MSc Acoustics and Music Technology Final Project

A Computational Model of Avian Vocalizations

Author: Yining Xie
Student Number: s2172442
Supervisor: Dr Michael Newton
Date: 04/09/2023

Introduction – Background and Motivation

- The beautiful and diverse nature of birdsong that attracts human’s attention

- Desire to understand the song production mechanism in birds (Mindlin et al.)

- Avian syrinx as a musical instrument to introduce (Smyth et al.)

- Inspiration for sound design artists

Introduction – State of Art

Physics-based models:

- Early version based on membrane vibration (Fletcher)
- Rescaling from human's vocal model (Fry)
- Progressive evolution starting from one mass model (Mindlin)

(validated with Zebra Finch by successfully triggering real birds' responses)

Control Parameter Fitting:

- Maximum likelihood and minimum action (Smyth)
- Look-up table (Mindlin)

Both are based on power spectrum

Table I Comparison of models in literature. A. Brackenbury (1979b); B. Fletcher (1985); C. Fee et al. (1998); D. Fry (1998); E. Gardner et al. (2001); F. Laje et al. (2002); G. Fee (2002). [1] SD; Spring Damper model, [2] TM; Tympaniform Membrane

Model type	aeroacoustical			modified oscillators			
Model	A	B	C	D	E	F	G
Aim of model development							
Neuromuscular control			•	•	•	•	
Morphology							
General mechanisms	•	•					•
Bird group							
Passeriformes (song birds)		•	•	•	•	•	•
Non-songbirds	•						
MTM modelled by							
Moving piston	•						
Edge clamped drum		•					
Labia modelled by							
One/two mass SD[1]			•	•	•	•	
Multiple mass SD[1]							•
Input parameters							
Bronchial pressure	•	•	•	•	•	•	
Complex pressure wave		•					
Labial / membrane stiffness		•		•	•	•	•
Gating of flow						•	
Model output tested							
Generation sounds	•	•	•	•	•	•	
Acoustical power output		•		•			
Amplitude modulation				•			
Pressure gradient over TM[2]		•					

Introduction – This Work

- Built computational birdsong model in Matlab based on Mindlin’s physical model
- Developed finite difference scheme which is new for this model, which runs >8 times faster than the existing implementation by Runge Kutta method (which I also have tested)
- Conducted an exhaustive assessment of the model within the parameter space, unveiling correlations between perceptual parameters (f0, SCI) through innovative remapping and visualization techniques, offering a complementary perspective from existing methods.
- Inverted the model using techniques such as look-up tables, interpolation, and machine learning. The introduction of interpolation is a novelty for this model, and the use of machine learning offers a new direction in the realm of birdsong and physics-based synthesis. This approach, using perceptual parameter input for control parameter output, improves controllability and inversion efficiency. It can hopefully support more complex models and be easily extended to musical instrument design.

Introduction – Outline of This Work

- Literature review
- The physics of birdsong
  - Song production mechanism
  - Mindlin's model
- Numerical implementation
  - Runge Kutta 4th Order (compared different coding logics, schemes in literature)
  - Finite Difference Method (schemes new of this model designed by myself)
- Evaluation
  - Influence of bifurcation
  - Influence of control parameters
  - Correlation between perceptual parameters discovered by remapping and visualization (method of my own as a compliment to the published method)
  - Inverse with a comparison of lookup table, interpolation, machine learning (new, to be finished)

The Physics of birdsong - Song Production Mechanism

Key structures:

Syrinx, Trachea, Oropharyngeal-esophageal cavity (OEC)

Diversity:

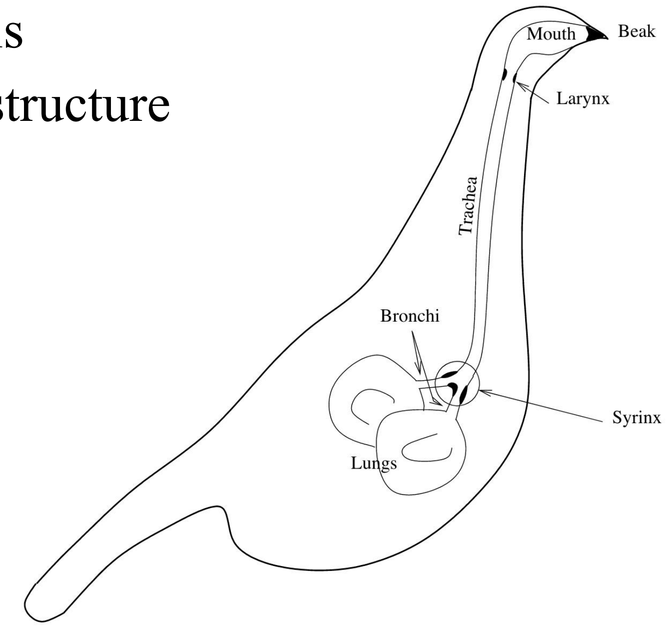
2 voice theory

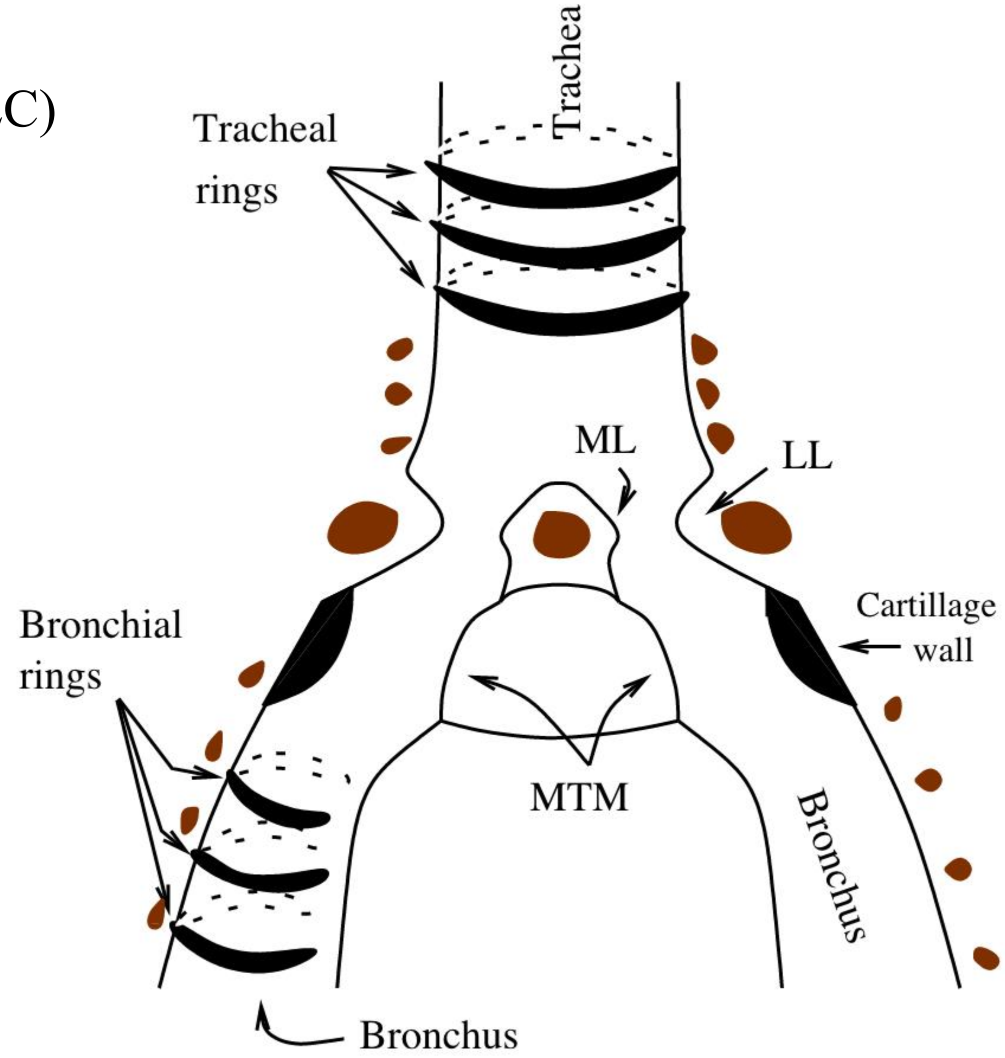
Different theories for song production mechanism

