THE ESTIMATION OF BIRDSONG CONTROL PARAMETERS USING MAXIMUM LIKELIHOOD AND MINIMUM ACTION

Tamara Smyth, Jonathan S. Abel, Julius O. Smith III

Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
tamara/abel/jos@ccrma.stanford.edu

ABSTRACT

In this research, a method is presented for extracting the control parameters of an avian syrinx synthesis model from recorded birdsong. A look-up table pairs combinations of pressure and tension parameters with the model's corresponding output power spectra. At each time frame, a generalized likelihood ratio fills a pressure-tension matrix indicating similarity between the birdsong power spectrum and the tabulated spectra. Successive pressure-tension matrices are stacked and points exhibiting a good fit to the data align to form trajectories corresponding to changes in pressure and tension over time which can then be used to control the model. In the event a range of trajectories matches the data well, the selected trajectory is that of least action.

1. INTRODUCTION

A physical model of the bird's vocal tract was developed using waveguide synthesis techniques for the bronchi and trachea tubes and finite difference methods for the nonlinear vibrating syringeal membranes [1, 2]. In order to judge the model's ability to produce realistic birdsong, and also to improve playability of the model by limiting the parameter space, we have devised a method for extracting the two primary control parameters of the model, pressure and tension, from recorded birdsong.

We employ a *maximum likelihood* approach illustrated in Fig. 3 to extract pressure and tension trajectories. Recorded birdsong is segmented and the power spectra frames are compared to tabulated model power spectra using a likelihood function to indicate goodness of fit. Those control parameter sets having large likelihood at each frame are then aligned to form trajectories over time. Finally, the trajectories are adjusted to account for the ability of the bird to generate and change pressure and tension.

We begin by briefly outlining aspects of the avian syrinx model to give context to the discussion that follows. The maximum likelihood approach is described and the likelihood function detailed. The method is then applied to a field recording of a zebra finch.

2. THE SYNTHESIS MODEL

The bird's airway consists of a trachea which divides into the left and right bronchi at its base, and a membrane which forms a valve near the top of each bronchus. During voiced song, the membrane is set into motion by air flow, vibrating at a frequency determined partly by the mass and tension of the membrane and partly by the resonance of the air column to which it is connected [3].

The valve model has the following four variables, illustrated in Fig. 1, which evolve over time during sound production:

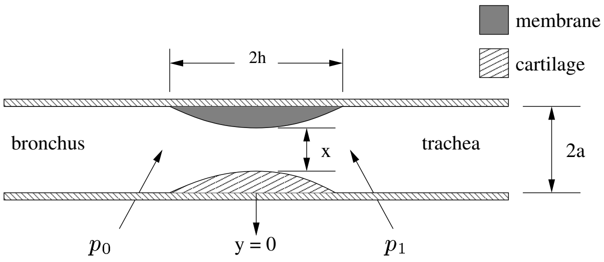

Figure 1: *The transverse model of a pressure controlled valve.*

1. Pressure on the bronchial side of the valve, $p_0(t)$.
2. Air volume flow through the valve channel, $U(t)$.
3. Displacement of the membrane, $x(t)$.
4. Pressure on the tracheal side of the valve, $p_1(t)$.

The model of the valve displacement and the resulting pressure through the constriction is based on the mechanical properties of the membrane and the Bernoulli equation for the air flow. The methods used for digitally simulating the avian vocal tract model are more thoroughly described in [1] and [2], but a signal flow diagram is presented in Fig. 2 for convenience.

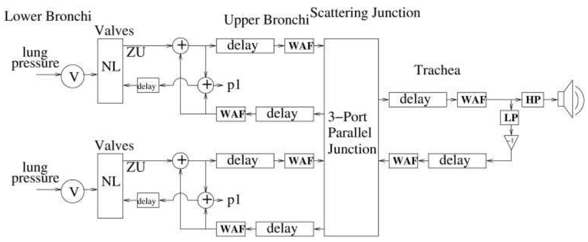

Figure 2: *Signal flow diagram of the model*

The syrinx shows tremendous variation in structure between different bird species; example values of some anatomical parameters used in the model for various bird sizes are given in Table 2. The two primary control parameters of the model (and the bird) are the blowing pressure from the lungs and the tension of the membrane [4] which the bird controls by contracting the syringeal muscles and by raising the pressure in the interclavicular air sac which encases the syrinx.

