Jennifer Cole
Northwestern University

Unlocking prosody: Discovering structured variation and rich context effects

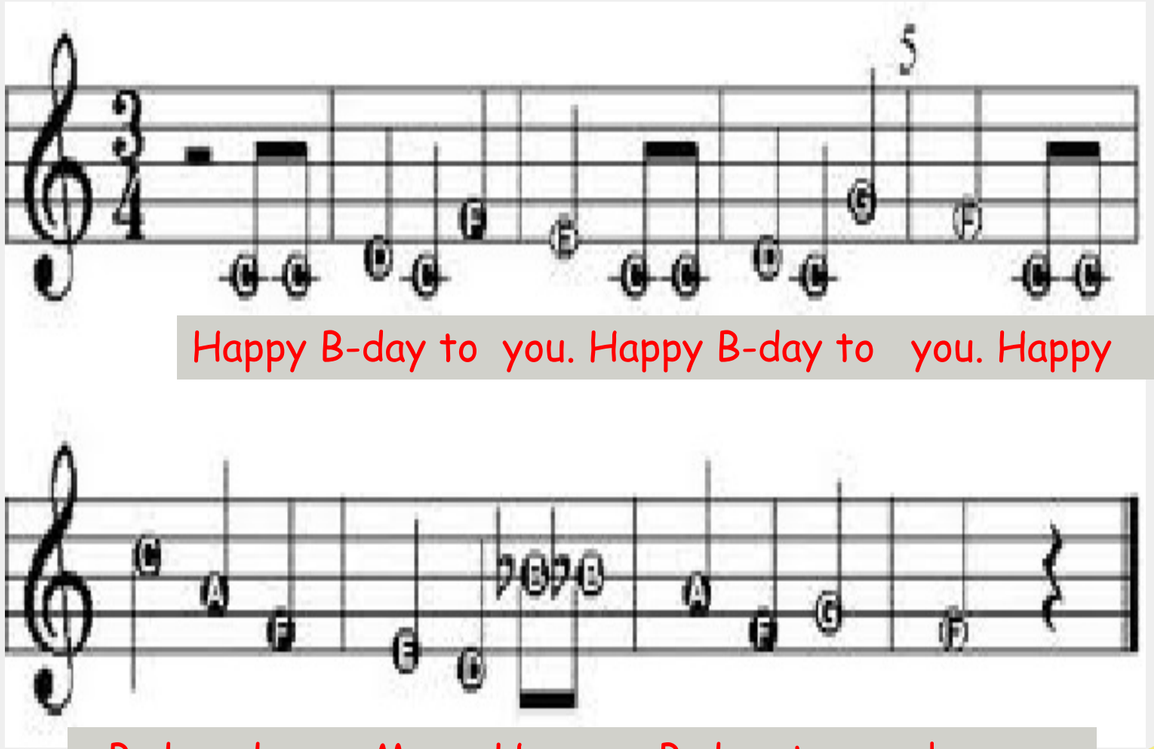
Northwestern

What is prosody?

the ‘music of speech’:

patterning of pitch, timing, loudness, voice quality over spoken words and phrases

(Byrd & Krivokapić 2021; Cho & Keating 2009; Cole 2015; de Jong 1995; Fletcher 2010; Wennerstrom 2001; *many others*)

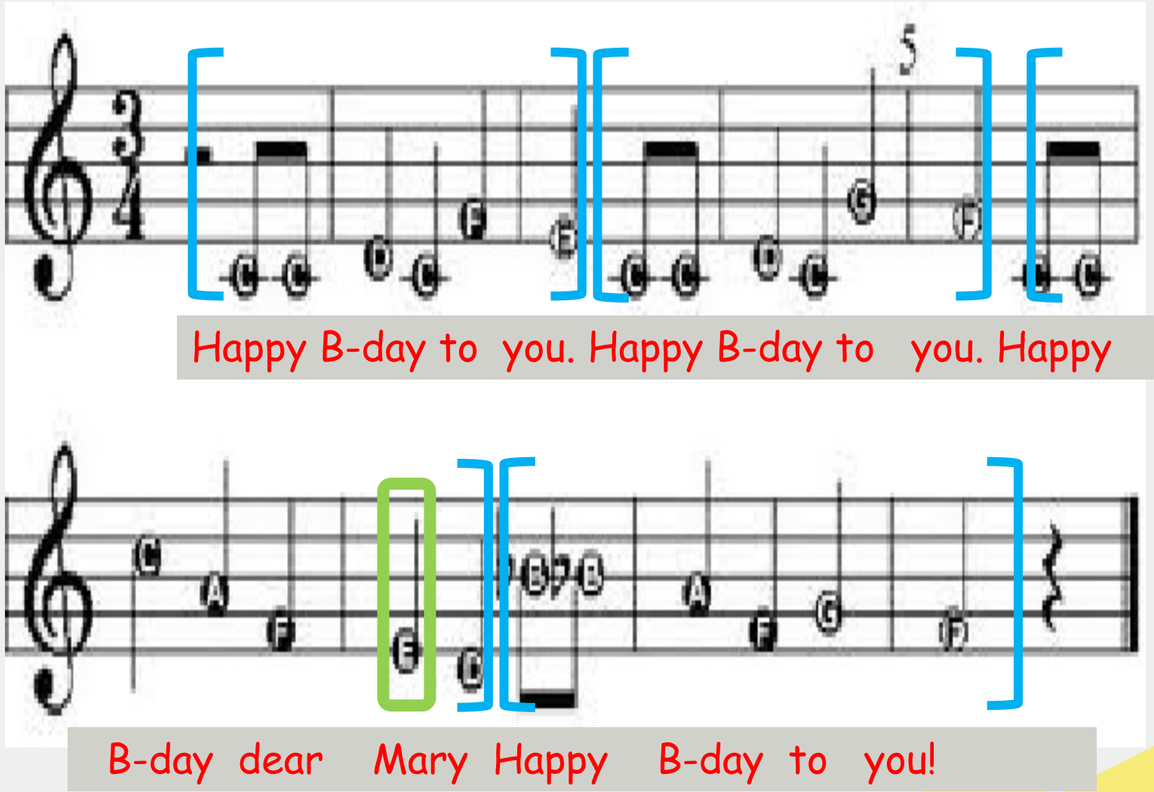


What is prosody?

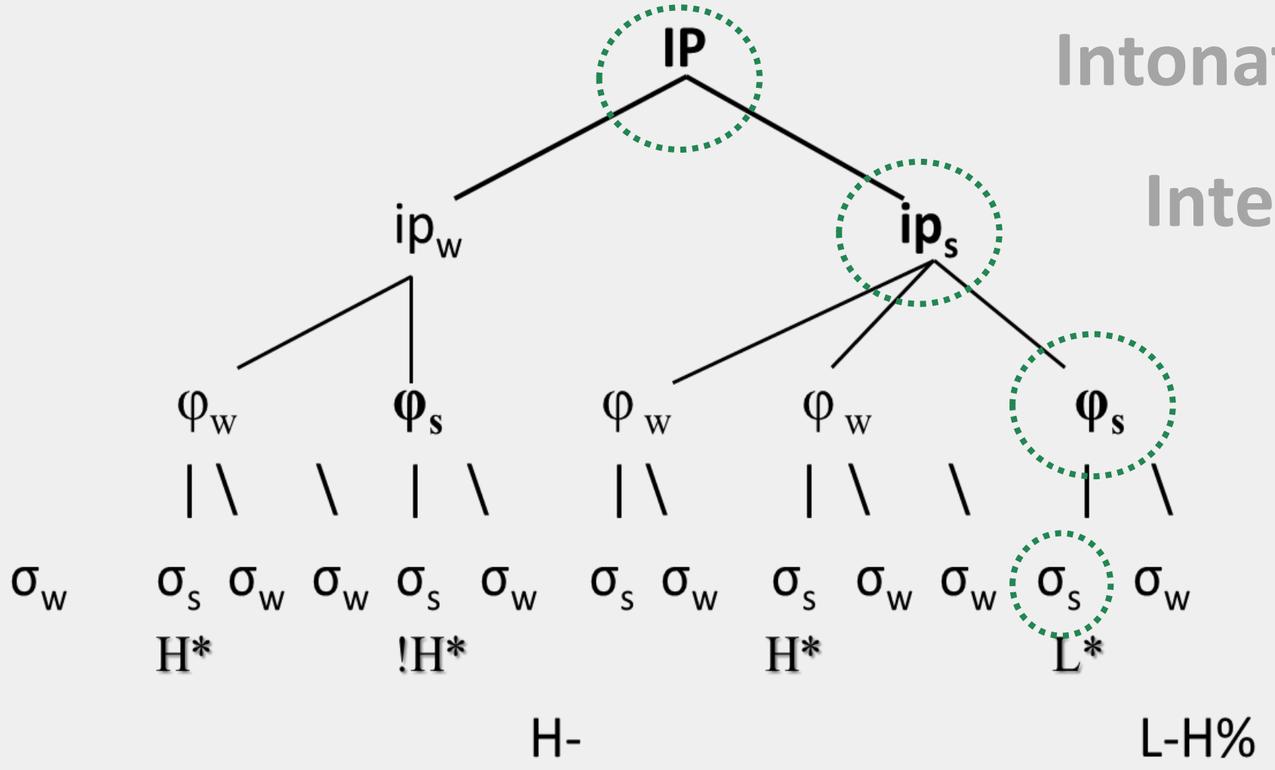
Underneath:

Structures that group linguistic units (syllables, words, phrases) into larger units that determine rhythm, **phrasing** and **prominence**.

(Ladd 2008; Gussenhoven 2004; Beckman 1996)



Hierarchical prosodic structure



Intonational phrase

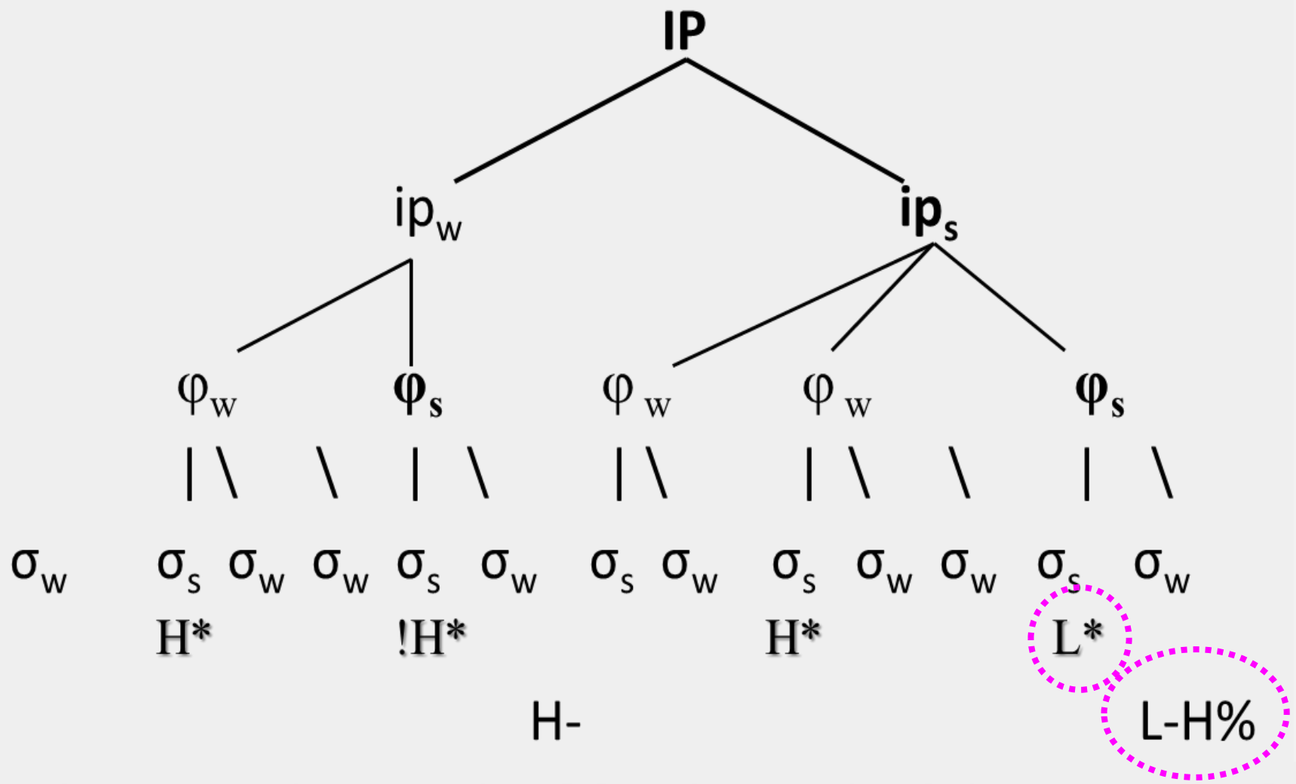
Intermediate phrase

Foot

Syllable

[You're gonna be standing] [at the peak of the mountain]

Hierarchical prosodic structure



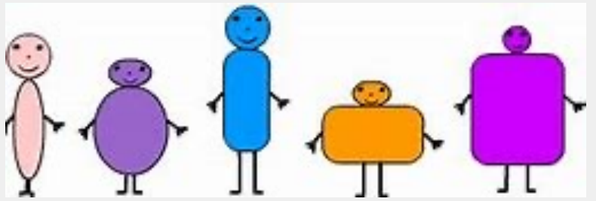
Pitch accent
Boundary tone

[You're gonna be standing] [at the peak of the mountain]

Preview

I. **Prosody plays a key role** in the identification and comprehension of words, phrases, and discourse units.

II. **Prosody is highly variable** in how it is produced in speech, in how listeners perceive it, and in how it is interpreted.



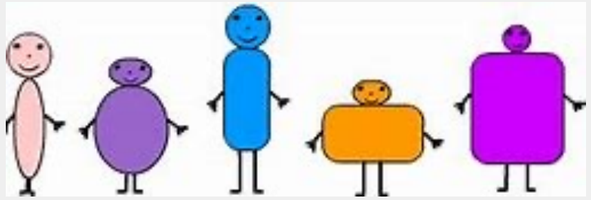
III. **Prosody is multi-faceted**, signaling the linguistic and situational context, speaker's communicative intentions and affect.

IV. **Our challenge**, a way **forward**, and two examples of what that looks like.

I. **Prosody plays a key role** in the identification and comprehension of words, phrases, and discourse units.

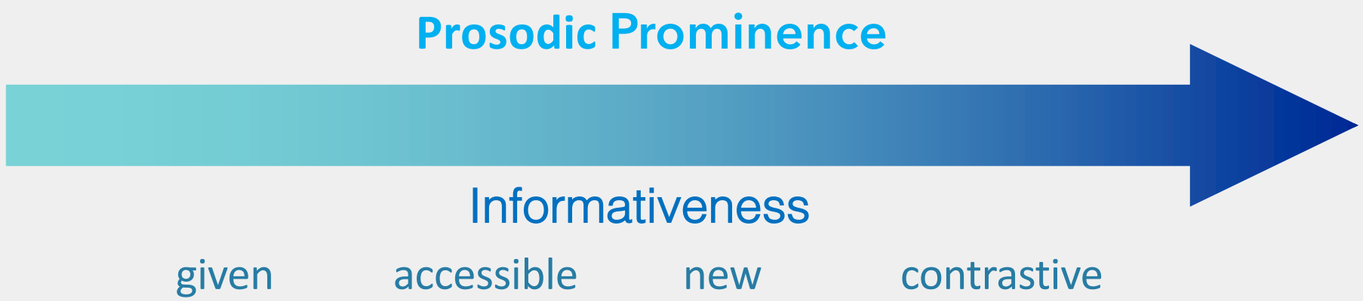
1. Rhythmic expectations affect downstream **word recognition**
2. Prosodic phrase boundaries cue **word segmentation, syntactic parsing, discourse structure (turn & topic ending)**
3. Prominence of accented words influences **inferences about their referents**
4. Phrase-final intonation conveys **speaker's communicative intentions & beliefs**

II. **Prosody is highly variable** in how it is produced in speech, in how listeners perceive it, and in how it is interpreted.



