

Computational model for vocal tract dynamics in a suboscine bird

M. F. Assaneo and M. A. Trevisan

Laboratorio de Sistemas Dinámicos, Departamento de Física, FCEyN, Universidad de Buenos Aires, Ciudad Universitaria 1428EGA, Buenos Aires, Argentina

(Received 4 May 2010; revised manuscript received 28 July 2010; published 16 September 2010)

In a recent work, active use of the vocal tract has been reported for singing oscines. The reconfiguration of the vocal tract during song serves to match its resonances to the syringeal fundamental frequency, demonstrating a precise coordination of the two main pieces of the avian vocal system for songbirds characterized by tonal songs. In this work we investigated the Great Kiskadee (*Pitangus sulfuratus*), a suboscine bird whose calls display a rich harmonic content. Using a recently developed mathematical model for the syrinx and a mobile vocal tract, we set up a computational model that provides a plausible reconstruction of the vocal tract movement using a few spectral features taken from the utterances. Moreover, synthetic calls were generated using the articulated vocal tract that accounts for all the acoustical features observed experimentally.

DOI: [10.1103/PhysRevE.82.031906](https://doi.org/10.1103/PhysRevE.82.031906)

PACS number(s): 87.19.-j, 43.80.Ka

I. INTRODUCTION

In the already long and well-established literature of birdsong, attention has been focused on the generation of sound by the avian vocal organ. This organ, the *syrinx*, consists of tissue membranes that can be set into oscillatory motion by the passage of air flow. Important progress has been made by mathematical models of the syringeal dynamics, accumulating strong evidence that points toward the idea that complex sounds may be a consequence of the syringeal structure [1,2]. Such ideas may have consequences on diverse biological fields, especially on animal communication, where complex vocal patterns do play a major role as behavioral traits. However, little is known about the importance of a tunable vocal tract coupled to the bird syrinx in the generation of complex sounds. In the field of birdsong, this lack of attention could be due in part to the fact that many species are characterized by the tonality of their songs, with minimal, if any, spectral richness.

Interestingly, a recent work supported by x-ray cinematography showed a very precise tuning of the vocal tract to the rapid changes in the fundamental frequency in the song of the Northern Cardinal (*Cardinalis cardinalis*) [3,4]. Moreover, even for birds like the White-Throated Sparrow (*Zonotrichia albicollis*) whose song consists of a series of almost constant frequency notes, the oropharyngeal cavity and cranial end of the esophagus maintain a relatively constant volume, corresponding to a major resonance following the fundamental frequency [5].

In this work we analyze a particular sound produced by the Great Kiskadee (*Pitangus sulfuratus*) known as *call*. This short acoustic production is typically emitted by birds under some kind of stress, presumably carrying information to other members of the community. The case is interesting because, unlike many species belonging to the oscine suborder, the sonogram reveals that this suboscine produces sounds of rich spectral content. For us humans, the generation of voiced sounds rests essentially on the possibility of altering the vocal tract configuration, enhancing some frequencies out of a spectrally rich signal giving rise, for instance, to the different vowels. The question of whether or

not these birds use the vocal tract as an active filtering device to alter the rich signal produced by their syrinx will be addressed here.

This work is organized as follows. In Sec. II we present and characterize the calls produced by the Great Kiskadee. Section III is devoted to the models for song generation: we analyze a recently developed dynamical model for the syrinx [2,6] in combination with a model for a mobile vocal tract. In Sec. IV we set up a computational technique based on a genetic algorithm and a least-cost search algorithm in order to reconstruct the anatomical parameters of the moving vocal tract, using spectral features of the calls. Complete synthetic calls for this bird are shown in Sec. V, along with the conclusions of this work.

II. ACOUSTIC ANALYSIS OF THE CALLS

The Great Kiskadee is a medium size American bird characterized by its nearly fixed three syllable song (from which the onomatopoeia *kis-ka-dee*). These birds also produce other sounds, notably the calls, which are very simple stereotyped continuous sounds of about 0.7 s long, as the ones shown in Fig. 1. Calls are produced by the birds under stress, presumably as a sign of the proximity of danger. Speculatively, it could be advantageous for these calls to be constrained to specific spectral bands in order to unmask the uttered sounds from noisy environments, optimizing the transfer of acoustical energy, as observed in other biological systems [7,8]. In that case, it is expected that the vocal tract, considered as a spectral filtering device, would play a central role.

