Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds

Dmitriy Aronov, **Lena Veit**, **Jesse H. Goldberg**, and **Michale S. Fee**[†]
McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Accurate timing is a critical aspect of motor control, yet the temporal structure of many mature behaviors emerges during learning from highly variable exploratory actions. How does a developing brain acquire the precise control of timing in behavioral sequences? To investigate the development of timing, we analyzed the songs of young juvenile zebra finches. These highly variable vocalizations, akin to human babbling, gradually develop into temporally-stereotyped adult songs. We find that the durations of syllables and silences in juvenile singing are formed by a mixture of two distinct modes of timing – a random mode producing broadly-distributed durations early in development, and a stereotyped mode underlying the gradual emergence of stereotyped durations. Using lesions, inactivations, and localized brain cooling we investigated the roles of neural dynamics within two premotor cortical areas in the production of these temporal modes. We find that LMAN (lateral magnocellular nucleus of the nidopallium) is required specifically for the generation of the random mode of timing, and that mild cooling of LMAN causes an increase in the durations produced by this mode. On the contrary, HVC (used as a proper name) is required specifically for producing the stereotyped mode of timing, and its cooling causes a slowing of all stereotyped components. These results show that two neural pathways contribute to the timing of juvenile songs, and suggest an interesting organization in the forebrain, whereby different brain areas are specialized for the production of distinct forms of neural dynamics.

Introduction

Many vertebrate behaviors, such as object manipulation, social displays, and vocalizations, are characterized precisely-timed, stereotyped motor sequences (Lashley, 1951; Levelt, 1993). Often, these sequences are not innate, but are acquired by juvenile animals via exploratory learning, during which the timing of behaviors can be highly variable. Little is known about the neural circuit dynamics that underlie variably timed juvenile behaviors or those that underlie the developmental emergence of stereotyped timing in adult behaviors.

Birdsong is an excellent model of motor timing and its development. Zebra finches, like most songbirds, learn their songs by imitating an adult (Immelmann, 1969; Marler, 1970). At an early learning stage, they produce subsong – highly variable vocalizations similar to human babbling (Marler, 1970; Oller et al., 1976; Doupe and Kuhl, 1999). Gradually, during the "plastic song" stage, vocalizations become more stereotyped until, in adulthood, singing consists of a highly precise sequence of syllables and silent gaps (Tchernichovski et al.,

---

2001; Glaze and Troyer, 2006). We use the term "early song" to refer to the earliest developmental phase, encompassing subsong and early plastic song.

Previous studies have elucidated the neural mechanisms of timing in adult songs. A key "cortical" area for mature singing is HVC (used as a proper name)(Nottebohm et al., 1976; Simpson and Vicario, 1990; Yu and Margoliash, 1996; Hahnloser et al., 2002). HVC projects to RA (robust nucleus of the arcopallium), which in turn projects to brainstem vocal and respiratory nuclei (Wild, 2004). Recent work demonstrated that mild cooling of HVC results in uniform slowing of all song components, whereas cooling of RA has no effect on song timing (Long and Fee, 2008). Because temperature affects the speed of all cellular processes, this result suggests that the neural dynamics generated by intrinsic circuitry within HVC control song timing in adulthood (Long and Fee, 2008; Fee and Long, 2011).

In contrast to adult song, subsong does not require HVC activity; in fact, following bilateral HVC elimination, birds of all ages produce subsong-like vocalizations (Aronov et al., 2008). Rather, a premotor area for subsong production is the lateral magnocellular nucleus of the nidopallium (LMAN)(Aronov et al., 2008). LMAN is the "cortical" output stage of an avian basal-ganglia-thalamocortical loop (Perkel et al., 2002; Farries et al., 2005) which, like HVC, also projects to RA (Bottjer et al., 1984). During later developmental stages, LMAN continues to contribute to song production by introducing variability to singing (Kao et al., 2005; Ölveczky et al., 2005) – a necessary feature for exploration in trial-and-error learning (Doya and Sejnowski, 1995; Tumer and Brainard, 2007; Andalman and Fee, 2009).

Where are the neural dynamics that generate the timing of early vocalizations? Although progress has been made in understanding the development of timing later in development (Tchernichovski et al., 2001, 2004), little is known about the earliest song timing. Thus, we first characterize the temporal features of early singing on a detailed quantitative level. We then use various circuit manipulations of premotor brain areas to understand their contributions to early song timing.

Materials and Methods

Subjects were juvenile male zebra finches between 33 and 59 days-post-hatch (dph). Birds were obtained from the Massachusetts Institute of Technology breeding facility. Animal care and experiments were carried out in accordance with the National Institute of Health guidelines and approved by the local Institutional Animal Care and Use Committee.

Sound recordings

Juvenile birds were raised in the MIT breeding facility, in single cages with their parents and siblings. At age 33–42 dph they were removed and placed in custom-made sound isolation chambers. Vocalizations were recorded with custom-written Matlab software or with Sound Analysis Pro (Tchernichovski et al., 2000), which was configured to ensure that recordings were triggered on all quiet vocalizations of young birds (Aronov et al., 2008).

Song analysis

All data analyses were performed with custom Matlab software. 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, Matlab functions fir1 and filtfilt), squaring, and smoothing with a 2.5 ms sliding square window. Relative sound level was converted to units of decibels (dB) by computing the logarithm (base 10) and multiplying by 10. The result is referred to in the text and figures as song or syllable "amplitude." Sound amplitude

produced during singing was bimodally distributed, corresponding to syllables and gaps. Means and standard deviations of these two modes were estimated by fitting a Gaussian mixture (two Gaussians) to the sound level distribution using expectation maximization (EM).

To segment sound into syllables and silent gaps, we defined an "upper threshold" as the Fisher discriminant of the two identified Gaussian modes and a "lower threshold" as 2 standard deviations above the mean of all values below the upper threshold. Candidate syllable onsets were defined as upward crossings of the lower threshold followed by upward crossings of the upper thresholds. Similarly, candidate syllable offsets were defined as consecutive downward crossings of the upper and then lower 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. All recordings were manually examined to eliminate calls and cage noise from the data set.

Syllables were initially analyzed by fitting an exponential function to their duration distribution using 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 duration 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. This MLE fitting algorithm was also used to fit exponential functions to various components of gap duration distributions, described in the Results section.

Syllable duration distributions were generally not exponential below 25 ms, possibly due to peripheral mechanical constraints on the production of brief sounds. Thus, 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 one day of singing (~1,000–10,000 syllables). The goodness-of-fit of the exponential was estimated using 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 less than 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.

In plastic song, syllable duration distributions exhibited peaks at durations of 40–200 ms. To quantify these peaks, we first fit an exponential to the long tail of the syllable duration distribution (200–400 ms). This exponential was then subtracted from the distribution, calculated in 1 ms bins. The peak in the residual was typically well-fit by a Gaussian; thus the best-fit Gaussian was used to estimate the amplitude and location of the protosyllable peak. In 8 of 40 plastic-song birds, the single-Gaussian fit was relatively poor ($r^2 < 0.9$) because two peaks were present in the distributions. For these birds, we fit a mixture of two Gaussians to the data instead. The total peak size was then determined as the sum of the two individual amplitudes, and both center values were included in the analysis of syllable durations.

Air sac pressure recordings

Air sac pressure measurements during subsong production were performed using methods similar to those previously described for adult birds (Franz and Goller, 2002; Goller and Cooper, 2004). A device was constructed using a piezoresistive pressure transducer chip (Fujikura FPM-02PG) together with an instrumentation amplifier (Burr Brown INA122).

Prior to surgery, birds were anesthetized with 1–2% isofluorane in oxygen. A small hole was made in the posterior thoracic air sac and a silastic tube (o.d. 2 mm, length 5 mm) was inserted and sutured to the rib cage. The other end of the tube was attached to the inlet tube of the pressure transducer. Another silastic tube (o.d. 0.9 mm, length 16 mm) containing a thin wire was secured subcutaneously on the bird's back, and the device was tied to this tube. Birds began to sing within a few days after surgery, at which point the pressure sensor was connected by a thin cable to a custom mercury commutator. Sound and air sac pressure were digitized at 40 kHz and recorded simultaneously using custom software in Matlab.

We asked whether singing was influenced by the implantation of air sac pressure recording devices. Between the population of subsong birds that received device implants (N=12) and birds that were not implanted (N=25) there were no significant differences in any of the parameters that we used to quantify duration distributions of silent gaps (p_1, τ_1, p_2, τ_3, μ, σ and γ; see Results); there was also no difference in the time constants of exponential fits to syllable duration distributions (p>0.06 for all 8 parameters). These measurements suggest that device implants for pressure recordings do not significantly affect the timing of syllable and gap durations.

Lesions and inactivations

Lesions and details of the inactivation methodology, as well as the histological and electrophysiological confirmations of these techniques are described elsewhere (Aronov et al., 2008). Briefly, birds were anesthetized with 1–2% isofluorane and placed into a stereotaxic apparatus. Electrolytic HVC lesions were made by passing current (100 μA current for 60 s) at 10–13 locations, spaced at 500 μm. Inactivations of HVC or LMAN were made with muscimol or tetrodotoxin, delivered using custom-built untethered reverse microdialysis probes.

Cooling

We used two custom-built cooling devices – one for cooling HVC (a superficial brain area) and another for cooling LMAN (a nucleus ~2 mm deep in the brain). These devices were modified from those described in (Long and Fee, 2008) to eliminate water cooling, making them less restrictive and thus more suitable for use in young juvenile birds. Design, construction and testing of these devices are described elsewhere (Aronov and Fee, 2011). Briefly, both devices consisted of two bismuth telluride thermoelectric elements (Custom Thermoelectric) connected in series between a 'hot plate' and a 'cold plate.' Running electric current across these elements in one direction resulted in cooling of the cold plate; current in the opposite direction resulted in warming of the cold plate.

For HVC cooling, two gold-plated annealed silver pads (1×2 mm) were soldered to the cold plate. During surgery, an area of the cranium above HVC was thinned using a dental drill, and the pads were placed in contact with the thinned areas. For LMAN cooling, thermally-conductive "probes" (gold-plated annealed silver, 250 μm diameter wire) were soldered to the cold plate. These probes were thermally insulated with polyimide tubing (1 mil thick) and a ~35 μm air gap between the wire and the tubing. The bottom 1 mm of the probes were left uninsulated. During surgery, small craniotomies and duratomies were made above

LMAN, and the probes were slowly inserted into LMAN, such that the un-insulated segments spanned the dorso-ventral extent of the nucleus.

To remove waste heat generated by current flow through the device (Long and Fee, 2008; Aronov and Fee, 2011), a convective copper-mesh heat sink was attached to the hot plate of the device. Additionally, a "body-coupled" heat sink (silver sheets) was attached to the hot plate; the other end of this heat sink was brought in close proximity to the surface of the cranium to further stabilize the temperature of the device. Details of heat sink construction and heat removal are described in (Aronov and Fee, 2011).

We directly measured temperature changes produced by these devices in 3 birds with the HVC cooling device and in 4 birds with the LMAN cooling device (see (Aronov and Fee, 2011) for details). Briefly, small thermocouples (40-gauge) were either attached to the cold plate of the device or implanted into the brain at various distances from the cooling surface. Following surgery, birds were placed into a foam restraint and allowed to recover from anesthesia until fully alert (~10 min). Electric current was then applied, alternating between 0.5 and −1.5 A using a programmable current source (Lambda ZUP14-60). Following a switch in currents, temperature typically reached a steady-state value within about 10 s. At 0.5 A, the device was maintained close to normal body temperature (41°C). At −1.5 A, the center of HVC (500 μm deep) was cooled by about 7°C below normal body temperature. The amount of cooling in LMAN strongly depended on distance from the probe; for experimental birds, the average temperature change achieved was therefore estimated from a histological analysis of the implanted probe location (see below).

