sound in these clips matched syllables in the bird’s song, templates were formed by aligning and averaging four to five manually chosen clips corresponding to each syllable in the repertoire for that bird-by-age sample; exemplars were time aligned by finding the peaks in a standard cross-correlation of syllable spectrograms. All recorded clips were then matched against each syllable template using a sliding algorithm (Glaze and Troyer 2007). For each template and each time point (t) in the clip, a match score [c(t)] was computed as the reciprocal of the mean squared difference between template and song log amplitudes at matched time-frequency points, as follows:

\[ c(t) = n \times m \sum_{i=1}^{n} \sum_{j=1}^{m} [s(i, j) - t(i, j)]^2 \]

where \( n \) is the number of time bins in the template, \( m \) is the number of frequency bins, \( i \) indexes time, \( j \) indexes frequency, \( s \) is the clip spectrogram, and \( t \) is the template spectrogram.

The peak match score was determined as the maximum of \( c(t) \) over time alignments \( t \). Matches were only considered potentially valid if the peak match score exceeded an initial fixed threshold of 0.5 (manually chosen based on visual inspection) and if the onset and offset for the clip and template differed by 20 ms or less. If a clip had multiple syllable matches, the template with the highest match score was chosen. This yielded a median of 14,720 (range: 1,686–120,873) matches per syllable in each bird-by-age sample.

Template matching thresholds. After all clips were initially matched to a template, template matching thresholds were recalculated for each syllable, based on a simplified minimum error rate classification scheme (Duda et al. 2000): match scores were grouped into 10 bins from the 0.5 minimum score through a value of 1.5, with values greater than 1.5 placed in the last bin. A random sample of 20 clips per score bin were then gathered. A few syllables that had particularly high average match scores had fewer than 20 clips in the low-match bins. In these cases, the clips were grouped with the next highest bin.

Within each bin, each of the 20 clips was manually judged as being either correctly classified or a false positive. Error rates from this process were extrapolated to all clips in each bin, and an optimal threshold was set to minimize the estimated combined number of false positives and false negatives (in this case, correct matches below threshold) for the entire sample (thresholds set by this method necessarily occur at bin boundaries). Across syllables, this yielded a median 1.8% false positive rate (range: 0–34.2%) and a median 0.9% false negative rate (range: 0–34.3%). Approximately 86% and 87% of syllables, respectively, had false positive and rejection rates of <10%. The highest error rates occurred for brief, noisy syllables that resemble introductory notes; these templates frequently yielded modestly high-match scores even for vocalization not directly involved in song, such as “tet” calls (Zann 1996). Importantly, we found no systematic changes in error rates by age.

Motif selection and transition probabilities. We considered a transition between two syllables to be valid if 1) both syllables had a match score above optimal threshold and 2) the time between the offset of one syllable and the onset of the next was no more than 100 ms. We defined the “forward transition probability” between syllables \( X \) and \( Y \) as the total number of valid \( X-Y \) transitions divided by the total number of valid transitions from \( X \) to another syllable. Following previous analysis of sequence variability (e.g., Foster and Botteri 2001; Kao and Brainard 2006), we excluded song stops (i.e., transitions to silence) from all probabilities. This analysis yields a transition probability for each pair of syllables in each bird-by-age sample.

Further timing analysis relied on the application of our timing variability model and was confined exclusively to syllables and gaps contained within each bird’s main song motif, defined as the most common sequence of song syllables sung at 1 yr of age. Across samples, 1,972–14,057 recordings were made (median: 7,132 record-
ings); of these, 12.6–80.4% (median: 41.9%) contained at least one motif, yielding 439–9,927 recordings with song. All of the bird-by-age samples with <33% matched recordings came from birds in the first three age groups (between the ages of 65 and 135 dph), which produced sequences of vocalizations consisting of frequently interrupted portions of song and/or long sequences of calls.

Given the sensitivity of our timing analysis, we further screened the data by omitting any sequence containing a song intervals (syllables and silent gaps) whose length was >5 SDs from the mean of the length distribution for that interval. The vast majority of these outliers could be attributed to acoustic interference from other noises in the case (e.g., wing flaps or vocalization from the tutor). We omitted a median 4.4% bird-by-age sequences for this reason (range: 0.5–25.6%). Without omitting outliers, we found no significant changes in basic statistics such as mean length and change in SD across age periods; however, the accuracy of our statistical modeling may be sensitive to the presence of these outliers.

Across birds-by-age samples, the final median sample size was 3,390 (range: 352–9,879). Across the song motifs of all 7 birds, the sample included a total of 33 unique syllables and 26 gaps of silence.

Song timing calculations. Timing analysis used a more fine-grained algorithm and was restricted to syllables found within identified motifs. Fine-grained spectrograms were recalculated for all syllables using FFTs with a 128-point window slid forward in 4-point steps, yielding 0.16-ms time bins. Although previous research (Glaze and Troyer 2006, 2007) used log amplitudes, here we used raw amplitudes, which we have found to be more reliable. The resulting spectrograms were then smoothed in time with a 64-point window formed by truncating a Gaussian with a 25.6-point (~5 ms) SD. Time derivative spectrograms (TDSs), calculated as differences in amplitude in time-adjacent bins, were used in the rest of the analysis.

