Intonational tunes as compositional, but autonomous, dynamical objects

Variance and Invariance Workshop

Khalil Iskarous, University of Southern California

Jennifer Cole, Northwestern University

Thomas Sostarics, Northwestern University

Northwestern University

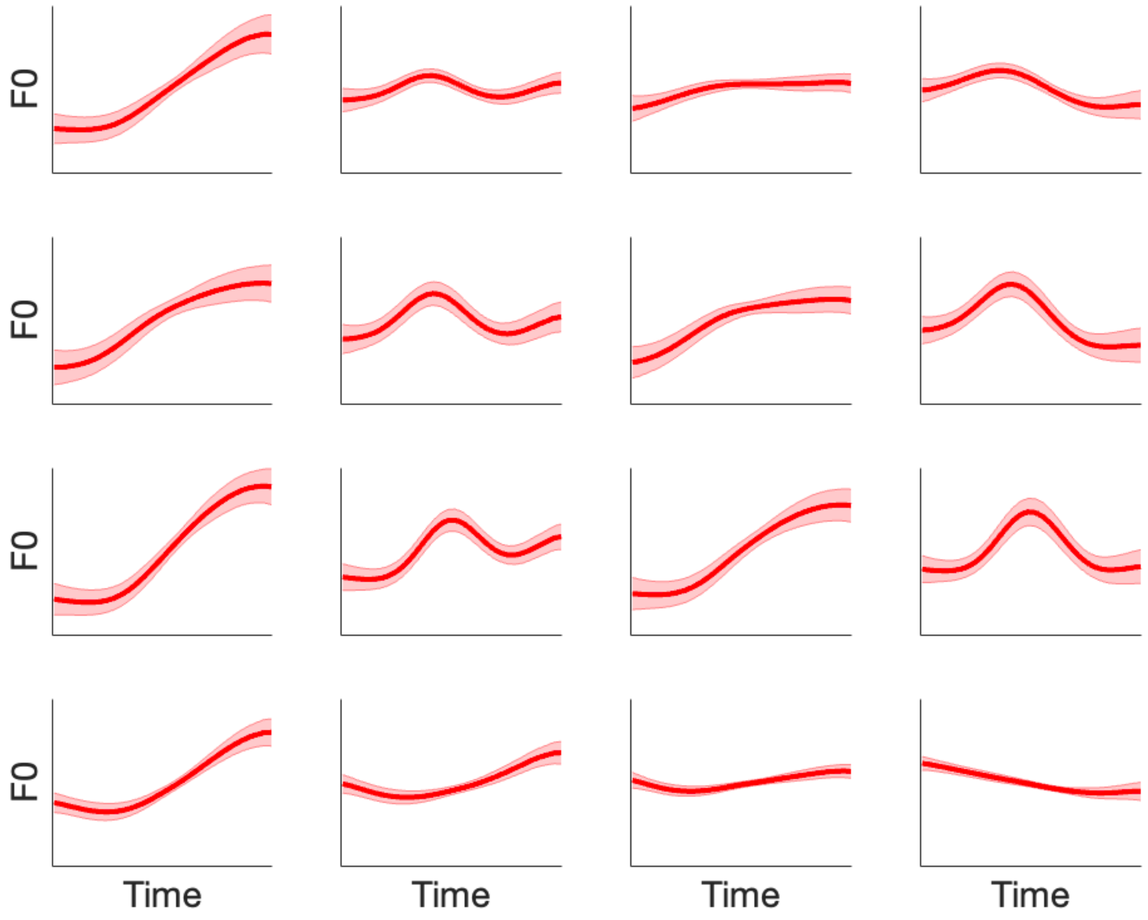
USC University of Southern California

Variance-Invariance

Chao, Pike, Bolinger, others:

There is **a pattern** to the variable pitch events in the languages of the world, e.g., towards the ends of AE utterances, *nuclear tunes*.

Desire for a **grammar**: A simple **invariant** mechanism from which the time-varying tunes emerge from some specification.



AE tunes from Northwestern database

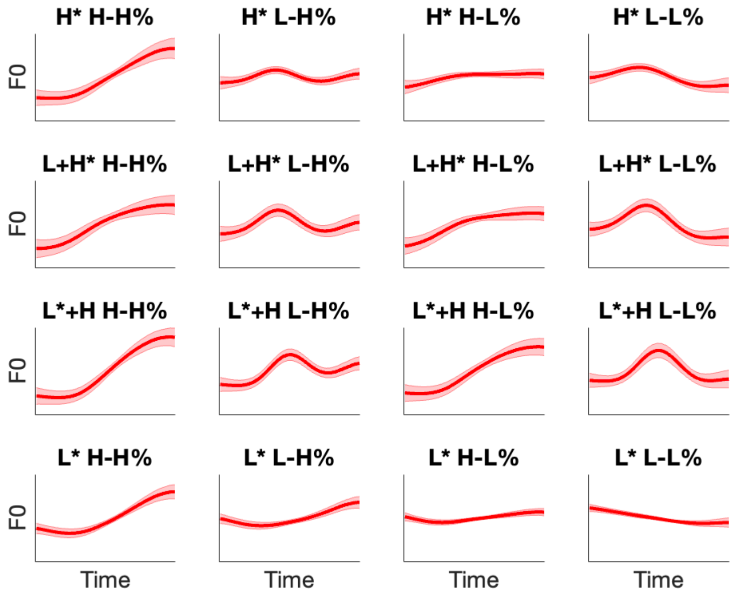
Structuralist Postulates and AM Phonology

Bloch (1948)

53.3. Corollary. A phoneme is a class of events.

Postulate 55. In any phrase, the phonemes occur in a particular order.

High pitches are variable in and across tunes, but they belong to one invariant *class*, H; so also for L. The birth of H, L.



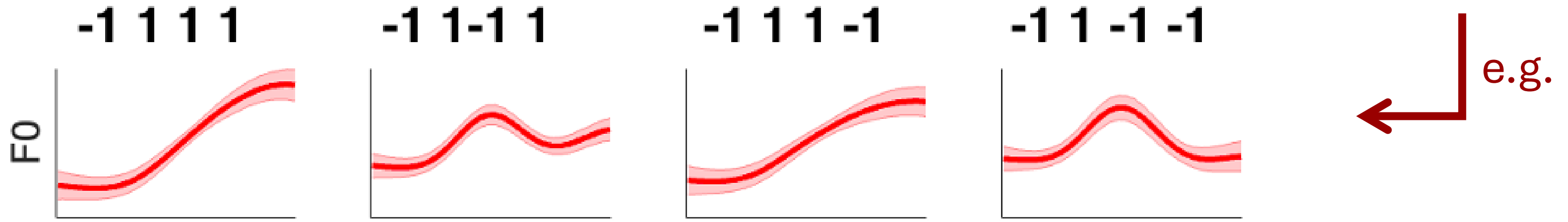
Goldsmith (1975)
Liberman (1976)
Bruce (1978)
Pierrehumbert (1980)

Pitch accents, phrase accent, boundary tone as local events in Bloch-order that compose to form a tune. An F0 Algorithm generates trajectory from the phonemes in order.

Dynamical Approach to Invariance

Fowler et al (1980)
Browman and Goldstein (1989)
Saltzman and Munhall (1989)

- Linguistic Variance and Invariance are an instantiation of motor variability (Variable tunes) + “equifinal” target achievement (Contrast).
- H, L → Constant targets of a *stable Equi* DS
- Equi DS’s executed in Bloch-Order via a **Task Dynamics** that *solves* this string of potentially overlapped systems for a predicted F0 trajectory



Claim: **Invariant** linguistic F0 patterning emerges from a general approach to motor control, dynamical systems

Problem 1: Targets H, L as input vs. *output*

Pierrehumbert & Pierrehumbert (1990)

Come on guys… Task Dynamics is just an interpolation algorithm, given H (1),L (-1) sequence, generate F0!

For *emergence* of a trajectory from a truly invariant dynamical account, H and L should be in the *output*… For TD, they're in the *input*.