Prosodic marking of information structure

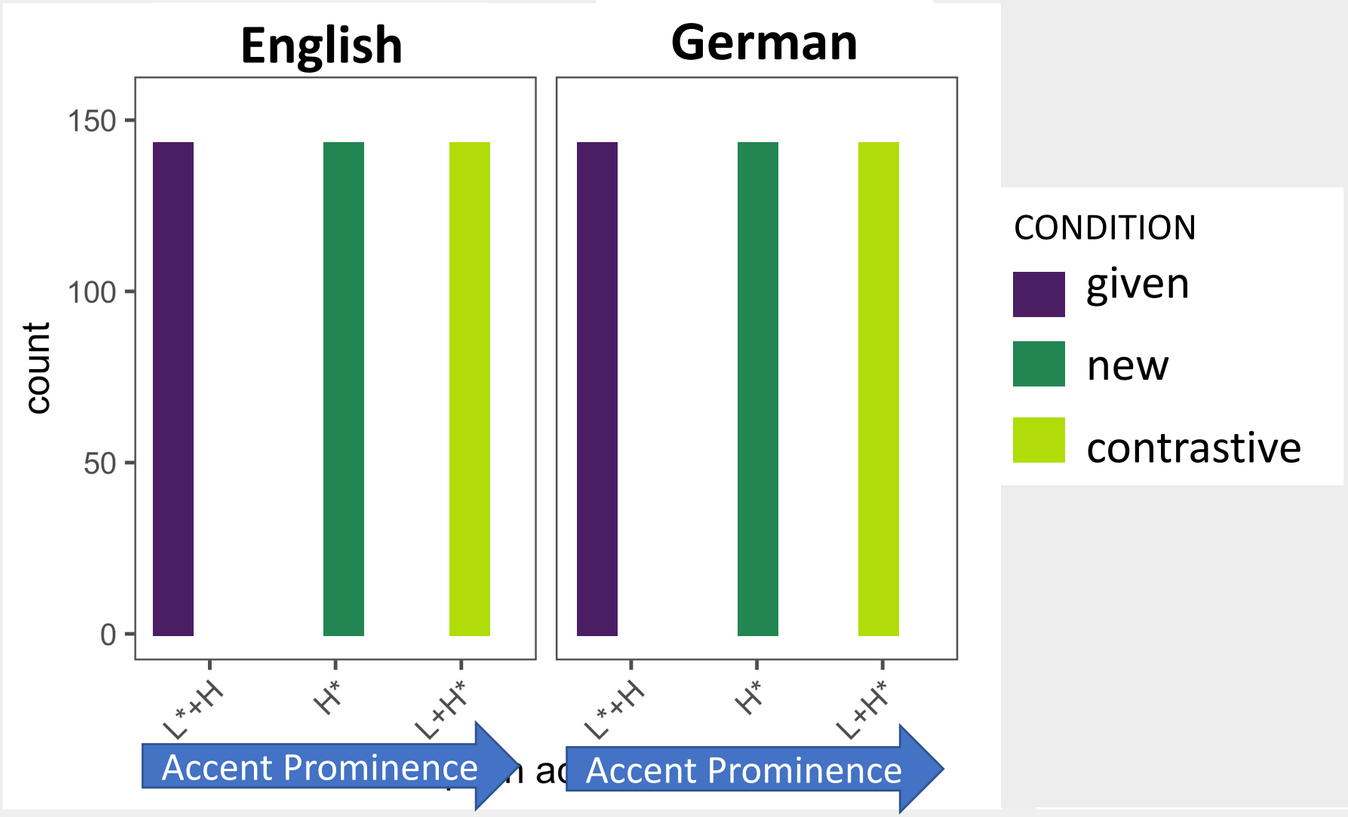
(Chodroff, Cole & Baumann 2021)



Her **nana** loved the **marmalade**.

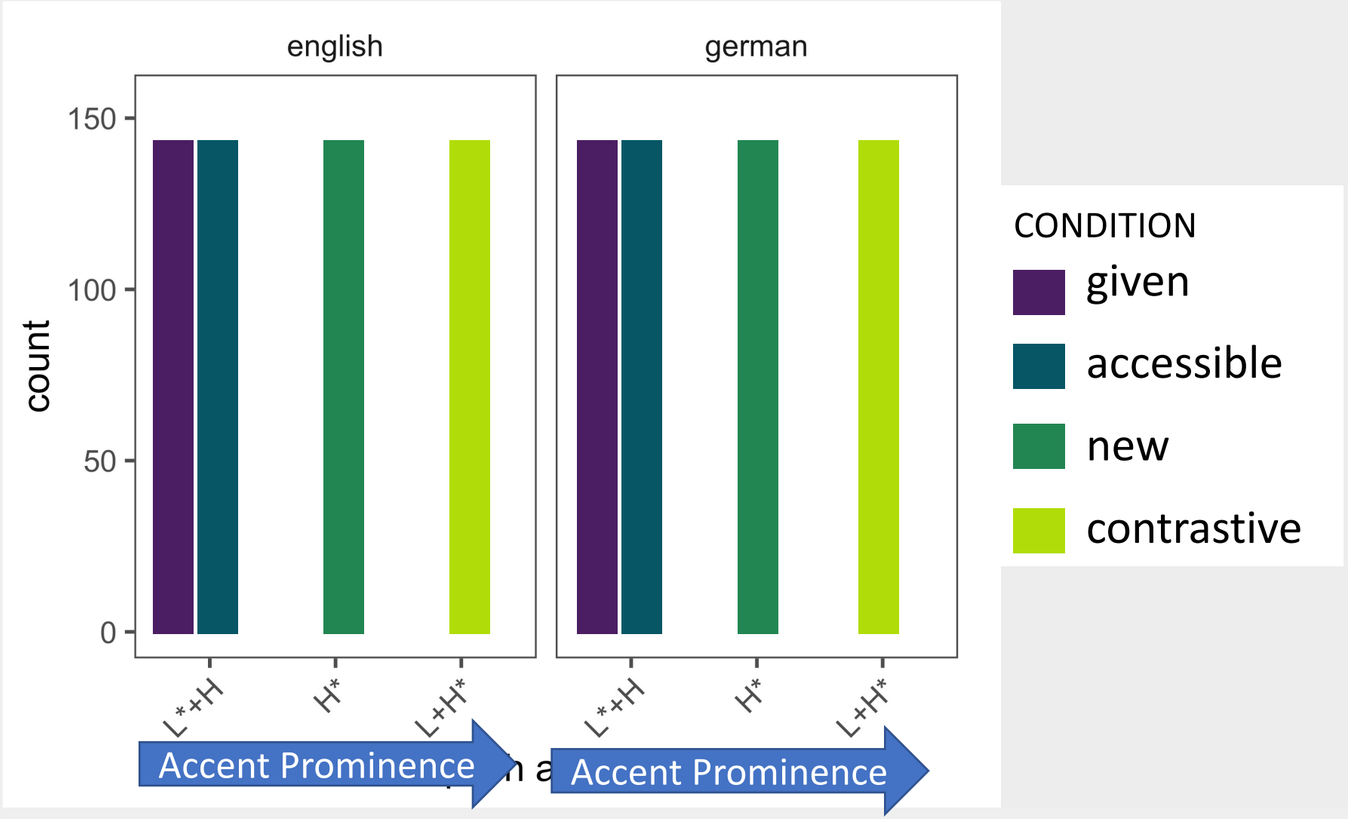
Expected patterns

(Chodroff, Cole & Baumann 2021)



Expected patterns

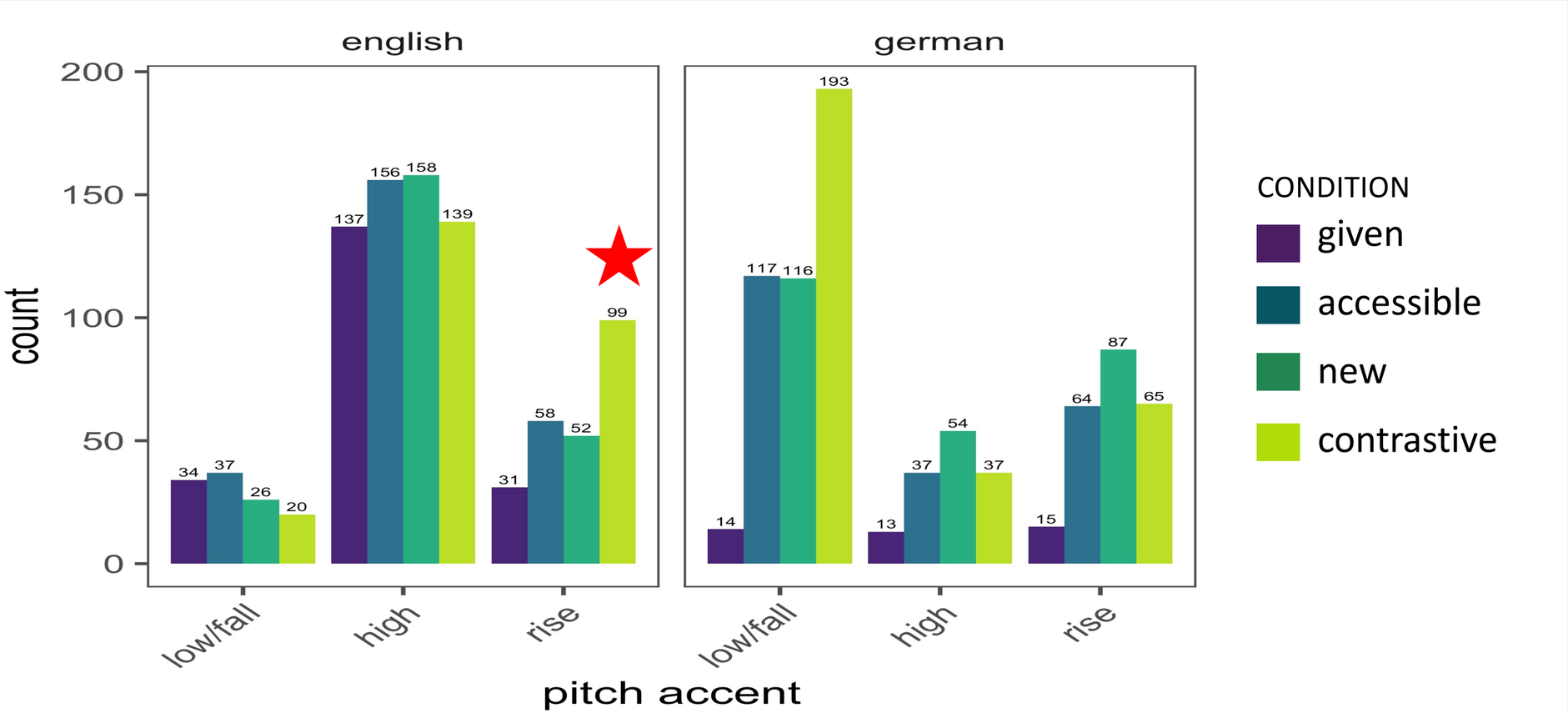
(Chodroff, Cole & Baumann 2021)



Observed patterns (32 speakers, each language)

Chodroff, Cole & Baumann 2021

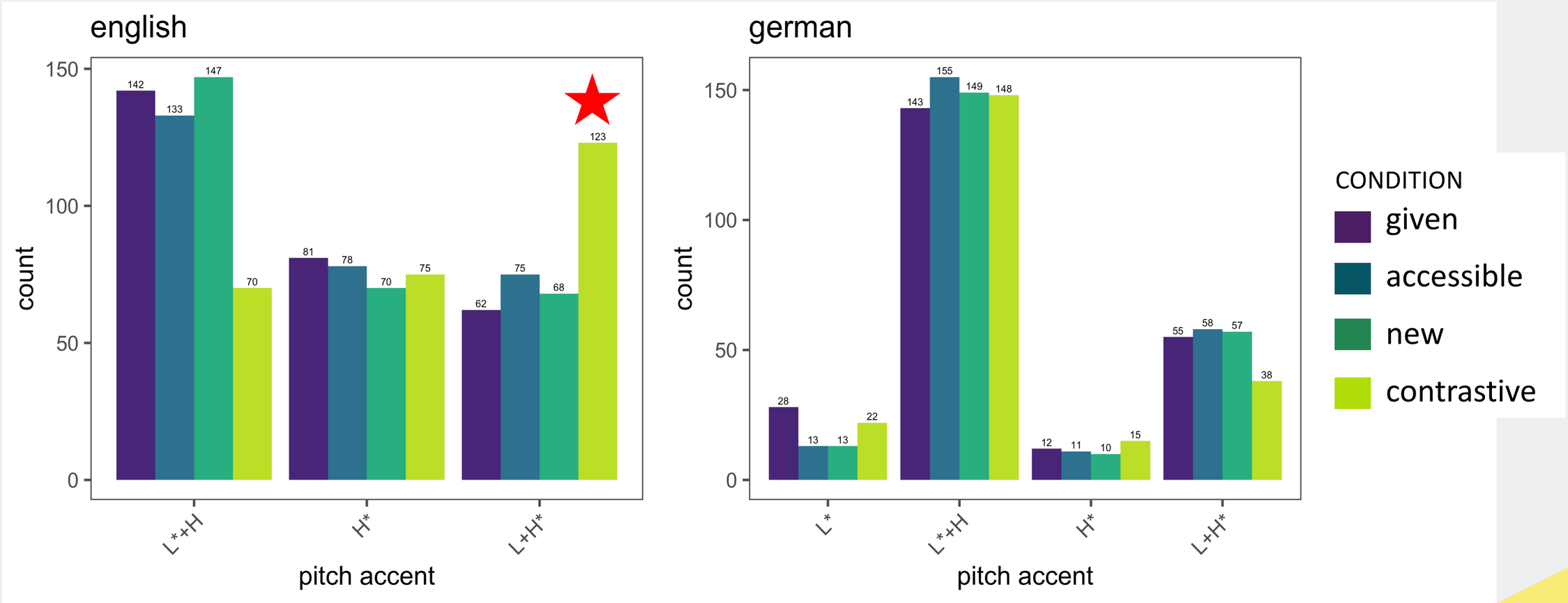
Nuclear pitch accents: many-many associations



Observed patterns

(Chodroff, Cole & Baumann 2021)

Preuclear pitch accents: many-many associations



In brief,

American English & German (Chodroff, Cole & Baumann 2021)

Prosody weakly and probabilistically encodes information structure.

Information structure licenses prosodic enhancement & reduction but does not require it.

Sources of variation?

- **Competition:** linguistic vs. psycho-social factors

- Speaker’s **choice** in phonetic implementation

There is also variation in how prosody is perceived and comprehended.

Prosodic congruence in dialogues

Roettger, Mahrt & Cole 2019

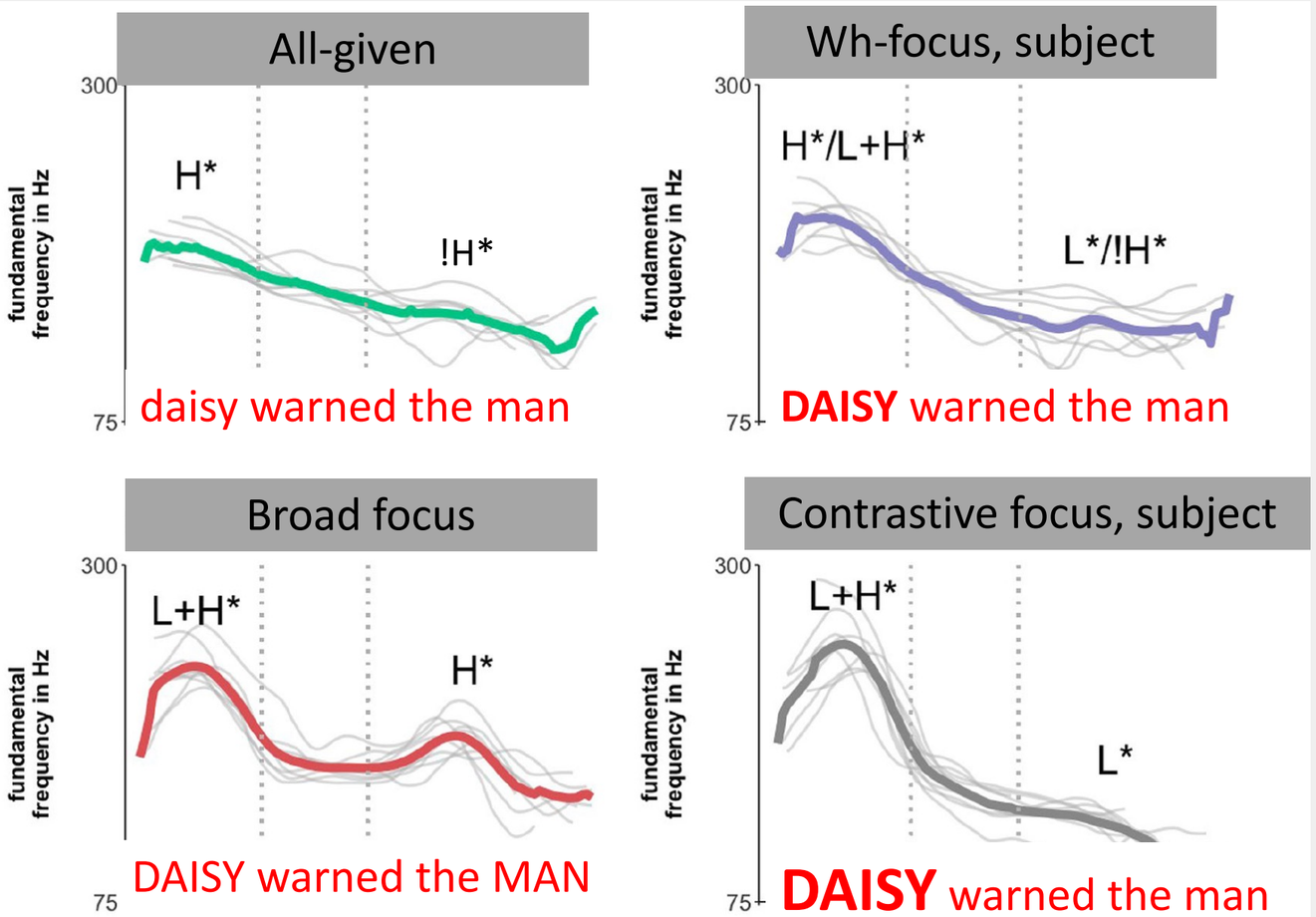
Do pitch accents convey distinctions in givenness and focus?