Different models

Different song structure

…





The Physics of birdsong - Mindlin's model

Syrinx:

$$M\ddot{x} = -Kx - B\dot{x} + P_s - F_0$$

M: mass of the labium

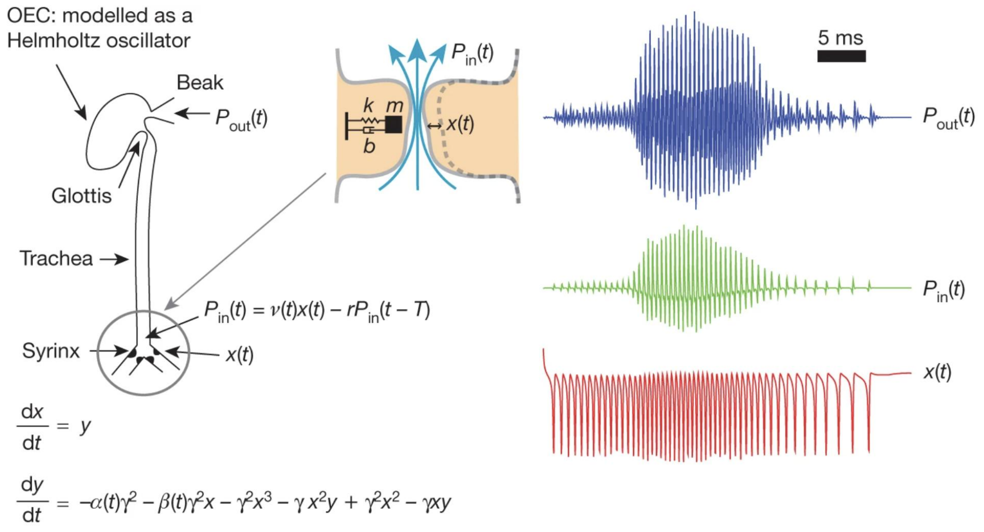
K: stiffness

B: dispersion term

P_s: interlabial pressure.
$F_{0:}$ force term that controls the labial's stationary position, usually takes 0

$$\dot{x} = y,$$

$$\dot{y} = -kx - \beta y + p_s$$



Interlabial pressure

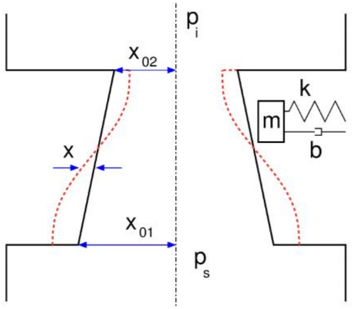
$$a_1 = a_{10} + x + \tau y,$$

$$a_2 = a_{20} + x - \tau y,$$

a_1: half the separation of upper edge

a_2: half the separation of lower edge

τ: the time taken for propagating wave to vertically travel half the distance of labia



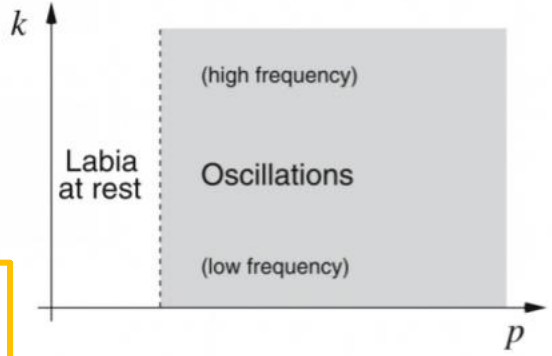
$$p_s = p_b\left(1 - \frac{a_2}{a_1}\right)$$

p_s: average pressure

p_b: sublabial pressure

Nonlinear dispersion

$$\boxed{\begin{aligned}\dot{x} &= y,\\ \dot{y} &= -kx - (\beta - p_b)y - cx^2 y\end{aligned}}$$



Self-oscillation when $p_b > \beta$

The Physics of birdsong - Mindlin’s model

Syrinx - updated:

Bifurcation and reduction to normal form

Fixed Point: in a dynamical where the system remains unchanged over time. It represents a point of equilibrium.

Stability: a stable fixed point will return to its original state after a small disturbance; an unstable one will diverge from it.

Bifurcation: in dynamics where the qualitative structure of a system is changed as the parameters are varied

Saddle Node Bifurcation: where 2 fixed point merge, collide and annihilate each other. Here it is called a saddle node in a limit cycle (SNILC) bifurcation

Hopf Bifurcation: where a system transitions from a stable fixed point to a stable limit cycle, leading to oscillatory behavior.

Bogdanov-Takens Bifurcation: a point in parameter space where both saddle node and Hopf bifurcations occur simultaneously.

Normal Form Reduction: way to simplify the equations around a bifurcation point to more easily analyze and predict the system's behavior near the bifurcation.

$$\dot{x} = y,$$

$$\dot{y} = -kx - \beta y - cx^2 y + p_b\left(\frac{\Delta a + 2\tau y}{a_{01} + x + \tau y}\right)$$

$$k = k_1 + k_2 x^2 \qquad \beta = \beta_1 + \beta_2 y^2$$

Another form more commonly used in later works:

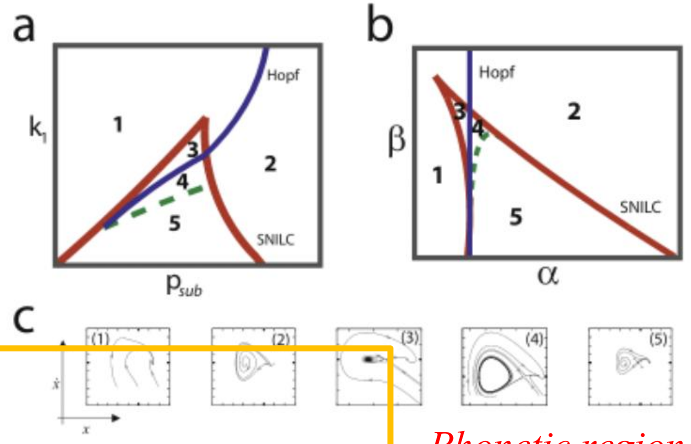
$$\frac{dx}{dt} = y,$$

$$\frac{dy}{dt} = (1/m)\left[-k(x)x - \beta(y)y - cx^2 y + a_{lab} p_{sub}\left(\frac{\Delta a + 2\tau y}{a_{01} + x + \tau y}\right)\right]$$

a_{lab}: lateral labial area

p_{sub}: subglottal pressure

$\Delta a = a_{01} - a_{02}$



$$\boxed{\begin{aligned}\frac{dx}{dt} &= y,\\ \frac{dy}{dt} &= -\gamma^2\alpha - \gamma^2\beta x - \gamma^2 x^3 + \gamma^2 x^2 - \gamma xy - \gamma x^2 y\end{aligned}}$$

Phonetic region: 2

γ: time scaling factor, α and β are associated with air sac pressure(p) and labial tension (k) respectively

The Physics of birdsong - Mindlin's model

Vocal Tract:

Trachea

The trachea is approximated as a tube that has

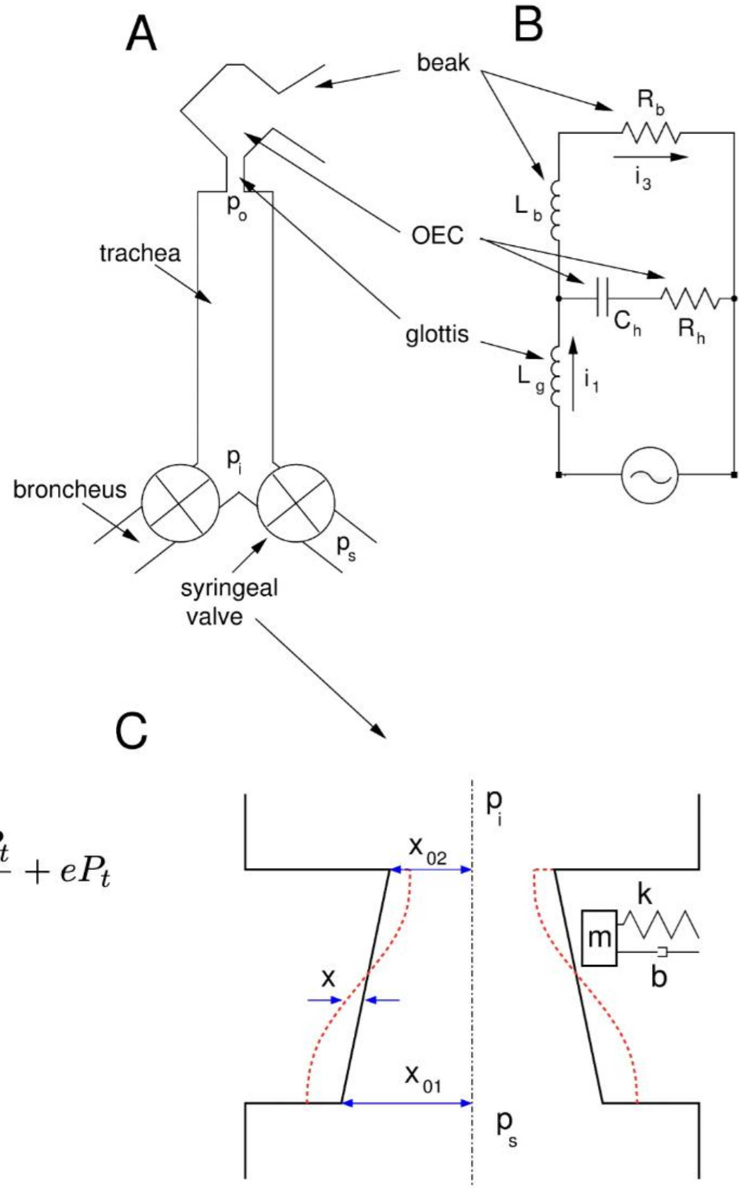
$$p_i(t) = \alpha_{env}(t)x(t) - rp_i(t - \frac{2L}{c})$$