A dynamical system that *predicts* linguistic structuring should not *postulate* that structure, targets and Bloch-order (anti-phase)

Problem 2: *Lessen* loan on intelligence

Perrier et al. (1988) | True trajectories of task change are more complex in their time-variability than predicted by stable equilibrium linear 2nd order dynamical systems.

Byrd & Saltzman (1998) | *Nonautonomous* approach: The linear equations should have in them time-varying components, designed by the intelligent researcher, that make the solution conform to reality.

Sorenson & Gafos (2016) Tilsen (2022) | *Autonomous* approach: An explanatory dynamic is *autonomous:* it predicts complex time-variation through feedback *from system state to system dynamics,* not a time-varying force from the researcher. So one or more coefficients now depends on the state, hence *nonlinearity*.

Our Aim: An autonomous tune

Step 1: Kinetic approach to targets

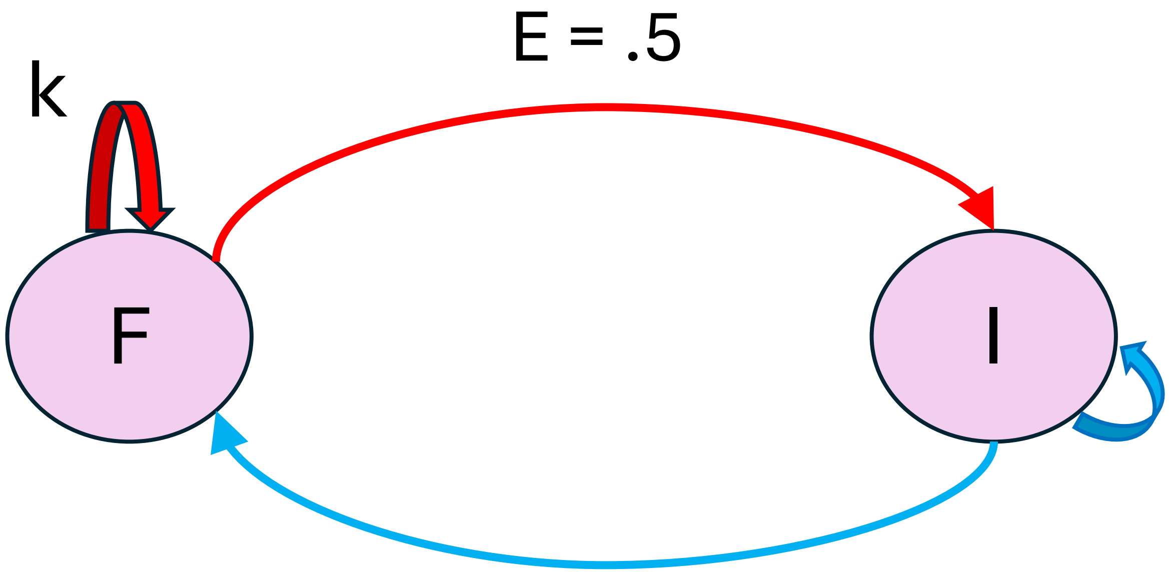
Iskarous, Cole, and Steffman (2024) advanced a dynamical model of the pitch accent part, which in AM include H*, L+H*, L*+H, L*, from which H and L are claimed to *emerge* from an embodied dynamics (Simko and Cummins, 2010).

- There are 2 dynamical variables:

  F: *pitch setting forces on laryngeal muscles*

  I: *pitch inhibiting forces on laryngeal muscles*
- The two variables dynamically interact, kinetically, in a circuit, conditioned by the phonology of a language
- Parametric kinetics of interaction lead to pitch accents (cf. Goldsmith, 1994).

The Kinetics



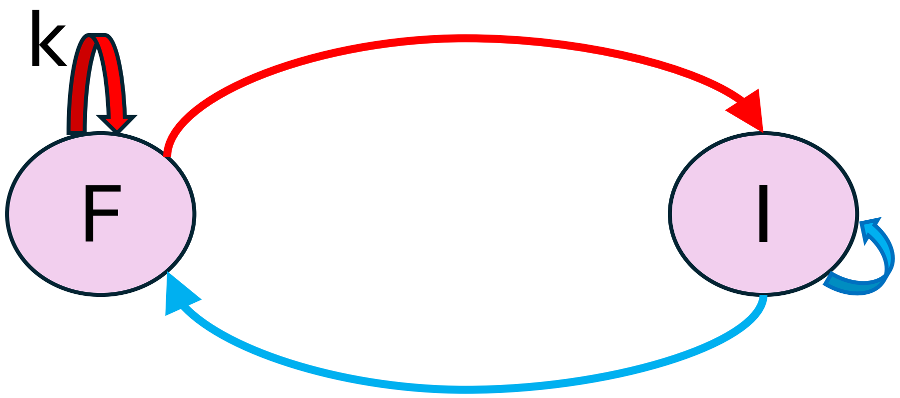
- **F** makes itself grow by a coefficient k
- **F** excites ***I by constant E***
- **I** inhibits **F** and itself

$$\frac{dF}{dt} = kF - F^3 - I + .5$$
$$\frac{dI}{dt} = EF - I$$

- **Read:**
  - **Slope of F is a function of F through a phonological constant k, I, and a constant energy source.**
  - **Slope of I is a function of F & I.**

<u>The true F0 trajectory is the One that conforms most to this dynamic *at every single point in time*</u>

Predictions of the Kinetic Theory



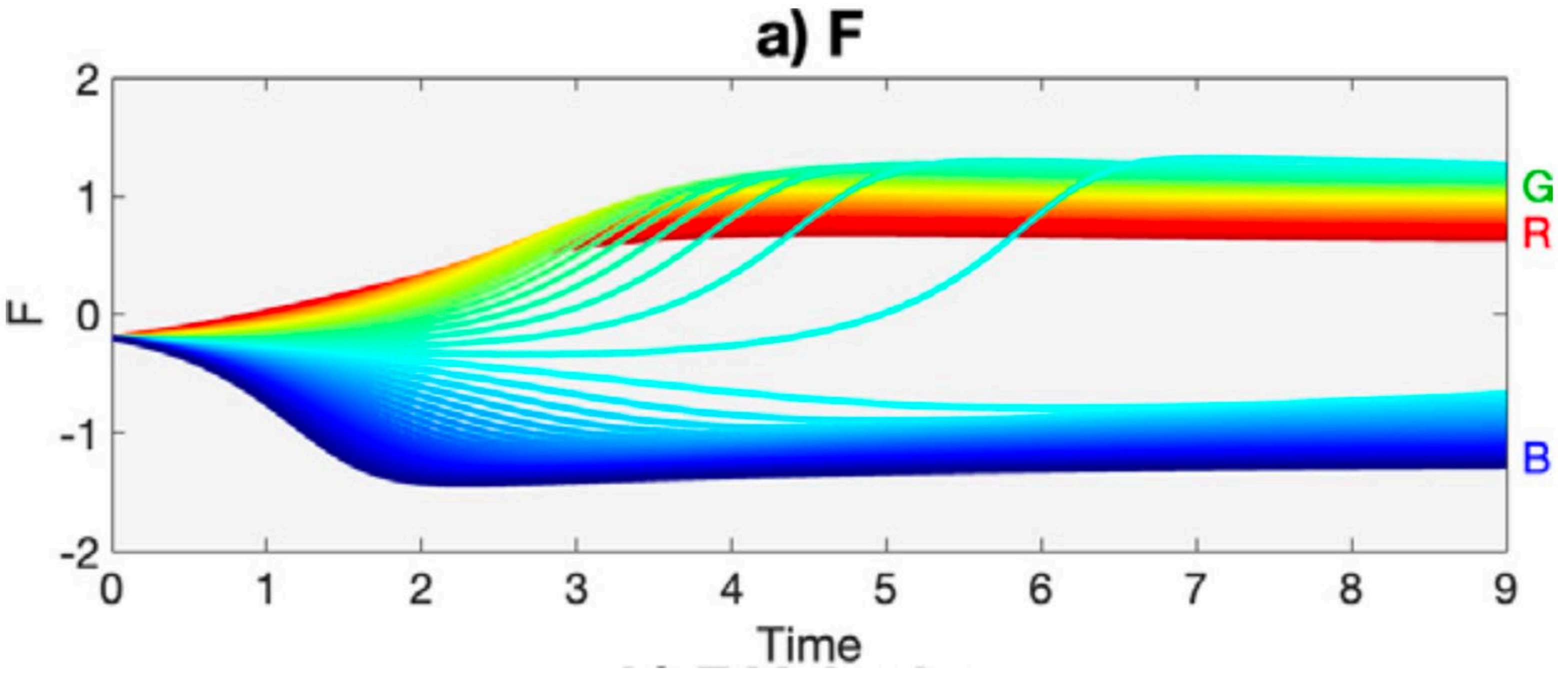
When k is phonology-fixed, we get one optimal *F-trajectory*, here color-coded by k from lowest (red) to (highest).

