

---

# An Optimal Control Model of Zebra Finch Vocalization

---

**Mike Schachter \***

Helen Wills Neuroscience Institute  
University of California, Berkeley  
Berkeley, CA 94720  
mike.schachter@gmail.com

## Abstract

In this work, a nonlinear oscillator modeling the syringeal folds of the Zebra Finch is controlled by the state of higher level linear dynamical system. We formulate an optimal control cost function and solution for the learning of bird song.

## 1 Introduction

First we'll discuss the Zebra Finch vocalization system and it's mathematical formulation, and then the control of that system. All code used to generate figures and results in this paper can be found at:

<http://github.com/mschachter/birdy>

### 1.1 The Zebra Finch Vocalization System

Like human speech, Zebra Finch song is generated at it's source by oscillations in airflow generated by vibrating vocal cords. A model for these oscillations is given as:

$$\begin{aligned}\dot{x} &= v \\ \dot{v} &= \gamma^2 \alpha + \gamma^2 \beta x - \gamma^2 x^3 - \gamma x^2 v + \gamma^2 x^2 - \gamma x v\end{aligned}$$

The control parameters of the model are  $\alpha$  and  $\beta$ .  $\gamma = 23500$  is a constant.

### 1.2 Control of the Syrinx Model

The goal is to control the parameters  $\alpha$  and  $\beta$  in order to produce an observed vocalization. Let  $\phi(t) = [\alpha(t) \beta(t)]^T$  be the state vector.

Assume that the temporal evolution of  $\phi(t)$  is defined by a controlled linear dynamical system:

$$\dot{\phi} = A\phi + Bu$$

where  $A$  is a 2x2 matrix, chosen to make the passive dynamics of the control system decay to some physiologically relevant rest state.  $B$  is a 2x2 control matrix, chosen to be the identity matrix for

---

\*With a metric ton of help and code from Hédi Soula: [hsoula@gmail.com](mailto:hsoula@gmail.com)

simplicity. Note that the time-scale for simulation the oscillator is much finer than that of the control. Let  $\Delta\tau$  be the time step for simulation of the control system, and define  $t_k = k\Delta\tau$ .

In optimal control theory, a cost function is specified and is to be minimized over time to produce an optimal control law. But what form should the cost function take? We'll make the following assumptions:

1. The control wants to keep the system's instantaneous energy low:  $\phi(t_k)^T \phi(t_k)$ .
2. The control wants to keep it's instantaneous energy low:  $\mathbf{u}^T \mathbf{u}$
3. The control varies smoothly, i.e.  $(\phi(t_k) - \phi(t_{k-1}))^2$  is small.
4. The control wants to produce an instantaneous fundamental frequency that matches that of a stored template.

To elaborate on the last assumption, say we are given the time-varying fundamental frequency of a song syllable that we would like to learn, represented as the function  $F_f(t)$ . Let  $\hat{F}_f = g(\phi)$  be the function that gives the steady state fundamental frequency for a given control  $\phi$ . The function  $g$  can be empirically estimated through simulation and approximated through interpolation. To follow this frequency means to keep the quantity  $(F_f - g(\phi))^2$  small.

Given these assumptions, we can construct the following instantaneous cost function:

$$C(\phi(t_{k-1}), F_f(t_k), \mathbf{u}) = c_\phi \phi(t_k)^T Q \phi(t_k) + c_u \mathbf{u}^T \mathbf{u} + c_s (\phi(t_k) - \phi(t_{k-1}))^2 + c_F (F_f - g(\phi))^2$$

where  $Q$  is a diagonal 2x2 scaling matrix and the  $c_* \in \mathbb{R}$  are weights to keep the magnitudes of each cost within a reasonable range.  $\phi(t_k)$  is given as a forward Euler step:

$$\phi(t_k) = (A\mathbf{x} + B\mathbf{u}) \Delta\tau + \phi(t_{k-1})$$

### 1.3 Temporal Hierarchy of Representation

Generically, we want to determine a joint probability between  $\phi$  and a set of variables that are related to the acoustic representation. These features may span several time scales. For example, instead of just relating  $F_f(t)$  and  $\phi$ , i.e. looking at the joint distribution  $p(F_f(t), \phi)$ , we might want to look at the running variance of  $F_f(t)$  for a specified time window.

Let  $\{\sigma_1, \dots, \sigma_m\}$  be a set of statistics for some acoustic variables. Construct this set so that they are ordered by time scale. By this, we mean that computing  $\sigma_i$  requires a larger window of time than computing  $\sigma_j$  if  $i < j$ .

Finding the optimal control requires maximizing a conditional probability:

$$\underset{\mathbf{u}}{\operatorname{argmin}} C_f = \underset{\phi(\mathbf{u})}{\operatorname{argmax}} p(\phi(\mathbf{u}) | \sigma_1, \dots, \sigma_m)$$

There may be a "telescoping" algorithm to maximizing  $p(\phi(\mathbf{u}) | \sigma_1, \dots, \sigma_m)$  quickly. Start with  $\sigma_1$ , which is the statistic with the longest time scale. There should be many instantaneous values of  $\mathbf{u}(t)$  that give a nonzero probability of occurrence with  $\sigma_1$ . But not all the values will, so restrict the search for all  $\mathbf{u}$  to that space. Do the same for  $\sigma_2$ , which will reduce the size of the space even further. Continue this process until the space of actual  $\mathbf{u}$  is small enough to do a more efficient optimization, and then perform that optimization to find the optimal control.

### Acknowledgments

Thank you to Hédi Soula, who created a C++/Python implementation of this model in our lab and has provided ideas, expertise, and feedback on this project.

## References