Vocal development in marmoset monkeys: neuromechanics and social interactions

SIAM-DS 19, Snowbird, Utah, May 19, 2019.
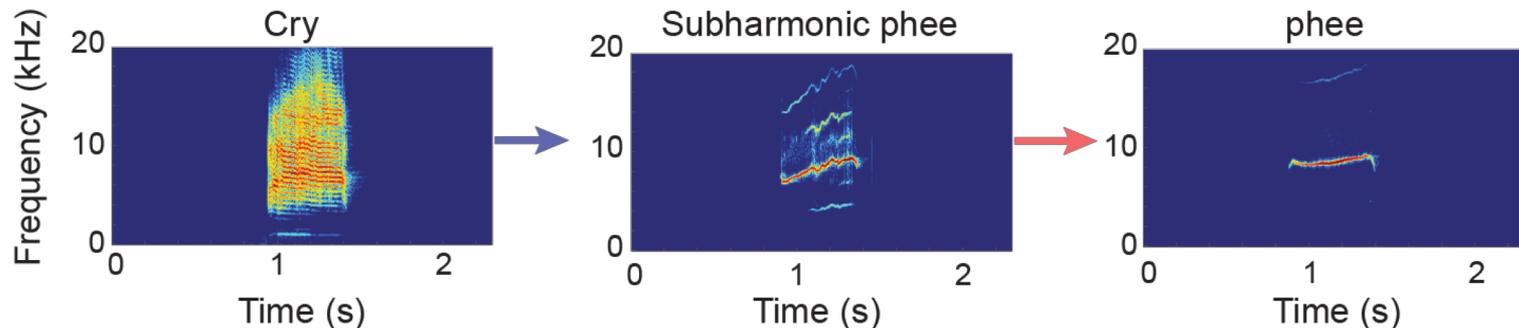
Yayoi Teramoto[1], Daniel Y. Takahashi[1,2], Philip Holmes[1,3] and Asif A. Ghazanfar[1,2,4]

[1]Princeton Neuroscience Institute, Princeton University;
[2]Department of Psychology, Princeton University;
[3]Program in Applied and Computational Mathematics and Department of Mechanical and Aerospace Engineering, Princeton University;
[4]Department of Ecology and Evolutionary Biology, Princeton University.

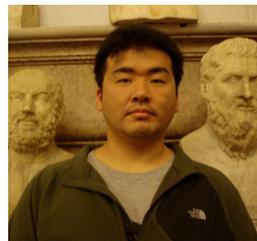D.Y. Takahashi et al. *Science 2015: 349, 734-738* and Y. Teramoto et al. *eLife 2017: 6, e20783.*
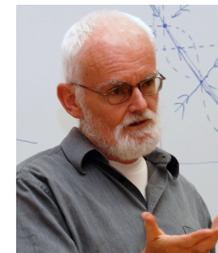
Contents

Yayoi Teramoto   Daniel Takahashi   Asif Ghazanfar   P. H.

I. Marmosets grow like human families, but 12x faster.

Marmosets* are small new world monkeys (approx 400 g as adults). Babies are raised by their parents, with much social interaction. They are considered infants for 2-3 months, during which their **broad spectrum cries and subharmonic calls** transition into mature-sounding **phee calls**. The timing of this transition is influenced by vocal feedback from parents.

Parents and twin infants.   Thanks to D.Y. Takahashi.

Callithrix jacchus.

mostly
cries

cries +
subharm
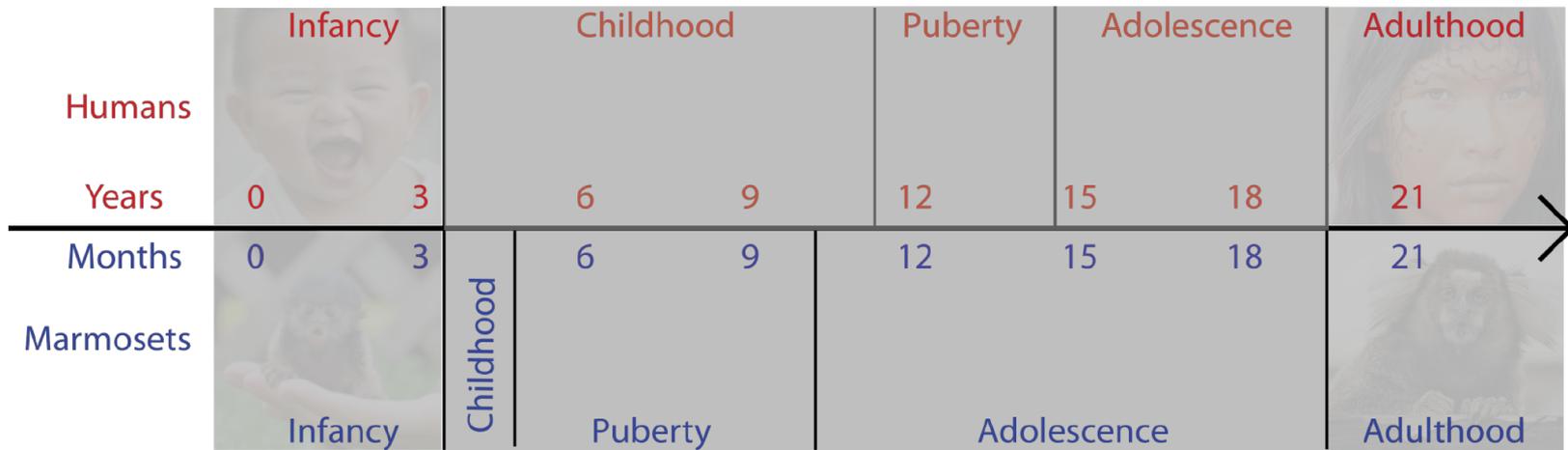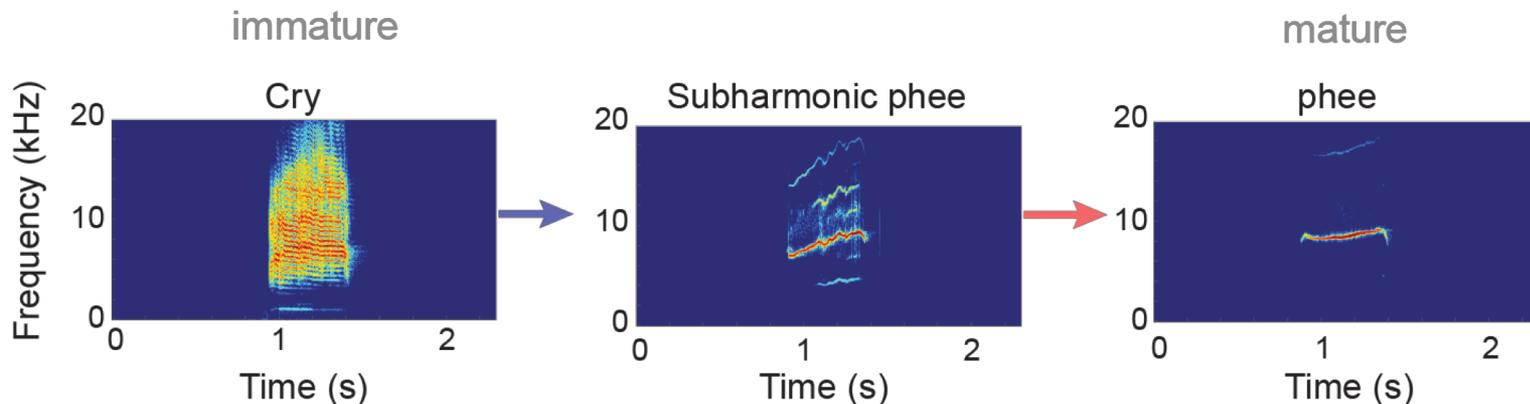
mostly
phees

We construct a model of the larynx and supra-glottal tract, and use it first to study the effects of growth on vocalizations. We find that it cannot predict the transition from cries to phees, so we introduce a **cost function** by adding factors accounting for **muscle and nervous system development**, and for **interactions with parents**. We thereby create a **developmental landscape** on which the animals evolve.

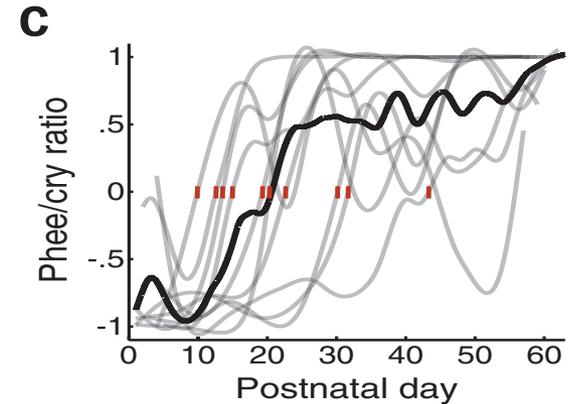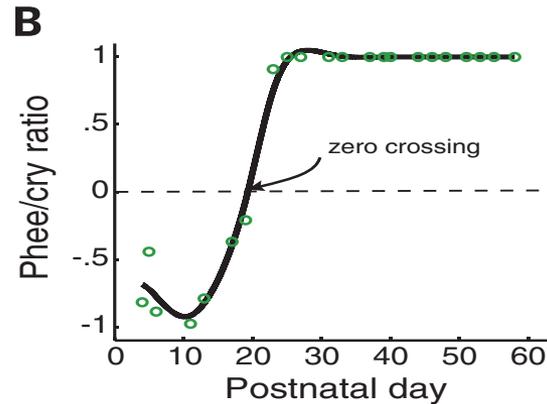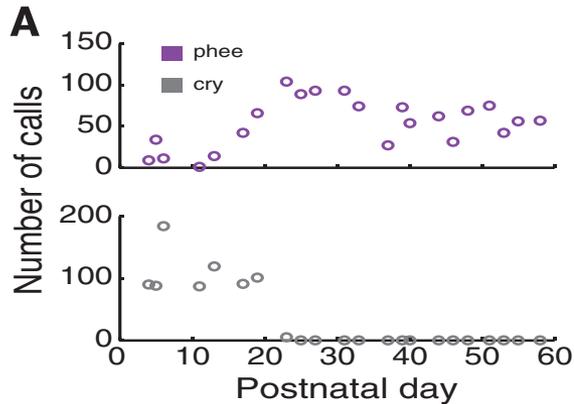# Marmosets provide a good model for early development in humans.