$$p_o(t) = (1-r)p_i(t - \frac{L}{c})$$

r: reflection coefficient, L: length of trachea

c: velocity of sound in air

$\alpha_{env(t)}$: a coefficient proportional to the mean velocity of the flow, amplitude of sound to be matched or a constant in pure synthesizing



Oropharyngeal-esophageal cavity (OEC)

modelled as Helmhoz resonator and written into equivalent circuit as

$$\frac{di}{dt} = \Omega_1$$

$$\frac{d\Omega_1}{dt} = -\frac{1}{L_g C_h} i_1 - R_h \left(\frac{1}{L_b} + \frac{1}{L_g}\right)\Omega_1 + i_3\left(\frac{1}{L_g C_h} - \frac{R_b R_h}{L_b L_g}\right) + \frac{1}{L_g}\frac{dV_{ext}}{dt} + \frac{R_h}{L_g L_b} V_{ext}$$

$$\frac{di_3}{dt} = -\frac{L_g}{L_b}\Omega_1 - \frac{R_b}{L_b} i_3 + \frac{1}{L_b} V_{ext}$$

or

$$\frac{di_1}{dt} = \Omega_1$$

$$\frac{d\Omega_1}{dt} = ai_1 + b\Omega_1 + ci_3 + d\frac{dP_t}{dt} + eP_t$$

$$\frac{di_3}{dt} = f\Omega_1 + gi_3 + hP_t$$

V_{ext} is proportional to the pressure at the end of the trachea $p_o(t)$

The output of the beak is proportional to $V_3 = R_b i_3$ hence proportional to i_3.

Numerical implementation

Runge-Kutta 4th Order Method

A very concise version of Runge-Kutta 4th order method is used here, as introduced in relevant books for both dynamics and birdsong, and implemented in existing models:

$$\dot{x} = f(x)$$

$$x_{n+1} = x_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where:

$$k_1 = f(x_n)\Delta t$$

$$k_2 = f\left(x_n + \frac{1}{2}k_1\right)\Delta t$$

$$k_3 = f\left(x_n + \frac{1}{2}k_2\right)\Delta t$$

$$k_4 = f(x_n + k_3)\Delta t$$

The above can be easily written into a function that works the same for all ordinary differential equations in the model

Finite Difference Method

Finite difference basics:

Approximate u(t) using u^n at intervals t = nk, n is an integer and k is a set time step.

Definition of difference operators:

Unit shifts $\quad e_{t+}u_n = u_{n+1}, \quad e_{t-}u_n = u_{n-1}.$

Difference approximations

$$\delta_{t+} = \frac{e_{t+} - 1}{k}, \quad \delta_{t-} = \frac{1 - e_{t-}}{k}, \quad \delta_{t\cdot} = \frac{e_{t+} - e_{t-}}{2k}$$

$$\delta_{tt} = \frac{e_{t+} - 2 + e_{t-}}{k_2}$$

Averaging operators

$$\mu_{t+} = \frac{e_{t+} + 1}{2}, \quad \mu_{t-} = \frac{1 + e_{t-}}{2}, \quad \mu_{t\cdot} = \frac{e_{t+} + e_{t-}}{2}$$

Numerical implementation

Finite Difference Method

Developed by myself part by part

* this works here since the model uses source-filter separation assumption

Syrinx:

Rewrite the equation into

$$\frac{d^2x}{dt^2} = -\gamma(x + x^2)\frac{dx}{dt} - \gamma^2\alpha - \gamma^2\beta x - \gamma^2 x^3 + \gamma^2 x^2$$

Operator form (the cubic term x^3 is represented by $x^2\mu_{t}.x$):

$$\delta_{tt}x = -\gamma(x + x^2)\delta_{t}.x - \gamma^2\alpha - \gamma^2\beta x - \gamma^2 x^2\mu_{t}.x + \gamma^2 x^2$$

Update form:

$$\begin{aligned} x^{n+1} = &\frac{1}{k\gamma(x^n + (x^n)^2) + k^2\gamma^2(x^n)^2 + 2} \\ &\times\left(-2k^2\gamma^2\alpha(i) - 2k^2\gamma^2\beta(i)x^n + 2k^2\gamma^2(x^n)^2 + 4x^n\right. \\ &\left.+(-k^2\gamma^2(x^n)^2 + k\gamma x^n + k\gamma(x^n)^2 - 2)x^{n-1}\right) \end{aligned}$$

Trachea:

A set of delay lines $p_i^n = \alpha_{\text{env}}^n x^n - r p_i^{n-2L_d}, \quad p_o^n = (1-r)p_i^{n-L_d}$

where T is the sampling period, L_d=L/(cT) is the number of samples corresponding to the delay L/c.

OEC:

Wrap into state-space formalism:

$$\mathbf{x} = \begin{bmatrix} i_1 & \Omega_1 & i_3 \end{bmatrix}^T, \quad \mathbf{p} = \begin{bmatrix} \frac{dP_t}{dt} & P_t \end{bmatrix}^T$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{p}$$

$$\mathbf{y} = \mathbf{C}^T\mathbf{x}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ a & b & c \\ 0 & f & g \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ d & e \\ 0 & h \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

with trapezoid rule there is (in this model Re(eig(A))<0, unconditionally stable):

$$\mathbf{x}^{n+1} = \left(\mathbf{I} - \frac{k\mathbf{A}}{2}\right)\left(\left(\mathbf{I} + \frac{k\mathbf{A}}{2}\right)\mathbf{x}^n + \frac{k}{2}\mathbf{B}\left(\mathbf{p}^{n+1} + \mathbf{p}\right)\right)$$

Numerical implementation

Comparison

Within the given equations, the Runge-Kutta method boasts higher accuracy.

Yet, in synthesizers with built-in approximations, as far as both methods are able to give almost the same waveforms, we care more about computational efficiency.

Below compares 3 ways of using the Runge-Kutta scheme and the finite difference scheme:

A. All equations in one ODEs

B. ODEs 1 for syrinx, delayline for trachea, ODEs 2 for OEC

C. Delayline for trachea, ODEs for syrinx and trachea

All 'ODEs' above are solved by the same RK function – stability not tested (not tested in literature either)

D. Finite Difference Method – stable in the desired parameter range

TI: time-invariant alpha, beta

TV: time-varying alpha, beta

Step size: 1/(4*48000)

Run Time Comparison

	A	B	C	D
TI, 0.2s sound	0.619281s	0.764654s	0.543868s	0.063981s
TV, 0.5s sound	1.571319s	1.745776s	1.235279s	0.146235s

FDM 8.5 times faster than RK!

