COMPUTER SYNTHESIS OF BIRD SONGS AND CALLS

Mark Kahrs[1]

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O. Box 3000
FIN-02015 HUT
`kahrs@acoustics.hut.fi`

Federico Avanzini

Universita' degli Studi di Padova
Dip. di Elettronica ed Informatica
Via Gradenigo 6/A
35131 Padova Italy
`avanzini@dei.unipd.it`

"Among the artistic hierarchy, the birds are probably the greatest musicians to inhabit the planet"
Messiaen[1]

ABSTRACT

Bird songs are fascinating acoustic phenomena. In this paper, we first examine the acoustic mechanisms of sound production in birds and contrast them with their anatomy. Next, we describe the simulation of a one-mass source together with a simple transmission line model for a psittacine bird. In concluding, we discuss future areas for research.

1. INTRODUCTION

Bird songs have provided inspiration to composers and listeners for centuries. But the acoustic mechanism of sound production remains a point of scientific contention. In this paper, we study the mechanisms of sound production in birds and compare and contrast their anatomy and their sounds using existing acoustical models. We implemented a simple model of a single tract bird with a modified human vocal source. In concluding, we will also point out areas for future research.

2. PAST WORK

The formal study of the acoustics of bird song originated with Greenewalt[2]. Greenewalt proposed the *syrinx* as the sound source in birds as opposed to the vocal folds in humans and other primates. However, anatomy is destiny... Different avian families have different vocal tracts. For example, song birds (Oscines) have two syrinxes below a common trachea as shown in figure 1. Psittacines, on the other hand, resemble the human vocal tract by having a single syrinx, trachea and tract as shown in figure 2. The syrinx can be located either above or below the trachea.

[1]Current address: Dept. of Electrical Engineering, Univ. of Pittsburgh, Pittsburgh, PA 15217 (USA); `kahrs@ee.pitt.edu`

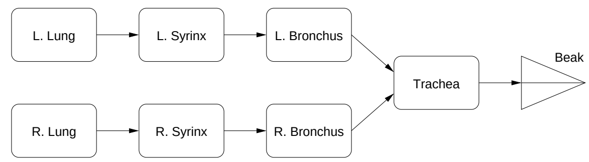

Figure 1: Oscine tract model

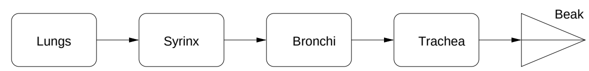

Figure 2: Psittacine tract model

Using the standard waveguide physical model[3], we can divide the synthesis problem into four aspects:

1. Syrinx

   The Syrinx is believed to be the source of oscillation in the avian vocal tract. It was first studied in detail by Casey and Gaunt [4]. The exact acoustical functioning of the syrinx remains controversial and is still the subject of scientific inquiry.

2. Trachea

   As stated above, in Oscines, there are two syringeal valves (four membranes total) connected via a common trachea. Because the oscillations can interact, we must model the trachea as a 3-port junction in this case. Three port junctions, for instance, were used by Välimäki[5] to model toneholes in the flute. The Psittacines, being simpler, use a modified human vocal tract model without the 3-port junction.

3. Tract

   We can simply model the tract as a smooth frictionless tube. Of course, this is as realistic as similar models of the human vocal tract but it suffices for a first cut. We consider the tract to include the larynx and the tongue. The tract can have varying length due to physical movement of the head[6].

4. Beak

   Since birds do not have lips, we must model the opening of beaks. Simple models for beaks were developed by Fletcher[7]. In essence, beaks act in a similar manner to horns[8]: they exhibit a cutoff frequency that acts as a variable high pass filter. Beak size is correlated with pitch[9] as one might expect. Beak opening (gape) can change the pitch by shortening the vocal tract as the oral cavity becomes flared. It could also change the vocal tract impedance by closing the tube[6] as the beak closes.

In this paper, we will examine each one of these synthesis problems one by one and finally demonstrate how they can be interconnected to simulate the sounds of birds.

