Featured in Physics

Unusual Avian Vocal Mechanism Facilitates Encoding of Body Size

Gonzalo Uribarri,[1] María José Rodríguez-Cajarville,[2] Pablo Luis Tubaro,[2] Franz Goller,[3] and Gabriel B. Mindlin[1,*]

[1]*IFIBA, CONICET and Departamento de Física, FCEyN, UBA, Buenos Aires 1428, Argentina*

[2]*División Ornitología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN-CONICET), Buenos Aires 1405, Argentina*

[3]*F.G. Institute of Zoophysiology, University of Münster, Münster 48143, Germany*

(Received 6 November 2019; accepted 3 February 2020; published 4 March 2020)

In this work we study the sound production mechanism of the raspy sounding song of the white-tipped plantcutter (*Phytotoma rutila*), a species with a most unusual vocalization. The biomechanics involved in the production of this song, and scaling arguments, allowed us to predict the precise way in which body size is encoded in its vocalizations. We tested this prediction through acoustic analysis of recorded songs, computational modeling of its unusual vocal strategy, and inspection of museum specimens captured across southeastern and south-central South America.

DOI: 10.1103/PhysRevLett.124.098101

Introduction.—Birdsong is one of the most exquisite forms of animal communication [1]. For approximately 50% of the close to 10 000 known bird species, there is a clear need of exposure to a tutor in order to learn proper conspecific songs. In the case of these species (known as vocal learners), complex and subtle motor instructions control the configuration of the vocal organ in order to produce rich and complex sounds. On the other hand, for those species outside the group for which there is strong evidence of vocal learning (known as nonlearners), a variety of strategies have evolved in order to enrich their vocal capacities. For example, tracheophones have developed a third sound source that enables them to add low frequency modulations to some vocalizations [2], something that learners achieve through subtle motor control of the muscles in charge of the vocal organ's configuration [3,4]. For this reason, the biomechanical strategies involved in the song production by nonlearners are particularly interesting.

In this Letter we analyze the song of the white-tipped plantcutter (*Phytotoma rutila*), a suboscine whose vocalizations do not present geographical dialects, and therefore it is presumably not a vocal learner. It is found widely in woodland and scrub of southeastern and south-central South America. The song of this species is a most singular vocalization, consisting of a long and rough creak, comparable to that of a rusty hinge (see the movie in Ref. [5]). We find that the dynamical ingredients necessary to reproduce its remarkable acoustic features include the interaction between a sound source that starts to oscillate in a saddle node in limit cycle bifurcation [6] and a filter provided by the vocal tract [constituted by the trachea and the oroesophageal cavity (OEC)].

Beyond reproducing the acoustic features of this species' song, the proposed dynamical scenario, together with scaling arguments, allowed us to predict the precise way in which body size is acoustically represented. This coding is an important aspect in the evolution of animal communication and has received much attention. Yet, for many species, the correlation between acoustic features and size tends to be weak and variable, in part due to the ability of animals to adjust filter properties dynamically [7–9]. Here we describe how the vibratory dynamics of the sound generating structure, and resonance properties of the upper vocal tract of this species, lead to a very clear and reliable indication of body size.

Song of Phytotoma rutila.—In Fig. 1 we display the sound trace of a song segment (approximately 0.2 s). The time waveform consists of a sequence of discrete sound elements that follow each other at rates ranging from 50 to 100 Hz at the beginning of the song to 300 Hz. In the second panel of Fig. 1, we display 12 ms of the waveform to illustrate two elements in detail. Each of these can now be described as a damped, fast oscillation of approximately 4 kHz.

A sound with a 4 kHz spectral component, slowly modulated at about 300 Hz, can be achieved by different mechanisms already described in the birdsong literature. Typically, the fast oscillations in birdsong production are generated when airflow passes through pairs of elastic labia located at the juncture between the bronchi and the trachea. These are flow-induced oscillations, whose frequency can be controlled through changes in the configuration of the avian vocal organ that affect the tension of the elastic labia [4,10]. In this paradigm, a low frequency modulation can be the result of an active muscle instruction, since the syrinx contains fast muscle fibers [11], which can contract at a rate of almost 300 Hz. Within this picture, the role of the vocal tract is to enhance the fast oscillations.