Distinct syllable templates were constructed for each bird-by-age sample. In the first step, we applied the same sliding algorithm used previously to the fine-grained TDSs. For each syllable, we selected a random sample of 200 example TDSs. Using the sample with the highest (coarse grained) template matching score as an initial template, each of the 200 TDSs were aligned to this sample using our sliding algorithm. An updated template was formed by averaging the aligned TDSs, and the process of alignment followed by averaging process was repeated.

In the second step, we repeated this strategy of template refinement by aligning TDSs to the template using a custom DTW algorithm (Glaze and Troyer 2007). The final template was based on the average of these warped syllable TDSs. This final DTW step proved to be crucial for the younger developmental periods (65–85 dph), in which timing within syllables is quite variable.

Measuring the variability of interval lengths requires that interval boundaries be identified in a consistent manner across syllable renditions. To accomplish this, we manually marked “events” within the TDS templates and then used DTW to determine the precise times at which these events were observed in individual syllables. Syllable onsets/offsets were generally identified as salient time derivative peaks/troughs in the template TDS. These correspond to peaks in the rise and fall of energy at the beginning and ends of syllables and result in syllables that are slightly shorter than would be expected from measurements based on exceeding a threshold of power. However, the differences are generally small, and we find that our method gives much more reliable measurements.

Markers were first identified manually in templates extracted from oldest age period (365–375 dph). To define template syllable onsets/offsets at earlier ages using the same events chosen for 365–370 dph, we used DTW to map syllable templates from younger ages onto the 365-dph templates.

Timing variability model. Our analysis focused on tracking distinct patterns of timing variation through development. These patterns of timing cannot be measured directly but are “latent” and can be extracted by fitting the timing variability model to the data, a form of factor analysis explicitly tailored to the analysis of timing in action sequences (Glaze and Troyer 2012). The model fitting procedure can be viewed as decomposing the matrix of length covariances into the sum of several components (Fig. 1). The model extracts four basic components of covariance. The first is the “independent variability” that is unique to each interval. This contributes to the individual variance of each interval, found along the diagonal of the covariance matrix. Second, the model estimates the jitter in the onset/offset border between adjacent intervals. Such jitter contributes to the variance in the length of each interval but also leads to a negative covariance between adjacent intervals. These two components of timing variance are “local” in the sense that the variations in each interval and each onset/offset border are assumed to be governed by a distinct stochastic variable.

The third and fourth timing components are each assumed to be governed by a single stochastic variable, but that variable has a “global” influence spread across all intervals in the sequence. The magnitude and direction of the effect of such a global factor interval (k) is captured by a weight ($W_i$), and each factor will contribute $W_i^2$ to the covariance between interval i and interval j. In our analysis, we used two global factors. For the first factor, the weightings to all

Fig. 1. Example of a timing covariance matrix and its decomposition in one bird [age: 365 days posthatch (dph)]. A: data covariance matrix. The darkness of each square shows the pairwise covariance of the durations of the respective song elements listed along the x and y axes (letters denote syllables; dashes denote silent gaps). B: decomposition of the covariance matrix using the timing variability model described in MATERIALS AND METHODS. The leftmost matrix is the total covariance generated by the model, while each matrix on the righthand side of the equation is the covariance generated by the respective timing components. Color limits are the same as in A.
intervals in the motif have the same sign. Thus, as the motif lengths
and shortens, all intervals lengthen or shorten together, although not
by the same amount. For this reason, we will refer to the first global
factor as the tempo factor and the weighting for each interval as the
tempo weight.

For the second global factor, most of the time (but not always)
syllables and gaps have weightings of opposite sign. We will refer to
this factor as the syllable/gap factor. Such a factor will change the
lengths of syllables and gaps across the song in opposite directions
and will contribute a “checkerboard” pattern of covariances. We chose
a sign convention where the syllable/gap weights for most syllables
were positive and the weights for most gaps were negative.

Errors in the measurements of syllable and onset and offset will
contribute to the jitter component. Therefore, we focused our analysis
on the other three factors: independent, tempo, and syllable/gap
timing variability. Jitter is mostly estimated to improve the estimates
of the other components of timing variability. Because the estimate of
jitter depends on knowing the length of the two surrounding intervals,
it is impossible for our algorithm to estimate the jitter in the onset of
the first syllable and the offset of the last syllable in the motif. As a
result, this jitter must be captured by the other factors in the model.
We found no systematic differences in the developmental trends for
variability estimates for these syllables, and, therefore, we lumped
them with other syllables. Finally, while both jitter and the syllable/
gap factor contribute to negative covariances between syllables and
gaps, jitter acts locally and for any given song can act to lengthen one
syllable and shorten another.

Model details and data fitting. For each interval, we calculated the
mean length of that interval. To analyze timing variability, we sub-
tracted off the mean length and wrote the deviation from the mean of
the kth interval in sequence n as \( x_{kn} \). We separated this variability into
a sum of components, as follows:

\[
x_{kn} = W_{kn} + W_{zn} + \eta_{kn} + \left\{ \begin{array}{ll}
-\frac{u_{k-1,n}}{u_{kn}} + \frac{u_{nk}}{u_{kn}} & 1 < k < K \\
\frac{u_{kn}}{u_{kn}} & k = 1 \\
\frac{u_{nk}}{u_{kn}} & k = K
\end{array} \right.
\]