3. VOCAL SOURCE MODEL

Bird song has been a topic of investigation for many years among ornithologists[10] and composers such as Messiaen (e.g., *Catalogue d'Oisseaux*). But it was Greenewalt[2] who first proposed an acoustical model for sound production. His work forms the basis of subsequent work on the acoustics of the avian vocal tract.

3.1. System models

The system model begins at the lungs and then follows the air flow from the syrinx forward. This can be modelled in the same manner as the human vocal tract[11]. Greenewalt[2] proposed a simple acoustical model of a time-varying constriction resulting in turbulent flow. Casey and Gaunt[4] proposed three different models: a vibrating membrane, a vibrating string and an aerodynamic model. They concluded that the vibrating membrane model was useful for harmonic spectra whereas the aerodynamic model was more suited to pure tone or noisy spectra. Fletcher[7] disputes a number of these claims.

3.2. Modeling the syrinx

One avenue for vocal source modeling is to take the existing human models and change them to accommodate the anatomy of birds. The most common human model is Ishizaka and Flanagan's two mass model[12]. This was modified by Mergell, et al.[13] for mammalian non-human sources. Rodet[14] also demonstrated how one and two mass models can be used as sources in musical instruments.

Fletcher[15] developed a numerical model of the vocal source of Ravens that tries to duplicate some of the nonlinear effects seen during large amplitude vibration. He noted that the syrinx model he produced was incapable of producing whistled song. Further studies by Fletcher[16] examine the generation of jet flow in a restricted orifice. He also mentions the possibility of "mode locking" between two vibrating syringeal membranes (This was mentioned earlier in the context of musical oscillators[17]). Fletcher[18] also explores the concept of *independent* oscillators resulting from vastly different oscillating frequencies. He suggests that such oscillations might produce "chaotic" calls.

An accurate description should regard the syringeal membrane as a 2-D distributed system with a infinite number of resonant modes, and model it using a second order partial differential equation. However, this is far too detailed for our purposes, and Fletcher[16] suggests that much simpler models are able to capture the basic features of pressure controlled valves. The syrinx is therefore treated as a lumped mass subject to elastic restoring forces and internal dissipation. Nonlinear interaction with the airflow is described following the Ishizaka and Flanagan[12] model for the human glottis. This is also used to model complete closure of the syrinx: during closure an additional restoring force is added and dissipation is increased. It must be stressed that this also introduces a strong nonlinearity in the system. During the whole closed phase the syringeal flow is zero, and consequently its spectrum is broadened and higher partials are generated. The main difference with the Ishizaka and Flanagan model is that the syrinx is modelled as a single mass, whereas their standard human vocal fold model is modelled as two coupled masses.

Furthermore, it must also be acknowledged that the human vocal tract is a source of nonlinear interaction; the avian vocal tract no doubt also exhibits such behavior.

3.3. Trachea and Tract models

Brittan-Powell, et al.[19] summarize three different theories:

1. The syrinx is a multiphonic source and the output is bandpass filtered by the vocal tract – in essence the syrinx and tract are decoupled.

2. The syrinx and tract are acoustically decoupled, but they track each other through neuro-coordination.

3. The syrinx and tract are coupled directly so that the tract suppresses harmonics generated at the vocal source. This was also proposed as a model for the coloratura soprano [20].

Fletcher[7] analyzed the Oscine tract model shown in figure 1. Here, the lungs are connected to two different syringeal membranes. Each syrinx is connected to the trachea. The trachea model is connected to the model, which has a tongue. Finally, the beak is connected to the mouth. Fletcher gives the acoustical impedance as

$$Z = B_{11} - \frac{B_{12}^2(B'_{11}T_{22} + T_{11}T_{22} - T_{12}^2)}{B'_{11}(B_{22}T_{22} + T_{11}T_{22} - T_{12}^2) + B_{11}(T_{11}T_{22} - T_{12}^2)} \quad (1)$$

where B_{ij} denotes the impedance of the bronchi from ports i and j and T denotes the impedance of the trachea from ports i and j. B' denotes the second brochus. Among the simplifications made here is the assumption that the walls are smooth and frictionless. Fletcher claims that such an adjustment leads to a difference of "only" 10%, therefore it is neglected in his results. Note that if the second bronchus is closed, i.e., the single bronchus case, then $B_{11} = \infty$.

