Simple Motor Gestures for Birdsongs

Tim Gardner, G. Cecchi,* and M. Magnasco

Center for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021

R. Laje and Gabriel B. Mindlin

Departamento de Física, FCEN, UBA, Ciudad Universitaria, Pab. I (1428), Buenos Aires, Argentina

We present a model of sound production in a songbird's vocal organ and find that much of the complexity of the song of the canary (*Serinus canaria*) can be produced from simple time variations in forcing functions. The starts, stops, and pauses between syllables, as well as variation in pitch and timbre are inherent in the mechanics and can often be expressed through smooth and simple variations in the frequency and relative phase of two driving parameters

Human language and the song of many bird species are both learned by juveniles through experience. The ontogeny of learned song begins with early food-begging calls, continues through more complex bablings known as subsong, to eventually reach an adult form of song which is first plastic, and later stereotyped [1]. This development depends on auditory experience, for abnormal forms of song develop if a juvenile is deafened and unable to hear its own sound production or is isolated from the experience of any adult song model to copy. The time course of development and the dependence on auditory experience are similar to aspects of human speech, suggesting there may be common principles of learning and memory underlying human speech and the song of birds [2]. A variety of current experimental approaches are directed towards a better understanding of how song perception and production are represented by the brain and ultimately how experience shapes the neural changes which establish these representations [3,4]. However, the broad extent of neural activity involved in song presents an increasing challenge for interpretation [5]. In order to understand the link between brain activity and song, a physical understanding of the process of sound production in the songbird's vocal organ is essential. In the following, we demonstrate a simple mechanism which can account for the acoustic structure of many song elements. The physical model we describe involves a minimum number of parameters and can account for a wide range of observed birdsong elements. Fitting of the model's parameters in order to reproduce recorded songs suggests that the acoustic complexity of individual elements in the song of canaries arises from simple modifications of a very generic gesture in lung pressure and vocal fold tension.

The vocal organ of songbirds, known as the syrinx, has been the subject of a long history of interesting studies. The existence of vibrating membranes capable of generating sound waves was established by Ruppel in 1933 [6]. At present, the precise mechanisms of sound production in the syrinx are still a matter of debate (see [7] and [8,9]). Current experiments suggest that the syrinx generates sound primarily through oscillation of the lateral labia—tissue folds which open and close the air passage from the bronchi to the trachea [10]. Through direct videography of a phonating syrinx, Larsen and Goller give evidence that these labia function in a manner homologous to the human vocal folds [11]. In support of this, Fee *et al.* have demonstrated that the transition from periodic to chaotic vibrations in the Zebra finch syrinx can be modeled by a classic two-mass model of human vocal fold oscillation [12,13]. In contrast to these aperiodic sounds, in what follows, we examine vocalizations composed of locally periodic waveforms, typical of canaries and many other species. The primary element of the canary lexicon is the syllable, a 15–300 ms vocalization of which the canary, Serinus canaria, typically has a few dozen types.

One of the simplest models to account for the transfer of the kinetic energy of air to vocal fold oscillations was introduced by Titze [14]. It is built upon experimental observation that the human vocal folds support both lateral oscillations and an upward propagating surface wave [14], and it successfully predicts the fundamental frequencies of voiced sounds in terms of sensible physiological parameters. There is no direct evidence in birds, but this "flapping" mode of oscillation is consistent with recent videography of the bird's syrinx during song, and in the isolated syrinx [11,13]. As illustrated in Fig. 1, the opposing labia have a convergent profile when they move away from each other and a more planar profile when they move towards each other. This results in a greater pressure on the labia during the opening phase and an overall gain in energy in each cycle of oscillation. This mechanism does not depend on the geometrical details of the folds, but the calculations are easier with a simplifying geometric hypothesis. By approximating the labia shape with straight edges as illustrated in Fig. 1, the assumption of a flapping mode can be written in terms of a phenomenological constant τ as follows:

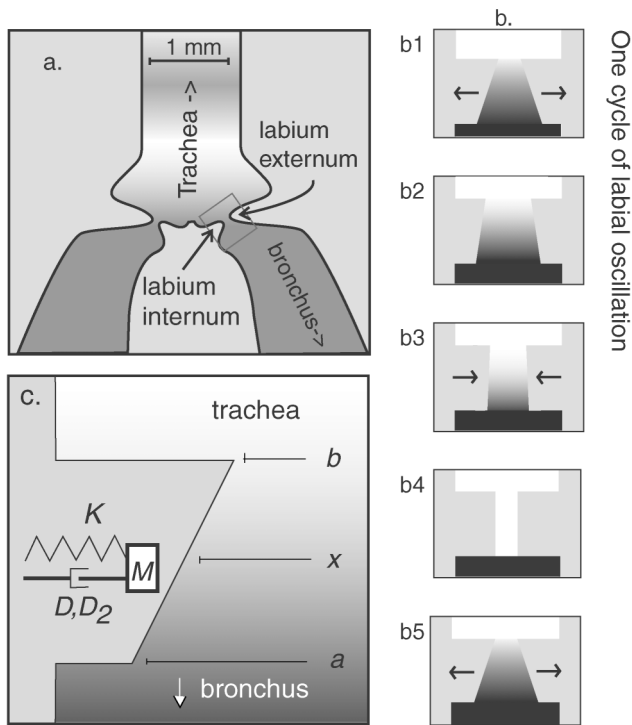$$a = a_0 + x + \tau \frac{dx}{dt}, \qquad (1)$$

FIG. 1. Illustration of the syrinx, terms of the model, and labial dynamics. Panel (a) illustrates the songbird syrinx which has a phonatory labium at the top of each bronchus. Panel (b) illustrates our assumption of the shape of the labia during a cycle of oscillation. Panel (c). is a diagram of the terms in our model. (Here we have drawn only one side of the symmetric pair of labia in the model.) In each figure, darker shading indicates higher pressure according to Eq. (3).

$$b = b_0 + x - \tau \frac{dx}{dt}. \tag{2}$$

The upper edge of the labium is denoted a, the lower edge b, and the center of the labium x. Following Titze, who uses a (phenomenologically modified) Bernoulli equation [14], the pressure averaged over the surface of the labia can be written as a function of bronchial pressure P_b and labial position:

$$P_f = P_b(1 - a/b). \tag{3}$$

Substituting equations for b and a into that for P_f and equating this averaged pressure with the driving force of a damped harmonic oscillator, we have

$$M\ddot{x} + D\dot{x} + D_2(\dot{x})^3 + Kx = P_b \frac{a_0 - b_0 + 2\tau\dot{x}}{x + b_0 + \tau\dot{x}}. \tag{4}$$

The parameters M, K, D, and D_2 describe the mass, restitution constant, and coefficients of a nonlinear dissipation—all per unit area. The nonlinear dissipation term $(\dot{x})^3$ is introduced *ad hoc* so that the variable x can take values only between precise boundaries, mimicking collisions [15].

In this model, we make the assumption that the bird controls vocalizations through the bronchial pressure P_b and the labial elasticity K. We assume that the elasticity term corresponds to the concerted activity of one or more

syringeal muscles which change the radial stiffness of the labia. Presumably, this term can be related to the activity of the muscles most correlated with changes in the fundamental frequency of the sound [16]. In what follows, we refer to this term as tension.

The behavior of the system for different parameters of pressure and tension is displayed in Fig. 2, where the left boundary of the shaded region indicates the critical values at which self-oscillations are induced. As the pressure P_b is increased, the equilibrium state loses its stability in a Hopf bifurcation and energy is transferred into the labia (there is an effective negative dissipation). Near the boundary, the oscillations have very little spectral content. As pressure is increased to points deeper in the region of oscillation, spectral content of the waveform increases monotonically. Given an estimation of physical parameters consistent with experimental measurements (see values in the caption of Fig. 2), we find that the simulations produced by the model have a frequency range and spectral content comparable to natural vocalizations.

The final feature of the model is the filtering of the sound produced by the oscillations through the acoustic properties of the vocal tract [17]. The filter was constructed by approximating the trachea and beak by two tubes of lengths L_1, L_2 and areas A_1 and A_2. The input pressure P_i (proportional to the time derivative of the flow at the base of the trachea) generates a wave which impinges on the boundary between the two tubes and is partially reflected back towards the trachea and partially
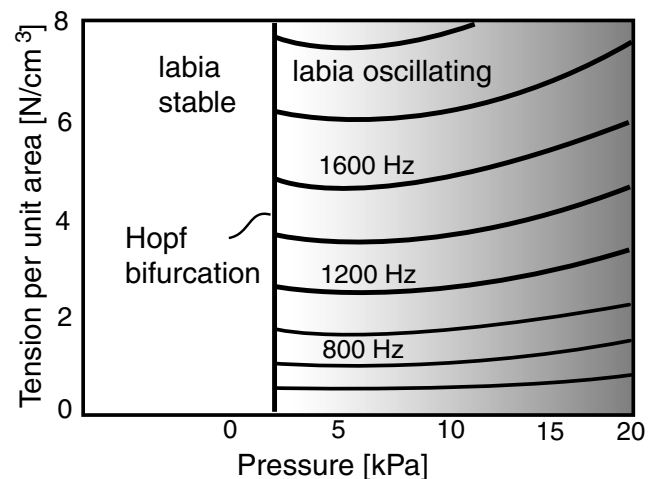