H*: **k ~ .5**

L+H*: **k ~ 1**

L*+H: **k ~ 1.5**

L*: **k ~ 2**

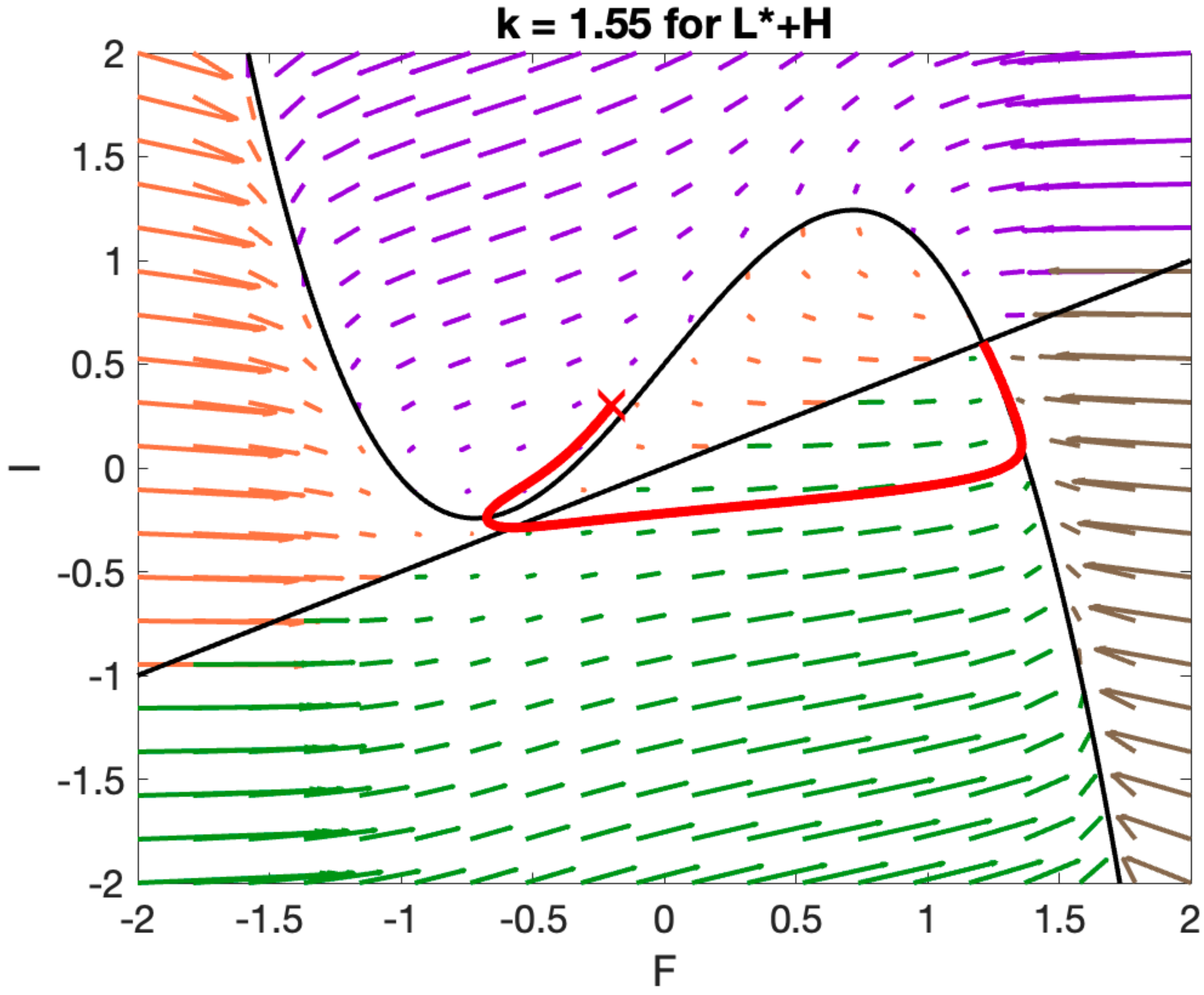


Iskarous et al. (2024)

Is this just another F0 algorithm? No, due to Equivalence Classes

$$\frac{dF}{dt} = kF - F^3 - I + .5$$
$$\frac{dI}{dt} = EF - I$$

<u>Dynamics *divides* infinitely continuous phonetic space into equivalence classes</u>

