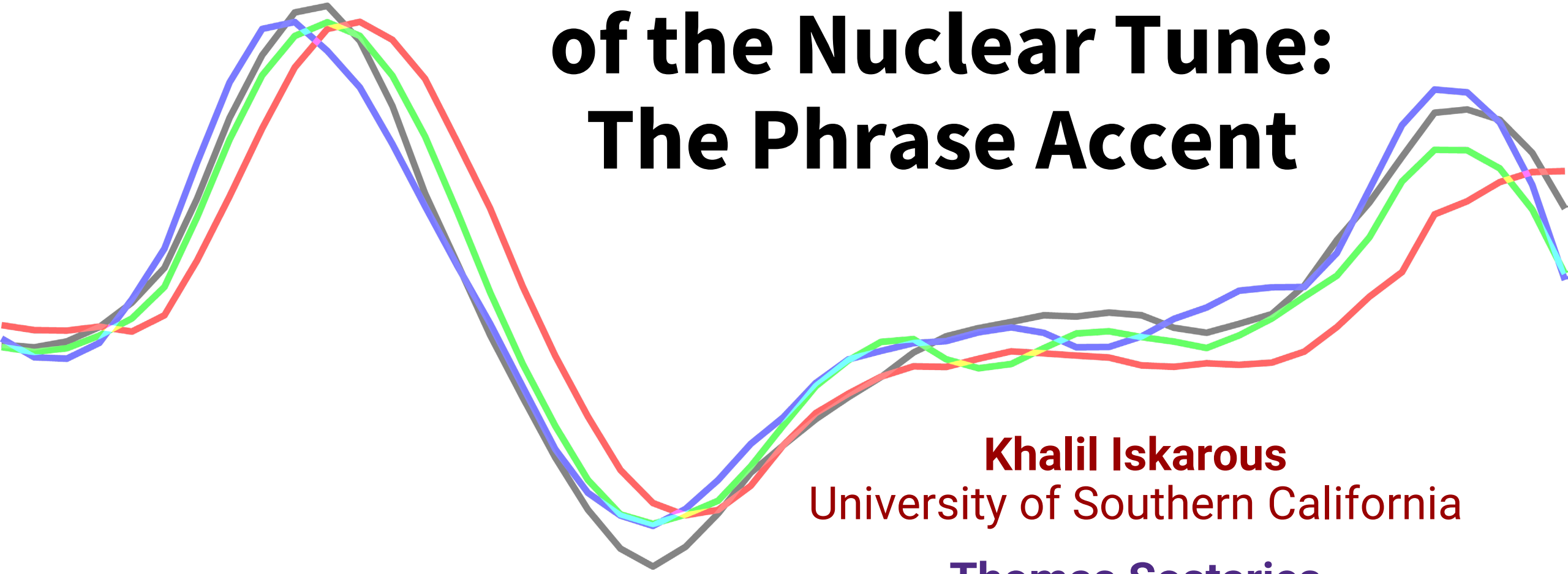
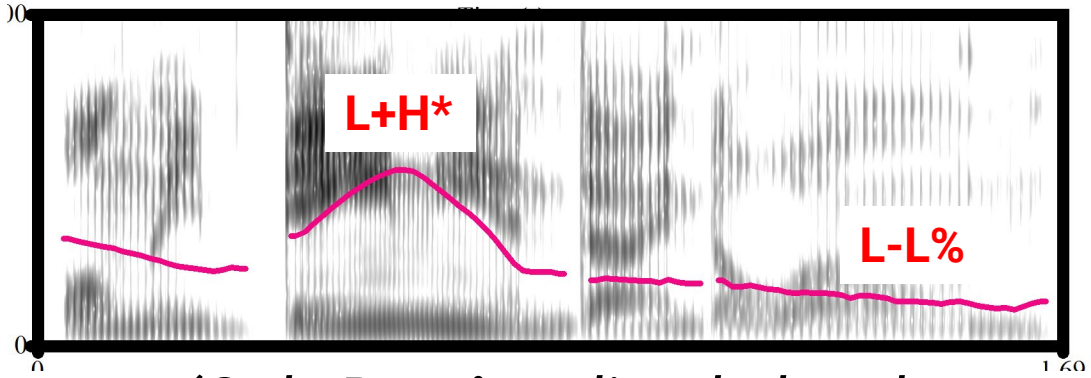
The Dynamical Structure of the Nuclear Tune: The Phrase Accent



Khalil Iskarous
University of Southern California

Thomas Sostarics
Jennifer Cole
Northwestern University

Bridging Continuous & Discrete Representations



*'Only **Damian** dined alone'*

Traditional Account

L+H* L- L% $\xrightarrow{F0\ Algorithm}$ $F0\ shape$

Discrete phonology

Continuous phonetic signal

Proposal: An alternative dynamical theory that *explains* the continuous aspects, *and* from which discrete aspects *emerge*.

A Dynamical Approach to Explaining F0

F0 trajectories arise from a dynamical system with two variables:

These are biologically ***interacting*** neural/muscular agents that jointly determine F0 trajectories

These systems interact via a small number of free parameters

These parameters are what is culturally/phonologically-specified

Examples: Pointing, head nods, finger manipulations in eating, involve muscle synergies informed by cultural norms

Part 1. AE Pitch Accent dynamics

How do ***F*** (F0↑) and ***I*** (F0↓) dynamically interact from *t* to *t+1?*

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

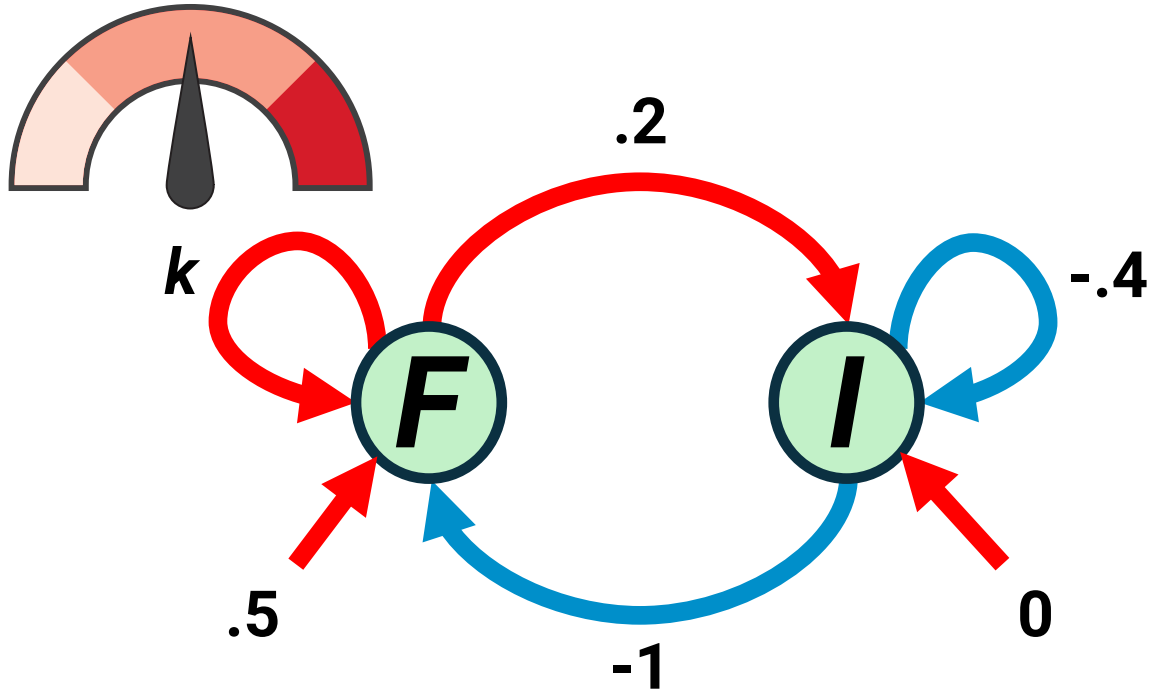
F increases from *t* to *t+1* → *growth*

Starting from small values: 0 < F < 1

I causes ***F*** to decrease → *inhibition*

$$\frac{dI}{dt} = .2F - .4I$$

F causes ***I*** to increase

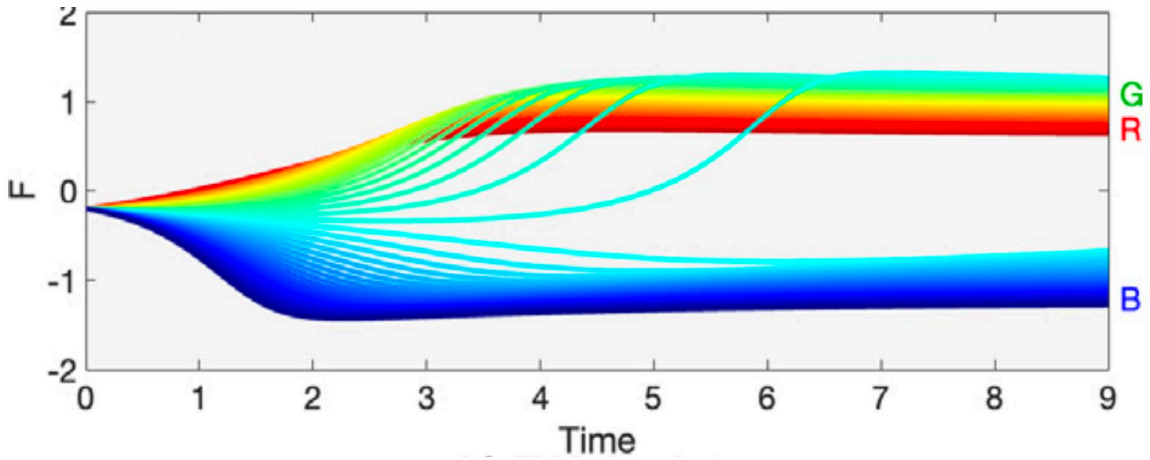


Iskarous, Steffman, Cole (2024) argue that this is the simplest dynamical interaction that generates the observed phonetic F0 trajectories for AE pitch accents.

Part 1. AE Pitch Accent dynamics

$$\frac{dF}{dt} = \boldsymbol{k}F - F^3 - I + .5 \qquad \frac{dI}{dt} = .2F - .4I$$

One free parameter ***k***. Varying only ***k*** generates many F0 trajectories:

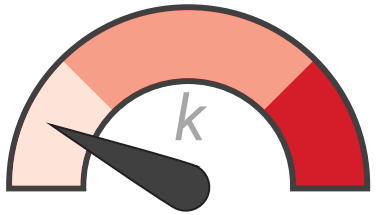


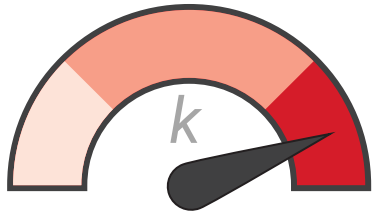
H*: **k ~ .5**

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

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

L*: **k ~ 2**





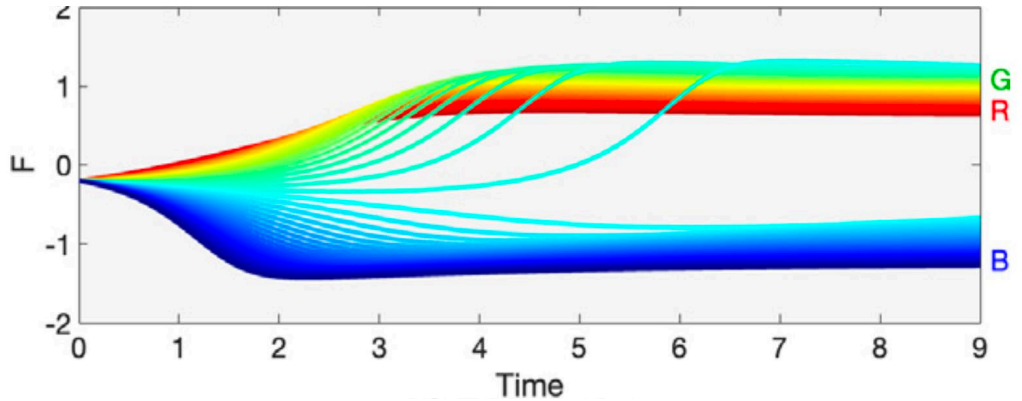
Claim for AE: A few ***k*** values, associated with pragmatic meaning

Part 1. Pitch Accent Dynamics

Phonology is *emergent*

Why?