	Infancy	Childhood		Puberty	Adolescence		Adulthood
Humans							
Years	0 3	6	9	12	15	18	21
Months	0 3	6	9	12	15	18	21
Marmosets							
	Infancy	Childhood	Puberty		Adolescence		Adulthood

We focus on the development of **adult phee** calls from newborns' **broad spectrum cries and subharmonic** calls during the first 60 days of infancy. These changes are partly due to **biomechanical growth of the vocal apparatus**, but are also substantially influenced by **neural development** and **learning from parents**.

immature → mature
Cry → Subharmonic phee → phee

Call types are determined by frequency content in **spectrograms**: red = loud, blue = quiet
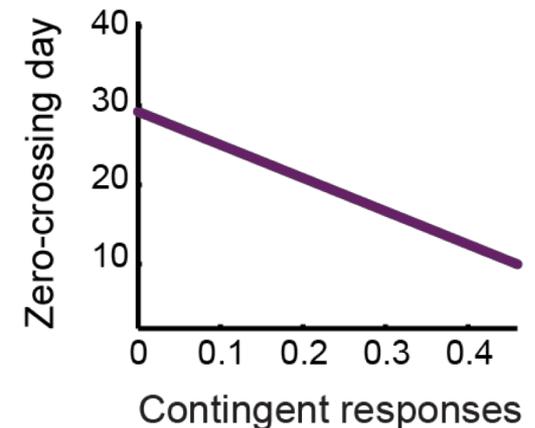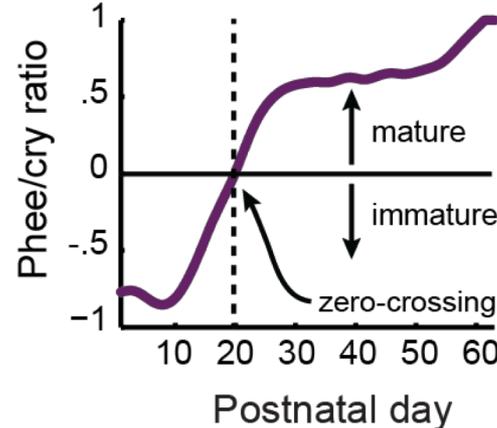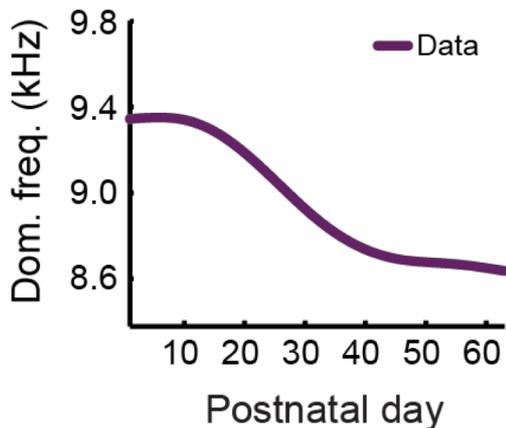
Early development is gauged by plotting dominant frequencies and phee/cry ratios vs. postnatal day. Data shown at top was collected from infant marmosets.

A

B

C
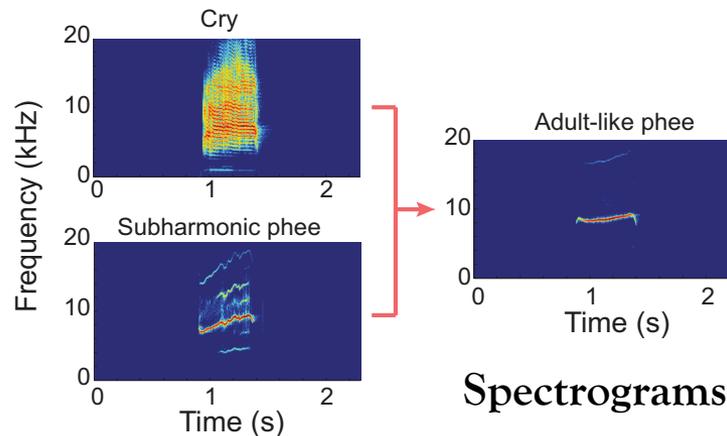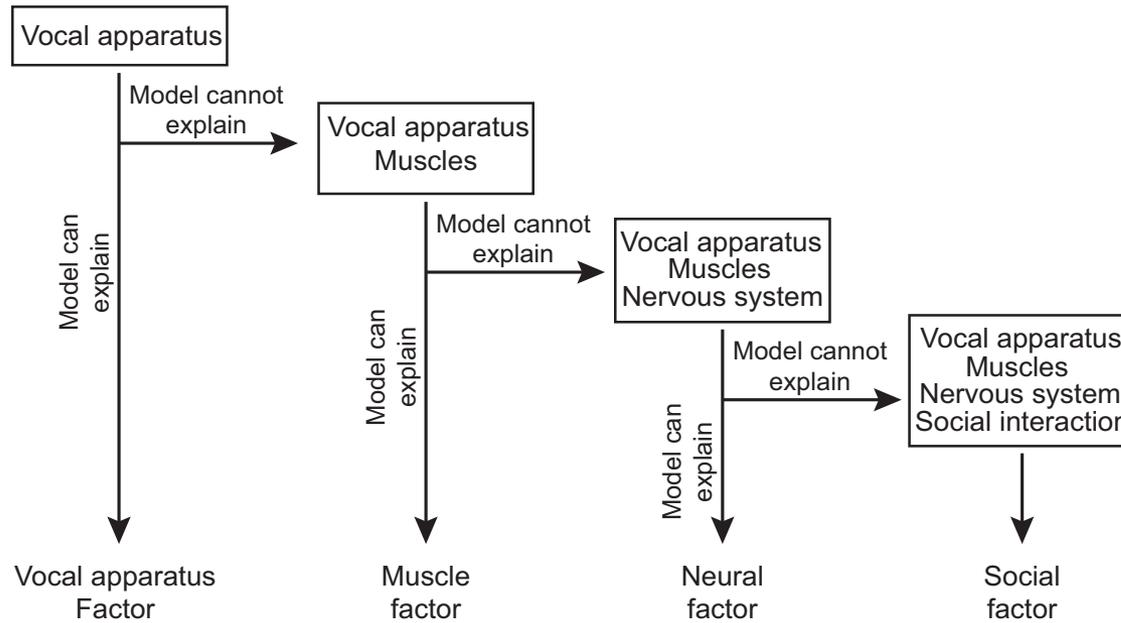
A and B show data for one infant;
C shows data for 10 infants.

$$\text{phee/cry ratio} = \frac{\text{Prob(phee)} - \text{Prob(cry)}}{\text{Prob(phee)} + \text{Prob(cry)}}.$$

Cartoons appear below. Zero crossing day depends on interaction with parents.
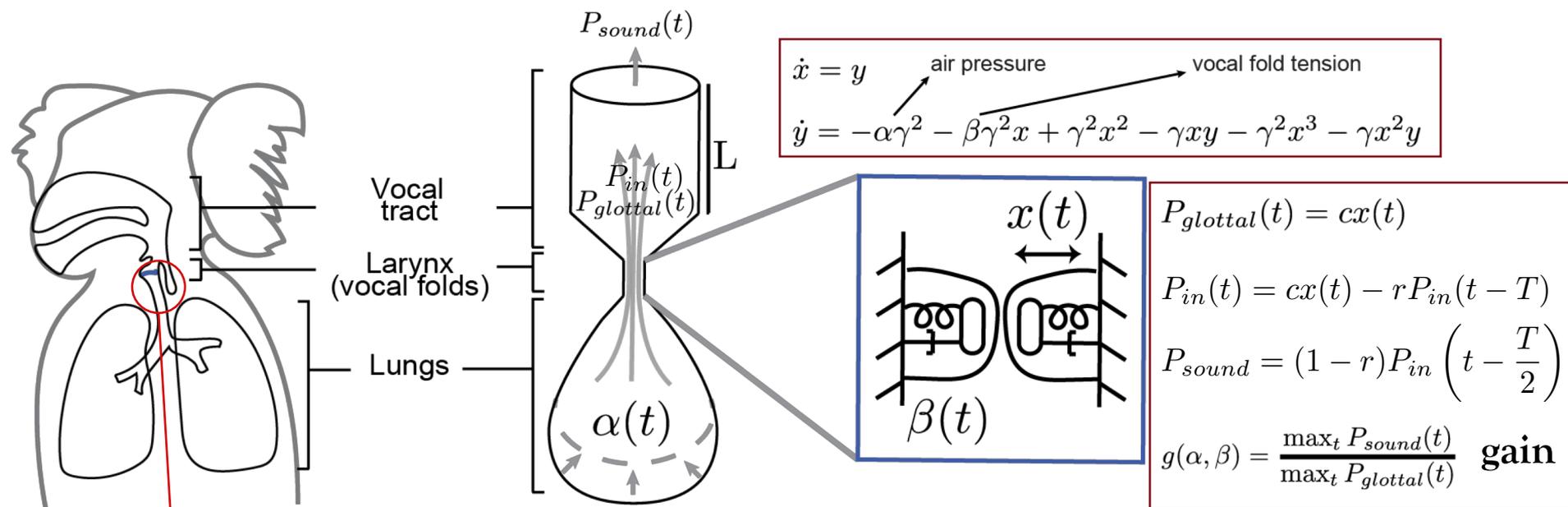
The main idea: start with biomech elements of the vocal apparatus, then add developing muscles, nervous system, and social interactions to take each factor into account.

Vocal apparatus

Model cannot explain → Vocal apparatus Muscles

Model can explain ↓

Vocal apparatus Factor

Model cannot explain → Vocal apparatus Muscles Nervous system

Model can explain ↓

Muscle factor

Model cannot explain → Vocal apparatus Muscles Nervous system Social interaction

Model can explain ↓

Neural factor

Model can explain ↓

Social factor

Cry

Frequency (kHz)

Subharmonic phee

Adult-like phee

Time (s)

Spectrograms

. . . etc. (iterate the process).
We use **simple linear models,** except for the **nonlinear vocal apparatus,** described next.

II. Biomechanical model of the larynx and vocal tract.

$P_{sound}(t)$

$$\dot{x} = y \quad \text{air pressure} \qquad \text{vocal fold tension}$$
$$\dot{y} = -\alpha\gamma^2 - \beta\gamma^2 x + \gamma^2 x^2 - \gamma xy - \gamma^2 x^3 - \gamma x^2 y$$

Vocal tract

Larynx (vocal folds)

Lungs

$P_{in}(t)$
$P_{glottal}(t)$

L

$x(t)$

$\alpha(t)$

$\beta(t)$

$$P_{glottal}(t) = cx(t)$$
$$P_{in}(t) = cx(t) - rP_{in}(t - T)$$
$$P_{sound} = (1 - r)P_{in}\left(t - \frac{T}{2}\right)$$
$$g(\alpha, \beta) = \frac{\max_t P_{sound}(t)}{\max_t P_{glottal}(t)} \quad \textbf{gain}$$

P_m

x_{02}

a_2, P_2

a_g

a_1, P_1

x

x_{01}

P_L

The biomechanical model is a nonlinear oscillator:

$$\dot{x}_1 = x_2,$$
$$\dot{x}_2 = \frac{1}{m}\left[-(k_1 + k_2 x_1^2)x_1 - (b_1 + b_2 x_1^2)x_2 + \frac{a_g P_L(x_{01} - x_{02} + 2\tau x_2)}{k_t(x_{01} + x_1 + \tau x_2)}\right].$$

We use a vocal tract model for mammals due to Titze. Non-dimensionalization and nonlinear coordinate changes produce the third order **normal form** (top), proposed earlier by Perl, Amador and Mindlin to model zebra finch song.