Inspired from these ideas, we created a database of 45 calls recorded in several sessions within 1 month of four wild-caught Great Kiskadees. The birds were kept in acoustical isolation while recorded with a TAKSTAR SGC 568 microphone and filmed with a Sony HDR-SR11/SR12 HD camera. In all cases we observed very stereotyped sounds, characterized by a fast increase in the fundamental frequency followed by a plateau and terminated by a rapid decrease in the fundamental frequency (see Fig. 1). We also performed direct anatomic measures on the anatomy on the birds, anes-

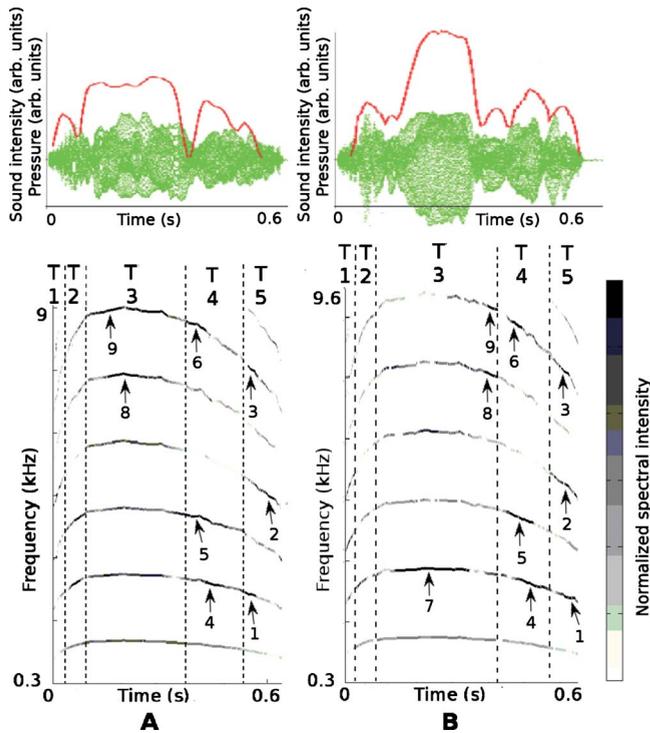


FIG. 1. (Color online) Sound time series envelope, sound intensity (solid line), and reassigned spectrograms of two different types of calls. (A) An example of a call that does not present a boost in the total sound intensity, observed in 20% of the analyzed calls (type A). (B) A call presenting an overshoot in total sound intensity, produced in roughly the 80% of the utterances of each bird (call of type B). Arrows indicate the emphasized frequencies (darker traces) found systematically in all the calls of our database [1: 2.58 kHz; 2: 5.08 kHz; 3: 7.50 kHz; 4: 2.83 kHz; 5: 4.39 kHz; 6: 8.43 kHz; 7: 3.20 kHz (present only in calls of type B); 8: 7.50 kHz; 9: 9.00 kHz]: ± 0.04 kHz. Other emphasized frequencies are found at the beginning of the calls, as a reflection of resonances 1–6 with respect to a vertical axis passing through the highest value of the fundamental frequency, and shrunk into the initial fast upsweep, and therefore not visible in the figure.

thetized with intramuscular injections of ketamine/xylazine (the protocol can be found elsewhere [2]).

The spectral representation shown in the lower panels of Fig. 1 is not the sonogram, which is the standard spectral tool in birdsong, but a representation described in [9,10] called *reassigned spectrogram*. Given the complex Gabor transform $\chi(t, \omega) = |\chi(t, \omega)| e^{i\phi(t, \omega)}$,

$$\chi(t, \omega) = \int x(\tau) e^{(t-\tau)^2/2\sigma^2} e^{i\omega(t-\tau)} d\tau,$$

the sonogram is computed using the modulus $|\chi(t, \omega)|$, while the reassigned spectrogram is derived from phase information, $\omega_{ins} = \partial\phi/\partial t$ and $t_{ins} = t - \partial\phi/\partial\omega$. Briefly, there exists a

transformation T , $(\omega, t) \rightarrow (\omega_{ins}, t_{ins})$, such that in the new variables a signal will have optimum precision and resolution in the spectral domain, in the sense that a given frequency trace is represented by a line of zero thickness (infi-

nite precision), and different traces will not interfere with each other if they are further apart than the Fourier uncertainty, which corresponds to the optimal spectral resolution. This representation allows an accurate detection of slight variations in spectral intensity. In fact, short dark traces can be identified within the harmonics of the call, marked by arrows in Fig. 1. These traces are much harder to visualize in the sonogram because of the finite precision of the fast Fourier transform.