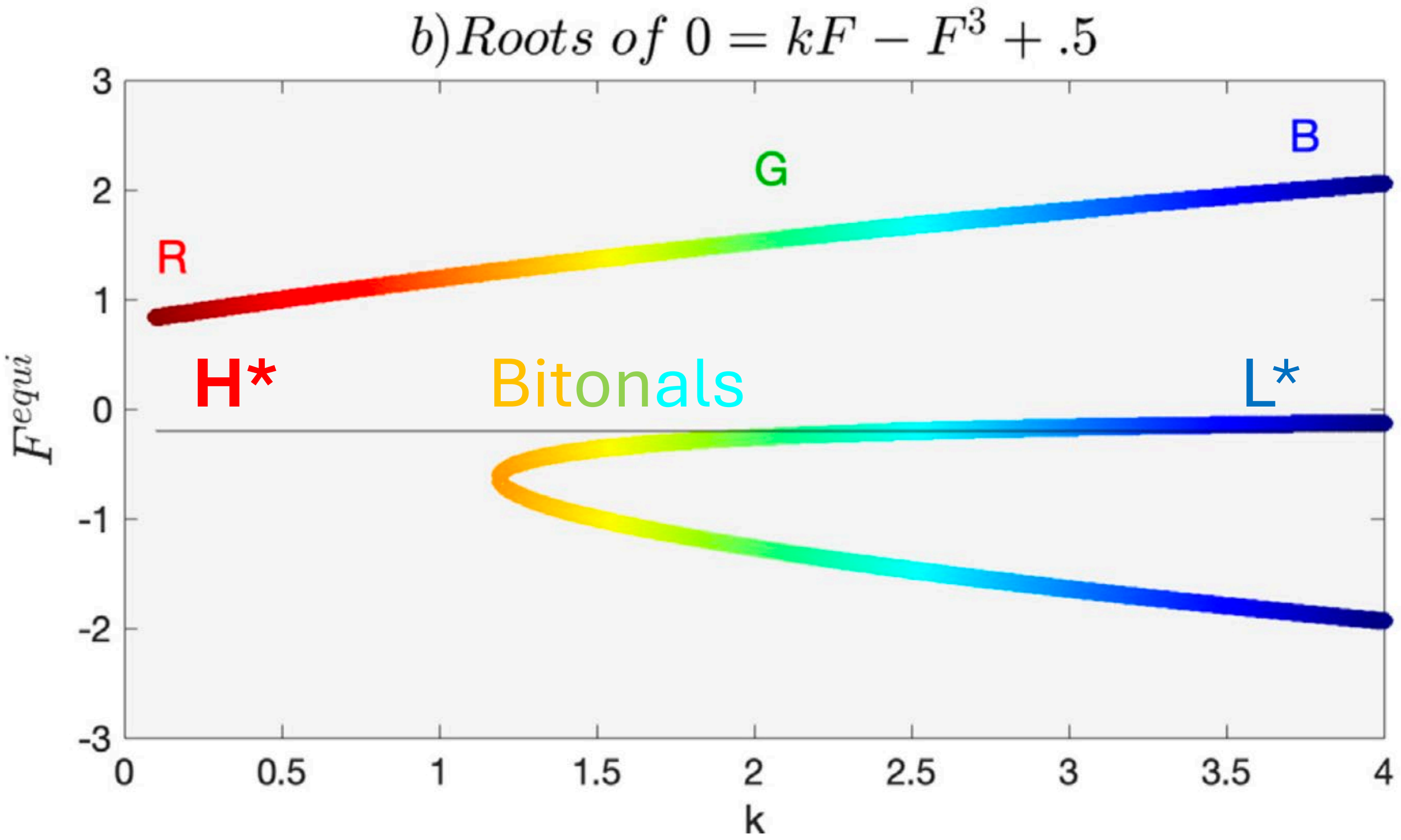


Note: E is slope of I-Nullcline

Is this just another F0 algorithm? No, due to quantal behavior

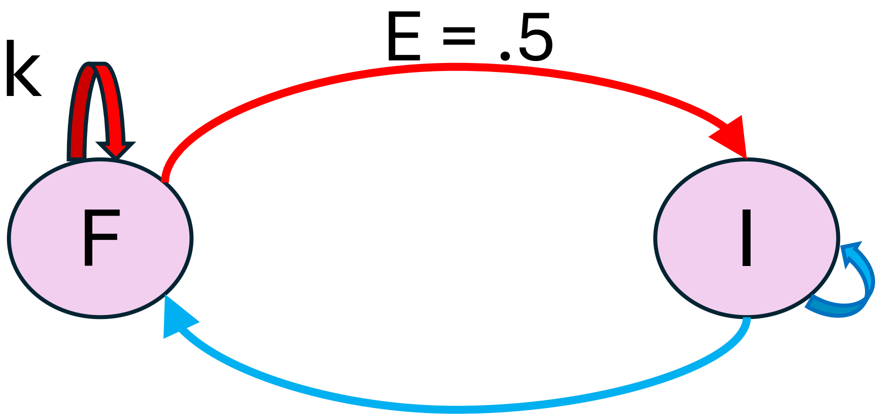
Iskarous et al. (2024) show that this model predicts phonological quantal-theory like discreteness as k varies

(Figure shows 1-variable model simplification)

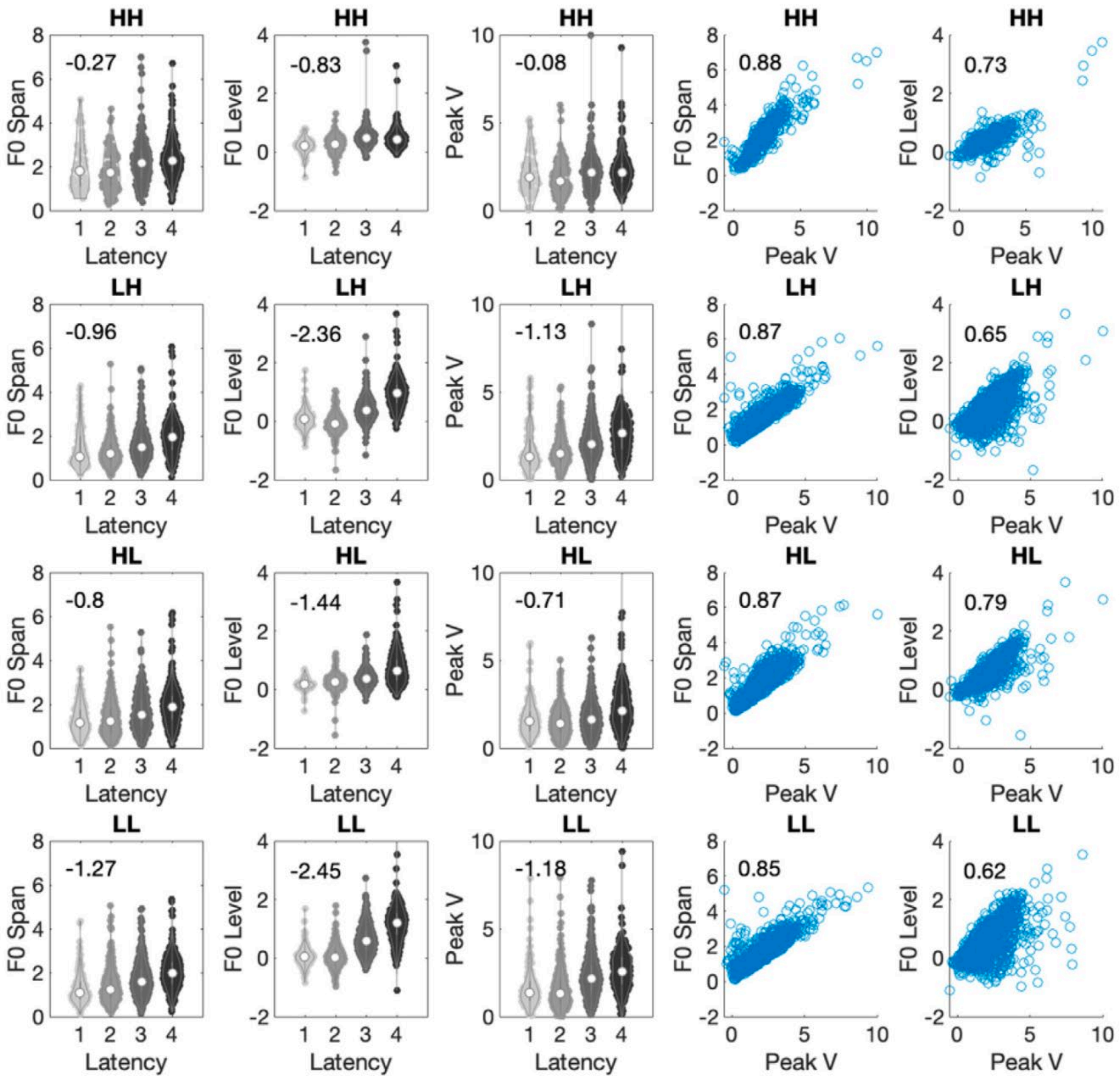


Claim: **Variance-Invariance in one model**

And several phonetic facts on details of time-variability from a corpus of 130 speakers are predicted by model

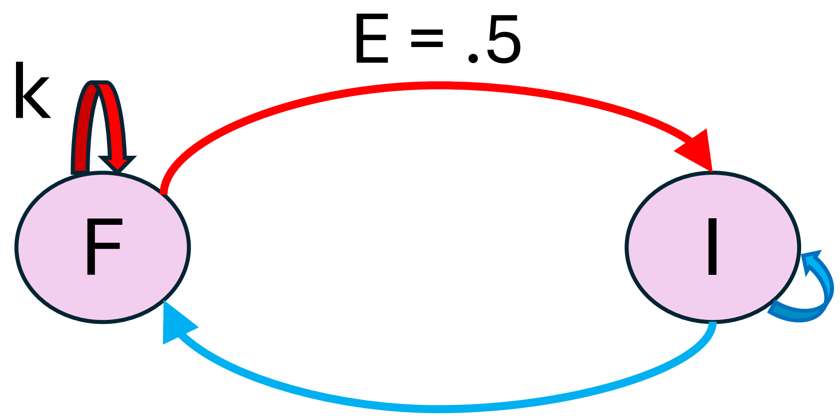


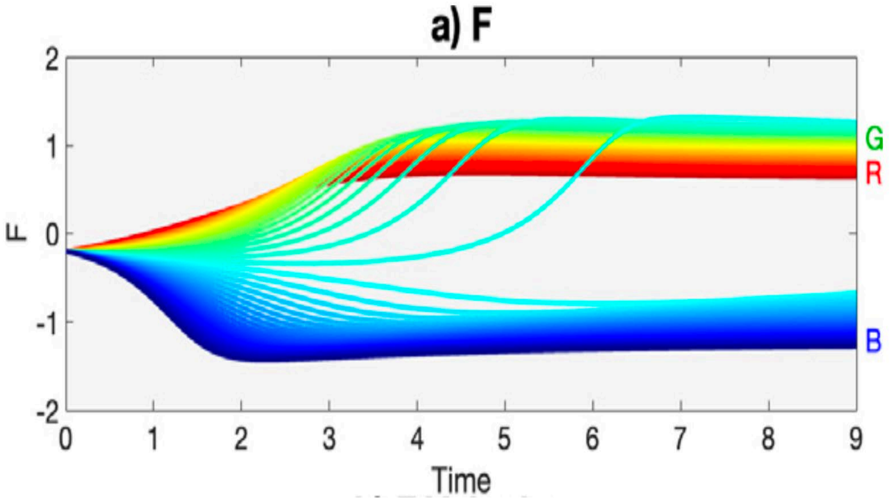
Iskarous et al. (2024)