Y. Teramoto et al. *eLife 2017:* 6, e20782, after
I.R. Titze, *J. Acoustical Soc. Amer. 1988: 1536-1552*.

Y.S. Perl et al. *Phys. Rev. E. 2011: 84, 051909*.
A. Amador et al. *Nature 2013: 495, 59-64*.

Normal form transformations via Lie brackets.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \sum_{j=2}^{n} \mathbf{f}_j(\mathbf{x})$$

Let $\mathbf{x} = \mathbf{y} + \mathbf{P}(\mathbf{y}) := \mathbf{y} + \sum_{j=2}^{n} \mathbf{P}_j(\mathbf{y})$

$$\Rightarrow \dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \sum_{j=2}^{n} \mathbf{g}_j(\mathbf{y}) + \mathcal{O}(|\mathbf{y}|^{n+1})$$

J. Guckenheimer & P.H. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields.* Springer, NY, 1983.
Y.A. Kuznetsov. *Elements of Applied Bifurcation Theory,* Springer, NY, 1995.

Choose $\mathbf{P}_j(\mathbf{y})$ to make $\mathbf{g}_j(\mathbf{y})$ simpler:

$\mathbf{g}_2(\mathbf{y}) = \mathbf{A}\mathbf{P}_2(\mathbf{y}) - \mathbf{D}\mathbf{P}_2(\mathbf{y})\mathbf{A}\mathbf{y} + \mathbf{f}_2(\mathbf{y})$, etc. The biomechanical model

$$
\begin{aligned}
\dot{x}_1 &= x_2, \\
\dot{x}_2 &= \frac{1}{m}\left[-(k_1 + k_2 x_1^2)x_1 - (b_1 + b_2 x_1^2)x_2 + \frac{a_g P_L (x_{01} - x_{02} + 2\tau x_2)}{k_t (x_{01} + x_1 + \tau x_2)} \right].
\end{aligned}
$$

transforms, and expanding in a Taylor series about the **Takens-Bogdanov bifurcation point**, we can remove all but the following terms at second and third orders:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix};\ \mathbf{g}_2^a = \begin{pmatrix} 0 \\ y_1^2 \end{pmatrix},\ \mathbf{g}_2^b = \begin{pmatrix} 0 \\ y_1 y_2 \end{pmatrix};\ \mathbf{g}_3^a = \begin{pmatrix} 0 \\ y_1^3 \end{pmatrix},\ \mathbf{g}_3^b = \begin{pmatrix} 0 \\ y_1^2 y_2 \end{pmatrix}.$$

Unfolding this bifurcation point gives the cubic order Normal Form as in the previous slide:

$$
\begin{aligned}
\dot{y}_1 &= y_2 \\
\dot{y}_2 &= -\alpha\gamma^2 - \beta\gamma^2 y_1 + \gamma^2 y_1^2 - \gamma y_1 y_2 - \gamma^2 y_1^3 - \gamma y_1^2 y_2.
\end{aligned}
$$

Bifurcations of the reduced normal form model.

Phase portraits in open regions of the bifurcation set are shown.

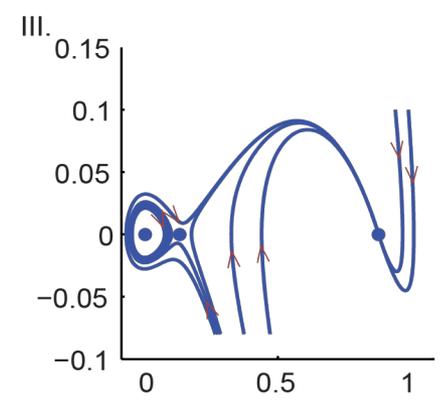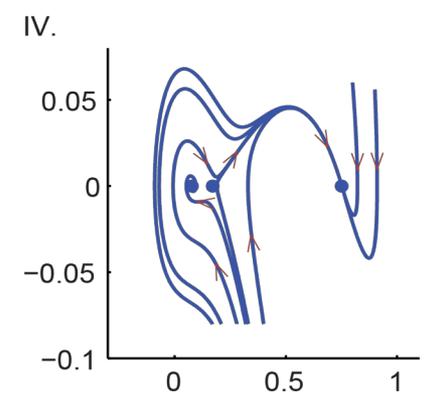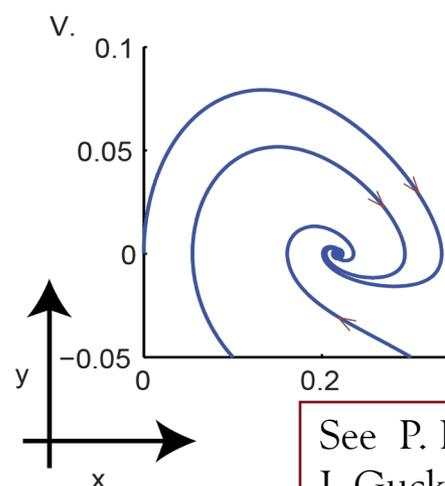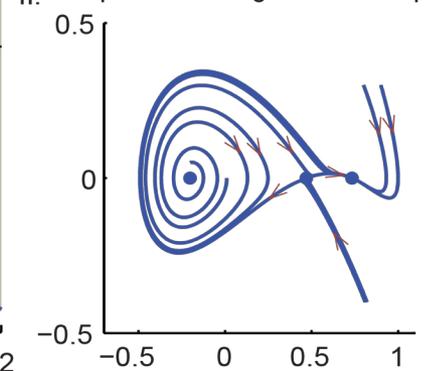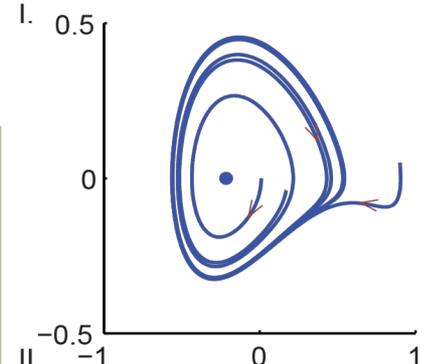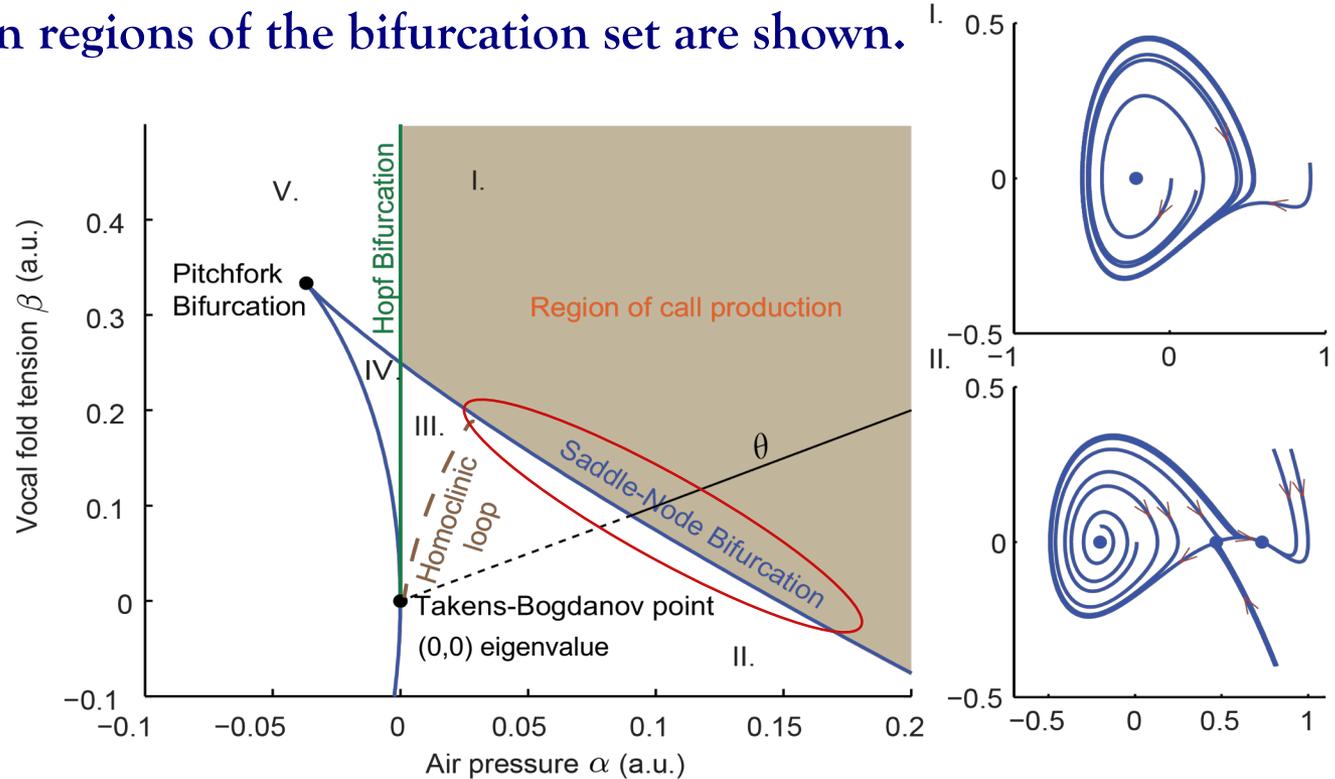Region I: Calls – stable limit cycle and source.

Region II: source, saddle, sink.

Region III: source, small stable limit cycle, saddle, sink.
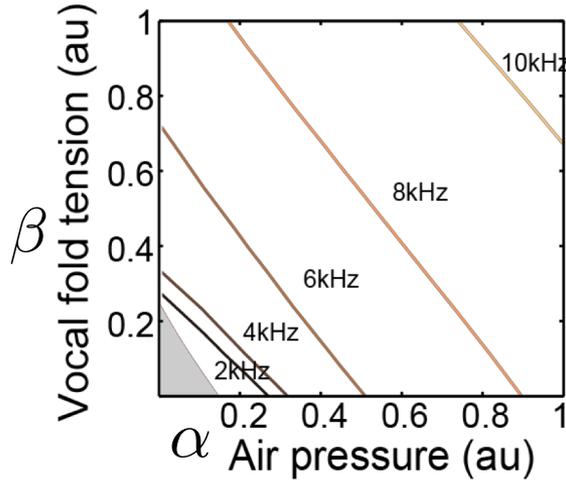
Region IV: 2 sinks, 1 saddle.

Region V: 1 sink.

