J Neurophysiol 106: 1747–1765, 2011.
First published June 22, 2011; doi:10.1152/jn.00247.2011.

Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds

Lena Veit, Dmitriy Aronov, and Michale S. Fee
McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

Submitted 24 March 2011; accepted in final form 19 June 2011

Veit L, Aronov D, Fee MS. Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds. *J Neurophysiol* 106: 1747–1765, 2011. First published June 22, 2011; doi:10.1152/jn.00247.2011.—How do animals with learned vocalizations coordinate vocal production with respiration? Songbirds such as the zebra finch learn their songs, beginning with highly variable babbling vocalizations known as subsong. After several weeks of practice, zebra finches are able to produce a precisely timed pattern of syllables and silences, precisely coordinated with expiratory and inspiratory pulses (Franz M, Goller F. *J Neurobiol* 51: 129–141, 2002). While respiration in adult song is well described, relatively little is known about respiratory patterns in subsong or about the processes by which respiratory and vocal patterns become coordinated. To address these questions, we recorded thoracic air sac pressure in juvenile zebra finches prior to the appearance of any consistent temporal or acoustic structure in their songs. We found that subsong contains brief inspiratory pulses (50 ms) alternating with longer pulses of sustained expiratory pressure (50–500 ms). In striking contrast to adult song, expiratory pulses often contained multiple (0–8) variably timed syllables separated by expiratory gaps and were only partially vocalized. During development, expiratory pulses became shorter and more stereotyped in duration with shorter and fewer nonvocalized parts. These developmental changes eventually resulted in the production of a single syllable per expiratory pulse and a single inspiratory pulse filling each gap, forming a coordinated sequence similar to that of adult song. To examine the role of forebrain song-control nuclei in the development of respiratory patterns, we performed pressure recordings before and after lesions of nucleus HVC (proper name) and found that this manipulation reverses the developmental trends in measures of the respiratory pattern.

HVC; motor development; respiration; zebra finch

IN HUMANS AND SONGBIRDS, successful vocalization requires the precise coordination of respiratory and vocal motor systems (Levelt 1993; Riede and Goller 2010). How does this coordination develop, and what is the involvement of the forebrain in this process? The songbird is a unique model to study these and other aspects of vocal communication (Doupe and Kuhl 1999; Fee and Scharff 2010). The earliest babbling vocalizations of juvenile birds, called subsong, exhibit highly variable timing and acoustic fluctuations (Marler and Peters 1982; Marler 1991). In a later stage called plastic song, vocalizations acquire stereotyped temporal and acoustic structure (Tchernichovski et al. 2001, 2004). With continued practice, songbirds ultimately learn a precisely timed, stereotyped song that can be a close copy of the adult song (Immelmann 1969; Marler and Peters 1977; Price 1979).

Address for reprint requests and other correspondence: M. S. Fee, MIT, 46-5133, 77 Massachusetts Ave., Cambridge, MA 02139.

The neural systems that control the production and learning of song have been well described. A key brain area for song production is nucleus HVC (used as a proper name) (Nottebohm et al. 1976; Vu et al. 1994), which is involved in the control of stereotyped timing in adult song (Hahnloser et al. 2002; Long and Fee 2008; Long et al. 2010; Schmidt 2003; Yu and Margoliash 1996). In contrast, lesion studies have shown that HVC is not required for the normal production of subsong syllables (Aronov et al. 2008). The anatomy of brain systems involved in the control of respiration is also well described in birds (Kubke et al. 2005; Reinke and Wild 1998; Wild 1993a, 1993b; Wild et al. 1998) and shows parallels to similar nuclei in mammals (Hage and Stefan 2009; Wild 1997).

A great deal is now known about respiratory patterning in adult song (Riede and Goller 2010). Respiration during singing is highly dynamic and includes rapid alternation between periods of strong expiratory and inspiratory pressure pulses (EPs and IPs, respectively) (Brackenbury 1980; Hartley and Suthers 1989; Hartley 1990; Wild et al. 1998). EPs are usually completely filled by a single syllable, and the silent gaps between syllables are completely filled with an IP, or "mini-breath," resulting in a precise one-to-one coordination of EPs and IPs with syllables and gaps (Franz and Goller 2002). However, less is known about respiratory patterns during subsong. Recordings in juvenile cardinals have revealed that respiratory pressure and syringeal airflow are poorly synchronized with phonation, including periods of nonvocalized expiratory flow and multiple syllables within a single EP (Suthers and Goller 1998; Suthers 1999, 2004). Here we set out to quantitatively describe the nature of respiratory patterns in subsong and early plastic song in the zebra finch. We address three main issues: *1*) How well are respiratory and vocal patterns coordinated? *2*) How does this coordination develop as the bird progresses from subsong into plastic song? *3*) Is the song motor pathway involved in these early developmental changes? To address these questions, we recorded thoracic air sac pressure in young juvenile zebra finches, quantified the developing relation between vocal and respiratory patterning, and carried out lesions of HVC to determine which aspects of this behavior involve this key forebrain nucleus.

MATERIALS AND METHODS

Animal care and experiments were carried out in accordance with National Institutes of Health guidelines and were reviewed and approved by the Massachusetts Institute of Technology (MIT) Institutional Animal Care and Use Committee. Subjects were juvenile male zebra finches obtained from the MIT breeding facility. Birds were raised by their parents in individual breeding cages, during

which time the birds were presumably tutored by their father. Male juveniles were moved at age 31–33 dph to custom sound isolation chambers (maintained on a 12:12-h day-night schedule). After the birds began to sing, typically within a few days, a pressure sensor was surgically implanted. Birds typically began to sing again 1–2 days after surgery. Because our experiments were carried out in very young birds before any song imitation was learned, the song or the pressure pattern of the tutor was not recorded.

Pressure Recordings

To monitor the breathing pattern during subsong and early plastic song, pressure sensors were implanted and chronically carried by the birds, by a method similar to that previously described for adult birds (Franz and Goller 2002; Hartley and Suthers 1989). Before surgery the animals were anesthetized with 1–2% isoflurane in oxygen. A thin, flexible cannula of nonbioreactive Silastic tubing (0.9-mm OD, RenaSil, Braintree Scientific, Braintree, MA) was inserted through the abdominal wall into the posterior thoracic air sac on one side of the body. This tube was passed between the most posterior two ribs (1.5-mm intercostal spacing) and was sutured to one rib. Before surgery, the end of this tubing to be placed in the air sac was inserted into and attached to (with a nonbioreactive silicone elastomer, Kwik-Kast, WPI) a short length (5 mm) of a larger diameter tube (2-mm OD) that was fully inserted into the air sac and helped maintain a clear internal opening of the tubing. The other end of the 0.9-mm tube was fed to a miniature piezoresistive pressure transducer (Fujikura FPM-02PG). The transducer was placed externally on the bird's back and held in place by a loop of thin silicone tube under the skin on the bird's back. The pressure signal was amplified directly on the device and transmitted to the recording computer by a custom-made cable and mercury commutator, which allowed free movement of the bird inside the cage. Before implantation, the pressure sensor's output voltage was calibrated in units of centimeters of water (cmH_2O) with a water column.

Sound Recordings

Vocalizations were recorded with a miniature microphone (Knowles Electronics, Itasca, IL) that was mounted on the same cable used for pressure recordings, close to the bird's head. The birds were constantly monitored by custom-written software, which triggered simultaneous sound and pressure recordings at the onset of song. Recordings were digitized at 40 kHz. For the analysis of high-frequency inspiratory sounds, we additionally used high-sensitivity microphones (model 40AE, G.R.A.S. Sound and Vibration, Holte, Denmark), positioned above the bird's cage, recorded with Sound Analysis Pro software (Tchernichovski et al. 2000).

HVC Lesions

In some animals ($n = 7$) HVC lesions were performed in a second surgery at ages 46–66 dph, after enough developmental data had been recorded. Before surgery, birds were anesthetized with 1–2% isoflurane in oxygen and placed in a stereotaxic apparatus. Craniotomies were made bilaterally above HVC. Lesions were made with a platinum-iridium electrode (Micro Probe; 100-μA current for 60 s). A two-dimensional lattice of 13 lesions spaced at 250 μm was made in each hemisphere for complete bilateral lesions (Aronov et al. 2008). Birds were returned to isolation chambers, and vocalizations resumed typically 1–2 days after the surgery. After the experiment, animals were deeply anesthetized and perfused with phosphate-buffered saline (PBS) followed by 3–4% paraformaldehyde. The brain was extracted and sliced parasagittally for histological confirmation of the lesions.

Data Analysis

All analyses were performed with custom-written software in Matlab.

Sound segmentation, analysis of amplitudes and transition times. Sound recordings were inspected manually to remove calls and cage noise. Song syllables were segmented and characterized based on sound between 1 and 4 kHz as follows: In each recorded sound file, the sound level was determined by band-pass filtering the microphone signal between 1 and 4 kHz (order-80 linear-phase FIR filter created with Matlab function fir1), squaring, and smoothing with a 2.5-ms sliding window. Sound level was converted to units of decibels by computing the logarithm (base 10), multiplying by 10, and normalizing with reference to a calibrated source placed in the recording cage (with a Brüel & Kjaer model 2236 calibrated microphone). The result was expressed in units of decibels sound pressure level (dB_{SPL}) and is referred to in the text and figures as song or syllable "amplitude."

Song amplitude in the 1–4 kHz band during singing was bimodally distributed, corresponding to syllables and gaps (see Fig. 4). Means and SDs of these two modes were estimated by fitting a mixture of two Gaussians to the amplitude distribution with expectation maximization (EM). To segment sound into syllables and gaps, we then defined an "upper threshold" as the Fisher discriminant of the two identified Gaussian modes and a "lower threshold" as 2 SDs above the mean of all values below the upper threshold. Candidate syllable onsets were defined as upward crossings of the lower threshold that were followed by an upward crossing of the upper threshold. Similarly, candidate syllable offsets were defined as downward crossings of the lower threshold that were preceded by a downward crossing of the upper threshold. Candidate syllables shorter than 7 ms were eliminated, and the surrounding gaps were merged. Similarly, candidate gaps shorter than 7 ms were eliminated and the surrounding syllables merged.

To calculate the rate of change of amplitude at syllable onsets and offsets (see Fig. 4), 100 syllables from each recording day were selected randomly, and the average of the onset- and offset-aligned amplitude traces were calculated. Since we are particularly interested in respiratory-vocal coordination, it was especially important to demonstrate the robustness of our identification of onsets and offsets of those syllables near the onsets and offsets of EPs. For this reason, the onset rate analysis (see Fig. 4*C*, *left*) was done by analyzing only first syllables in an EP, and the offset rate analysis (see Fig. 4*C*, *right*) was done by analyzing only last syllables in an EP. The slope of the onset trace (in dB/ms) was then calculated in a 2-ms window after syllable onset, and the slope of the offset trace was calculated in a 2-ms window before syllable offset. Similar results were obtained for syllable offsets and onsets within the interior of EPs.

Song bouts were defined as sequences of consecutive syllables with no silences longer than 300 ms. All analyses were carried out on the segments of time between the beginning of the first vocalized EP to the end of the last vocalized EP in each song bout. Note that song bouts often begin or end with nonvocalized EPs. Thus the histogram in Fig. 5*B* represents an underestimate of the occurrence of entirely nonvocalized EPs because it does not include nonvocalized EPs that occur before the first syllable or after the last syllable of a bout.

High-frequency inspiratory sounds (HFIS) were detected within gaps by filtering the microphone signal between 5 and 19.5 kHz, squaring and smoothing with a 2.5-ms sliding window, and calculating the sound level in decibels. HFIS onsets and offsets were detected as threshold crossings of this signal >5 dB above baseline.

The position and width of the two peaks of the bimodal pressure distributions were determined from the fit of a sum of two Gaussians to the pressure distribution. The central minimum was found as the minimum of the distribution between the two peaks. All probability densities (e.g., Fig. 1, *B–G*, Fig. 2, *B–D*) have units that are the inverse of the quantities expressed on the *x*-axis. The distributions are first calculated as probability per bin and then

divided by bin size. Thus pressure distributions (e.g., Fig. 1*B* and *C*) have units of $[\mathrm{cmH_2O}]^{-1}$.

The "peakiness" of the EP distribution was calculated by fitting the EP distribution to an asymmetric Gaussian distribution. A function of the form

$$P(\tau) = \alpha\tau e^{-\left(\frac{\tau+\beta}{\beta}\right)^2}$$

was found to fit the distribution of subsong EP durations well, where P is the probability per unit time of finding an EP of duration τ, and α and β are free parameters specifying the amplitude and width of the distribution, respectively. As EP distributions develop narrow peaks during early plastic song, this simple function fails to fit the distribution. Thus the squared fitting error of this function is used as a measure of the peakiness of the EP distribution. The fitting was carried out with the Matlab curve fitting toolbox.

The protosyllable-containing EP (PSEP) distributions show the duration of all EPs that carry at least one syllable within a 10-ms window around the protosyllable duration. The peaks of this distribution were determined by smoothing with a 50-ms full-width Gaussian window and fitting three Gaussians to the distribution.

Definition of subsong. Syllable durations were initially analyzed by fitting an exponential function to their duration distribution with maximum-likelihood estimation (MLE). On a finite interval (a, b), maximum-likelihood analysis yields the equation

$$\langle S_i \rangle_{a<S_i<b} = \tau' + a - \frac{(b-a)e^{-(b-a)/\tau'}}{1-e^{-(b-a)/\tau'}}$$

Here the left side of the equation indicates the mean of all syllables with durations s_i between a and b, and τ' is a time constant of the exponential distribution that has the maximal probability of producing the observed data. We used the Matlab zero-finding algorithm (fzero function) to solve the above equation for τ'.