where \( z_{wn} \) is a zero-mean, unit-variance latent variable representing
tempo, \( W_{kn} \) is the sensitivity of interval k to that variable, and \( z_{wn} \)
and \( W_{zn} \) are the respective latent variables and sensitivities for the sylla-
ble/gap factor. The third term, \( \eta_{kn} \), represents the independent variable
and is zero mean but has a nonunitary variance; \( u_{k-1,n} \) and \( u_{nk} \)
represent jitter in the onset and offsets of the interval and are similarly
zero mean with nonunitary variance. It is important to note that the
tempo and syllable/gap latent variables (\( z_{1} \) and \( z_{2} \)) come naturally from a
single, \( 2 \times 1 \) latent variable (\( z \)) that is shared by all song intervals,
whereas the respective weights \( W_{1} \) and \( W_{z1} \) are the columns of a single
\( K \times 2 \) weight matrix (\( W \)). Furthermore, we can rewrite the timing
jitter variables with a single vector \( u \) multiplied by the \( K \times (K - 1) \)
differencing matrix (\( D \)), which has ones along the diagonal (\( D_{k,k} = 1 \)
since boundary causes k is determined in the offset of interval k), negative
one along the subdiagonal (\( D_{k,k-1} = -1 \) since boundary \( k - 1 \)
causes deviations in the onset of interval k), and zero elsewhere.

With these two substitutions, we now have the following:

\[
x_{n} = W_{zn} + \eta_{n} + Du_{n}
\]

We fit parameters to the data using a previously described expecta-
tion-maximization (EM) algorithm (Glaze and Troyer 2012). The
algorithm fits parameters to the measured covariance matrix (\( Si \)), so we
can now write the parameters in matrix form. The covariance data can
now be written as follows:

\[
S = WW^{T} + \Psi + DD^{T}S
\]

where \( \Omega \) is a diagonal matrix of jitter variance and \( \Psi \) is a diagonal
matrix of independent variability. Once parameters were fit using EM,
we transformed \( W \) so that the tempo weights (first column) captured
all of the accumulating variability shared by song intervals. We then
determined interval variability due to the tempo and syllable/gap
factors along the first and second columns of \( W \), respectively. These
values were in units of milliseconds and may be thought of as SDs
with respect to those factors. We determined variance due to the
independent factor along the diagonal \( \Psi \) and, to keep units consistent
with the other variables, computed the square root to yield an effective
SD with respect to that factor.

For each bird-by-age sample, we ran the EM algorithm from 100
randomly drawn initial parameter estimates (Glaze and Troyer 2012). To
assess model fits, we computed the standardized root mean-squared
residual (SRMR) (Glaze and Troyer 2012; Hu and Bentler 1998). This
measure is the average difference between the pairwise correlations in
the data and those predicted by the model covariance matrix. Across samples
at 65–70, 85–90, 125–130, and 365–370 dph, the respective median
SRMR \( \pm \) median absolute deviation was 0.027 \( \pm \) 0.015, 0.015 \( \pm \) 0.015,
0.015 \( \pm \) 0.009 and 0.015 \( \pm \) 0.010. While median SRMR appeared higher
at 65 dph than at other ages, this trend held for only three of seven birds.
Thus, the model fit a large majority of birdsongs in our sample reasonably
well, and there did not appear to be any reliable trend by age for fits to get
better or worse.

Reported statistics. To test for significant age-dependent changes in
a given parameter, we evaluated the difference between 365–375 and
65–70 dph using a nonparametric Wilcoxon signed-rank test (WSR). Tests
for parameter differences between syllables and gaps were performed
at each age range and tested using the Wilcoxon rank-sum test (WSM),
which was Bonferroni corrected for four repeated comparisons. For relationships among parameters themselves, we used
Spearman’s correlation. Throughout the text, we report parameter
distributions using medians \( \pm \) SEs, where we computed the SE from
200 bootstrapped samples per parameter.

RESULTS

We analyzed the development of temporal structure in zebra
finch song in seven males from the following four time periods
spanning the later stages of juvenile song learning through the
(For convenience, we will denote each period using the first
age only.) Across birds, we analyzed a total of 33 unique
syllables and 26 gaps of silence. Final bird-by-age samples
contained 414–9,876 sequences that contained at least 1 full
motif. We used our previously published timing variability
model (Glaze and Troyer 2012) to track three different sources
of timing variability across development (Fig. 1). Previous
studies have shown important differences in the timing
of syllables and intersyllable gaps (Glaze and Troyer 2006; Coo-
per and Goller 2006; Andalman et al. 2011), with syllables
generally being longer but less variable as measured by the
coefficient of variation (CV), equal to the SD divided by the
mean. Therefore, we analyzed syllables and gaps separately.
Our main findings concerning the development of individual
parameters are shown in Table 1.

