Experimental support for a model of birdsong production

G. B. Mindlin,[1,5] T. J. Gardner,[2] F. Goller,[3] and R. Suthers[4]

[1]*Institute for Nonlinear Science, UCSD, 9500 Gilman Drive, La Jolla , California 92093, USA*
[2]*Center for Studies in Physics and Biology, Rockefeller University, New York, USA*
[3]*Department of Biology, University of Utah, Salt Lake City, Utah, USA*
[4]*School of Medicine, Indiana University, Bloomington, Indiana, USA*
[5]*Departamento Fisica, FCEyN, UBA, Argentina*

In this work we present an experimental validation of a recently proposed model for the production of birdsongs. We have previously observed that driving the model with simple functions of time, which represent tensions in vocal muscles, produces a wide variety of sounds resembling birdsongs. In this work we drive the model with functions whose time dependence comes from recordings of muscle activities and air sac pressure. We simultaneously recorded the birds' songs and compared them with the synthetic songs. The model produces recognizable songs. Beyond finding a qualitative agreement, we also test some predictions of the model concerning the relative levels of activity in the gating muscles at the beginning and end of a syllable.

I. INTRODUCTION

Birdsongs contain some of the most complex and beautiful sounds in nature. Moreover, for the species known as songbirds (or Oscines), learning plays an important role in the development of their complex vocal communication. The possible existence of common principles of learning and memory underlying human speech and birdsong has made Oscines an important animal model with which to study the mechanisms of vocal learning and sound production. Part of the research effort has focused on understanding how the representation and execution of vocal communication is manifested in the brain. This line of research deals with the description of the connectivity of the different nuclei involved in these processes (for a review, see Ref. [1]).

Between the behavioral output (song) and this complex representation at the level of the brain (central motor control) stands the avian vocal organ. A continuous research effort has been carried out to investigate the motor activity of various muscles and their role in controlling the vocal organ [2]. The avian vocal organ, the syrinx, is located at the junction of the primary bronchi and trachea (see Fig. 1). Freely moveable connective tissue masses, labia, are set in vibratory motion by expiratory airflow. Their position in the airstream can be actively controlled by syringeal muscles (gating muscles). Similarly, action by syringeal muscles is thought to allow control of the tension of the vibrating labia [see Fig. 1(a)]. The electromyographic activity of the muscles involved in these processes has been recorded for some species together with the vocal behavior. Instructions generated by the nervous systems control the syringeal muscles and, therefore, ultimately drive the physical device, which generates the sound.

Recently, a dynamical model of the processes involved in birdsong production was presented [3,4]. The parameters are, in principle, the time varying forces produced by vocal muscles. In this work, we measured electromyographs (EMGs) from muscles and chose magnitudes for these functions that come closest to replicating the birdsongs. Beyond

the qualitative agreement between synthetic and natural songs, we test a specific prediction of the model concerning the relative levels of activity of gating muscles at the beginning and end of individual song syllables [4].

A review of results concerning the neuromuscular control of birdsong and their relationship with the model is presented in Sec. II. In Sec. III we show the result of driving the model with actual data from Northern cardinals (Cardinalis cardinalis). Section IV deals with the validation of the predictions of the model concerning gating activity. In Sec. V, we present our conclusions.

II. NEUROMUSCULAR CONTROL OF BIRDSONG AND ITS RELATION WITH A PHYSICAL MODEL OF THE SYRINX

Most birdsongs are composed of repetitions and alternations of simple continuous sounds, called syllables [5]. Syl-
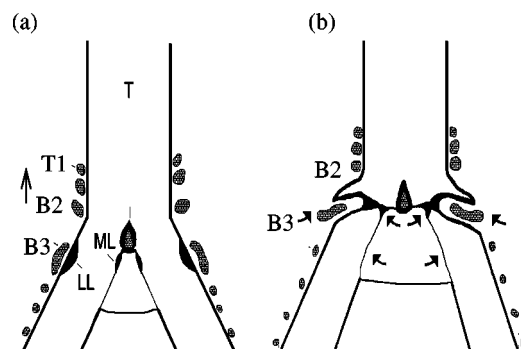