Because syllable duration distributions in subsong were consistently exponential above 25 ms, we used the MLE procedure to fit exponential distributions to syllable durations between $a = 25$ ms and $b = 400$ ms. This analysis was carried out on song data collected during 1 day of singing (1–10 thousand syllables). The goodness of fit of the exponential was estimated with the Lilliefors statistic (Lilliefors 1969), a procedure similar to the Kolmogorov-Smirnov test (but suitable for evaluating fits). This involves calculating the maximum difference between the cumulative density functions of the data distribution and the best exponential fit and then normalizing this difference by the square root of the number of syllables. Distributions that were well-fit by exponentials typically had a goodness-of-fit metric <2, whereas distributions that were just beginning to exhibit a protosyllable peak typically had values >2. We therefore used 2 as the threshold for distinguishing subsong from plastic song.

It has been shown (Aronov et al., 2010) that the distribution of syllable durations is well-fit by an exponential distribution in young birds. With the transition to early plastic song and the development of a consistently timed "protosyllable" (Liu et al. 2004; Tchernichovski et al. 2004; Aronov et al., 2010) the syllable distribution develops a peak at the durations of the protosyllable, and the maximum deviation from the cumulative distribution function of the best-fit exponential distribution (Lilliefors goodness-of-fit statistic) shows a sudden increase. The precise age at which birds reach this point in their development differs among birds, so we chose to align birds based on their relative song maturity to compare their developmental progression in this study. The first day that crossed our limit for nonexponentiality, and therefore showed development of a protosyllable, was designated as relative *age 0*. All recording days were then aligned relative to *age 0* for developmental population analysis. We quantified the features of early vocalizations in three groups of different developmental stages: subsong (*ages −7* to *−2*), early plastic song (*ages 2* to *4*), and late plastic song (*ages 8* to *25*). We allowed a 3-day gap between subsong and early plastic song (*ages −1*, *0*, and *1*) to more clearly highlight the population differences between these age groups. The values of the parameters for these different developmental stages were obtained by first calculating an average for each bird and then calculating the mean and SE across all birds that were recorded in the respective age bracket. The daily change was calculated from *age −8* to *2*. If a bird was not recorded continuously in this period, we divided the change between the recorded data points by the number of days separating them to get an estimate of the daily change. In two birds the pressure signal deteriorated after ~1 day, presumably because of fluid that entered the tube. These birds were excluded from the analysis of daily change.

RESULTS

Characterization of Respiratory Patterns

We recorded thoracic air sac pressure in juvenile zebra finches ($n = 14$) during subsong and plastic song stages (ages 38–78 dph). To compare some aspects of the juvenile behavior to adult song, thoracic air sac pressure was recorded in adult birds as well ($n = 5$). During nonsinging periods, air sac pressure in subsong birds showed characteristic rhythmic eupneic (quiet) breathing at roughly 3 Hz (range: 1.5–3.5 Hz; Fig. 1, *A* and *B*). During subsong production, respiratory patterns had a larger amplitude and exhibited a bimodal distribution of pressures (Fig. 1*C*; the central minimum of the distribution was 65.2 ± 3.2% of the smaller negative peak), similar to that observed in adult song (Fig. 2, *A* and *B*). The positive peak during subsong was at a pressure 9.5 ± 2.0 times higher than that during quiet respiration. The negative peak during singing was at a pressure 3.1 ± 0.6 times higher in magnitude than during quiet respiration.

What are the timescales of the inspiratory and expiratory components of early vocal behavior? While periods of positive (expiratory) pressure alternated with periods of negative (inspiratory) pressure in a highly irregular way during subsong, a number of regularities were noted. Inspiratory pulses (IPs) were usually characterized by a single negative deflection of pressure (Fig. 1*A*), similar to those observed in the minibreaths of adult song. IPs in subsong had an average duration of 49.3 ± 3.2 ms ($n = 8$ subsong birds), significantly longer than IPs in adult song (Fig. 2*C*, Table 1; $P < 0.01$), and the distribution of IP durations within subsong birds had an average width of 22.2 ± 1.8 ms (SD) (Fig. 1, *D* and *E*; 10–90% range of IP durations from 22.1 ± 1.6 to 77.8.0 ± 5.7 ms).

In contrast to IPs, expiratory pulses (EPs) were longer (average duration 188.7 ± 8.2 ms) and had more complex pressure waveforms, often exhibiting multiple peaks within a single EP, as observed in adult song. However, subsong EP waveforms were highly variable (Fig. 1*A*), with no evidence of the stereotyped pressure patterns previously described for individual adult song syllables (e.g., Fig. 2*A*). Subsong EP durations were broadly distributed (Fig. 1, *F* and *G*; 115.8 ms SD, 10–90% range of durations from 65.8 ± 4.1 ms to 340.1 ± 20.6 ms), without the narrow peaks characteristic of EP duration distributions in adult birds (Fig. 2*D*).

Vocal signals During Expiratory and Inspiratory Pressures

We were interested in how the respiratory pressure relates to the sounds that are produced during subsong. Vocal signals were

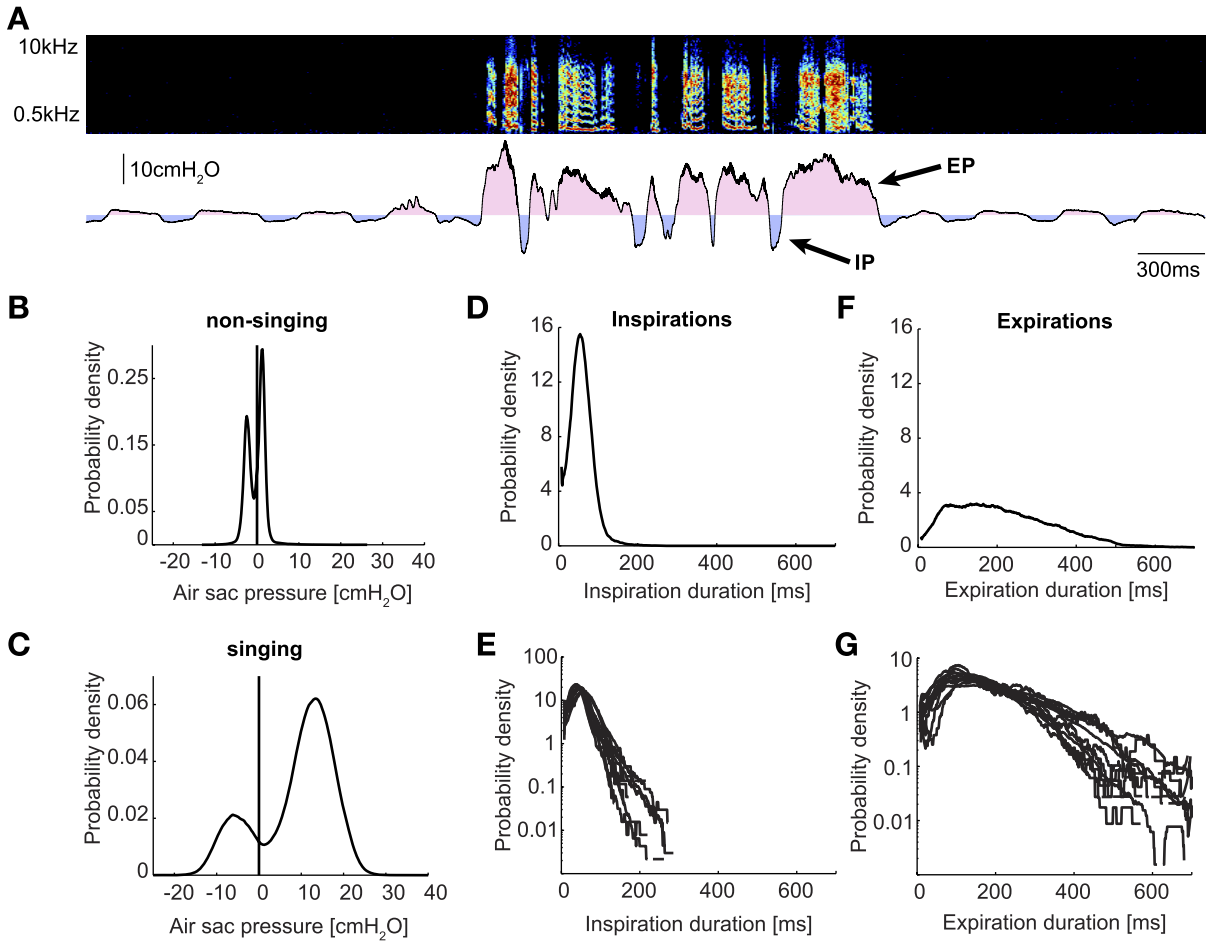


Fig. 1. Respiratory patterns in subsong. *A*: song spectrogram (*top*) showing a bout of singing in a young zebra finch (*bird 1*, 47 dph) and recording of thoracic air sac pressure during subsong (*bottom*). Note the shallow regular eupneic breathing before and after the song bout and the high-intensity inspiratory and expiratory periods during singing. The pressure pattern during singing is divided into inspiratory pressure pulses (IPs, blue shading) and expiratory pressure pulses (EPs, pink shading). *B*: distribution of air sac pressure values during nonsinging. Expiratory and inspiratory peaks were, on average, located at -1.5 ± 0.3 cmH_2O and $+1.4 \pm 0.2$ cmH_2O, respectively. *C*: distribution of pressure values during singing. On average, expiratory and inspiratory peaks were located at 14.3 ± 1.1 cmH_2O and -4.4 ± 0.5 cmH_2O, respectively, and the central minimum was located at 1.4 ± 0.9 cmH_2O (not significantly different from ambient pressure). The bimodal pressure distributions during singing suggest the presence of distinct inspiratory and expiratory phases. *D*: distribution of IP durations from 1 bird, showing a stereotyped duration around 50 ms. *E*: IP duration distributions from all subsong birds on a semilogarithmic plot. *F*: distribution of EP durations from 1 bird, showing a much broader distribution centered around 100 ms. *G*: EP duration distributions from all subsong birds (semilog plot). Data in *A–D* and *F* are from *bird 1*, 47 dph.

observed over a wide range of respiratory pressures (Fig. 3*A*). To characterize the relation between respiratory pressure and sound produced, we first examined average acoustic power (in a 0.5–15 kHz band) of the song signal at different air sac pressure levels (Fig. 3*B*). The acoustic signal had high power primarily at positive air sac pressure levels, a minimum of average sound power produced at ambient pressures, and a smaller peak of power at large negative pressures.

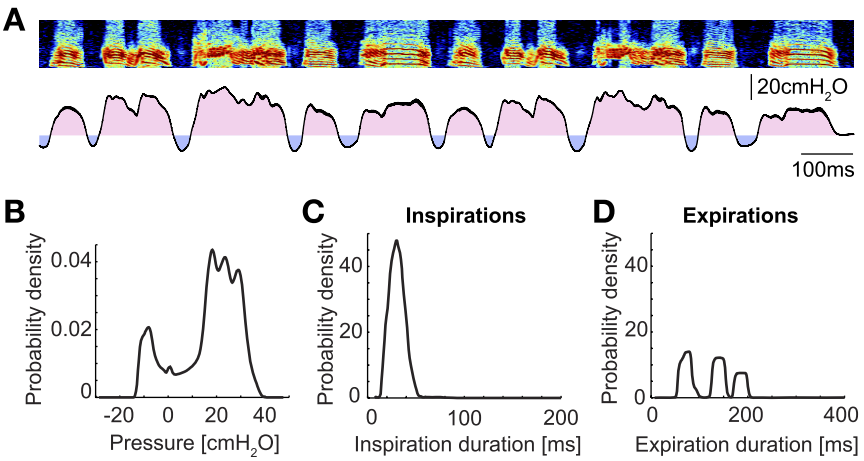


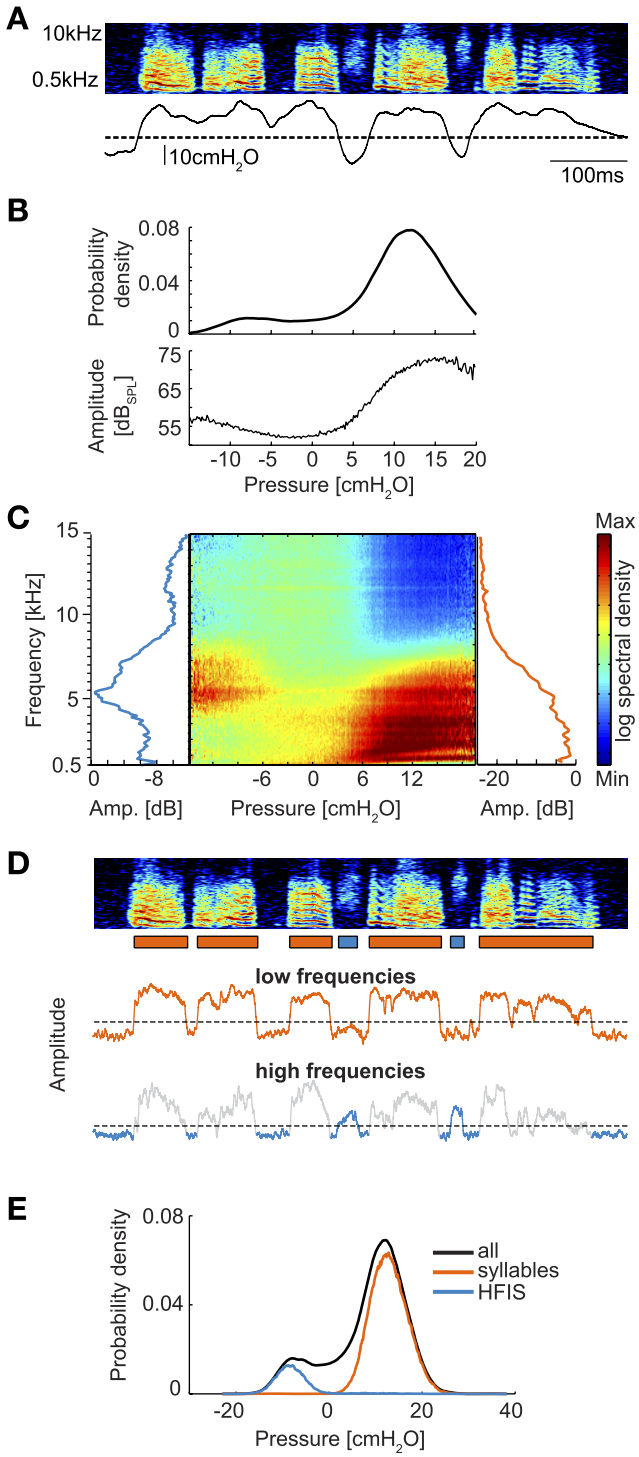
Fig. 2. Respiratory parameters in adult song. *A*: song spectrogram (*top*) showing 2 motifs of singing in an adult zebra finch and recording of thoracic air sac pressure during singing (*bottom*). *B*: distribution of air sac pressure values during singing. As for subsong, the distribution is bimodal for expiratory and inspiratory pressures, but the positive peak exhibits multiple small peaks possibly associated with different syllables. *C*: distribution of IP durations during singing for the same bird. Note that the time axis is different from Fig. 1*D*. *D*: distribution of EP durations during singing for the same bird. There are distinct peaks corresponding to the different syllables. All examples in this figure are from *adult 1*.