For cooling experiments, devices were implanted and birds were allowed to recover after surgery until the onset of singing (typically within 1 week). Birds were then tethered with light, flexible cables to custom mercury commutators. Electric current (0.5 and −1.5 A, alternating every 100 s) was delivered to the devices continuously throughout the day using the programmable current supply. Only songs produced during the last 80 s of each current pulse were used for analysis, such that only data at steady-state temperatures were included.

Histological analysis

After all experiments, animals were sacrificed and perfused with 4% paraformaldehyde. Brains were extracted and sliced parasagittally for histological examination of the tracks created by probes (for inactivation and cooling experiments) or the extents of lesions. HVC lesions were confirmed by the absence of retrogradely-labeled RA-projecting cells (see (Aronov et al., 2008) for details).

Temperature in the brain is strongly dependent on the distance from the cooling or heating probe (Jafari and Higgins, 1989; Long and Fee, 2008). Consequently, the size of LMAN or positioning of probes in it can have a substantial effect on the average amount of cooling achieved in this nucleus and may vary across birds. To approximate the amount of cooling in each bird, we estimated the location of the cooling probes relative to LMAN boundaries.

To estimate the average amount of cooling in LMAN, we simulated the 3-dimensional spatial profile of temperature around the probes in each bird using the standard model of thermal diffusion in cylindrical coordinates ((Jafari and Higgins, 1989); also see (Aronov and Fee, 2011) for details):

$$\Delta T(r) = \Delta T_P \frac{K_0(r/\lambda)}{K_0(r_P/\lambda)}$$

where *r* is the radius (distance from the central axis of the cooling probe), $\Delta T(r)$ is the steady-state temperature change (with respect to normal brain temperature) at that radius, ΔT_p is the steady-state temperature at the surface of the cooling probe (also with respect to normal brain temperature), K_0 is the modified zero-order Bessel function of the second kind, and λ is the space constant of the brain (estimated from the data in (Long and Fee, 2008) to be approximately 1.59 mm). We estimated ΔT_p using thermal modeling and calibrations (Aronov and Fee, 2011) to be approximately $-12°C$. For calculating the amount of temporal stretch due to cooling in $\% \ °C^{-1}$, we obtained a single-value estimate of cooling by averaging the estimated amounts across the two hemispheres.

Results

Behavioral timescales of subsong

Adult zebra finch songs contain distinct syllable types of highly stereotyped durations (Figure 1a,b). Whereas the timing of syllables and silent gaps in adult song is well described (Glaze and Troyer, 2006), less is known about the temporal structure of early song. We therefore sought to characterize the temporal features of subsong and to identify the underlying timescales of this behavior. We recorded the songs of 68 birds between the ages of 33 and 59 dph (see Methods). Figure 1c illustrates a typical song bout of a 40 days-post-hatch (dph) juvenile bird. As described in detail elsewhere (Veit et al., 2011), in birds of all ages, even the youngest recorded (<40 dph), the distribution of song amplitudes exhibited a bimodal shape, with two peaks corresponding to vocalization and silence (Figure 1d). Thus, even early song appeared to consist of two distinct acoustic states, which we refer to as syllables (continuous segments of sound) and gaps (continuous silent intervals, Figure 1c). We separately analyzed the distributions of syllable and gap durations.

In the youngest of the recorded birds, syllable durations were broadly distributed (Figure 1e). Short syllables (<50 ms) occurred with the highest probability density, and long syllables – some as long as 400–800 ms – occurred with the lowest probability density. Above ~25 ms, the distributions were well-fit by an exponential function, as illustrated by their linearity on a semi-logarithmic plot (Figures 1e). Most birds under the age of 40 dph exhibited exponentially-distributed syllable durations (16 of 24 birds, see Methods for the goodness-of-fit statistic). This fraction of birds decreased with age over a brief developmental period, such that none of the 15 birds older than 50 dph exhibited exponentially-distributed syllable durations (Figure 1f,g). For the remainder of this paper, we define "subsong" as the production of syllables whose durations are well fit by an exponential distribution. In older birds, syllable durations deviated from an exponential distribution due to the appearance of consistently-timed syllables during the early plastic song phase (Tchernichovski et al., 2004; Liu et al., 2004; Aronov et al., 2008); these syllables formed peaks in duration distributions (Figure 1f) and will be addressed later.

An exponential distribution of durations is remarkable because it is indicative of a highly random process (Poisson process), in which the probability of terminating a syllable is uniform in time, and is thus independent of when the syllable was initiated. The time constant of this exponential is a single timescale that characterizes the syllable durations of an individual bird. Across the population of subsong birds, the time constant was 89.5±4.5 ms (N=37 birds, Figure 1h,i).

We next sought to characterize the timing of silent gaps in subsong. Because subsong appears to be a series of randomly-timed transitions between vocal and silent periods, we expected gap durations to be distributed similarly to those of syllables. Surprisingly, however, gap duration distributions in subsong birds did not show a monotonic decrease at greater durations, but typically exhibited a more complex structure, often including a

secondary peak near 60 ms (Figure 2a–d). Thus, even in young subsong birds with no stereotypy in syllable durations, gaps often appeared to exhibit some consistency in timing.

Even outside the peak at ~60 ms, the distributions of gap durations were more complex than the single-exponential shape of syllable duration distributions. Specifically, distributions of short gaps (<~30 ms) appeared to be approximately exponential (time constant of 27.9±2.4 ms, N=34 subsong birds), as did the distributions of gaps between 100 and 200 ms (time constant 54.9±1.2 ms). However, these time constants were different (p<0.001, paired t-test, Figure 2c), indicating a possible presence of two mechanistically distinct modes in gap durations. Furthermore, although gap distributions were well-described by an exponential function at long durations (100–200 ms), they strongly deviated from this exponential fit for gaps longer than ~300 ms (Figure 2c). These long gaps approximately followed a power-law distribution (Figure 2d), indicating the possible presence of yet another mode in gap timing.

Respiratory patterns underlying behavioral timescales in early song

What accounts for this complexity in gap timing? In particular, what causes gaps to be relatively precisely timed at ~60 ms durations, even in young birds with no equivalent precision in syllable timing? Because inspirations occur primarily during silent intervals (Franz and Goller, 2002), we hypothesized that the distinct modes of gap timing are related to different respiratory patterns during the gaps. To characterize these patterns, we recorded air sac pressure in 12 young juvenile birds (ages 39–53) during subsong production (Figure 2e, see Methods). Silent periods between bouts typically contained a periodic pressure fluctuation at 2–3 Hz corresponding to eupneic breathing. Subsong, like adult song, was accompanied by an increase in the amplitude of both expirations and inspirations. Syllables were almost always produced during periods of positive pressure (in 98.4±0.7% syllables the majority of time points were expiratory, N=12 birds). Gaps, however, were less homogenous in their respiratory patterns. On the basis of these patterns, we were able to identify roughly four distinct "modes" of gap production, described below.

Many of the gaps were expiratory ("mode-1" gaps, 33.2±2.3% of all gaps, N=12 birds, Figure 2f) – that is, air sac pressure stayed positive at all times between the two flanking syllables during these gaps. Although some of these gaps were up to ~100 ms long (e.g., bottom example in Figure 2f), most had a short duration (mean duration 31.0±1.9 ms). Conversely, most short gaps were expiratory (91.9±2,2% of all gaps <25 ms). Expiratory gaps can be thought of as non-vocalized periods that break a single expiratory gesture into multiple syllables. In fact, individual expiratory pressure pulses in subsong often contained more than one syllable, with some pulses containing as many as 8 syllables (mean 1.50±0.06 syllables, Figure 2e). The presence of so many expiratory gaps is unique to juvenile song; in adult zebra finches nearly every gap contains an inspiration (Franz and Goller, 2002).

The remaining gaps contained at least one inspiratory pressure pulse (IP), although for gaps that were longer than ~300 ms, this typically involved a return to eupneic breathing; these long gaps will be discussed below. Gaps that were shorter than 300 ms and contained at least one IP comprised 52.6±2.1% of all gaps (N=12 birds, Figure 2g,h). In these gaps, periods of negative air sac pressure were brief (average IP duration 49.4±2.3 ms) and appeared to be distinct events with rapid onsets and offsets. Many gaps containing an IP also included positive pressure regions. These silent expiratory periods were on average 40.7±2.9 ms long and could occur prior to or after the IP (e.g., examples in Figure 2h). In order to describe these events, we loosely refer to those gaps that contain a single IP immediately preceded and followed by syllables as "mode-2" gaps (Figure 2g); we refer to gaps that contain prolonged positive pressure regions in addition to at least one IP as "mode-3" gaps (Figure 2h). Mode-2 and mode-3 gaps form a continuum, depending on the duration of the positive-pressure region, and, aside from illustrating them in Figure 2, we did not classify individual

gaps as belonging to one of these modes. The reasons for using two modes to describe these gaps will be further clarified when we develop a quantitative description of gap duration distributions.

Finally, 14.2±1.3% of all gaps were longer than 300 ms and generally contained eupneic breathing ("mode-4" gaps, N=12 birds, Figure 2i). The shortest of these consisted of what appeared to be a single period of the breathing cycle. However, most mode-4 gaps were long periods of silence between subsong bouts (median duration 2.06±0.28 s, ±bootstrap standard error of the median, but often tens of seconds long). Interestingly, long gaps followed a power-law duration distribution with a long tail extending to ∼30 s (power-law exponent γ= −1.37±0.03 for gaps 300 ms – 30 s long, Figure 2d).

Quantitative description of gap duration distributions

Our recordings suggest a model for the structure of the complex gap duration distribution in early singing (Figure 3a): (1) The initial sharply-decreasing segment of the overall gap distribution at short durations is produced by mode-1 gaps. These are produced during EPs, as are syllables, and exhibit a roughly exponential duration distribution (Figure 3b,c). In support of this hypothesis, we found that the time constant of the short-duration gap duration distribution across all subsong birds (27.9±2.4 ms, N=34 birds) was consistent with the average time constant of expiratory gaps measured from birds in which air sac pressure was recorded (31.0±1.9 ms, N=12 birds, p=0.48, Figure 3d). (2) The peak in the gap duration distribution is produced by mode-2 gaps containing a consistently-timed IP tightly coupled to the preceding syllable offset and the following syllable onset. This peak was roughly Gaussian in shape (Figure 3e,f) and was centered at 55.5±1.4 ms (see Methods), similar to the average IP duration (49.4±2.3 ms, Figure 3f,g). The small but significant difference between the means of these durations (p<0.05, t-test), is likely due to the fact that mode-2 gaps also contain a brief transition period between syllable offset and IP onset, as well as between IP offset and syllable onset. (3) The long tail in the overall gap duration distribution is produced by gaps that contain an IP surrounded by an extended period of non-vocalized positive pressure (mode-3 gaps). The duration of these gaps is formed by the sum of an IP duration and the durations of the positive pressure periods before and after the IP, the sum of which exhibited a roughly exponential distribution with a time constant of 48.3±3.3 ms (measured in the tail between 50–150 ms, Figure 3h). This time constant was, in fact, not significantly different from the time constant of an exponential fit to the gap distribution (54.9±2.0 ms for gaps 100–200 ms long, p=0.10, Figure 3i,j). (4) Finally, gaps >300 ms are primarily generated by transitions to eupneic breathing (mode-4 gaps). Due to the presence of these gaps, the overall gap distribution deviated from an exponential at long durations.