Prior studies report mixed results.

(Gussenhoven 1983; Welby 2003; Breen et al. 2010)

Prosodic congruence in dialogues

Roettger, Mahrt & Cole 2019



Do pitch accents convey distinctions in givenness and focus?

Prosodic congruence in dialogues

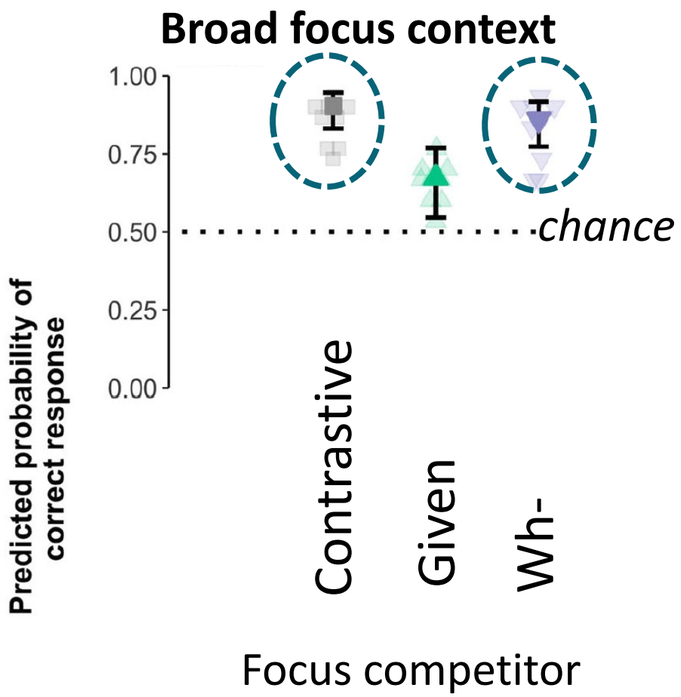
Roettger, Mahrt & Cole 2019

Q: Do you know what happened?
A1: Yes, [Daisy warned the man]$_F$

Broad focus context
Broad focus response

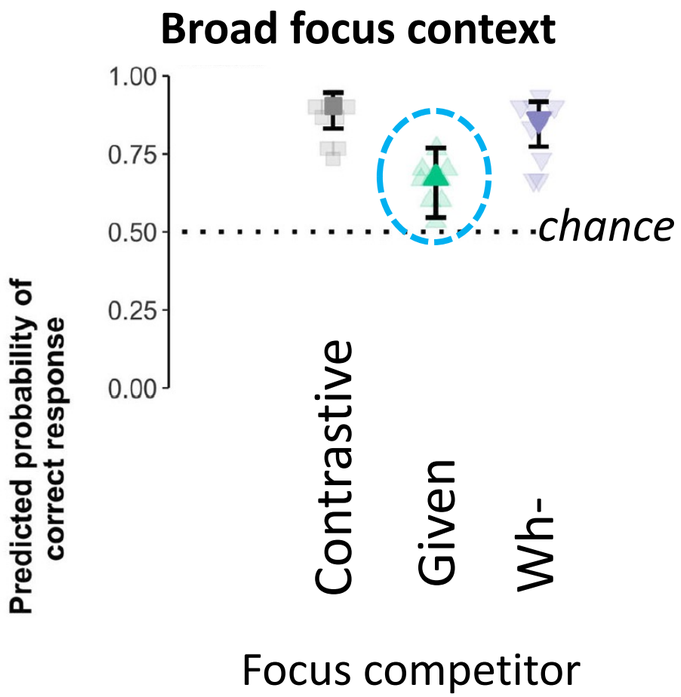
Q: Do you know what happened?
A2: Yes, [**DAISY**]$_F$ warned the man.

Broad focus context
Contrastive Subject response



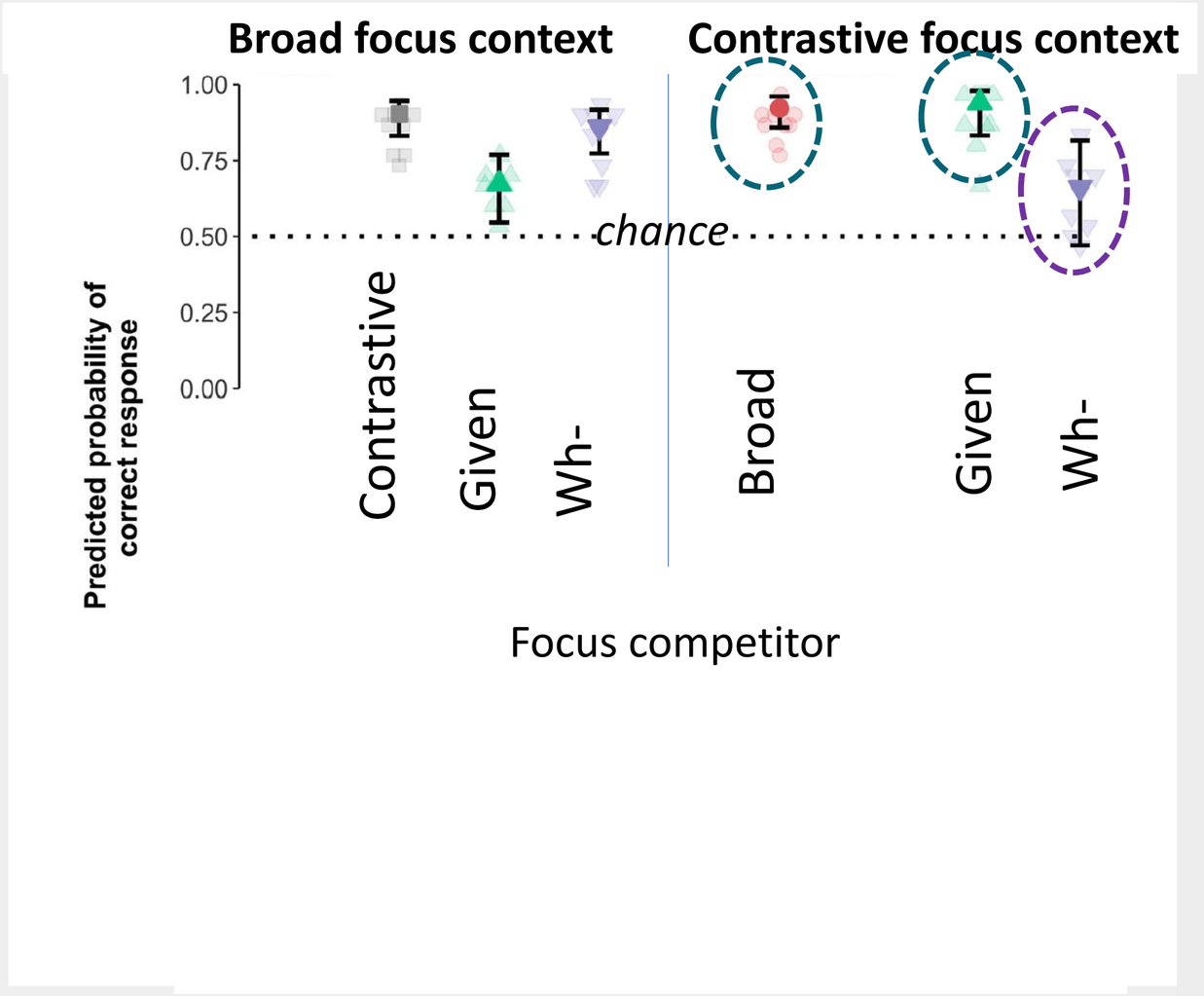
RESULTS: In a broad focus context,
listeners correctly choose broad focus intonation,
rejecting contrastive or Wh-focus intonation.

Roettger, Mahrt & Cole 2019



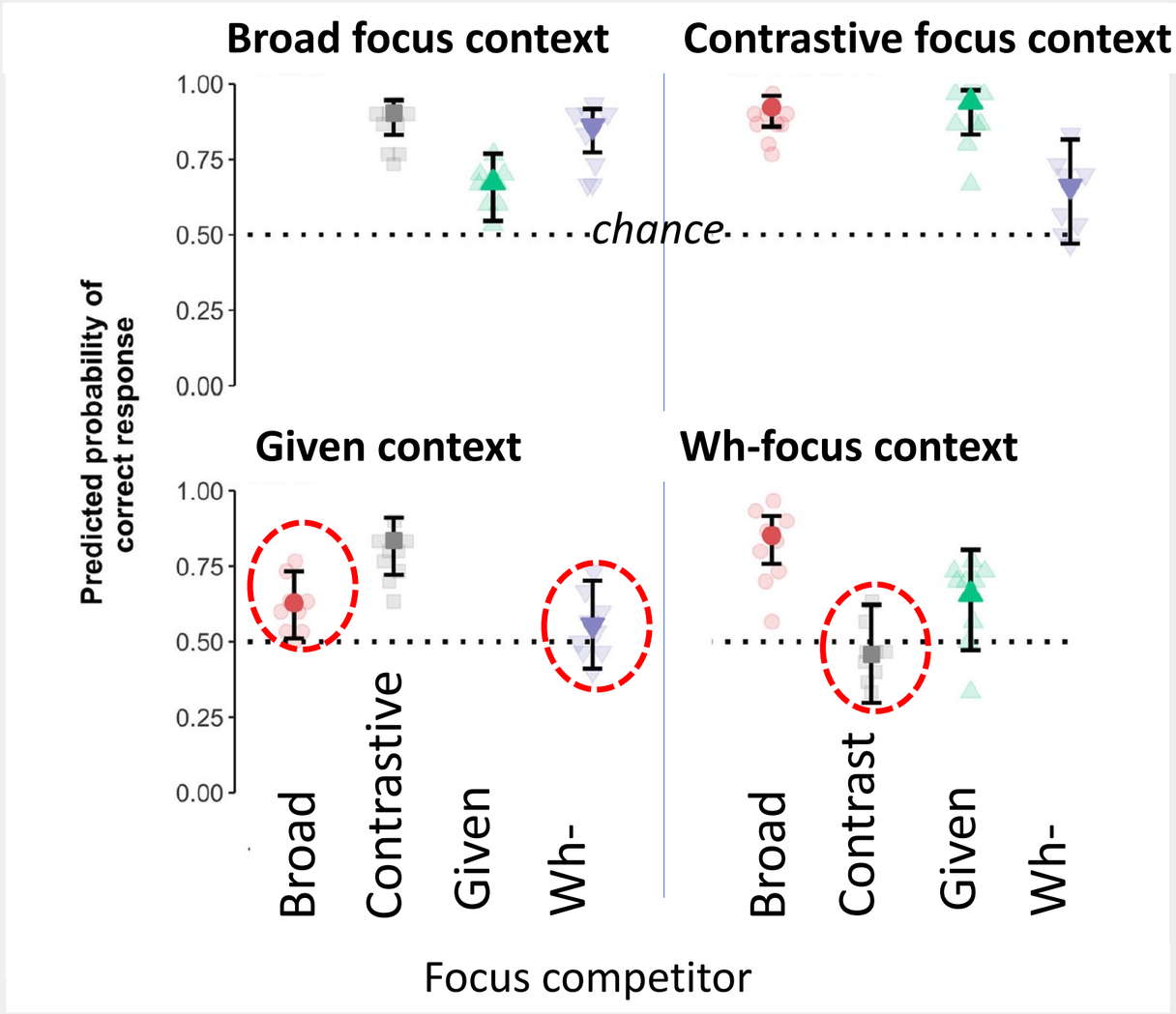
Listeners were less accurate rejecting the **given** intonation pattern.

Roettger, Mahrt & Cole 2019



RESULTS: In a <u>contrastive</u> focus context, listeners correctly choose contrastive intonation when the competitor was broad or Wh-focus, but were less accurate rejecting the **Wh- focus** intonation pattern.

Roettger, Mahrt & Cole 2019



RESULTS: In the given and Wh-focus contexts, listeners were less accurate in choosing the congruent intonation.

Roettger, Mahrt & Cole 2019

In brief,

American English Roettger, Mahrt & Cole 2019

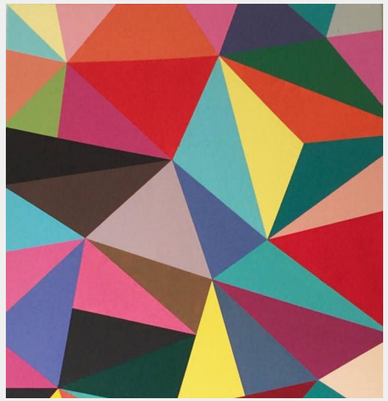
Listeners perceive some information structure distinctions better than others, based on intonational cues

Given ≈ Broad focus Wh-focus ≈ Contrastive focus

Why is prosody so challenging?

III. **Prosody is multi-faceted**, signaling the linguistic and situational context, speaker's communicative intentions and affect… making it difficult to isolate the linguistic signal

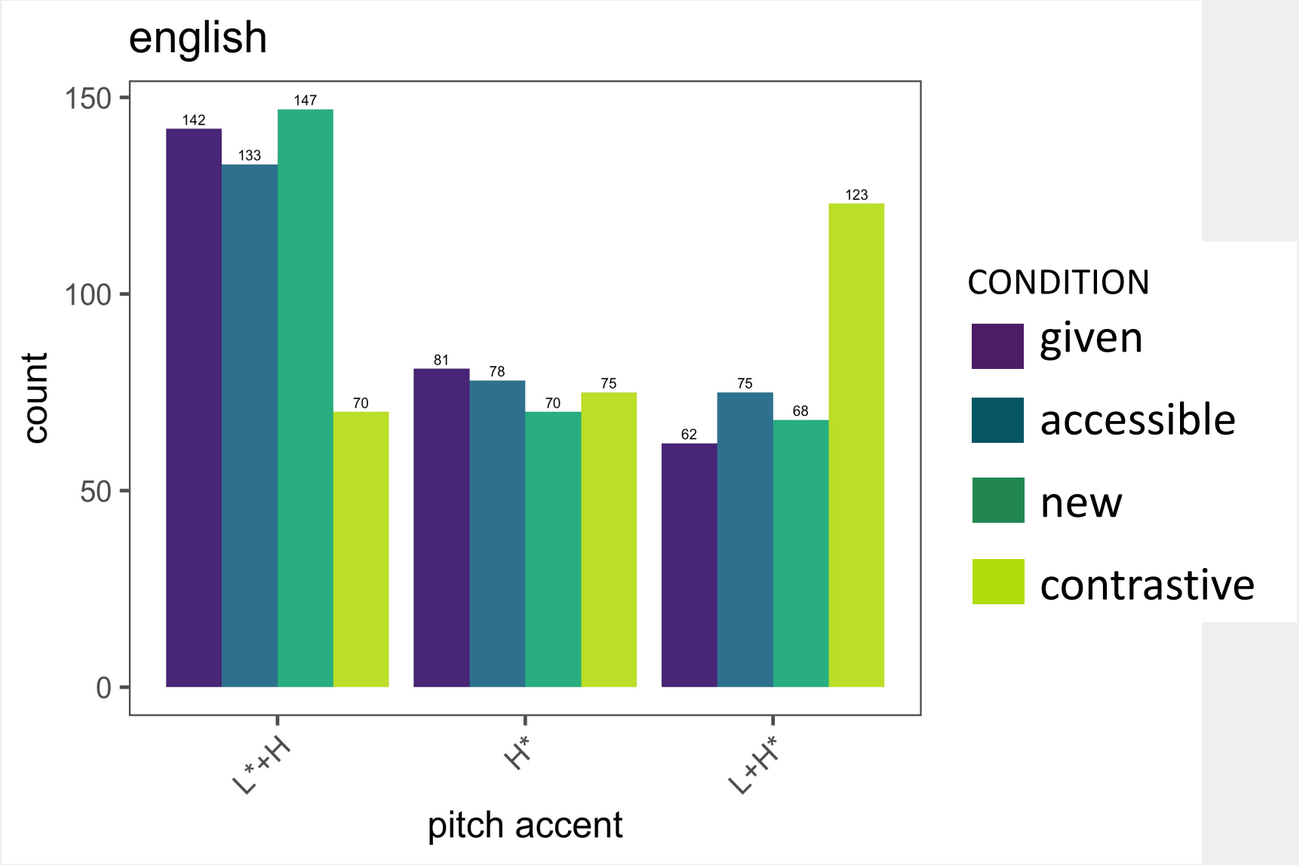
Why? How?