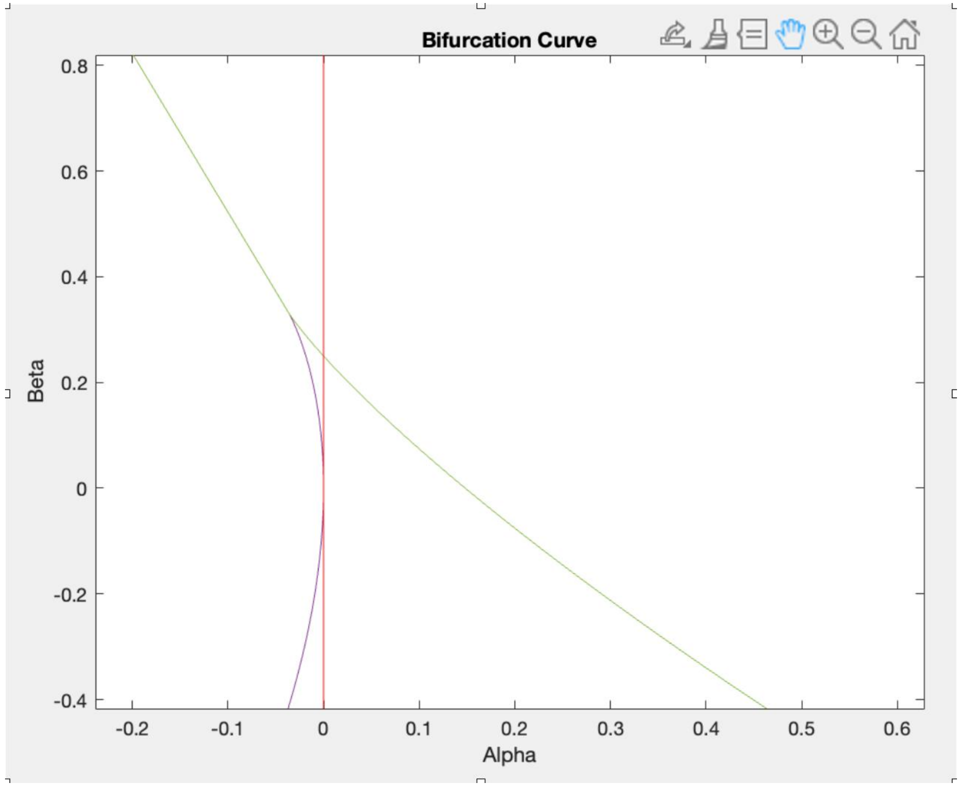
Could be accelerated even more when generating data in TI case with further vectorization!

Evaluation – Parameter Space

Here we take Zebra Finch as an example, gamma is found to be fixed to 24000

Control parameters are identified as alpha, beta, corresponding to air sac pressure and labial tension in the unnormalized form, phonetic region is found on the top right or the bifurcation plot

Bifurcation plot and SNILC calculation



Calculation:
linearize around equilibrium point:

$$\alpha = -\beta x + x^2 - x^3$$

the tangents of $f(x) = \alpha$ and $g(x) = -\beta x + x^2 - x^3$ are equal which gives:

$$\beta = 2x - 3x^2$$

In coding we find the **real** x roots first then plot corresponding alpha, beta

Or with further calculation to eliminate x:

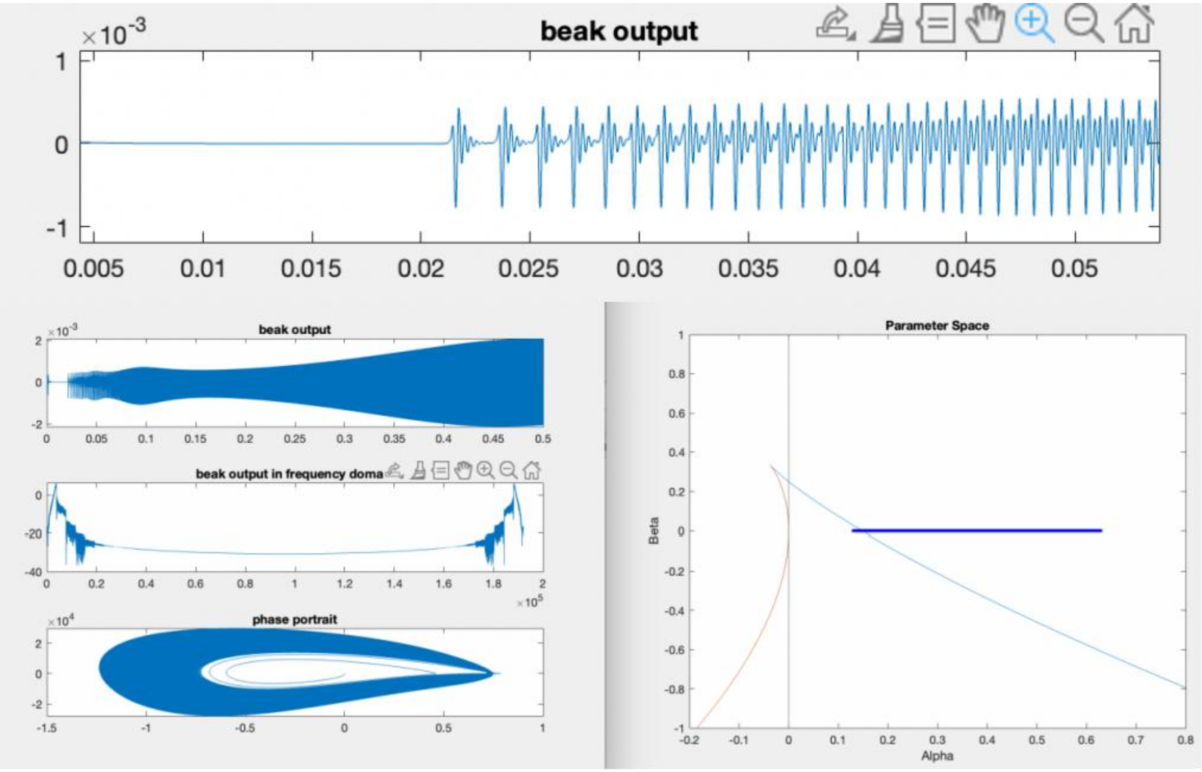
$$a = -\frac{1}{27}\left(-2 + 9\beta \pm (-2 + 6\beta)\sqrt{1 - 3\beta}\right)$$

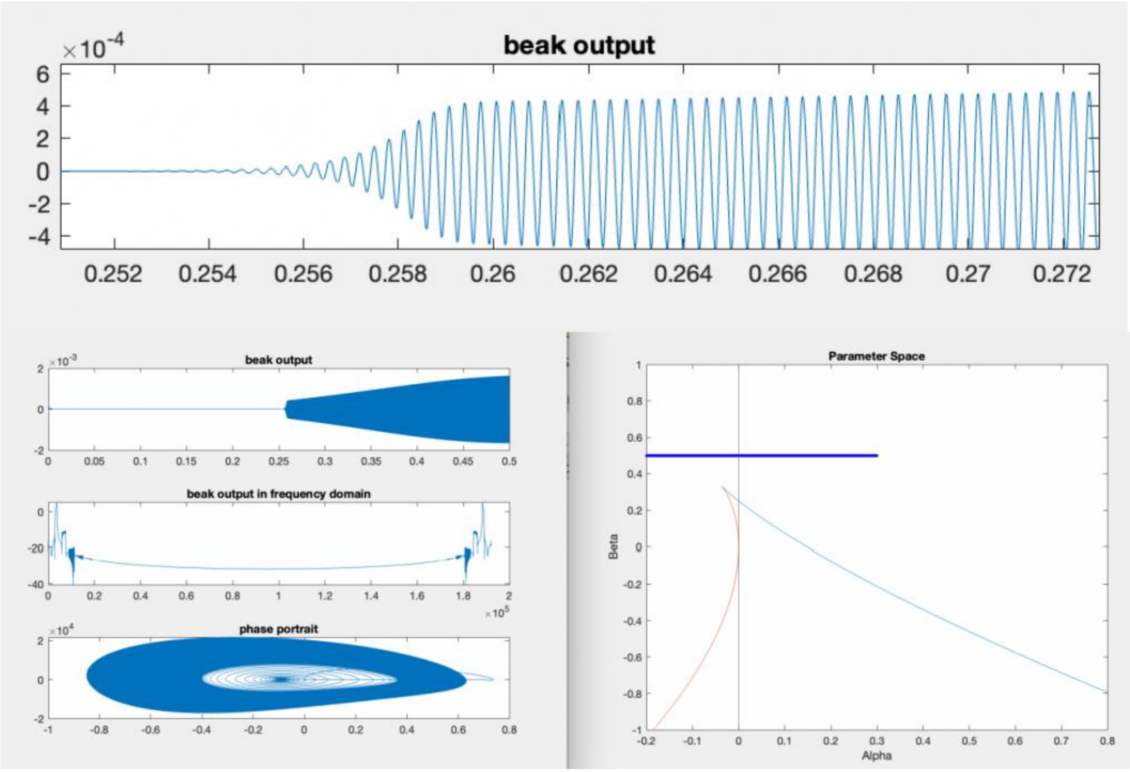
Evaluation - Waveform

Waveforms identical to those in literature

Left: Born in SNILC bifurcation spike-like wave with infinite period and spectrally rich