Understanding how gaps durations are formed by the underlying respiratory patterns allowed us to quantify several aspects of respiratory timing from songs alone (Figure 4). Because eupneic-breathing gaps tended to be long (>300 ms), we first concentrated on gaps shorter than 300 ms. We assumed that some fraction p_1 of these were mode-1 gaps. Since the durations of mode-1 gaps were approximately exponentially-distributed, we wanted to estimate both p_1 and the time constant τ_1 of this exponential. Because almost all gaps shorter than 30–40 ms were expiratory mode-1 gaps (see above), the overall gap distribution at these short durations was also well-fit by an exponential and provided a good estimate of the exponential time constant of mode-1 gaps. At longer durations, the gap duration distribution tended to increase, forming a peak at ∼60 ms. We identified the duration t_{min} at which distribution began to increase by finding the minimum of the smoothed distribution (1 ms bins, 20-ms square-window smoothing). (Because on some recording days in younger birds a minimum was not present, we excluded those days when determining t_{min}). An exponential function was then fit to all gaps shorter than $0.75t_{min}$ using an MLE procedure

(see Methods; Figure 4, Step 1). This fit provided a value of τ_1 (time constant of the exponential) and p_1 (area under the exponential curve).

The remaining fraction (1-p_1) of all gaps shorter than 300 ms were assumed to be mode-2 or mode-3 gaps. To estimate their distribution, we subtracted the mode-1 exponential fit described above from the overall gap distribution (calculated in 1 ms bins). The residual distributions tended to exhibit a peak near 60 ms and were strongly asymmetric, with a long exponential tail >100 ms (Figure 4, Step 2). Because mode-2 and mode-3 gaps form a continuum, we did not attempt to classify individual gaps as belonging to one of these modes. Rather, we assumed that some fraction p_2 of the residual distribution was formed by mode-2, while the remaining 1-p_2 was formed by mode-3. Mode-2 gaps closely matched the durations of IPs, the durations of which exhibited approximately Gaussian distributions. We therefore described the duration distribution of mode-2 gaps as a Gaussian of mean μ and standard deviation σ (Figure 3e,f). Mode-3 gaps contained an IP plus additional expiratory components, the durations of which were approximately exponentially-distributed. Thus, we modeled the duration distribution of mode-3 gaps as a sum of a Gaussian-distributed variable and an exponential-distributed variable, forming an ex-Gaussian distribution (Hohle, 1965). We made the further assumption that the Gaussian component of mode-3 gaps has the same center and width as mode-2 gaps. To describe the combined distribution of mode-2 and mode-3 gaps we therefore needed four parameters: p_2, mean μ and standard deviation σ of the Gaussian, and time constant τ_3 of the exponential.

The time constant τ_3 could be estimated from the long tail of the overall gap distribution (Figure 4, Step 3). The distribution was typically uncontaminated by the Gaussian at durations >100 ms. However, it deviated from the exponential at durations >~300 ms due to the presence of mode-4 gaps. We thus estimated τ_3 by fitting an exponential to the distribution between 100–200 ms. The remaining parameters (p_2, μ, and σ) were estimated using nonlinear optimization to minimize the mean squared difference of the fit from the distribution.

It is not obvious, *a priori*, that mode-2 and mode-3 gaps should be modeled as two distinct modes of the distribution. Can the distribution of gap durations be adequately explained with only the ex-Gaussian model used for mode-3 gaps? By fitting a general Gaussian/ex-Gaussian mixture to the data, we found that the extracted mode-2 (Gaussian) component is often a sizable fraction of the mixture, and that this component is necessary to accurately describe the peak of the distributions (Figure 4, Step 4). Furthermore, as will be shown later the amplitude of this mode-2 component increases with respect to the mode-3 component during development, and this relative increase is abolished by HVC lesions. In summary, a combination of four temporal modes accurately describes the complex distributions of gaps durations in early song (Figure 4, Step 5). Because these temporal modes arise from distinct respiratory patterns, this model allowed us to extract various features of respiratory timing from singing alone. The goal of such extraction is to allow one to quantify the effects of various brain manipulations (described later) on respiratory timing, even in the absence of simultaneous air sac pressure measurements.

Early development of stereotyped timing

We next sought to understand how the temporal structure of singing changes during the period of early development. Figure 5a shows the songs of a typical bird recorded during the period of transition from subsong to early plastic song. Although the spectral changes of songs during this period were relatively subtle, the temporal changes were pronounced. At ~45 dph, birds began to produce syllables with consistently-timed durations, which resulted in the appearance of peaks in syllable duration distributions (Figure 1f, 5b,c). Adult and late plastic songs in zebra finches usually exhibit multiple peaks in syllable duration

distributions, corresponding to distinct syllable types (Tchernichovski et al., 2001, 2004). In the early plastic song phase we analyzed, however, distributions commonly exhibited only one peak (32 of 40 birds had no detectable second peak, see Methods). Such single peaks diversify into multiple distinct syllable types later in development (Tchernichovski et al., 2004; Liu et al., 2004). We therefore refer to the syllables of this earliest stereotyped duration as "protosyllables." Protosyllables became more common over time, resulting in the growth of the peak in syllable duration distributions. The duration of protosyllables varied greatly across birds (50–174 ms, mean 93.8±32.0 ms, ±s.d., N=48 syllable duration peaks in 40 birds, Figure 5d).

As discussed above, the peak in gap duration distributions corresponded primarily to mode-2 gaps (Gaussian component of the distribution); by analogy we refer to these consistently-timed gaps as "protogaps." Just like protosyllables, protogaps also became more common during development (Figures 4b, 5b,c). Specifically, this was observed as an increase in the contribution of mode-2 gaps (Gaussian component of the duration distribution, parameter p_2 described above) relative to the contribution of mode-3 gaps (ex-Gaussian component, 1-p_2) in the quantitative model described above. In contrast to the peaks in syllable duration distributions, the protogap peaks (mode-2) in gap duration distributions were at shorter durations and more narrowly distributed across birds (mean 53.3±8.0 ms, ±s.d., N=59 birds, Figure 5d).

Protosyllables and protogaps appeared to be strung together into long, rhythmic sequences. For example, the bird in Figure 5e exhibited a peak in gap durations at ~60 ms and a peak in syllable durations at ~150 ms. Syllables with durations close to this peak occurred almost exclusively before and after the ~60 ms protogaps; they were rarely present before and after gaps <25 ms or >100 ms (Figure 5f). To analyze this sequencing, we identified protosyllables and protogaps as those syllables and gaps whose durations were within 20 ms of the corresponding peaks in the distributions. Across 23 plastic-song birds, 35.3±1.8% of syllables were identified as protosyllables. Incidences of protosyllables before and after protogaps were higher than this value (43.5±2.2% and 40.6±2.0% respectively; p<0.05 and p<0.01). Similarly, the incidence of protogaps before and after protosyllables (54.2±2.4% and 51.3±2.9% respectively) were also higher than expected from the overall fraction of protogaps (44.0±2.0%; p<0.05 for both comparisons). These numbers indicate that song transitions in which a protosyllable follows a protogap are 16% more likely than expected from random ordering of syllables and gaps; transitions from a protosyllable to a protogap are 24% more likely. Consistently with these transition probabilities, the incidence of protosyllables following other protosyllables was slightly higher than expected by chance (40.0±2.1% vs. 35.3±1.8%, p<0.001).

Involvement of HVC and LMAN in early song timing

Our results indicate that subsong involves an interplay of processes that produce randomly-timed durations of syllables and gaps with those that produce consistently-timed syllables and gaps. Which brain areas are involved in the timing and development of these processes? It has previously been reported that subsong persists in the absence of HVC, and that subsong syllables are essentially unchanged by bilateral HVC lesions. HVC lesions also abolish the production of distinct stereotyped syllables in plastic song and adult song (Aronov et al., 2008). In contrast LMAN lesions or inactivation produce an increase in song stereotypy in plastic song and adult birds (Kao et al., 2005; Ölveczky et al., 2005). Here we examine the different roles of HVC and LMAN in the control of the temporal structure of early song vocalizations (Figure 6, 7 and 8).

To characterize the role of HVC in early song, we quantified the distributions of syllable and gap durations in young birds before and after bilateral elimination of HVC (see Methods).

As previously reported (Aronov et al., 2008), HVC elimination did not change syllable duration distributions in subsong birds (N=8 birds, Figure 6a,b). However, it abolished the peaks corresponding to protosyllables in plastic song (N=11 birds, peak size 10.9 ± 2.6 s^{-1} across birds, 1.3 ± 0.2 s^{-1} in the same birds following HVC elimination, p<0.001, Figure 6d,e, population results are all summarized in Figure 8a).

Similarly, the peak in gap duration distributions corresponding to protogaps was largely abolished in both subsong and plastic-song birds by HVC elimination (Figure 6c,f). Analysis of these distributions (Figure 6g,h) showed that HVC elimination specifically abolished the production of mode-2 gaps, as determined from the fitting parameter p_2 in the quantitative model described above (peak size 2.8 ± 0.7 s^{-1} across birds, 0.35 ± 0.05 s^{-1} after HVC elimination, p<0.01, Figure 8b). This result indicates that HVC is necessary for the production of consistently-timed protogaps whose duration is tightly coupled to the onsets and offsets of IPs. (Respiratory mechanisms of HVC's involvement in protogap production are detailed in (Veit et al., 2011)).

In contrast to protogaps, short expiratory (mode-1) gaps in subsong birds were largely unaffected by HVC elimination (Figure 6c,f). The peak probability density for short gaps was 13.6 ± 1.0 s^{-1} across all birds and 13.4 ± 1.0 s^{-1} in the same birds following HVC elimination (N=8 bird, p=0.84, Figure 8c), indicating that expiratory gaps may be produced by an HVC-independent mechanism distinct from that which produces inspiratory gaps.

Our results so far indicate that HVC is necessary for the production of consistently-timed protosyllables and protogaps, but not for the production of the exponentially-distributed subsong syllables and mode-1 gaps. Because LMAN plays a major role in early song production (Aronov et al., 2008; Ölveczky et al., 2005), we hypothesized that LMAN activity is necessary for the generation of these song components. To test this possibility, we inactivated LMAN in birds producing early plastic song (N=9 birds, ages 45–56 dph, see Methods). As previously shown (Ölveczky et al., 2005), plastic songs during LMAN inactivation were strikingly less variable that those prior to inactivation (Figure 7a). LMAN-inactivated birds produced distinct syllable types, which formed one or more peaks in the distributions of syllable durations (Figure 7b). These peaks were the dominant feature of syllable timing during LMAN inactivation (peak size 8.8 ± 2.2 s^{-1} across birds, 25.9 ± 5.3 s^{-1} in the same birds following LMAN inactivation, p<0.01, Figure 8d).

The mode-2 peak in gap duration distributions also increased during LMAN inactivation (peak size 9.3 ± 3.2 s^{-1} across birds, 28.0 ± 6.8 s^{-1} in the same birds following LMAN inactivation, Figures 7c, 8e), while mode-1 gaps were largely eliminated (peak probability density 15.0 ± 2.6 s^{-1} across birds, 1.48 ± 0.49 s^{-1} following LMAN inactivation, Figure 8f). Thus, the effect of LMAN inactivation is essentially opposite of HVC elimination, which left mode-1 gaps intact but abolished the peaks in gap duration distributions corresponding to protogaps.