There is a definite relationship between the pitch of the sound produced by the bird and the tension of the syringeal membrane.

Fletcher writes the displacement of the membrane as a function of time $x(t)$ for mode n as [5]

$$m_n \left[\frac{d^2 x_n}{dt^2} + 2\kappa \frac{dx_n}{dt} + \omega_n^2 (x_n - x_0) \right] = \epsilon_n F, \qquad (1)$$

where the frequency of the first mode ω_1 is given by

$$\omega_1 = \left(\frac{5T}{\rho_M a H d} \right)^{\frac{1}{2}}, \qquad (2)$$

with T being tension, ρ_M the membrane density, a the radius of the bronchus, H the membrane width and d the membrane thickness with example values given in Table 1. It can be seen therefore, that when the bird increases the tension of the membrane the pitch of the song will also increase. Likewise, an increase in air pressure from the lungs will cause a general increase in amplitude of the produced sound.

anatomical parameters	**small**	**medium**	**large**
membrane density (kg/m^3)	1000	1000	1000
membrane width (mm)	1.9	3.5	4.1
membrane thickness (μm)	100	100	100
left bronchus length (mm)	5	14	30
right bronchus length (mm)	5	14	30
trachea length (mm)	17.1	23	35.6
left bronchus radius (mm)	1.47	2.5	3.5
right bronchus radius (mm)	1.47	2.5	3.5
trachea radius (mm)	1.9	3.5	4.1

Table 1: *Examples of fixed anatomical parameters for different bird sizes. Though it is the case here, it is not necessary that the left and right bronchus be symmetrical*

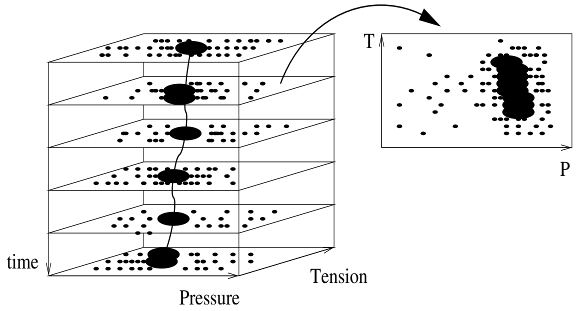

Figure 3: *Stack of likelihood images.*

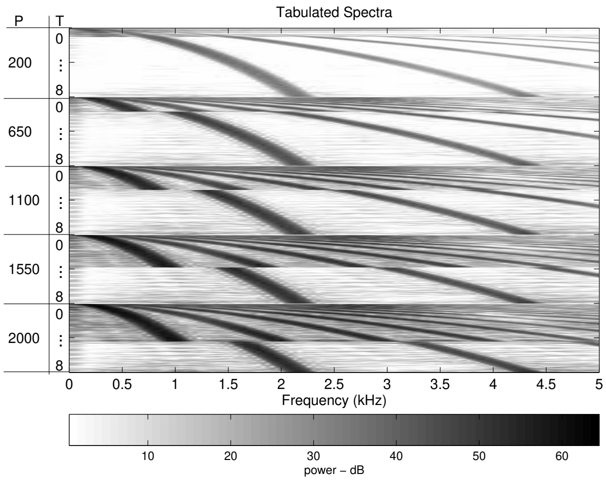

Figure 4: *Model output power spectra for different pressure and tension values.*

The mapping of pressure and tension to loudness and pitch respectively is not straightforward however, since nonlinearities intrinsic to the dynamics of the syrinx cause less predictable behaviour [6]. A slight change in one parameter, for instance, can cause effects such as period doubling, mode-locking and transitions from periodic to chaotic behaviour. It would also be somewhat unreasonable to ask a player of the model to control the parameters with the speed and virtuosity of a songbird. It is clear therefore, another mapping strategy is necessary.

3. MAXIMUM LIKELIHOOD MODEL

We begin by specifying fixed anatomical parameters of the model corresponding to the bird species whose song we wish to simulate and of which we have a recording. In this case our technique is illustrated making reference to a recording of a zebra finch and makes use of the small bird anatomical parameters from Table 1.