Table 1. *Overview of all quantified parameters at different ages and after HVC lesion*

	HVC Lesion ($n = 7$)	Subsong (relative *age* −7 to −2) ($n = 8$)	Early Plastic (relative *age* 2 to 4) ($n = 8$)	Late Plastic (relative *age* 8 to 25) ($n = 6$)	Adult ($n = 5$)	% Daily Change	Subsong vs. Early Plastic Song
EP dur, ms	**213.9 ± 22.8**	**188.7 ± 8.2**	**141.8 ± 12.8 (*S)**	**125.9 ± 4.8**	**134.6 ± 19**	**−6.1 ± 1.1 (*)**	***P* < 0.01**
IP dur, ms	**65.3 ± 7.1**	**49.3 ± 3.2**	**39.9 ± 2.4 (*S)**	**36.3 ± 1.5**	**26.6 ± 1.0 (*L)**	**−2.3 ± 1.1 (*)**	***P* < 0.05**
IP rate, Hz	**3.8 ± 0.3**	**4.1 ± 0.2**	**5.7 ± 0.4 (*S)**	**6.2 ± 0.2**	**6.6 ± 0.6**	**6.8 ± 1.3 (*)**	***P* < 0.01**
Max. pressure, cmH_2O	**25.4 ± 4.0**	**27.0 ± 3.4**	**32.1 ± 1.7**	**33.4 ± 3.2**	**37.8 ± 2.7**	**6.1 ± 1.5 (*)**	***P* = 0.15**
Min. pressure, cmH_2O	−12.2 ± 2.6	−18.3 ± 2.0	−16.1 ± 1.4	−12.7 ± 0.9	−14.1 ± 0.9	−0.8 ± 1.2	*P* = 0.39
Syllable amp, dB_{SPL}	69.3 ± 1.1	71.1 ± 1.0	72.7 ± 1.2	75.1 ± 1.6	81.5 ± 0.6 (*L)	−0.8 ± 0.5	*P* = 0.30
Syll. onset speed, dB/ms	6.4 ± 0.6	6.2 ± 0.3	6.6 ± 0.4	6.3 ± 0.3	7.0 ± 0.8	2.1 ± 2.5	*P* = 0.38
Syll. offset speed, dB/ms	3.0 ± 0.3 (*P)	4.4 ± 0.2	3.9 ± 0.2	4.8 ± 0.4	6.0 ± 0.2 (*L)	−0.3 ± 2.1	*P* = 0.16
Syllables per EP	**1.5 ± 0.1**	**1.5 ± 0.1**	**1.3 ± 0.0 (*S)**	**1.2 ± 0.1**	**1.0 ± 0.0 (*L)**	**−2.3 ± 0.8 (*)**	***P* < 0.01**
Expiratory gaps, %	**39.2 ± 5.8**	**41.6 ± 2.5**	**25.9 ± 2.4 (*S)**	**20.2 ± 4.6**	**0.8 ± 0.6 (*L)**	**−8.0 ± 2.0 (*)**	***P* < 0.001**
Vocalized fraction, %	**66.8 ± 2.3**	**67.3 ± 2.9**	**74.5 ± 3.7**	**80.8 ± 3.2**	**96.1 ± 0.6 (*L)**	**2.8 ± 0.7 (*)**	***P* = 0.15**
EP onset dur, ms	**24.7 ± 2.6**	**19.0 ± 1.5**	**13.2 ± 1.2 (*S)**	**12.3 ± 2.3**	**4.4 ± 0.5 (*L)**	**−6.6 ± 1.6 (*)**	***P* < 0.01**
Expiratory gap dur, ms	**33.3 ± 3.2**	**31.0 ± 2.1**	**28.2 ± 2.6**	**23.6 ± 2.0**	**3.6 ± 2.2 (*L)**	**−4.8 ± 1.4 (*)**	***P* = 0.40**
EP offset dur, ms	**33.6 ± 5.3**	**25.3 ± 3.4**	**11.8 ± 2.1 (*S)**	**6.1 ± 0.7 (*E)**	**1.9 ± 0.5 (*L)**	**−14.9 ± 2.4 (*)**	***P* < 0.01**
EPs filled by syllable, %	**23.1 ± 8.5 (*P)**	**44.1 ± 4.5**	**72.2 ± 3.8 (*S)**	**82.9 ± 3.8**	**98.6 ± 0.6 (*L)**	**18.8 ± 4.3 (*)**	***P* < 0.001**
Gaps filled by IP, %	**27.8 ± 5.7 (*P)**	**46.0 ± 4.7**	**70.0 ± 4.5 (*S)**	**78.2 ± 2.2**	**99.2 ± 0.3 (*L)**	**12.0 ± 2.4 (*)**	***P* < 0.001**
EP fitting error, s^{-1}	**0.5 ± 0.3**	**0.1 ± 0.0**	**1.0 ± 0.2 (*S)**	**2.2 ± 0.2 (*E)**	**6.0 ± 1.7 (*L)**	**92.1 ± 29.4 (*)**	***P* < 0.001**

Values are means ± SE. EP dur, average duration of expiratory pressure pulses (EPs); IP dur, average duration of inspiratory pressure pulses (IPs); IP rate, average rate of IPs; Max. pressure, 99th percentile of pressure distribution; Min. pressure, 1st percentile of pressure distribution; syllable amp, amplitude of the syllable peak in the amplitude distribution; Syll. onset speed, speed of amplitude change for 2 ms after syllable onset; Syll. offset speed, speed of amplitude change for 2 ms before syllable offset; syllables per EP, mean no. of syllables per EP; expiratory gaps, fraction (%) of gaps that do not contain IPs; vocalized fraction, average fraction of each EP that is filled with syllables; EP onset dur, time between EP onset and syllable onset; expiratory gap dur, average duration of expiratory gaps; EP offset dur, time between syllable offset and EP offset; EPs filled by syllable, fraction (%) of EPs that are completely filled by a single syllable (syllable duration within 25 ms of EP duration); gaps filled by IP, fraction of gaps that are filled by a single IP (IP duration within 25 ms of gap duration); EP fitting error, error of fitting an asymmetric Gaussian to the EP distribution (measures "peakiness" of EP duration distribution); %daily change, mean within-bird daily change over a 10-day period from day −8 to day 2 (see MATERIALS AND METHODS) in %; subsong vs. plastic: *t*-test between subsong ($n = 8$) and early plastic song ($n = 8$). Symbols (*S), (*E), and (*L) mean that the indicated quantity is significantly different from the subsong, early plastic song, and late plastic song values, respectively. Rows indicated in boldface showed a significant within-bird daily change between subsong and early plastic song ($P < 0.05$, *t*-test). Asterisks without a letter indicate that the mean within-bird daily change is significantly different from zero.

Sounds produced during high positive air sac pressure differed substantially in spectral content from the sounds during negative pressures. The spectral composition of sounds at different pressures was determined by binning and averaging time slices of the song spectrogram by pressure value (Fig. 3*C*). The sounds produced at pressures between 10 and 18 cmH_2O (which tended to produce the loudest sounds) were primarily noisy sounds, often with rich harmonic structure (Fig. 3*A*), that had dominant spectral components at low frequencies between 1 and 5 kHz (Fig. 3*C*, orange trace at *right*). Subsong syllables were defined on the basis of high acoustic power in a band from 1 to 4 kHz (Fig. 3*D*; see MATERIALS AND METHODS). Defined in this way, subsong syllables were produced exclusively during expiratory pulses (Fig. 3*E*; 99.6 ± 0.2% of all time points within detected syllables occurred at positive pressures).



In contrast to subsong syllables, sounds produced at negative (inspiratory) pressures (−7 to −15 cmH_2O) were dominated by a high-frequency broadband noisy peak between 4 and 15 kHz (peaked at 5 kHz for the bird shown in Fig. 3*C*; blue trace at *left*). These high-frequency inspiratory sounds (HFISs) were observed in most subsong birds and resembled fricatives or breath sounds when played at a lower speed. HFISs showed a sufficiently distinct spectral composition that they could be detected within the silent intervals between syllables as peaks in power of the acoustic signal filtered between 5 and 19.5 kHz (Fig. 3*D*; see MATERIALS AND METHODS). Almost all pressures below the negative peak in the pressure distribution contained detected HFISs, suggesting that most IPs could be detected by using only the acoustic signal (Fig. 3*E*).

Subsong Syllables Exhibit Rapid Onsets and Offsets

Our approach to characterizing respiratory-vocal coordination in early juvenile song relies on an analysis of the timing between syllable onsets and offsets and EP onsets and offsets. For such an analysis to be meaningful, it is necessary that syllable onsets and offsets can be reliably determined and that this determination is consistent over development. Greatly simplifying the assignment of subsong syllable onset and offset times is the fact that the low-frequency (1–4 kHz) sounds used to define subsong syllables exhibited a strongly bimodal distribution of sound amplitudes—one peak corresponding to the presence of a syllable and the other peak corresponding to silent gaps between syllables (Fig. 4, *A* and *B*). In subsong, the separation between the two peaks was typically about three times the width of the vocalized peak, indicating a fairly clean separation of the amplitude of syllables from gaps, even in

Fig. 3. Relation between song acoustic structure and air sac pressure. *A*: spectrogram (*top*) of a segment of subsong (*bird 8*, 52 dph, all examples in this figure) together with simultaneously recorded air sac pressure (*bottom*). Note that vocal output occurs over a wide range of pressures, including during inspiration. *B*: plot of average song amplitude [in units of decibels sound pressure level (dB_{SPL}), *bottom*; see MATERIALS AND METHODS], as a function of air sac pressure. Note the increased amplitude at both high and low absolute pressures (pressure distribution at *top*). *C*: average spectrum (*y*-axis) of vocal output produced at different air sac pressures (*x*-axis). At *left* is the spectrum averaged over inspiratory pressures (−15 to −7 cmH_2O), and at *right* is the spectrum averaged over expiratory pressures (10 to 18 cmH_2O). Note that the loud sounds produced at positive pressures show a concentration of power at low frequencies between 1 and 5 kHz. Broadband high-frequency sounds produced during inspirations are concentrated at higher frequencies, and sound like sharp inward breaths. *D*: subsong syllables are segmented based on acoustic intensity in the band from 1 to 4 kHz (low-frequency sounds, orange trace and bars). High-frequency inspiratory sounds (HFISs) are detected between song syllables by high acoustic intensity in the band from 5 to 19.5 kHz (blue trace and bars). *E*: distribution of air sac pressures during detected syllables (orange curve) and during detected HFISs (blue curve) and full pressure distribution (black trace).

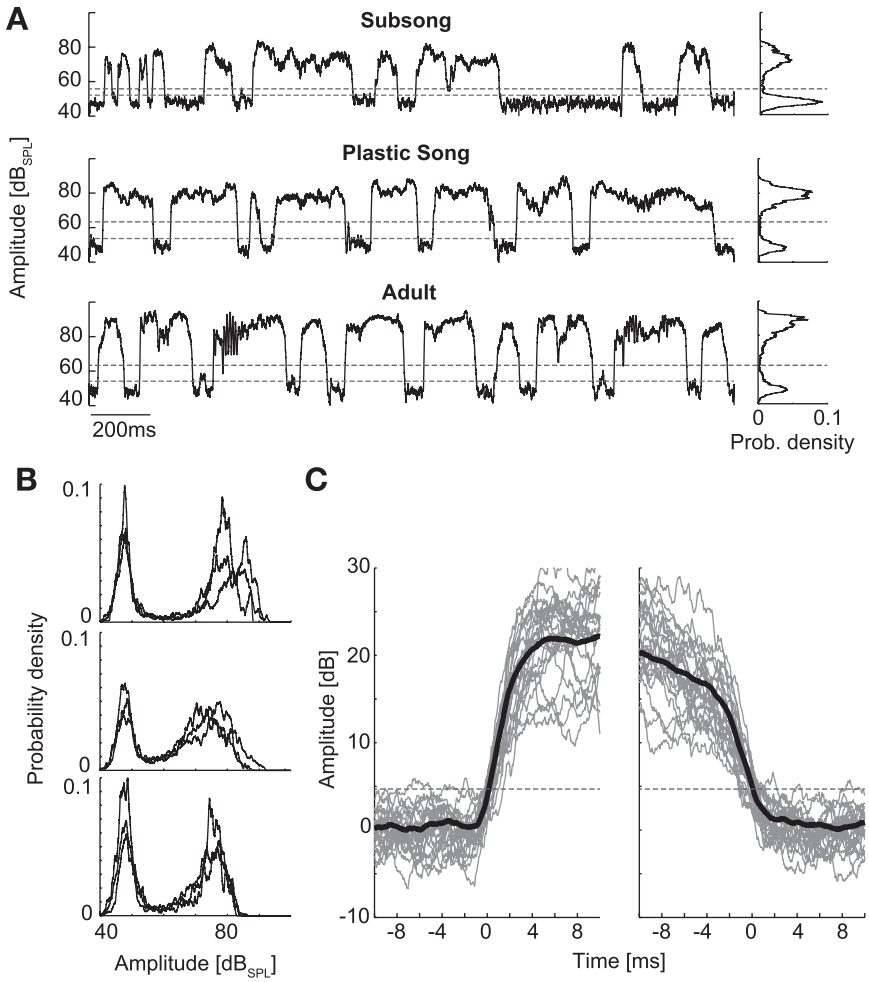


Fig. 4. Characterization of subsong syllables. *A*: traces of song amplitude (log units, dB_{SPL}) during a bout of subsong (*top*), plastic song (*middle*), and adult song (*bottom*). Note the strongly bimodal nature of song amplitude at all developmental stages (distributions at *right*). *B*: distribution of sound amplitude for 3 subsong birds (*top*: *bird 7*, 49 dph; *middle*: *bird 13*, 46 dph; *bottom*: *bird 3*, 43 dph). For each bird, the distribution is shown (overlaid traces) for 3 different files recorded on the same day. *C*: average temporal profile (thick line) of song amplitude at syllable onset (*left*) and offset (*right*). Also shown are amplitude profiles (gray traces) from 25 individual examples. Amplitude is shown relative to average amplitude of background noise (0 dB) for all traces. Also shown is the lower threshold used to determine syllable onset time (dashed line, see MATERIALS AND METHODS).

the earliest song [see Table 1; average separation between peaks in subsong, 24.4 ± 1.0 dB; average (SD) of vocalized peak, 7.6 ± 0.2 dB].