In summary, Figure 8 quantifies across birds the opposing effects of HVC and LMAN elimination on syllable and gap production. These results suggest that different temporal components of syllables and gaps may be produced by distinct neural mechanisms, one of which is HVC-dependent, while the other is LMAN-dependent.

Role of HVC and LMAN dynamics in the control of song timing

The HVC and LMAN elimination experiments show that these brain areas are necessary for the production of stereotyped and randomly-timed durations in early song, respectively. However, they do not tell us how the timing of these song components is controlled. One possibility is that the biophysical dynamics within HVC control the durations of

protosyllables and protogaps, just as the timing of syllables and gaps in adult birds is controlled by HVC (Long and Fee, 2008). An alternative hypothesis is that the timing of these early stereotyped song features is controlled elsewhere in the song system, perhaps by brain areas upstream of HVC such as nucleus Uvaeformis (Nottebohm et al., 1982; Ashmore et al., 2008). To distinguish between these alternatives, we used localized cooling of HVC to slow its intrinsic biophysical dynamics (Long and Fee, 2008; Fee and Long, 2011). We developed a new lightweight Peltier device applicable for temperature manipulations in young juvenile birds (see Methods, Figure 9). HVC was cooled bilaterally by an average of ~7°C (Figure 9a), and the songs produced during cooling were compared to those produced when HVC was maintained at normal body temperature. We isolated various components of gap and syllable duration distributions and quantified their fractional change ("stretch") divided by the average amount of temperature change in HVC (% °C^{-1}). A negative stretch indicates a lengthening of song features at lower temperatures.

In subsong birds, HVC cooling had no effect on the durations of syllables (Figure 10a, 0.06±0.08% °C^{-1}, N=5 birds, p=0.45, paired t-test for all cooling comparisons), consistent with the finding that HVC is not necessary for the production of normal subsong syllables. There was also no effect on the long exponential tail of syllable durations in young plastic-song birds (N=8 birds, p=0.84). In contrast, HVC cooling in plastic-song birds caused the peak in syllable duration distributions (protosyllables) to be shifted toward longer durations (Figure 10b, −1.36±0.16% °C^{-1}, p<0.001). Thus, HVC dynamics are involved in controlling the durations of protosyllables in plastic song, whereas they have no influence on the broadly distributed durations of syllables at any developmental stage.

Like protosyllables, protogaps also became longer with HVC cooling (Figure 10c); the peak in gap durations changed by −1.44±0.11% °C^{-1} (N=8 birds, p<0.001). However, the durations of mode-1 (expiratory) gaps did not change with HVC cooling (0.18±0.30% °C^{-1}, p=0.58). There was also no effect of HVC cooling on the time constant of long (100–200 ms) gaps, which quantifies the expiratory periods in mode-3 gaps (p=0.90). These data suggest that biophysical dynamics within HVC are involved in controlling the duration of protogaps early in development, but not the randomly-timed durations of mode-1 or mode-3 gaps.

Our results so far indicate that HVC dynamics are involved in the timing of both protogaps and protosyllables – the stereotyped components of early singing – but appear not to be involved in the control of randomly-timed subsong syllable and gap durations. What controls the duration of these song components? We hypothesized that biophysical dynamics within LMAN may be involved in the timing of these latter song components. To test this possibility, we cooled LMAN bilaterally in subsong and plastic-song birds by ~6–8°C (see Methods, Figure 9b–d) and analyzed the effects of cooling on gap and syllable durations.

LMAN cooling indeed prolonged the durations of subsong syllables (Figure 10d). Although distributions of syllable durations remained exponential (5 of 5 birds, Lilliefors statistic, see Methods), the time constant of this exponential became longer during LMAN cooling (−1.86±0.52% °C^{-1}, N=5 birds, p<0.05). This suggests that dynamics in LMAN are actively involved in the timing of subsong syllables; i.e., determining when a syllable is terminated relative to its onset. In plastic-song birds, LMAN cooling had no effect on the duration of protosyllables (Figure 10e, −0.60±0.32% °C^{-1}, N=6 birds, p=0.12). However, the long tails of the syllable distributions (longer than the protosyllable) became longer when LMAN was cooled (p<0.05, data not shown).

The effects of LMAN cooling on gaps were similar: cooling did not change the durations of protogaps (Figure 10f, −0.005±0.178% °C^{-1}, N=6 birds, p=0.98). However, mode-1 gaps

became longer with LMAN cooling (change in slope of exponential distribution of $-4.13\pm0.60\%$ $°C^{-1}$, p<0.002). Similarly, LMAN cooling stretched long (100–200 ms) gaps, indicating that LMAN is involved in timing of positive-pressure periods of mode-3 gaps as well (significant decrease in slope of exponential distribution, p<0.02, data not shown).

In summary, the biophysical dynamics intrinsic to HVC and LMAN both play a role in the timing of subsong and early plastic song, but these roles are distinct (Figure 10g–j). LMAN dynamics are involved in controlling the durations of subsong syllables and the positive-pressure components of gaps (mode-1 and mode-3). The timing of these events appears to be independent of HVC. On the other hand, HVC dynamics play a role in the timing of protogaps and protosyllables in early song. Durations of these HVC-controlled events appear to be independent of LMAN.

Discussion

Early singing is characterized by two distinct modes of timing

Characterizing the temporal structure of a behavior is necessary to understand how the brain learns and generates the timing of behavioral sequences (Lipkind and Tchernichovski, 2011). We found that syllable durations in young birds are exponentially-distributed, indicating an extreme level of randomness in the syllable generation process. Over a brief developmental period, peaks formed by "protosyllables" emerge and gradually become the dominant components of syllable duration distributions (see also (Veit et al., 2011)) – a phenomenon that represents the earliest consistent syllable timing. Importantly, this process is not a simple developmental reduction in the variability of syllable durations. Rather, it suggests the presence of two mechanistically distinct processes: one that produces random, exponentially-distributed durations and another that produces repeatable durations from one rendition of a syllable to another.

Silent gaps in early singing also exhibited durations formed by a mixture of distinct temporal modes, including exponentially-distributed expiratory gaps and consistently-timed (~60 ms) "protogaps" that contain inspiratory pressure pulses (IPs). Such features are consistent with previous reports of respiratory patterns in juvenile cardinals (Suthers, 2004). Protogaps became increasingly common during development, perhaps arising from an increasing coordination between vocalizations and respiration (Veit et al., 2011) that eventually leads to the precise coordination of these processes in adult song (Hartley and Suthers, 1989; Franz and Goller, 2002).

LMAN and HVC contribute differentially to the modes of song timing

By manipulating two premotor nuclei – LMAN and HVC – we characterized their roles in producing random and stereotyped modes of early song timing. Eliminating LMAN activity specifically disrupted the production of randomly-timed components of syllable and gap duration distributions, but left intact the consistently-timed protosyllables and protogaps. Furthermore, mild cooling of LMAN prolonged randomly-timed song components (subsong syllables and expiratory components of gaps), but had no effect on protosyllable and protogap durations. These results suggest that LMAN is involved in producing randomly-timed syllables and gaps, and that the biophysical dynamics intrinsic to LMAN circuitry are actively involved in controlling the durations of these song elements.

In contrast to LMAN manipulation experiments, we found that neither lesions nor cooling of HVC affect exponentially-distributed syllable durations in subsong, but that HVC cooling increases the durations of protosyllables and protogaps, while HVC elimination entirely abolishes their production. These results are consistent with experiments in adult birds showing that adult motifs, including syllables and gaps, are uniformly slowed by HVC

cooling (Long and Fee, 2008). Because sequences of protosyllables and protogaps developmentally transform into adult song motifs (Tchernichovski et al., 2001; Tchernichovski and Mitra, 2002), these results suggest that the biophysical processes within HVC controlling the timing of early song elements may be related to those controlling the timing of adult songs.

Our results do not eliminate the possibility that other brain areas are also involved in generating the dynamics that control early song timing. LMAN is reciprocally interconnected with other brain areas as part of a basal ganglia-forebrain loop (Farries and Perkel, 2002; Person et al., 2008; Gale and Perkel, 2010). Specifically, LMAN receives input from the medial part of the dorsolateral thalamus (DLM), which in turn receives inputs from a basal ganglia nucleus Area X (Okuhata and Saito, 1987; Bottjer et al., 1989; Luo and Perkel, 1999). Recent work has shown that lesions of DLM largely eliminate vocal variability in young birds, while lesions of Area X have little effect on variability (Goldberg and Fee, 2011), suggesting that DLM plays a key role in generating subsong vocalizations. Yet, the slowing of subsong timescales by LMAN cooling suggests that the circuit dynamics underlying subsong timing are not entirely upstream of LMAN (e.g., in DLM). One possibility is that DLM is not involved in timing syllables and gaps, but is necessary to tonically "activate" LMAN during singing. Alternatively, it remains possible that dynamics are produced by neural activity that quickly cycles between multiple brain areas. For example, LMAN forms a loop with RA and DLM (Vates et al., 1997). Thus, cooling LMAN could slow down subsong syllables by increasing the time for signals to travel around this loop (Fee and Long, 2011).

Similar arguments can be made regarding the role of HVC in controlling the timing of protosyllables and gaps. HVC receives feedback connections from RA via a pathway through the midbrain and thalamus (Schmidt et al., 2004; Ashmore et al., 2008). Thus, while it seems likely that biophysical dynamics within HVC are involved in the control of protosyllable and protogap timing, we cannot rule out the possible involvement of other brain regions afferent to HVC. More detailed lesion and cooling experiments within other areas of the song system are necessary to address these possibilities.

A separate issue is that our analysis specifically concerns the temporal patterning of syllables and gaps in early songs. Some finer aspects of timing in the subsyllabic acoustic or respiratory structure may be influenced by areas downstream of HVC and LMAN (Andalman et al., in press) or even by peripheral mechanisms of song production (Fee et al., 1998; Goller and Cooper, 2004; Riede and Goller, 2010).

How might the LMAN- and HVC-dependent modes we identified interact during early song development? We found that protosyllables and protogaps tend to follow one another, perhaps accounting for the previously described early appearance of rhythmicity in singing (Saar and Mitra, 2008). However, this tendency is only modestly stronger than expected by chance, suggesting that components with stereotyped and random durations are largely intermingled during singing. Consistent with this idea, single-neuron recordings in LMAN suggest that this area is active throughout subsong (Aronov et al., 2008), plastic song (Ölveczky et al., 2005), and even adult song (Kao and Brainard, 2006). A more detailed understanding of the specific roles of HVC and LMAN in generating stereotyped and randomly-timed components will require an analysis of neuronal activity in these areas during subsong and the transition to plastic song.

Brain areas are specialized for producing distinct modes of neural dynamics

Our results suggest that two forms of circuit dynamics are produced by distinct regions of the avian forebrain, allowing the production of both stereotyped and randomly-timed song

components. How might the circuits within HVC and LMAN produce these distinct forms of neural dynamics? In adult birds, neurons in HVC produce a brief burst of spikes at one moment in the song, and, as a population these neurons generate a sparse sequence of bursts (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Long et al., 2010) that is thought to arise from a synaptically-connected chain of neuronal groups (Abeles, 1991; Vogels et al., 2005; Li and Greenside, 2006; Jin et al., 2007; Long et al., 2010). Such a process within HVC might produce sequential dynamics during protosyllables and possibly protogaps early in development. Theoretical work suggests that a synaptic-chain organization could arise early in development from simple spike timing-based synaptic learning rules, even before sensorimotor learning (Jun and Jin, 2007; Fiete et al., 2010).