A look-up table is generated by pairing combinations of the control parameters pressure and tension with the model's corresponding output power spectra. Fig. 4 shows an excerpt of the spectrum table for tension ranging from zero to eight at five pressure values. The actual table has many more entries and since it only need be computed once (or at least once per bird species), it can be made as large and dense as is permitted by computer memory.

With the tabulated model spectra in place, the recorded birdsong is processed via a short-time Fourier transform [7] to form a sequence of power spectra over time. An example spectrogram in Fig. 5, taken from a field recording, shows a sequence of gestures for the song of a zebra finch. As is typical of field recordings, it contains a significant amount of background noise concentrated in the low frequencies.

At each time frame, a normalized *log likelihood function* [8] is used to indicate the similarity between the birdsong power spectrum and each of the entries in the spectrum table, filling a pressure-tension matrix similar to the one shown in Fig. 6. The likelihood function used here is normalized so that entries having a value close to one—indicated with dark pixels in Fig. 6—correspond to pressure-tension settings producing spectra very similar to the measured frame spectrum. Those producing entries close to zero—indicated with lighter pixels—have very different spectra from the measured frame under consideration.

The likelihood function was chosen for its statistical properties: In the limit of small estimation errors, its peak location is known to be unbiased with minimum variance [8]. In other words, the likelihood function will, on average, peak at the correct control parameter values, and the peak location will be as insensitive as possible to measurement noise.

We assume that the recorded signal $s(t)$ consists of the model output $\mu_\theta(t)$ with measurement and/or modeling error $\nu(t)$ which

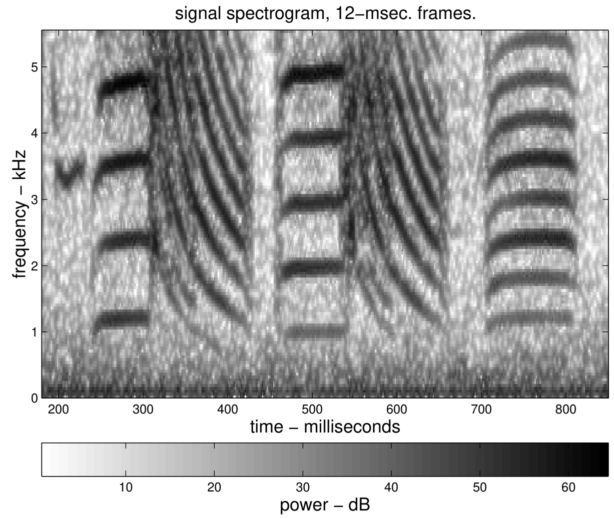

Figure 5: *Power spectra for field recording of a zebra finch.*

is additive and Gaussian-distributed,

$$s(t) = \alpha\mu_\theta(t) + \nu(t), \tag{3}$$

where α is an unknown scaling factor and $\theta = [P, T]$ is the parameter vector containing pressure P and tension T.

Assuming the measured signal is stationary over the duration of an analysis frame, the power spectrum of the measured signal is given by

$$S(\omega) = \alpha^2 M_\theta(\omega) + N(\omega) \tag{4}$$

with M_θ representing the model output power spectrum for parameters θ and $N(\omega)$ the noise power spectrum at frequency ω. In the case that the noise power $N(\omega)$ is much smaller than the signal, the probability of observing a power spectrum $\mathbf{S}$ given control parameters θ may be approximated by

$$\begin{aligned} \wp(\mathbf{S};\theta) &\approx \mathcal{N}(\alpha^2\mathbf{M}_\theta, \Sigma) \\ &= \frac{\exp\{-\frac{1}{2}(\mathbf{S}-\alpha^2\mathbf{M}_\theta)^\top\Sigma^{-1}(\mathbf{S}-\alpha^2\mathbf{M}_\theta)\}}{\det(2\pi\Sigma)^{\frac{1}{2}}} \end{aligned} \tag{5}$$

where $\mathbf{S}$ is the stack containing power spectrum bins $S(\omega_i)$ and similarly $\mathbf{M}_\theta$ is the stack of model power spectrum bins.