Prosody: the ultimate multi-tasker

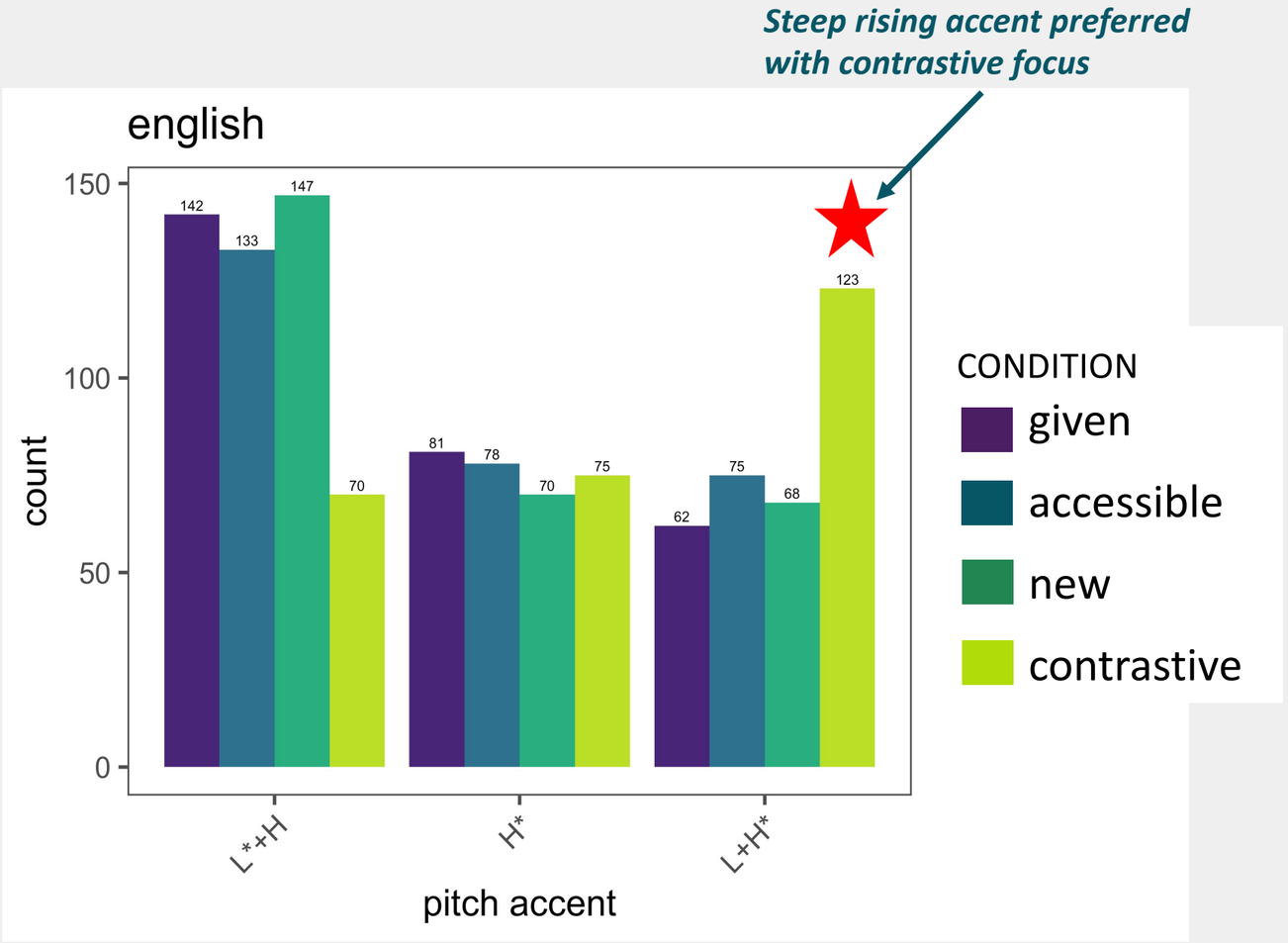
1. Focus
2. Discourse-givenness
3. Situational context
4. Speaker’s communicative intentions
5. Speech style
6. Speaker’s affect

1 & 2. Focus & Discourse givenness (Chodroff, Cole & Baumann 2021)



Subtle effects of focus & givenness

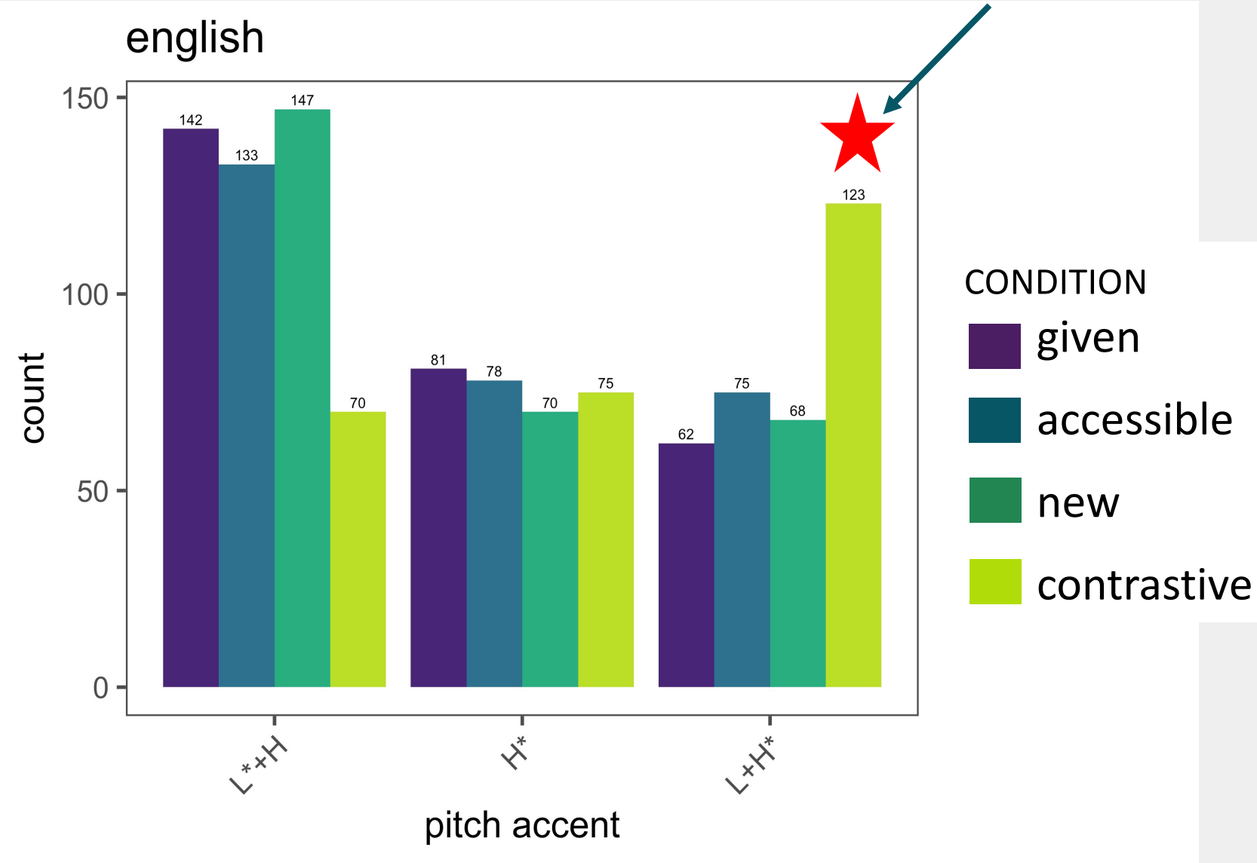
(Chodroff, Cole & Baumann 2021)



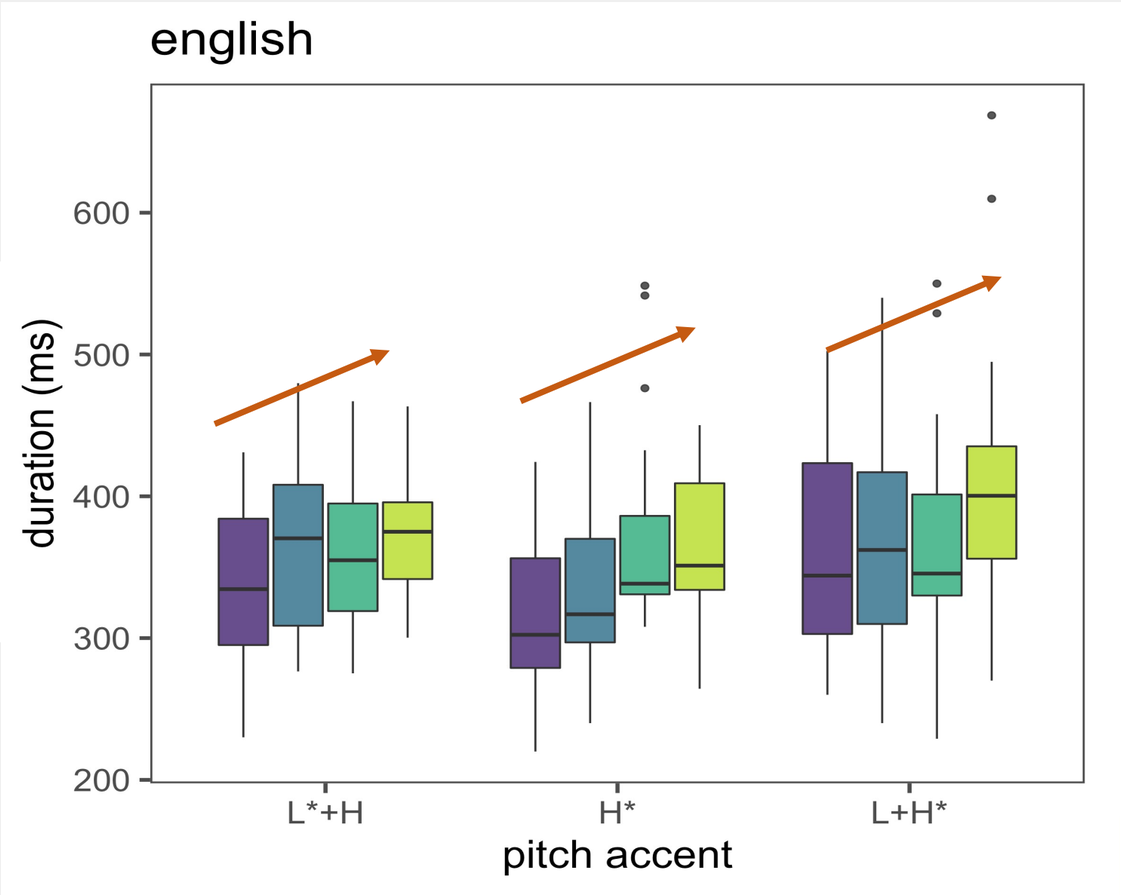
Subtle effects of focus & givenness

(Chodroff, Cole & Baumann 2021)

Steep rising accent preferred with contrastive focus



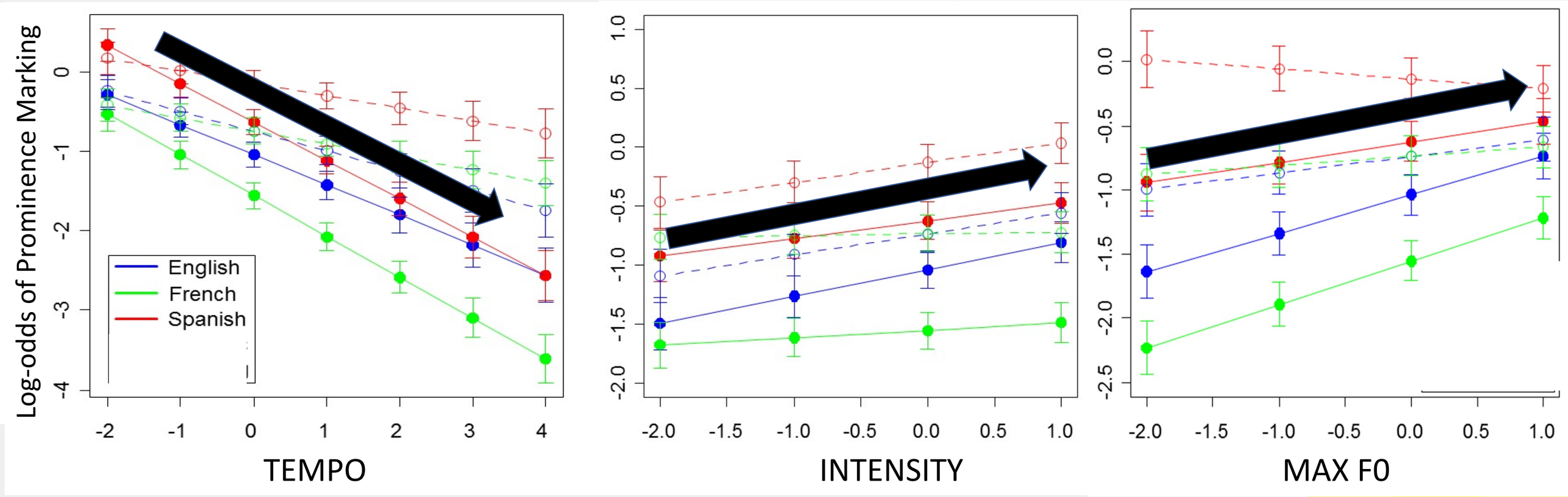
Word duration increases with informational prominence



Listeners pay attention

Cole, Hualde, Smith, et al. 2019

Am. English, French, Spanish: Words that have slower tempo, higher intensity, and higher F0 peak are more likely to be rated as prominent by native listeners.



3. Situational context

Listeners perceive acoustic cues to prosody in relation to talker-specific patterns of variation (Xie, Buxó-Lugo, Kurumada 2021)

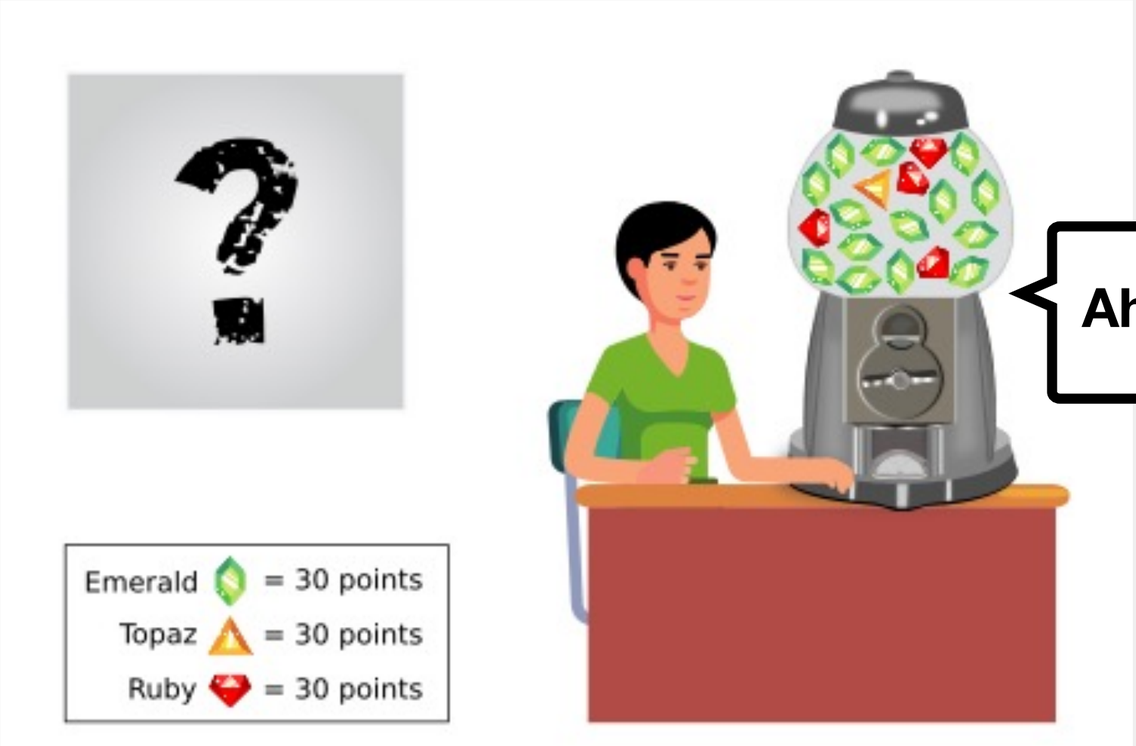
Listeners infer prosodic meaning in relation to the immediate situational context (Roettger & Rimland; Roettger & Cole 2020, *in prep.)*