FIG. 1. A schematic ventral view of a songbird syrinx. $T1$, $B2$, and $B3$ stand for the first tracheal cartilage, second and third bronchial cartilages, respectively. ML and LL stand for medial and lateral labia. The ventral muscle vS is attached to $B2$. Therefore, as vS is contracted, $B2$ is moved upwards. As the separation between $B2$ and $B3$ is increased, the tension of the labia also increases (a). Dorsal muscles induce the rotation of bronchial cartilages ($B3$), as illustrated in (b). Adapted from Ref. [14].

68 041908-1

lables are produced by the syrinx, which is a bipartite structure located at the junction between the primary bronchi. In each syringeal half, there is a sound generator creating the potential for generating two independent sounds simultaneously. In the cardinal, however, a male will use only one side to produce a sound at a given time. The sound sources are pairs of labia, which vibrate when air from the lungs is forced between them [6,7]. The modulated airflow is responsible for the pressure fluctuations which generate the sound. Each syllable is accompanied by a pulse of expiratory air pressure, and in between syllables, birds take short inspirations, mini-breaths (i.e., air sac pressure drops below atmospheric pressure) [8,9].

Each sound generator is controlled by a set of six syringeal muscles. In spontaneously singing brown thrashers (*Toxostoma rufum*), the role of different muscles was investigated with electromyographic techniques combined with simultaneous monitoring of airflow and air sac pressure [2]. The amplitude of EMG of the ventral syringeal muscle (vS) was highly correlated with the fundamental frequency of the vocalization. Based on this strong correlation and consistent with anatomical considerations this muscle is believed to control the tension of the oscillating labia.

Activity of the ventral and dorsal tracheobronchialis muscles was correlated with syringeal abduction (opening of airways) and adduction (closing of airways), respectively. The lateral part of the dorsal syringeal muscle also acts as an adductor of the lateral labium. The medial portion of the dorsal syringeal muscle most likely effects adduction of the medial labium [10], although no electromyographic data have been recorded from this muscle during spontaneous song.

Recently, a model for the physical processes involved in the production of sound by the syrinx was proposed [4]. This model contains parameters which can be qualitatively related to the activities of these muscles.

In this model, the labia are assumed to support both lateral oscillations and an upward propagating surface wave [11]. In this way, the opposing labia have a convergent profile when they move away from each other and a divergent profile when they move towards each other. This results in a higher pressure on the labia during the opening phase, and an overall gain in energy during each cycle of oscillation [3]. The reason is that while presenting a convergent profile, the average pressure between the labia is closer to the bronchial pressure, whereas interlabial pressure is closer to atmospheric pressure for a divergent profile. This results in a force in the same direction as the velocity of displacement of the labia, which might overcome the dissipation. The labia are also subjected to other forces: restitution (due to the elasticity of the tissues) and eventually a high dissipation if the labia displace too much from their equilibrium position. Labial oscillation can be interrupted if a constant force pushes the labia against each other (or against the walls). Adding up all these effects, a simple set of equations can be written to describe the motion of the variable describing the departure of the midpoint of a labium from the prephonatory position, x [3,4]:

$$x' = y, \tag{1}$$

$$y' = -\epsilon x - Cx^2 y + By - D_0 \tag{2}$$

where ϵ stands for the restitution constant of the labium, B is the total dissipation constant (constituted by both the negative dissipation induced by the interlabial pressure as discussed above, and the positive linear dissipation of the system), and D_0 measures a constant force (which displaces the equilibrium of the system). These parameters can be associated with the muscular activities mentioned above. The parameter C is the nonlinear dissipation constant. For large departures from the equilibrium position (i.e., large absolute values of x), the nonlinear dissipation term becomes very large. This models the effect of containing walls, as described in Ref. [4]. Throughout this work, we take $C = 2 \times 10^8 (1/\text{cm}^2 \text{ s})$.