Assuming the impedance of trachea and bronchi can be modeled with tubes, then the classical approximation[8] is:

$$Z_{11} = Z_{22} = -jZ_B \tan kl \quad (2)$$

$$Z_{12} = Z_{21} = -jZ_B \csc kl \quad (3)$$

where $Z_B = \frac{\rho c}{S}$, k is the wavenumber $\frac{\omega}{c} + j\alpha$, ω is the well known angular frequency $2\pi f$, S is the cross sectional area, and f is the frequency in Hertz. α is the attenuation constant 2×10^{-5}.

3.4. Mouth and Beak model

Fletcher[7] also presents models of both the mouth and the beak. The mouth can be modeled in much the same manner as the human mouth, with the exception that the tongue is (of course) less flexible than the human equivalent. A simple model is as a short piece of pipe with a time varying cross sectional area. Fletcher gives the input impedance as follows:

$$Z = T_{11} - \frac{T_{12}^2(M_{22} + K)}{(T_{22} + M_{11} + L)(M_{22} + K) - M_{12}^2} \quad (4)$$

where M denotes the impedance of the mouth, K denotes the beak input impedance and L denotes the impedance of the larynx.

Fletcher provided two models of the beak: first he presents a simple slotted cylindrical model and then a conical model. His final beak model is:

$$K(f,g) = jZ_B[-\cot\frac{k\delta}{2} + \frac{\csc^2\frac{k\delta}{2}}{\cot(\frac{k\delta}{2}) - \frac{k\delta}{2}}] \quad (5)$$

and g is the beak opening (gape) and the end correction $\tilde{\delta}(f,g) \approx 0.05\, l + 10^{-5}\, f\, l^2/g$.

4. COMPUTER SIMULATION

The acoustic waveguide model is as shown in figure 3. As shown in figure 3, the vocal source produces a forward going pressure wave. The return from the tract is the negative going wave shown on the bottom.

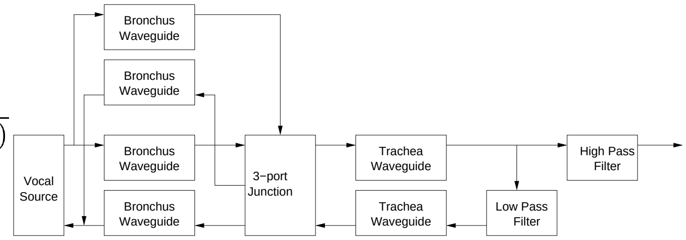

Figure 3: Waveguide implementation

4.1. Glottal source (syrinx)

The syrinx is modelled according to the discussion in section 3.2. The values for the masses come from the paper by Fletcher[15] although he does not give the mass of the membrane or spring constant directly. Therefore we compute the membrane volume, assuming that the membrane density is approximately 1000 Kg/m^3 (as in Titze[21] for vocal folds), and find the mass m. Next, we choose the spring constant k such that the resonance frequency $2\pi f_0 = \sqrt{k/m}$ of the syrinx matches a desired value. In our simulations, k is a time-varying control parameter, so that the pitch of the tone can be adjusted over time. The discretization technique is based on the bilinear transformation and the K-method [22] is used for dealing with non-linearities in the system.

4.2. Transmission lines

After the source, we model the trachea and bronchii. Since anatomy plays such a critical part in the simulation, the various parameters (all in mm) can change tremendously: l is the length and r is the radius of the tube.

parameter	oil bird	raven	budgie
l_{trachea}	100	70	50
r_{trachea}	2.5	3.5	2
l_{bronchus_1}	10	0	0
r_{bronchus_1}	1.5	0	0
l_{bronchus_2}	15	0	0
r_{bronchus_2}	1.5	0	0
l_{beak}	20	20	10

Note that for the raven, the bronchi are zero; this is because the raven has a single syrinx and so we can place the bronchus on the measurement of the trachea. The transmission lines should be implemented with fractional delay lines[5], particularly if the length of the transmission line is to be made variable. All of the transmission lines are assumed lossless, but could be made lossy [23].