Addressing P & P (1990) and S & G (2016)

- The **k**-mediated struggle between **F** and **I** determines whether pitch goes low or high.
- We derive the scooped L*+H (down then up), from a single specification of **k**, just as for H*, L*
- **Bitonals are *saddles, not stable equilibria*** 😇
- Tones arise autonomously without time-variation in **k**, through the kinetics of the phonologically-mediated motor circuit
- Btw: this circuit has been fundamental to biological development and neuroscience since its initial description by Turing (1952), *The Chemical Basis of Morphogenesis,* see also Iskarous (2019) on the morphogenesis of gestures.

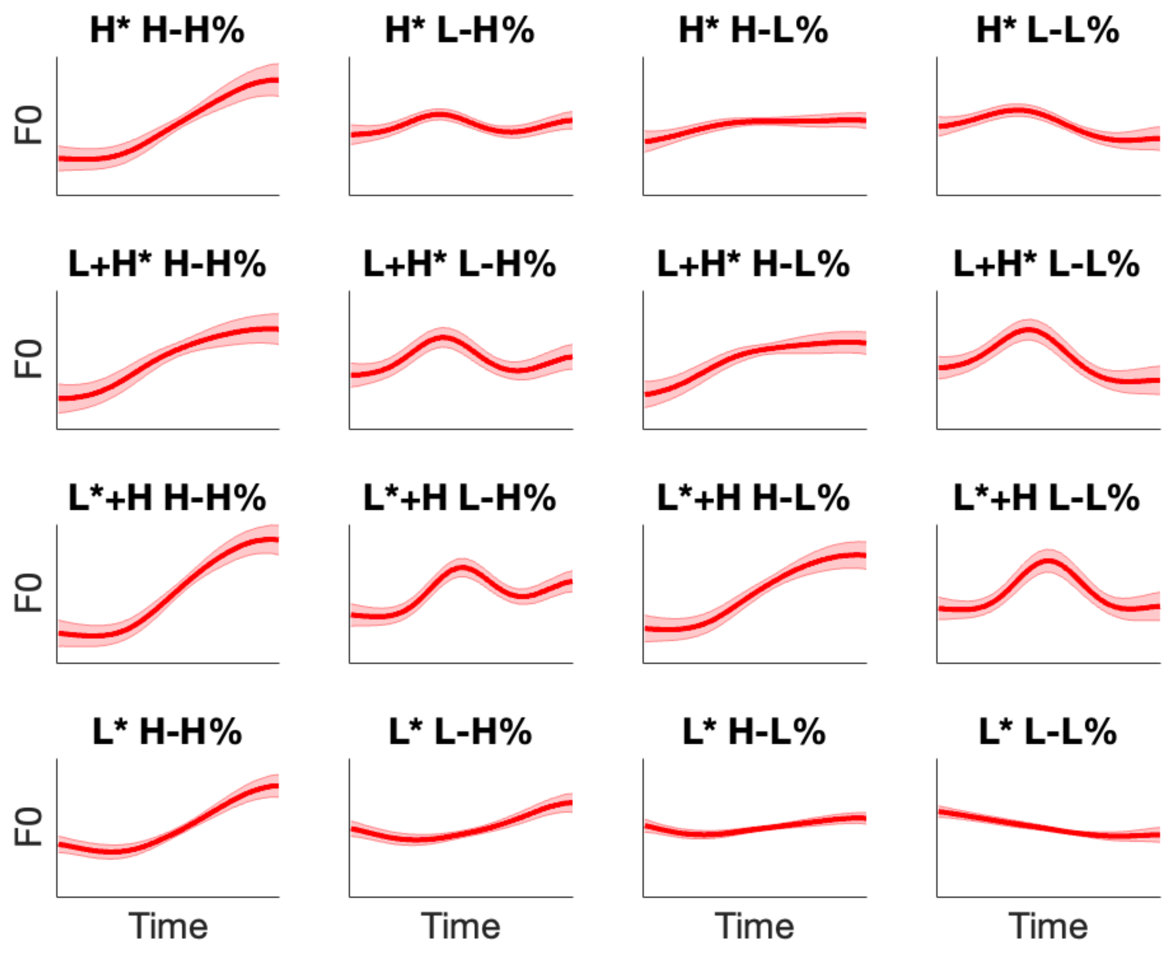




Challenge: What about full nuclear tunes?

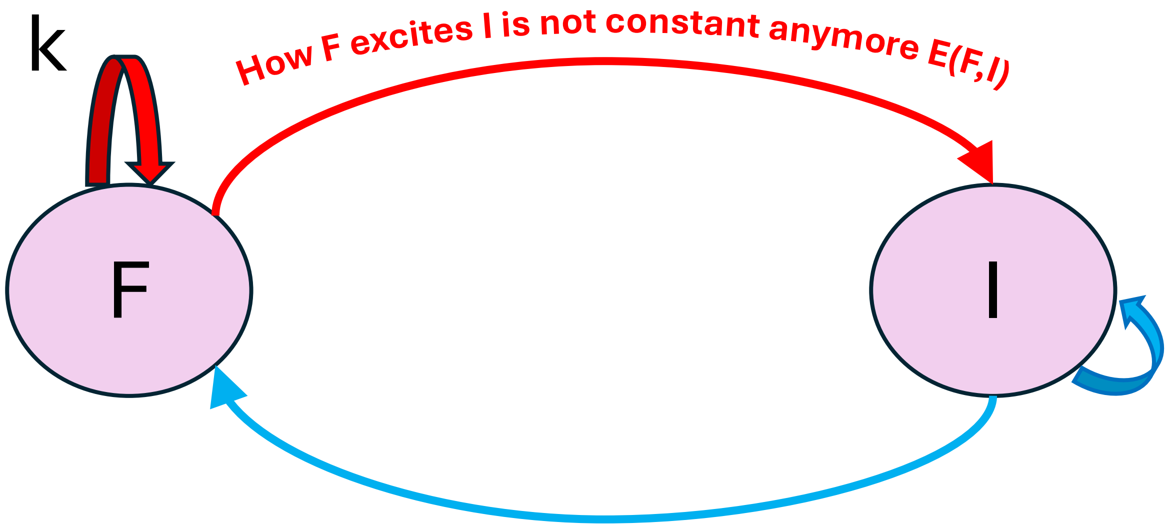
Is it possible to extend the dynamical interaction to account for full tunes?

Specifically, can one heed the advice of P& P and account for all targets in a tune in the ***output***, not ***input*** of dynamics, *and* S & G on autonomy of entire tune?