Under the assumption that the additive spectral noise is independent and identically distributed, the log likelihood for a parameter set θ given measurements $\mathbf{S}$ is

$$l(\theta;\mathbf{S}) = -\frac{1}{2\sigma^2}(\mathbf{S}-\alpha^2\mathbf{M}_\theta)^\top(\mathbf{S}-\alpha^2\mathbf{M}_\theta). \tag{6}$$

For convenience, we normalize the likelihood so that it ranges from zero, indicating an unlikely fit, to one, indicating a good match:

$$\bar{l}(\theta;\mathbf{S}) = 1 - \frac{(\mathbf{S}-\alpha^2\mathbf{M}_\theta)^\top(\mathbf{S}-\alpha^2\mathbf{M}_\theta)}{(\mathbf{S}+\alpha^2\mathbf{M}_\theta)^\top(\mathbf{S}+\alpha^2\mathbf{M}_\theta)}. \tag{7}$$

The second term from (7) may be interpreted as the squared distance between the scaled measured data and the table entry, normalized by the squared length of their sum. When the parameters are such that the measured and model spectra coincide, the numerator above will be small and the likelihood close to one. When the measured and model spectra are very different, the numerator is large—it can never be larger than the denominator as both spectra are positive—and the likelihood is close to zero.

4. MINIMUM ACTION

Typically there will be a range of likelihood ratios that match the data well. The pressure-tension likelihood functions tend to have maxima that are wide in pressure and narrow in tension. On occasion there will be well separated likelihood maxima corresponding to different registers or modes of oscillation with a division between the two. We assume our bird will not expel any unnecessary energy during song performance and in the event of multiple well separated matching maxima we choose the one requiring the least effort on the part of the bird.

Minimum action is introduced into the likelihood function in two ways: one takes into account the effort involved to move from one parameter value to another over time while the other simply incorporates the instantaneous effort where higher values of tension and pressure are considered to require more effort.

For the instantaneous effort we assume that it is more difficult to produce higher values of pressure and tension and represent this added difficulty with a penalty function added to the likelihood function. On the assumption that it is difficult for the bird to rapidly slew control parameters, the sequence of likelihood function maximizers are median filtered to eliminate sporadic, short lived, jumps in the trajectories. A further slew limiting filter is applied to ensure rates of parameter change that are physical plausible.

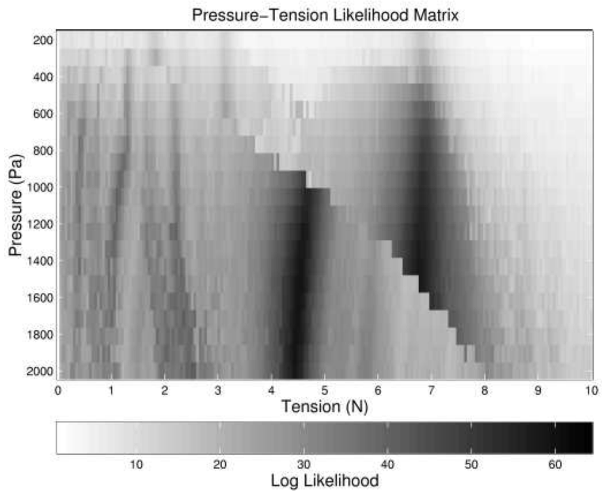

Figure 6: *A pressure-tension matrix for one time frame shows two disparate operating registers.*

5. CONCLUSIONS

This research serves two purposes: 1) to judge the model's ability to produce realistic birdsong by calibrating it to recorded birdsong and 2) to restrict and scale the control parameter space so as to