- L and H values and their **sequencing** emerge from the interaction of F and I: They are not specified in the input (cf., Autosegmental Phonology, Articulatory Phonology)
- The dynamic can generate an H* by itself ***and*** group L + H together into a complex intonational event, e.g., ***k*** ~ 1.5
- This dynamical account also predicts phonetic details of F0 trajectory **shape**, which are not predicted by the AM model (Iskarous, Steffman, Cole, 2024)



H*:	k ~ .5
L+H*:	k ~ 1
L*+H:	k ~ 1.5
L*:	k ~ 2

Part 2. The Nuclear Tune in English

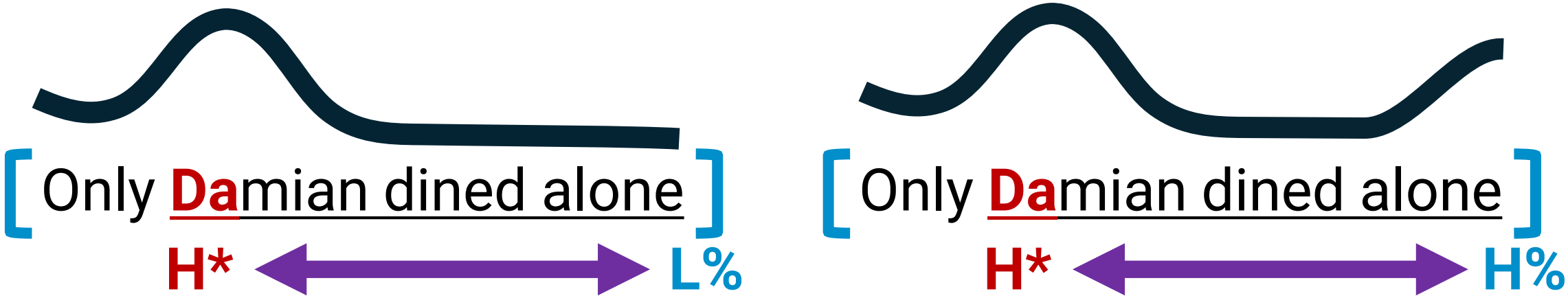
Pitch trajectory at the end of the intonational phrase (IP):

Only Damian dined alone

Only Damian dined alone

Part 2. The Nuclear Tune in English

Pitch trajectory at the end of the intonational phrase (IP):



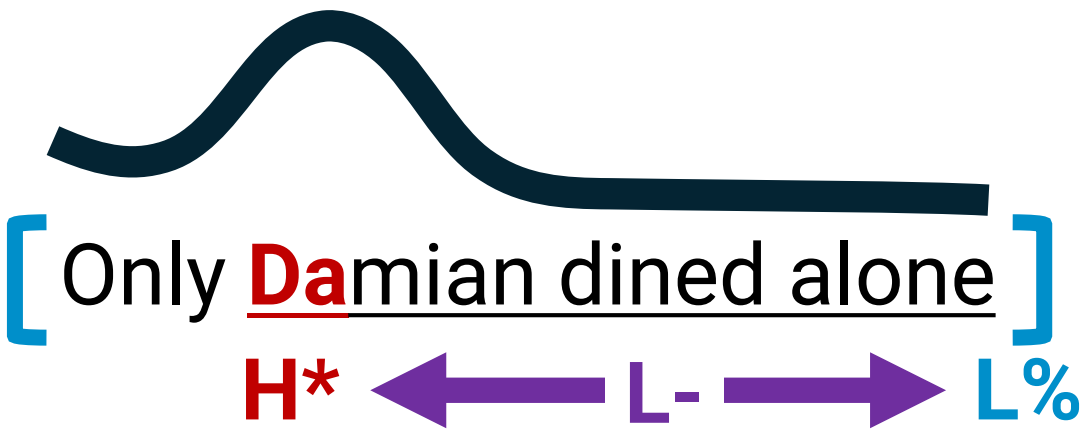
Pitch accents lend **prominence to a stressed syllable (*)**

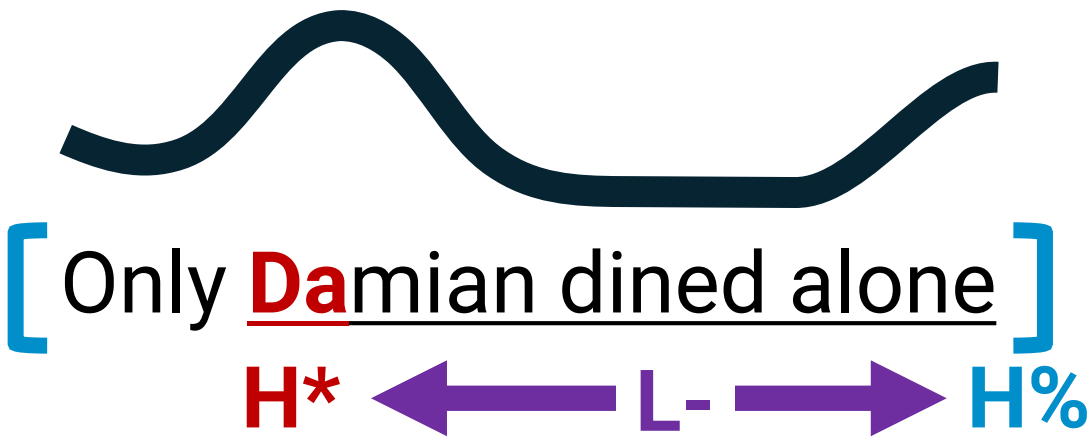
Boundary tones mark the **right edge of an IP (%)**

What happens in between?

Part 2. The Nuclear Tune in English

What happens in between?





In the Autosegmental-Metrical (AM) model, this middle region is governed by the **phrase accent (-)**

Grouping Hypotheses

Only Damian dined alone

No grouping: **H*** **L-** **L%**

Pitch accent & Phrase accent: **[H*** **L-]** **L%**

Phrase accent & Boundary tone: **H*** **[L-** **L%]**

Aim: Test speakers' productions for F0 patterns that reveal grouping preferences through **stable timing patterns**

F0 Timing: Data from Tune Imitation

Participants imitate resynthesized F0 trajectories where we have perturbed the location of the phrase accent “target” (High/Low)

→ Introduces variation in timing and F0 curvature in the phrase accent interval

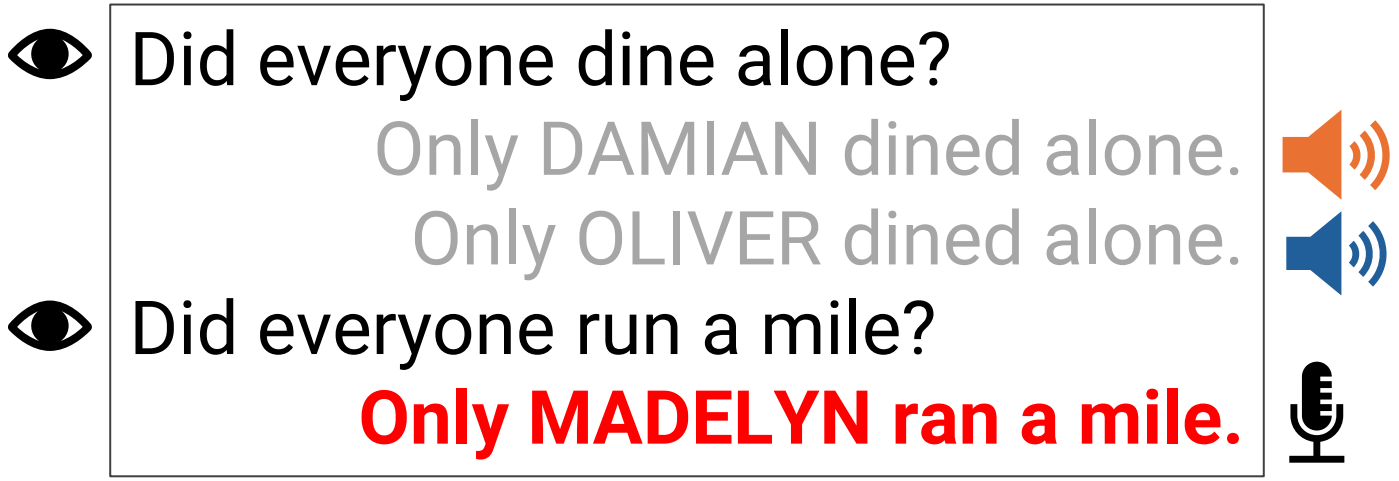
Prior work compared productions of phonologically specified tunes

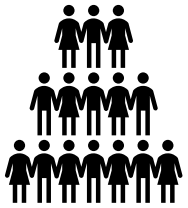
Steffman, Cole & Shattuck-Hufnagel, 2024; Cole, Steffman, Shattuck-Hufnagel & Tilsen, 2023; Steffman & Cole, 2024

Here: Test imitated productions for distinctions in timing patterns among different F0 implementations of an individual tune

F0 Timing: Data from Tune Imitation

Participants listen to 2 stimuli that present the same variant of the same tune, then produce a new sentence with the tune they heard.