The activity of the vS muscle is positively correlated with the frequency of the vocalization. Therefore, in the model it is associated with the parameter ϵ. It is likely that the longitudinal tension of the labia, caused by contraction of the vS muscle, alters their elasticity [see Fig. 1(a)]. Therefore, it is natural to associate it with the restitution constant in the model. Large values of D_0 represent active adduction, since it prevents oscillations by placing the membranes at their minimum distance. It is then related to the activity of the muscle tracheobronchialis dorsalis(dTB) or lateral part of the dS muscle. The details of the mechanics are not completely known, but the dTB presumably rotates the bronchial ring pushing the inner labium into the lumen as shown in Fig. 1(b).

With these elements in hand, previous work focused on reconstructing the instructions needed in order for this model to produce vocalizations of an appropriate acoustic structure. The typical feature to be reproduced in a simulation is the changing vibratory behavior over time within a syllable, i.e., the fundamental frequency of the labial oscillations. For example, "upsweeps" (downsweeps) denote syllables in which the oscillations are turned on at a lower (higher) frequency than that at which they are turned off. In the following section, we will use electromyographic data as slowly time dependent parameters of the model. Since these data were collected in singing birds, we can compare the synthetic vocalizations of the model to the actual vocalizations. The dynamical evolution of the fundamental frequency within a syllable is one of the most noticeable acoustic features of birdsong, and we concentrate on reproducing this feature. Yet, in order to produce synthetically realistic sounding solutions, it is important to model the filtering of the vocal tract. As in Ref. [4], we use a simple one-tube approximation of length $L = 1.9$ cm.

III. DRIVING THE MODEL WITH RECORDED MUSCLE ACTIVITIES

In the model discussed in the preceding section, the main acoustic parameters were correlated with the air sac pressure, the activity of the vS and the activity of the dTB. Therefore, in order to test the model we recorded these variables and
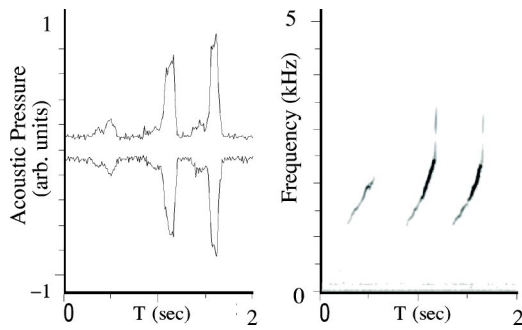
FIG. 2. Wave form (envelope) and sonogram of a song fragment produced by a cardinal (Cardinalis cardinalis). The grayscale in the sonogram represents the amplitude of the spectral peak, the darker one representing higher amplitude.

use them to drive the model and compare synthetic and re-corded songs. Electromyographic data were recorded using a procedure described in Ref. [2]. Briefly, we proceeded as follows. Air sac pressure was recorded with a flexible can-nula inserted into the anterior thoracic air sac and connected to a piezoresistive pressure transducer. In order to record muscle activities, wire electrodes were implanted into sy-ringeal muscles. The electrodes were prepared from insulated stainless steel wire and secured to the tissue with a micro-drop of tissue adhesive. Before closing the air sac, all the wires were led out and routed to the back. The data used in this work consist of 13 song elements with simultaneous recordings of dTB activity, vS activity, and air sac pressure (P). The data were taken from two cardinals (Cardinalis car-dinalis), hereafter called bird 1 and bird 2. In Fig. 2 we display the changes in sound pressure (envelope) over time as recorded with a microphone, together with the spectro-gram of this segment. The spectrogram shows the spectral content as a function of time for three syllables in which the fundamental frequency monotonically increases during vo-calization (upsweeps). Figure 3, on the other hand, shows the pressure and activities of vS and dTB during the same 2 s in which the vocalization was produced.