It is less clear how LMAN-dependent dynamics might generate exponentially-distributed syllable and gap durations. One possibility is that LMAN (possibly in conjunction with other brain areas) exhibits two types of population activity states – some that generate premotor activity for vocal production and others that generate silences. By continuously evolving in time, population activity could "wander" between vocalizing and non-vocalizing states, producing the observed transitions between syllables and gaps. Another possibility is that distinct populations of neurons within LMAN actively trigger switching between vocal and silent states. This idea is consistent with a recent finding that many LMAN neurons exhibit a burst of activity prior to the onsets or offsets of subsong syllables (Aronov et al., 2008).

A further question is what type of functional network architecture could generate random timing within LMAN. Some theoretical studies suggest that stochastic activity in neural circuits can be produced by chaotic dynamics (van Vreeswijk and Sompolinsky, 1998; Vogels et al., 2005; Wilson, 1999). Such dynamics may be generated by recurrent networks with sparse, random connectivity, or even by biophysical processes in single neurons. Whether these models are applicable to the LMAN circuit, or possibly other systems underlying random exploratory behavior in the brain, is an open question.

Our analysis of HVC- and LMAN-dependent dynamics suggests an interesting organizational scheme in the song system, and possibly in other motor or sensory systems. The anatomical segregation of HVC-dependent stereotyped dynamics and LMAN-dependent random dynamics raises the possibility that distinct brain regions are specialized not only for the receptive-field or tuning properties they express, but also for the types of dynamics they generate in a given behavior. Furthermore, just as cortical properties like maps and receptive fields are thought to form under the influence of local learning rules (Miller, 1996), it is possible that distinct forms of circuit dynamics – such as stereotyped neuronal sequences or stochastic patterns – similarly result from the application of simple learning rules (Jun and Jin, 2007; Fiete et al., 2010). What processes govern the formation of distinct neural dynamics in the song system, and whether similarly segregated processes exist in other systems, remains to be determined.

Acknowledgments

References

Abeles, M. Corticonics. Cambridge University Press; 1991.

Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Natl Acad Sci. 2009; 106:12518–12523. [PubMed: 19597157]

Andalman AS, Foerster JN, Fee MS. Control of vocal and respiratory patterns in birdsong: Dissection of forebrain and brainstem mechanisms using temperature. PloS One. 2011 in press.

Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science. 2008; 320:630–634. [PubMed: 18451295]

Aronov D, Fee MS. Analyzing the dynamics of brain circuits with temperature: design and implementation of a miniature thermoelectric device. J Neurosci Methods. 2011; 197(1):32–47. [PubMed: 21291909]

Ashmore RC, Renk JA, Schmidt MF. Bottom-up activation of the vocal motor forebrain by the respiratory brainstem. J Neurosci. 2008; 28:2613–2623. [PubMed: 18322104]

Bottjer SW, Halsema KA, Brown SA, Miesner EA. Axonal connections of a forebrain nucleus involved with vocal learning in zebra finches. J Comp Neurol. 1989; 279:312–326. [PubMed: 2464011]

Bottjer SW, Miesner EA, Arnold AP. Forebrain lesions disrupt development but not maintenance of song in passerine birds. Science. 1984; 224:901–903. [PubMed: 6719123]

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci. 1999; 22:567–631. [PubMed: 10202549]

Doya, K.; Sejnowski, T. A novel reinforcement model of birdsong vocalization learning. In: Tesauro, G.; Tourezky, DS.; Leen, TK., editors. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press; 1995. p. 101-108.

Farries MA, Ding L, Perkel DJ. Evidence for "direct" and "indirect" pathways through the song system basal ganglia. J Comp Neurol. 2005; 484:93–104. [PubMed: 15717304]

Farries MA, Perkel DJ. A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. J Neurosci. 2002; 22:3776–3787. [PubMed: 11978853]

Fee MS, Shraiman B, Pesaran B, Mitra PP. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. Nature. 1998; 395:67–71. [PubMed: 12071206]

Fee, MS.; Long, MA. [Accessed August 25, 2011] New methods for localizing and manipulating neuronal dynamics in behaving animals. Curr Opin Neurobiol. 2011. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21763124

Fiete IR, Senn W, Wang CZH, Hahnloser RHR. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron. 2010; 65:563–576. [PubMed: 20188660]

Franz M, Goller F. Respiratory units of motor production and song imitation in the zebra finch. J Neurobiol. 2002; 51:129–141. [PubMed: 11932954]

Gale SD, Perkel DJ. Anatomy of a songbird basal ganglia circuit essential for vocal learning and plasticity. J Chem Neuroanat. 2010; 39:124–131. [PubMed: 19596062]

Glaze CM, Troyer TW. Temporal structure in zebra finch song: implications for motor coding. J Neurosci. 2006; 26:991–1005. [PubMed: 16421319]

Goldberg JH, Fee MS. Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. J Neurophysiol. 2011; 105:2729–2739. [PubMed: 21430276]

Goller F, Cooper BG. Peripheral motor dynamics of song production in the zebra finch. Ann N Y Acad Sci. 2004; 1016:130–152. [PubMed: 15313773]

Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature. 2002; 419:65–70. [PubMed: 12214232]

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

Hohle RH. Inferred components of reaction times as functions of foreperiod duration. J Exp Psychol. 1965; 69:382–386. [PubMed: 14286308]

Immelmann, K. Song development in the zebra finch and other estrildid finches. In: Hinde, RA., editor. Bird vocalizations. Cambridge University Press; 1969. p. 61-74.

Jafari F, Higgins PD. Thermal modeling in cylindrical coordinates using effective conductivity. IEEE Trans Ultrason Ferroelectr Freq Control. 1989; 36:191–196. [PubMed: 18284967]

Jin DZ, Ramazanoğlu FM, Seung HS. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. J Comput Neurosci. 2007; 23:283–299. [PubMed: 17440800]

Jun JK, Jin DZ. Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. PLoS ONE. 2007; 2:e723. [PubMed: 17684568]

Kao MH, Brainard MS. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. J Neurophysiol. 2006; 96:1441–1455. [PubMed: 16723412]

Kao MH, Doupe AJ, Brainard MS. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature. 2005; 433:638–643. [PubMed: 15703748]

Kozhevnikov AA, Fee MS. Singing-related activity of identified HVC neurons in the zebra finch. J Neurophysiol. 2007; 97:4271–4283. [PubMed: 17182906]

Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. Cerebral Mechanisms in Behavior. Wiley; 1951. p. 112-136.

Levelt, WJM. Speaking: from intention to articulation. MIT Press; 1993.

Li M, Greenside H. Stable propagation of a burst through a one-dimensional homogeneous excitatory chain model of songbird nucleus HVC. Phys Rev E Stat Nonlin Soft Matter Phys. 2006; 74:011918. [PubMed: 16907138]

Lilliefors HW. On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. J American Stat Assoc. 1969; 64:387–389.

Lipkind, D.; Tchernichovski, O. [Accessed August 31, 2011] Quantification of Behavior Sackler Colloquium: Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. Proc Natl Acad Sci U S A. 2011. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21436035

Liu W, Gardner TJ, Nottebohm F. Juvenile zebra finches can use multiple strategies to learn the same song. Proc Natl Acad Sci USA. 2004; 101:18177–18182. [PubMed: 15608063]

Long MA, Fee MS. Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature. 2008; 456:189–194. [PubMed: 19005546]

Long MA, Jin DZ, Fee MS. Support for a synaptic chain model of neuronal sequence generation. Nature. 2010 in press. 10.1038/nature09514

Luo M, Perkel DJ. Long-range GABAergic projection in a circuit essential for vocal learning. J Comp Neurol. 1999; 403:68–84. [PubMed: 10075444]

Marler P. Birdsong and speech development: could there be parallels? Am Sci. 1970; 58:669–673. [PubMed: 5480089]

Miller, K. Receptive fields and maps in the visual cortex: Models of ocular dominance and orientation columns. In: Domany, E.; van Hemmen, J.; Schulten, K., editors. Models of Neural Networks III. New York: Springer-Verlag; 1996. p. 55-78.

Nottebohm F, Kelley DB, Paton JA. Connections of vocal control nuclei in the canary telencephalon. J Comp Neurol. 1982; 207:344–357. [PubMed: 7119147]

Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, Serinus canarius. J Comp Neurol. 1976; 165:457–486. [PubMed: 1262540]

Okuhata S, Saito N. Synaptic connections of thalamo-cerebral vocal nuclei of the canary. Brain Res Bull. 1987; 18:35–44. [PubMed: 3030511]

Oller DK, Wieman LA, Doyle WJ, Ross C. Infant Babbling and Speech. J Child Lang. 1976; 3:1–11.

Ölveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol. 2005; 3:e153. [PubMed: 15826219]

Perkel DJ, Farries MA, Luo M, Ding L. Electrophysiological analysis of a songbird basal ganglia circuit essential for vocal plasticity. Brain Res Bull. 2002; 57:529–532. [PubMed: 11923022]

Person AL, Gale SD, Farries MA, Perkel DJ. Organization of the songbird basal ganglia, including area X. J Comp Neurol. 2008; 508:840–866. [PubMed: 18398825]

Riede T, Goller F. Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. Brain Lang. 2010; 115:69–80. [PubMed: 20153887]

Saar S, Mitra PP. A technique for characterizing the development of rhythms in bird song. PLoS ONE. 2008; 3:e1461. [PubMed: 18213370]

Schmidt MF, Ashmore RC, Vu ET. Bilateral control and interhemispheric coordination in the avian song motor system. Ann N Y Acad Sci. 2004; 1016:171–186. [PubMed: 15313775]

Simpson HB, Vicario DS. Brain pathways for learned and unlearned vocalizations differ in zebra finches. J Neurosci. 1990; 10:1541–1556. [PubMed: 2332796]

Suthers, RA. How birds sing and why it matters. In: Marler, PR.; Slabbekoorn, H., editors. Nature's Music: The Science of Birdsong. San Diego, CA: Academic; 2004. p. 272-295.

Tchernichovski O, Lints TJ, Deregnaucourt S, Cimenser A, Mitra PP. Studying the song development process: rationale and methods. Ann N Y Acad Sci. 2004; 1016:348–363. [PubMed: 15313784]

Tchernichovski O, Mitra PP. Towards quantification of vocal imitation in the zebra finch. Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology. 2002; 188:867–878.

Tchernichovski O, Mitra PP, Lints T, Nottebohm F. Dynamics of the vocal imitation process: how a zebra finch learns its song. Science. 2001; 291:2564–2569. [PubMed: 11283361]

Tchernichovski O, Nottebohm Ho, Pesaran, Mitra. A procedure for an automated measurement of song similarity. Anim Behav. 2000; 59:1167–1176. [PubMed: 10877896]

Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature. 2007; 450:1240–1244. [PubMed: 18097411]

Vates GE, Vicario DS, Nottebohm F. Reafferent thalamo- "cortical" loops in the song system of oscine songbirds. J Comp Neurol. 1997; 380:275–290. [PubMed: 9100137]

Veit, L.; Aronov, D.; Fee, MS. [Accessed August 26, 2011] Learning to breathe and sing: Development of respiratory-vocal coordination in young songbirds. J Neurophysiol. 2011. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21697438

Vogels TP, Rajan K, Abbott LF. Neural network dynamics. Annu Rev Neurosci. 2005; 28:357–376. [PubMed: 16022600]

van Vreeswijk C, Sompolinsky H. Chaotic balanced state in a model of cortical circuits. Neural Comput. 1998; 10:1321–1371. [PubMed: 9698348]

Wild JM. Functional Neuroanatomy of the Sensorimotor Control of Singing. Ann N Y Acad Sci. 2004; 1016:438–462. [PubMed: 15313789]

Wilson, HR. Spikes, decisions, and actions. Oxford University Press; 1999.