Despite their stereotyped spectral shape, the calls of the birds can be easily classified in two subsets [Figs. 1(A) and 1(B)], depending on the presence or absence of an overshoot in the total sound intensity, roughly located in the middle of the call, as seen in the higher panels of Fig. 1. Each bird produces this overshoot in about 80% of its calls (that we labeled as type B) that also display an emphasized spectral intensity in the first harmonic (labeled 7 in Fig. 1), features that are absent in the other 20% of the calls (type A), characterized by a slightly lower fundamental frequency. In the next sections we will set up a minimal model of the avian vocal system that accounts for all the acoustic features found in the experimental records of the calls, allowing us to give a plausible explanation for the existence of both acoustic behaviors.

III. SYRINX AND VOCAL TRACT MODELS

The Great Kiskadee (*Pitangus sulfuratus*) presents a tracheosyringeal syrinx with two independently controlled sound generators, each of which consists of a pair of oscillating labia. One of the simplest models that captures the physical principle of energy transfer from the airflow to the tissue membranes is based on the observation of surface waves traveling upward through the labia during the oscillatory cycle, presenting a syrinx of convergent profile when the labia move away from each other and a syrinx of divergent profile in the closing semicycle. To account for these observations, an equation of motion for the midpoint displacement x of the labia was proposed in [2,6], which can be written as

$$m\ddot{x} = -kx - \beta\dot{x} + a_1 p_s \frac{\Delta + 2\tau\dot{x}}{a_{01} + x + \tau\dot{x}}, \quad (1)$$

where $k = k(x) = k_1 + k_2 x^2$ and $\beta = \beta(x, \dot{x}) = \beta_1 + \beta_2 x^2 + \beta_3 \dot{x}^2$ are first-order nonlinear elastic and dissipative coefficients, respectively. The last term is the intraglottal pressure acting on the labia that depends on the air sac pressure p_s and the profile of the syrinx, with a_1 as the labial area, a_{01} and Δ as geometric parameters of the syringeal profile, and τ as a characteristic time of the tissue wave [1,11]. The labial movement is the source of pressure perturbations at the input of the vocal tract, $p_i(x, \dot{x}) = \sqrt{p_s/\rho} x$ [2,12], with ρ as the air density.

In order to complete the sound production process, we are now ready to incorporate a computational vocal tract model to the output of the syrinx. Vocal tract effects can be of two types: *filtering* and *coupling* effects. In this work we will consider only the filtering effects, according to the source-filter theory [1,11].

From the theory of sound propagation in narrow tubes [13], a number of computational models of the vocal tract were studied in the literature [3,14,15]. The sound wave meeting the interface between tubes n and $n \pm 1$ will split up into a reflected and a transmitted wave, characterized by coefficients $r_{n,n\pm 1} = a_n/a_{n\pm 1}$ and $t_{n,n\pm 1} = 1 - r_{n,n\pm 1}$, respectively.

The sound resulting from the pressure traveling through a system of four tubes can be written as

$$\begin{aligned}
 a(t) &= p_i(t) + b_b(t - \tau_1), \\
 b_b(t) &= r_{1,2}a(t - \tau_1) + t_{2,1}c_b(t - \tau_2), \\
 b_f(t) &= t_{1,2}a(t - \tau_1) + r_{2,1}c_b(t - \tau_2), \\
 c_b(t) &= r_{2,3}b_f(t - \tau_2) + t_{3,2}d_b(t - \tau_3), \\
 c_f(t) &= t_{2,3}b_f(t - \tau_2) + r_{3,2}d_b(t - \tau_3), \\
 d_b(t) &= r_{3,4}c_f(t - \tau_3) + t_{4,3}e_b(t - \tau_4), \\
 d_f(t) &= t_{3,4}c_f(t - \tau_3) + r_{4,3}e_b(t - \tau_4), \\
 e_b(t) &= r_e d_f(t - \tau_4), \tag{2}
 \end{aligned}$$

where τ_n is the time it takes to the sound to travel along the n th tube at speed c , $\tau_n = l_n/c$, and r_e is the reflection coefficient at the interface with the atmosphere. When supplied with a time series of air sac pressure p_s and time varying parameters a_n and l_n for the vocal tract, the model of Eqs. (1) and (2) is capable of generating synthetic sounds proportional to $d_f(t)$.