Situational context

Roettger & Cole, 2020, *in prep.*

Situational context

Roettger & Cole, 2020, *in prep.*

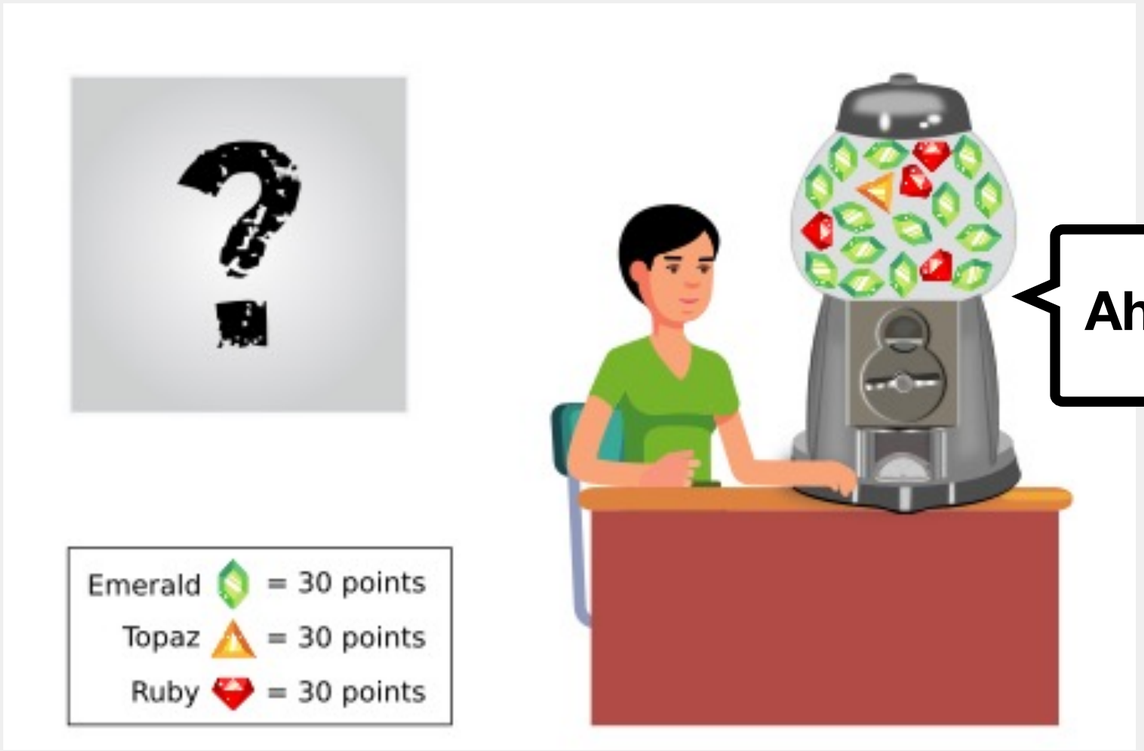


"What did Jones receive from the gumball machine?"

- Emerald
- Topaz
- Ruby

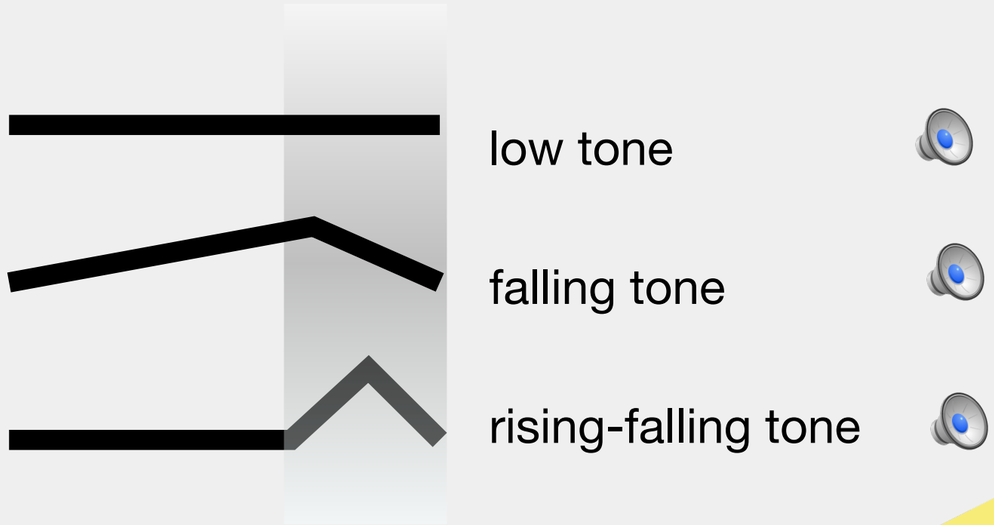
Situational context

Roettger & Cole, 2020, *in prep.*

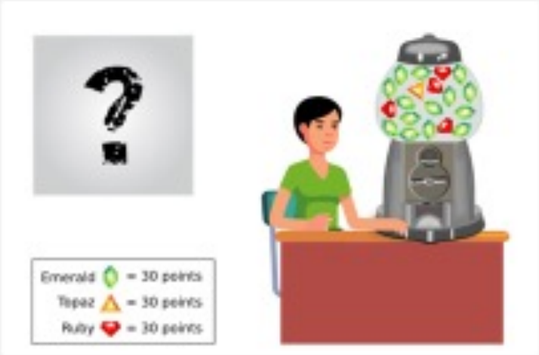


"What did Jones receive from the gumball machine?"

- Emerald
- Topaz
- Ruby



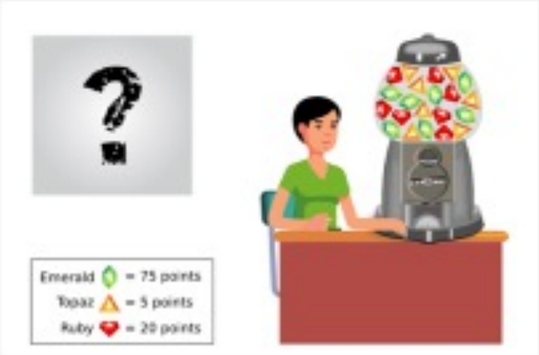
Group 1: predictability



Balanced points

Skewed gem frequency

Group 2: importance



Skewed points

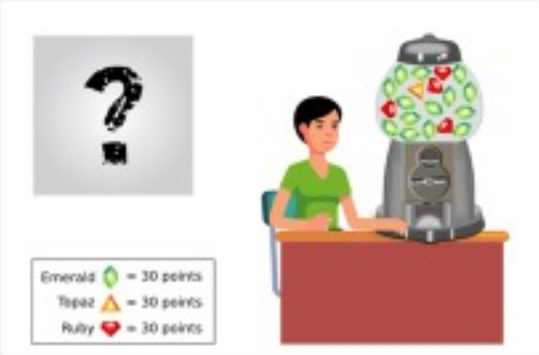
Balanced gem frequency

Group 3: control

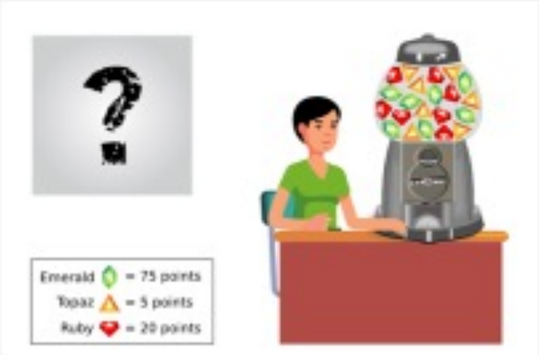
Balanced points

Balanced gem frequency

Group 1: predictability



Group 2: importance



Group 3: control

IF intonational prominence is related to information value

Prediction: low tone → the most predictable OR least important gem
rising-falling tone → the least predictable OR most important gem

Robustly confirmed!

4. Speaker’s communicative intention

Inferences probabilistically related to the ‘nuclear tune’

Declaratives:	rising	→ questions	She’s from Canada?
	falling	→ assertions	She’s from Canada.

(Jeong 2018)

4. Speaker’s communicative intention

Inferences probabilistically related to the ‘nuclear tune’

Declaratives: rising → questions

falling → assertions

She’s from Canada?

She’s from Canada.

(Jeong 2018)

low-flat → withholding speaker commitment

Q: Can anyone volunteer?

A: I will

(Sostarics 2021)

4. Speaker’s communicative intention

Inferences probabilistically related to the ‘nuclear tune’

Declarative: rising → questions (Jeong 2018)

falling → assertions

low-flat → withholding speaker commitment (Sostarics 2021)

Imperatives: rising → suggestions (Sandberg 2021)

falling → commands

Bake the bread for half an hour.

5. Speech style

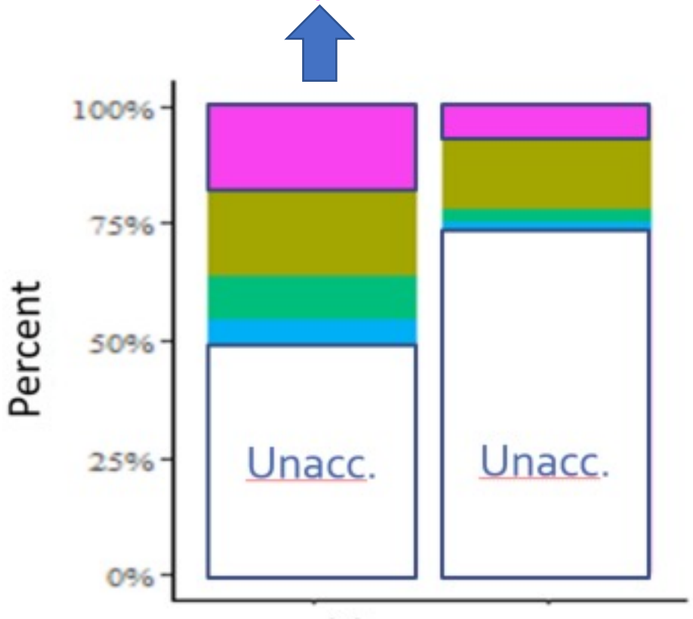
Highly engaged TED talk speakers are emphatic, with frequent use of high-prominence pitch accents, and acoustic enhancement of accented words.

(Im, Cole & Baumann 2018; in prep.)

TED Talks vs. conversational speech

(Im, Cole & Baumann 2018; in prep.)

Prominent accents are more frequent



Listeners calibrate!

Listeners recognize the influence of **speech style** and **speaker engagement** prosody and perceive prominence differently depending on these contextual factors.

6. Speaker's affect

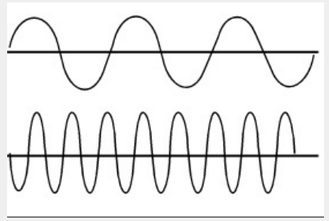
In real life settings, prosody reflects the speaker's emotional state, mediated via neural pathways connecting the auditory system and vocal tract with the parasympathetic nervous system. (Porges, 2007)

Similar effects can sometimes be captured in the lab through enacted emotion or affect.

Prosody in mother-child interactions

-an observational study, *62 dyads*

(Cole, Berry, McElwain et al. 2015)



Mom’s pitch

positive affect

sensitivity to child

Prosody in enacted speech (a laboratory experiment):

Cole, Chodroff, Baumann *in prep.*

Prominent pitch accents
Acoustic enhancement

Recap

In its linguistic function, **prosody is highly variable** in how it is produced in speech, in how listeners perceive it, and in how it is interpreted.

Maybe because…

Prosody is multi-faceted, signaling the linguistic and situational context, speaker's communicative intentions and affect.

IV. **Our challenge**, a way **forward**, and two examples of what that looks like.

Our challenge: To build a theory of the mapping between prosodic form and its linguistic functions that explains when, where and why variability arises, and how listeners cope with variable input in comprehending speech.

How? Embrace varation, expand the empirical horizon

1. CONTEXT: Examine prosody in relation to the linguistic context of the utterance, the social context of the communication, and the psychological context of the speaker-hearer.
2. STRUCTURED VARIATION: Analyze the structure of prosodic variation within and across individual speaker-hearers

using Big Data methods

Examine prosody in rich CONTEXTS

What features of prosody matter for communication?

What do listeners track?

We look into these questions with studies of **prosodic entrainment** in unscripted, interactive speech

(game-playing)

(Cole, Reichel, Roettger *in prep;* Patel, Cole et al. 2021)

Prosodic entrainment

Record spontaneous speech from two people playing a picture-matching game, cooperatively and competitively.

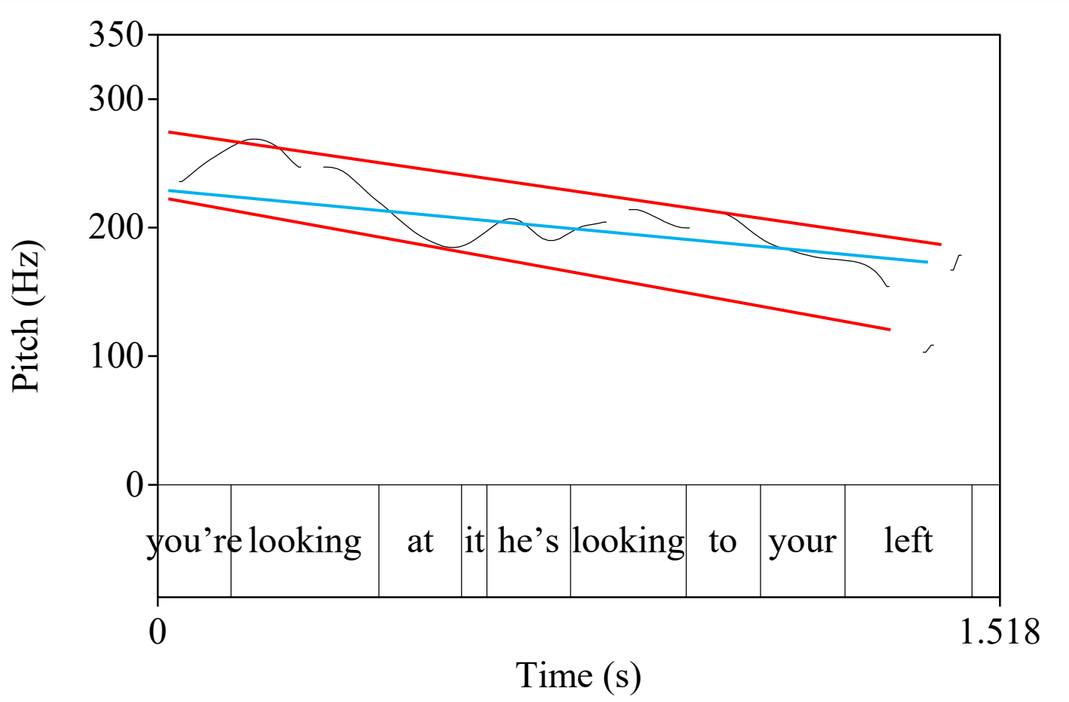
Model F0 trajectories over phrases and words for both conversation partners.