FIG. 2. Illustration of sound production as a function of P_b and K. The left boundary of the shaded region indicates the critical values of pressure and tension at which self-oscillations are induced. The dark bands in this figure are numerically calculated contours of isofundamental frequency. Higher spectral content is schematically indicated by darker shading. (No vocal tract filtering is applied.) $M = 0.005$ g/cm^2, $D = 5.0$ dyne sec/cm^3, $D_2 = 0.01$ dyne sec^3/cm^5, $\tau = 0.00015$ sec, $b_0 = 0.04$ cm, and $a_0 = 0.02$ cm. With these, and P_b ranging from 0 to 3 kPa (as experimentally observed), and K from 0 to 8 N/cm^3, the model generates frequencies within canary range, and with oscillations in the range of 0 to 250 μm, consistent with experimental evidence [10,13].

transmitted to the second tube. The proportion of sound energy reflected at the junction is defined by a ratio of areas: $r_{1,2} = (A_1 - A_2)/(A_1 + A_2)$, and the portion of the sound energy transmitted through the junction is simply $t_{1,2} = 1 - r_{1,2}$. At the interface between the second tube and the atmosphere, the wave is again partially reflected and partially transmitted. Calling $a(t)$ $[b_b(t)]$ the forward [backward] wave in the first tube and $b_f(t)$ $[c_b(t)]$ the forward [backward] wave in the second tube, the equations accounting for the boundary conditions are [18]

$$a(t) = P_i(t) + b_b(t - \tau_1), \qquad (5)$$

$$b_b(t) = r_{1,2}a(t - \tau_1) + t_{1,2}c_b(t - \tau_2), \qquad (6)$$

$$b_f(t) = t_{1,2}a(t - \tau_1) + r_{1,2}c_b(t - \tau_2), \qquad (7)$$

$$c_b(t) = \alpha b_f(t - \tau_2), \qquad (8)$$

where α accounts for the reflection coefficient of the interface between the third tube and the atmosphere (no losses being $\alpha = -1$), and the time τ_i is the time it takes a sound wave to travel L_i. The aperture of the beak alters frequency content of the sound in a manner similar to that observed experimentally [19,20]. For simplicity, in each simulation illustrated in this Letter, beak aperture was fixed.

Given a physically based model with few parameters, what kind of control of P_b and K might the bird employ to generate the elements of its song? The oscillations that are established in a typical vocalization are in the order of 1 or 2 kHz while a syllable duration is between 10 and 300 ms. With this separation of time scales, if the paths in parameter space are swept slowly, at each time t^* the system will behave basically as expected in a steady condition with $P_b(t^*) = P_b^*$ and $K(t^*) = K^*$. Given the simplicity of the parameter space defined in Fig. 2, this implies that the time course of K will essentially trace out the time course of the fundamental frequency of the syllable. Pressure will control not only sound amplitude but also the relative strength of the higher harmonics.

For example, an upsweep syllable like the second one displayed in Fig. 3 requires that the control trajectory begin at a point of low pressure and tension and enter the region of oscillation pictured in Fig. 2 by increasing pressure. While inside the region of oscillation, tension must be incremented to produce the upsweep in frequency. Finally, at a position of high K, the pressure must be reduced to end the vocalization by crossing out of the region of oscillation. In accord with this simple example, we found that many canary syllables could be approximated by excursions as simple as harmonic oscillations of pressure and tension. Specifically, we generated elliptical excursions in pressure and tension which had the following form:

$$P_b = P_o + A\cos[\phi(t) + \phi_i], \qquad (9)$$

$$K = K_o + B\cos[\phi(t) + \phi_j]. \qquad (10)$$

For the figures in this Letter, the parametrization of time is either linear or else involves a slight slowing for
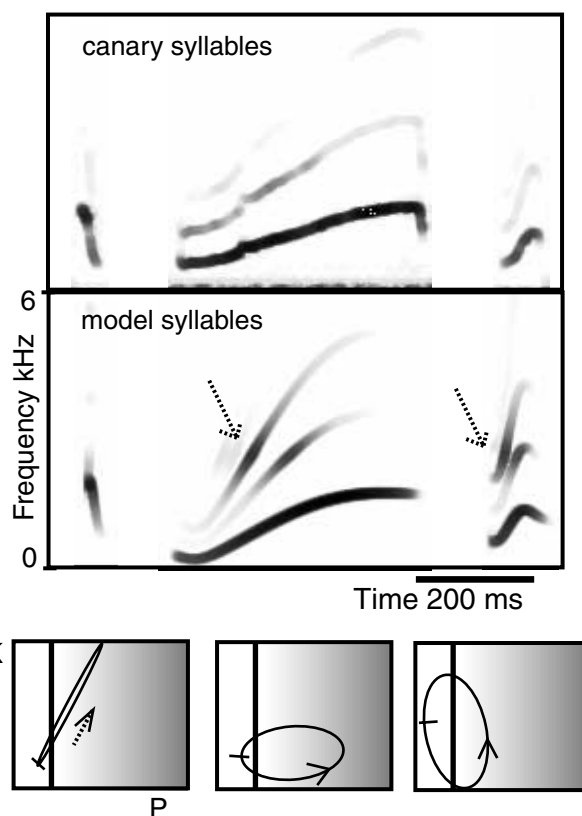

FIG. 3. Three natural syllables, three artificial syllables, and their corresponding control trajectories. The elements are illustrated in time-frequency format, which was computed through sliding, overlapped Fourier transforms. (Pixel grey level represents the power of sound on a logarithmic scale.) Parameters are drawn with respect to Fig. 2. Arrows indicate the direction of traversal over the curve. In each case, the trajectory begins at the minimum pressure value and traverse the ellipse in a counterclockwise direction. As discussed in the text, the parametrization of time in each ellipse is chosen to cause a slowing of traversal rate near the point of maximum pressure. For each syllable, the vocal tract filter has the following parameters: $L_1 = 3.0$ cm, $L_2 = 1.0$ cm, $A_1 = 0.15$ cm, $A_2 = 0.4$ cm, and $\alpha = 0.8$. This filter leads to a transient emphasis of the power of the second and third harmonics as indicated by the arrows. A further point can be illustrated with this figure. For fast syllables, sound production at a particular point P_b and K is not equivalent to the steady state sound production. Were it equivalent, syllable one would be almost symmetric in time. In effect, there is some overlap between the time scale of equilibration of the oscillation and the time scale of the fastest syllables.

high pressure. $[\dot{\phi}(t) = c$, or $\dot{\phi}(t) \simeq 1 - e^{[\phi(t) - \phi_0]^2/\sigma^2}$, for constants ϕ_0, and σ.] We want to emphasize that this parametrization of time is the same for both pressure and tension. This means that the time scale and the complexity of the two control parameters are equivalent. In this form, an ellipse in the space of pressure and tension characterizes the identity of each syllable. Figure 3 illustrates the similarity between three elliptical control trajectories and three notes chosen from a single bird's song. For this figure, the fitting procedure was qualitative in nature. The range of K could be estimated from the frequency contours of Fig. 2, and the phase and amplitude of the ellipse were chosen to
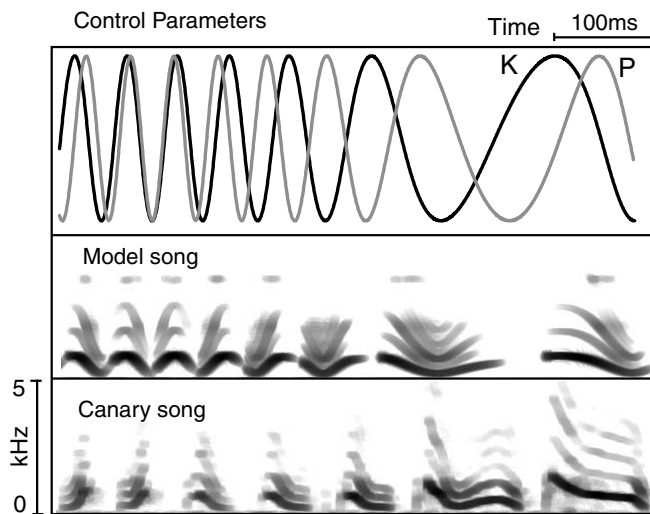
FIG. 4. Synthetic signals generated by slow modulations of pressure and tension. (P_b and K are taken from a brief sample of the positions of two identical, spring-coupled pendulums subject to a harmonic rescaling of time. P_b ranges between 0 and 2 kPa. K ranges between 0 and 7 N/cm³.) On the bottom of the figure, a consecutive sequence of syllables from natural canary song is included for comparison with the synthetic signals.

reproduce the qualitative shape of the sonogram. Work is in progress to build an algorithmic fitting procedure.