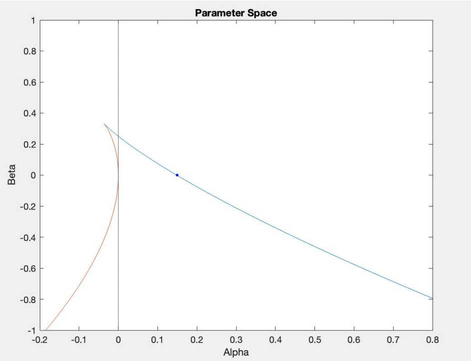
Right: Born in Hopf bifurcation Tonal sound with a defined frequency and zero amplitude

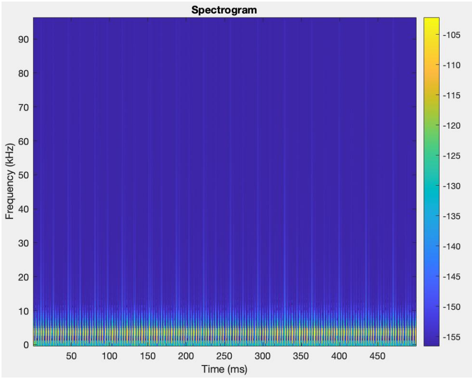


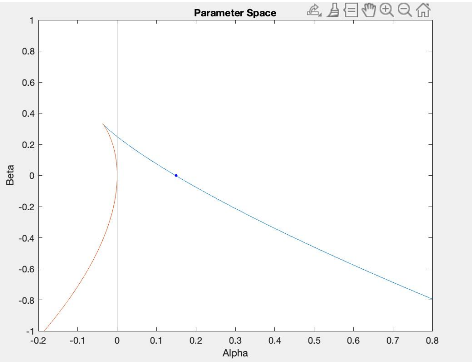


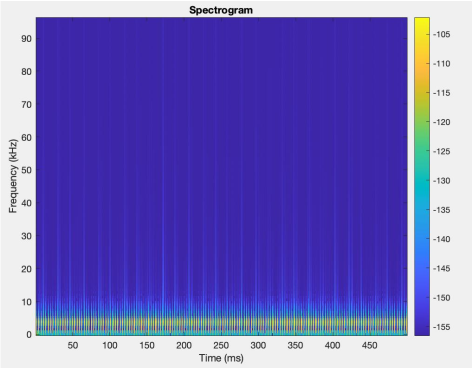
Evaluation - Sound

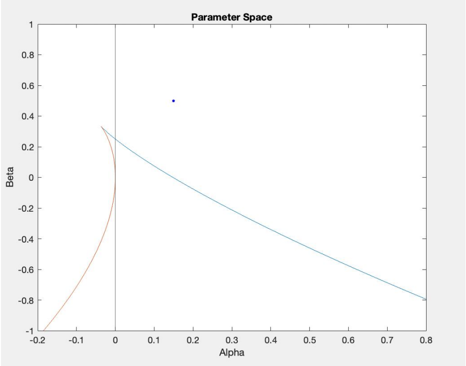
Other sound samples

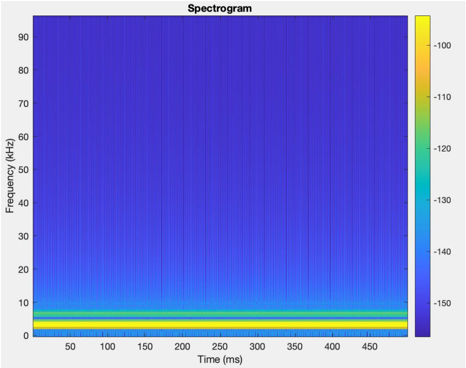


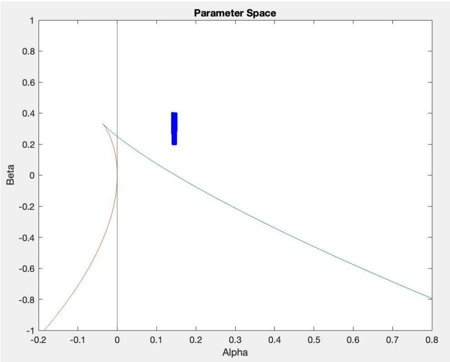


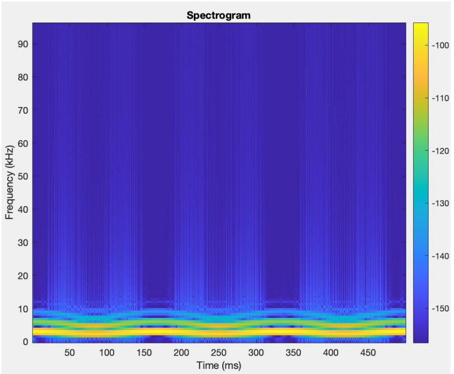


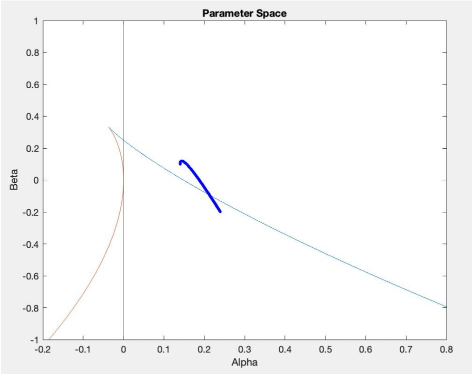


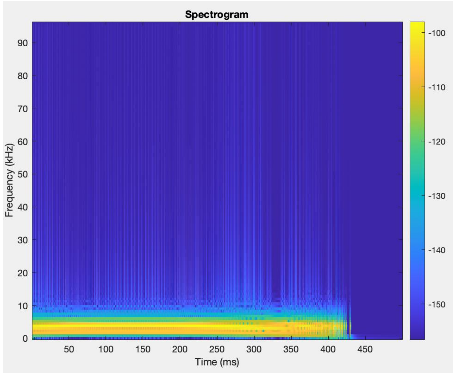












Evaluation – Perceptual Mapping

Playability?

Introduce the following perceptual parameters:

f_0: *fundamental frequency*

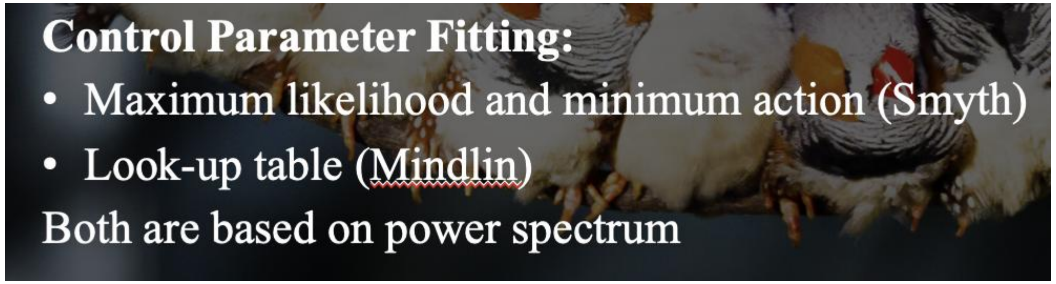
SCI: spectral richness index calculated by Spectral Centroid / Fundamental Frequency

where Spectral Centroid = $\frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$

Calculation of f_0

Experimented with methods including autocorrelation, YIN etc, in birds' case it's the simplest first peak of the power spectrum / frequency plot that gives the best result, same as methods in literature