Although the width of the distribution of syllable amplitudes showed no significant developmental trend ($P > 0.5$, *t*-test, subsong compared with adult song), there was a trend toward higher average syllable amplitude, which increased by 1.6 ± 1.5 dB between subsong and early plastic song (see MATERIALS AND METHODS for definition of song stages) and by 10.4 ± 1.6 dB between subsong and adult song ($P < 0.05$, subsong compared with adult song). In adult song, the separation between the vocalized and nonvocalized peaks in the amplitude distribution was roughly 4.5 times the SD of the vocalized peak.

Further simplifying the identification of subsong syllable onsets and offsets is that the transition times were extremely fast, typically within 2–4 ms (Fig. 4*C*). In particular, the rate of amplitude change was 6.2 ± 0.3 dB/ms at syllable onsets and 4.4 ± 0.2 dB/ms at syllable offsets. The mean rate did not change significantly between subsong and later plastic song stages ($P < 0.05$, *t*-test) and was only slightly faster in adults (see Table 1). Because of the fast transitions, the determination of syllable onset and offset times is highly robust to overall changes in song amplitude during development, or to small differences in the choice of threshold. For example, assuming a fixed syllable-detection threshold, the 1.6-dB change in song amplitude between subsong and plastic song could produce at most a 0.3-ms change in the estimate of syllable onset times or a 0.4-ms change in syllable offset times.

Expiratory Pressure Pulses Contain Nonvocalized Regions

Similar to what has been reported in juvenile cardinals (Suthers 2004), we found that in young zebra finches some expiratory pulses did not contain any syllables, while others contained more than one syllable (range 0–8 syllables, Fig. 5, *A* and *B*). To further quantify the relation between the vocal and respiratory events during subsong, we examined four timescales of interest (Fig. 5*A*): *1*) the duration of the nonvocalized period between the onset of the EP and the beginning of the first syllable (EP onset period, average 19 ± 1.5 ms); *2*) the duration of the syllables (average 72.8 ± 4.5 ms); *3*) the duration of silent gaps between syllables within an EP (expiratory gaps, average 31.0 ± 2.1 ms); and *4*) the nonvocalized period between the last syllable in the EP and the offset of the EP (EP offset period, average 25.3 ± 3.4 ms). The durations of these events were highly variable during singing, and all exhibited a broad distribution from ~10 ms to hundreds of milliseconds (Fig. 5, *C–F*). The distributions decreased monotonically over this range, indicating that long durations were less common than shorter durations. For syllable and expiratory gap durations, the decrease was approximately linear on a semilog plot, indicating

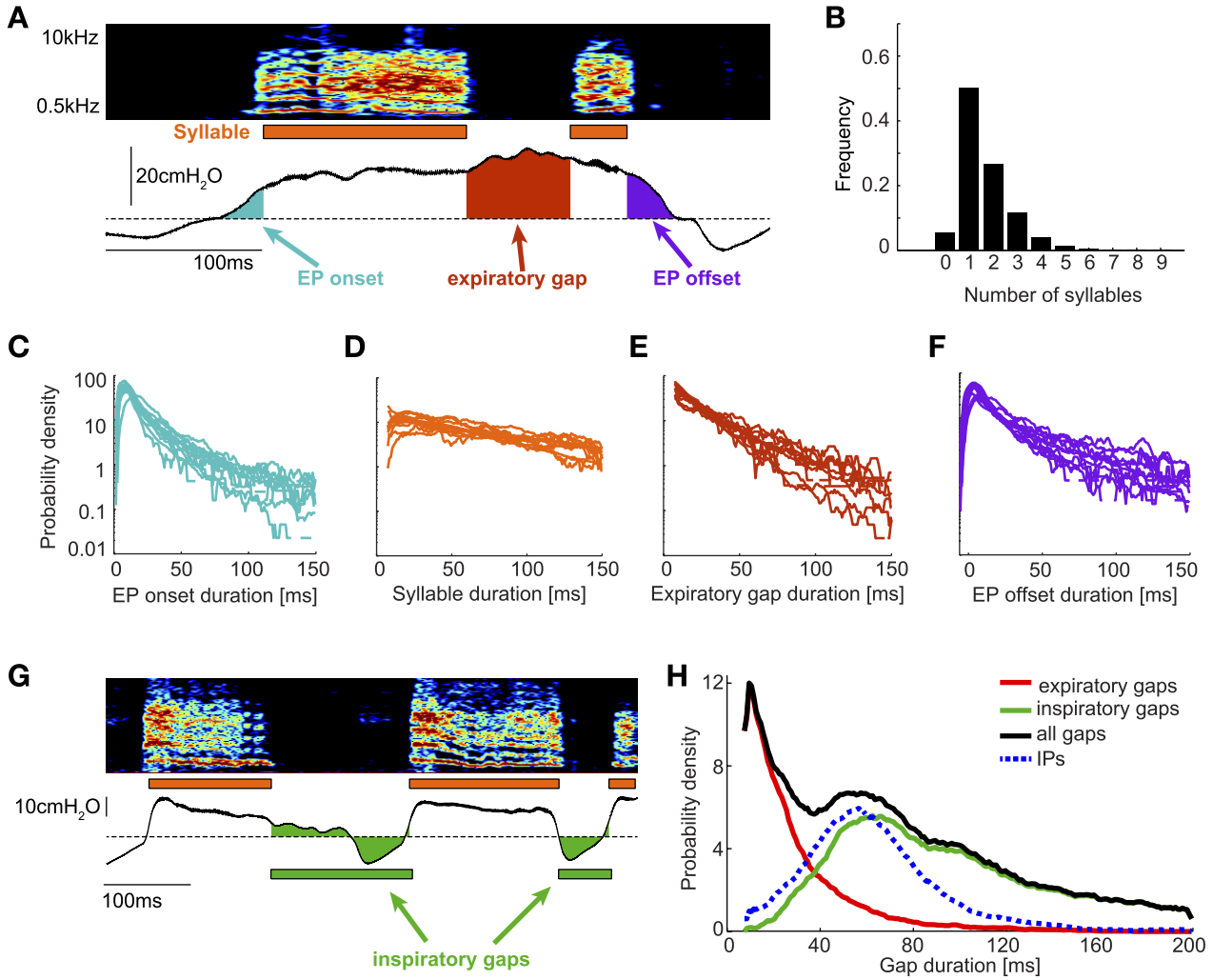


Fig. 5. Relation between subsong respiratory and vocal events. *A*: example of an EP that contains multiple syllables (*bird 9*, 39 dph). The EP can be divided into 4 distinct components: EP onset period, prior to the onset of the first syllable (blue); syllables (orange bars); expiratory gaps between syllables within an EP (red); and EP offset period after the last syllable (purple). *B*: distribution of the number of syllables per EP for all subsong birds. *C–F*: distribution of EP onset durations, syllable durations, expiratory gap durations, and EP offset durations in 12 subsong birds. *G*: examples of 2 gaps that contain inspirations (inspiratory gaps) (*bird 3*, 43 dph). The first gap is much longer than the IP because it contains a long EP offset period. The second gap has very short EP onset and offset periods, and its duration is close to the duration of the IP. *H*: duration distribution of all gaps (black trace), expiratory gaps (red), and inspiratory gaps (green) for the same bird. Note that the peak in the overall gap distribution is from inspiratory gaps that have a duration close to the IP duration (IP duration distribution, blue dashed trace).

an exponential distribution of durations (Fig. 5, *D* and *E*). The shape of these distributions appeared generally consistent across all subsong birds examined (Fig. 5, *C–F*).

In adult song, the silent gaps between syllables are tightly and uniquely linked to inspiratory pulses (Wild et al. 1998). Consistent with this, in the adult zebra finches we recorded, the vast majority of gaps contained an inspiration (99.3%) and the inspirations nearly completely filled the gaps. In contrast, in subsong only 58.4% of gaps within bouts of singing contained an inspiration (Fig. 5, *A* and *G*), the remainder being the expiratory gaps described above (also see MATERIALS AND METHODS). While some inspiratory gaps in subsong were completely filled with IPs, many inspiratory gaps included long periods of nonvocalized expiratory pressures before or after the IP (Fig. 5*G*). As a result, the distributions of gap durations were complex, containing a short component due to brief expiratory gaps and a longer broad component due to inspiratory gaps (Fig. 5*H*).

Developmental Changes in Respiratory-Vocal Patterns in Subsong and Early Plastic Song

We now turn to the developmental changes leading to the coordination of respiratory and vocal patterns. One of the earliest identifiable components of early song vocalizations is syllables with fairly consistent timing, visible by the appearance of a peak in the syllable duration distribution (Fig. 6*A*) (Tchernichovski et al. 2004). We refer to all syllables within this peak as "protosyllables" (Aronov et al., 2010). The development of protosyllables occurs well before any acoustic stereotypy is apparent in the song; thus protosyllables are defined here only on the basis of their duration. The protosyllable peak appears at roughly age 45 dph (range: 35–50 dph; Aronov et al., 2010) and serves as an easily detectable developmental milestone signaling the transition to early plastic song (see MATERIALS AND METHODS). Typically, several days prior to the appearance of the protosyllable peak, a peak in the

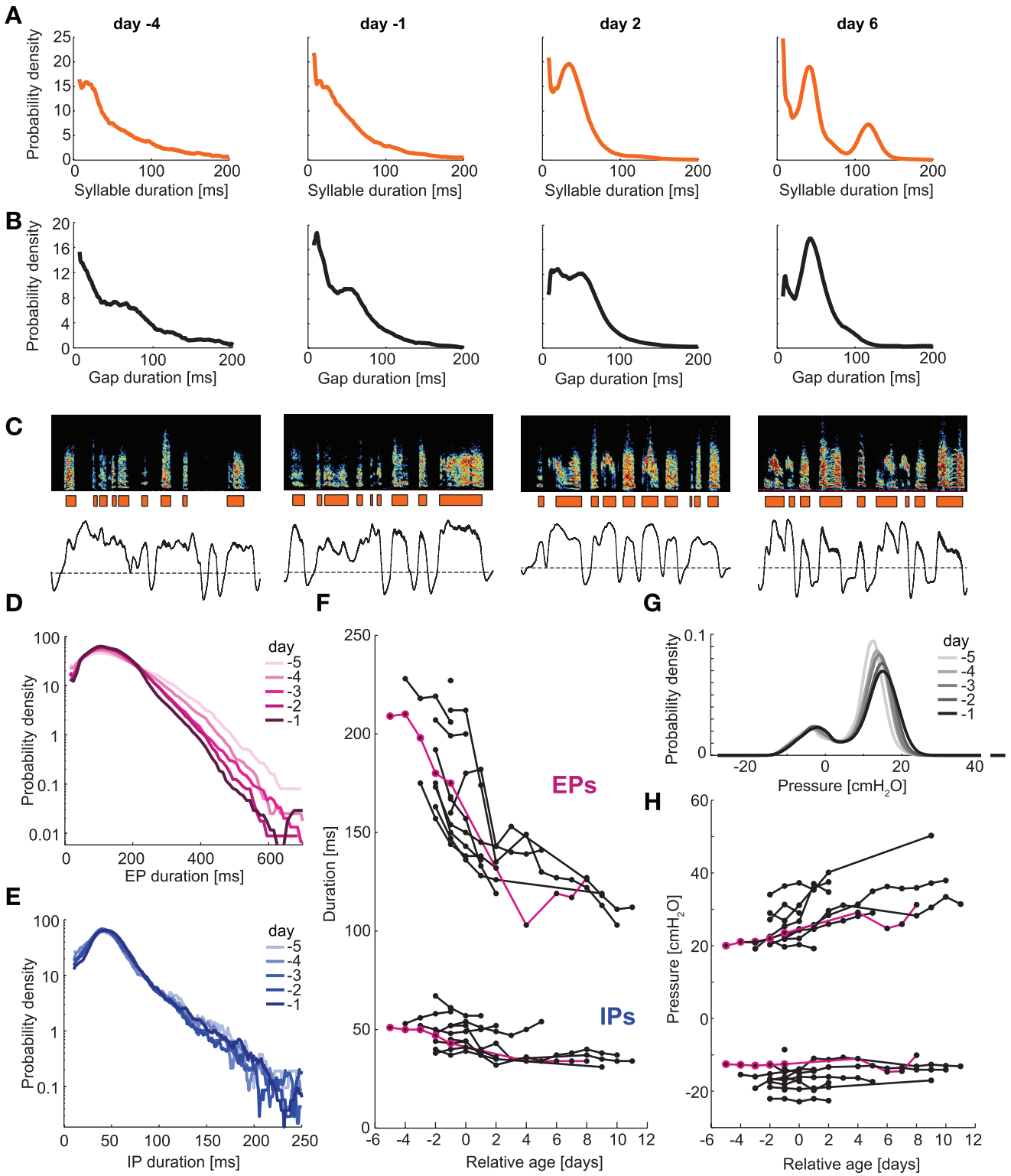


Fig. 6. Developmental progression of vocal and respiratory song features. *A* and *B*: examples of distributions of syllable durations (*A*) and gap durations (*B*) on 4 days during subsong (40 dph and 43 dph) and early plastic song (46 dph and 50 dph) in 1 bird (all examples in this figure are from *bird 9*). Note the early appearance of a peak in the gap distribution (protogaps) and the later appearance of a peak in the syllable distribution (protosyllables). *C*: spectrograms and air sac pressure waveform for singing during the same 4 days shown above. *D*: distribution of EP durations during 5 sequential days of song development (39–43 dph). Note the progressive loss of long EPs. *E*: distribution of IP durations during the same 5 days. *F*: summary plot of developmental progression of average EP durations (*top*) and average IP duration (*bottom*) for all birds recorded in this age range ($n = 12$). Results from different birds are aligned at relative *age 0* according to the day of the first appearance of a peak in the syllable distribution (protosyllable peak, see MATERIALS AND METHODS). Data points from the same bird are shown connected, and points from *D* and *E* are shown in red. *G*: distribution of air sac pressures during singing, showing a developmental increase of peak positive air sac pressures but no change in the peak negative pressures. *H*: summary plot of developmental progression of peak positive (*top*) and negative (*bottom*) pressures.

gap duration distribution appears at roughly 60 ms, corresponding to the duration of inspiratory pressure pulses (Aronov et al., 2010). As with protosyllables, the peak in the gap duration distribution grows rapidly, becoming more pronounced and narrower on a day-to-day basis (Fig. 6*B*). We note that the terms "protosyllable" and "protogap" refer only to the vocal pattern, not to the underlying EPs and IPs.