Changes to average tempo are found in gaps. We began by
confirming previous reports that average song tempo increases
systematically as a function of age (Lombardino and Notte-
bohm 2000; Brainard and Doupe 2001; Pytte et al. 2007). In
fact, between 65 and 365 dph, motif length decreased by a
median of 4.11 \( \pm \) 3.36\% (\( P = 0.0313 \) by WSR), with the trend
holding across five of seven birds and the change ranging from
-18.94\% to 3.14\%. Over half of this decrease appeared to be
explained by changes between 65 and 85 dph, with motif
length decreasing over that period in all seven birds by a
median of 2.68 \( \pm \) 0.75\% (range: -16.52\% to 1.75\%).
Table 1. Developmental changes in timing and sequencing

<table>
<thead>
<tr>
<th></th>
<th>65 dph</th>
<th>85 dph</th>
<th>125 dph</th>
<th>365 dph</th>
<th>Change from 65 to 365 dph</th>
<th>Change from 65 to 85 dph</th>
<th>Change from 85 to 365 dph</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration, ms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motifs*</td>
<td>590.092 ± 32.934</td>
<td>579.975 ± 31.437</td>
<td>569.658 ± 41.755</td>
<td>565.853 ± 47.697</td>
<td>-4.11 ± 3.36%</td>
<td>-2.68 ± 0.75%</td>
<td>-2.43 ± 2.66%</td>
</tr>
<tr>
<td>Syllables*</td>
<td>65.694 ± 11.916</td>
<td>63.198 ± 11.163</td>
<td>63.449 ± 12.218</td>
<td>63.059 ± 12.141</td>
<td>2.13 ± 2.26%</td>
<td>0.31 ± 0.79%</td>
<td>-0.07 ± 1.58%</td>
</tr>
<tr>
<td>Gaps*</td>
<td>62.603 ± 5.01</td>
<td>61.261 ± 6.46</td>
<td>60.131 ± 6.062</td>
<td>58.471 ± 4.807</td>
<td>-10.94 ± 2.81%</td>
<td>-6.10 ± 2.90%</td>
<td>-5.67 ± 2.43%</td>
</tr>
<tr>
<td>Tempo variation, ms</td>
<td>0.334 ± 0.313</td>
<td>0.742 ± 0.169</td>
<td>0.651 ± 0.135</td>
<td>0.954 ± 0.163</td>
<td>0.040 ± 0.183</td>
<td>-0.062 ± 0.208</td>
<td>0.149 ± 0.088</td>
</tr>
<tr>
<td>Gaps*</td>
<td>1.265 ± 0.230</td>
<td>0.943 ± 0.239</td>
<td>0.875 ± 0.122</td>
<td>0.817 ± 0.052</td>
<td>-0.613 ± 0.337</td>
<td>-0.225 ± 0.157</td>
<td>-0.094 ± 0.173</td>
</tr>
<tr>
<td><strong>Tempo coefficient of variation, %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllables*</td>
<td>0.786 ± 0.453†</td>
<td>1.244 ± 0.170</td>
<td>1.014 ± 0.136†</td>
<td>1.132 ± 0.160</td>
<td>-0.022 ± 0.287</td>
<td>-0.089 ± 0.378</td>
<td>0.216 ± 0.081</td>
</tr>
<tr>
<td>Gaps*</td>
<td>2.013 ± 0.487†</td>
<td>1.495 ± 0.254</td>
<td>1.612 ± 0.253†</td>
<td>1.317 ± 0.163</td>
<td>-0.752 ± 0.656</td>
<td>-0.373 ± 0.354</td>
<td>-0.179 ± 0.375</td>
</tr>
<tr>
<td>Independent variation, ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllables*</td>
<td>2.629 ± 0.412†</td>
<td>1.285 ± 0.168</td>
<td>1.304 ± 0.168</td>
<td>1.179 ± 0.245</td>
<td>-0.970 ± 0.299</td>
<td>-0.479 ± 0.325</td>
<td>-0.149 ± 0.162</td>
</tr>
<tr>
<td>Gaps*</td>
<td>3.790 ± 0.448†</td>
<td>2.492 ± 0.264†</td>
<td>1.844 ± 0.313</td>
<td>1.516 ± 0.199</td>
<td>-2.447 ± 0.364</td>
<td>-1.543 ± 0.295</td>
<td>-1.013 ± 0.342</td>
</tr>
<tr>
<td><strong>Independent coefficient of variation, %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllables*</td>
<td>3.286 ± 0.398†</td>
<td>1.719 ± 0.312†</td>
<td>1.578 ± 0.278†</td>
<td>1.196 ± 0.324</td>
<td>-1.255 ± 0.315</td>
<td>-0.876 ± 0.286</td>
<td>-0.288 ± 0.112</td>
</tr>
<tr>
<td>Gaps*</td>
<td>6.251 ± 0.643†</td>
<td>3.71 ± 0.477†</td>
<td>3.585 ± 0.507†</td>
<td>2.353 ± 0.409</td>
<td>-3.669 ± 0.829</td>
<td>-2.038 ± 0.504</td>
<td>-1.423 ± 0.311</td>
</tr>
<tr>
<td>Syllable/gap factor</td>
<td>0.411 ± 0.199†</td>
<td>0.547 ± 0.152†</td>
<td>0.463 ± 0.173†</td>
<td>0.773 ± 0.363†</td>
<td>0.303 ± 0.212</td>
<td>-0.053 ± 0.177</td>
<td>0.087 ± 0.133</td>
</tr>
<tr>
<td>Gaps</td>
<td>-0.652 ± 0.311†</td>
<td>-0.587 ± 0.171†</td>
<td>-0.812 ± 0.197†</td>
<td>-0.933 ± 0.219†</td>
<td>-0.421 ± 0.294</td>
<td>-0.120 ± 0.225</td>
<td>-0.132 ± 0.210</td>
</tr>
<tr>
<td>Transition probability*</td>
<td>0.730 ± 0.098</td>
<td>0.989 ± 0.028</td>
<td>0.986 ± 0.018</td>
<td>0.986 ± 0.007</td>
<td>0.080 ± 0.064</td>
<td>0.047 ± 0.030</td>
<td>0.003 ± 0.000</td>
</tr>
</tbody>
</table>

Reported values are medians ± SE, calculated from 200 bootstrapped samples. Duration changes are reported as percent changes; all others in the same units as the raw measures. *Significant change in that variable over development [65–365 days posthatch (dph), P < 0.05 by Wilcoxon signed-rank test]; †significant group differences (P < 0.05 by Bonferroni-corrected Wilcoxon rank-sum test).