The elliptical paths in parameter space described above are characteristic of the solutions of a wide range of coupled oscillators. This suggests that the structure of some canary songs, determined by the ordering of a sequence of syllables, may be modeled by the slow dynamics of changing phase relationships in the oscillations controlling P_b and K. To illustrate this hypothesis, Fig. 4 demonstrates the response of the model to parameter oscillations characterized by two oscillators whose frequency and relative phase are slowly changing. The resulting sequence of vocalizations demonstrates elements and transitions qualitatively similar to a sequence of syllables measured experimentally. Specifically, the starts, stops, and pauses between syllables, as well as variation in pitch and timbre, are reproduced by the simple oscillations of P and K.

In this work we found that the spectral and temporal diversity of many canary vocalizations can be modeled by the response of a nonlinear equation to simple forcing functions. One particular implication of our model is that the control of syringeal tension does not need to be any more detailed than the recorded pressure in a particular air sac which is known to be smooth for most canary syllables [21,22]. For many songs, simple non-self-intersecting cycles in the P_b and K parameter space were able to reproduce the data. These cycles could be approximated by ellipses, which ultimately suggests that a fundamental variable to be controlled in order to build a song is the phase difference between simple oscillations in bronchial pressure and vocal fold tension.

*Present address: IBM T. J. Watson Research Center, Yorktown Heights, NY.
  Email address:
[1] F. Nottebohm, M. E. Nottebohm, and L. Crane, Behav. Neural. Biol. **46**, 445–71 (1986).
[2] M. S. Brainard and A. J. Doupe, Nature Rev. **1**, 31–40 (2000).
[3] A. C. Yu and D. Margoliash, Science **273**, 1871–1875 (1996).
[4] S. Ribeiro, G. A. Cecchi, M. Magnasco, and C. V. Mello, Neuron **21**, 359–371 (1998).
[5] F. Nottebohm, in *The Design of Animal Communication,* edited by M. D. Hauser and M. Konishi (MIT Press, Cambridge, MA, 2000), pp. 37–62.
[6] W. Ruppel, J. Ornithol. **81**, 433–522 (1933).
[7] C. H. Greenewalt, *Bird Song: Acoustics and Physiology* (Smithsonian Institution Press, Washington, DC, 1968).
[8] N. H. Fletcher, J. Theor. Biol. **135**, 455–481 (1988).
[9] P. Mergell, W. T. Fitch, and H. Herzel, J. Acoust. Soc. Am. **105**, 2020–2028 (1999).
[10] F. Goller and O. N. Larsen, Proc. Natl. Acad. Sci. U.S.A. **94**, 14 787–14 791 (1997).
[11] O. N. Larsen and F. Goller, Proc. R. Soc. London B **266**, 1609–1615 (1999).
[12] K. Ishizaka and J. L. Flanagan, Bell Syst. Tech. J. **51**, 1233–1268 (1972).
[13] M. S. Fee, B. Shraiman, B. Pesaran, and P. P. Mitra, Nature (London) **395**, 67–71 (1998).
[14] I. R. Titze, J. Acoust. Soc. Am. **83**, 1536–1552 (1988).
[15] The system $\dot{u} = v - \alpha u^3 + Du, \dot{v} = -\epsilon u$ is a classic example of the theory of relaxation oscillations, and it describes a bounded motion of the variable v. Writing $v = x$ and $y = -\epsilon u$, we obtain a standard system that reads $\dot{x} = y, \dot{y} = -\epsilon x + Dy - \frac{\alpha}{\epsilon^2}y^3$. Written as a second order oscillator, it reads as $\ddot{x} + \epsilon x - D\dot{x} + \frac{\alpha}{\epsilon^2}(\dot{x})^3 = 0$. The cubic term guarantees a precise bounded motion of x. See V. I. Arnold, V. S. Afrajmovich, Yu. S. Ilyashenko, and L. P. Shilnikov, *Bifurcation Theory and Catastrophe Theory* (Springer, Berlin, 1999).
[16] R. A. Suthers, Neurobiology **33**, 632–652 (1997).
[17] I. R. Titze, *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ, 1994).
[18] M. A. Trevisan, M. C. Eguia, and G. B. Mindlin, Phys. Rev. E **63**, 026216 (2001).
[19] W. J. Hoese, J. Podos, N. C. Boetticher, and S. Nowicki, J. Exp. Biol. **203**, 1845–1855 (2000).
[20] R. Suthers and F. Goller, in *Current Ornithology,* edited by V. Nolan, Jr. *et al.* (Plenum Press, New York, 1997), Vol. 14.
[21] W. A. Calder, Comp. Biochem. Physiol. **32**, 251–258 (1970).
[22] R. S. Hartley and R. A. Suthers, J. Comp. Physiol. **165**, 15–26 (1989).