Yu AC, Margoliash D. Temporal hierarchical control of singing in birds. Science. 1996; 273:1871–1875. [PubMed: 8791594]
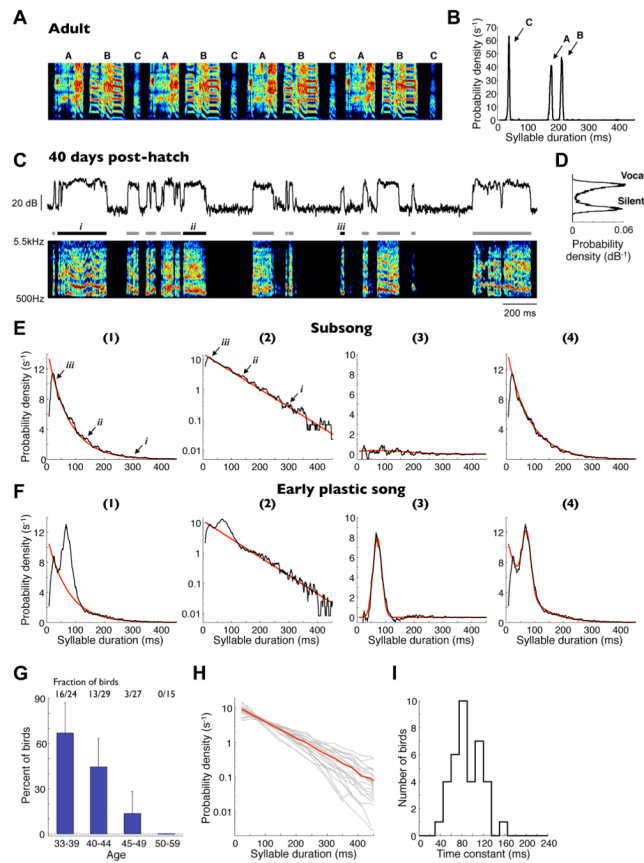
Figure 1.

Subsong exhibits exponential distributions of syllable durations. (a) Spectrogram of a typical adult zebra finch song. Three distinct syllable types are labeled A–C. (b) Syllable duration distribution of all recorded renditions of this song illustrating the highly stereotyped syllable durations. (c) Bottom: spectrogram of a typical song produced by a 40-dph juvenile zebra finch. Bars above indicate detected syllables. Top trace: Song amplitude (log units, dB) in the 1–4 kHz band. (d) Distribution of song amplitudes for the song sample shown. (e) Syllable duration distribution (black traces) for the same bird on (1) a linear and (2) a semi-log scale. The distribution is well-fit by an exponential function (red traces). Labels *i–iii* indicate durations of syllables shown in (c). (3) The residual (black trace) of the distribution following the subtraction of the exponential fit. Also shown is a Gaussian fit to this residual (red trace). (4) Overall gap duration distribution (black trace) and a sum of the exponential and Gaussian fits (red trace). (f) Same fits for syllable durations of a bird in the early plastic song phase. The distribution deviates from an exponential due to the appearance of a peak (Gaussian fit). (g) Fraction of birds whose syllables are well-fit by exponential distributions (Lilliefors statistic, see Methods) for different age groups. Error bars: 95% confidence intervals for a binomial distribution. (h) Syllable duration distributions for all subsong birds (gray lines) and the population average (red line). For clarity, only birds with >2000 recorded syllables are shown. (i) Distribution of exponential time constants for all distributions shown in (h).
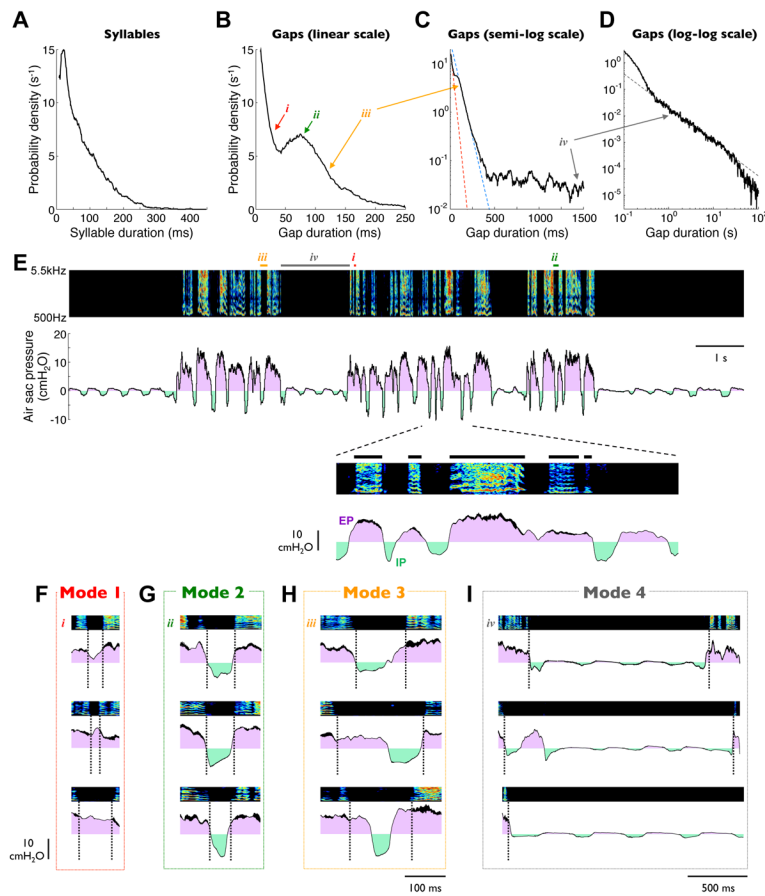
Figure 2.
Silent gaps in subsong are shaped by a diversity of respiratory patterns. Except panel (d), all examples shown are from a single typical subsong bird (47 dph). (a) Syllable duration distribution, exhibiting an exponential shape. (b) Gap duration distribution, exhibiting a complex mixture of timescales including a peak at ~70 ms. (c) Same gap duration distribution shown with a logarithmic vertical axis. Red line: exponential fit to short gaps (<30 ms). Blue line: exponential fit to gaps at 100–200 ms. The long tail deviates from an exponential fit at ~300 ms. (d) Gap duration distribution on a log-log scale. Dashed line: power-law fit. Because the tail contains very few data points, data are shown for a bird from which an exceptionally large amount of singing was recorded. (e) Spectrogram showing several bouts of subsong with simultaneous air sac pressure recording. Periods of positive and negative pressure are indicated with magenta and green, respectively. Gaps labeled *i–iv* correspond to those labeled in (b)–(d). Bottom: detail of the recording, showing four expiratory pressure pulses (EPs) during which 1, 1, 3 and 0 syllables (black bars) are produced. (f) Examples of mode-1 gaps, which contain continuous expiratory pressures. (g) Examples of mode-2 gaps, which are filled by a single IP. (h) Examples of mode-3 gaps, which contain an IP and extended intervals of non-vocalized positive pressure. Note that mode-2 and mode-3 gaps form a continuum; although we illustrate them as two extremes here, the analysis carried out in the paper does not classify individual gaps as belonging to mode-2 or mode-3. (i) Examples of mode-4 gaps, which contain eupneic breathing. The top examples in (f–i) are the gaps highlighted in panels a–c (labeled *i–iv).
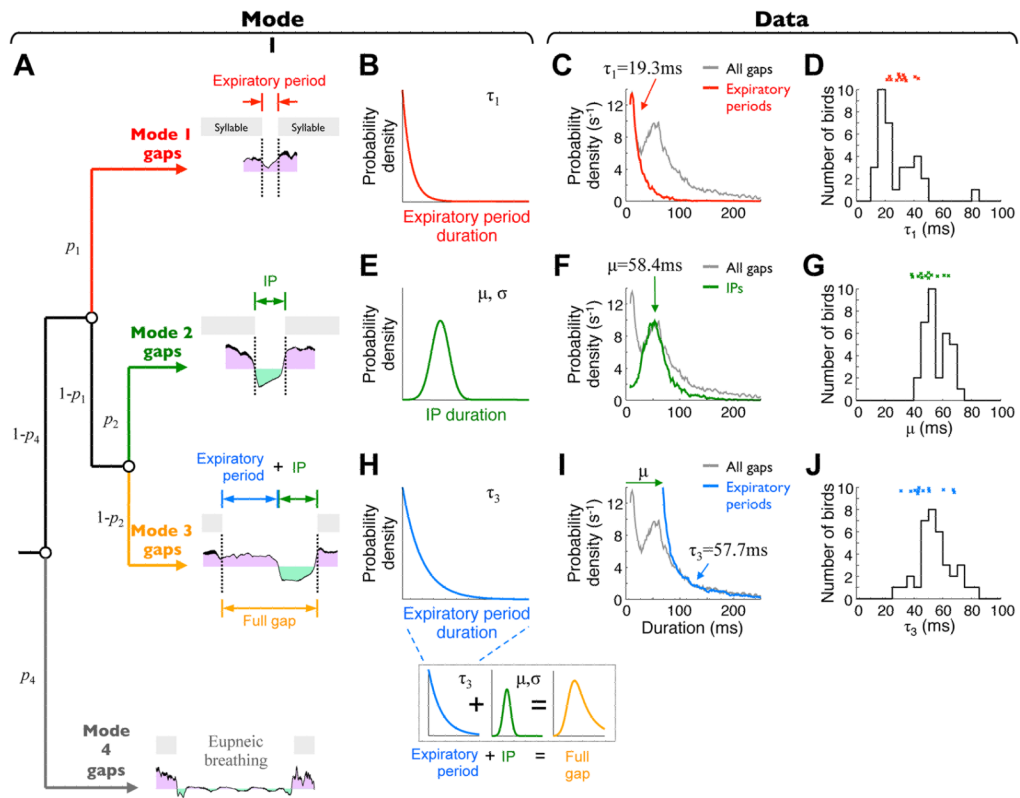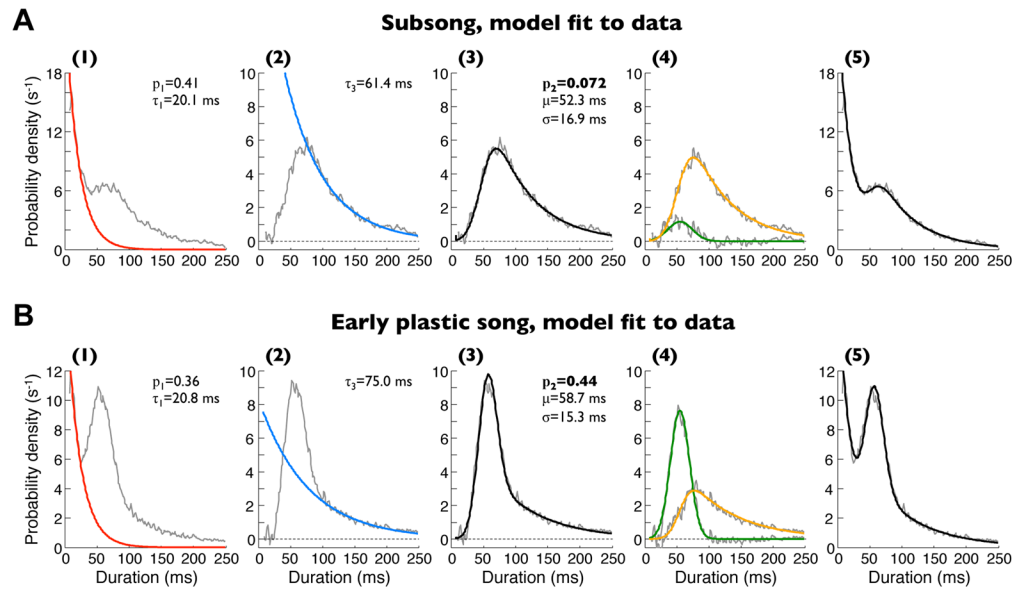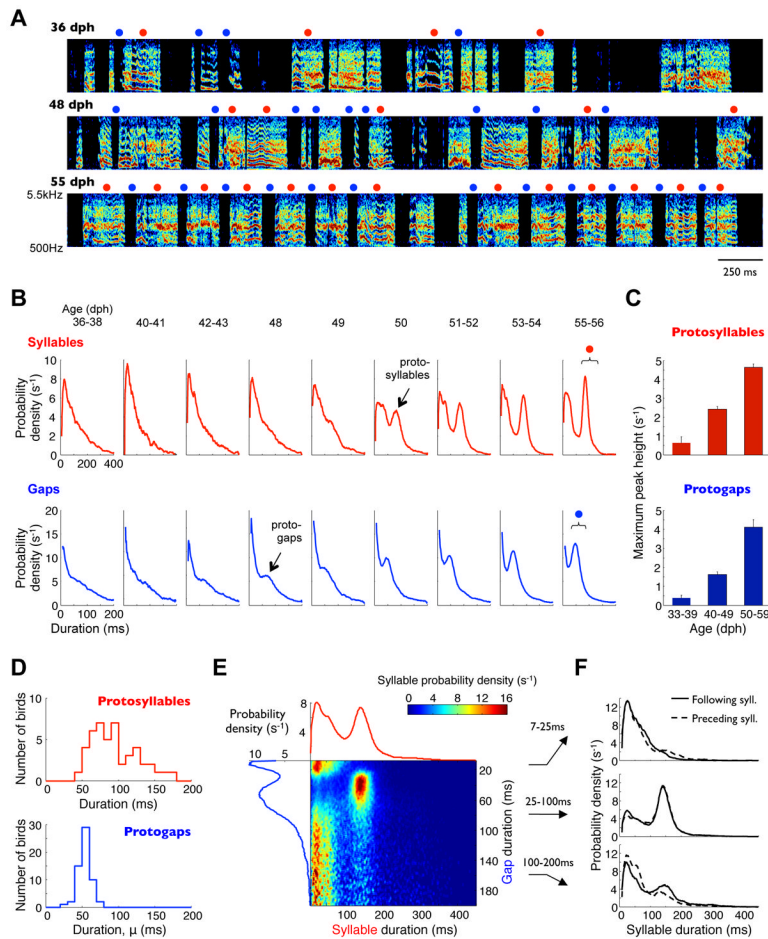
Figure 3.