Given the distinction between syllable- and gap-based timing in adult song, we investigated the extent to which these developmental changes occurred across both types of time intervals (Fig. 2). On average, the dominant increases in tempo occurred during gaps (Fig. 2D). Mean gap length decreased in 22 of 26 gaps, with a median decrease of -10.94 ± 2.81% (P < 3.96 × 10⁻⁴ by WSR). Calculated per bird, median gap length decreased in six of seven birds. As with motif duration, more than half of this change occurred between 65 and 85 dph, during which gaps decreased in duration by 6.10 ± 2.90% vs. 5.67 ± 2.43% at 85–365 dph.

In contrast to gaps, syllables as a whole failed to show a consistent pattern of length changes, with changes in length of 2.13 ± 2.26% from 65 to 365 dph (P = 0.851 by WSR).

Changes to tempo variability occur in gaps. We next asked whether the variability in tempo also decreased over late development. For each bird-by-age sample of songs, we used the timing variability model to extract tempo weights that measure tempo variability for each interval. Interestingly, we found a pattern similar to the development of average tempo: silent gaps selectively decreased in tempo variability between 65 and 365 dph, whereas syllables did not (Fig. 2E). Median

Fig. 2. Changes to average song duration and tempo variability over 65–365 dph. A: spectrograms from *bird 257* from 65 and 85 dph. Each spectrogram depicts the song motif whose duration is closest to the average computed for that age period. Syllable labels are indicated between the motifs. B and C: changes in average duration (D) and tempo coefficients of variation (CVs; C) for the same bird over 65–85 dph. Syllable labels match those in A. D–F: median ± SE values for duration (D), tempo variability (E), and tempo CV (F) across development. Durations were normalized by subtracting the duration at 365 dph.
gap tempo variability was 1.265 ± 0.230 ms during the former period and 0.943 ± 0.239 ms during the latter, with average decreases over that period of 0.613 ± 0.337 ms (P = 0.0107 by WSR), more than one-third of the original 65-dph values. As with average tempo, the bulk of this change occurred between 65 and 85 dph, with tempo variability decreasing by 0.225 ± 0.157 ms over that age range vs. 0.094 ± 0.173 ms at 85–365 dph.

In contrast to gaps, syllable tempo variability failed to show significant decreases over late development, with respective medians of 0.334 ± 0.313 and 0.954 ± 0.163 ms at 65 and 365 dph and no significant change across the age range (P = 0.339 by WSR).

It is natural to expect that longer intervals will show more variability in their length. Indeed, across all interval types, we found positive Spearman’s correlations between average length and global weight of 0.450, 0.626, 0.608, and 0.703 at 65, 85, 125, and 365 dph respectively. These values were significantly positive over all four periods (Bonferroni-corrected P < 0.005 in all cases). To control for changes in average length when examining changes in timing variability, we divided tempo weights by mean duration to derive a tempo-based CV (Glaze and Troyer 2012). Given that gaps decrease in length average and CV over development, we expected that gaps would have higher tempo CV at all ages examined. This was indeed the case (see Fig. 2), although given the relatively limited size of our data, only data from 65 and 125 dph reached statistical significance (P < 0.05 by Bonferroni-corrected WSM).

Independent variability decreases for both syllables and gaps. In contrast to tempo variability, independent variability decreased across both syllables and gaps between 65 and 365 dph (Fig. 3), decreasing from 3.790 ± 0.448 to 1.516 ± 0.199 ms among gaps and from 2.629 ± 0.412 to 1.179 ± 0.245 ms among syllables, with respective median decreases of 2.447 ± 0.364 and 0.970 ± 0.299 ms (P < 0.0001 by WSR in both cases). For both syllables and gaps, most of this change occurred 65–365 dph. As with tempo, independent timing variability scales with average interval duration in adults (Glaze and Troyer 2012). This relationship also held across all developmental periods, with Spearman’s correlation ranging from 0.291 to 0.450 (Bonferroni-corrected P < 0.05 in all cases except at 365 dph, where uncorrected P = 0.026). Therefore, we examined independent CV, as we did for tempo, by dividing independent variability by mean duration and examined whether that measure also changed over development. Independent CV decreased significantly between across the first year of life for both gaps and syllables, from 6.251 ± 0.643% to 2.353 ± 0.409% among gaps and from 3.286 ± 0.398% to 1.196 ± 0.324% among syllables (P < 0.0001 by WSR in both cases). As with tempo, gaps tended to have higher independent CVs than syllables across the age ranges examined, although the difference failed to reach significance for 365 dph (P < 0.05 by Bonferroni-corrected WSM).