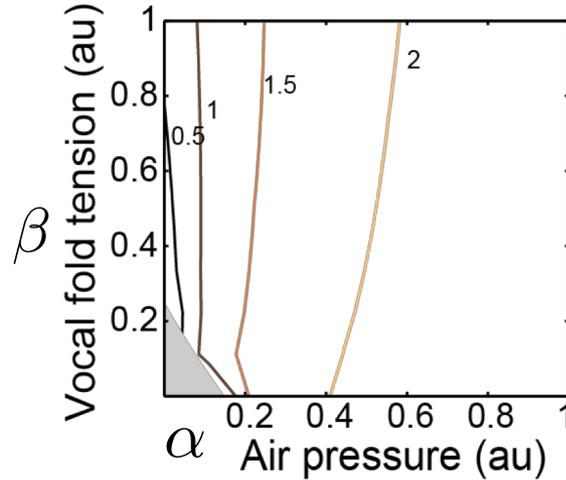Sensitive behavior occurs above and near the upper SN bifurcation curve, in Region I.

See P. H. and D. Rand, *Quart. Appl. Math. 1978: 35, 495-509*; J. Guckenheimer and P. H., *Nonlinear Oscillations, etc. 1983.*

Two parameters (α, β) suffice to produce a range of calls; for simplicity, we reduce to one parameter, θ.

Iso-frequency curves

Iso-amplitude curves

Gain = $P(t)_{sound}/P(t)_{glottal}$

These figures show maps of behavioral outputs in the bifurcation parameter plane.

Marmoset calls optimize

↑ Distance

↑ Efficiency

$$g(\theta) = \frac{\max_t P_{sound}(t)}{\max_t P_{glottal}(t)}$$

$$C(\theta) = -\log g(\theta)$$

The biomechanical model's growing vocal tract matches the dominant frequency, using a closed/open tube model to calculate the fundamental (first) resonance.

How the model makes cries and phee calls.

Slow and fast limit cycles respectively produce weak, decaying resonant oscillations and strong resonant oscillations in the supraglottal vocal tract (P_{sound}). Former occur in the **sensitive region of parameter space near the upper SN bifurcation curve** (slow passage).
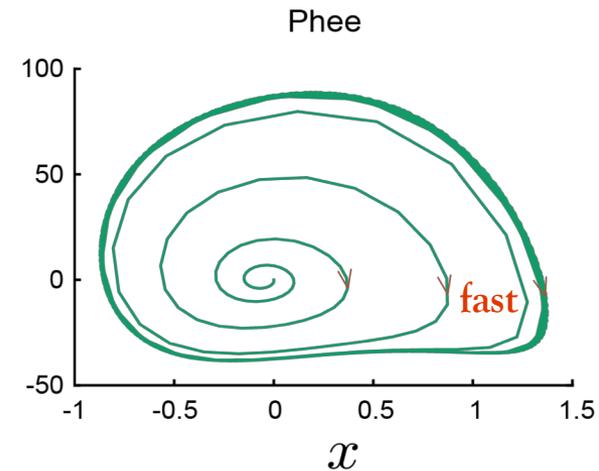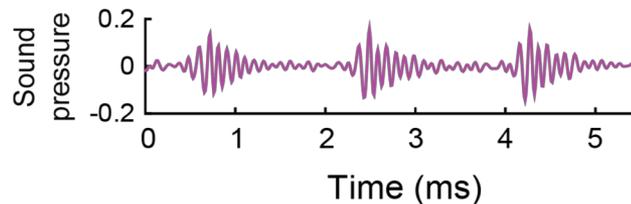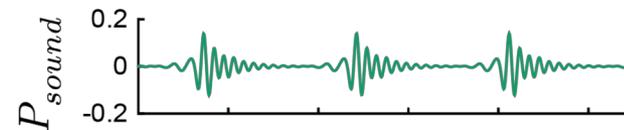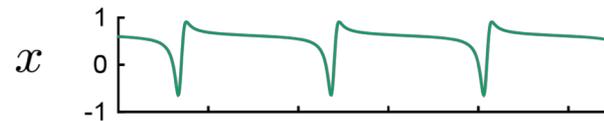
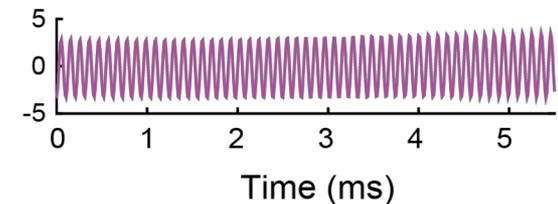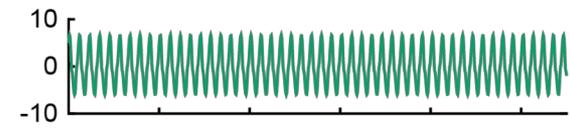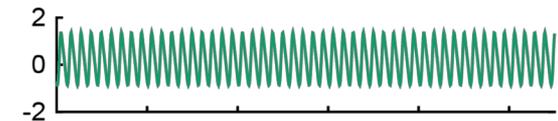**Model results in green.**

Marmoset call sound pressures in purple.

Note different ordinate scales.

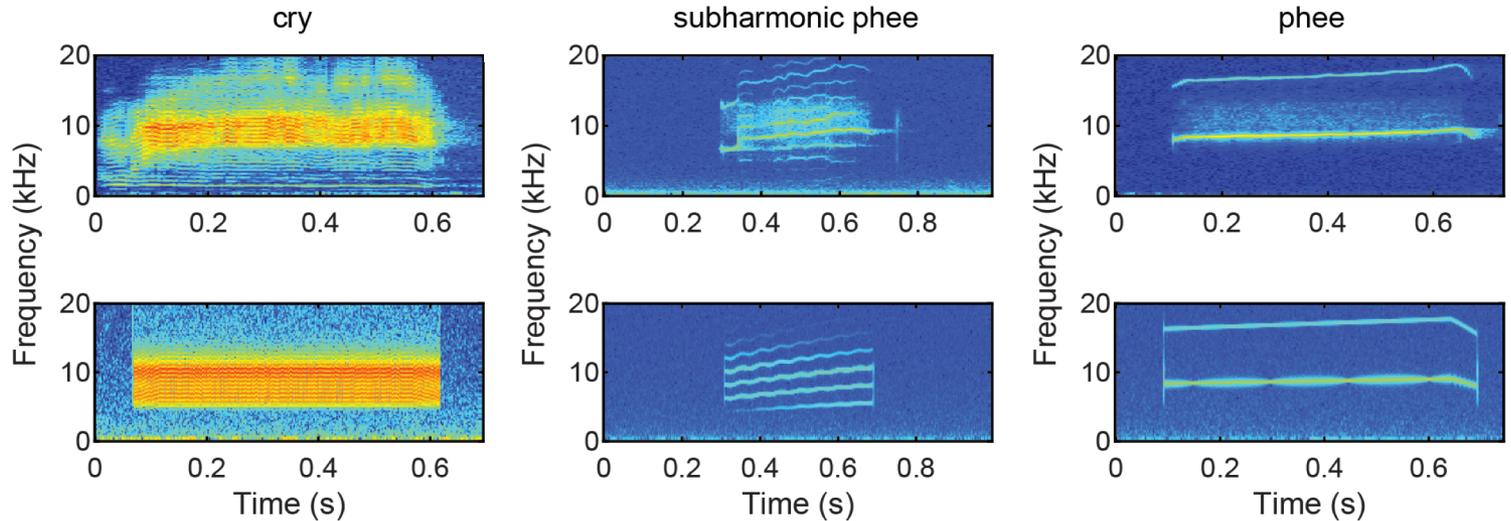small θ large θ

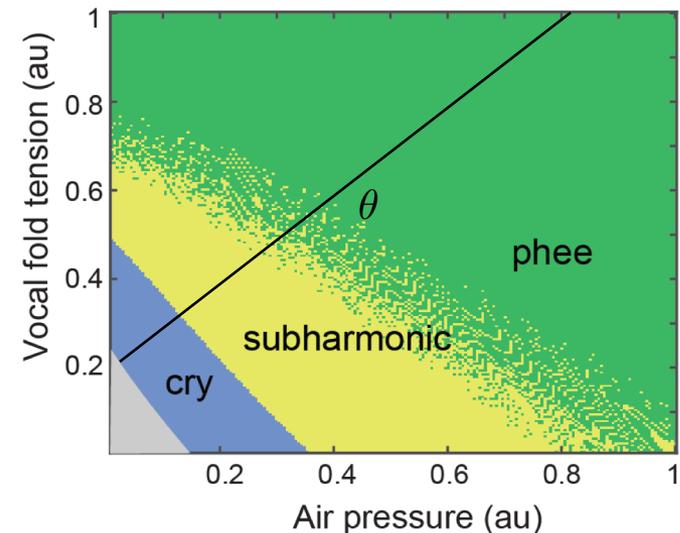# The model can reproduce infant and adult marmoset calls.
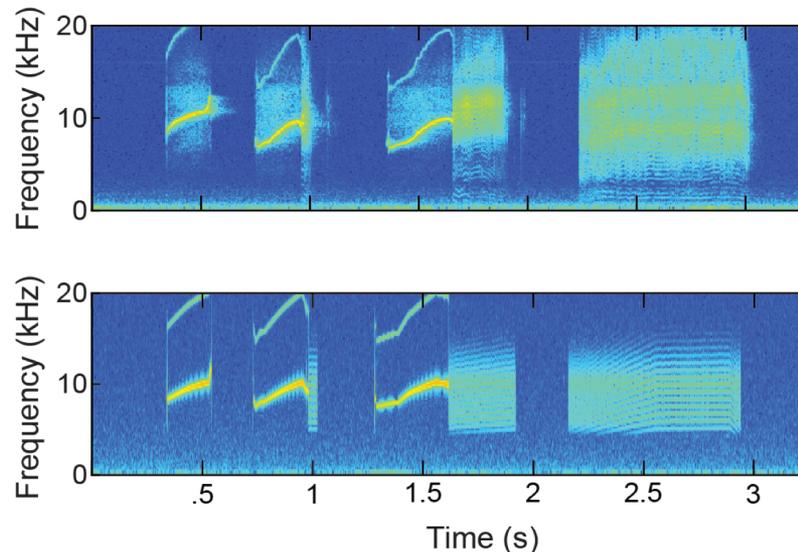
Spectrograms

Marmoset:

Model:

Marmoset:

Model:

Constant θ and piecewise-linear changes in θ yield a range of marmoset calls. **We have a good biomechanical model.**

Note color codes:
cry subharm phee.