From the complete mathematical model for sound production and the analysis of the reassigned spectra calculated from the calls, we will assume the following:

(1) The frequencies at which a harmonic displays sudden short traces of high spectral intensity will be considered resonances of the vocal tract. This assumption is supported by numerical simulations of the syrinx model. In a recent work, the syrinx model was used to successfully synthesize bird sounds when driven by experimental records of air sac pressure p_s recorded from singing Kiskadees [2,6]. Driving Eq. (1) with a varying air sac pressure p_s , we found that higher values of p_s enhance the spectral intensity of every harmonic of the acoustic pressure p_i . Moreover, the relative intensities of the different spectral components of p_i hold almost invariant for different values of p_s . According to this, variations of air sac pressure affect the complete set of harmonics, enhancing or attenuating them as a whole. Therefore, when a single harmonic suddenly increases its spectral intensity, we assume that it coincides with a resonance of the vocal tract. The same applies to a subset of the harmonics (as indicated, for instance, by arrows 4–6 in the time frame T_4 ; Fig. 1).

(2) The Great Kiskadee articulates the vocal tract while uttering the calls. The resonant frequencies of a narrow continuously deformed open-closed tube are determined by its geometrical properties. In particular, the maximum number of resonances can be estimated using a uniform tube of the same length, while the specific resonant values are deter-

mined by its cross-section departures from the uniform tube [14,15].

The simplest model for a vocal tract is then a uniform tube of length L with closed-open endings corresponding to the syringeal (glottal) exit and the beak (mouth) for birds (humans). Such a quarter-wave tube presents resonances at $f_k = (2k-1)c/4L$ ($k \in \mathbb{N}_{\geq 1}$), with $c \sim 340 \times 10^3$ mm/s, the speed of sound.

Using $L \sim 90$ mm, we can approximate the number of resonances of a vocal tract for a medium size bird like the one examined here. We get $f_k \sim (2k-1) \times 0.94$ kHz. Therefore, within the bandwidth $0 \leq f \leq 10$ kHz, we expect to find at most five resonances ($f_6 \sim 10.39$ kHz), independently from the specific shape of the vocal tract. Numerical simulations using $n=2$ to $n=20$ tubes of different shapes and a total length of $L \sim 90$ mm display three to five resonances in that bandwidth.

However, in Fig. 1 we recognize eight (nine) resonances for calls of type A (type B), as we would observe in much larger vocal tracts of around 180 mm long, like the human [14,15]. We interpret this as a signature of a moving vocal tract such that, for a given time frame, a maximum of five resonances will be found, corresponding to a specific vocal tract configuration of $L \sim 90$ mm. Following Fig. 1, we consider a time interval including resonances 7–9 for the higher fundamental frequencies (T_3), then another interval associated with resonances 4–6 (T_4), and a final interval that includes resonances 1–3 (T_5).

In addition to the resonances shown in Fig. 1, the reassigned spectrograms display six more resonances that are not marked in Fig. 1. These resonances correspond to a symmetrical reflection of the resonances 1–6 with respect to a vertical axis located at the higher fundamental frequency value and concentrated at the beginning of the call. Therefore, we will associate the complete set of resonances of the type B call to five time intervals: $T_1 \rightarrow (1,2,3)$, $T_2 \rightarrow (4,5,6)$, $T_3 \rightarrow (7,8,9)$, and again $T_4 \rightarrow (4,5,6)$ and $T_5 \rightarrow (1,2,3)$. In the next section, we set up an algorithm that allows us to extract vocal tract parameters a_n and l_n from the experimental sound recordings.

IV. PARAMETER FITTING

Now we face the following problem. Take, for instance, the call of type A, whose fundamental frequency is slightly lower than the fundamental frequency of the call of type B. Note that, if the resonance marked as 7 in the call of type B was present in the call of type A, it would lie outside of the frequency range of its first harmonic and, consequently, we cannot conclude that this resonance (or others in between the harmonics) are absent from the vocal tract. Therefore, due to this lack of spectral information, many different tube configurations, as calculated from Eq. (2), will succeed in matching the experimental spectral resonances. Despite the apparent contradiction, this overexpression of tube configurations for each time frame can be useful in the reconstruction of the moving vocal tract, provided we add the biologically reasonable constraint of “minimal anatomical effort,” i.e., by minimizing the variations of the vocal tract shape along the call.