What changes in the respiratory pattern are associated with the dramatic changes that occur in the syllable and gap patterning during the transition from subsong to plastic song? To address this question, in a subset of birds ($n = 12$) we recorded thoracic air sac pressure continuously through the progression from subsong into plastic song (Fig. 6*C*). Here we carry out a detailed quantitative analysis of the developmental changes to the pressure waveform and its relation to syllables and gaps. All of the developmental analyses we present here are aligned to the day on which the protosyllable peak appears (*day 0*, see MATERIALS AND METHODS).

Developmental Changes in EPs and IPs

One developmental change visually apparent in the respiratory pattern is an increase in the rate of minibreaths (IPs; Fig. 6*C*). The rate of IPs during subsong was only ~4.1 s^{-1}, which increased by 40 ± 8.5% by early plastic song ($P < 0.01$, *t*-test) and further increased into adulthood (Table 1). This change in IP rate was a consequence of changes in the duration of both EPs and IPs (Fig. 6, *D–F*). Overall, expiratory pulses became significantly shorter between subsong and early plastic song (Fig. 6*D*; decrease of 25 ± 10%, $P < 0.01$, *t*-test). The duration of inspiratory pulses also decreased significantly during the same period (Fig. 6*E*; decrease of 19 ± 9%; $P < 0.05$, *t*-test). Both EPs and IPs continued to shorten into adulthood, as determined from air sac pressure measurements made in five separate adult birds (average EP duration, Table 1).

Interestingly, the EPs not only shortened but also increased in amplitude during development (Fig. 6, *G* and *H*). The peak pressure during EPs (defined as the 99th percentile of the distribution of positive pressures) showed a significant daily increase between subsong and early plastic song and continued to increase into adulthood (37% higher in adult than during subsong; see Table 1). In contrast to EPs, the peak inspiratory pressures (the minima of the IPs) did not show a significant developmental change.

EPs also became more completely vocalized with development. On average, the total duration of all subsong syllables within an EP, as a fraction of the EP duration (vocalized fraction), was only 67.3 ± 2.9%. Between subsong and early plastic song, the vocalized fraction showed a significant daily fractional increase (Fig. 7*A*; $P < 0.01$, *t*-test). The vocalized fraction continued to increase throughout song learning, reaching a value of 96.1 ± 0.6% in the adult birds ($P < 0.01$ compared with late plastic song). Consistent with this, we found that expiratory gaps became less frequent throughout development (Fig. 7*B*; 41.6 ± 2.5% of gaps in subsong and only 0.8 ± 0.6% of gaps in adult song, see Table 1). Despite the increased fraction of each EP vocalized, across birds the average duration of syllables did not show a consistent or significant developmental change through subsong into early plastic song ($P = 0.76$; Fig. 7, *D* and *H*).

What is the developmental progression of the relation between syllable onsets and offsets and respiratory transitions? The average EP onset period became significantly shorter between subsong and early plastic song (Fig. 7, *C* and *G*; decrease of 31 ± 12%; $P < 0.01$, *t*-test). The largest developmental change was seen in the EP offset periods, which decreased significantly by 53 ± 22% during this period (Fig. 7, *F* and *K*; $P < 0.001$, *t*-test). The rapid reduction in EP onset and offset intervals during subsong indicates an increasing relative coordination between the onsets and offsets of EPs and syllables.

Changes to Coordination of Syllables and Expiratory Pulses

The developing coordination between respiratory and vocal patterns can also be seen in the joint distribution of syllable durations and EP durations (Fig. 8, *A* and *B*). In subsong, there was only a very weak relation between syllable duration and the duration of the EP on which it is produced. Over the course of a few days, syllables became more closely associated with EPs of similar length, evident from the accumulation of syllables at the diagonal of the joint distribution, indicating EPs that were almost completely filled by a single syllable (Fig. 8*A*). To quantify this, we calculated the fraction of EPs that contained a single syllable with a duration within 25 ms of the EP duration. In subsong, this fraction of "filled EPs" was on average 44.1 ± 4.5%. This fraction increased significantly by early plastic song (Fig. 8*B*; $P < 0.001$) and ultimately reached an average of 98.6 ± 0.6% in adult birds. Similar to the developmental "filling" of EPs with syllables, IPs gradually filled inspiratory gaps more completely (Fig. 8, *C* and *D*). The fraction of IPs that had a duration within 25 ms of the associated gap was only 44.3 ± 4.1% in subsong and increased significantly during plastic song, reaching a value of 99.2 ± 0.3% in adult song.

We wondered whether there is a developmental change in the duration distribution of EPs that parallels the appearance of protosyllables in early plastic song. In all birds, the appearance of a protosyllable peak was consistently associated with the appearance of a peak in the EP duration distribution at a similar duration (Fig. 9*A*). To quantify the presence of narrow peaks in the EP duration distribution, we fit EP distributions to a broad asymmetric Gaussian function that fit extremely well to the EP distribution in subsong birds (Fig. 9*B*; see MATERIALS AND METHODS). In subsong, the residual square error of this fit was low, and as the EP distribution developed peaks the asymmetric Gaussian produced larger squared errors, which we used as a metric of "peakiness" (Fig. 9*C*; $P < 0.001$, *t*-test; Table 1). The location of peaks in the EP duration distribution corresponded to the duration of protosyllables, resulting in distinct clusters along the diagonal in the joint syllable-EP duration distribution (Fig. 9*D*; $n = 8$ of 8 birds), similar to those observed in adult song (Fig. 9*E*).

The presence of these distinct clusters along the diagonal of the EP-syllable joint duration distribution might suggest that each protosyllable is uniquely associated with an EP of the same duration, even at the earliest stages of protosyllable development. However, in contradiction to this hypothesis, we observed that EPs containing protosyllables (PSEPs) exhibited a characteristic and surprising distribution of durations—in all birds recorded at the transition to plastic song

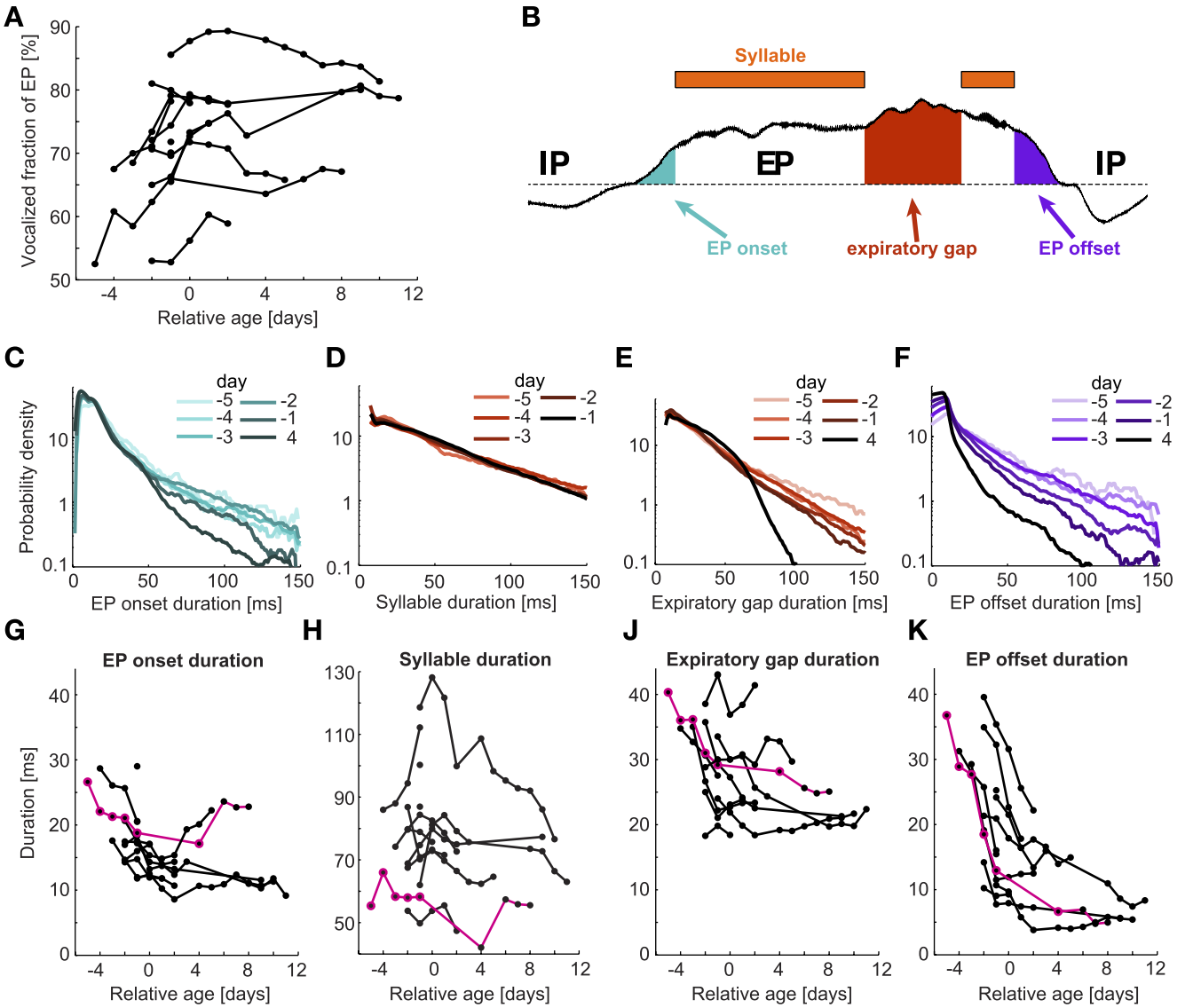


Fig. 7. Developmental progression of respiratory-vocal coordination. *A*: summary plot of developmental progression of the average fraction of each EP that is filled with syllables for all birds. *B*: schematic diagram showing the definition of EP onset, expiratory gap, and EP offset period, for reference to panels below. *C–F*: developmental changes in the distributions of EP onset duration (*C*), syllable duration (*D*), expiratory gap duration (*E*), and EP offset period (*F*) for *bird 9*, 39–43 and 48 dph. The distributions are plotted over 5 sequential days and another 4 days later. Note the developmental shortening of all durations except syllable durations. *G–K*: summary plots of developmental progression of the 4 quantities shown above. *G*: average duration of EP onset during each recorded day, shown for all birds. Points for the bird shown in *C* are highlighted in red. *H*: average syllable duration shows no consistent trend. *J*: average duration of expiratory gaps. *K*: average duration of EP offset period shows the most dramatic developmental change during subsong.

(relative *age 1–4*, $n = 8$), the duration distribution of PSEPs exhibited one or more peaks at longer durations than the primary protosyllable-related peak [Fig. 10, *A–C*; average height of secondary peak relative to primary peak, 30.2 ± 18.1% (±SD)]. In all birds, the spacing between these multiple PSEP peaks appeared to occur at roughly integer multiples of the primary EP peak duration [Fig. 10*D*; correlation of spacing between the multiple peaks and the duration of the primary EP peak, $r^2 = 0.88$, slope = 0.95 ± 0.18 (SE), not significantly different from 1, $P > 0.2$]. Consistent with these measures, an examination of the song spectrograms and pressure waveforms revealed that the longer peaks in the PSEP distribution appeared to be cases in which two or more protosyllables were produced sequentially but an intervening IP failed to occur (Fig. 10*C*).

HVC Lesions Reverse Developmental Changes

It has previously been shown that birds revert to the production of subsong-like vocalizations after bilateral lesions of nucleus HVC (Aronov et al. 2008). In addition, HVC lesions in young plastic birds eliminate both the peak in syllable duration distribution (Aronov et al. 2008) and the peak in the gap duration distribution (Aronov et al., 2010). Because protogaps appear to be associated with IPs and, as shown above, protosyllables appear to be associated with stereotyped EP durations, we wondered what role HVC plays in the development of respiratory patterns in early song.