The syllable/gap factor does not change over late development. Previous analysis of adult songs revealed a second global timing factor shared across the song in which syllables generally have positive weights and gaps have negative weights (Glaze and Troyer 2012). The result is a timing component that induces a negative covariation between syllable and gap durations, independently of the other factors reported thus far. We asked whether this syllable/gap factor was present across development and, if so, whether it changed at all. Indeed, during all four developmental periods, syllable weights tended to be positive, whereas gap weights tended to be negative (Fig. 4). The differences between syllables and gaps were significant for all four periods (P < 0.0005 by Bonferroni-corrected WSM). Furthermore, we found no significant changes in factor weight among either gaps or syllables at 65–365 dph, with median changes of −0.421 ± 0.294 and 0.303 ± 0.212 ms (P = 0.201 and 0.237 by WSR).
Links between changes in gap timing and syllable transition probabilities. We have shown that gaps, defined as the period of silence between two syllables in the motif, become shorter and less variable over development with changes in both tempo and independent variability disproportionately occurring between 65 and 85 dph. To investigate whether these changes in timing might correspond to a tighter link between those syllables, we examined developmental changes in the transition probabilities for each consecutive syllable pair in the motif. Transition probabilities showed a developmental trajectory that was similar to timing parameters (Fig. 5). Specifically, we found significant increases at 65–365 dph, from a median of 73.0 ± 9.8% to 99.6 ± 0.7% (P < 0.0001 by WSR). Of the 20 gaps with transition probabilities <95% at 65 dph, 18 gaps increased, with 12 gaps by over 10%. Qualitatively, increases appeared to reflect jumps to a probability near one, irrespective of what those probabilities were at 65 dph (Fig. 5A). As with changes to timing parameters that showed significant development, most of this change occurred at 65–85 dph, with transition probabilities increasing by 4.7 ± 3.0% vs. 0.3 ± 0.0% at 85–365 dph.

We next asked whether, on a gap-by-gap basis, the increases in transition probabilities from 65 to 85 dph were linked with changes in the timing parameters that showed significant differences across development (Fig. 6), specifically average duration, tempo CV, and independent CV. Indeed, we found that increases in transition probability were correlated with decreases in duration and tempo CV, with respective Spearman’s correlations of −0.557 and −0.342 (Bonferroni-corrected P < 0.01 in both cases). The correlation between the increase in transition probability and independent CV was also negative, although very weak (Spearman’s correlation = −0.136, P = 0.506).

DISCUSSION

We investigated the development of zebra finch rhythmic stereotypy from the late plastic period through 1 yr of age. We collected hundreds to thousands of songs from each of four age periods and measured song timing and syllable transition probabilities within each age period to probe how the motor code changes over this age range.

We found increases in song tempo that occurred almost entirely during the gaps of silence between syllables. Most of the tempo increase occurred at 65–85 dph, and, on a gap-by-gap basis, increased speed was linked with increases in the reliability of corresponding syllable transitions. Increases in transition probability between 65 and 85 dph were also linked with decreases in tempo variability. For syllables, we found no systematic increase in tempo or decrease in tempo variability. But as with gaps, syllables showed a sharp reduction in independent variability between 65 and 85 dph followed by continued reductions over the first year of life.

Overall, the data suggest a phase of late development in which song production becomes faster and more automated. Such a process of motor consolidation has been previously suggested for zebra finch song based on other data (Brainard and Doupe 2001). Interestingly, one recent study has indicated opposite changes in Bengalese finch song timing in the later years of life, i.e., that songs slow down, mostly due to increases in gap duration (Cooper et al. 2012).

Syllable lengths. We found that gaps, but not syllables, shortened over the first year of life. Brainard and Doupe (2001) reported a shortening of both syllables and gaps in young birds (>100 dph), with the fractional shortening of syllables roughly half as much as for gaps. However, only the syllable changes reached statistical significance, as gap changes were more variable. Pytte et al. (2007) found a significant shortening of syllables over the 1.5 yr of life, but no overall shortening of gaps. It is unclear what accounts for the discrepancies between the three studies. One point of difference is our use of a reliable peak in the amplitude derivatives to measure syllable durations rather than threshold crossing of an amplitude threshold. Syllables sometimes have significant amplitudes extending past the onset of inspiration (Andalman et al. 2011); it is likely such low-amplitude “tails” of a syllable would be included in our measure of gap rather than syllable length. If the greater coordination between syllable production and respiration that occurs over development reduces these low-amplitude portions of syllables, this might lead to the shortening of syllable length as measured in previous studies without having much effect on syllable length as measured here.

The chaining hypothesis. The most common view of sequence generation is that of a chain with neurons active at one point in the sequence, exciting the neurons active at the next point in the sequence, and so on. Previous modeling work has shown that the combination of spike timing-dependent plasticity and synaptic pruning can lead to the development of chain-like patterns of activity in initially randomly connected networks (Jun and Jin 2007). Under such a scenario, one might expect that the strength of connections along the chain to be...
correlated with several behavioral parameters. First, stronger connections would decrease the latency to activity in the next link in the chain, leading to a positive correlation between synaptic strength and tempo. Second, stronger synaptic connections would be associated with more reliable propagation along the chain and, hence, less subject to extraneous “noisy” inputs, yielding lower timing variability.

Behavioral data from adult birds are consistent with a version of this hypothesis in which the parts of the chain corresponding to gaps remain more weakly connected than the parts of the chain corresponding to syllables. First, birds startled by flashes of light interrupt their songs at transitions between syllables (Cynx 1990), consistent with a greater sensitivity to perturbation for the more weakly connected gap portions of the chain. Timing variability is greater for gaps than for syllables (Cooper and Goller 2006; Andalman et al. 2011) both in the tempo and independent components of variability (Glaze and Troyer 2006, 2012), consistent with less sensitivity to outside perturbation for the syllable portions of the chain.