From a computational point of view, both the problem of generating a population of compatible vocal tracts for each frame and the problem of finding the lower anatomical effort between them can be tackled. To solve the first one, we set up a genetic algorithm to explore the vocal tract configurations.

A genetic algorithm is an optimization procedure inspired by biological evolution. A caricature of natural selection consists of making the most adapted individuals of a species prevail in reproduction, generating offspring that mixes the genetic information of their parents and changes it at random (using the genetic operators of crossover and mutation, respectively). This double mechanism that privileges good genetic information and explores new possibilities in reproduction is a very efficient way to search the genetic space for the best adapted individuals [16]. This caricature can be used to find the parameters of a mathematical model that best fit a desired output. To do so, a *fitness* is assigned to every output of the model as a positive real number quantifying its similarity with the desired solution. The algorithm then starts with a generation of parameters chosen at random, generating a population of outputs. The next step is to choose pairs of outputs with a probability proportional to their fitness and to apply the genetic operators of crossover and mutation in the *genetic space*. In this way, a new population is generated (offspring) that will be used as the new input to the algorithm, restarting the process until the distance between outputs and the solution reach some desired threshold.

In our specific problem, a point in the parameter space $\{L, A\}$ of vocal tract lengths $L=(l_1, l_2, \dots, l_n)$ and cross sections $A=(a_1, a_2, \dots, a_n)$ is used to generate a sound time series from Eq. (2). We calculate the fast Fourier transform and compute its numerical resonant frequencies. The corresponding fitness is inversely proportional to the distance (in Hz) between the numerical and the experimental resonant frequencies. In order to apply the genetic operators, the parameters $\{L, A\}$ are coded in a genetic space formed by a string of numbers that result from normalizing the parameter values with respect to their corresponding ranges, keeping the first four decimals and putting them in a row. After a pair of outputs is selected, the genetic operators of crossing over (interchange of genetic information from the two parents with respect to a random point in the string) and mutation (random change of any number from the string) are applied, generating a new generation and repeating the process for 30 generations (more details are in the caption of Fig. 2).

We ran our genetic algorithm in search for a configuration of tubes displaying the complete set of the nine experimental resonances indicated by the arrows in Fig. 1. We used $n=2$ to $n=20$ tubes, finding only configurations for total lengths $L \geq 180$ mm, i.e., twice as long as the estimated vocal tract for these birds.

On the other hand, vocal tracts fitting the resonant frequencies within each frame (for instance, frequencies pointed by arrows 7–9 for frame T_3 ; Fig. 1) converged to total lengths $L \leq 110$ mm, consistent with the estimated length for the bird’s vocal tract. We then ran simulations for $2 \leq n \leq 20$ tubes and total lengths $70 \leq L \leq 110$ mm. In particular, the vocal tracts obtained with $n=2$ tubes failed to match the experimental frequency values, while tracts with $n \geq 3$ tubes