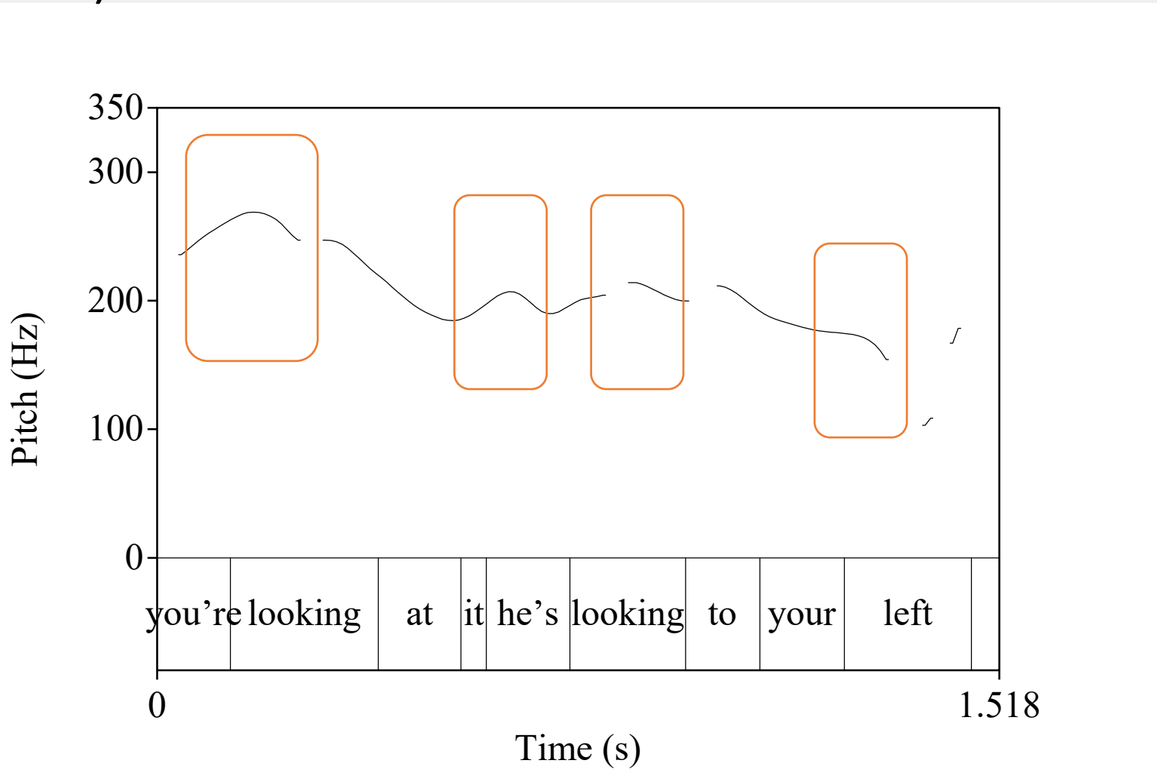
Compare F0 trajectories used in utterances with matched dialog act, for partners and for randomly paired (non-interacting) speakers

Tested with young adult speakers**--neurotypical** (*N*=22) and with **clinically diagnosed autism** (N=23) and age-matched neurotypical controls.

(Cole, Reichel, Roettger *in prep;* Patel, Cole et al. 2021)

F0 modeled in global (phrase) and local (syllable) domains, using CoPaSul software for parameterizing F0 (Reichel 2017)





Global measures:
F0 trends (baseline, midline and topline)
F0 range

Local measures:
F0 mean, max, s.d.
Deviation from F0 trend lines & range
F0 contour shape (3rd order polynomial fit)

Briefly, we found:

Entrainment is **not automatic**—varies by the function of the utterance in the dialogue.

Reduced entrainment for individuals with ASD.

F0 features that are entrained must be **perceptually salient**, and have a **cognitive representation**:

Global:	*F0 register*	-- speaker affect
	baseline slope	-- speech act
Local:	*F0 contour deviation*	-- prominence relations, informativity
	F0 max	-- informativity, speaker commitment

(Cole, Reichel, Roettger *in prep;* Patel, Cole et al. 2021)

Look for STRUCTURE in patterns of variation

Which acoustic cues vary,
across speakers and for individuals?

What can variation patterns tell us
about prosodic categories?

We address these questions by examining variation patterns in **imitative sentence intonation**.

Imitative productions of ‘nuclear tunes’ in American English

The claim:

there are 24 distinct tunes, composed from three tonal features:

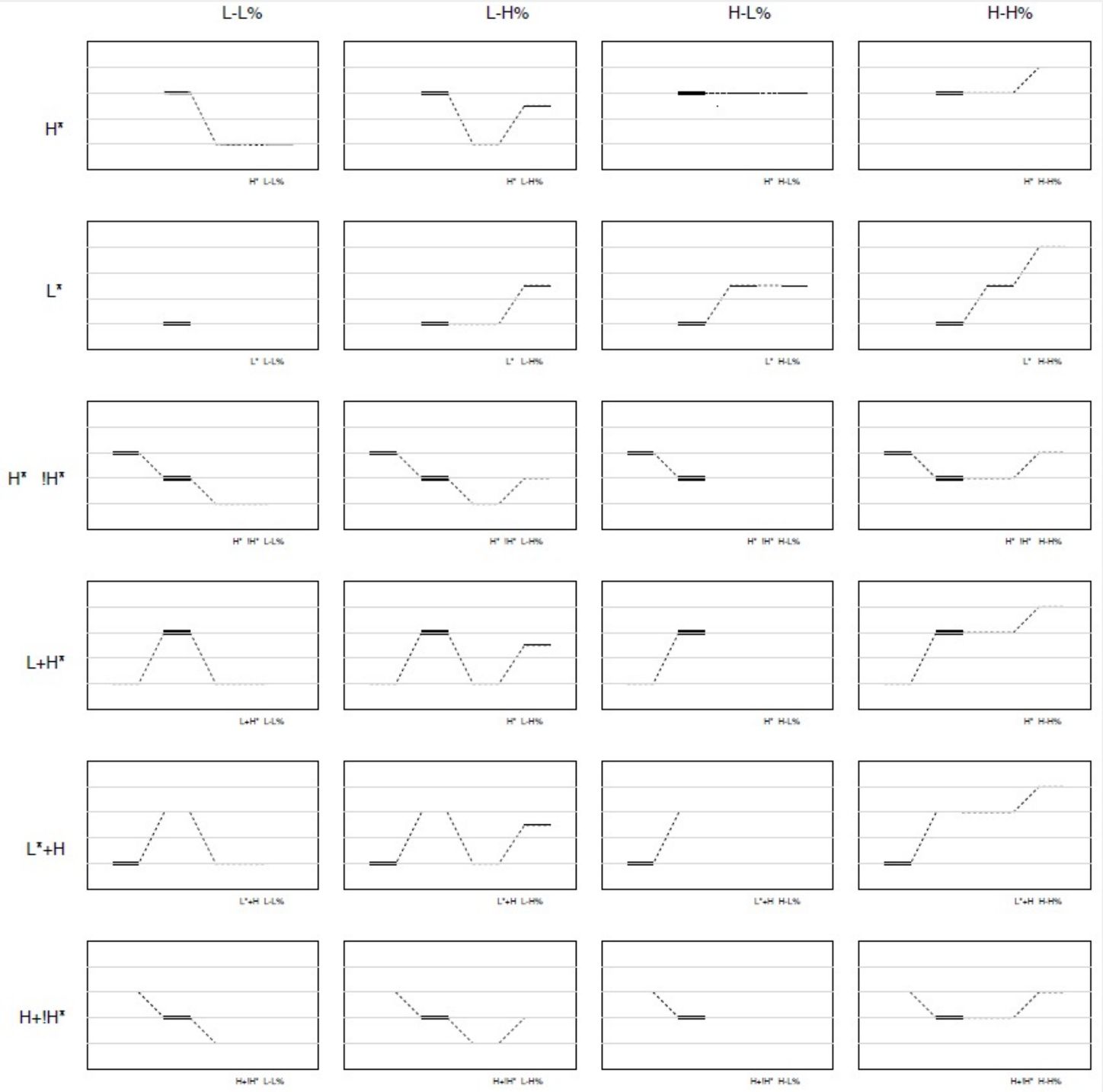
nuclear pitch accent (6) + phrase accent (2) + boundary tone (2)

H*
L*
!H*
L+H*
L*+H
H+!H*

H-
L-

H%
L%

(Chodroff & Cole 2019; Cole, Steffman, Tilsen, Shattuck-Hufnagel, 2021)

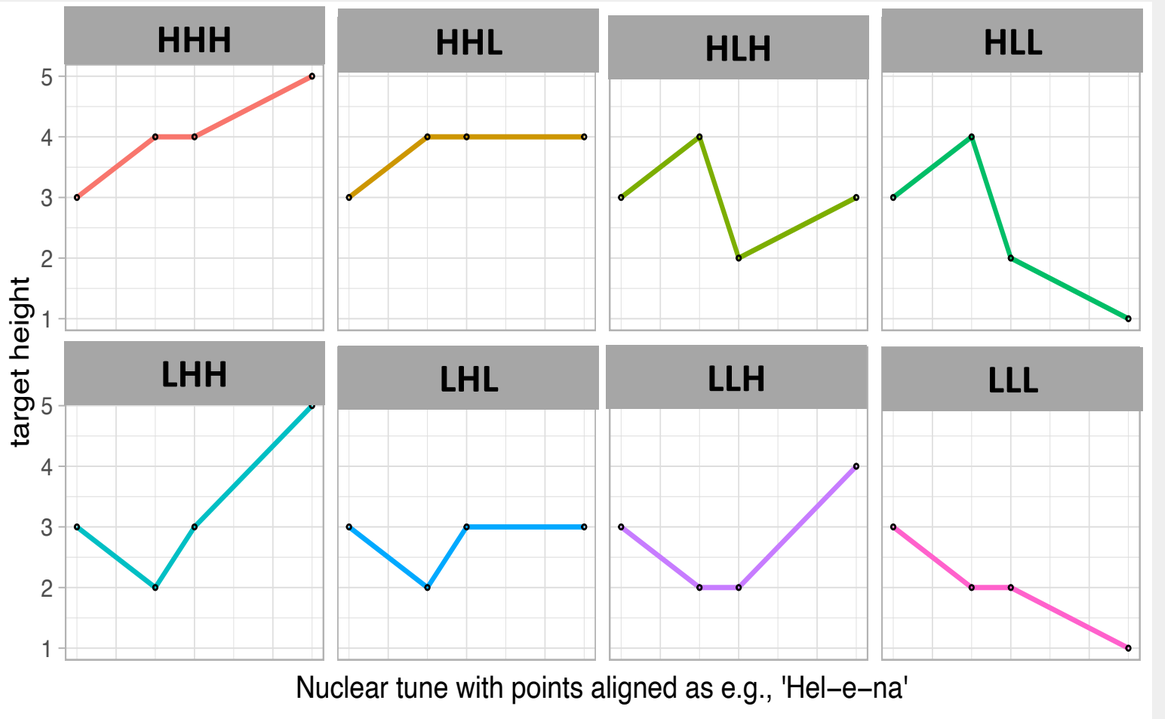


Schematic diagram of F0 trajectories of nuclear tunes in American English

Veilleux, Shattuck-Hufnagel & Brugos (2006), MIT Opencourseware “Transcribing Prosodic Structure of Spoken Utterances with ToBI”

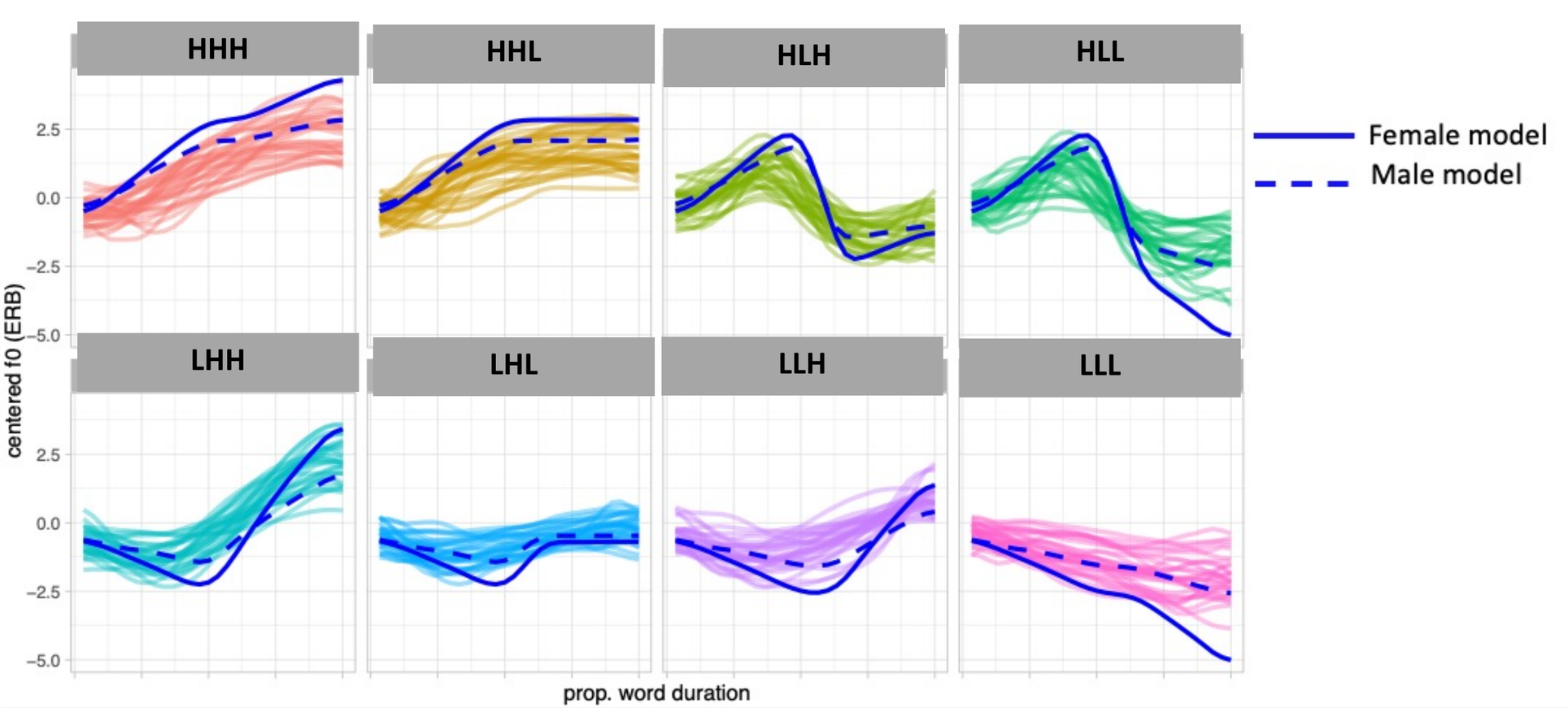
Tested: 30 participants imitating **8 basic tunes,** transposing tunes from auditory model utterances to new sentences.

*"She quoted **Helena**"*



(Chodroff & Cole 2019; Cole, Steffman, Tilsen, Shattuck-Hufnagel, 2021)

Auditory model tunes (1 M & 1F speaker) and imitative productions (30 speakers)