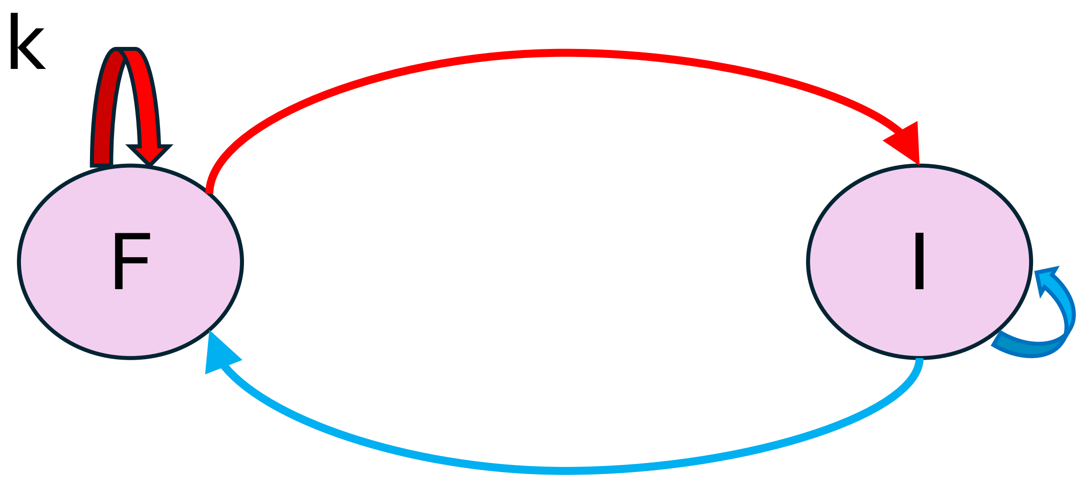
Full AE Tune Dynamics

- Phrase Accent and Boundary Tone are part of the *prosody* of language
- As many before us, we look at prosody as structure (Beckman, 1996)
- We interpret structure as dynamical modulation
- But the modulation doesn't come from a time-varying intelligent nonautonomous force; it is itself triggered by F and I via feedback: dependence of a coefficient on the state (S&G, 2016).



Current theory : <u>F and I tell F how to excite I</u>

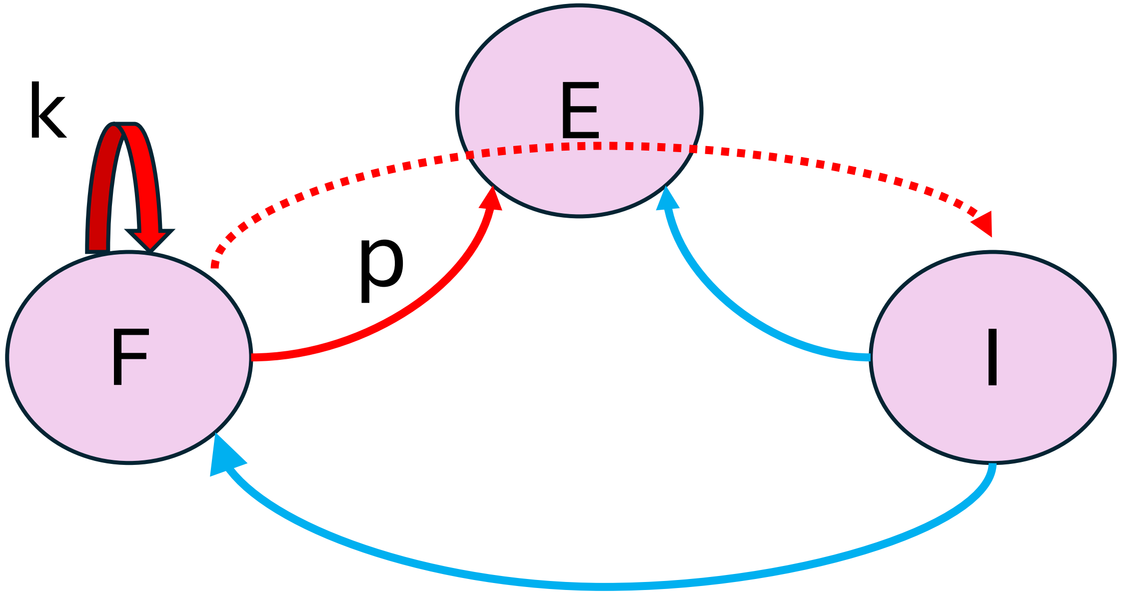
Pitch Accent Dynamics



$$\frac{dF}{dt} = kF - F^3 - I + .5$$
$$\frac{dI}{dt} = EF - I$$

Iskarous et al. (2024)

AE Prosodic Dynamics

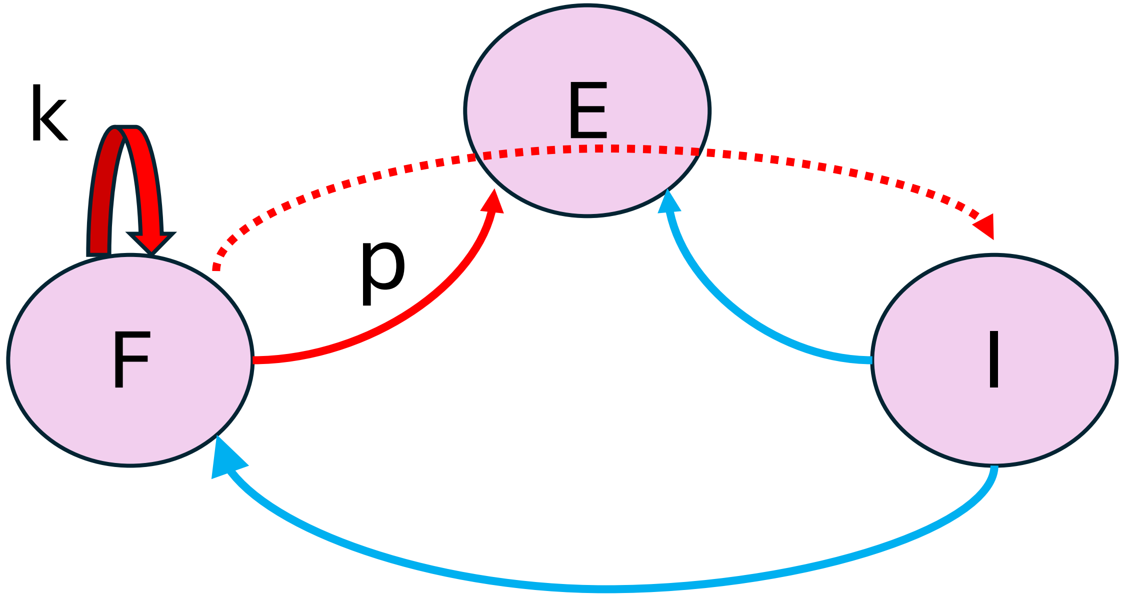


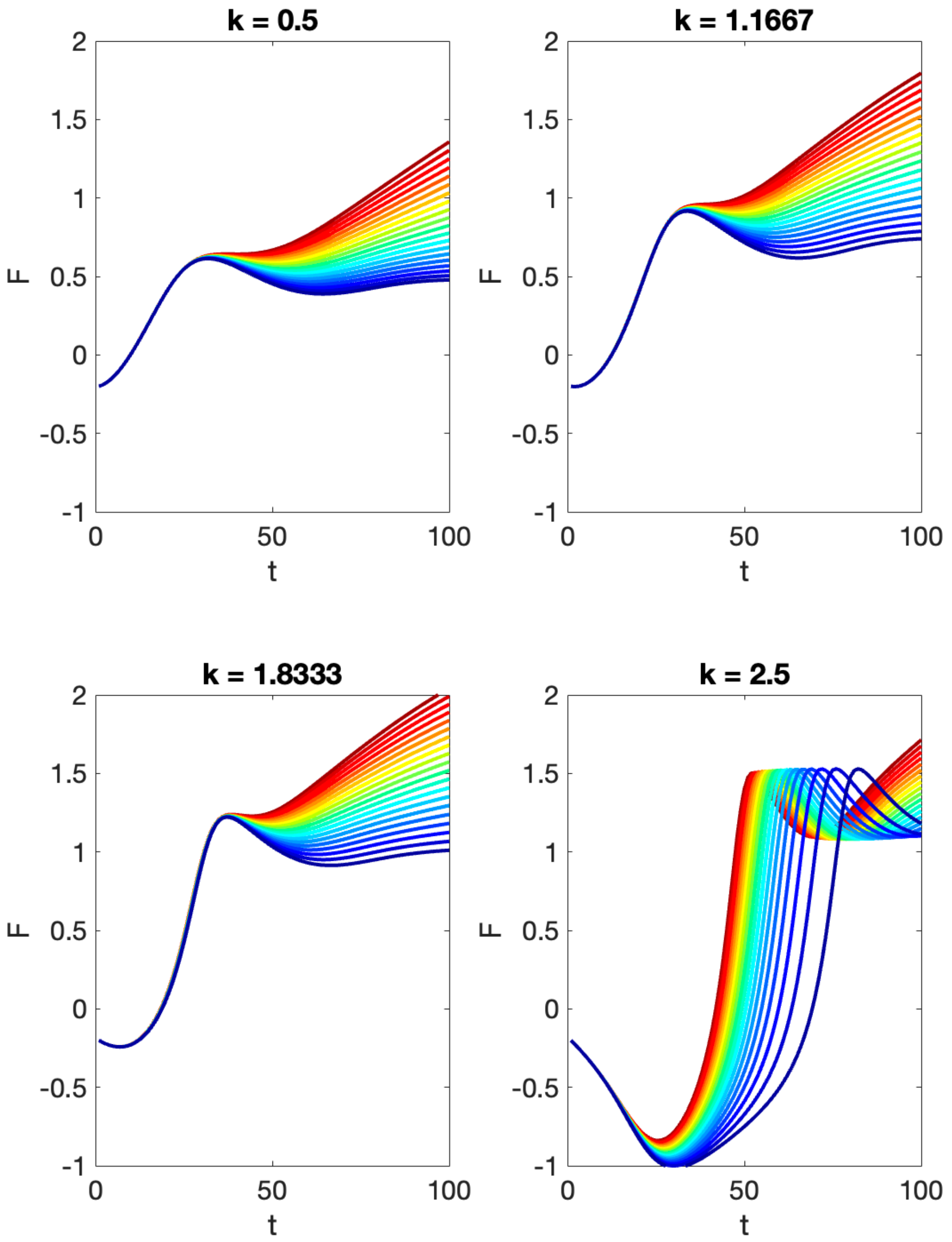
$$\frac{dF}{dt} = kF - F^3 - I + .5$$
$$\frac{dI}{dt} = 3EF - I$$
$$\frac{dE}{dt} = pF - .05I$$