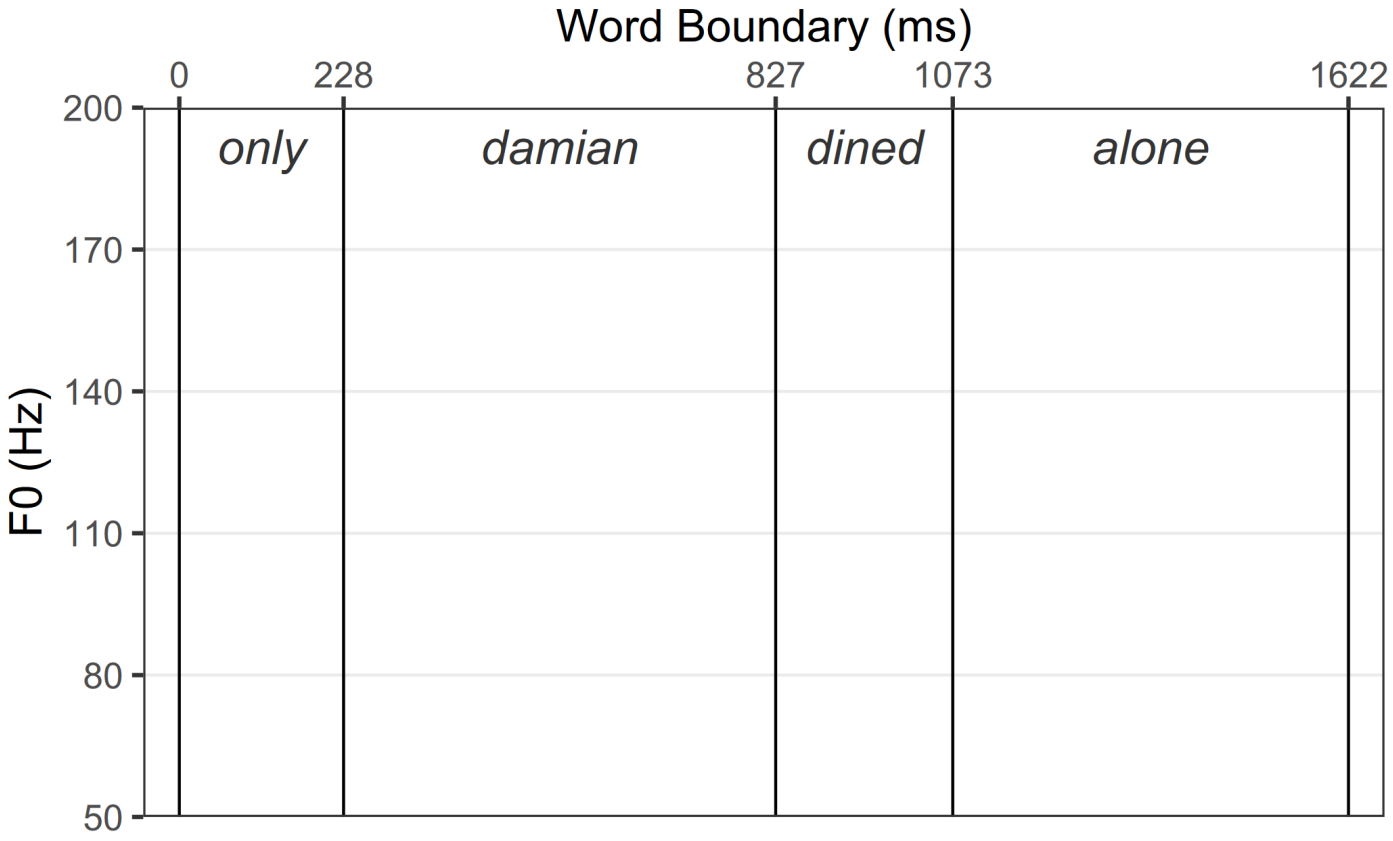




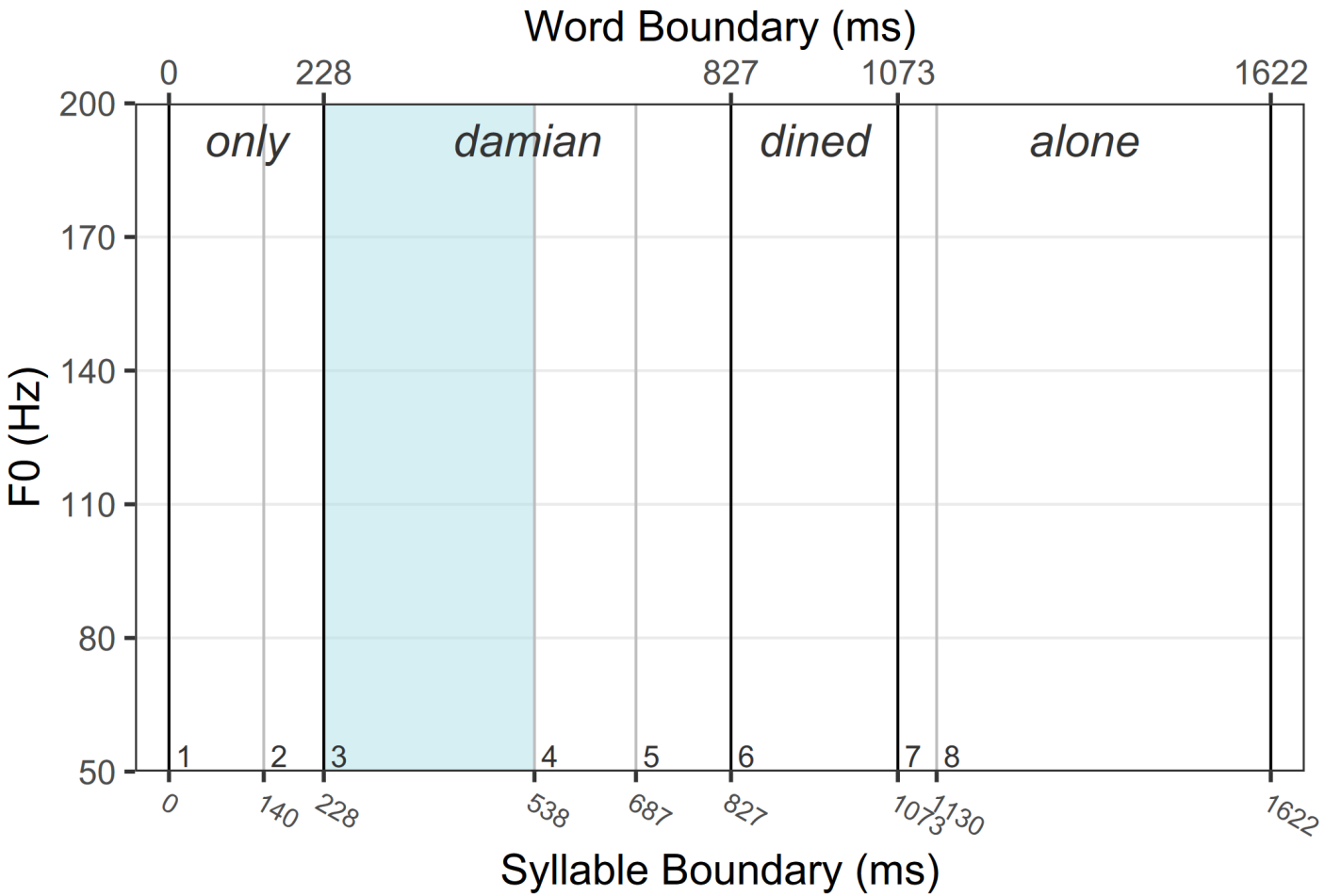
24 participants (Prolific)

144 trials: 3 tunes * 4 shapes * 2 lengths * 2 sentences * 3 reps

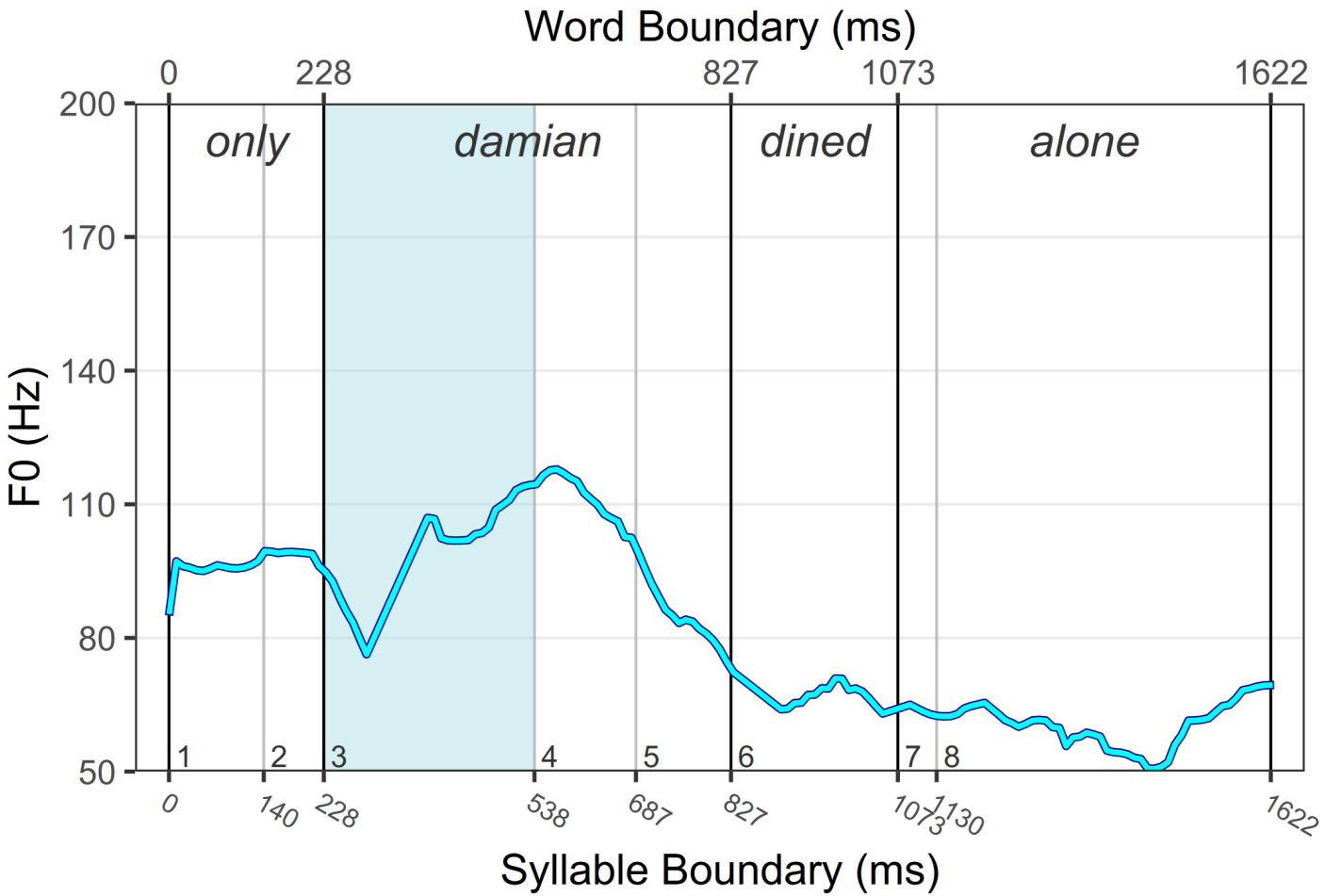
Materials Ex: “Only Damian dined alone”