To address this question, we recorded air sac pressure after bilateral HVC lesions in early plastic song birds that we had

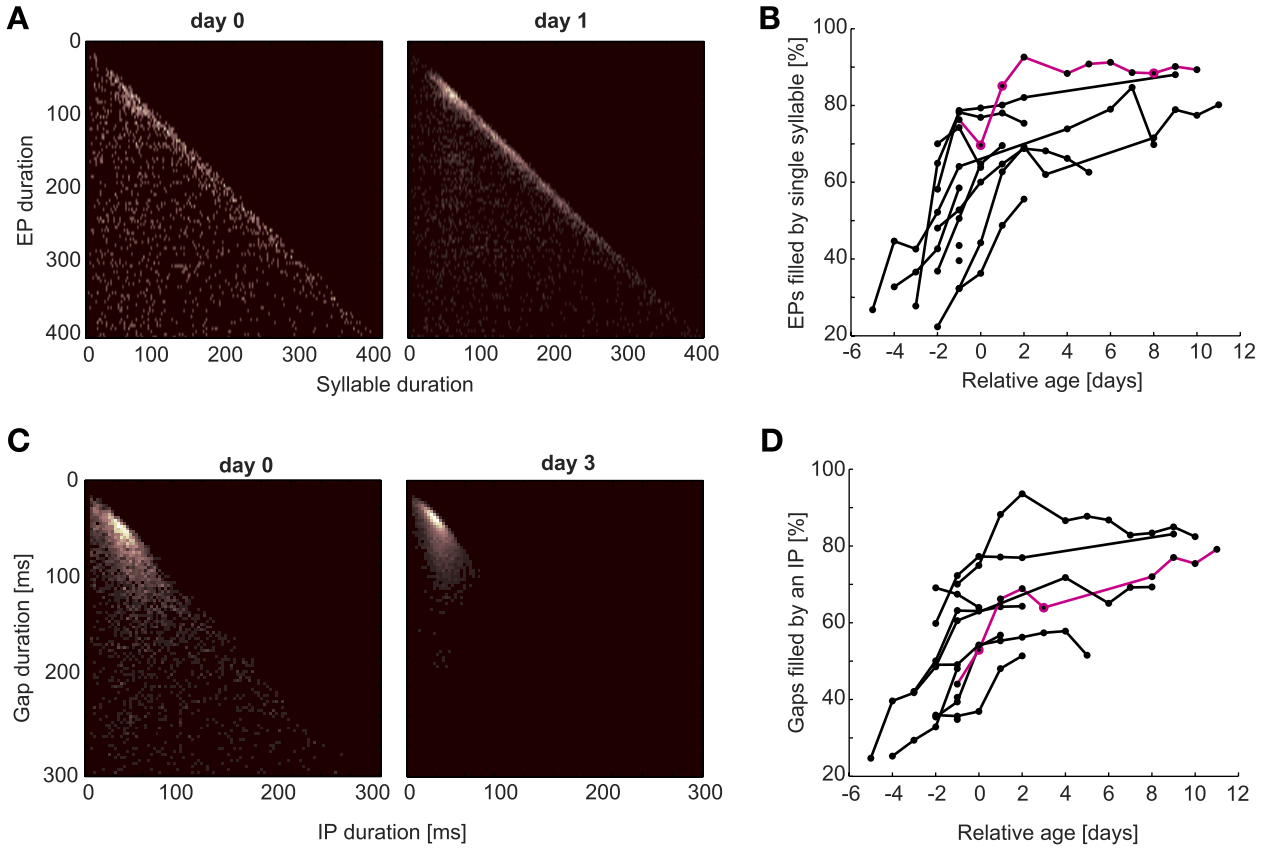


Fig. 8. Development of association between syllables and EPs. *A*: joint distribution of syllable and EP durations in a bird singing subsong (*left*) and 1 day later (*right*) (*bird 10*, 48 and 49 dph). Note the higher density of syllable-EP points along the diagonal, indicating an increased incidence of EPs that are completely filled with a syllable. EP-syllable coordination is quantified as the fraction of EPs that contain a single syllable with a duration within 25 ms of the EP duration. This example shows a change from 69.6% (*left*) to 85.1% (*right*) over 1 day. *B*: summary plot of developmental changes in the fraction of EPs filled with a single syllable for all birds. *C*: joint distribution of gap and IP durations in a bird singing subsong (*left*) and 3 days later (*right*) (*bird 8*, 52 and 55 dph, respectively). Note the accumulation of gaps and IPs near the diagonal, indicating short gaps that are completely filled by an IP. *D*: summary plot for the whole data set of the proportion of gaps that were filled by a single IP with a duration within 25 ms of the gap duration.

tracked developmentally before the lesion ($n = 7$). Surprisingly, strong inspiratory and expiratory pulses persisted after HVC lesions (Fig. 11*A*). Although the extremes of the pressure intensities were reduced (Fig. 11, *B* and *G*), the pressure distribution remained strongly bimodal after lesion (Fig. 11*B*) and even exhibited a significantly more pronounced bimodality than in subsong birds (height of central minimum was 47.0 ± 6.4% of smaller peak, smaller than in subsong birds; $P < 0.05$). This finding suggests that HVC is not necessary for the distinct inspiratory and expiratory phases of respiration during singing.

To further examine the effect of HVC lesions, we quantified EP and IP durations, as well as several measures of respiratory-vocal coordination described above. After HVC lesions, the average length of EPs and IPs was significantly increased, opposite to the observed developmental changes (Fig. 11, *C*, *D*, and *G*; Table 1). Furthermore, the shape of the EP and IP duration distributions also showed an apparent reversal of developmental trends, acquiring distributions more characteristic of early subsong (Fig. 11, *C* and *D*). A similar effect was also seen in the durations of EP onset and offset periods (Fig. 11, *E* and *F*) and of expiratory gaps. For each of these measures, we quantified the developmental changes in the days prior to HVC lesion and then quantified the change between pre- and postlesion song (Fig. 11, *G* and *H*). In all cases, the change between pre- and postlesion song was in the opposite direction to the developmental changes prior to the lesion. Examination of the joint syllable duration and EP duration distribution shows that HVC lesions also abolish the coordination between syllable duration and EP duration, returning the fraction of "filled" EPs to a level consistent with early subsong (Fig. 11, *J* and *K*). HVC lesions also eliminated the peaks in the EP duration distributions (Fig. 11*L*). This, together with the increased duration of IPs and EPs following HVC lesion, suggests that HVC does not just play a developmental role in the coordination of syllables and respiration but also increasingly controls the timing and duration of respiratory events. The effects of HVC lesion and developmental changes from subsong, through early and late plastic song, and in adult song are summarized in Fig. 12.

DISCUSSION

We set out to describe the respiratory patterns in young zebra finches singing early "babbling" (subsong) vocalizations and to examine the developmental time course by which respiratory and vocal patterns become coordinated. We found that respiratory patterns in zebra finches exhibited distinct brief (50 ms) inspiratory pulses, similar to minibreaths (Hartley and Suthers 1989), alternating with longer, more variable pulses of sustained expiratory pressure (50–500 ms). Consistent with earlier descriptions of air sac pressure in juvenile cardinals

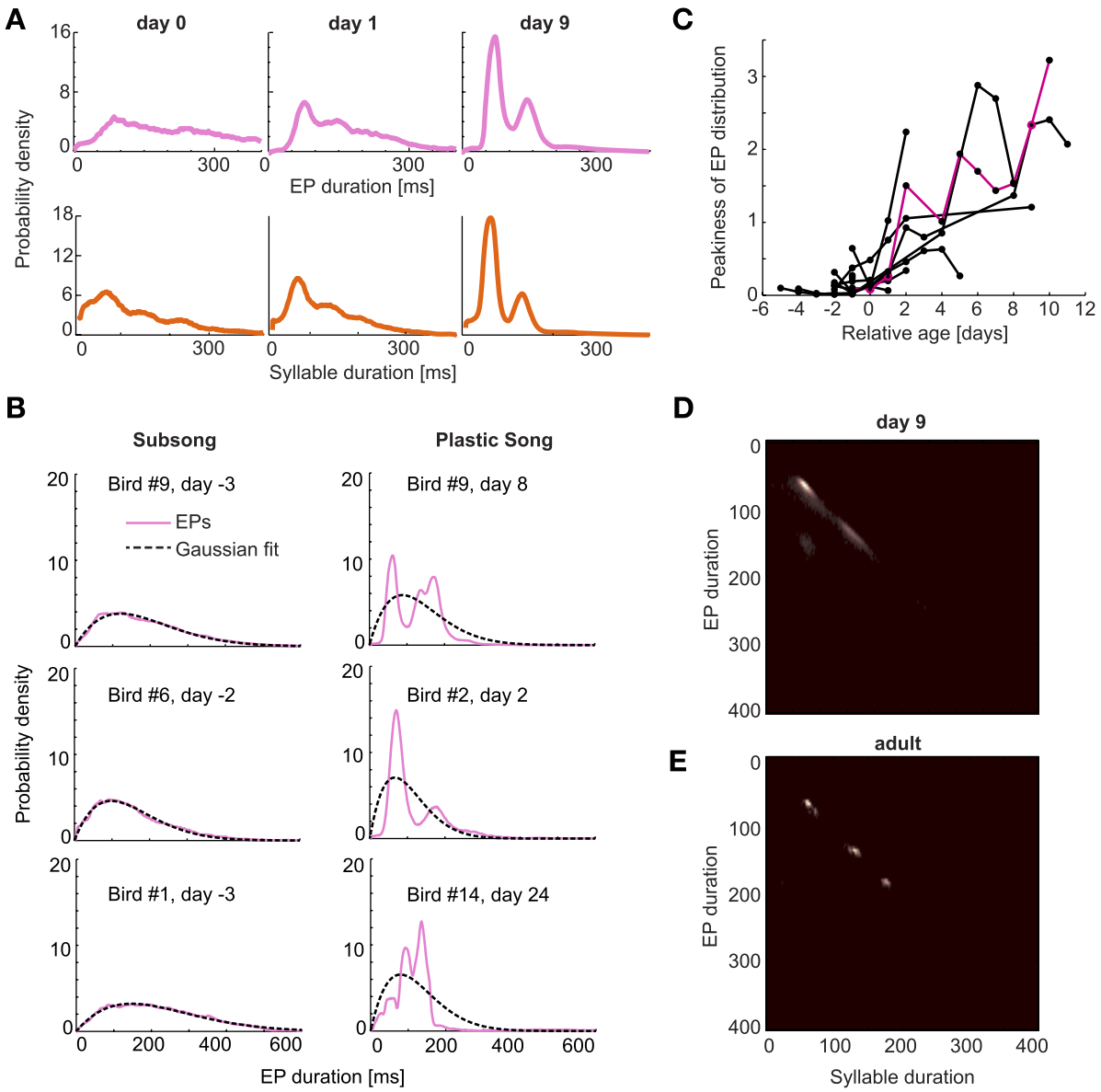


Fig. 9. Development of stereotyped EP durations. *A*: distributions of EP durations (*top*) and syllable durations (*bottom*) in 1 bird on 3 different days (*bird 10*, dph 48, 49, and 57). Note the appearance of 1 and then 2 narrow peaks in the syllable duration distribution and the concurrent appearance of peaks in the EP durations. *B*: the subsong EP duration distributions are well fit by a single broad asymmetric Gaussian function (see MATERIALS AND METHODS), but the narrow peaks in the plastic song EP distribution are not well fit. Thus the "peakiness" of the distribution was quantified as the fitting error of this function to the EP distribution. *C*: summary plot of the developmental progression of peaks in the EP duration distribution, shown for all birds (pink points indicate bird in *A*). On the *y*-axis is plotted the total squared error of the best-fit asymmetric Gaussian function to the EP distribution. Note the sudden deviation from a good fit after the appearance of a protosyllable peak (relative *age 0*). *D*: joint distribution of EP duration and syllable duration for *day 9* of the bird shown in *A* (same bird shown in Fig. 8*A*). Note the appearance of distinct clusters of EP/syllable durations along the diagonal, indicating that the peaks in the EP distribution correspond to the duration of protosyllables, as is typical for an adult bird (*E*, *adult 1*).

(Suthers and Goller 1998; Suthers 1999, 2004), we found that the respiratory and vocal gestures in zebra finch subsong were poorly coordinated. We also found that this coordination exhibited striking developmental changes during the transition from subsong to plastic song.

We have characterized the spectra of vocalizations produced at different respiratory pressures during subsong. Subsong syllables in the zebra finch, characterized by high sound level between 1 and 5 kHz and rich harmonic structure, were produced exclusively during expiratory pulses. The mechanisms underlying these sounds is unclear but may involve the pulse-tone mechanism of the syrinx thought to underlie low-frequency harmonic syllables in the adult zebra finch (Fee 2002; Jensen et al. 2007) or other nonlinear oscillatory processes (Fee et al. 1998; Wilden et al. 1998; Suthers et al. 2011). In contrast, broadband high-frequency inspiratory sounds, with peak spectral intensity between 5 and 15 kHz, were produced during most inspiratory pulses and may be due to turbulent airflow through the trachea and syrinx during inspiration. In some cases, these sounds became more like high-frequency whistles as the bird progressed into plastic song, a process that could underlie the development of inspiratory whistled notes observed in adult zebra finches (Goller and Daley 2001).