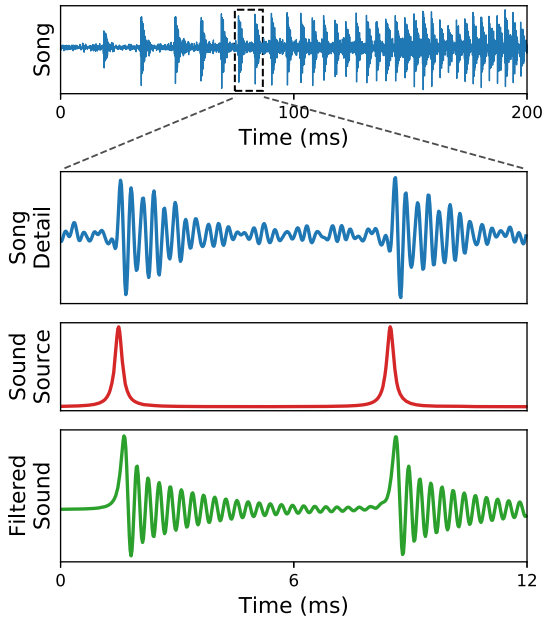

FIG. 1. The song and a synthesis by our dynamical model. In the first panel, we show the waveform of the song of *Phytotoma rutila*. In the second panel, we display a detail of two sound segments. A rapidly decaying fast oscillation is compatible with a pulse being filtered by a dissipative resonator. A pulse, as synthesized by our dynamical model, is shown in the third panel. Time traces representing pulses, filtered by a damped oscillator of the appropriate resonant frequency, are shown in the bottom panel. In this simulation, $(\gamma_1, \gamma_2, r, \omega_r) = (9\times10^3, 18.0, 0.3, 148.2)$. For the simulation displayed in the bottom panel, $\mu \sim 1.005$.

The scenario described above is not the only one compatible with low-rate modulated, high frequency sounds. The onset of the labial oscillations could be a saddle node in limit cycle bifurcation [6,12]. For parameters close to this bifurcation, the oscillations are explosive. In fact, at the bifurcation they are born with zero frequency, and a large spectral content. In this scenario, the airflow pulses are generated as the labia explosively open, exciting the rest of the vocal tract. In our problem, there is an element that strongly suggests that this second mechanism is responsible for the generation of these rough sounds: the amplitude decay of the fast oscillations in each sound element (see second panel of Fig. 1). This rapidly decaying fast oscillation is compatible with a pulse being filtered by a dissipative resonator. In fact, the fast frequency in the waveform is within the expected range of resonance frequencies of the OEC for a bird of the size of *Phytotoma rutila* (about 4 kHz). For this reason, we propose a simple model where pulsatile oscillations are generated by the labia at a slow rate (50–300 Hz), and are subsequently filtered by the OEC, acquiring its high frequency component of approximately 4 kHz.

Model.—Mathematically, a pulse can be described in terms of a variable that spends a long time close to zero, rapidly peaks to reach a nonzero value, and finally returns to the vicinity of zero, to start the cycle again. In terms of our physical problem, when the labia are pressed against each other, the airflow is interrupted. If the air sac pressure is continuously increased, the air makes its way through the lumen and an airflow pulse is generated. Then, Bernoulli forces are responsible for the lumen closing again, and the cycle repeats itself.

A simple mathematical model displaying this kind of pulse behavior can be written in terms of a phase variable θ, describing oscillations in a phase space $(x, \dot{x})$, with x parametrizing the labial separation. Defining $x \equiv 1 - \sin(\theta)$, $x = 0$ corresponds to a closed lumen. The model reads:

$$\dot{\theta} = \gamma_1[\mu - \sin(\theta)],$$

with γ_1 a temporal scaling parameter and $\mu \geq 1$ a bifurcation parameter proportional to the air sac pressure ($\mu = 1$ corresponds to a pressure for which air does not manage to pass between the labia) [12]. Since the lumen's area is proportional to x, the rate of mass injection at the base of the trachea is proportional to $\dot{x}$ [13] and acts as a source of pressure fluctuations at the base of the trachea P_b. This pressure can therefore be computed, up to a scaling factor, as

$$P_b(t) = \dot{x}(t) - rP_b(t - 2T),$$

with r the reflection coefficient at the juncture between the trachea and the OEC. The constant T stands for the time that it takes the sound to travel from the base of the trachea to the OEC. Finally, the transmitted component of the pressure wave, $P_T = (1 - r)P_b(t - T)$, excites the OEC. Modeling this cavity as a Helmholtz resonator, the inward displacement z of the gas obeys [14,15]

$$\frac{dz}{dt} = y, \qquad \frac{dy}{dt} = -\gamma_2^2\omega_r^2 z - 60\gamma_2 y - \gamma_2^2 P_T,$$

with γ_2 a temporal scaling parameter. In this way, $z = z(t)$ constitutes the synthetic waveform generated by our model.