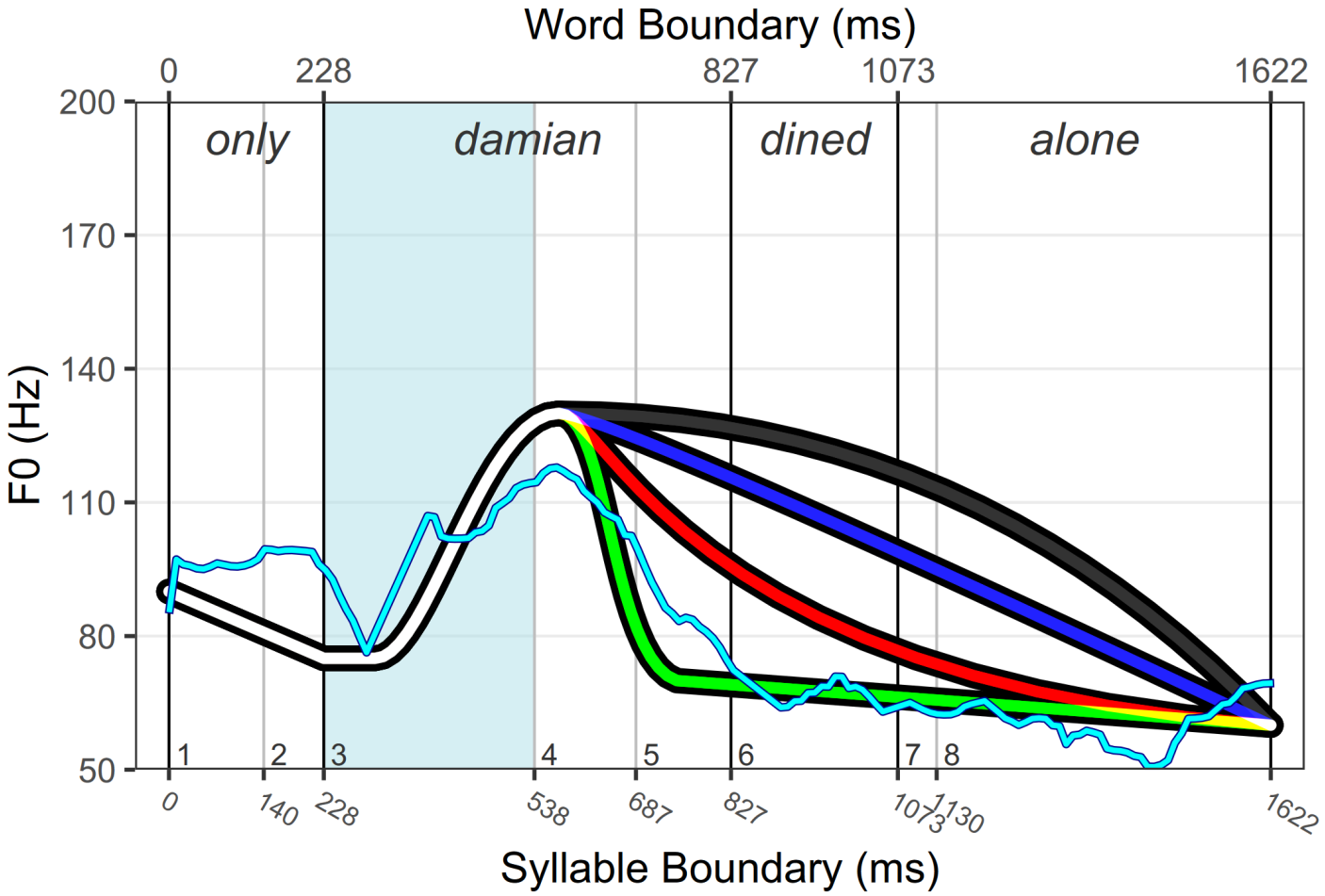
Materials

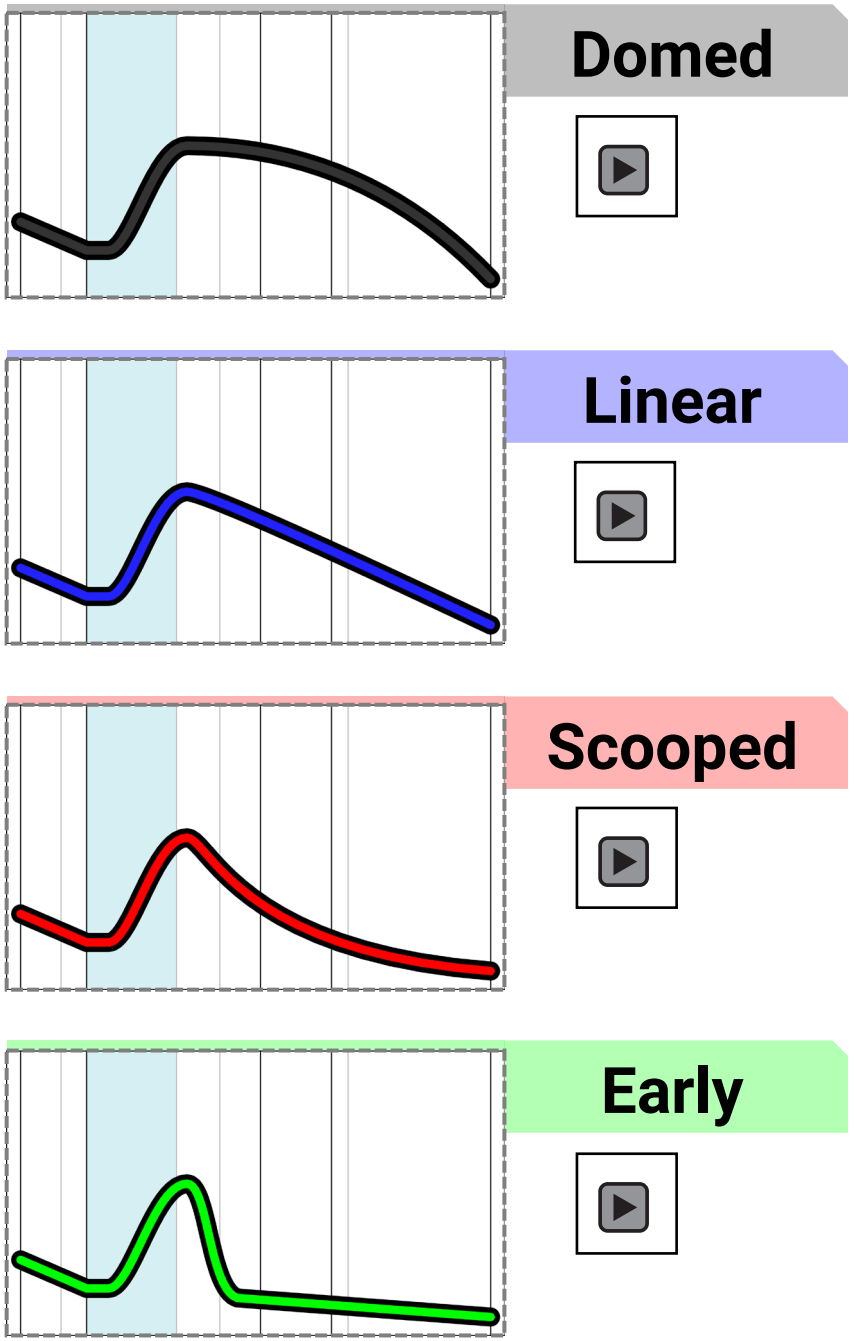

Word Boundary (ms)
0
228
827
1073
1622
200
170
140
110
80
50
F0 (Hz)
only
damian
dined
alone
1
2
3
4
5
6
7
8
0
140
228
538
687
827
1073
1130
1622
Syllable Boundary (ms)

Materials

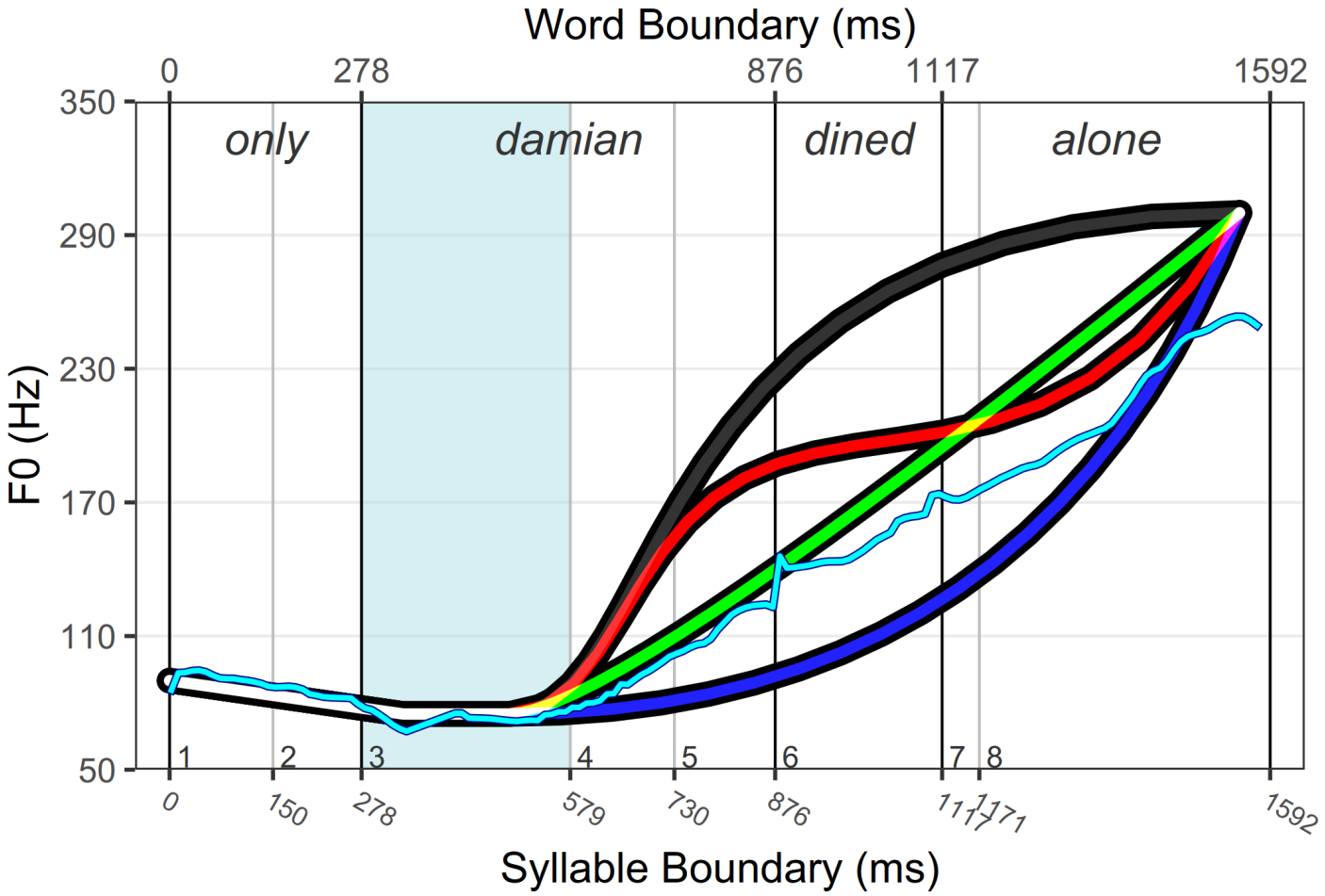

Word Boundary (ms)
0
228
827
1073
1622
only
damian
dined
alone
200
170
140
110
80
50
F0 (Hz)
1
2
3
4
5
6
7
8
0
140
228
538
687
827
1073
1130
1622
Syllable Boundary (ms)

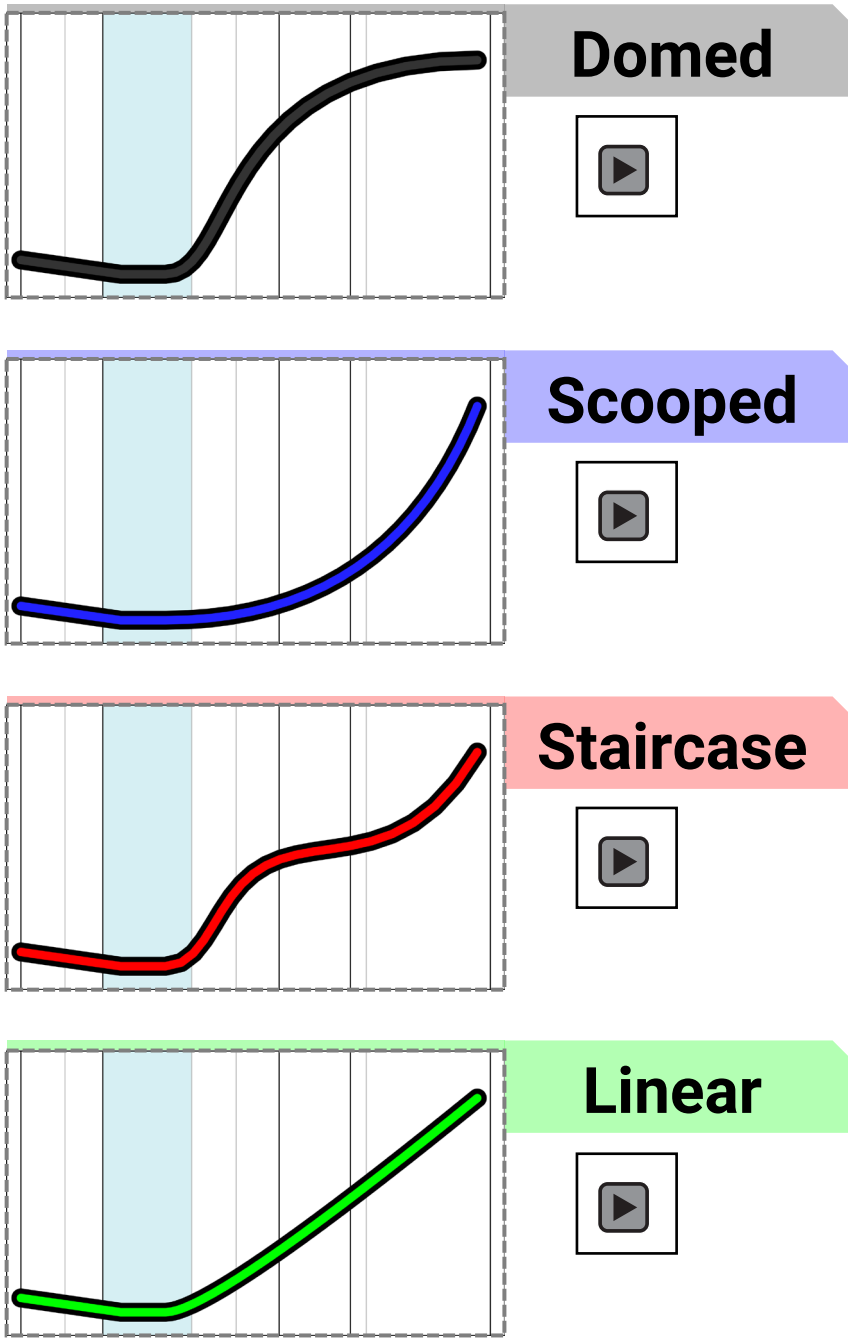
Materials (Falls, LH*LL)