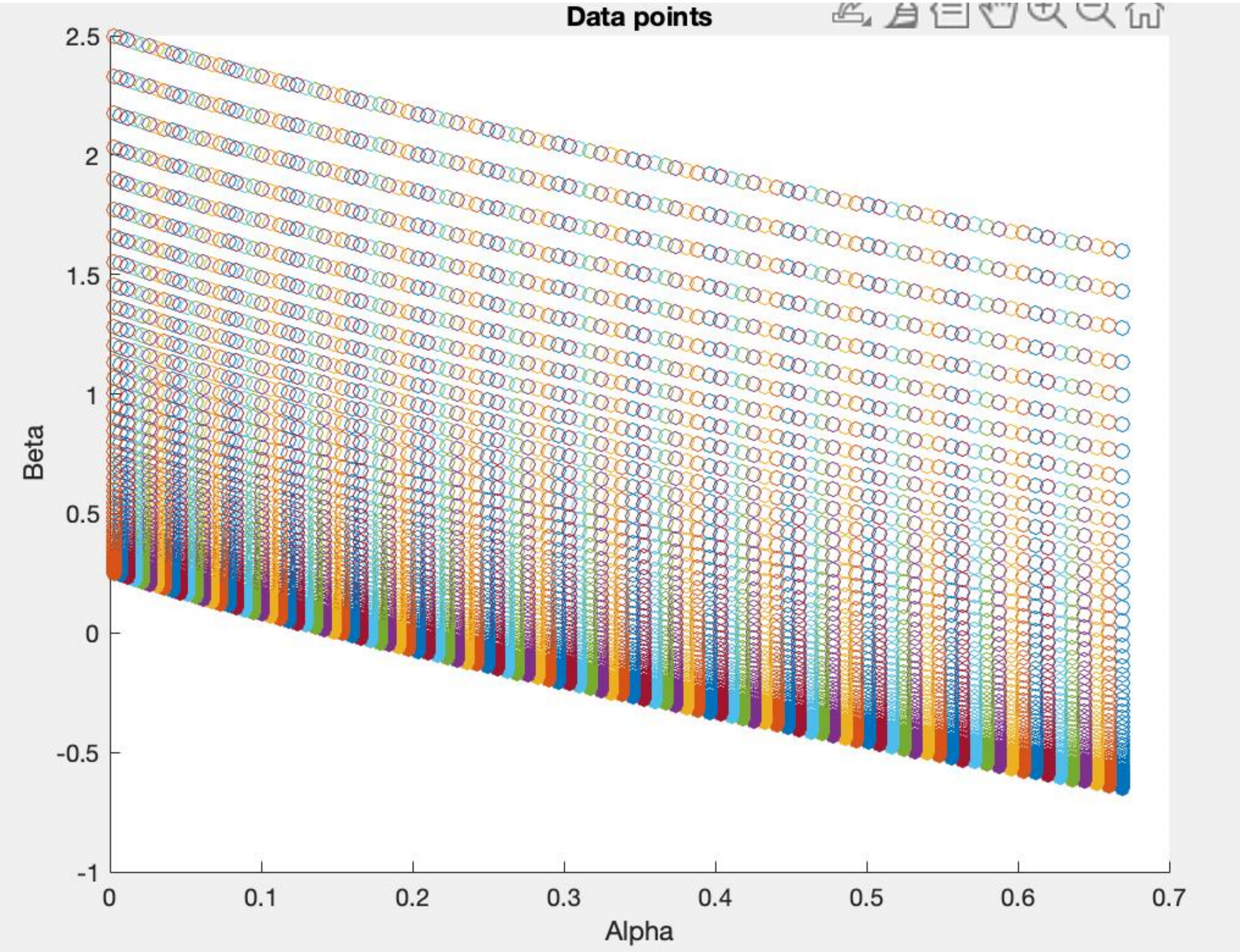
Range: 375Hz ~ 6047Hz



Evaluation – Perceptual Mapping

Visualization

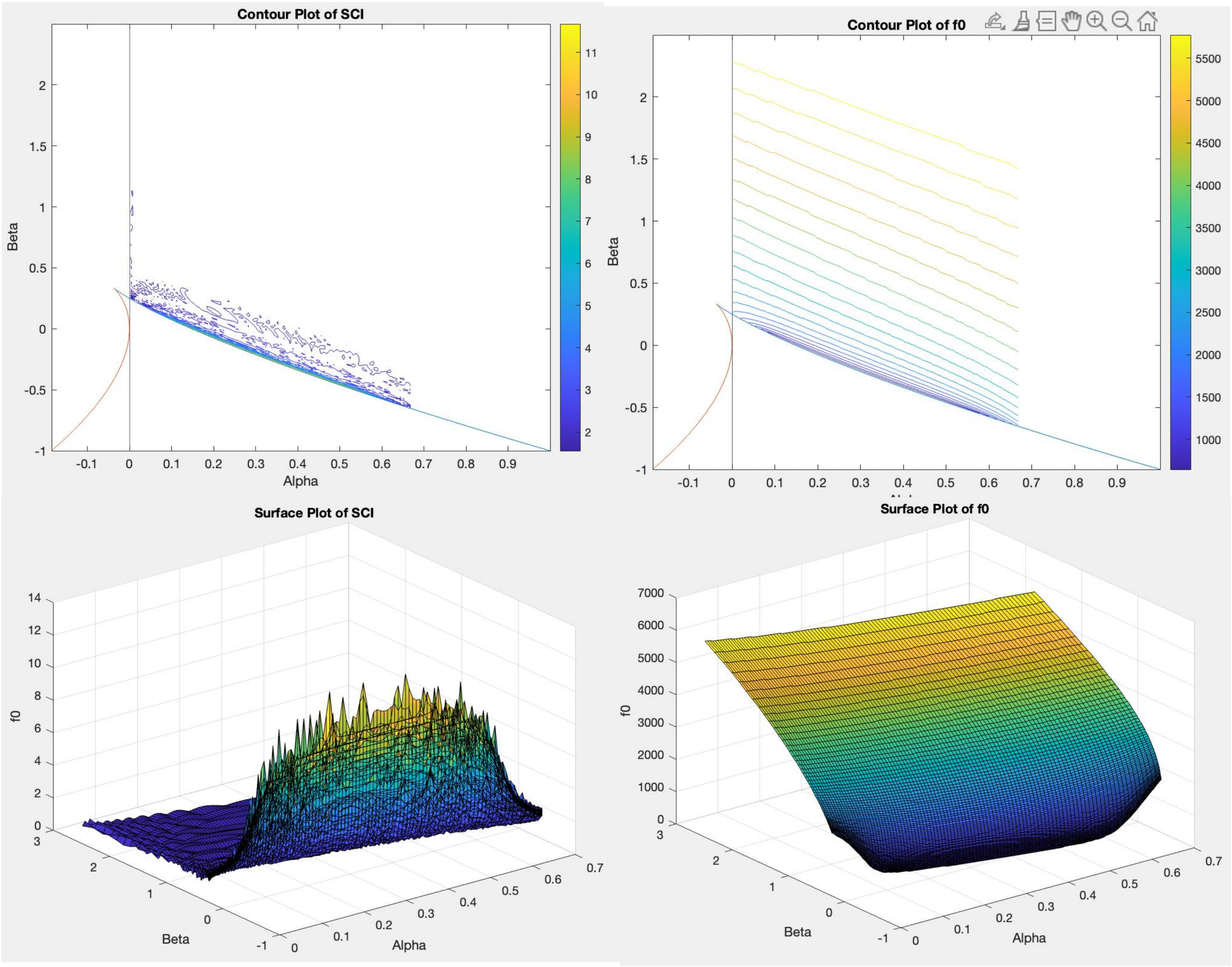
alpha, beta



Evaluation – Perceptual Mapping

Visualization

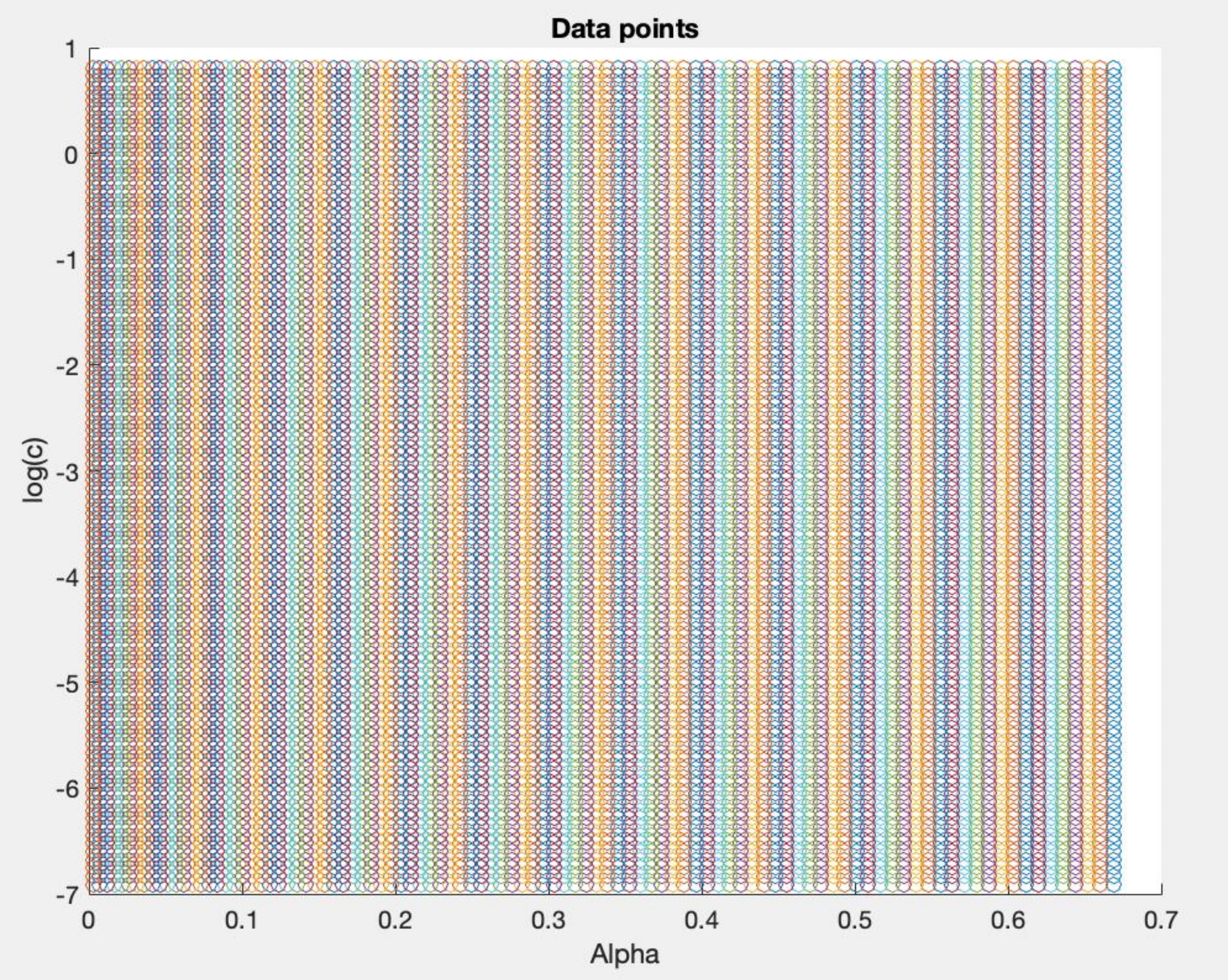
alpha, beta – SCI, f0