Approximation I

Communication System hands constants *k* and *p* to the phonology

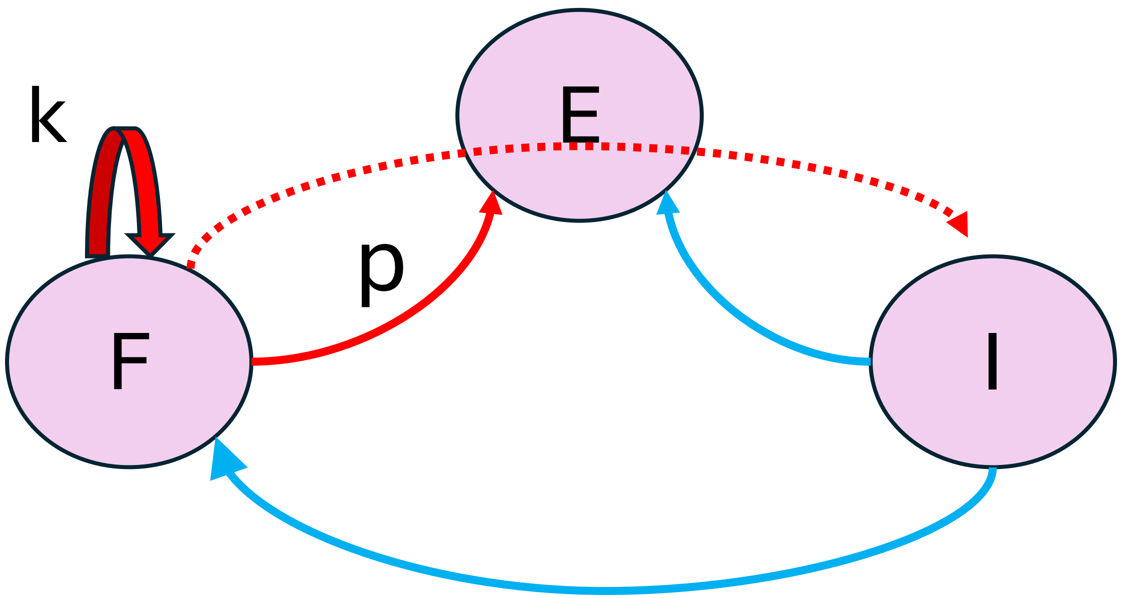
Predictions of the Prosodic Kinetic Theory Approx I

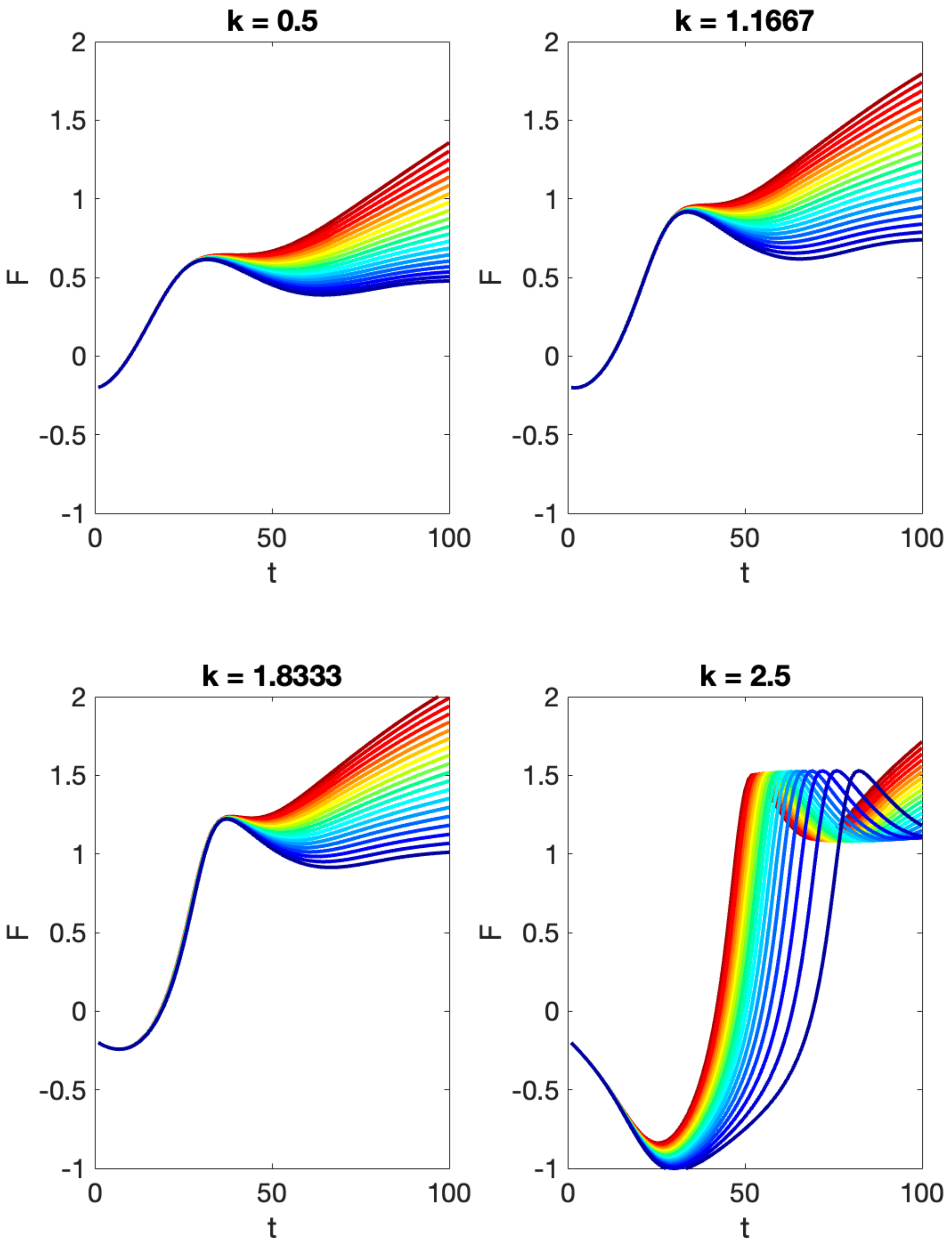




p choice allows F to leave its equilibrium value, and go higher, as in HH or lower than higher, as in LH.