As can be seen in Figs. 3(b) and 3(c), the EMG recordings (measure of the electrical activity in the muscles) are spiky signals, which are difficult to interpret. A standard procedure to process these signals is to rectify the data and to extract the smoothed envelope [12]. The rectification (the negative values are converted to positive ones by taking the absolute value) eliminates the artificial biphasic nature of the record-ings, which result from the bipolar nature of the electrodes.

EMGs were sampled at 40 000 points per second. We took the absolute value and then replaced each point by the aver-age of its 2000 neighbors, between ±0.1 s. The rectified, smoothed envelope of the EMG provided a relative measure of the muscle activity. This could be used in the model to understand how muscles contribute to the movements re-sponsible for the slow evolution of the fundamental fre-quency within song syllables (see spectrogram in Fig. 2). In Fig. 4 we display the result of this smoothing procedure.

The relationship between the EMG signal and the tension generated in the muscle is likely to be a complex one. The EMG signals vary considerably from experiment to experi-
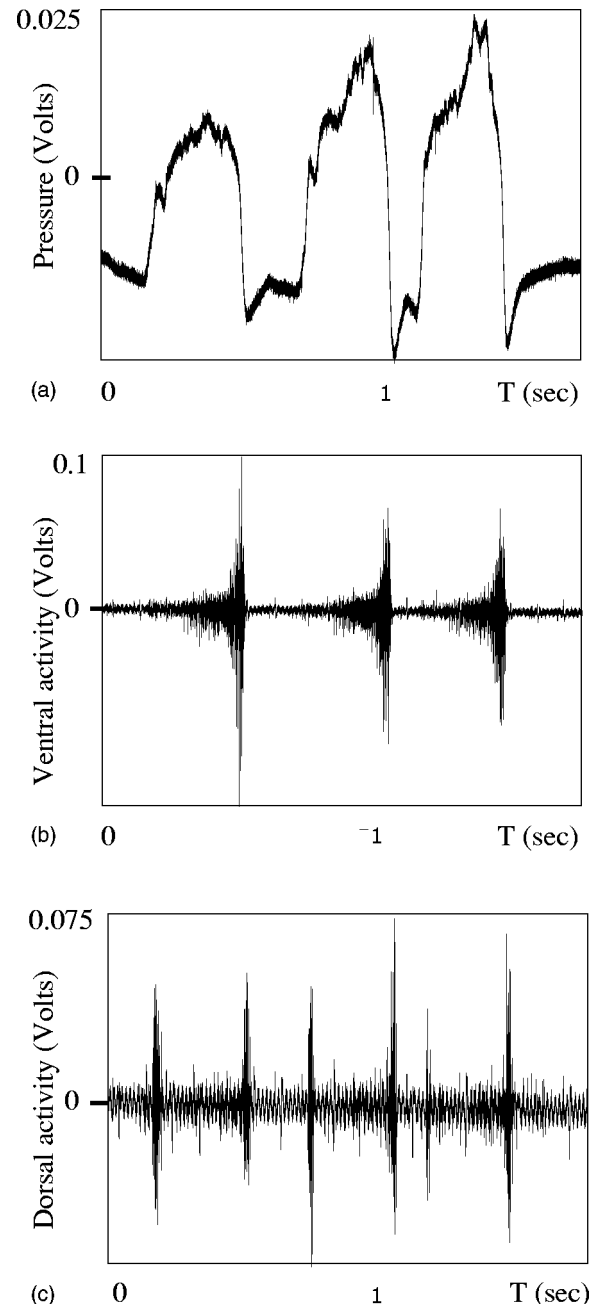

FIG. 3. (a) Pressure, (b) vS activity, and (c) dTB activity re-corded during the vocalization displayed in Fig. 1.

ment (even for the same subject), and therefore an absolute calibration is not possible. We assume that a monotonically increasing function relates the EMG linear envelope to the tension [12], which is sufficient to test the qualitative fea-tures of the song as predicted by the model.