Evaluation – Perceptual Mapping

Visualization

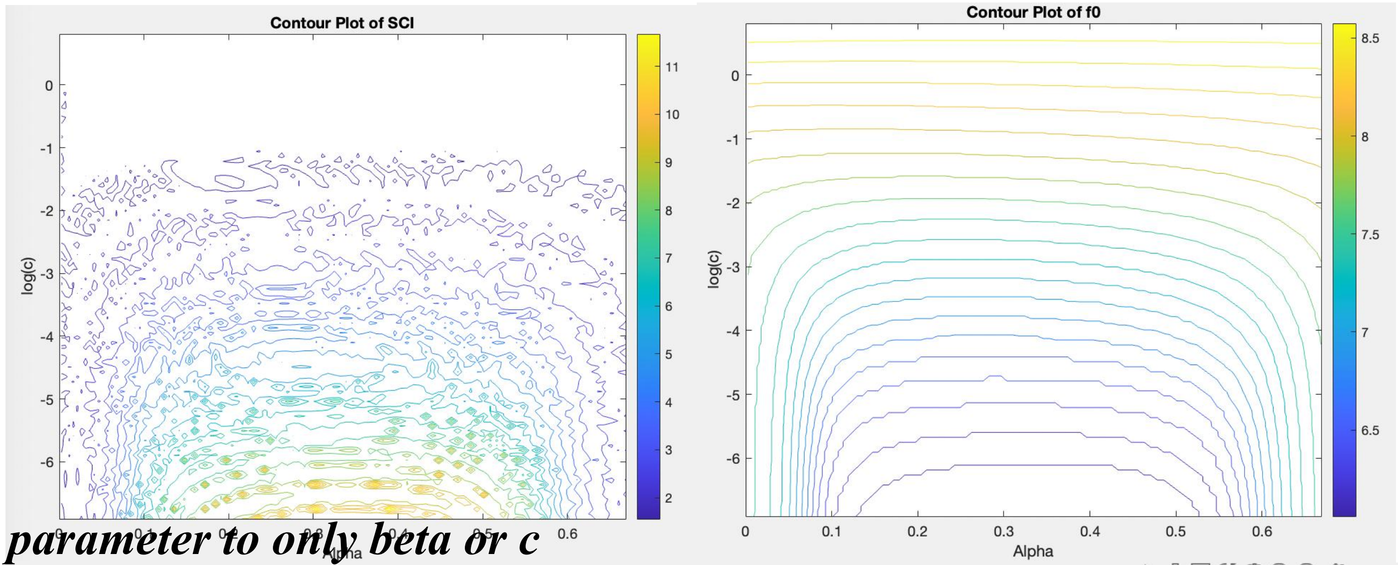
Remap beta to c that gives equally spaced alpha, log(c)



Evaluation – Perceptual Mapping

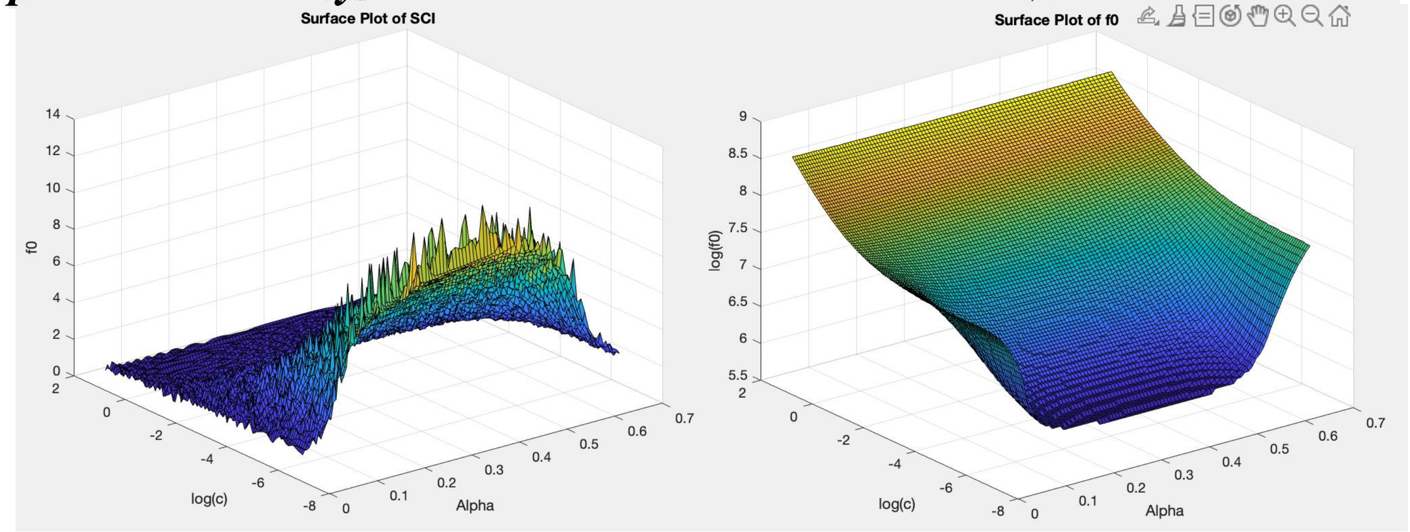
Visualization

Remap beta to c that gives equally spaced alpha, log(c)