Predictions of the Prosodic Kinetic Theory Approx I

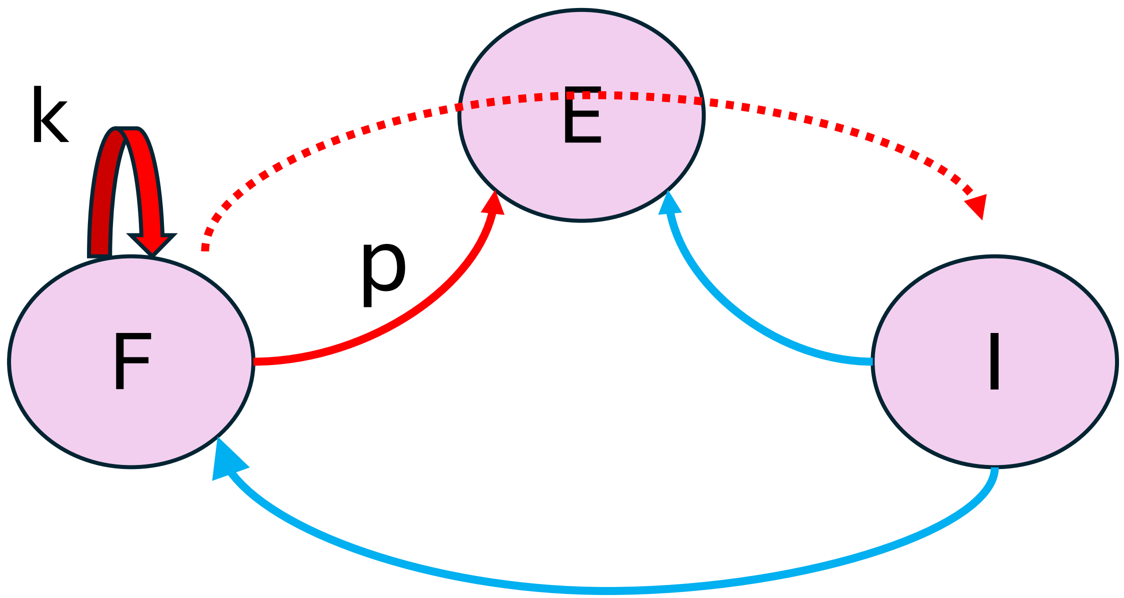


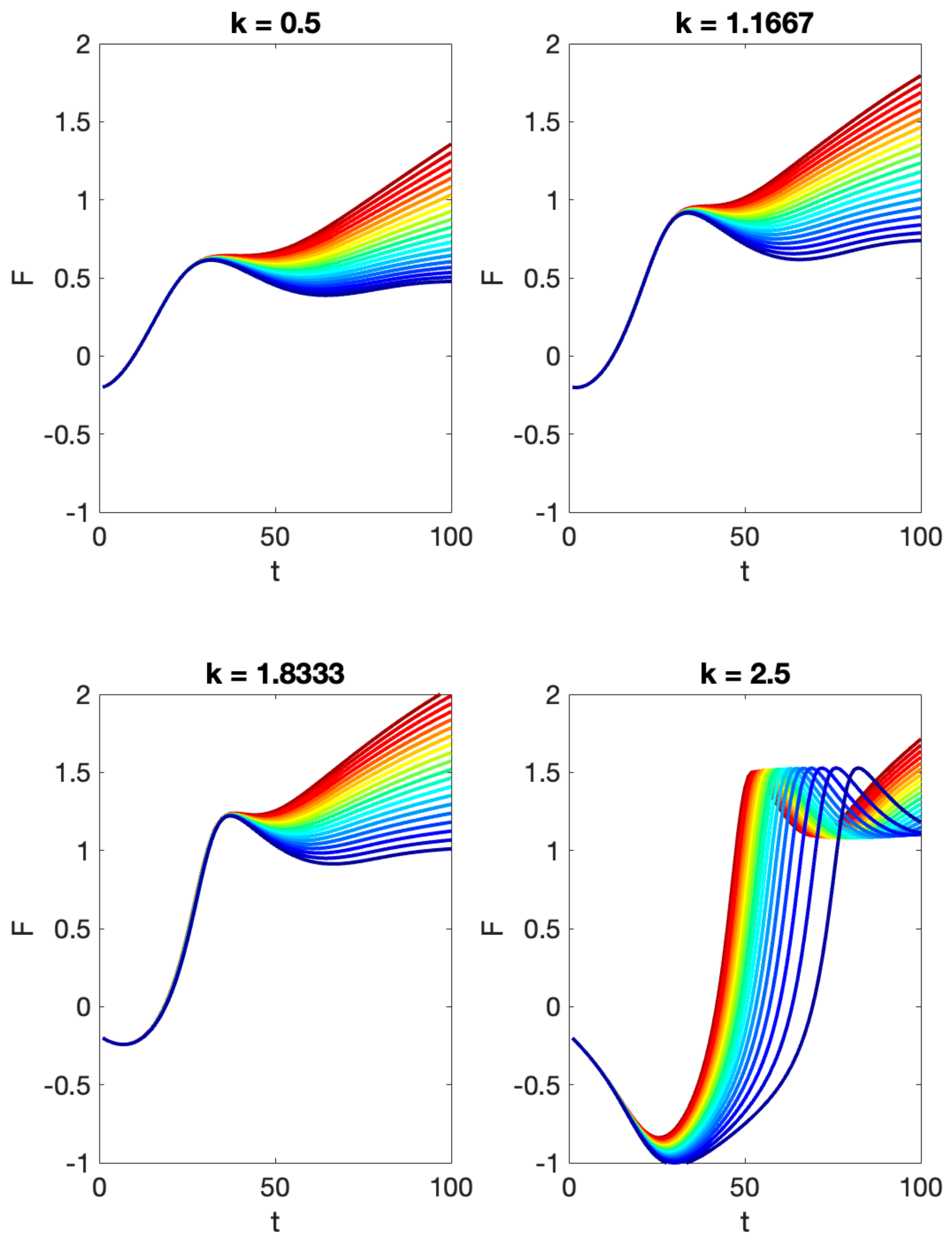


Problems → Approximation II:

- the L of LH is not sufficiently low, and we don't really get L-L%
- H-L% is already present in the Pitch Accent model alone—a problem for L* H-L%
- L* is too *excitable* (yet no oscillation)

Predictions of the Prosodic Kinetic Theory, II & III





Our ongoing work:
Loosening interactions to account for coarticulation between the pitch accent and "edge" regions of the tune

Where we are going…

- Continue work on the $\frac{dE}{dt}$ equation, resolving current limitations in a feedback-fashion (**not** nonautonomously)
- Following Kirkham (2024), we will try to infer the dynamics of F, I and E from our empirical F (f0) data, using SINDy regression (Brunto et al., 2016)
  - This type of regression can also model perception (Iskarous, 2016).
- Our current system is a type of 'winnerless competition' (Lauren, 2002):
  - Motor and cognitive systems perform tasks in sequence – no individual task 'wins', and 'targets' are modeled as temporary saddles.
  - One system of differential equations predicts both the **sequence** and **saddles**.
- Future extensions of this approach to model: Pre-nuclear pitch accents, downstep and interactions with speech timing, syllable, stress, systems without pitch accents, etc.

Where we stand: Local vs. Global Approaches

- Pierrehumbert (1980), Ladd (2008), Arvaniti (2011, 2021) criticize whole-tune global approaches of the IPO school and advocate a theory of local pitch events L and H that compose to form a tune.
- Our work supports the local approach, but is actually *hyperlocal*, as the circuit we propose applies at a far finer scale than the events H and L. It applies at every single point of time.
- However, like the AM theory it is able to predict a global tune.

Conclusions

- Kinetic Autonomy may allow us to *predict*, instead of *postulate,* many structural aspects of linguistic systems.
- Cognitive phonological and motor action can meet quite casually: phonology sets parameters of motor circuits, instead of through postulation of a cognitive dynamics separate, an execution dynamics, and a “translation” interface.
- Kinetic autonomy opens the road to thinking of syllables, words, and perhaps utterances as austonomous actions in and of themselves, not only as compositions of smaller actions.