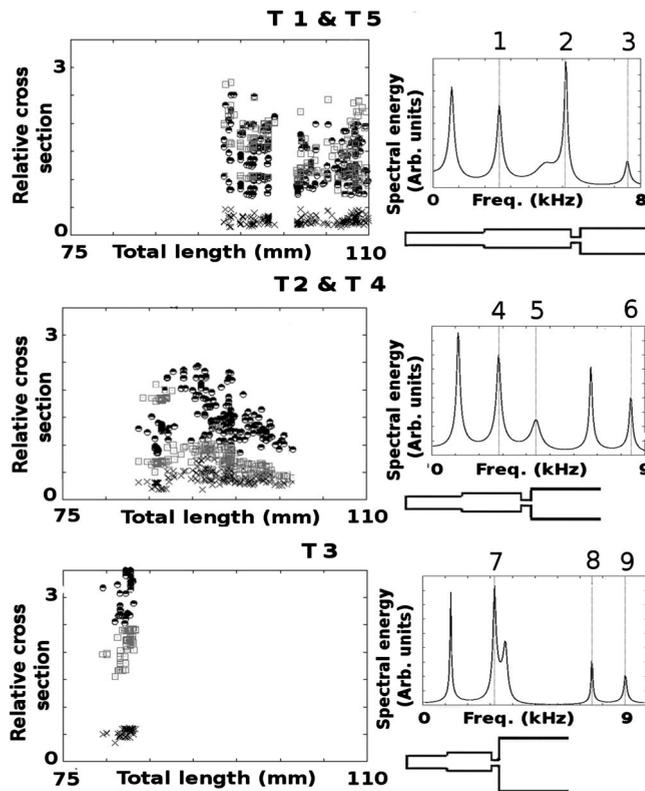


FIG. 2. Left: Vocal tract candidates projected in a two-dimensional parameter space, as retrieved by the genetic algorithm for the different frames. Circles correspond to $(\sum_{i=1}^4 l_i, a_4/a_1)$, crosses correspond to $(\sum_{i=1}^4 l_i, a_3/a_1)$, and squares correspond to $(\sum_{i=1}^4 l_i, a_2/a_1)$. We set the genetic algorithm to generate populations of 2000 vocal tract candidates, keeping the ones whose spectral resonances are less than 40 Hz away from the experimental resonances. We penalized tube configurations displaying resonances coinciding with low spectral intensity harmonics in the reassigned spectrogram. We used crossing-over and mutation rates of 0.85 and 0.1, respectively, and ran the algorithm for about 30 generations. Right: sketch and spectra of the final vocal tract configurations for each time frame. The specific tube parameters $\{l_1; l_2; l_3; l_4 | a_1; a_2; a_3; a_4\}$ are $T_1(T_5) \rightarrow \{35.2; 36.0; 3.7; 23.0 | 1.0; 2.1; 0.4; 1.6\}$, $T_2(T_4) \rightarrow \{34.3; 25.1; 3.7; 23.6 | 1.0; 2.0; 0.6; 2.0\}$, and $T_3 \rightarrow \{25.4; 24.0; 3.7; 29.6 | 1.0; 2.2; 0.6; 3.6\}$. Lengths are expressed in mm and relative areas are normalized with respect to a_1 .

succeeded. Moreover, vocal tracts with $n > 4$ tubes were obtained at higher computational cost and represent just small corrections to the overall shapes shown in Fig. 2, for $n=4$ tubes. In view of our numerical results, in this work we approximate the vocal tract by a series of $n=4$ tubes of variable cross sections a_i and lengths l_i ($1 \leq i \leq 4$). This choice is compatible with the anatomy of other bird species. In particular, the literature reports different articulated parts for the oscine vocal tract: trachea, glottis, oropharyngeal cavity, and beak [3,4].

The main advantage of using a genetic algorithm in our case is that it naturally generates a number of approximate solutions [16,17], therefore generating a population of vocal tract candidates for each frame. In Fig. 2 (left) we show the best solutions found by the genetic algorithm for vocal tracts

modeled by $n=4$ tubes, in the two-dimensional space $(\sum l_i, a_j/a_1)$ for the different frames. The solutions are clearly clustered for T_3 , showing vocal tracts of $L \sim 82.5$ mm long with very stereotyped shapes: a second tube wider than the first one (squares), then a constriction (crosses), and finally a wide tube (circles). Solutions for the other time frames (T_1 , T_2 , T_4 , and T_5) approximately follow this shape, presenting wider basins of attraction.

In this way, the genetic algorithm provides us with a pool of vocal tract candidates for each frame, such that choosing a candidate for each time frame $\{(L,A)^{T_1}, \dots, (L,A)^{T_5}\}$ will define an anatomical path for the vocal tract. Let us define a transition cost C from frame T_i to T_{i+1} as

$$C(T_i, T_{i+1}) = \sum_{n=1}^4 \frac{|a_n^{T_i} - a_n^{T_{i+1}}|}{a_n^{T_i}} + \frac{|l_n^{T_i} - l_n^{T_{i+1}}|}{l_n^{T_i}}, \quad (3)$$

and use it to find the globally lower cost path by means of dynamical programming, i.e., using a Viterbi algorithm [18].

In the right panels of Fig. 2 we show the final spectra and tube configurations as retrieved by our computational model for the different time frames of the type B call. Notably, the final configurations can be interpreted as variations of a unique configuration: a lower vocal tract, composed of the first two tubes (squares in Fig. 2), and an upper vocal tract (circles in Fig. 2), separated by a constriction (crosses in Fig. 2).