III. The maximum entropy (ignorance) principle.

Given a **cost function** $C(\theta)$ such as those to be defined, we ask what is the **probability** $P(\theta)$ that a marmoset will produce a call with specific lung pressure and laryngeal tension θ? The **maximum entropy principle** states that it is the probability distribution that makes the **fewest possible assumptions**, hence maximizing the entropy:

$$H(p) = -\int_{\theta} P(\theta) \log P(\theta) \, d\theta,$$

it also satisfies an **expectation constraint** (the mean cost):

$$\int_{\theta} C(\theta) P(\theta) \, d\theta = E,$$

where we integrate over the relevant range of θ. This gives the **exponential distribution:**

$$P(\theta) = \exp(-\eta C(\theta))/Z,$$

where Z normalizes so that $P(\theta)$ integrates to 1 and η is chosen to satisfy the cost constraint. We don't know the constraint, so will have to guess, or fit η to data.
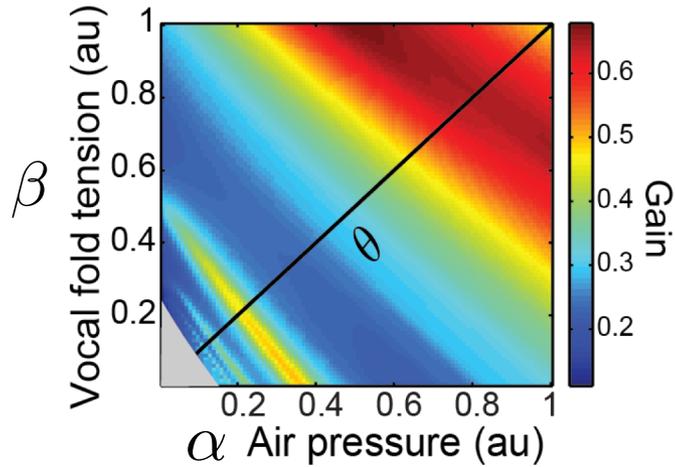
<div align="right">E.T. Jaynes Proc. IEEE 1982, 70:939–952.</div>

Moreover, the following stochastic differential equation (SDE)

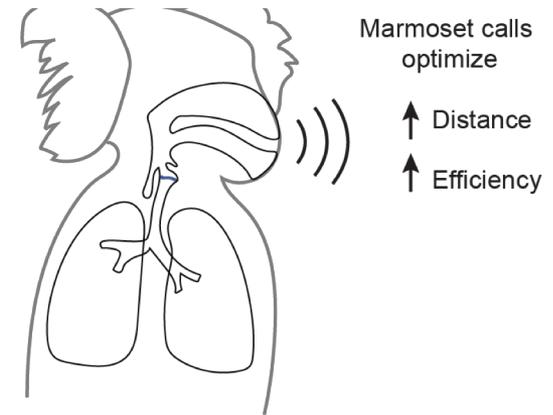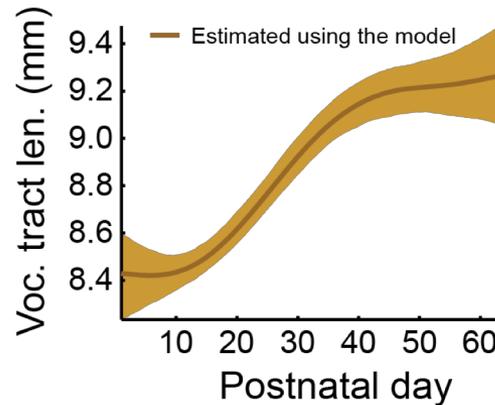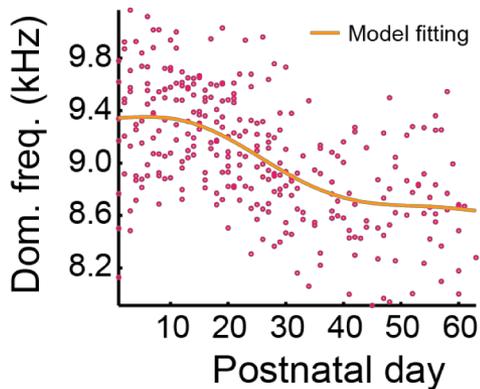$$d\theta_s = \frac{\partial C}{\partial \theta}(\theta_s) \, ds + \sqrt{2/\eta} \, dW_s$$

has the **equilibrium distribution** $P(\theta) = \exp(-\eta C(\theta))/Z$, so running it simulates call production. We'll return to this later. **Next we define cost functions.**

Recall that gain is predicted by the biomechanical model.

$$0 < g(\theta) = \frac{\max_t P_{sound}(t)}{\max_t P_{glottal}(t)} < 1$$

Also, the model predicts the average length of supraglottal vocal tracts as infants grow, which influences resonant coupling, thus changing gain $g(\theta)$.

Marmosets want to optimize calls for distance and efficiency, so gain seems a good quantity for our first attempt at a cost function.

Cost function 1.1.

We use the **maximum entropy principle***.
Taking **cost = – log (gain)** accounts for the effect of vocal tract length on dominant frequency, but fails to predict change in phee/cry ratios. Increasing η makes phees more likely, but with **insignificant dependence on growth** (postnatal day).
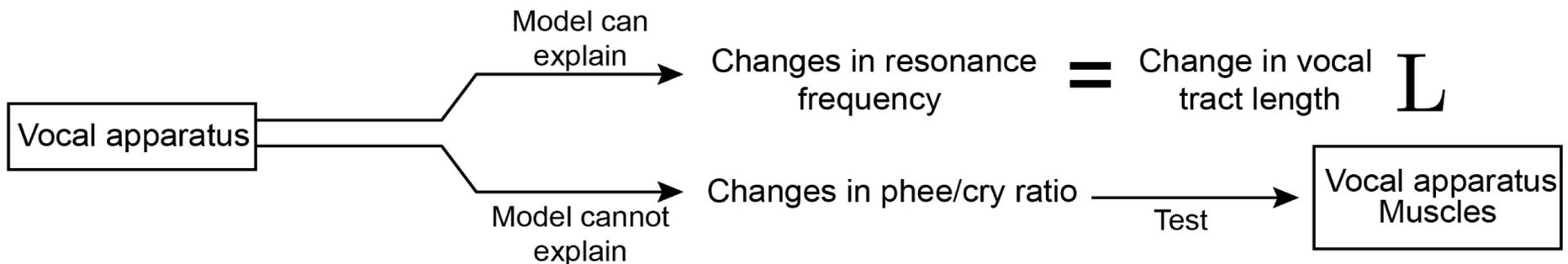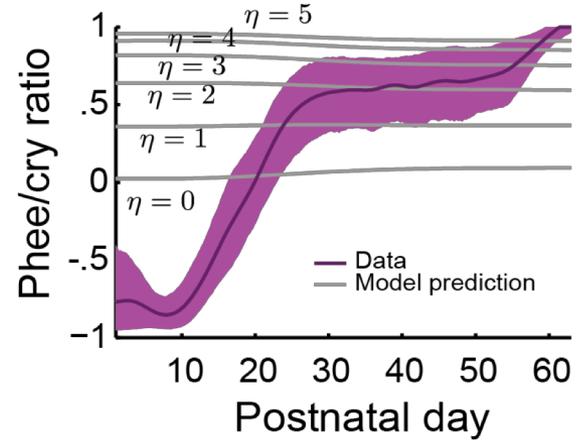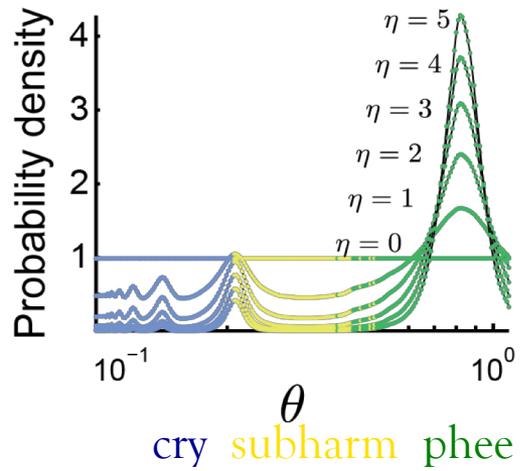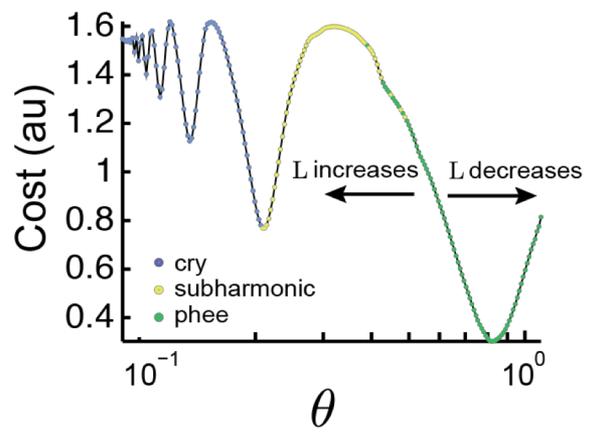
$$\text{Cost} = C(\theta) = -\log g(\theta),$$
$$\text{Prob}(\theta) = \exp(-\eta C(\theta))/Z,$$

normalizn: $Z = \int \exp(-\eta C(\theta)) d\theta.$

$$\text{phee/cry ratio} = \frac{\text{Prob(phee)} - \text{Prob(cry)}}{\text{Prob(phee)} + \text{Prob(cry)}}.$$

*E.T. Jaynes *Proc. IEEE 1982, 70:939–952.*

cry   subharm   phee

But cost function 1.1 was a poor candidate.

We use the **maximum entropy principle***. Taking **cost = – log (gain)** accounts for the effect of vocal tract length on dominant frequency, but fails to predict change in phee/cry ratios. Increasing η makes phees more likely, but with **insignificant dependence on growth** (postnatal day).
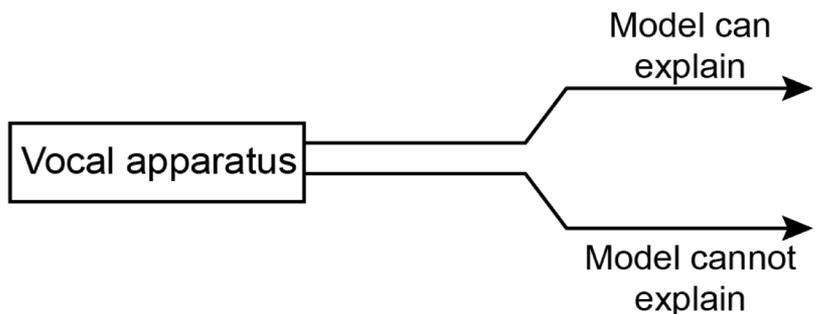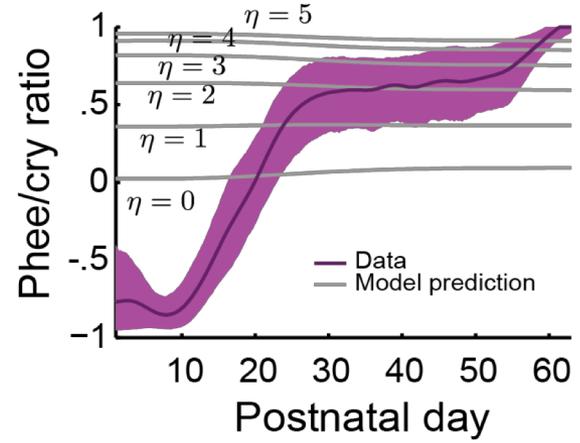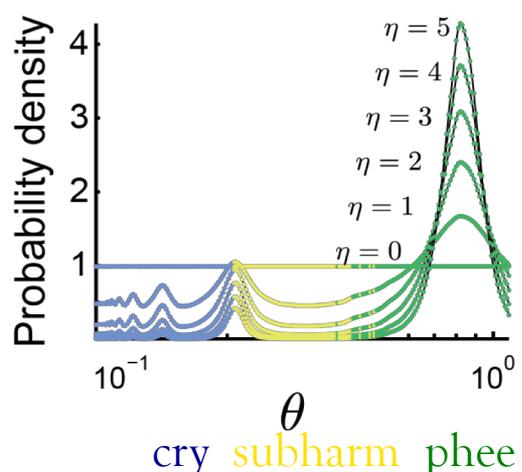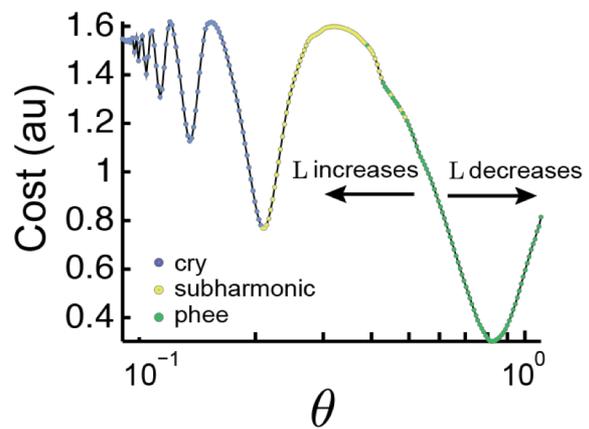
$$\text{Cost} = C(\theta) = -\log g(\theta),$$
$$\text{Prob}(\theta) = \exp(-\eta C(\theta))/Z,$$