We will now proceed to integrate the model for the dy-namics of the labia with the smoothed envelope of the ac-tivities of vS (K) and dTB (D) muscles, and the data (P). For simplicity, we assume a linear relationship between the smoothed envelope of the EMG recordings and the tension of the respective muscles, as well as a linear relationship between the air sac pressure and the driving average interla-bial pressure:
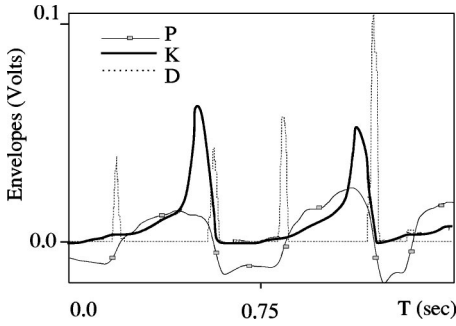
FIG. 4. Linear envelopes of the activities displayed in Fig. 3.

$$\epsilon = \epsilon_{1,1} + \epsilon_{1,2}K, \tag{3}$$

$$B = \beta_{1,1} + \beta_{1,2}P, \tag{4}$$

$$D_0 = \delta_1 D. \tag{5}$$

These set of linear relationships allow us to fit frequency ranges for the synthetic sounds generated by our model, when driven by ϵ, B, and D_0. The fitting was performed for each bird, since the amplitude of the EMG may vary between individuals depending on the input impedance of electrodes. For each bird, a syllable was used to find the coefficients. The coefficients in Eq. (3) were adjusted so that the frequencies at the beginning and end of the synthetic syllable would be those of the recorded one. The coefficients of Eq. (4) were chosen such that the synthetic and recorded syllable would start and end at the same time. The coefficient in Eq. (5) is such that D_0 has the minimum value at which an oscillation of the model can be prevented, when D reaches a local maximum, close to the start of the syllable [4]. For each bird in our study, the fitting obtained with one syllable was used for all utterances. For bird number 1, $\epsilon_{1,1} = 1.25 \times 10^8 (1/s^2)$, $\epsilon_{1,2} = 7.5 \times 10^9 (1/V \, s^2)$, $\beta_{1,1} = -2 \times 10^3 (1/s)$, $\beta_{1,2} = 5.3 \times 10^4 (1/V \, s)$, $\delta_1 = 15 \times 10^6 (cm/V \, s^2)$.

The results are displayed in Fig. 5, which shows spectrograms of four syllables sung by the bird [Figs. 5(a) and 5(b)] and generated by the model [Figs. 5(c) and 5(d)]. Figure 5(c) was obtained when the model was driven with the EMG data recorded, while bird 1 sang the segment displayed in Fig. 5(a). In Figs. 5(b) and 5(d) we performed the same analysis with a different segment vocalized by the same bird.
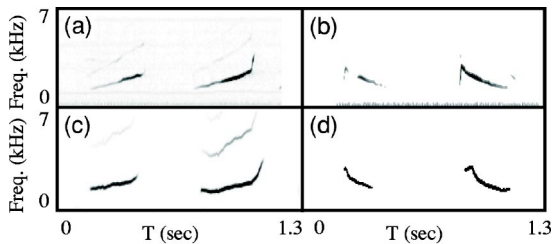

FIG. 5. The sonograms of recorded fragments of songs (a,b) and the sonograms of synthetic songs (c,d) generated when the model was driven with linear functions of the smoothed envelopes of the activities recorded during the vocalizations. Both vocalizations correspond to bird 1.
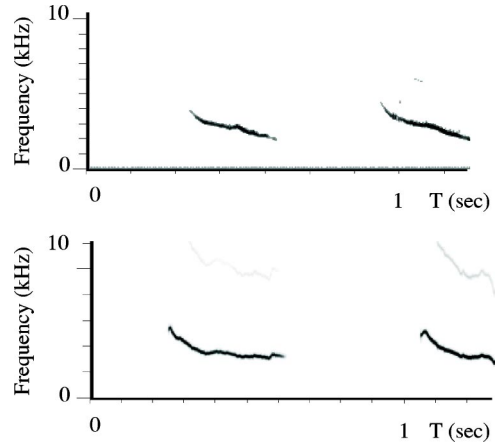

FIG. 6. The same as in Fig. 5, for bird 2.