The chaining hypothesis has been extended to songs with variable sequencing, by supposing that each syllable corresponds to a tightly linked subchain, and neurons at the end of chain corresponding to one syllable can synapse onto neurons at the head of the chain for more than one subsequent syllable; mutual inhibition from local interneurons then implements a winner-take-all competition between chains, ensuring that only one syllable-based chain could be active at any given time (Jin 2009). Under this hypothesis, transition probability is correlated with connection strength, since a chain receiving stronger input would be more likely to emerge as the winner in a winner-take-all competition. During late plastic song (~65 dph), these mechanisms may have created a loosely organized network of syllable-based chains capable of producing a small variety of song sequences with a bias consistent with the ordering in final motif. Over time, spike timing-dependent plasticity would favor the more likely transitions, and these could completely “win out” by adulthood.

We have found that increases in transition probability are correlated with decreases in gap duration and tempo CV on a gap-by-gap basis. Correlations with independent CV are also negative, although the correlation is weak. Negative correlations between transition probability and duration and timing variability are consistent with the chaining hypothesis in that increases in transition probability and decreases in timing parameters would result from increases in synaptic strength over development. This pattern is also consistent with data in Bengalese finches showing that the ability of auditory feedback perturbations to alter both the stereotypy of song sequence and the latency to the next syllable is smaller when perturbations are introduced at points in the song with more regular sequencing (Sakata and Brainard 2006, 2009).

**HVC as the chaining nucleus.** Previous electrophysiological recordings have shown that premotor neurons in the high vocal center (HVC) song nucleus (used as proper name) produce short bursts of precisely timed action potentials, both during the syllables and during the gaps (Hahnloser et al. 2002). Further correlational evidence suggests that inputs from HVC control burst timing in the RA (Fee et al. 2004). Cooling the HVC leads to a slowing of song tempo, whereas cooling the downstream RA does not (Long and Fee 2008). These results, along with intracellular recordings in HVC neurons during singing (Long et al. 2010), suggest that the HVC supports the chain-like bursting that controls the timing of song behavior. This hypothesis is challenged by the tight synchronization of both the left and right HVC, despite the lack of a corpus callosum in birds (Schmidt 2003). The most likely source of a synchronization signal is from bilateral connections from the midbrain and brain stem targets of the song motor pathway, which eventually feedback to HVC (Schmidt et al. 2004; Ashmore et al. 2005, 2006; Andalman et al. 2011).

**Chaining in Bengalese finches.** If true, the chaining hypothesis could also have one of several implications for changes to song in older Bengalese finch birds (Cooper et al. 2012): the slowing of tempo could reflect a gradual weakening of synapses, although this would also imply a similar correlation between increases in gap duration and weakening of associated transition probabilities. That study did find decreases in syllable repetition rate, although a specific correlation with gap duration was not reported for those transitions. Furthermore, another modeling study of Bengalese finch song has suggested that syllable repetitions are controlled by a process (adaptation) that is separate from what controls other syllable transitions (Jin and Kozhevnikov 2011). On this basis, one may not expect the transition-timing correlation among syllable repeats per se but rather among other transitions; such an expectation implies other changes to transition structure in aging Bengalese finch song. Alternatively, the slowing of tempo could reflect changes to influences on tempo, such as brain temperature (Long and Fee 2008), rather than network parameters, such as synaptic strength; under this scenario, there would not necessarily be changes to transition structure outside the decrease in repetitions.

**Challenges to a chaining explanation.** While increased synaptic strength is an obvious hypothesis for explaining developmental changes, it cannot alone account for the changes in timing variability we have found. First, whereas average tempo increases among gaps only, decreases in independent timing variability occur across both syllables and gaps. This contradicts one basic prediction of the chaining model, i.e., that changes in chain speed are coupled with sensitivity to external sources of timing variability. How can syllable tempo remain the same while independent timing variability decreases? Second, increases in synaptic strength would be expected to decrease the sensitivity of song intervals to perturbations from all possible noise sources, including neuromodulatory mechanisms that induce shared, global timing variability as well as background noise and inputs from areas of the song system that might contribute to independent variability. Why are the decreases in independent variability so much greater than the decreases in tempo variability?

One mechanism that could reconcile these differences would be a developmental increase in synaptic strength for gap portions of the chain coupled with an overall decrease in inputs from circuits that contribute to independent timing variability for both syllables and gaps. The former would account for changes in transition probability, mean length, and tempo variability for gaps, whereas the latter would contribute to reductions in independent variability for both syllable and gaps. Under such an hypothesis, the decreases in independent variability for gaps that come from external sources would act to mask any dependence of independent variability on transition probability.
One possibility for the source of input contributing to independent variability is the LMAN. This nucleus is the output nucleus of the cortical-basal ganglia loop for song and has been strongly implicated in song learning (Bottjer et al. 1984; Scharff and Nottebohm 1991). The LMAN produces reliable burst sequences with trial-to-trial timing variability (Hessler and Doupe 1999; Kao et al. 2008), whereas LMAN inactivation in developing birds yields adult-like stereotypy in the activity premotor RA during singing (Olveczky et al. 2011). Based on experiments combining behavioral or pharmacological manipulations of LMAN activity, it has been proposed that the LMAN is involved in generating the rendition-to-rendition variability necessary for trial and error learning (Kao and Brainard 2006; Kao et al. 2005; Olveczky et al. 2005; Stepanek and Doupe 2010; Charlesworth et al. 2012). Behavioral studies in developing birds have suggested that song variability driven by LMAN activity might be directed to more poorly learned elements of the song (Ravbar et al. 2012). Finally, whereas the LMAN influences song variability throughout a bird’s lifetime, the degree of this influence has been shown to progressively decline over development (Aronov et al. 2011), including the same age range we investigated here (Brainard and Doupe 2001).