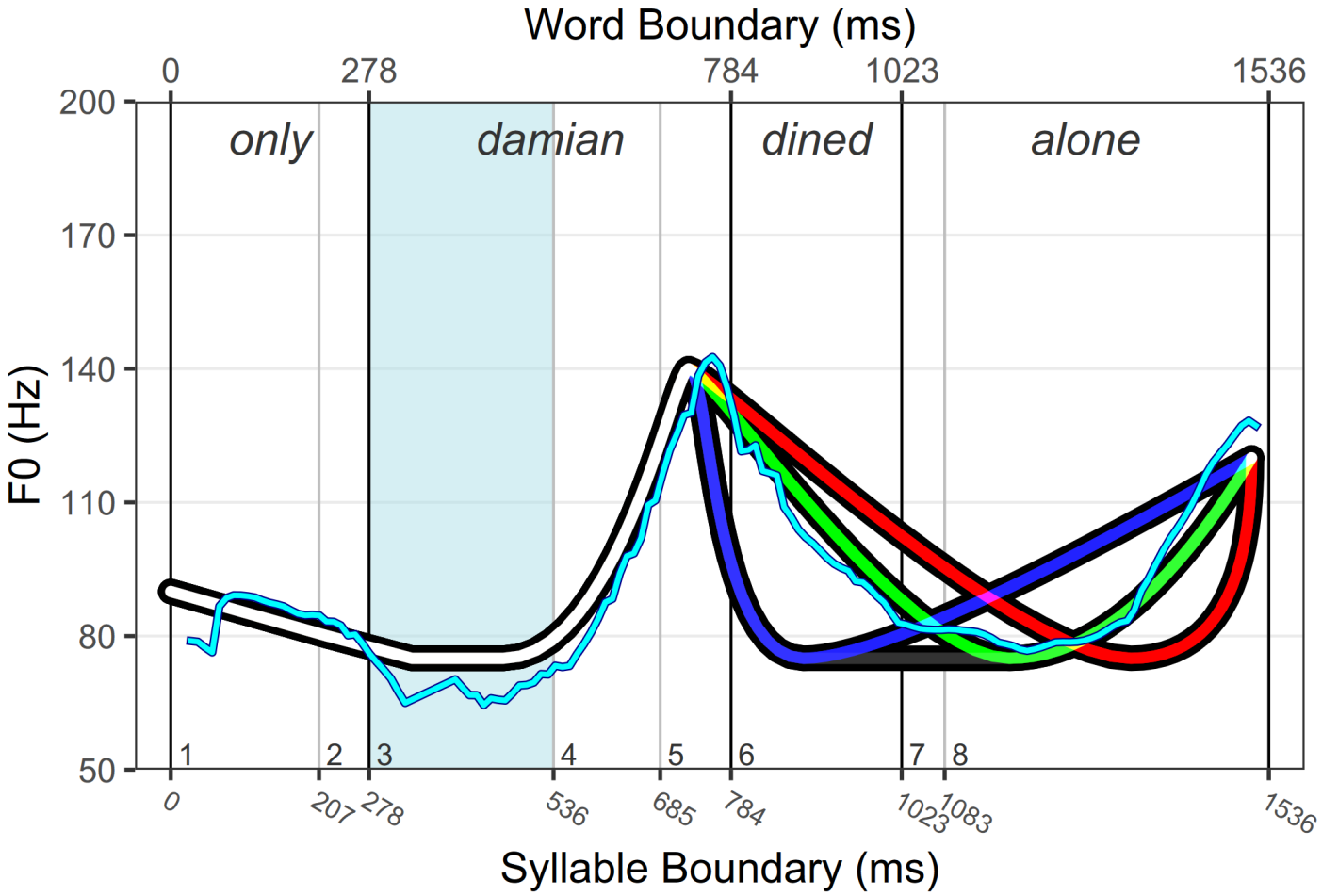


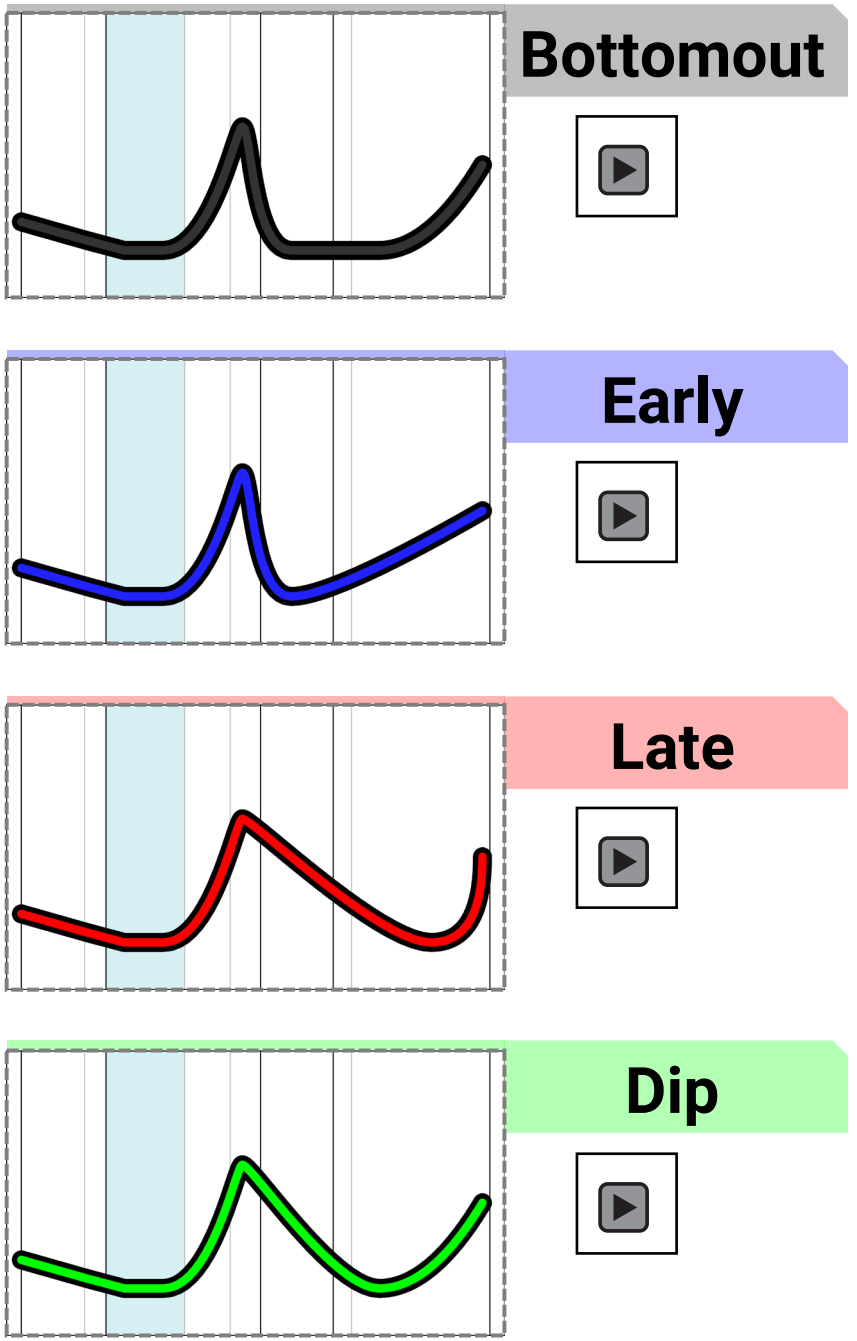
Materials (Rises, LHH)





Materials (RFRs, L*HLH)





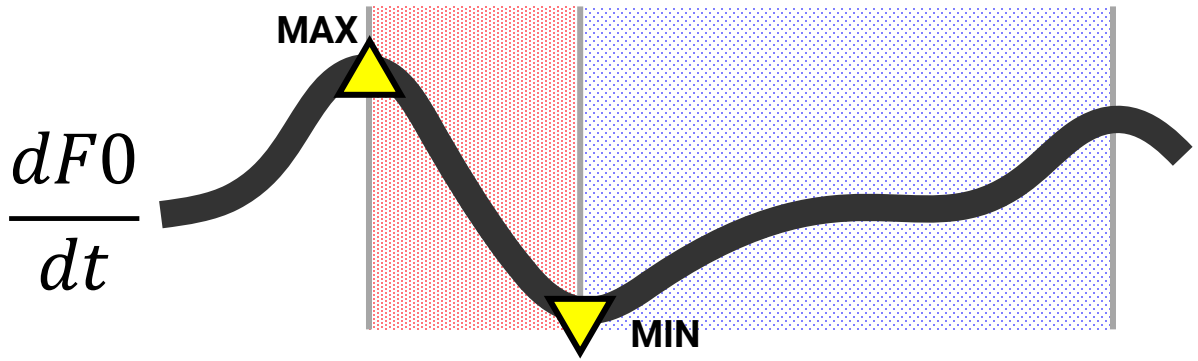
Mean F0 Contours

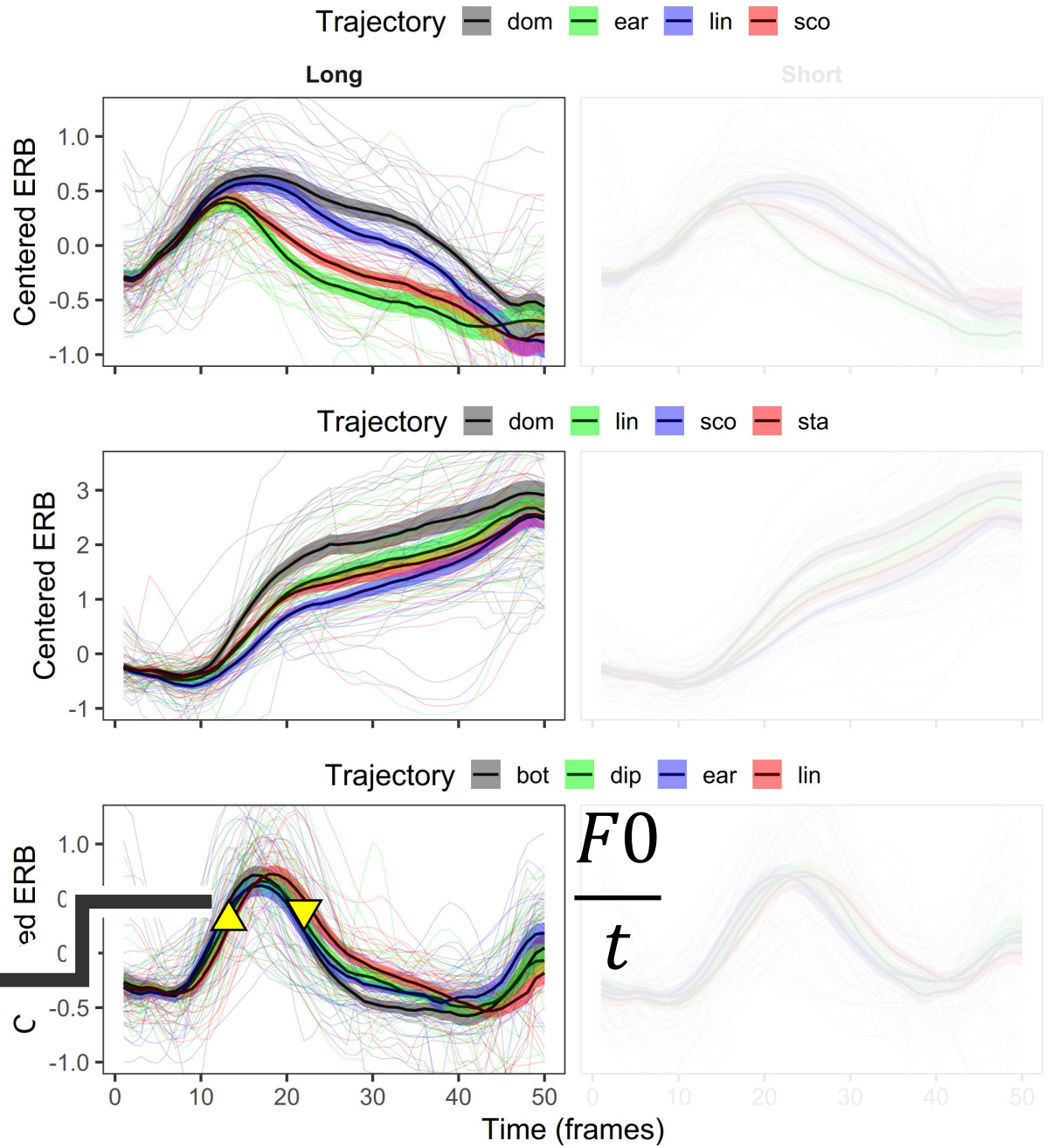
by-speaker, by-trajectory, by-length

What are the patterns of **coordinated stable timing** in the nuclear interval?