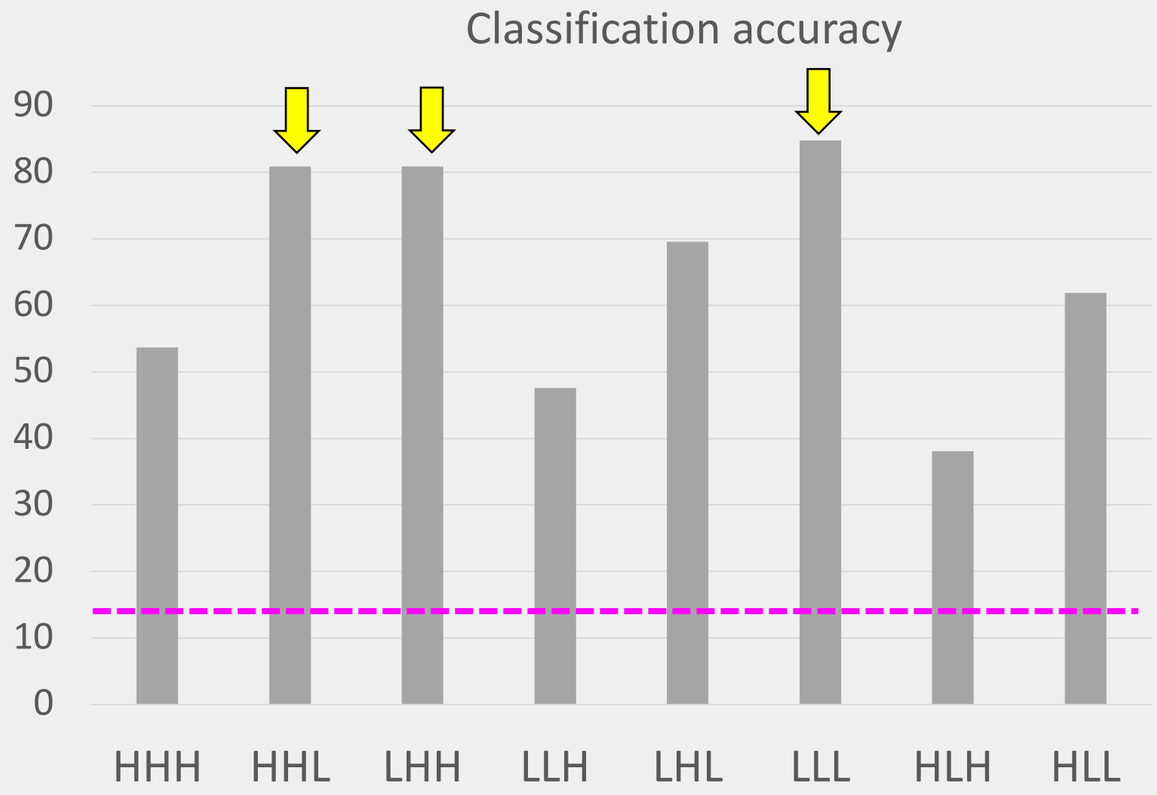


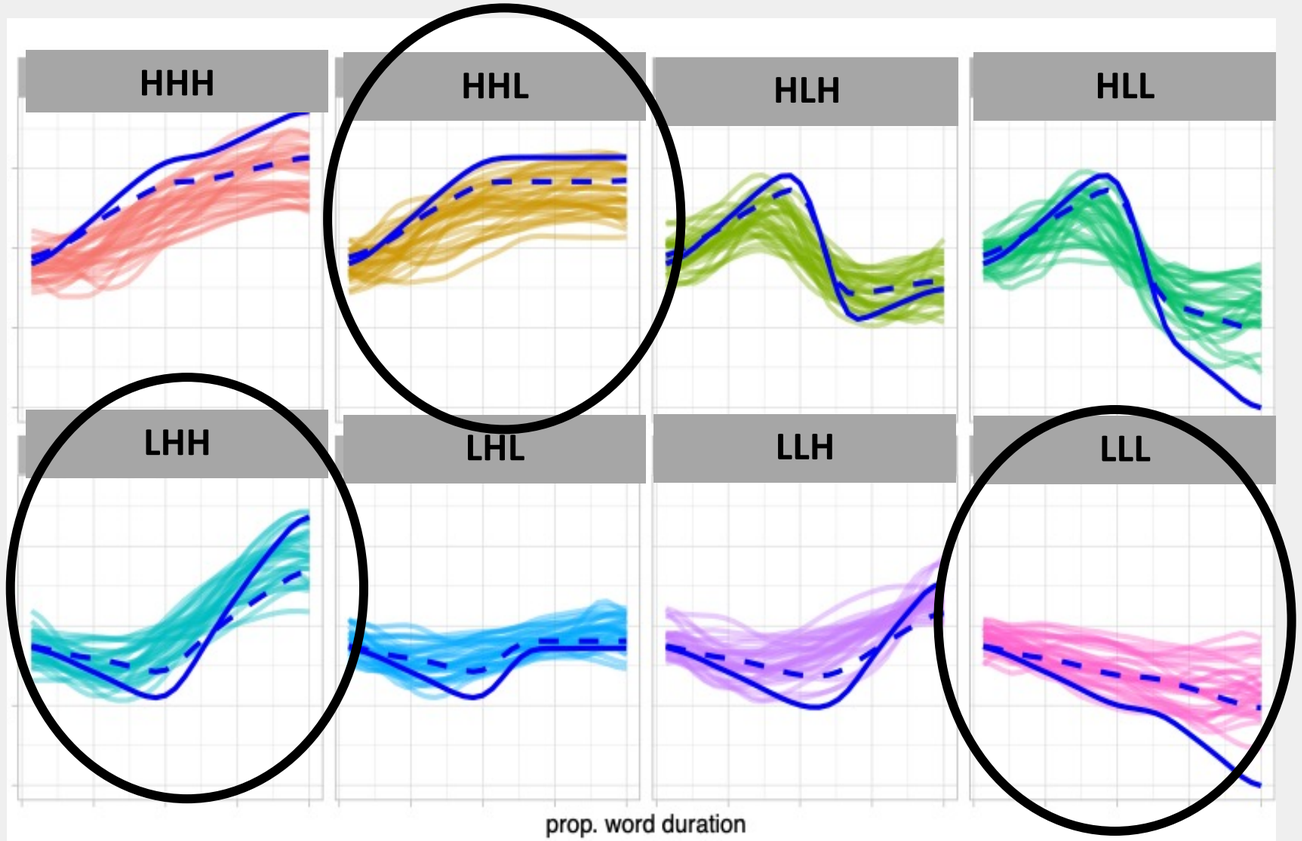
(Chodroff & Cole 2019; Cole, Steffman, Tilsen, Shattuck-Hufnagel, 2021)

Neural Net Classifier (LSTM)

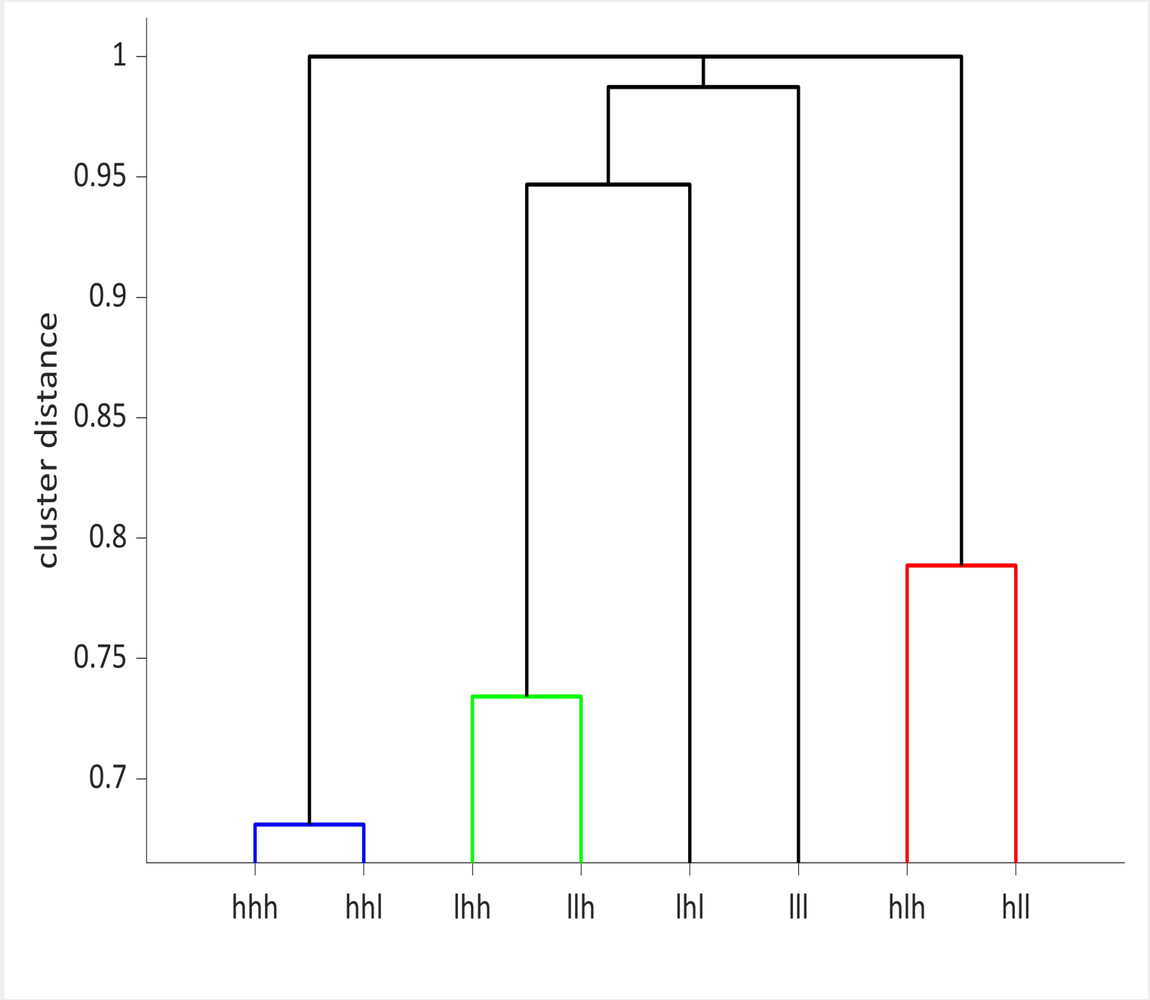
trained on labeled F0 trajectories (speaker centered)

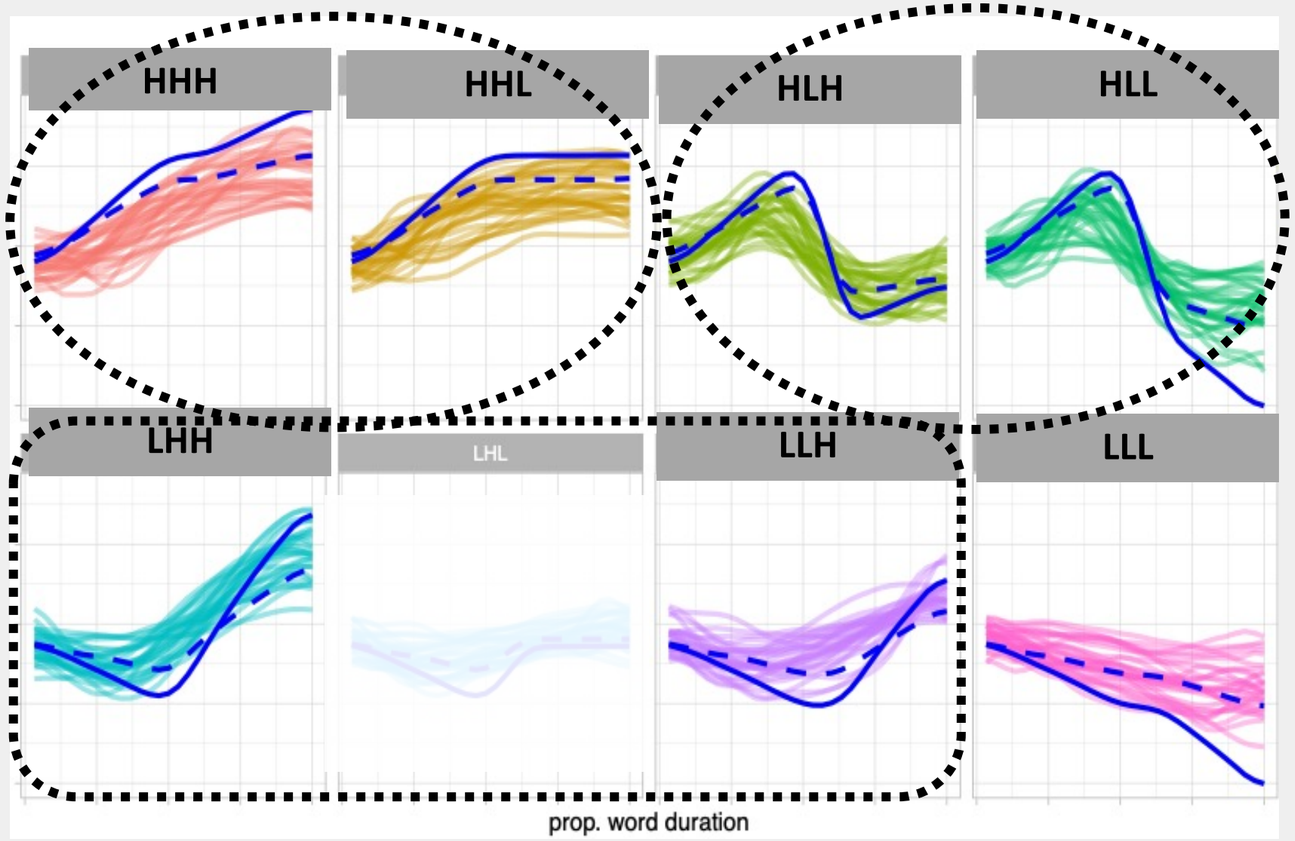
Overall accuracy: 0.61 (chance: 0.125)



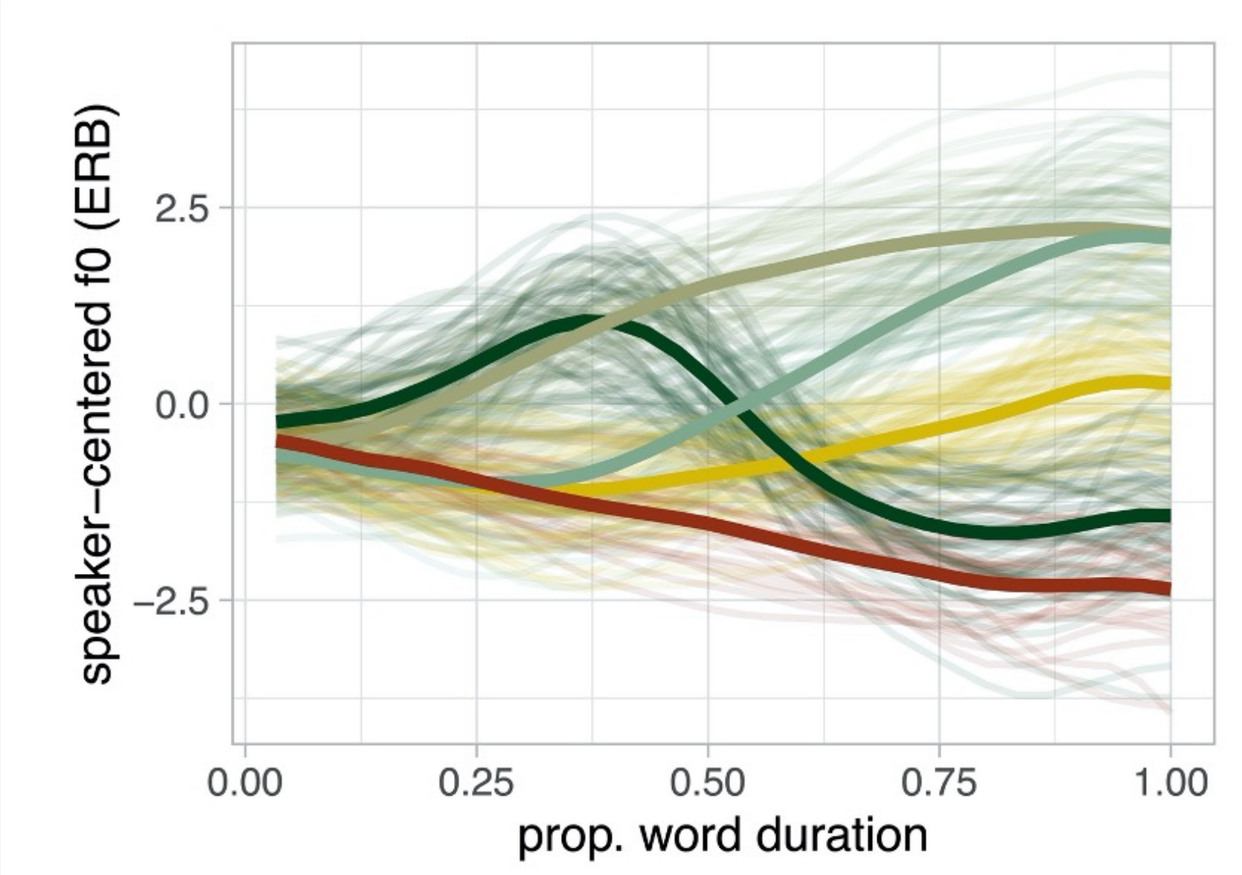


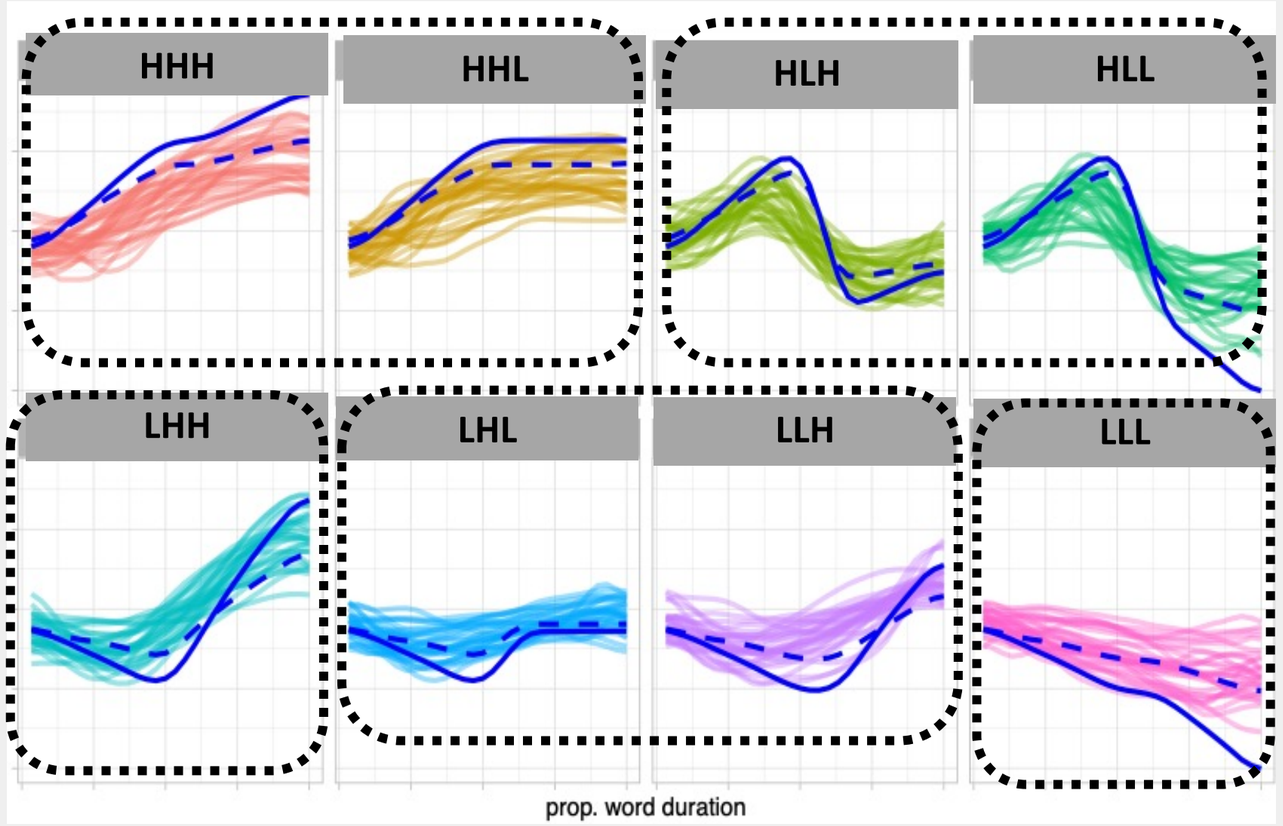
Pairwise tune similarity based on classifier output