Gap durations are formed by the underlying respiratory patterns. (a) Model of the respiratory timescales that generate different modes of gap durations. Schematics illustrate all parameters of the gap duration distributions described in the text. (b) An exponential distribution (time constant τ_1) is used to model the durations of expiratory (mode-1) gaps. (c) Gap duration distribution of a subsong bird overlaid with the distribution of mode-1 gaps (red trace). The distributions overlap well at short durations. (d) Histogram of mode-1 gap timescales (τ_1), determined from exponential fits to the overall gap duration distribution at short durations (<~30ms, see Methods). Red symbols show mode-1 gap timescales directly determined from birds with air sac pressure measurements. (e) A Gaussian distribution (mean μ, standard deviation σ) is used to model the distribution of gaps whose durations are tightly coupled to IP durations (mode-2 gaps). (f) Gap duration distribution and IP duration distribution (green trace), measured in the same bird and rescaled to match peak height. (g) Histogram of peak centers identified in gap duration distributions, μ. Green symbols show average IP durations directly determined from air sac pressure measurements. (h) Mode-3 gap durations are modeled by the sum of an exponentially-distributed expiratory period duration (time constant τ_3, blue trace) and a Gaussian-distributed IP duration (green trace), forming an ex-Gaussian duration distribution (orange trace). (i) Gap duration distribution plotted with the distribution of positive-pressure periods in mode-3 gaps, shifted by the mean IP duration (blue trace). (j) Histogram of long (100–200 ms) gap timescales (τ_3). Blue symbols show timescales of the positive-pressure periods in mode-3 gaps, determined from air sac pressure measurements.

Figure 4.
Analysis of gap durations in (a) subsong and (b) early plastic song. (1) Gap duration distribution with exponential fit to short gap durations (red trace, see Methods). The fit was used to estimate the fraction (p_1) and time constant (τ_1) of mode-1 gaps. (2) Residual following the subtraction of the mode-1 exponential fit from the gap duration distribution (gray trace). Overlayed is another exponential fit to this residual at long durations (100–200 ms, blue trace). The fit was used to estimate the exponential time constant of the ex-Gaussian model of mode-3 gaps (τ_3). (3) Fit of a weighted sum of a Gaussian and an ex-Gaussian distribution to the same residual, used to estimate the parameters of mode-2 and mode-3 gaps. (4) Gray traces: residual shown in (3), following subtractions of either of the individual components of the fit (Gaussian or ex-Gaussian). Green and orange traces show the Gaussian and ex-Gaussian components of the fit separately. Amplitude of the Gaussian component (p_2, green trace), describing the contribution of mode-2 gaps, is substantially greater in plastic song than in subsong. Note that individual gaps in this step have not been classified as mode-2 or mode-3; the distributions only show how much of the fit in (3) was assigned to the Gaussian and the ex-Gaussian modes. (5) Overall gap duration distribution from the same bird (grey trace) and the fit of the full model (black trace; weighted sum of an exponential, a Gaussian, and an ex-Gaussian).

Figure 5.
Early song development is characterized by appearance of consistent timing. (a) Spectrograms of songs produced by one bird at three developmental stages. Red circles: syllables 140–240 ms long. Blue circles: gaps 30–90 ms long. (b) Developmental progression of syllable and gap duration distributions from the same bird. Brackets indicate duration ranges marked above (red and blue circles). (c) Heights of peaks in syllable and gap duration distributions (corresponding to protosyllables and protogaps), quantified across all birds. Bars indicate median values across birds in a given age group; error bars are bootstrap standard errors of the median. (d) Distributions of protosyllable and protogap durations across all birds. (e) Probability distribution of all consecutive syllable and gap durations in an early plastic-song bird. Each row of the color-coded matrix is individually normalized to a sum of 1, such that values indicate syllable durations conditional on gap duration. Also shown are the overall gap duration distribution (blue trace, left) and syllable duration distribution (red trace, top). (f) Duration distributions of syllables that precede and follow gaps of three different duration ranges. Protogaps and protosyllables tend to follow each other.
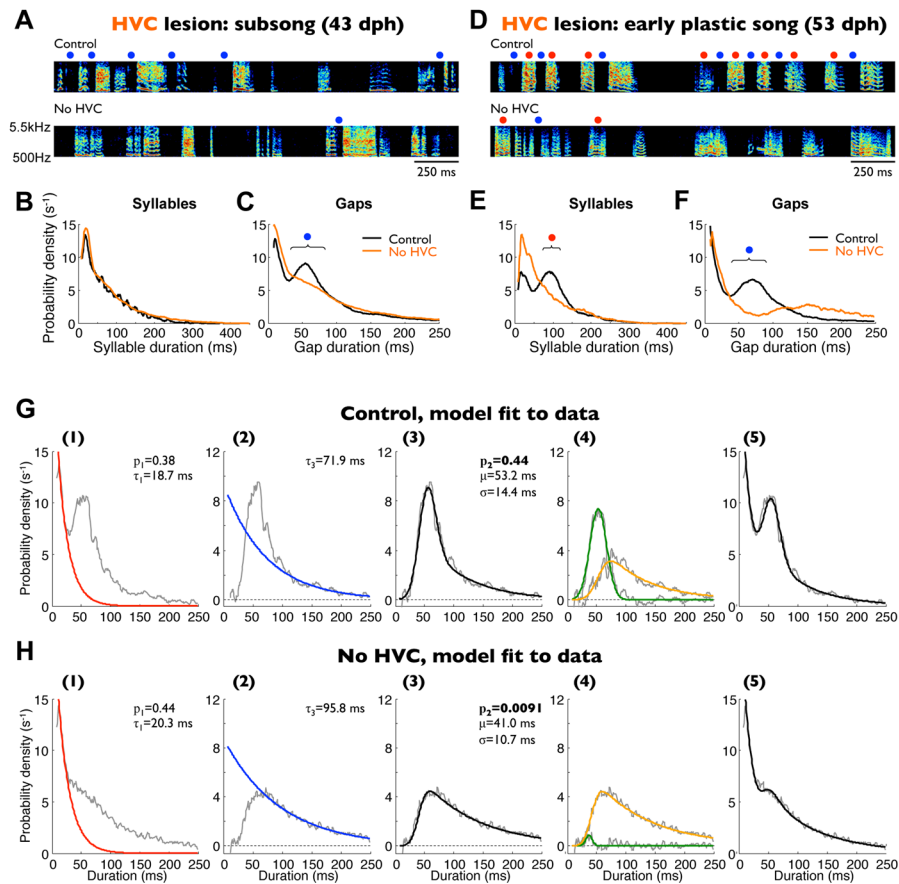
Figure 6.
Effects of HVC elimination on early singing. (a) Typical spectrograms of a subsong bird before and after complete bilateral HVC lesions. Blue circles: gaps 30–90 ms long. (b) Syllable duration and (c) gap duration distributions for the same bird before and after the lesions. Bracket indicates the duration range marked above. (d–f) Same plots for an early plastic-song bird before and after HVC lesion. Red circles: syllables 80–120 ms long. Blue circles: gaps 30–90 ms long. (g) Analysis of the temporal modes in normal plastic song gap distribution. Data are plotted the same way as in Figure 4. (h) Analysis of gaps produced by the same bird after bilateral HVC lesions. Note that the major effect of the lesion was the elimination of mode-2 gaps (panel 4, Gaussian fit, green trace).
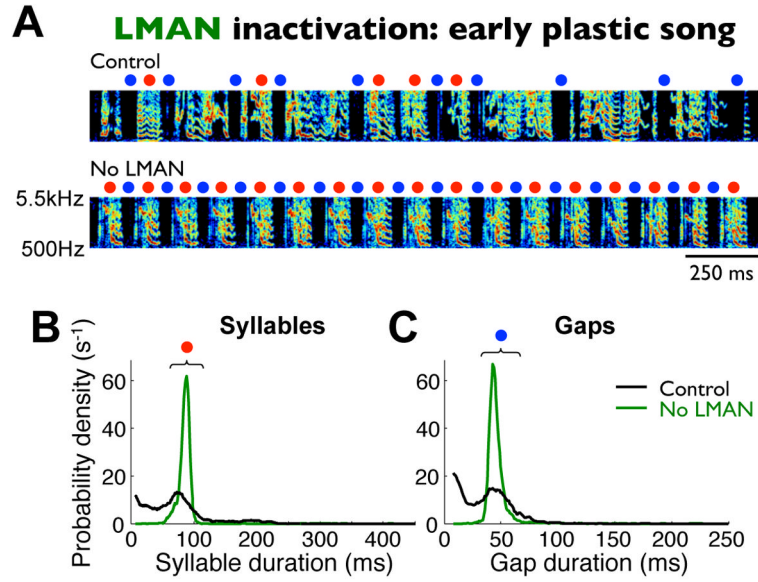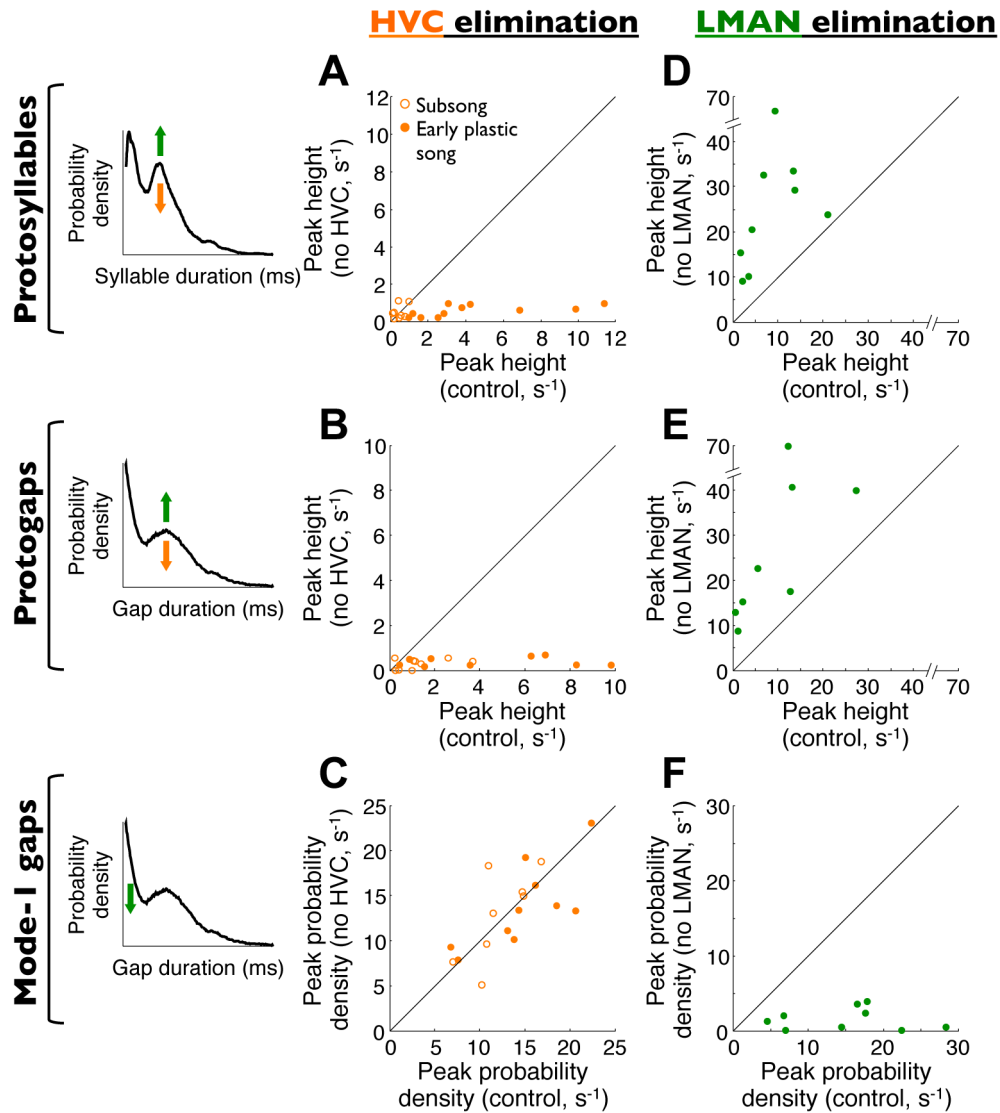
Figure 7.
Effects of LMAN elimination on early singing (a) Typical spectrograms of an early plastic-song bird before and after bilateral LMAN inactivation. Red circles: syllables 60–110 ms long. Blue circles: gaps 30–60 ms long. (b) Syllable duration and (c) gap duration distribution for the same bird before inactivation (black trace) and after inactivation (green trace). Brackets indicate duration ranges marked above.