Similar results are obtained if the same procedure is implemented with the data from bird 2. One example is displayed in Fig. 6. The window for the filtering of the data was the same as that used to process the data from the first bird. For bird number 2, we found $\epsilon_{1,1} = 1.25 \times 10^8 (1/s^2)$, $\epsilon_{1,2} = 12.5 \times 10^9 (1/V \, s^2)$, $\beta_{1,1} = -2 \times 10^3 (1/s)$, $\beta_{1,2} = 5.3 \times 10^4 (1/V \, s)$, $\delta_1 = 28 \times 10^6 (cm/V \, s^2)$.

For the two birds studied, fitting the coefficients with one syllable allows us to reproduce synthetically different syllable types vocalized by the same bird. Notice that the frequency range, as well as the time course of the fundamental frequency are remarkably similar. This indicates that the model is consistent with the data, and that the biological interpretation of its parameters is meaningful.

IV. A PREDICTION FROM BIFURCATION THEORY

In the data analyzed, the pressure grows rapidly as a syllable is vocalized. Therefore, increasing or decreasing frequencies in the sonogram are mainly controlled by increasing or decreasing activities of the vS muscle. For this reason, the qualitative agreement between synthetic and recorded songs does not require much from a model. Interestingly, the presence of the D_0 force allows us to make a precise prediction of the relative values of the gating muscle dTB at the onset and the end of a syllable. This term can cause the oscillations to stop by pushing the equilibrium position to a region of high dissipation (physically, it is responsible for the active adduction of the labia). As the value of D_0 is decreased, eventually the oscillations begin. In Ref. [4], the bifurcation diagram in terms of this parameter is discussed. This analysis indicates that the minimum value of D_0 needed to prevent oscillations is $D_0 = \epsilon \sqrt{B/C}$. In this way, the higher the ϵ (i.e., the higher the frequency of the oscillation being born) the higher the value of D_0. According to this observation, in an upsweep syllable, the activity of the dTB muscle controlling the precise onset of the oscillations should be smaller than the activity at the end, whereas this should be reversed for a frequency downsweep. This difference is caused by the fact that in an upsweep, the oscillations being prevented at the end are of higher frequency than those at the beginning, during the same range of pressure. Figure 7
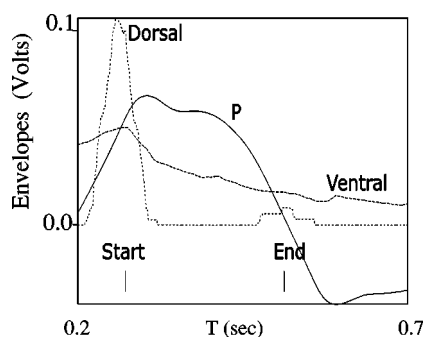
FIG. 7. The pressure and smoothed envelopes of smoothed, rectified dS and vS during a downsweep syllable.

shows this feature in two examples, although this prediction was verified for all the syllables of the analyzed song segments. In Fig. 8 we display the ratios between the maximum activities at the end and at the beginning of 17 syllables. The first six syllables are upsweeps, and the ratios are larger than one, while the rest of the syllables were downsweeps, with ratios smaller than one.