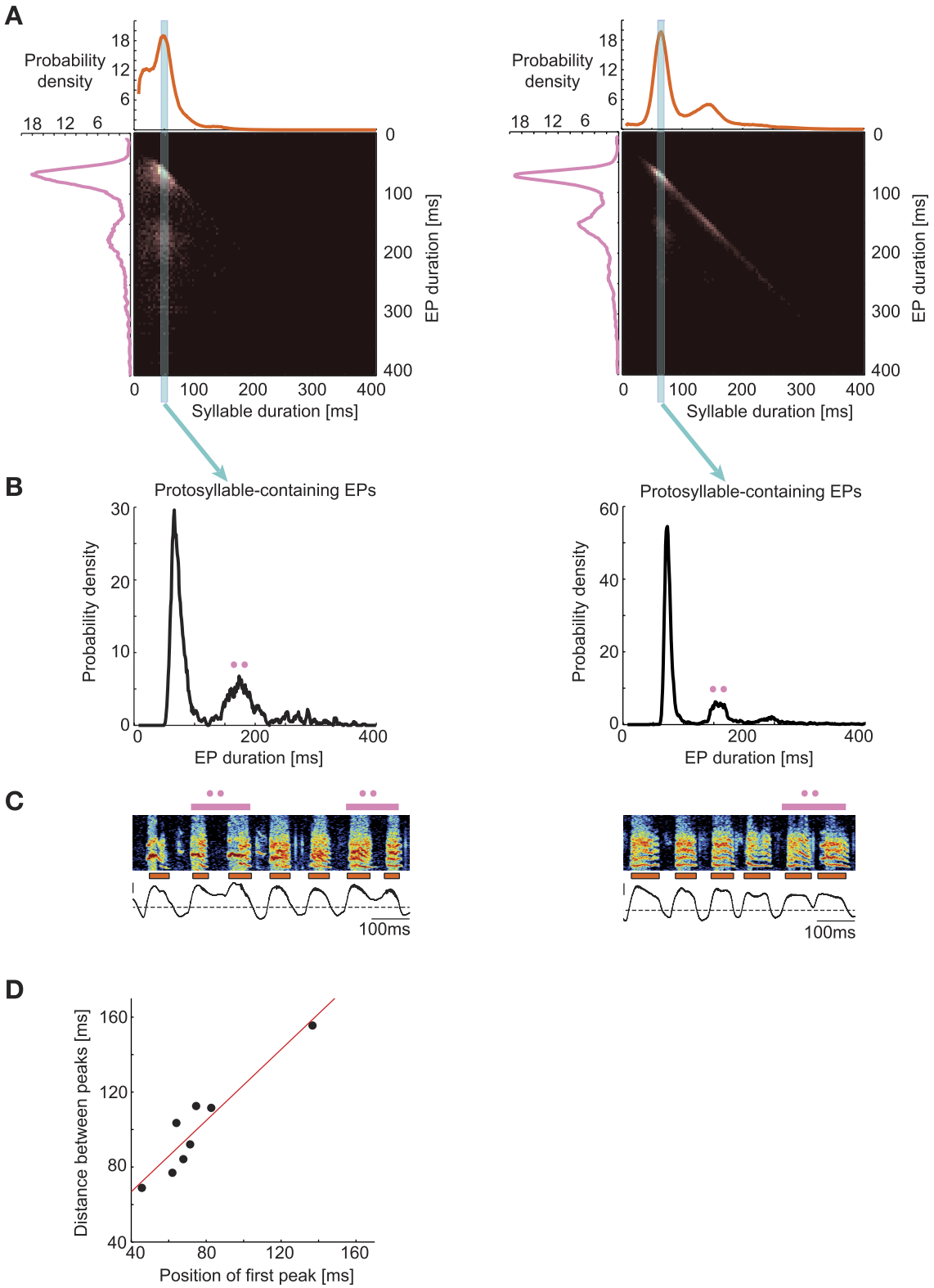


Fig. 10. Duration distribution of EPs containing protosyllables. *A*: joint distribution of EP durations (*y*-axis) and syllable durations (*x*-axis) for 2 birds (*bird 2*, relative *age 2*, *left*; *bird 10*, relative *age 2*, *right*). The protosyllable peak in the syllable duration distribution was identified (vertical blue bar). *B*: duration distribution of EPs containing syllables within 10 ms of the protosyllable peak. These EPs are identified as protosyllable-containing EPs (PSEPs). All birds recorded in early plastic song exhibited 1 or more peaks above the primary protosyllable-duration peak. *C*: spectrograms and air sac pressure traces for the 2 example birds. Most EPs contain 1 protosyllable. However, note the presence of long EPs that contain 2 protosyllables (marked with 2 dots). These EPs have a duration corresponding to the second peak in *B* (also marked with 2 dots). *D*: the spacing between multiple EP peaks (*y*-axis) is strongly correlated with the duration of the principal peak (*x*-axis) ($r^2 = 0.88$). Also shown is a linear fit to the points (red line, slope = 0.95).

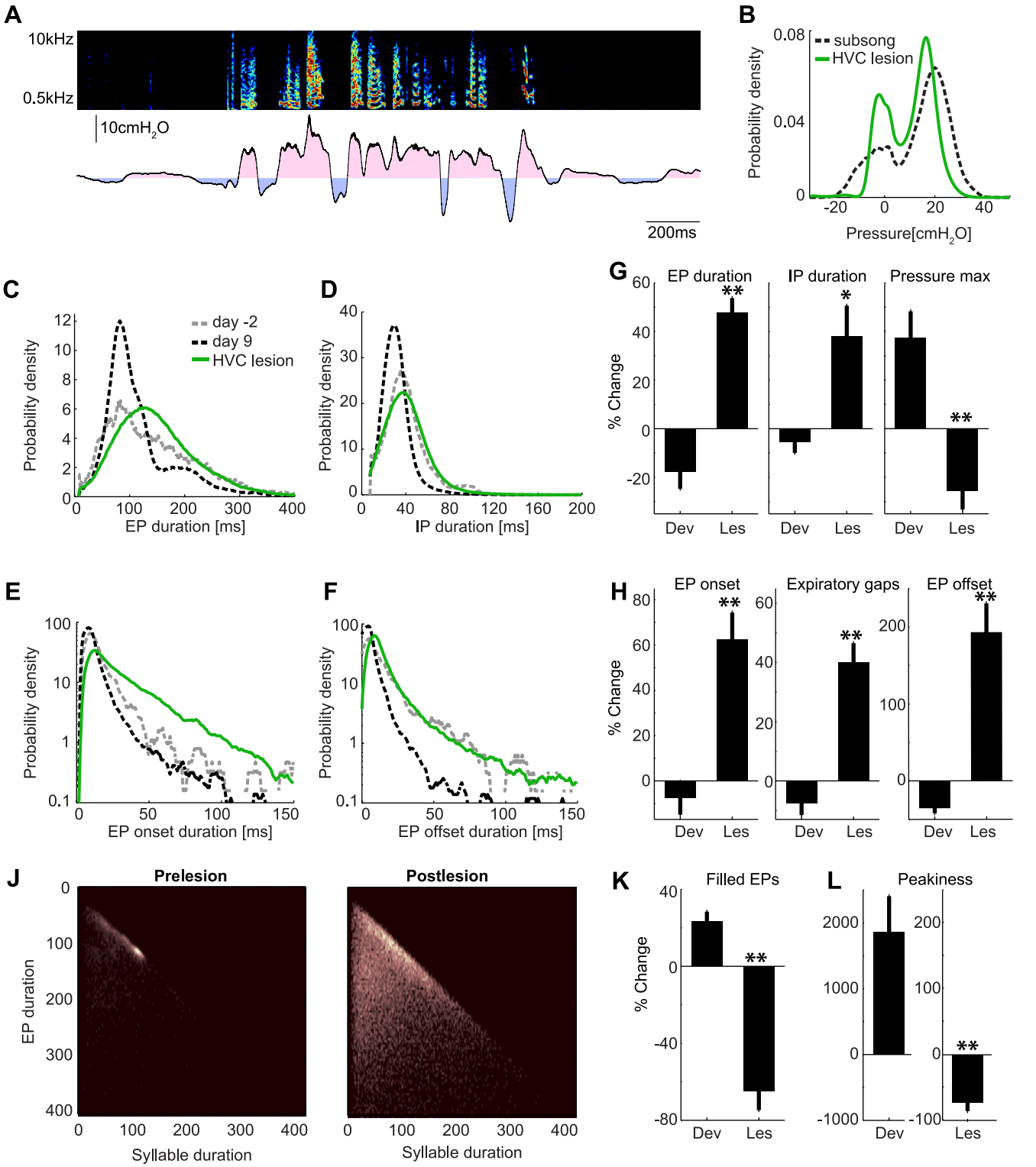


Fig. 11. Effect of HVC lesion on respiratory patterns in subsong. *A*: example of subsong-like vocalizations (*top*) and air sac pressure (*bottom*) in a bird 4 days after HVC lesion (*bird 6*, 55 dph). Note the continued presence of distinct IPs and EPs. *B*: distribution of air sac pressure before (dashed line) and after (green trace) HVC lesion for the bird in *A*. *C–F*: distributions of EP duration (*C*), IP duration (*D*), EP onset duration (*E*), and EP offset duration (*F*) before and after HVC lesion for 1 bird (*bird 7*). Shown are the distributions on 2 different days prior to lesion (gray and black dashed traces, 49 and 60 dph, respectively) and after HVC lesion (green trace). *G* and *H*: population summary of the effect of HVC lesions on respiratory-vocal features. For each feature, the average within-bird developmental trend during subsong is shown (Dev), as is the effect of the lesion (Les, postlesion minus prelesion, normalized by the prelesion value). *J*: joint distribution of EP duration and syllable duration before (*left*) and after (*right*) HVC lesion for the same bird (*bird 7*). Note the loss of EP-syllable durations along the diagonal and the loss of clustering, indicating the loss of EP-syllable coordination and loss of structure in the EP distribution following HVC lesion. *K*: population summary of the fraction of EPs filled by a single syllable during development and after HVC lesion. *L*: population summary of the total squared error (peakiness metric) of EP durations during development and after HVC lesion. All changes shown are statistically significant ($^*P < 0.05$, $^{**}P < 0.01$, *t*-test).

In adult song, nearly every gap contains an IP and every EP is completely filled with a syllable, such that respiratory patterns are tightly coordinated with the acoustic signal (Franz and Goller 2002; Hartley and Suthers 1989). In contrast, in zebra finch subsong, EPs contain multiple syllables and long nonvocalized periods at the onsets and offsets of the EP; thus respiratory patterns are distinct from, and largely uncorrelated with, the acoustic signal, reminiscent of the nonvocalized expiratory periods seen in cardinal subsong (Suthers 2004). This affords a unique opportunity to study the neural origins of IPs and EPs as behavioral events distinct from the onsets and offsets of syllables. For example, modulations in neural activity in the premotor areas LMAN and RA could be separately analyzed with respect to respiratory and syllable patterns to clarify the possible involvement of forebrain circuits in the generation or initiation of subsong respiratory patterns.

We have previously identified the transition from subsong to early plastic song with the appearance of a distinct peak at ~100 ms in the distribution of syllable durations (protosyllables) and the appearance of a peak at ~60 ms in the

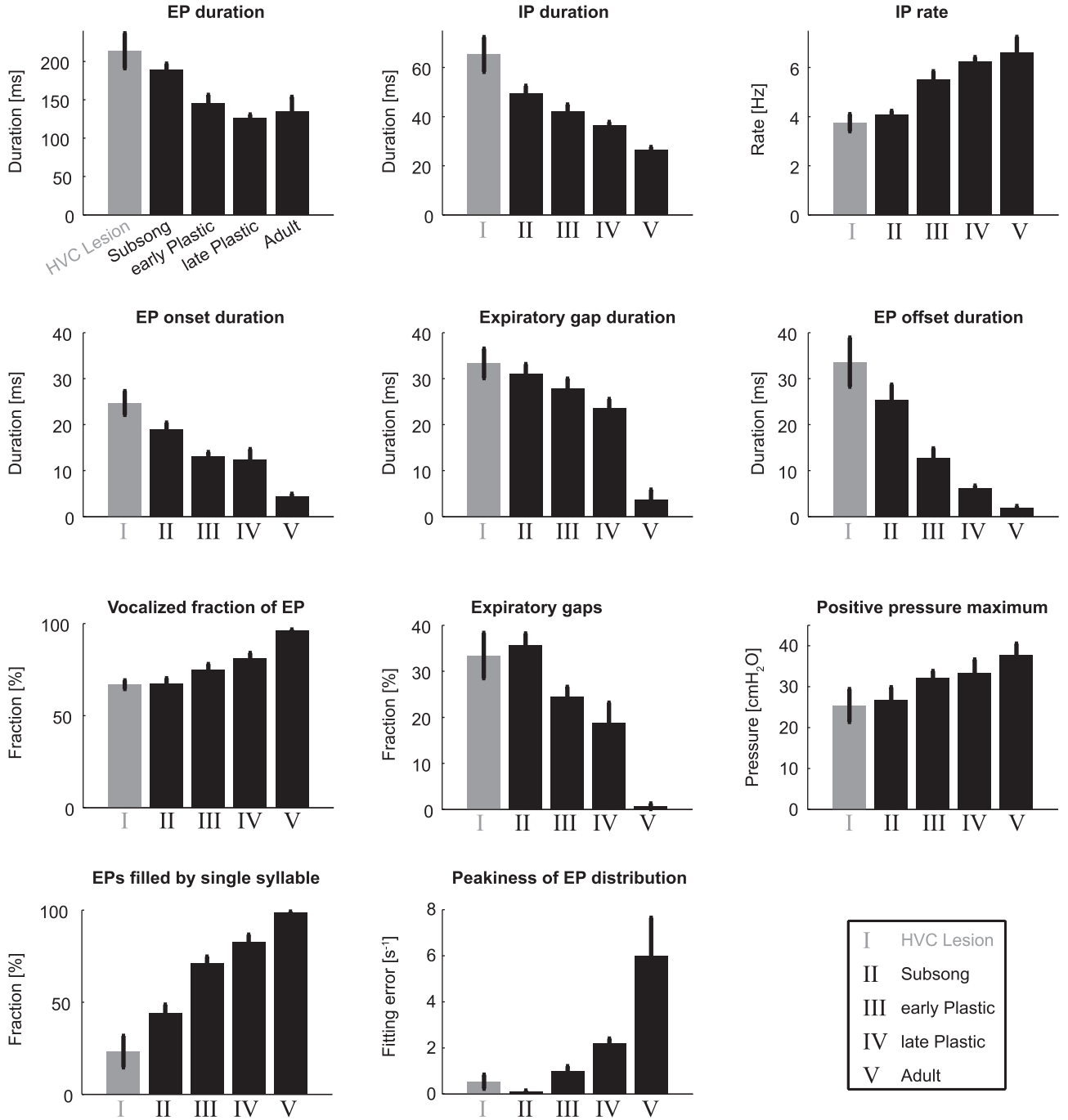


Fig. 12. Summary of developmental trends and effect of HVC lesion. For each respiratory-vocal feature measured, the population average value is shown in 5 conditions: after HVC lesion (I), subsong (II), early plastic song (III), late plastic song (IV), and adult (V). Note that all features show a consistent developmental trend from subsong through plastic song to adult. Note that after HVC lesion most features revert to a value developmentally "earlier" than the relatively late subsong recorded in our experiments. Error bars indicate SE. (See Table 1 for a summary of statistically significant differences.)

distribution of gap durations (protogaps) (Aronov et al., 2010). What is the basis of the appearance and rapid development of these early precisely timed song features? Protogaps may appear when the relatively stereotyped IP becomes closely flanked by syllables. As EP offset and EP onset periods shorten with development, each IP will be associated with a gap of similar duration, producing a peak in the gap distribution that closely matches the IP durations. Thus the dramatic shortening of EP onset and EP offset periods, combined with the relatively stereotyped duration of IPs, are likely key factors in the appearance of stereotyped gap durations.