From videos of the birds uttering calls from a lateral view, we can easily identify a repetitive gesture: right after the vocalization begins, the bird retracts its head while opening the beak, maintaining this gesture before inverting the movement to its original position at the end of the utterance [19]. We performed a video analysis with ImageJ [20], separating the video of a call in 28 frames from where two anatomical measures were extracted: one associated with the length of the neck (D_1) and the second with the beak opening (D_2) (Fig. 3, upper panel). A qualitative correlation can be established between the anatomical data and the dynamics resulting from our vocal tract model. In particular, D_1 is compatible with the time evolution of the lower vocal tract, and D_2 is compatible with the dynamics of the cross section of the upper vocal tract (Fig. 3, lower right inset).

V. SIMULATIONS AND DISCUSSION

Beyond the qualitative agreement between the reconstructed vocal tract and the anatomical measures and the observed bird's movement, we used our results to synthesize calls of the Kiskadee. In a recent work, song and air sac pressure p_s time series were measured simultaneously from singing Kiskadees [2]. The pressure was found to be a smooth linear function of the fundamental frequency of the call. In order to test our model, we constructed a smooth function $p_{sB}(t) = \alpha F_B(t)$, proportional to the fundamental frequency $F_B(t)$ of the call of type B, such that, when feeding Eq. (1) with $p_{sB}(t)$ we obtained a sound compatible with the experimental records.

Notably, the vocal tract movement proposed in this work provides a simple explanation for both call types, depending

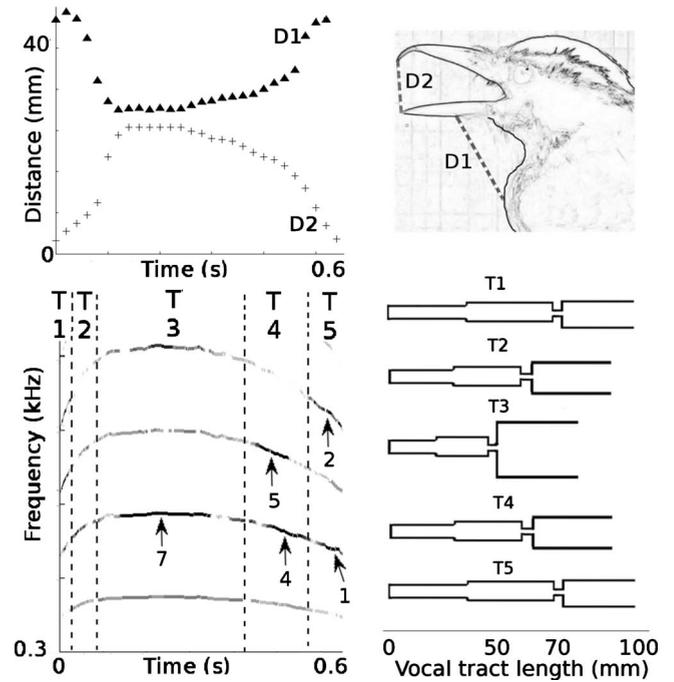


FIG. 3. Upper panels: time evolution of a measure associated with the neck extension D_1 and beak opening D_2 , as indicated in the lateral view (right inset). In particular, D_1 is the distance from the ventral end of the beak-skull transition to the keel and D_2 is the beak gape, calculated for 28 video frames. Lower panels: reassigned spectrogram of a call indicating the five time frames (left) corresponding to our model of the vocal tract configurations (right). Compatible with the evolution of the lower vocal for our model, direct measures of the trachea (an extensible tube located right beneath the bird skin), about 60 mm long for anesthetized birds, and variations of D_1 of about 20 mm are compatible with our computational results.

on whether or not the fundamental frequency of the call is high enough to tune its first harmonic to the second resonance of the vocal tract: reassigned spectra of the synthetic calls present all the spectral features observed in the experimental records for both air sac pressure functions $p_{sA}(t) = \alpha F_A(t)$ and $p_{sB}(t) = \alpha F_B(t)$, corresponding to calls of types A and B respectively, as shown in Fig. 4.