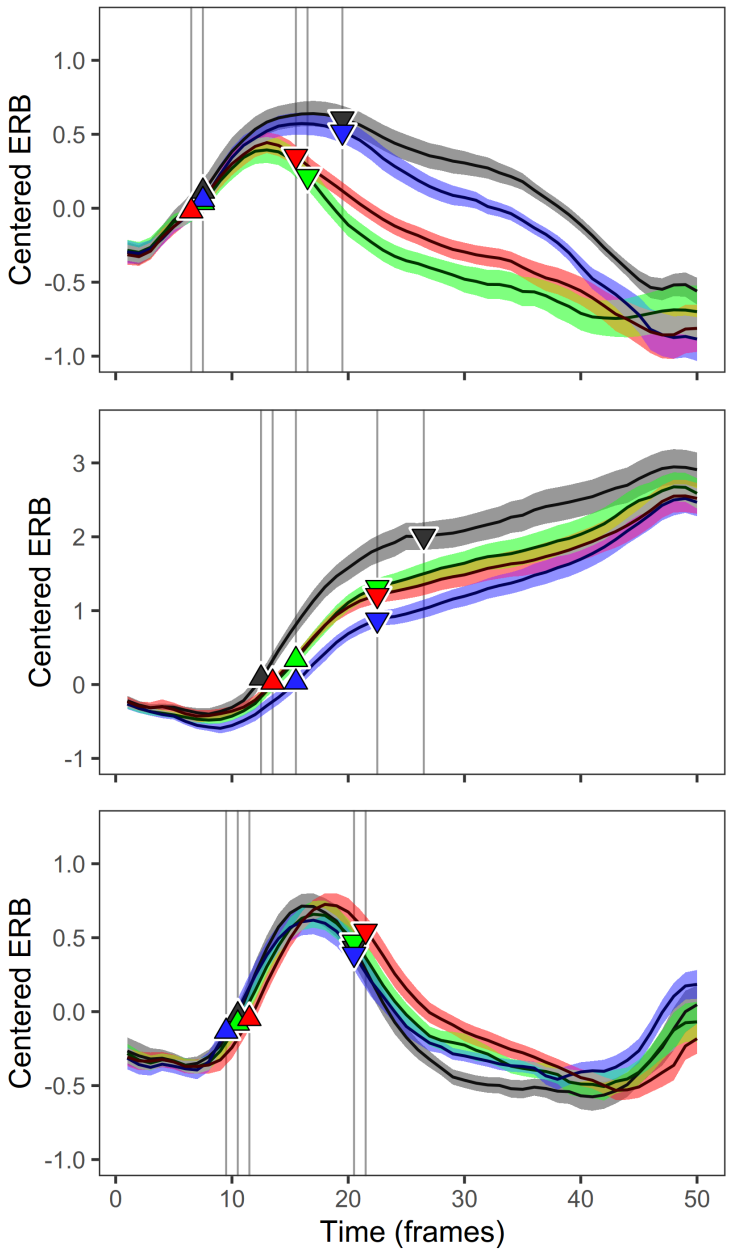
Need to look at the **timing between F0 events:** velocity maxima/minima

Pitch Accent → Phrase Accent → Boundary Tone

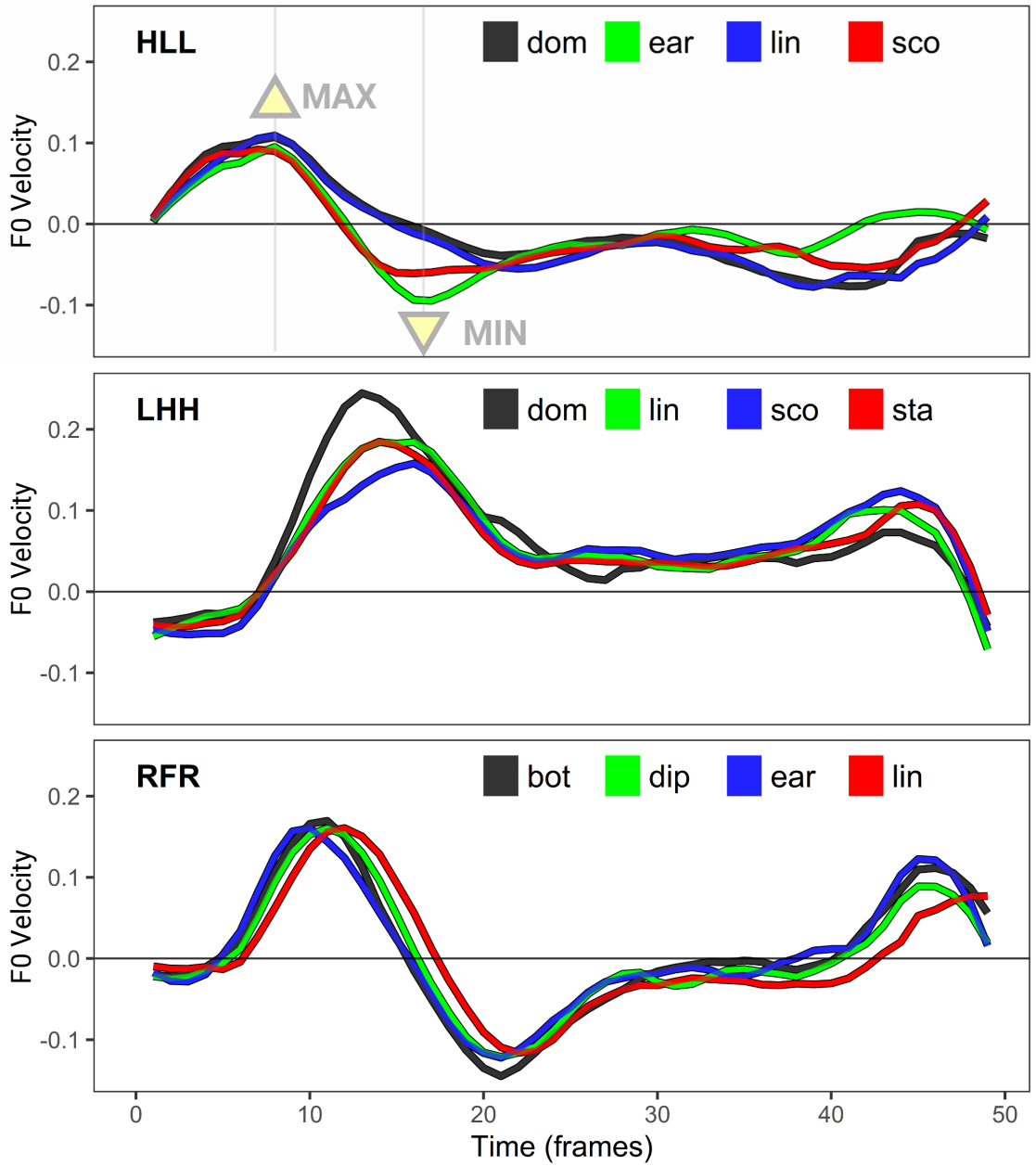




F0 Contours



Velocity Profiles



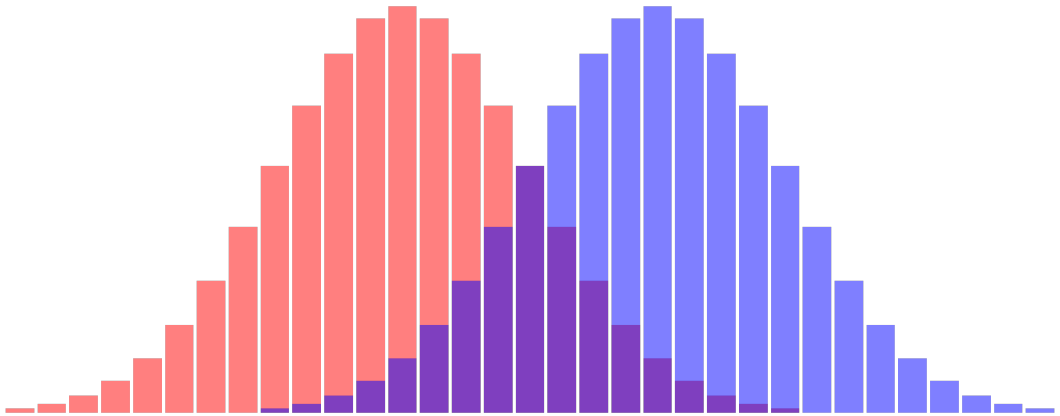
Relating *Grouping* to *Velocity Lags*

Pitch Accent & Phrase Accent

Phrase Accent & Boundary Tone

H* ⟷ **L-** ⟷ **L%**

Histograms plots the **distance between velocity inflection points** (max/min) for the distributions of these two measurements:



Lag (Frames)

Hypotheses

H*

L-

L%

[H*

L-]

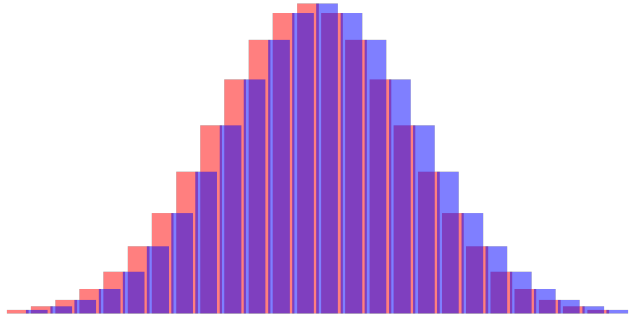
L%

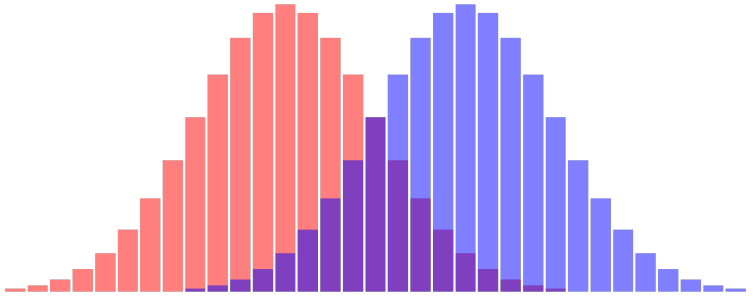
H*

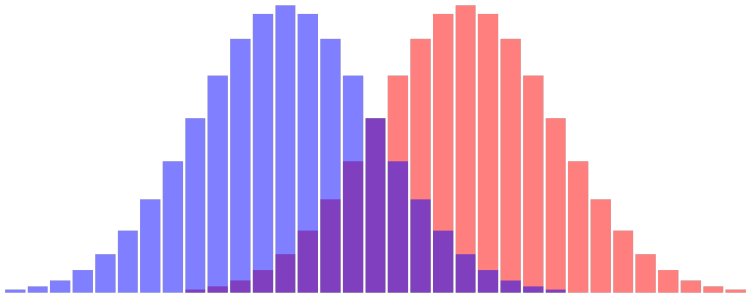
[L-

L%]