Figure 8.
Summary of bilateral HVC and LMAN elimination experiments across birds. Note that only HVC eliminations are shown for subsong birds because these do not sing following LMAN inactivations (Aronov et al., 2008). Left panels: schematic illustrations of the effects of HVC elimination (orange arrows) and LMAN elimination (green arrows) on syllable and gap distributions. (a) Scatter plot of the size of the peak in syllable durations (protosyllables) before and after HVC elimination. (b) Size of the peak in gap durations (protogaps) before and after HVC elimination, calculated as the magnitude of the components corresponding to mode-2 gaps. (c) Fraction of mode-1 (expiratory) gaps before and after HVC elimination. (d–f) Effects of LMAN elimination on the same quantities as those shown in (a–c).
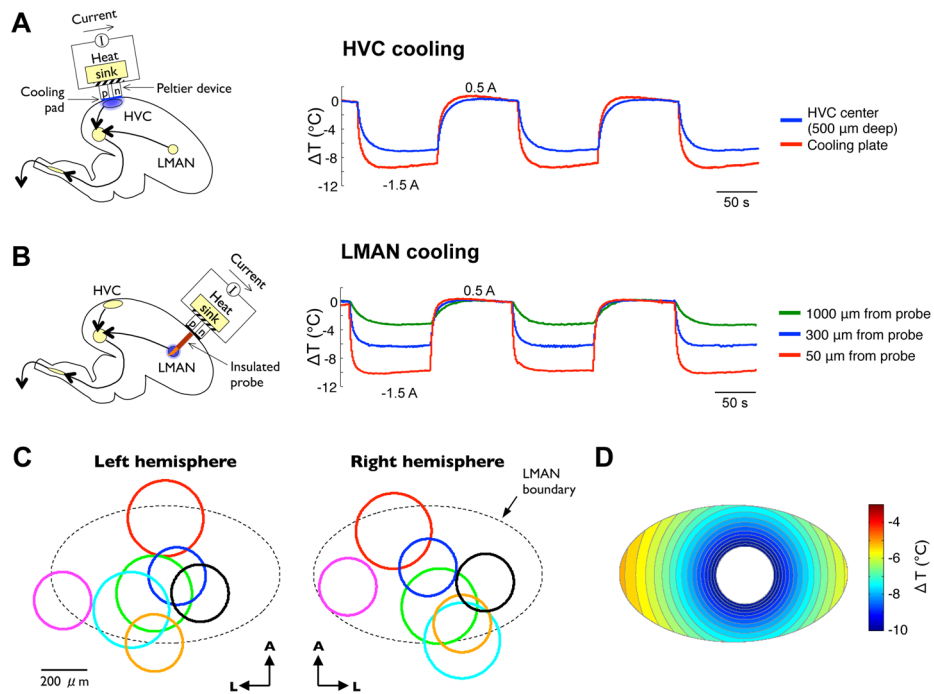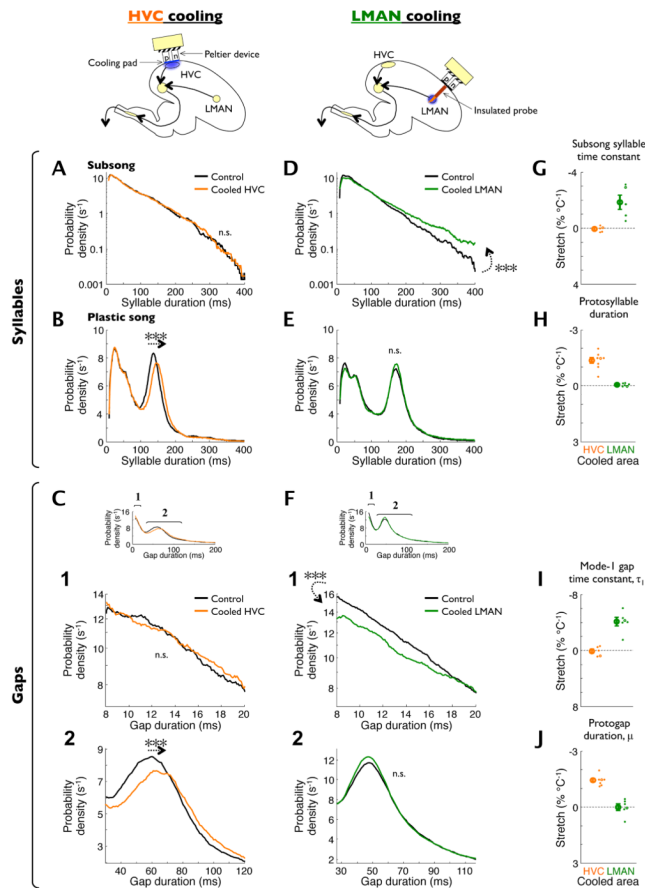
Figure 9.
Calibration of devices for cooling HVC and LMAN in juvenile zebra finches. (a) Left: schematic of the device for cooling HVC using cooling pads placed against thinned cranium above HVC. Right: temperature at the cooling plate of the device and in the center of HVC measured in an awake bird. Current is alternated between −0.5 and 1.5 A every 100 s. The same protocol is used for all cooling experiments reported here. (b) Left: schematic of the device for cooling LMAN using thermally conductive probes. Right: temperature in LMAN at various distances from the cooling probe of the device using the same electric current protocol. (c) Approximate locations of cooling probes in LMAN for all analyzed birds. The probes in the left and right hemisphere of each bird are indicated by circles of matching color. Note that probe diameter was 250 μm in some birds and 330 μm in others (indicated by smaller and larger circles). A: anterior; L: lateral. (d) Example of simulated temperature around the probe in a horizontal section of LMAN (see Methods).

Figure 10.
Biophysical dynamics intrinsic to HVC and LMAN are involved in timing different components of early singing. (a) Distributions of subsong syllable durations produced by a single bird at normal body temperature and during HVC cooling, plotted on a semi-logarithmic scale. HVC cooling had no effect on subsong syllable durations. (b) Syllable duration distributions for a plastic song bird on a linear scale. HVC cooling increased protosyllable durations. (c) Effect of HVC cooling on gap duration distributions. (1) Detail of the gap duration distribution at short durations, plotted on a semi-logarithmic scale. HVC cooling had no effect on mode-1 (expiratory) gaps. (2) Detail of the gap duration distribution on a linear scale, showing the protogap peak. HVC cooling prolonged protogaps. (d–f) Effects of LMAN cooling on syllable and gap durations, plotted as in (a–c). (d) LMAN cooling prolonged subsong syllables. (e) LMAN cooling had no effect on the durations of protosyllables in plastic song. (f) LMAN cooling increased the duration of mode-1 gaps (1), but had no effect on the durations of protogaps (2). (g) Population summary of the effect of HVC cooling (orange) and LMAN cooling (green) on subsong syllable durations. Each dot indicates the effect for an individual bird. Open dots indicate the examples shown in (a) and (d). Error bars: standard errors across all birds. (h–j) Population summary of the effects of HVC and LMAN cooling on protosyllable durations, mode-1 gap timescales, and protogap durations.