normalizn: $Z = \int \exp(-\eta C(\theta))d\theta.$

$$\text{phee/cry ratio} = \frac{\text{Prob(phee)} - \text{Prob(cry)}}{\text{Prob(phee)} + \text{Prob(cry)}}.$$

*E.T. Jaynes *Proc. IEEE 1982, 70:939–952.*

cry   subharm   phee

It can't capture the phee/cry ratio!
No choice of η fits the data.

Modified cost function 1.2 and probabilistic predictions.

The biomechanical model predicts that resonant frequency decreases with growth, but cannot account for muscle development. So we modify the cost function, **adding a penalty $\lambda\theta$ for muscular effort that decreases with growth**, and we calculate the probabilities of cries and phees. As λ **decreases, the cost function landscape rotates** and **the phee minimum gets deeper, making phees more probable** than cries or subharmonics.

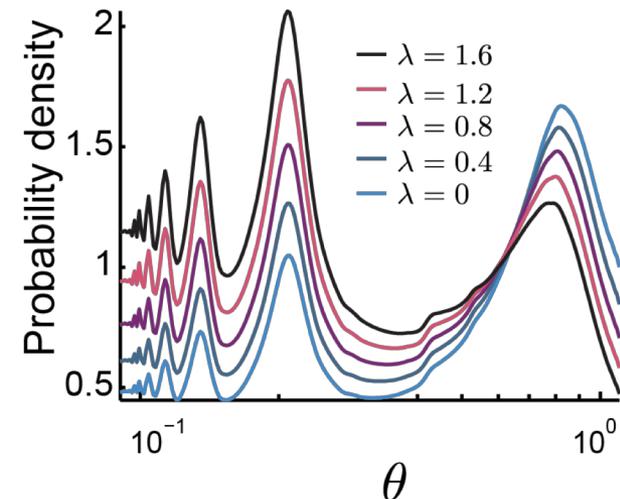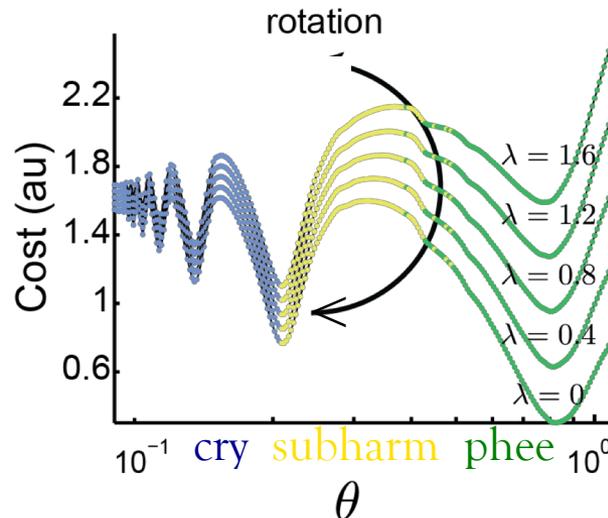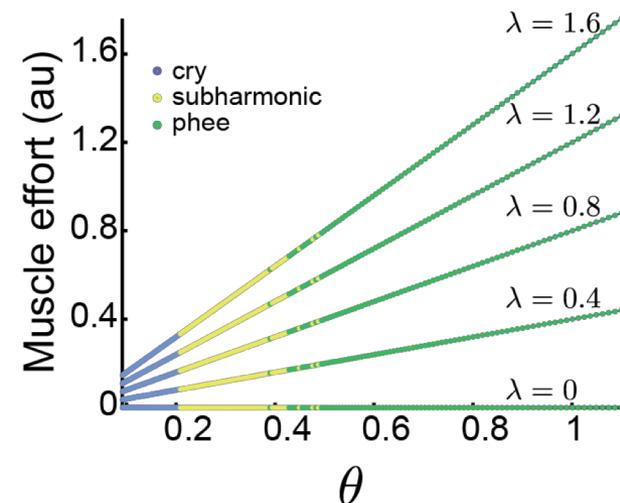$$\text{Cost} = C(\theta) = -\log g(\theta) + \lambda\theta,$$

$$\text{Prob}(\theta) = \exp(-\eta C(\theta))/Z,$$
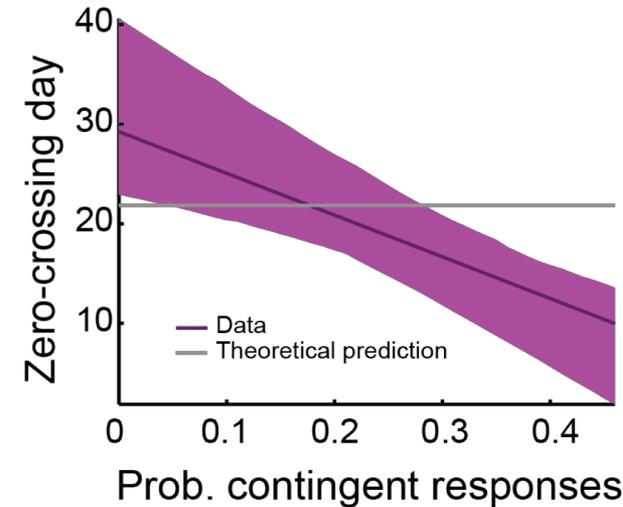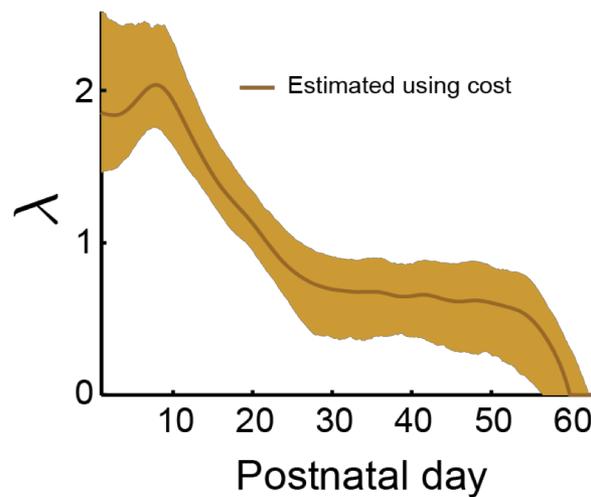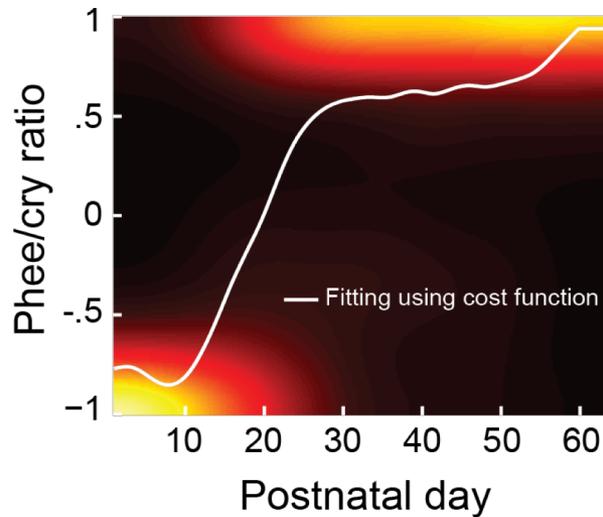
$$\text{normalizn: } Z = \int \exp(-\eta C(\theta))d\theta.$$

$$\text{Prob(cry)} = \int_{A_{\text{cry}}} \text{Prob}(\theta)d\theta,$$

$$\text{Prob(phee)} = \int_{A_{\text{phee}}} \text{Prob}(\theta)d\theta.$$

$$\text{phee/cry ratio} = \frac{\text{Prob(phee)} - \text{Prob(cry)}}{\text{Prob(phee)} + \text{Prob(cry)}}.$$
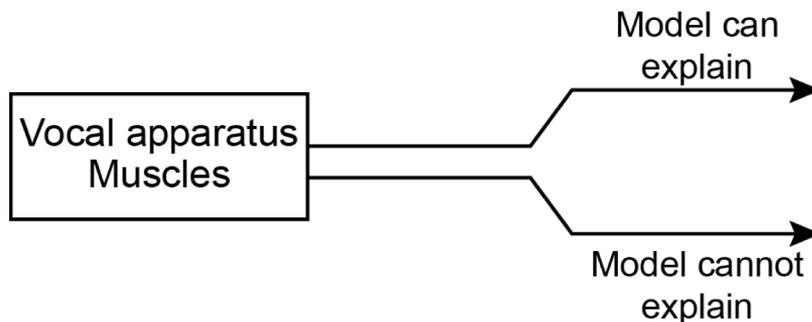
Cost function 1.2 can predict average phee/cry ratios.

Fitting phee/cry ratios reveals how lambda changes with growth, but **predicts transitions independent** of contingent parental responses to infants' calls.