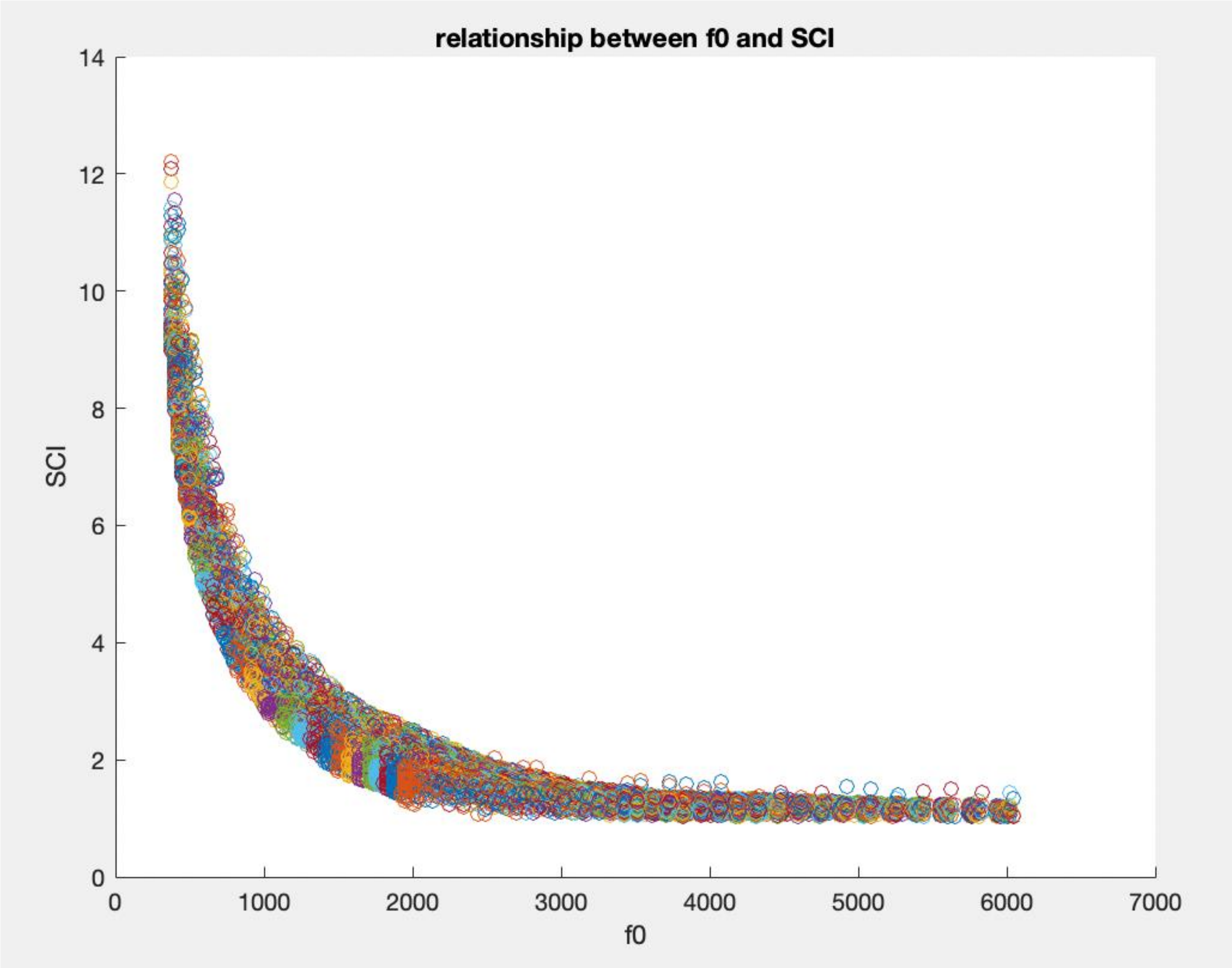
Correlation in f0 and SCI

Possible to reduce control parameter to only beta or c



Evaluation – Perceptual Mapping

Correlation



Evaluation – Reconstruction

f0 - Beta (c)

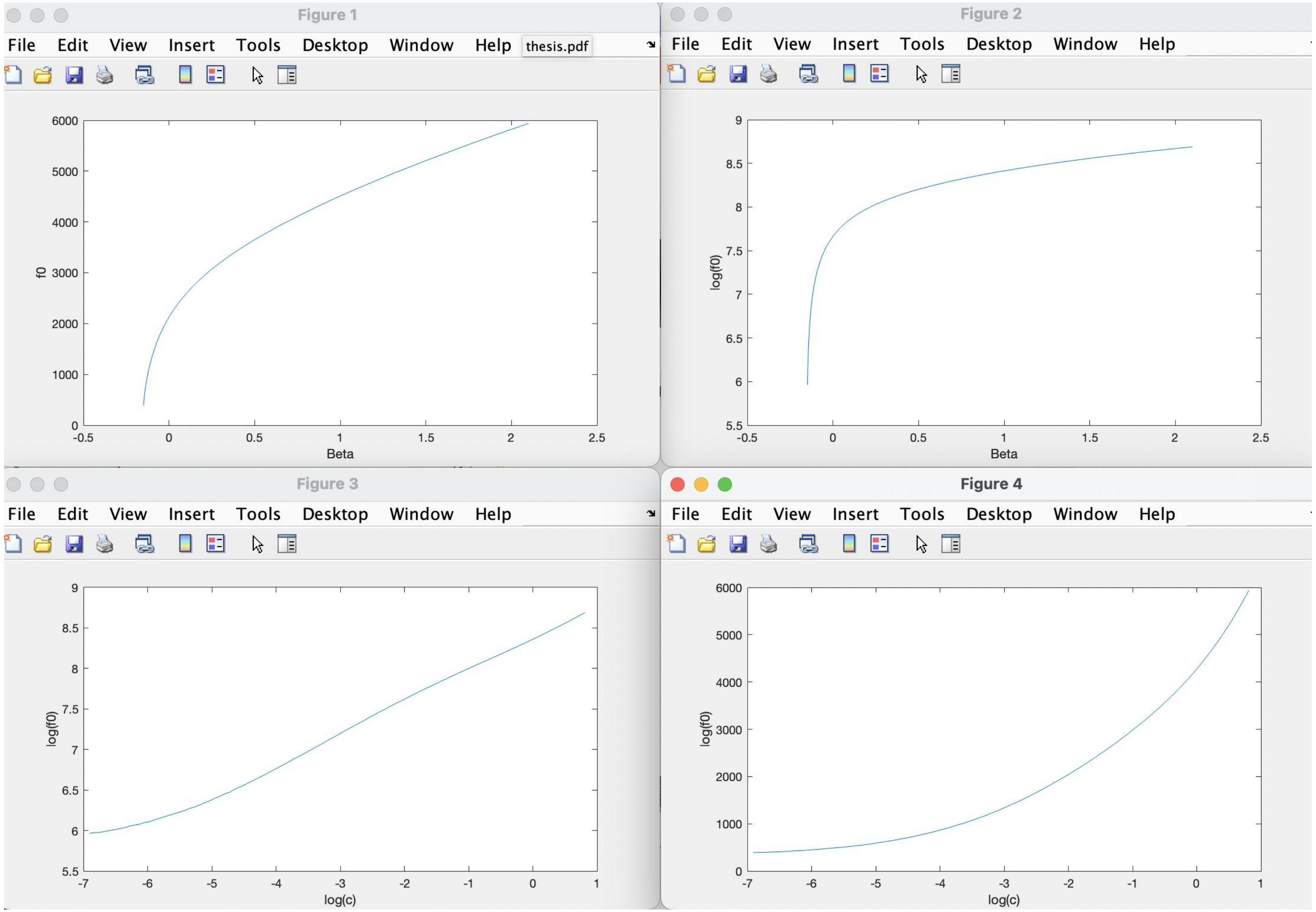
Methods:

look-up table

Interpolation – linear interpolation between each pair of neighbours

Machine learning – Support Vector Machine

(to be finalized – see dissertation for details)



Conclusion

- Successfully built a computational birdsong model in Matlab.
- Introduced an efficient finite difference scheme, outperforming existing methods.
- Discovered novel method to unveil correlations in the parameter space.
- Introduced the use of interpolation and machine learning for model inversion.

Future work

- Refine the physical model for better generalization
- Increase diversity (2 voices, more parameters, more species etc.)
- Calculate and compare with analytical frequency   solution
- Real-time implement in MaxMSP

Thank you!

Key References

Strogatz, Steven H. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And*

Mindlin, G. B., and Rodrigo Laje. *The Physics of Birdsong*. Biological and Medical Physics, Biomedical Engineering. Berlin: Springer, 2005.

Smyth, Tamara, and Julius O Smith Iii. 'THE SOUNDS OF THE AVIAN SYRINX – ARE THEY REALLY FLUTE-LIKE?', 2002.

Trevisan, M. A., and G. B. Mindlin. 'New Perspectives on the Physics of Birdsong'. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 367, no. 1901 (2009): 3239–54.

Amador, Ana, and Gabriel B. Mindlin. 'Low Dimensional Dynamics in Birdsong Production'. *The European Physical Journal B* 87, no. 12 (10 December 2014): 300. https://doi.org/10.1140/epjb/e2014-50566-5.

Boari et al., 'Automatic Reconstruction of Physiological Gestures Used in a Model of Birdsong Production'.

Smyth, Tamara, Jonathan Abel, and Julius Smith. 'THE ESTIMATION OF BIRDSONG CONTROL PARAMETERS USING MAXIMUM LIKELIHOOD AND MINIMUM ACTION', 1 January 2003.

Hsu, Jennifer, Tamara Smyth, and Third Author. 'Specifying Sounding Frequency of a Voice Model during Live Interactive Saxophone Performance', n.d.
Engineering. Reading: Westview Press, 2008. Print.