To qualitatively test the output of our model, we compared the synthetically generated pulses with actual song recordings. The synthetic pulses were generated to match the real ones by selecting appropriate γ_1, γ_2, r, ω_r, and μ parameters (see third panel in Fig. 1 and the caption for the values used in the simulations). We present the comparison between the model generated sounds and the real ones in the fourth panel of Fig. 1.

Quantifying the decay rates and filter frequencies for individuals.—In order to test our hypothesis of sound production, we analyzed song recordings from xeno-canto and the Macaulay Library at the Cornell Lab of Ornithology. Songs were included in this analysis if (1) the signal-to-noise ratio was sufficiently high to allow proper analysis, (2) the song was not substantially contaminated by songs of other bird species in the background, (3) sound intensity did not present an abrupt change during the song, and (4) there were

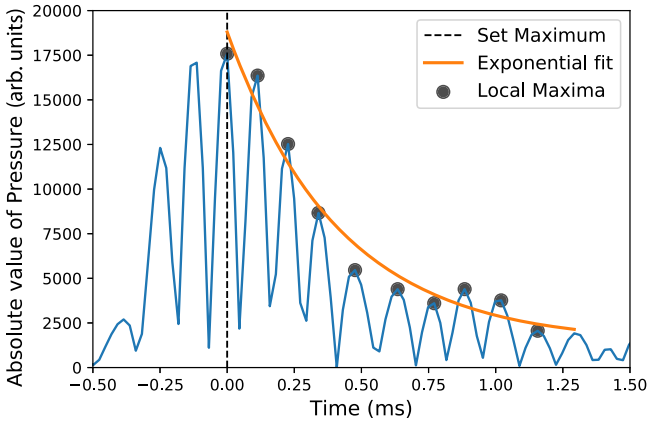

FIG. 2. Quantification of the decay time and frequency of the filter. We display a segment of the sound pressure time trace, after computing its absolute value. The maxima of ten consecutive peaks were computed. We fitted the points with a decreasing exponential function. The temporal differences between consecutive peaks were used to estimate the resonant frequency of the filter. The decaying amplitude of the peaks is attributed to the dissipative nature of the filter.

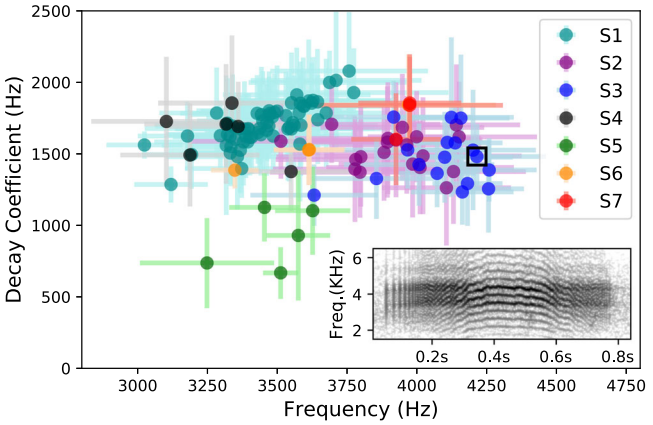

FIG. 3. Computed values of the decay times and filter frequencies for the songs in our study. Notice the remarkably low variation of the resonant frequency for each song. A point represents a song, with its color indicating the singing subject. Each song consists of a sequence of segments, and from each one a set of decaying amplitude peaks can be extracted. Point location indicates the mean decay coefficient and filter frequency for the whole song, and the sizes of the bars indicate their (2σ) dispersions. The inset shows the sonogram of a song. The black square indicates the selected song.

at least three songs per subject. A total of 123 songs of 7 different birds successfully passed the criteria.

We calculated the absolute value of the sound wave of each song and segmented it in sets of high frequency peaks of decreasing amplitude. The first 10 maxima in each set were fitted with a decreasing exponential function. This procedure is illustrated in Fig. 2. The average goodness of fit for each bird, measured in terms of r^2, ranged from 0.5 to 0.82, indicating that the exponential function successfully captures most of the functional form of the peak decay. In order to estimate the frequency of the filter, we computed the inverse of twice the time difference between the 10 local maxima following the absolute maxima for each set. Then we analyzed these estimated frequencies for all the sets in each song. In Fig. 3 we show the mean frequency and mean decay coefficient for every song in our study. Subject identity is indicated by color. The relative standard deviation of the frequency (average over all the birds and all the songs, ~6.5%) indicates that the filtering cavity remains stationary during the song. Similar numerical values are obtained considering all the pulses in all the songs together for each bird (average over all the birds ~6.9%). As an example, the inset in Fig. 3 displays the sonogram of a representative song, where it can be seen that the emphasized spectral band around 4 kHz remains unchanged.