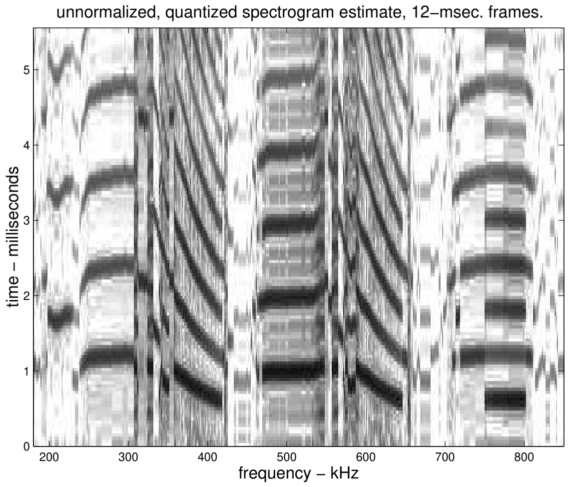

Figure 7: *Power spectra based on maximum likelihood control estimates.*

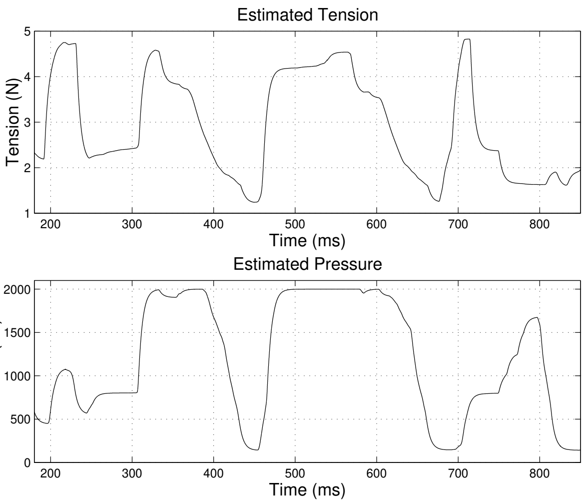

Figure 8: *Control trajectories for pressure and tension.*

improve the user's ability to interact with the model, e.g., by having a controller that follows predetermined trajectories through the matrix stack.

We applied our method to the zebra finch song shown in Fig. 5. Tabulated power spectra corresponding to the most likely parameters at each time frame are sequenced and shown in Fig. 7. The similarity between the two spectra show very good potential in the model's ability to produce accurate bird-like song.

The final control trajectories extracted from the zebra finch recording are shown in Fig. 8. The spectrogram of the corresponding model output is shown in Fig. 9. Though not identical to either Fig. 5 or Fig. 7 the sound output is perceptually similar and the song gesture is well captured. By mapping the trajectories of Fig. 8 to an input device with two individual continuous controls, the player of the model is able to reproduce bird-like song without requiring the bird's expert technique.

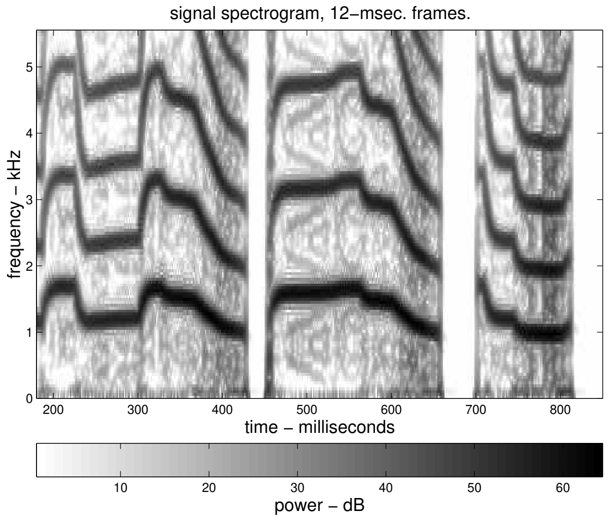

Figure 9: *Power spectra of model output using control trajectories from Fig. 8*

6. REFERENCES

[1] Tamara Smyth and Julius O. Smith, "The sounds of the avian syrinx—are they really flute-like?," in *DAFX 2002 Proceedings*, Hamburg, Germany, September 2002, International Conference on Digital Audio Effects.

[2] Tamara Smyth and Julius O. Smith, "The syrinx: Nature's hybrid wind instrument," in *CD-ROM Paper Collection*, Cancun Mexico, September 2002, Pan-America/Iberian Meeting on Acoustics.

[3] Neville H. Fletcher and A. Tarnopolsky, "Acoustics of the avian vocal tract," *Journal of the Acoustical Society of America*, vol. 105, no. 1, pp. 35–49, January 1999.

[4] Tim Gardner, G Cecchi, M. Magnasco, R. Laje, and Gabriel B. Mindlin, "Simple motor gestures for birdsongs," *Physical Review Letters*, vol. 87, no. 20, pp. 1–4, November 2001.

[5] Neville H. Fletcher, "Bird song – a quantitative acoustic model," *Journal of Theoretical Biology*, vol. 135, pp. 455–481, 1988.

[6] M. S. Fee, B. Shraiman, B. Pesaran, and P. P. Mitra, "The role of nonlinear dynamics of the syrinx in the vocalization of a songbird," *Nature*, vol. 395, pp. 67–71, 1998.

[7] J. B. Allen and L. R. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, November 1977.

[8] Louis L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, Inc., 1991.