The 3-port junction is also lossless, which makes it very easy to implement.

4.3. Reflectance filters

Finally, the beak is implemented as two 5th order Butterworth filters, one for the highpass and one for the lowpass. The pole trajectory for the filters should follow the nonlinear path of the beak cutoff as discussed by Fletcher (and seen in equation 5)

4.4. Simulation

Figure 4 shows the output of the syrinx simulation. The

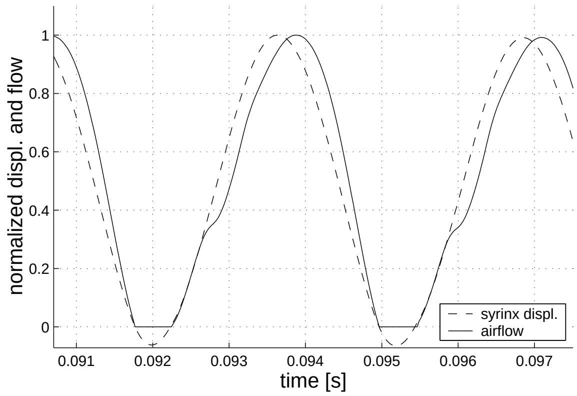

Figure 4: Syrinx output

output of the syrinx is used as the input to the transmission line. The transmission line of a bird is *small*, for example, a 20 mm tube will be only 2.56 samples at 44100 Hz. The low pass filter output is fed back to the syrinx model. The final output after the beak high pass filter is shown in figure 5. An individual period output is shown in figure 6.

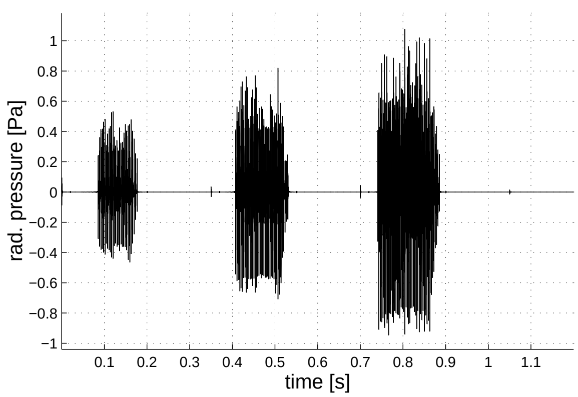

Figure 5: Pressure output

5. CONCLUSION

It must be acknowledged that in the absence of scientific certainty, it is adventurous to implement an avian vocal tract

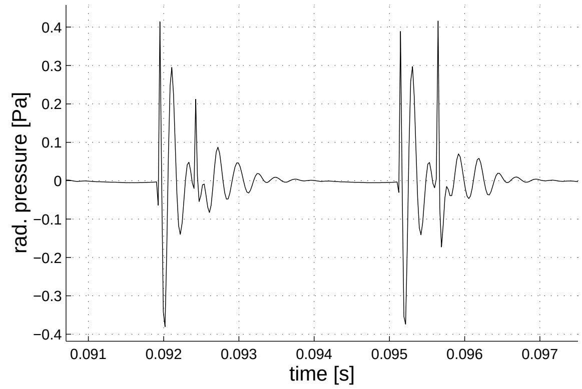

Figure 6: Zoom into pressure output

model. However, since we are interested in the sounds and not the scientific truth, this makes it easier. There are many directions for future work, including mode-locking of multiple syrinxes, the introduction of aerodynamic models for the three port junction, lossy transmission lines and tongue models.