Thus, changes in muscular control can explain the **population's change in phee/cry ratios,** but not the **individual timings** of this transition, which are influenced by **parental feedback.**
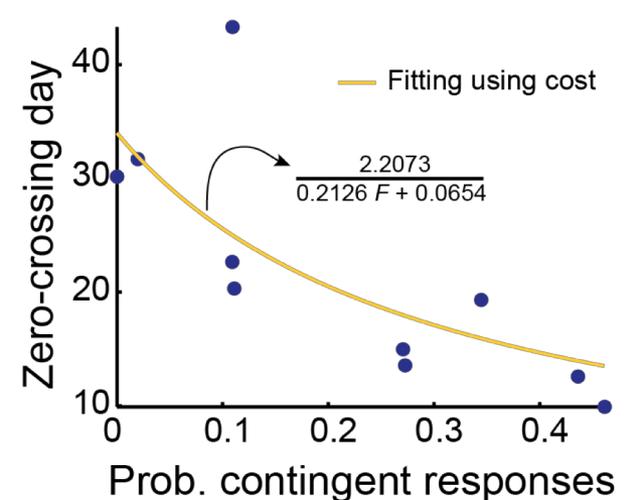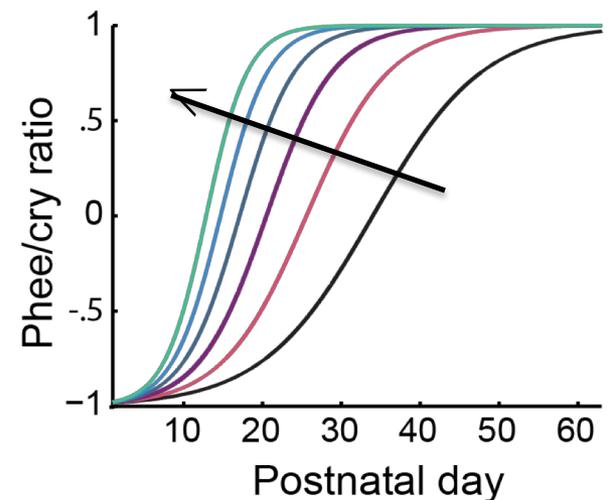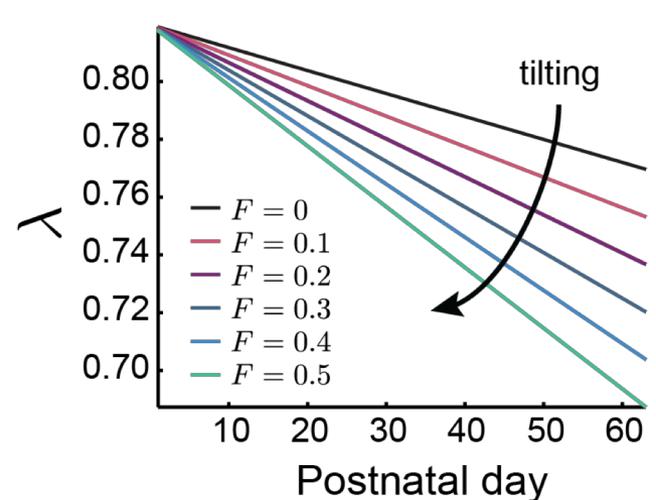

Better predictions, but individual infant behaviors are not addressed.

Cost function 1.3 accounts for learning from parents.

Add parameters δ, κ, F, let λ evolve with t.
δ = physical growth factor, κ = scaling factor,
F = contingent parental response to infants.

$$\text{Let } \lambda \to \lambda_t, \quad \lambda_t = \lambda_{t-1} - \delta - \kappa F,$$
$$C(\theta) = -\log g(\theta) + \lambda_0 \theta - (\delta + \kappa F)t\theta.$$

More parental feedback F makes λ decay faster, moving the phee/cry ratio zero crossing earlier. We can fit δ, κ to zero crossings for each infant, determining a **nonlinear curve of zero-crossing days.**

Partial success! But do infants' behaviors affect parental feedback?

We next ask if the infants' growth and phee call rates influence parental feedback. We find that they do not, supporting the hypothesis that **parental contingent responses influence infants' neural development, regardless of infants' growth and call rates.**

$r = 0.314$
$p = 0.378$

$r = -0.048$
$p = 0.896$

(n = 10.)

Prob. contingent responses

Weight change W (g/day)

Prob. contingent responses

Phee call rate N (calls/min)

$F =$ **Insignificant** $+$ noise ?

L δ κ F

Changes in resonance frequency

Changes in phee/cry ratio

Changes in zero-crossing day

Independent of infants' growth rates

IV. The maximum entropy method is a diffusion process with cost function as potential.

Theta tells how far the infant marmoset has moved through the region of call production. $\mathrm{Prob}(\theta) = \exp(-\eta C(\theta))/Z$, **with cost function 1.3,** defines a **landscape that evolves over postnatal days** as the larynx, muscles and nervous system develop. Each day (t), the infant produces cries and/or calls as described by cost function 1.3. This is equivalent to **a drift-diffusion process with potential** $C(\theta)$:
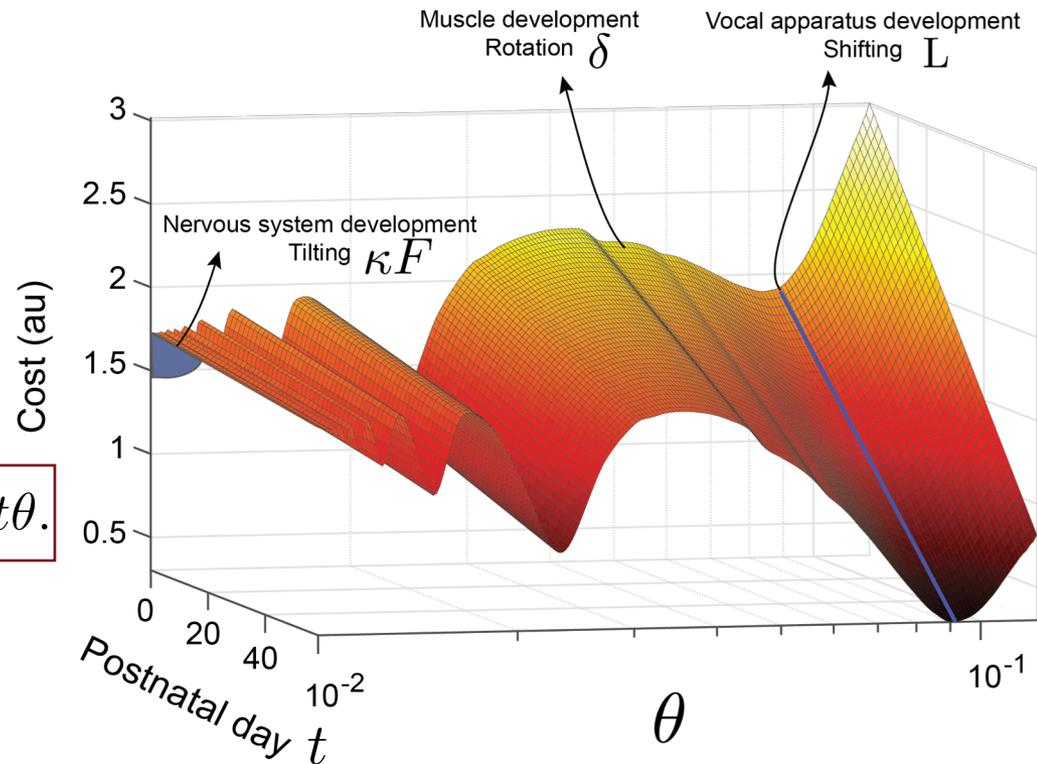
$$dθ_s = \frac{\partial C}{\partial \theta}(\theta_s)\, ds + \sqrt{2/\eta}\, dW_s \ ,$$

defined on the landscape shown here.

The slow **time** (t)-**dependent cost function 1.3** includes **muscle growth, nervous system** and **social interaction:**

$$C(\theta) = -\log g(\theta) + \lambda_0 \theta - (\delta + \kappa F)t\theta.$$

Each postnatal day (t), cries and/or calls are produced on the fast time scale (s).

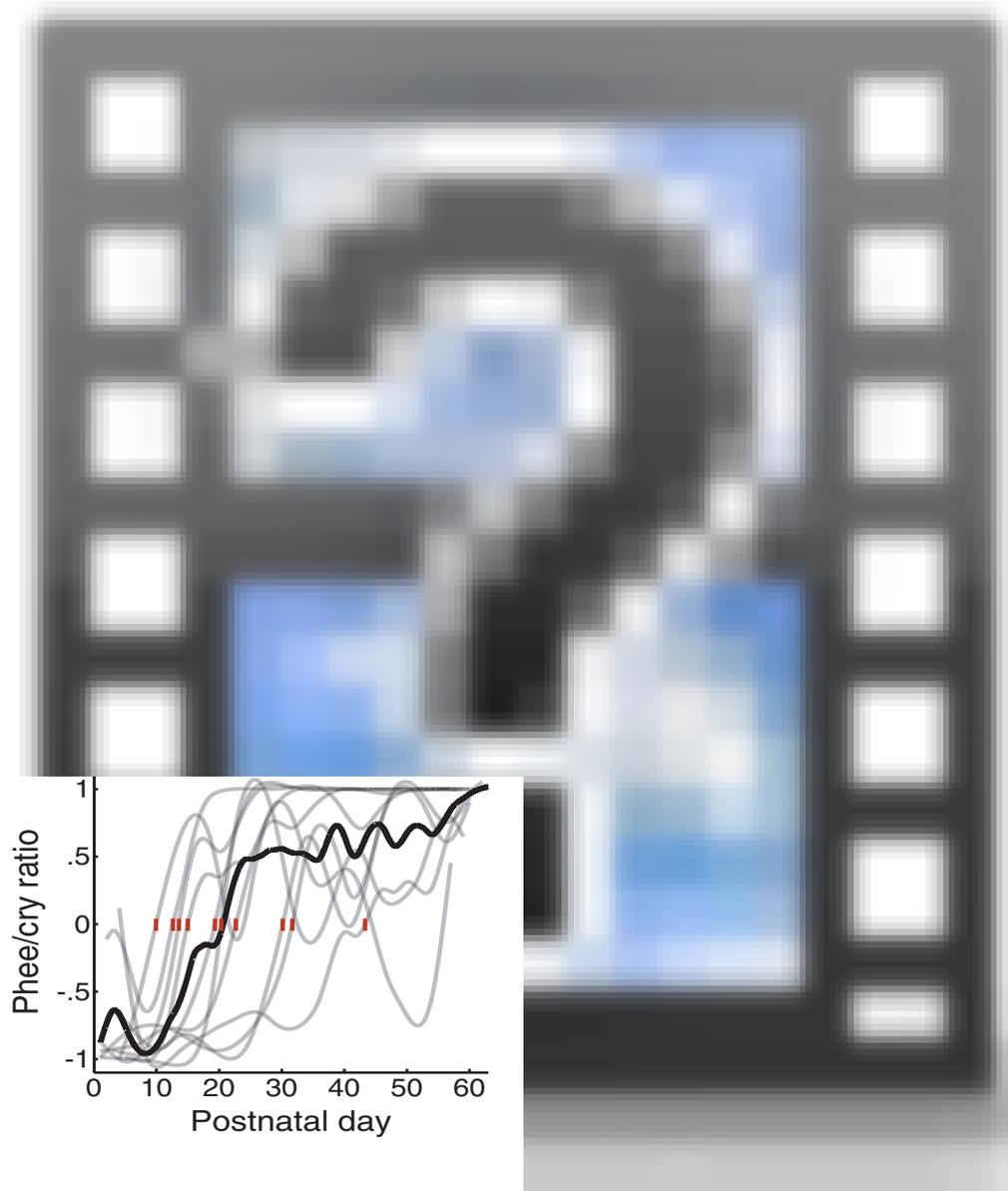# Animation of vocal dynamics on a Waddington landscape

The system's state diffuses on a probability function landscape that evolves daily due to changes in vocal apparatus, muscle strength, nervous system and social interactions as the animal grows. **In early days, mostly cries are produced and later, mostly mature phee calls.** Dynamics are shown here at 6-day intervals from one model growth run.