High frequency is an indicator of body size.—The frequency with highest energy in these time traces is the resonant frequency of a resonator, whose dimension reflects the bird's size. Approximating the OEC as a Helmholtz resonator, we can write its resonant frequency as

$$\omega = v_{\text{sound}}\sqrt{\frac{A}{LV}},$$

where v_{sound} stands for the sound velocity and V, A, and L stand for the volume of the OEC, the area and length of the resonator's neck, respectively. In this way, if the characteristic lengths of the animal are slightly varied through a factor $\lambda \sim (1+\epsilon)$, then the frequency of this resonance will change:

$$\omega \to \omega\sqrt{\frac{\lambda^2}{\lambda\lambda^3}} = \frac{\omega}{\lambda} \sim (1-\epsilon)\omega,$$

And, therefore,

$$\frac{\Delta\omega}{\omega} \sim -\epsilon.$$

This approximation indicates that a change of ϵ percent in the length scale will be reflected as a ϵ percent change in the resonant frequency.

Since we do not have size values for the birds whose songs we analyzed, we performed an indirect test. This species is broadly distributed across southeastern and south-central South America, from La Pampa, Argentina, to Bolivia. Therefore, we could test whether body size increases with increasing altitude, a hypothesis that is consistent with Bergmann's rule [16,17], which states that populations of larger size are found in colder environments. We then analyzed tarsus length in museum specimens (a conventional indicator of linear body size), and found that it increased markedly with increasing altitude (Fig. 4) as well. For the birds in the dataset captured in a range between 0

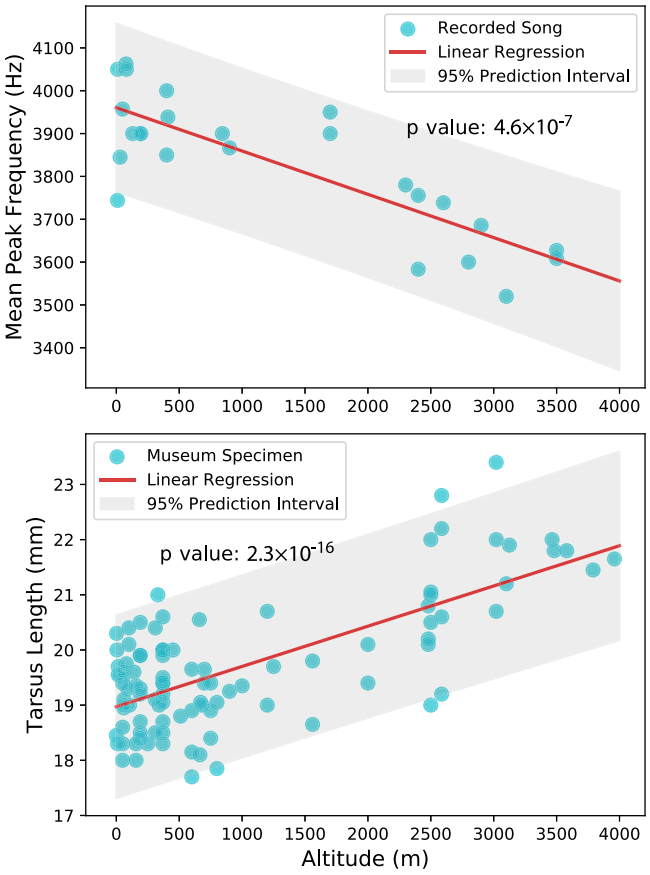

FIG. 4. Testing the predictions of the model. Mean peak frequency as a function of the altitude (top), and tarsus length as a function of the altitude (bottom). The regressions were computed using PYTHON's library "Statsmodels." The p values of the regressions are listed in the insets, and the 95% prediction intervals are indicated as the gray bands' widths along the fitted linear functions. The databases are described in the text.

and 3500 m of altitude, we got a 12.9% change in body size (linear regression: slope = 0.0007, $r^2 = 0.54$, $p < 0.0001$). Then, we analyzed the peak frequency from the song recordings from the xeno-canto and Cornell libraries for which the altitude could be determined either from the recorded metadata or by approximation using Google Maps. For the relationship between peak frequency determined from pulse oscillations and altitude, we also found a correlation between frequency and altitude (linear regression: slope = −0.101, $r^2 = 0.67$, $p < 0.0001$). The range between 0 and 3500 m yields a 13.5% change in resonant frequency. The similarity between the changes is consistent with the prediction derived from our scaling arguments.