V. DISCUSSION AND CONCLUSIONS

The physical processes involved in birdsong are very rich and yet it is possible to build quite simple models to account for important acoustic features of the vocalizations. Since the oscillating behavior of labia was identified as the principal mechanism for the generation of periodic obstructions of the flow responsible for the sound, the modeling of the phenomenon in terms of simple oscillators is only natural. However, the interpretation of the model parameters and their relationship to the actual biological forces is not a trivial task, and requires experimental validation. In this report, we addressed this issue by integrating the simple model with experimental data as its time dependent parameters.

The comparison of the synthetic songs generated by driving the system with biological data to the song actually vocalized while the driving parameters were recorded, yields a qualitative agreement between synthetic and natural songs or
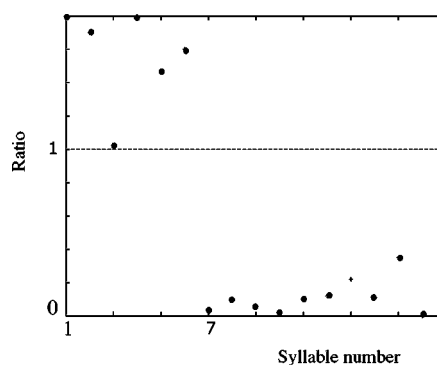

FIG. 8. The ratios between the local maximum activities of the dS muscle at the end and at the beginning of seventeen syllables from birds 1 and 2. The first six syllables are upsweeps, and the rest are downsweeps.

all inspected syllables. Moreover, specific predictions of the relative activities of dTB at the beginning and end of syllables were validated with experimental data. Since many species of birds require a tutor to learn their vocalizations, much effort is dedicated to understanding the representation of song at the level of the brain, as well as the neural architecture needed to generate the precise instructions involved in song production. By illuminating the mechanisms by which the activities of different syringeal muscles control the avian vocal organ for song production, we help to build the bridge between neural activity and the physical manifestation of the motor control, the vocal behavior. An implication of this model for neurophysiologists in the field is that in order to map the central neural activity onto peripheral motor action during song, it is necessary to relate neural activity with the proposed driving parameters, rather than the acoustic properties of the emitted sound [13].

[1] M.S. Brainard, A.J. Doupe, Nat. Hazards Rev. **1**, 31 (2000)

[2] F. Goller, R.A. Suthers, Neurophysiology **75**, 867 (1996)

[3] T. Gardner, G. Cecchi, M. Magnasco, R. Laje, and G.B. Mindlin, Phys. Rev. Lett. **87**, 208101 (2001)

[4] R. Laje, T. Garner, and G.B. Mindlin, Phys. Rev. E **65**, 051921 (2002)

[5] In the literature of birdsong, the minimal continuous sound is called a ''note,'' and syllables are repeated combinations of notes. In this work, all syllables consist of one continuous sound, and therefore we omit to use the term ''note,'' since in many other fields this term implies a constant fundamental frequency.

[6] C.K. Catchpole and P.J.B. Slater, *BirdSong, Biological Themes and Variations* (Cambridge University Press, Cambridge, 1995).

[7] F. Goller and O.N. Larsen, Proc. Natl. Acad. Sci. U.S.A. **94**, 14787 (1997)

[8] O.N. Larsen and F. Goller, Proc. R. Soc. London, Ser. B **266**, 1609 (1999)

[9] R.S. Hartley Respir. Physiol. **81**, 177 (1990)

[10] F. Goller and O.N. Larsen, J. Comp. Physiol., A **188**, 841 (2002).

[11] I.R. Titze, J. Acoust. Soc. Am. **83**, 1536 (1988).

[12] R.D. Keynes and D.J. Aidley, *Nerve and Muscle* (Cambridge University Press, Cambridge, 2001).

[13] R. Suthers and D. Margoliash, Curr. Opin. Neurobiol. **12**, 684 (2002)

[14] R. Suthers and F. Goller, in *Motor Correlates of Vocal Diversity in Songbirds*, edited by V. Nolan, Jr. *et al.*, Current Ornithology, Vol. 14 (Plenum Press, New York, 1997).