Lower left shows distributions of cries and phees for each day; lower right shows phee/cry ratios for the same postnatal days.

C.H. Waddington *The Strategy of the Genes*. Routledge, *1957.*
J. Wang, K. Zhang, L. Xu, E. Wang, *2011.*
PNAS 108:8257–8262.

Epilogue: Models of mind

I think I know that brain creates the mind,
but why is this so hard to see? Lips brush
and leave a touch encoded, or a taste.
Spikes run from heel to head; we feel
the clutch of signals taken up, released:
a world of sense and memory from which
all action springs. Yet here's the scruple:
is correlation cause? And where am *I*?

We've learned this much: if not bound
by muscle, then by nerve. Our bodies hold
the present in their pasts, afford a grasp
against the day's assault, the waste of light.
Yet it's provisional; our models shift and slide,
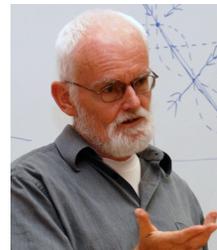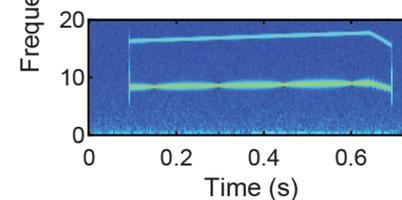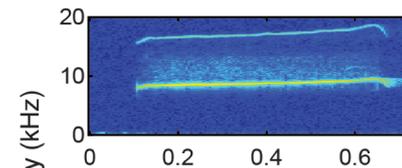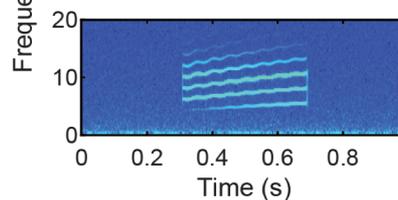leaving words to spare, but few to bring to mind.
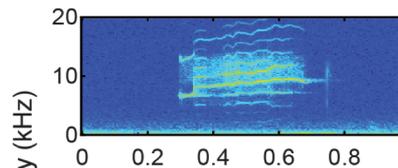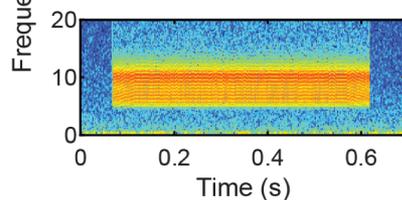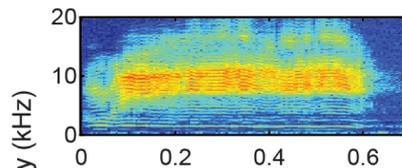
                                   Philip Holmes

V. Conclusions.

1. We built a biomechanical mathematical model of the marmoset vocal apparatus.

2. Together with behavioral data and maximum entropy methods, the model enabled an integrated view of the system, including muscle growth, neural development and social feedback from parents.

3. The general framework should adapt to other developmental examples.

Mathematical tools: deterministic and stochastic dynamical systems, bifurcations and normal forms, maximum entropy methods, some probability theory.
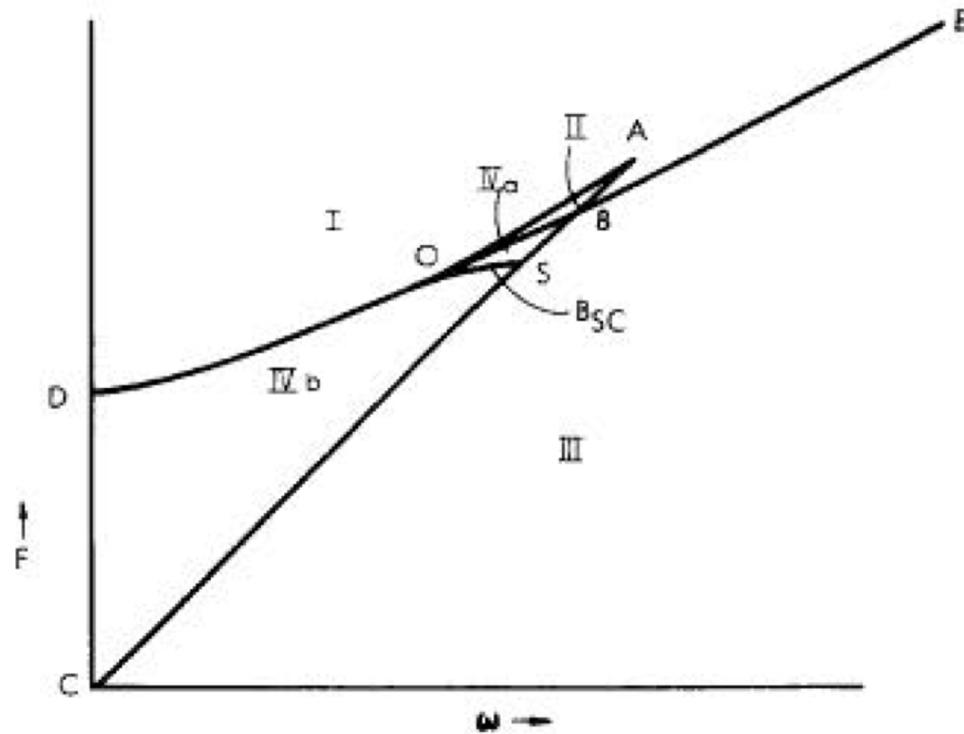
Morals: Neural systems learn from and act upon their environments. Integrative neuroscience needs biomechanics and mathematical models.

Y. Teramoto, D.Y. Takahashi, P. Holmes & A.A. Ghazanfar. Vocal development in a Waddington landscape. *eLife* 6: e20782, 2017.   E.D. Tytell, P. Holmes & A.H. Cohen. *Curr. Opin. in Neurobiol. 21: 816-822, 2011.*

E. Thelen & L.B. Smith. Dynamic system theories. In R. Lerner (Ed.) *Handbook of Child Psychology, Wiley:*

BIFURCATIONS OF THE FORCED VAN DER POL OSCILLATOR

See P. H. and D. Rand, *Quart. Appl. Math. 1978: 35, 495-509;*
J. Guckenheimer and P. H. *Nonlinear Oscillations, Dynamical
Systems, and Bifurcations of Vector Fields, Springer NY, 1983.*

Storage for potential illustrations and equations.

$$m\ddot{x} + b(x, \dot{x}, t)\dot{x} + k(x, t)x = f(x, \dot{x}, t).$$

$$\dot{x}_1 = x_2,$$
$$\dot{x}_2 = \frac{1}{m}\left[-(k_1 + k_2 x_1^2)x_1 - (b + cx_1^2)x_2 + \frac{a_g P_L(x_{01} - x_{02} + 2\tau x_2)}{k_t(x_{01} + x_1 + \tau x_2)}\right].$$
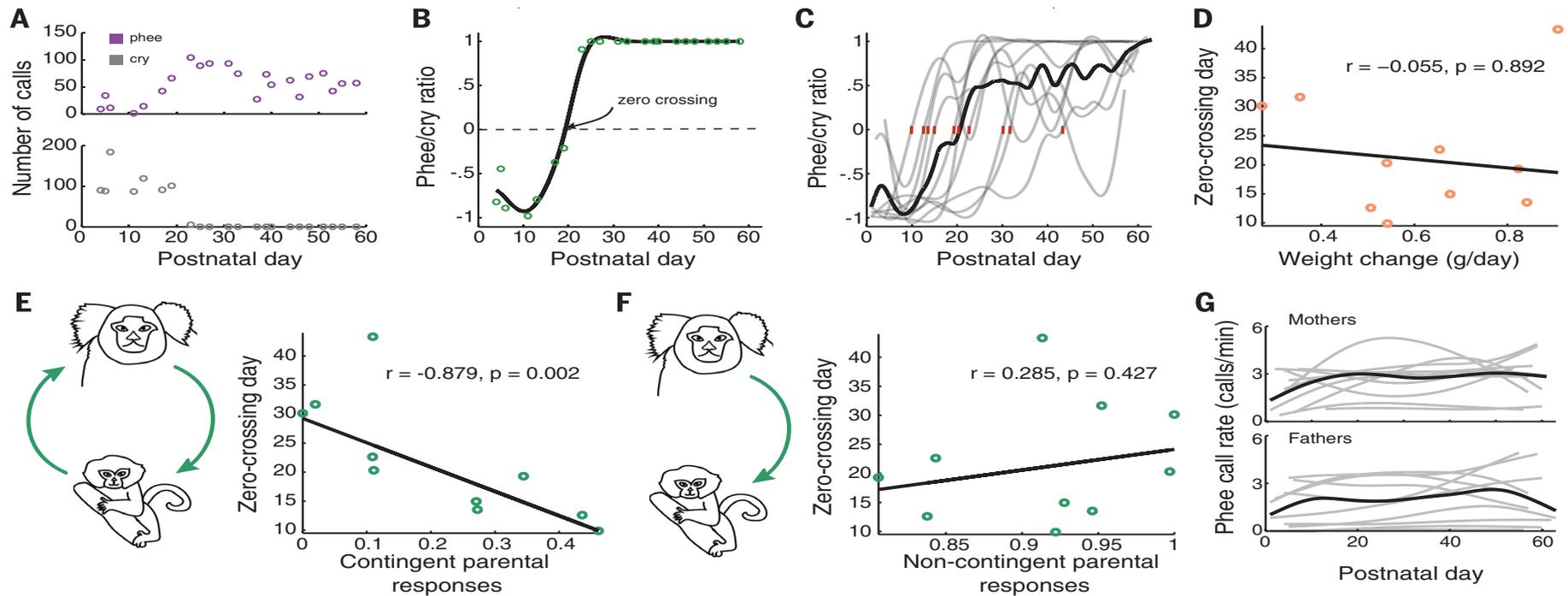

Fig. 4. Transition from cry to phee is influenced by contingent parental calls. (A) Numbers of cries and phees over 2 months for a single infant. (B) Phee/cry ratio for the infant in (A) across days. (C) Phee/cry ratios (gray curves) and zero-crossing days (red ticks) for each infant and for the population (black curve). Black and gray curves in (B) and (C) are cubic spline fits. (D) Correlation between the weight-change rate and the zero-crossing day among infants. (E and F) Correlations between the zero-crossing day and the proportion of contingent and noncontingent parental responses, respectively. (G) Rates of individual parental phee-call production during infant development (gray) and the population average (black).