6. ACKNOWLEDGEMENTS

This work was only made possible because the authors had the good fortune to be thrown together in the Helsinki acoustic hothouse, courtesy of Matti Karjalainen, who has encouraged this work. M.K. was supported by the Fulbright Foundation as well as Tekes. F.A. was funded by the Sound Source Modeling Project. Both authors were supported in part by the Academy of Finland.

7. REFERENCES

[1] Claude Samuel. *Entretiens avec Olivier Messiaen (Conversations with Olivier Messiaen)*. Stainer & Bell, London, 1976.

[2] C. H. Greenewalt. *Bird Song: Acoustics and Physiology*. Smithsonian Institution Press, Washington, D.C., 1968.

[3] J. O. Smith. Principles of digital waveguide models of musical instruments. In Mark Kahrs and Karlheinz Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*. Kluwer Academic Publishers, 1999.

[4] R. M. Casey and A. S. Gaunt. Theoretical models of the avian syrinx. *J. Theo. Biology*, 116:45–64, 1985.

[5] Vesa Välimäki. *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*. Ph.D. thesis, Helsinki Univ. Technology, December 1995.

[6] M. W. Westneat, J. H. Long, W. J. Hoese, and S. Nowicki. Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows. *J. Exp. Biology*, 182:147–171, 1993.

[7] N. H. Fletcher and A. Tarnopolsky. Acoustics of the avian vocal tract. *J. Acoust. Soc. Am.*, 105(1):35–49, January 1999.

[8] L. R. Beranek. *Acoustics*. McGraw-Hill, New York, 1954.

[9] J. Podos. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, 409:185–188, January 2001.

[10] C. K. Catchpole and P. J. Slater. *Bird Song - Biological themes and variations*. Cambridge University Press, 1995.

[11] G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, 1960.

[12] K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Sys. Tech. J.*, 51:1233–1268, 1972.

[13] P. Mergell, W. T. Fitch, and H. Herzel. Modeling the role of nonhuman vocal membranes in phonation. *J. Acoust. Soc. Am.*, 105(3):2020–2028, March 1999.

[14] X. Rodet. One and two mass model oscillations for voice and instruments. In *Proceedings of International Computer Music Conference (ICMC)*, Banff, Canada, September 1995.

[15] N.H. Fletcher. Bird song - quantitative model. *J. Theo. Biology*, 135:455–481, 1988.

[16] N. H. Fletcher. Autonomous vibration of simple pressure-controlled valves in gas flows. *J. Acoust. Soc. Am.*, 93(4):2172–2180, April 1993.

[17] N. H. Fletcher. Mode locking in non-linearly excited inharmonic musical instruments. *J. Acoust. Soc. Am.*, 64:1566–1569, 1978.

[18] N. H. Fletcher. A class of chaotic bird calls? *J. Acoust. Soc. Am.*, 108(2):821–826, August 2000.

[19] E. F. Brittan-Powell, R. J. Dooling, O. N. Larsen, and J. T. Heaton. Mechanisms of vocal production in budgerigars. *J. Acoust. Soc. Am.*, 101(1):578–589, January 1997.

[20] M. Rothenberg. Cosi fan tutte and what it means, or nonlinear source-tract interaction in the soprano voice and some implications for the definition of vocal efficiency. In T. Baer, C. Sasaki, and K. Harris, editors, *Laryngeal Function in Phonation and Respiration*, pages 254–264. College-Hill, Boston, MA, 1987.

[21] I. R. Titze. Parameterization of the glottal area, glottal flow, and vocal fold contact area. *J. Acoust. Soc. Am.*, 75(2):570–580, 1984.

[22] G. Borin, G. De Poli, and D. Rocchesso. Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *IEEE Trans. Speech and Audio Processing*, 8(5):597–606, September 2000.

[23] Vesa Välimäki, Tero Tolonen, and Matti Karjalainen. *Signal-Dependent Nonlinearities for Physical Models Using Time-Varying Fractional Delay Filters*. 1998 International Computer Music Conference (ICMC 98), Ann Arbor, Michigan, October 1998.