The hypothesis that the LMAN also contributes to timing variability is consistent with previous studies showing that inactivation of the LMAN reduces variability in both syllable and gap durations (Thompson et al. 2011; Goldberg and Fee 2011). At the level of song motifs, Thompson et al. (2011) found a reduction in timing variability after LMAN lesions, but this reduction failed to reach statistical significance. One possible explanation is that tempo variability, since it is shared across the song, dominates variations in motif duration, whereas the tempo and independent components make roughly equal contributions to timing variability for individual syllables or gaps (Glaze and Troyer 2012). As a result, the hypothesized reduction in independent variability after LMAN inactivation would be expected to cause relatively smaller reductions to variability in motif duration versus that seen at the level of individual syllables or gaps.

If the hypothesis that the LMAN contributes to timing variability is paired with the hypothesis that song timing is controlled in the HVC, then the LMAN must have a way to influence activity in the HVC. Given the absence of direct connections from the LMAN to the HVC, this influence must be carried by the possible routes from the RA, the premotor nucleus that receives input from the LMAN, back to the HVC. These include direct synaptic connections from the RA back to the HVC (Roberts et al. 2008), projections from the RA to the medial portion of the dorsolateral thalamus to the medial magnocellular region and back to the HVC (Vates et al. 1997), or through the circuits mentioned above that loop through the midbrain and brain stem (Schmidt et al. 2004).

An alternative mechanism that might reconcile differences in developmental timing changes for syllables and gaps would be a balancing of increased excitatory synaptic strength with commensurate increases in inhibitory input during syllables. In adults, inhibitory HVC interneurons have been demonstrated to be more active during syllable-based activity than for gaps (Kozhevnikov and Fee 2007), and it is possible that tempo increases expected from a strengthening of excitatory synapses over development are offset by increased inhibitory feedback during syllables. However, this mechanism does not explain the differences in tempo CV for syllables versus gaps. Furthermore, increased inhibitory feedback might also suppress local timing variability in timing, leading to a greater decrease in independent CV for syllables than gaps, which we did not find.

Other possible influences on the development of song timing. The declining influence of the LMAN at later stages in development may be a continuation of a switch from an early mode of vocalization known as subsong, in which young birds less than ~45 dph produce long rambling vocalizations without a clear segmentation into syllables. Subsong is known to be primarily driven by LMAN inputs to the RA with inputs from the HVC becoming more important as birds start to produce early versions of song syllables (Aronov et al. 2008, 2011). During these earlier stages of learning (~35–65 dph), lesions to the medial portion of the dorsolateral thalamus, which sends afferents to the LMAN, increase the temporal stereotypy and overall rhythmicity of these early song elements (Goldberg and Fee 2011).

During this early period, birds are also learning to coordinate their breathing patterns with syllable production, with long inspiratory pulses increasingly punctuated by short inspirations that eventually closely coincide with intersyllable gaps (Veit et al. 2011). It is possible that the shortening of gaps during the later period of development analyzed here may result from maturation of that respiratory and syringeal musculature, allowing birds to replenish their air supply with shorter inspirations. However, this mechanism does not explain the link between the variability in gap duration and sequence stereotypy.

New neurons are added to the HVC song nucleus throughout a zebra finch’s life, but the rate of new neuron incorporation declines precipitously over the first 2 yr after song crystallization (Wang et al. 2002). During this period, spectral features of song syllables become increasingly stereotyped, and it is possible that the decline in new neuron incorporation may be linked to changes in temporal stereotypy (Pytte et al. 2007). In particular, the incorporation of new neurons might contribute to the decline in independent timing variability for both syllables and gaps measured here.

Previous experiments have shown that song tempo is positively correlated with brain temperature (Long and Fee 2008; Andalman et al. 2011) and that changes in temperature may explain the increases in song tempo seen when male zebra finches direct their song to females (Aronov and Fee 2012). It is possible that developmental changes in the ability to regulate brain temperature may contribute to increased tempo over development. Indeed, Long and Fee (2008) reported that gap lengths are more sensitive to changes in temperature than syllable lengths, although the syllable/gap difference was much smaller than for the developmental changes in duration reported here.

Consolidation of sequencing and timing in other systems. Overall, the data suggest a phase of development in which syllables are eventually consolidated into a longer, precisely timed chaining mechanism that can run from beginning to end in a highly reliable order. A similar process of linking simpler chains to form more functional activity patterns has been proposed for the neocortex (Bienenstock 1995). Such a linking may be viewed from another perspective as chunking actions (in this case, syllables) into units of behavior (motifs) that are less susceptible to neural noise and interference from the environment. Such a process has been proposed and investigated for mammalian action learning (Graybiel 1998; Yin et al. 2009;