Predictions

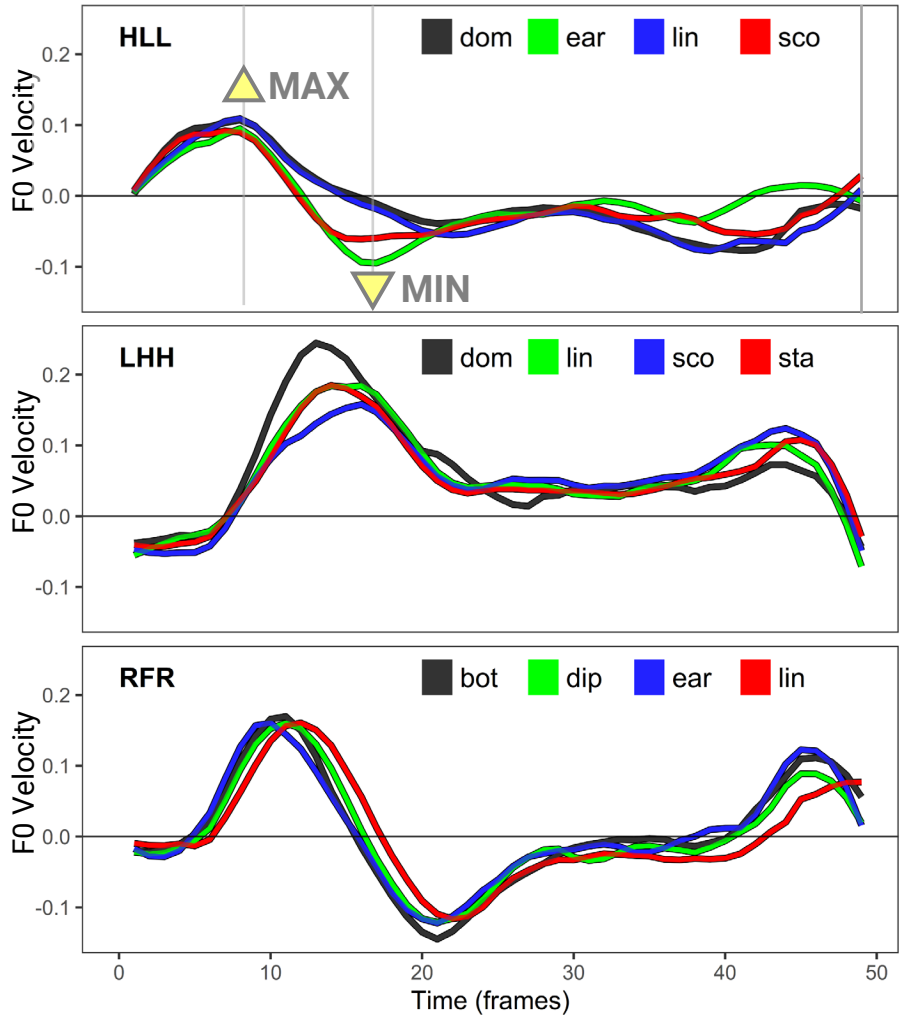


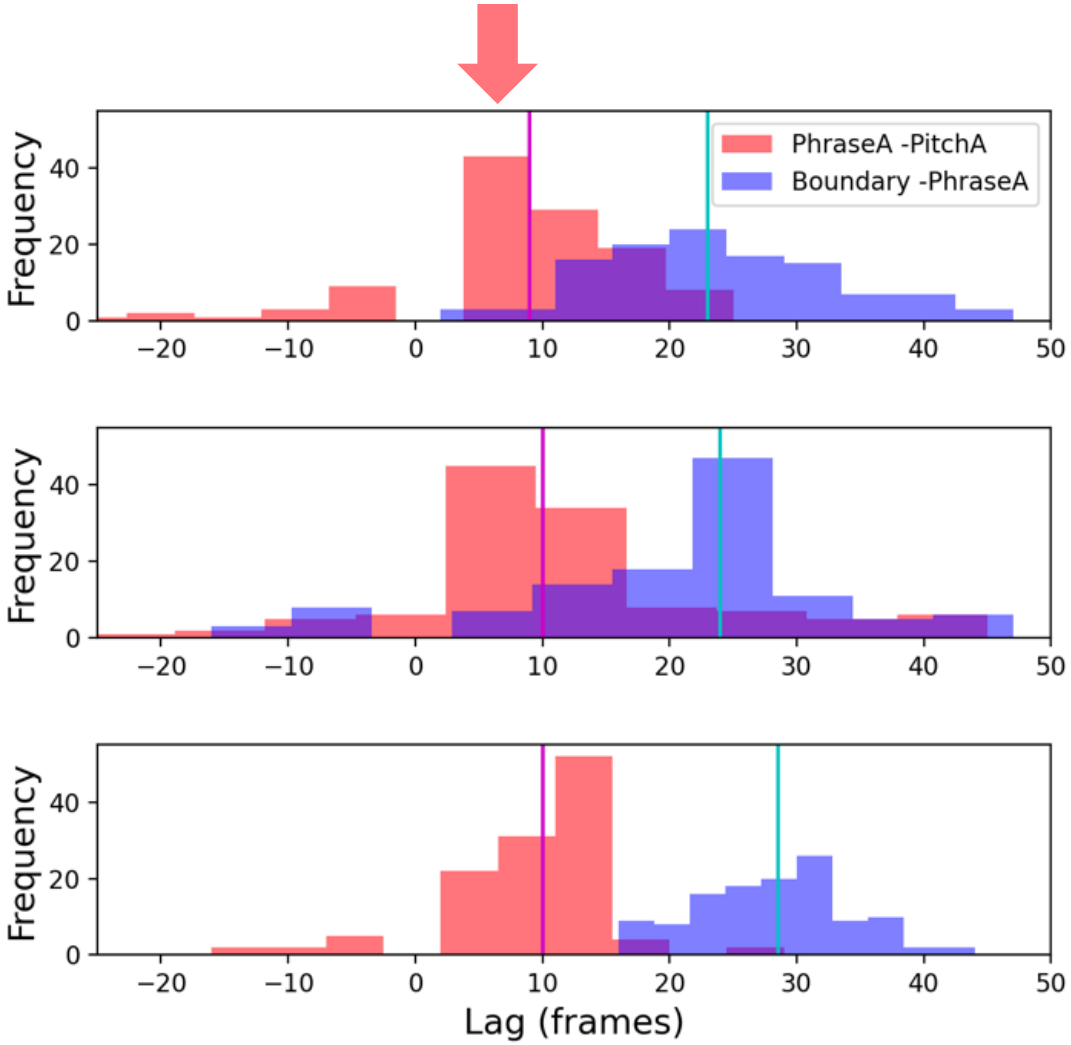




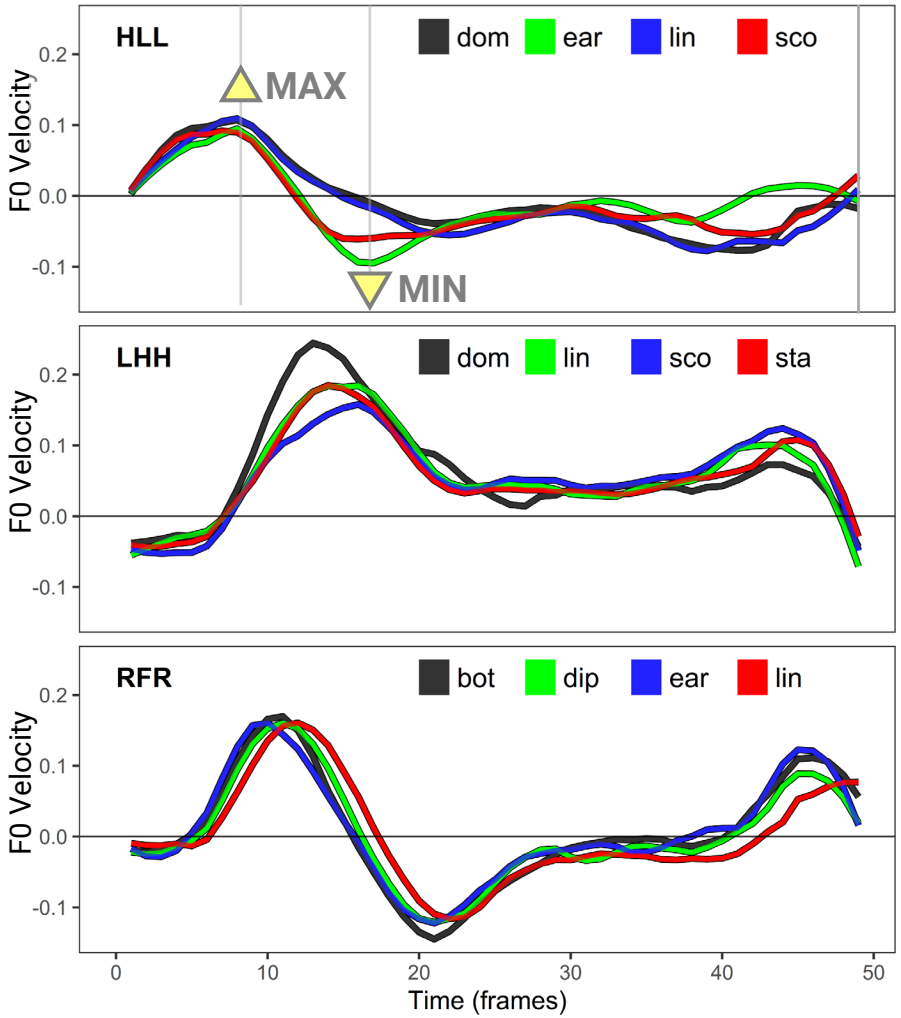
Velocity Lag Distributions

The phrase accent shows more consistent timing with the **pitch accent**, not the boundary tone