Discussion.—For bird species that do not show clear evidence of vocal learning, there is a wide variety of complex biomechanical and dynamical mechanisms that allow generating rich vocalizations. We found that the most singular, raspy song of the white-tipped plantcutter is generated by a pulse-tone mechanism of unusually low frequency, and that this sound is filtered by much higher resonances of a stationary upper vocal tract filter.

Encoding information in signals and its reliability are important aspects of the evolution of communication behavior [18–20]. Many acoustic signals such as birdsong are used in reproductive context, either in male-male contest or male-female communication, and are thought to contain information on the sender's fitness. Because body size may inform receivers about fighting potential or general fitness, it is important to unveil how body size is encoded in acoustic signals. To be effective, the encoding of body size has to be reliable and has to be present, even if other aspects of vocal communication change. Yet, due to the animals' ability to adjust filter properties of the upper vocal tract dynamically, the correlation between acoustic features and body size has remained elusive. The sound production mechanism that we report here for the raspy sounding song of the *Phytotoma rutila* consists of a pulsatile excitation of a static cavity. For this reason, we could expect a clear and reliable signature of body size in the peak frequency. We tested this hypothesis by measuring tarsus lengths in museum specimens and acoustic properties of songs from databases, allowing us to report that the unusual vocal strategy reported here facilitates the acoustic encoding of body size.

We gratefully acknowledge all sound recordists for use of their field recordings, Ana Amador for insightful comments, and Gustavo Sebastián Cabanne for data about white-tipped plantcutter body size. Grants from ANPCyT (agencia nacional de promoción científica y tecnológica) (Argentina) and UBACyT (Secretaría de Ciencia y Técnica, Universidad de Buenos Aires) are acknowledged. We thank Fabián Gabelli for the movie of the white-tipped plantcutter in Ref. [5].

*Corresponding author.
gabo@df.uba.ar

[1] C. K. Catchpole and P. J. B. Slater, *Bird Song: Biological Themes and Variations* (Cambridge University Press Cambridge, England, 2008), p. 335.
[2] S. M. Garcia, C. Kopuchian, G. B. Mindlin, M. J. Fuxjager, P. L. Tubaro, and F. Goller, Curr. Biol. **27**, 2677 (2017).
[3] F. Goller and T. Riede, J. Physiol. **107**, 230 (2013).
[4] T. Gardner, G. Cecchi, M. Magnasco, R. Laje, and G. B. Mindlin, Phys. Rev. Lett. **87**, 208101 (2001).
[5] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.098101 for a movie displaying a white-tipped plantcutter singing.
[6] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, Boca Raton, 2018).
[7] G. C. Cardoso, Behav. Ecol. **23**, 237 (2012).
[8] P. Linhart and R. Fuchs, Anim. Behav. **103**, 91 (2015).
[9] J. P. Liu, L. K. Ma, Z. Q. Zhang, D. H. Gu, J. J. Wang, J. J. Li, L. J. Gao, and J. H. Hou, Eur. Zool. J. **84**, 186 (2017).

[10] F. Goller and R. A. Suthers, J. Neurophysiol. **76**, 287 (1996).
[11] C. P. Elemans, A. F. Mead, L. C. Rome, and F. Goller, PLoS One **3**, e2581 (2008).
[12] A. Amador and G. B. Mindlin, Chaos **18**, 043123 (2008).
[13] G. B. Mindlin, Chaos **27**, 092101 (2017).
[14] T. Riede, R. A. Suthers, N. H. Fletcher, and W. E. Blevins, Proc. Natl. Acad. Sci. U.S.A. **103**, 5543 (2006).
[15] Y. S. Perl, E. M. Arneodo, A. Amador, F. Goller, and G. B. Mindlin, Phys. Rev. E **84**, 051909 (2011).
[16] M. J. Rodríguez-Cajaraville, L. Calderon, P. L. Tubaro, and G. S. Cabanne, J. Ornithol. **160**, 947 (2019).
[17] S. Meiri and T. Dayan, Journal of Biogeography **30**, 331 (2003).
[18] J. W. Bradbury and S. L. Veherencamp, *Principles of Animal Communication*, 2nd ed. (Sinauer Associates, Inc, Sunderland, MA, 2011).
[19] M. Lachmann, S. Szamado, and C. T. Bergstrom, Proc. Natl. Acad. Sci. U.S.A. **98**, 13189 (2001).
[20] W. A. Searcy and S. Nowicki, *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems* (Princeton University Press, Princeton and Oxford, 2005).