In analogy with recent works that revealed a precise tuning of the vocal tract to the fundamental frequency in oscines [3,4], our analysis suggests that also the suboscine Great Kiskadee coordinates the sound source with the vocal tract, tending to utter sounds of high intensity by making the first harmonic of the sound coincide with the second resonance of the vocal tract (calls of type B). According to our results, the bird's strategy would consist of raising the fundamental frequency by increasing the air sac pressure, and setting the vocal tract cavities to track the syringeal first harmonic by shortening the lower vocal tract and expanding the upper vocal tract area, increasing the second resonance. Whenever the first harmonic and the second vocal tract resonance coincide, a boost in the total sound intensity is produced, particularly around the band of 3200 Hz. On the other hand, if the bird fails to match these two frequencies, the calls

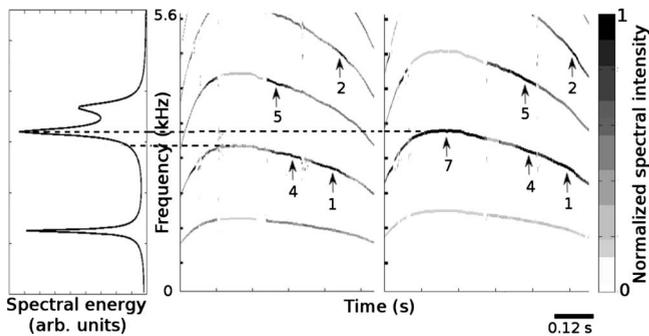


FIG. 4. Left: detail of the synthetic spectrum corresponding to T_3 . Right: reassigned spectrograms of the sounds synthesized with the model for air sac pressures p_{SA} (left) and p_{SB} (right). Pressure p_{SB} is enough to get the first harmonic to fall into the second vocal tract resonance at T_3 , boosting the total sound intensity of the call. The slightly lower air sac pressure p_{SA} is not enough to reach this second resonance, therefore losing the boost in total sound intensity, as in the recorded calls of type A.

produce lower sound intensities, which happens—to the degree of our exploration—for around 20% of the calls.

In this work we presented a simple computational algorithm that makes the most of few spectral features available from sound recordings, giving a plausible reconstruction of the movement of vocal tract during the emission of sound by a suboscine bird. If further validated, the picture that emerges from our computational model shows a strong coordination between the two main pieces of the avian vocal system in a suboscine bird.

ACKNOWLEDGMENTS

We wish to thank Gabriel Mindlin for enlightening conversations and Ana Amador for her kind help during the manipulation and care of the Kiskadees. This work was partially funded by the University of Buenos Aires, CONICET, ANPCyT, and NIH (through Grants No. R01 DC-06876 and No. R01 DC-04390).

-
- [1] G. B. Mindlin and R. Laje, *The Physics of Birdsong* (Springer, New York, 2005).
- [2] A. Amador and G. B. Mindlin, *J. Neurophysiol.* **99**, 2383 (2008).
- [3] T. Riede, R. A. Suthers, N. H. Fletcher, and W. E. Blevins, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5543 (2006).
- [4] N. H. Fletcher and T. Riede, *J. Acoust. Soc. Am.* **119**, 1005 (2006).
- [5] T. Riede and R. A. Suthers, *J. Comp. Physiol., A*, **195**, 183 (2009).
- [6] A. Amador and G. B. Mindlin, *Chaos* **18**, 043123 (2008).
- [7] G. Boncoraglio and N. Saino, *Funct. Ecol.* **21**, 134 (2006).
- [8] H. Slabbekoorn and M. Peet, *Nature (London)* **424**, 267 (2003).
- [9] T. J. Gardner and M. O. Magnasco, *J. Acoust. Soc. Am.* **117**, 2896 (2005).
- [10] T. J. Gardner and M. O. Magnasco, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6094 (2006).
- [11] I. R. Titze, *J. Acoust. Soc. Am.* **83**, 1536 (1988).
- [12] E. M. Arneodo and G. B. Mindlin, *Phys. Rev. E* **79**, 061921 (2009).
- [13] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Butterworth-Heinemann, Oxford, 1987).
- [14] I. R. Titze, *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
- [15] B. H. Story, *J. Acoust. Soc. Am.* **117**, 3231 (2005).
- [16] J. Holland, *Hidden Order: How Adaptation Builds Complexity* (Helix Books, Cambridge, MA, 1996).
- [17] M. A. Trevisan, M. C. Eguia, and G. B. Mindlin, *Phys. Rev. E* **63**, 026216 (2001).
- [18] Viterbi algorithms are commonly used in speech recognition applications as pitch contour identification. We adapted the algorithm as described by G. D. Forney, Jr., *Proc. IEEE* **61**, 268 (1973).
- [19] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevE.82.031906> for a video example of a Kiskadee while uttering a call.
- [20] W. S. Rasband, *ImageJ* (U.S. National Institutes of Health, Bethesda, MD, 1997–2009), <http://rsb.info.nih.gov/ij/>