Equilibria: Regions of Zero Velocity

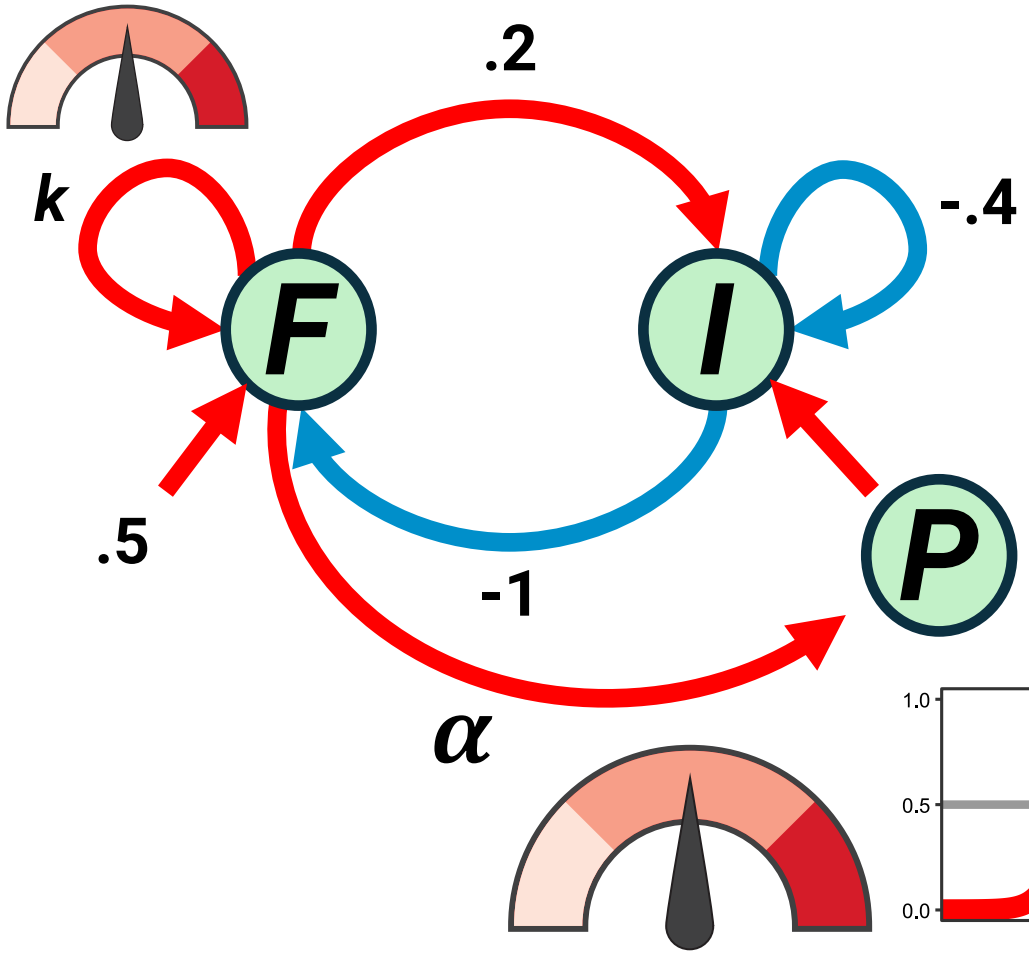


We find regions of near-zero F0 velocity (equilibrium) **after** the minimum velocity of the phrase accent

This pattern provides evidence for a **complex dynamical action** composed of the **pitch accent *and* phrase accent** events

This complex event reaches equilibrium long before the boundary tone

Pitch Accent + Phrase Accent: A Single Complex Event

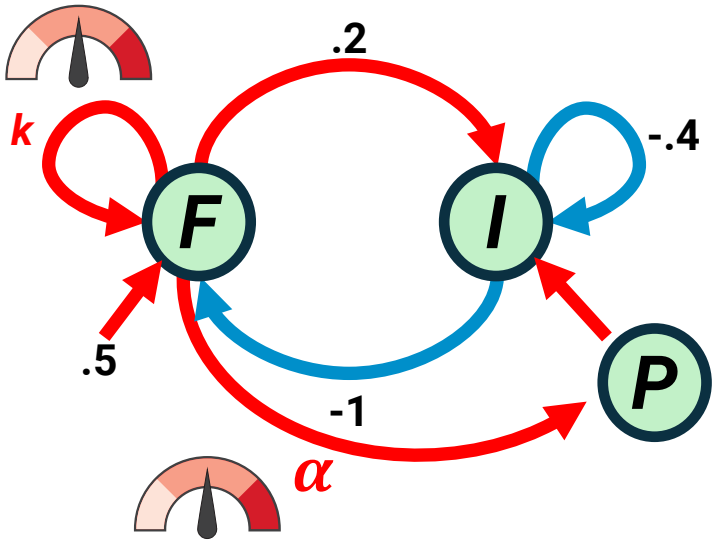


The extended system has a new parameter ***P*** which modulates the inhibiting effect of ***I*** on ***F***.

Modulation is achieved via a logistic function with a free parameter α, generating possible behaviors after the pitch accent.

$$\frac{dP}{dt} = \alpha\left(\frac{1}{1+e^{-(F-1)}}\right)$$

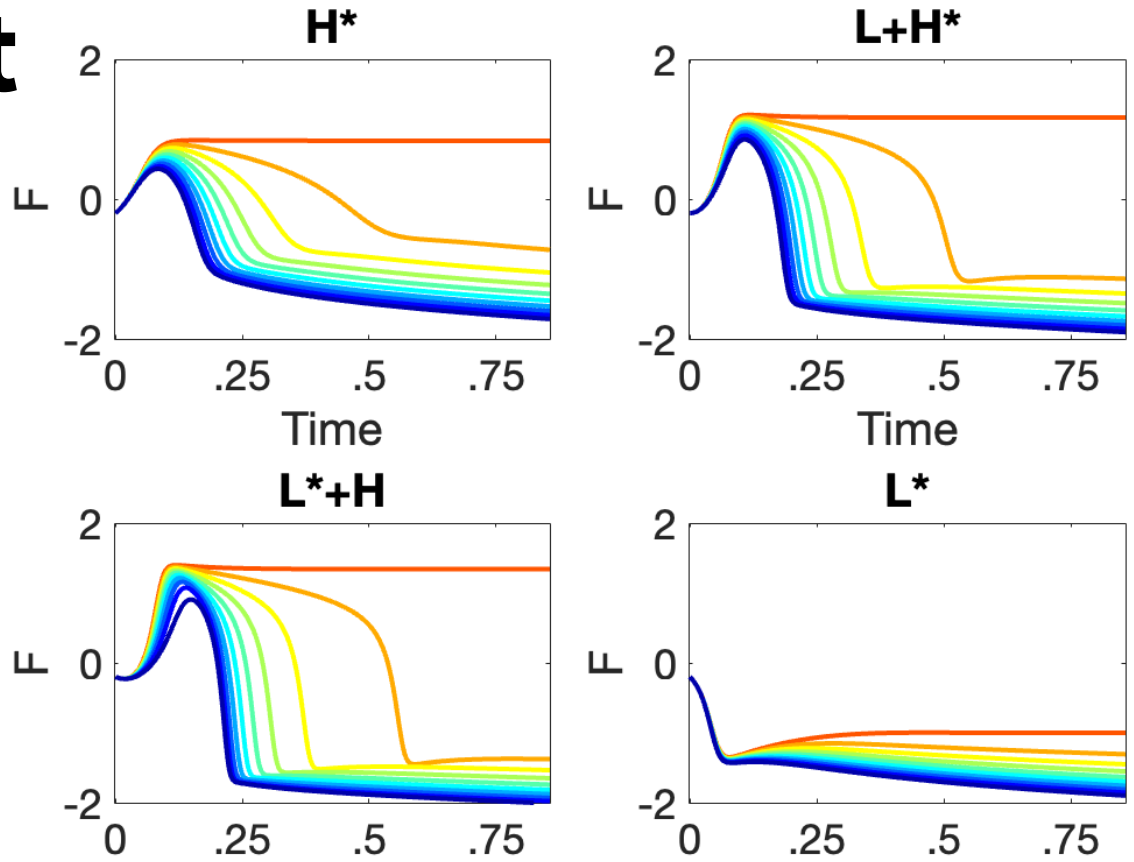
Pitch Accent + Phrase Accent: A Single Complex Event



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

$$\frac{dI}{dt} = .2F - .4I + P$$

$$\frac{dP}{dt} = \alpha\left(\frac{1}{1 + e^{-(F-1)}}\right)$$



The Pitch Accent + Phrase Accent interval is generated by this system with three parameters, ending in an equilibrium corresponding to the Phrase Accent 'target'

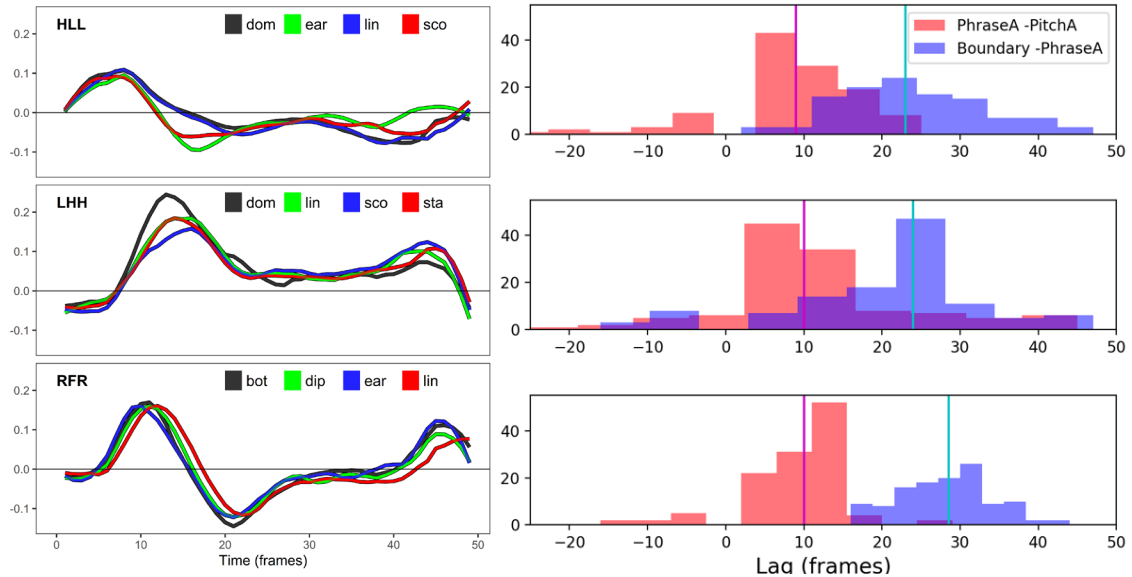
The boundary tone is not (yet) modeled by this system. *Stay tuned!*

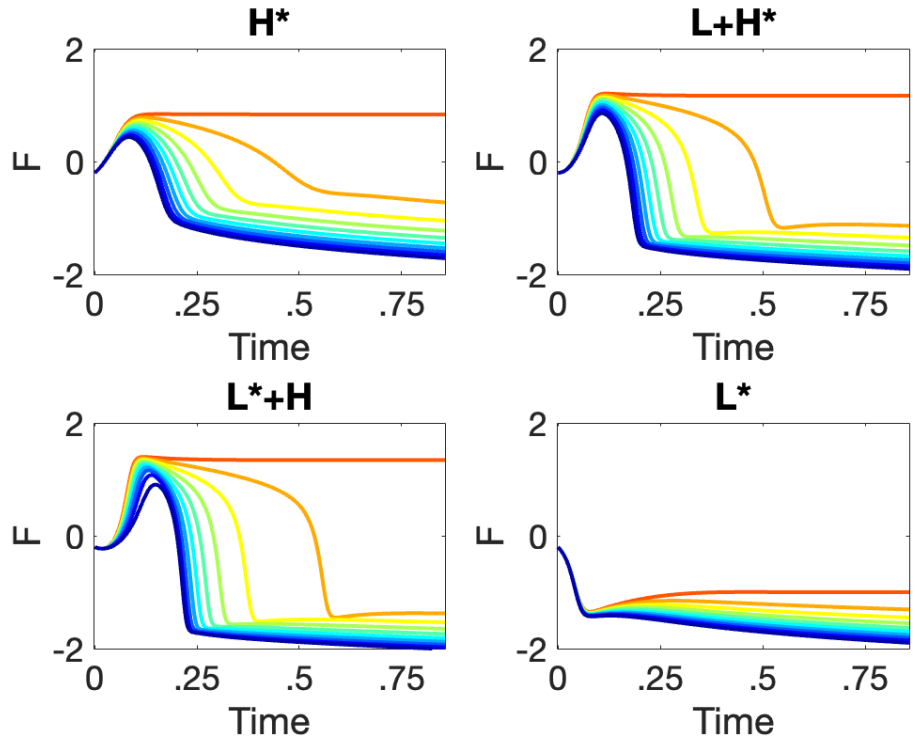
Conclusions

F0 velocity analysis suggests **grouping** of intonational features within the tune

Pitch Accents and Phrase Accents are tightly coordinated, suggesting a **single** dynamical unit that captures **both** events

Our ongoing work: Boundary tones; Prenuclear accents





We are grateful to the NSF (BCS-1944773), Chun Chan & the Prosody and Speech Dynamics Lab at Northwestern University