The shortening of EP onset and EP offset periods is part of a general "filling in" of EPs with vocalization during subsong development, which is apparent in the increasing fraction of each EP that is vocalized (Fig. 7) and the gradual disappearance of expiratory gaps (Fig. 12 and Table 1). The combined effect of these changes during the transition from subsong to plastic song is a dramatic increase in the fraction of EPs that are completely filled by a single syllable (Fig. 8). In fact, the one-to-one relation between syllables and EPs, so characteristic of adult song, is largely accomplished within a few days of the transition from subsong to early plastic song.

What mechanisms underlie the development of protosyllables? The appearance of distinct clusters along the diagonal in the joint distribution of syllable duration and EP duration suggest that two factors may primarily contribute: *1*) the "filling in" of EPs described above, which implies that the duration of each syllable will be similar to that of its EP (i.e., along the diagonal of the joint distribution), and *2*) the appearance of one (or more) peaks in the EP duration distribution, which implies stereotypy in the duration of expiratory pulses. The combination of these two factors will naturally cause the appearance of syllables of stereotyped durations (protosyllables).

In a separate study, we examined the role of HVC in the generation of stereotyped (peaked) structure in syllable and gap distributions (Aronov et al., 2010) and found that HVC lesions completely eliminate protosyllable and protogap peaks. Here we found that HVC lesions cause a substantial reversion of developmental changes in all measures of respiratory-vocal coordination examined. For example, the duration of EP onset and offset periods is greatly increased after HVC lesion, as are the duration and number of expiratory gaps, which could contribute to the loss of the protogap peak after lesion. We also found that the fraction of EPs that are filled by a single syllable is dramatically reduced by HVC lesion, and the sharp peaks in the EP distribution are completely eliminated. Both of these could contribute to the loss of the protosyllable peak after HVC lesion.

Together these findings suggest a critical role for HVC in the development of these earliest consistently timed features in zebra finch song. What specific functions could HVC implement that might explain the developmental changes described above? First, the decreased number and duration of expiratory gaps, as well as the general "filling in" of EPs with vocalization, could be explained if HVC begins to drive patterns of activity in RA that increase the overall expiratory pressure during EPs, thus increasing the fraction of time that the EP remains above the phonation threshold pressure (Riede and Goller 2010; Titze 1992). This hypothesis is consistent with the developmental increase of peak EP pressures (Fig. 6, *G* and *H*) and the loss of peak pressure after HVC lesion (Fig. 11*G*). HVC could also begin to maintain the syrinx in a configuration that has a low phonation threshold pressure, which would tend to eliminate those expiratory gaps that occur at fairly high positive pressures (Fig. 5*A*).

In addition, HVC also appears to be necessary for the developmental shortening (Fig. 11*G*) and increased stereotypy (Fig. 11*L*) of EP durations. The mechanism by which HVC drives or permits these developmental changes in respiration is unknown, but based on our understanding of the function of HVC in adult song, one might speculate that during subsong, circuit dynamics in HVC may generate a burst, or sequence of bursts, that has the duration of one protosyllable (~100 ms) (Long et al. 2010; Solis and Perkel 2005). In the later stages of subsong, this activity in HVC could begin to produce a coordinated activation of stereotyped-duration EPs (100 ms), an IP, and then another EP, enabling the rhythmic repetition of protosyllables and protogaps observed in early plastic song (Saar and Mitra 2008; Tchernichovski et al. 2001; Aronov et al., 2010). We might imagine that these events are probabilistic, i.e., that HVC occasionally fails to initiate either another syllable or an IP. In cases where HVC succeeds to activate another protosyllable but fails to initiate an IP, we would expect to see EPs that contain multiple protosyllables, precisely like those shown in Fig. 10. Of course, these ideas are speculative, and further experiments will be required to determine the precise role of HVC and other brain regions in the development of respiratory-vocal coordination. We note here that while the appearance of a "protosyllable peak" in the syllable duration distribution can be quite sudden, often occurring over the course of 1 day, it is likely that the underlying HVC-driven processes occur much more gradually, throughout subsong and early plastic song.

One important question not addressed in this study is the role of learning and auditory feedback in the development of early respiratory-vocal coordination. Does the appearance of protosyllables and protogaps require the neural circuits important for other aspects of song learning? It is interesting to note that peaks in the syllable duration distribution (protosyllables) can appear in socially isolated birds before the exposure of a bird to a tutor song (Tchernichovski et al. 2004), suggesting that the development of early temporal structure may not require auditory experience of conspecific song. On the other hand, the fact that these features are dependent on HVC, which is principally involved in the generation of learned vocalizations, suggests that auditory feedback may play a crucial role, but this hypothesis could be more directly tested by a combination of experiments, including deafening or lesions of area X, a key learning-related nucleus (Bottjer et al. 1984; Konishi 1965; Scharff and Nottebohm 1991).

Our findings may shed light on some aspects of early vocalizations in humans. Just as in songbirds, the breathing pattern during human speech is fundamentally different from nonspeech respiration, even for earliest babbling in infants (Connaghan et al. 2004), and continues to be refined throughout childhood (Boliek et al. 1996, 1997). Interestingly, respiratory-vocal coordination in children appears to exhibit a pattern different from that in young zebra finches.

In young children, vocalized expiratory pulses (called breath groups) tend to be filled with a single phonation. While zebra finches appear to produce fewer syllables per EP through development, children learn to produce more syllables per breath group (Boliek et al. 2009). This difference between speech development and song development in birds makes sense because adult speech involves the production of many words and silences within a single breath (Levelt 1993), whereas adult birdsong is produced with a tight one-to-one correspondence between breaths and syllables (Franz and Goller 2002). Despite these differences between human speech and birdsong, understanding the neural mechanisms underlying developmental changes in songbirds could shed light on the development of coordination between respiratory and vocal patterns in humans and, more generally, on the interaction of forebrain motor-learning circuits with brain stem motor systems.

ACKNOWLEDGMENTS

We thank Jesse Goldberg and Franz Goller for helpful discussions and for comments on earlier versions of the manuscript.

GRANTS

This study was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC-009183 to M. S. Fee and by a Hertz Foundation Silvio Micali Fellowship to D. Aronov.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* 320: 630–634, 2008.

Aronov D, Veit L, Goldberg JH, Fee MS. Two distinct forms of forebrain dynamics underlie the production of an early motor behavior. *Soc Neurosci Abstr* 411.2, 2010.

Boliek CA, Hixon TJ, Watson PJ, Jones PB. Refinement of speech breathing in healthy 4- to 6-year-old children. *J Speech Lang Hear Res* 52: 990–1007, 2009.

Boliek CA, Hixon TJ, Watson PJ, Morgan WJ. Vocalization and breathing during the first year of life. *J Voice* 10: 1–22, 1996.

Boliek CA, Hixon TJ, Watson PJ, Morgan WJ. Vocalization and breathing during the second and third years of life. *J Voice* 11: 373–390, 1997.

Bottjer SW, Miesner EA, Arnold AP. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* 224: 901–903, 1984.

Brackenbury J. Respiration and production of sounds by birds. *Biol Rev* 55: 363–378, 1980.

Connaghan KP, Moore CA, Higashakawa M. Respiratory kinematics during vocalization and nonspeech respiration in children from 9 to 48 months. *J Speech Lang Hear Res* 47: 70–84, 2004.

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567–631, 1999.

Fee MS. Measurement of the linear and nonlinear mechanical properties of the oscine syrinx: implications for function. *J Comp Physiol A* 188: 829–839, 2002.

Fee MS, Scharff C. The songbird as a model for the generation and learning of complex sequential behaviors. *ILAR J* 51: 362–377, 2010.

Fee MS, Shraiman B, Pesaran B, Mitra PP. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* 395: 67–71, 1998.

Franz M, Goller F. Respiratory units of motor production and song imitation in the zebra finch. *J Neurobiol* 51: 129–141, 2002.

Goller F, Daley MA. Novel motor gestures for phonation during inspiration enhance the acoustic complexity of birdsong. *Proc Biol Sci* 268: 2301–2305, 2001.

Hage SR, Stefan MB. Neuronal networks involved in the generation of vocalization. In: *Handbook of Mammalian Vocalization—An Integrative Neuroscience Approach,* edited by Brudzynski SM. Oxford: Academic, 2009, p. 339–349.

Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419: 65–70, 2002.

Hartley RS. Expiratory muscle activity during song production in the canary. *Respir Physiol* 81: 177–187, 1990.

Hartley RS, Suthers RA. Airflow and pressure during canary song: direct evidence for mini-breaths. *J Comp Physiol A* 165: 15–26, 1989.

Immelmann K. Song development in the zebra finch and other estrildid finches. In: *Bird Vocalizations*, edited by Hinde RA. New York: Cambridge Univ. Press, 1969, p. 61–74.

Jensen KK, Cooper BG, Larsen ON, Goller F. Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proc Biol Sci* 274: 2703–2710, 2007.

Konishi M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z Tierpsychol* 22: 770–783, 1965.

Kubke MF, Yazaki-Sugiyama Y, Mooney R, Wild JM. Physiology of neuronal subtypes in the respiratory-vocal integration nucleus retroambigualis of the male zebra finch. *J Neurophysiol* 94: 2379–2390, 2005.

Levelt WJM. *Speaking: from Intention to Articulation*. Cambridge, MA: MIT Press, 1993.

Lilliefors HW. On the Kolmogorov-Smirnov Test for the exponential distribution with mean unknown. *J Am Stat Assoc*. 64: 387–389, 1969.

Liu W, Gardner TJ, Nottebohm F. Juvenile zebra finches can use multiple strategies to learn the same song. *Proc Natl Acad Sci USA* 101: 18177–18182, 2004.

Long MA, Fee MS. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456: 189–194, 2008.

Long MA, Jin DZ, Fee MS. Support for a synaptic chain model of neuronal sequence generation. *Nature* 468: 394–399, 2010.

Marler P. Song-learning behavior: the interface with neuroethology. *Trends Neurosci* 14: 199–206, 1991.

Marler P, Peters S. Selective vocal learning in a sparrow. *Science* 198: 519–521, 1977.

Marler P, Peters S. Structural changes in song ontogeny in the swamp sparrow *Melospiza georgiana*. *Auk* 99: 446–458, 1982.

Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, *Serinus canarius*. *J Comp Neurol* 165: 457–486, 1976.

Price PH. Developmental determinants of structure in zebra finch song. *J Comp Physiol Psychol* 93: 260–277, 1979.

Reinke H, Wild JM. Identification and connections of inspiratory premotor neurons in songbirds and budgerigar. *J Comp Neurol* 391: 147–163, 1998.

Riede T, Goller F. Peripheral mechanisms for vocal production in birds—differences and similarities to human speech and singing. *Brain Lang* 115: 69–80, 2010.

Saar S, Mitra PP. A technique for characterizing the development of rhythms in bird song. *PLoS One* 3: e1461, 2008.

Scharff C, Nottebohm F. 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, 1991.

Schmidt MF. Pattern of interhemispheric synchronization in HVc during singing correlates with key transitions in the song pattern. *J Neurophysiol* 90: 3931–3949, 2003.

Solis MM, Perkel DJ. Rhythmic activity in a forebrain vocal control nucleus in vitro. *J Neurosci* 25: 2811–2822, 2005.

Suthers RA. Peripheral mechanisms for singing: motor strategies for vocal diversity. In: *Proceedings of the 22nd International Ornithological Congress, Durban*, edited by Adams NJ, Slotow RH. Johannesburg: Bird Life South Africa, 1999, p. 491–508.

Suthers RA. How birds sing and why it matters. In: *Nature's Music—The Science of Birdsong*, edited by Marler P, Slabbekoorn H. San Diego, CA: Academic, 2004, p. 272–295.

Suthers RA, Goller F. Respiratory-syringeal motor coordination during song learning in Northern cardinals. *Fifth Intl Cong Neuroethol* Abstract No. 298, 1998.

Suthers RA, Wild JM, Kaplan G. Mechanisms of song production in the Australian magpie. *J Comp Physiol A* 197: 45–59, 2011.

Tchernichovski O, Lints T, Deregnaucourt S, Cimenser A, Mitra PP. Studying the song development process: rationale and methods. *Ann NY Acad Sci* 1016: 348–363, 2004.

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

Titze IR. Phonation threshold pressure, a missing link in glottal aerodynamics. *J Acoust Soc Am* 91: 2926–2935, 1992.

Vu ET, Mazurek ME, Kuo YC. Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci* 14: 6924–6934, 1994.

Wild JM. The avian nucleus retroambigualis: a nucleus for breathing, singing and calling. *Brain Res* 606: 319–324, 1993a.

Wild JM. Descending projections of the songbird nucleus robustus archistriatalis. *J Comp Neurol* 338: 225–241, 1993b.

Wild JM. Neural pathways for the control of birdsong production. *J Neurobiol* 33: 653–670, 1997.

Wild JM, Goller F, Suthers RA. Inspiratory muscle activity during bird song. *J Neurobiol* 36: 441–453, 1998.

Wilden I, Herzel H, Peters G, Tembrock G. Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics* 8: 1–30, 1998.

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