Clustering analysis of F0 trajectories (*k-means clustering for longitudinal data*)
The optimal solution groups the tunes into five distinct clusters



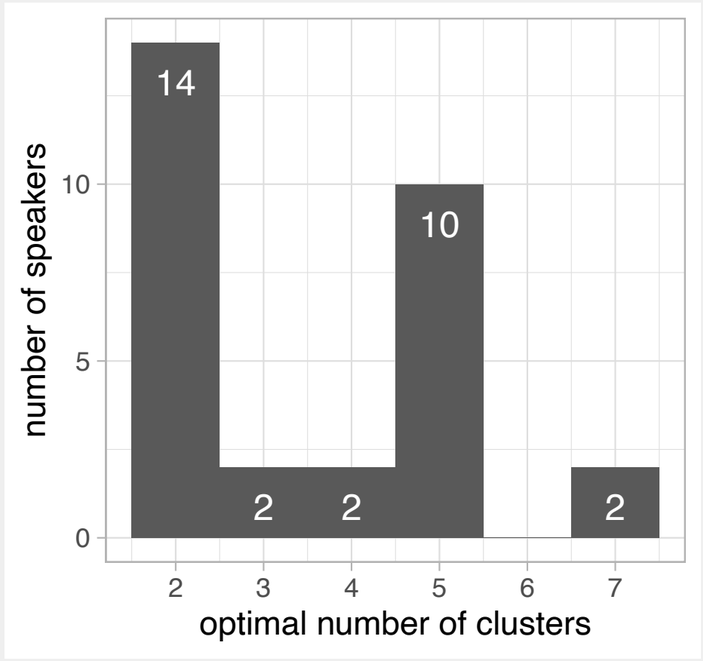


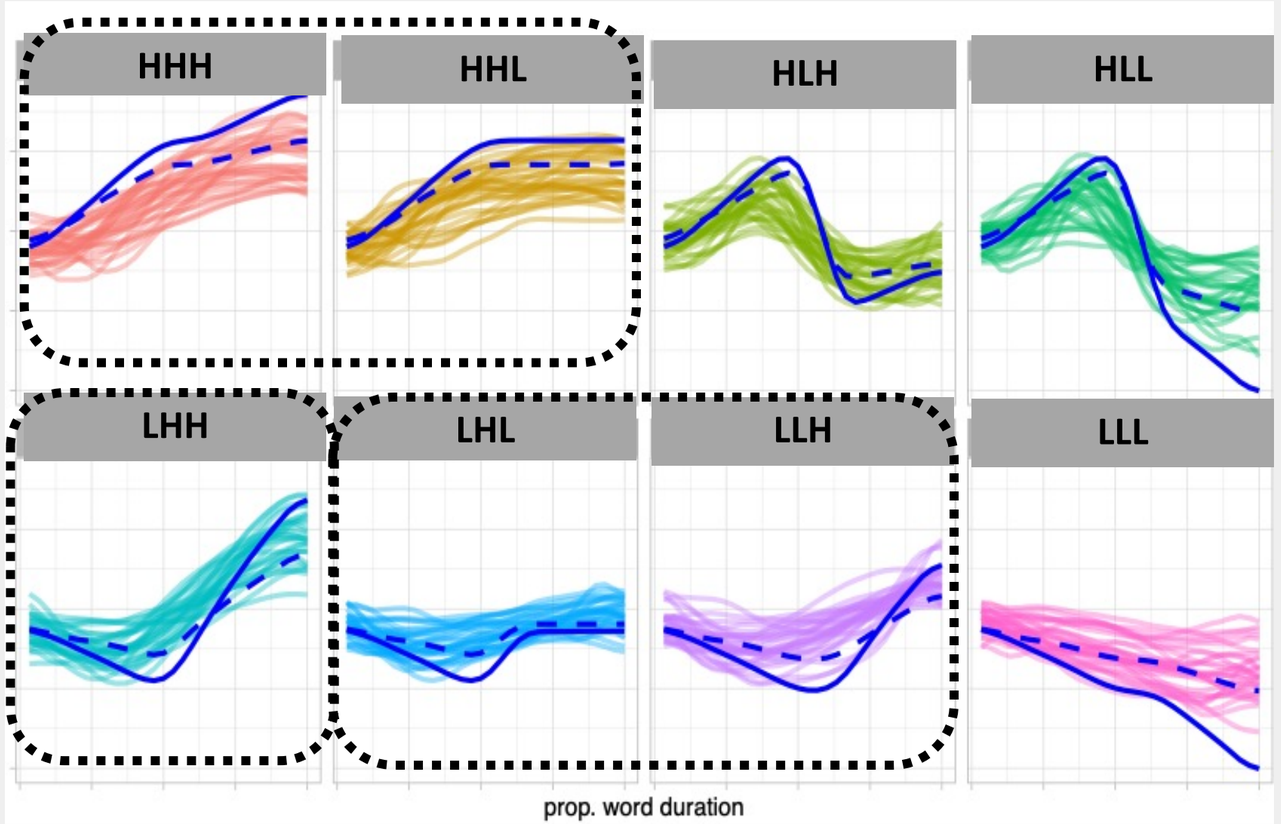
Every speaker has a rising cluster.

14 speakers distinguish only **RISING** and **NON-RISING** clusters.

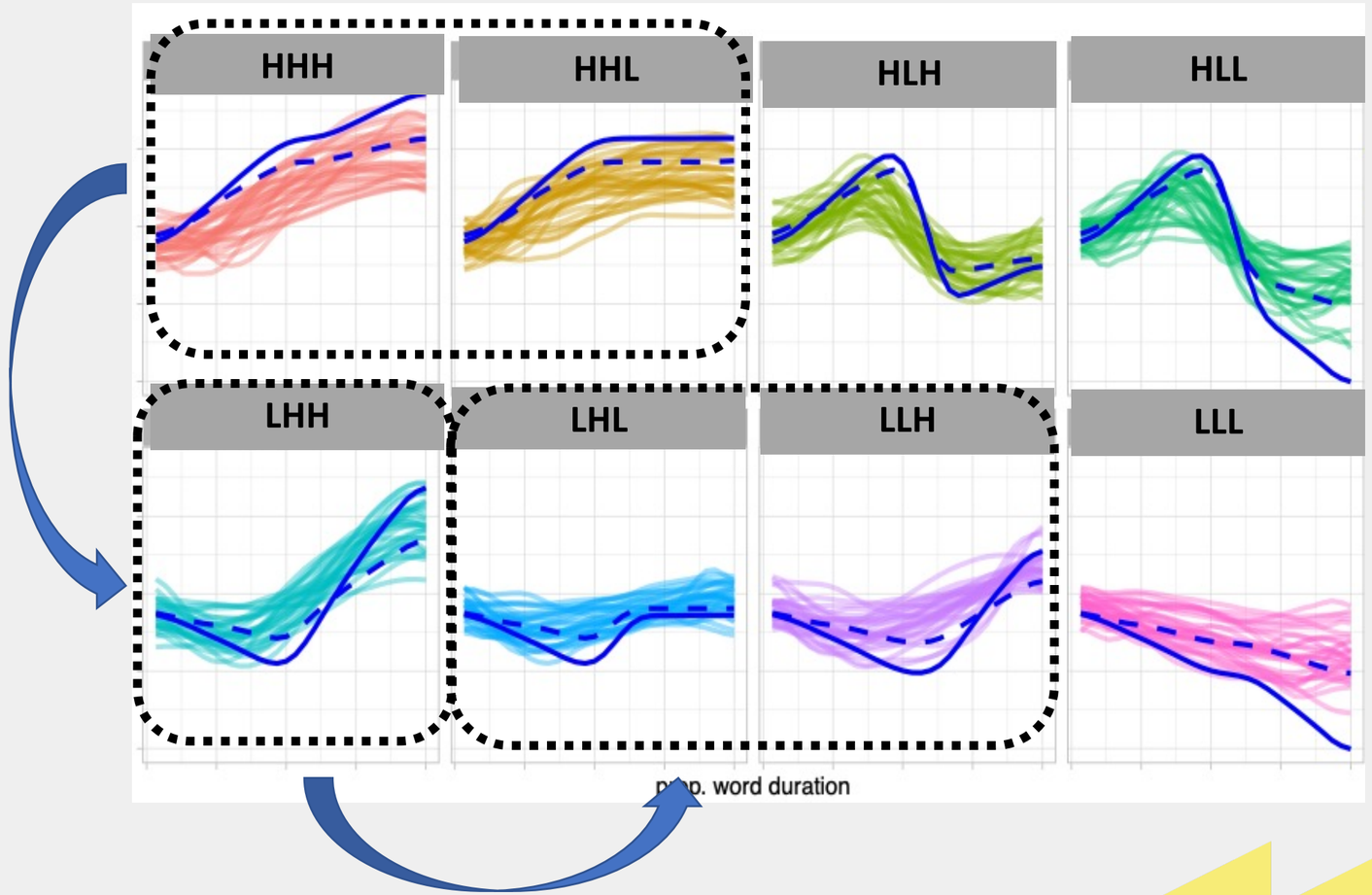
The rising cluster systematically generalizes.

# of clusters for individual speakers





A hierarchy of rising tunes



Summary

At the group level, 5 out of 8 hypothesized nuclear tunes are produced, capturing gross distinctions:

rising vs. non-rising
rise with low vs. high onset
low-to-mid rise (≈ mid-plateau)

Summary

At the group level, 5 out of 8 hypothesized nuclear tunes are produced, capturing gross distinctions:

rising vs. non-rising
rise with low vs. high onset
low-to-mid rise (≈ mid-plateau)

A rising tune hierarchy emerges...corresponds to distinctions in the tonal ‘center of gravity’ (preliminary analysis)

These tunes were produced as naturalized imitations, with no context. Will more distinctions emerge when context is present, or with interactive speech?

Wrapping up

Despite pervasive and substantial variation, prosody plays a key role in speech perception and comprehension.

Many factors influence speakers' production of prosody, and listeners cope with this variation, taking into account information about the linguistic and non-linguistic context.

Moving Forward

CONTEXT

STRUCTURED VARIATION

Moving Forward

Mounting the challenge with **data-intensive analyses**, using computational and statistical tools, maximizing automated methods.

Discoveries so far:

Listeners track and entrain to perceptually salient aspects of prosody associated to linguistic **and** non-linguistic meaning. *Jointly entrained, are linguistic and non-linguistic prosody linked in cognitive representation?*

Structured variation in nuclear tune production reveals a primary **dichotomy** between Rising and Non-Rising tunes, with individual variation in the tunes that are included in the Rising group, reflecting phonetic properties, e.g., rise center-of-gravity. *Suggests fewer tune categories; within-category variation due to non-linguistic factors?*

Our grand challenge

How do languages use prosody?

… expanding empirical methods, complex data, diverse languages

… brings opportunities for scientific discovery

…and applications with positive societal impact!

Thanks!

Thanks to my (former) postdocs, (former) students and collaborators!

Eleanor Chodroff
Timothy Mahrt
Shivani Patel
Uwe Reichel
Timo Roettger

Kate Sandberg
Stefanie Shattuck-Hufnagel
Thomas Sostarics
Jeremy Steffman
Sam Tilsen

AOVIVO.**ABRALIN**.ORG

Funding from NSF, the Volkswagen Stiftung and Northwestern University

linguists online

References

Barnes, J., Brugos, A., Veilleux, N., & Shattuck-Hufnagel, S. (2021). On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics*, *85*, 101020.

Beckman, M. E. (1996). The Parsing of Prosody. *Language and Cognitive Processes*, *11*(1–2), 17–68.

Brown M, Salverda AP, Dilley LC, Tanenhaus MK. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychon Bull Rev*, 18:1189–1196.

Brown M, Salverda AP, Gunlogson C, Tanen- haus MK. (2015). Interpreting prosodic cues in discourse context. Lang Cogn Neurosci 2015, 30:149–166.

Burdin, R. S., & Tyler, J. (2018). Rises inform, and plateaus remind: Exploring the epistemic meanings of “list intonation” in American English. *Journal of Pragmatics*, *136*(136), 97–114.

Buxó-Lugo, A., Toscano, J. C., & Watson, D. G. (2018). Effects of Participant Engagement on Prosodic Prominence. *Discourse Processes*, *55*(3), 305–323.

Calhoun S. The theme/rheme distinction: accent type or relative prominence? (2012). *J Phon*, 40:329–349.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual Differences in Speech Production and Perception* (pp. 123–145). Frankfurt am Main: Peter Lang. https://doi.org/10.3726/978-3-653-05777-5

Carlson K. How prosody influences sentence comprehension. (2009). *Lang Linguist Compass*, 3:1188–1200.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics,* 29, 155–190.

Cho T, McQueen JM, Cox EA. (2007). Prosodically driven phonetic detail in speech processing: the case of domain-initial strengthening in English. *J Phon*, 35:210–243.

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31.

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452.

Dahan D, Tanenhaus MK, Chambers CG. (2002). Accent and reference resolution in spoken-language comprehension. *J Mem Lang*, 47:292–314.

De Marneffe, M.-C., & Tonhauser, J. (2019). Inferring meaning from indirect answers to polar questions: The contribution of the rise-fall-rise contour . In M. Zimmermann, K. von Heusinger, & V. E. Onea Gaspar (Eds.), *Questions in Discourse: Volume 2 Pragmatics* (pp. 132–163). Leiden, The Netherlands: Brill.

Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences: Second Edition* (pp. 521–602). Chicester: John Wiley & Sons, Ltd.

Frazier L, Carlson K, Clifton C Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn Sci*, 10:244–249.

Geluykens, R., & Swerts, M. (1994). Prosodic cues to discourse boundaries in experimental dialogues. Speech Communication, 15, 69–77.

Goodhue, D., Harrison, L., Cl, Y. T., & Wagner, M. (2015). Toward a bestiary of English intonational tunes. In C. Hammerly & B. Prickett (Eds.), *Proc. North East Linguistics Society,* Vol. 46, pp. 311–320.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

References (continued)

Ito K, Speer SR. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *J Mem Lang*, 58:541–573.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2017). Effects of distributional information on categorization of prosodic contours. *Psychonomic Bulletin and Review*, 1–8.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge, UK: Cambridge University Press.

Porges, S. W. (2007). The Polyvagal Perspective. *Biological Psychology*, *74*(February 2007), 116–143.

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning – the case of information structure in American English. *Language, Cognition and Neuroscience*, *34*(7), 841–860.

Reichel, U.D. (2017). CoPaSul Manual: Contour-based, parametric, and superpositional intonation stylization, arXiv:1612.04765

Reichel, U. D., Beňuš, Š., & Mády, K. (2018). Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication*, 100, 46–57.

Salverda AP, Dahan D, McQueen JM. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90:51–89.

Salverda AP, Dahan D, Tanenhaus MK, Crosswhite K, Masharov M, McDonough J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, 105:466–476.

Swerts, M., Bouwhuis, D. G., & Collier, R. (1994). Melodic cues to the perceived “finality” of utterances. *Journal of the Acoustical Society of America*, 96, 2064–2075.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*(1993), 21–43.

Tomlinson, J. M., Gotzner, N., & Bott, L. (2017). Intonation and Pragmatic Enrichment: How Intonation Constrains Ad Hoc Scalar Inferences. *Language and Speech*, *60*(2), 200–223.

van Zyl, M., & Hanekom, J. J. (2012). When “okay” is not okay: Acoustic characteristics of single-word prosody conveying reluctance. Journal of the Acoustical Society of America, 133(1), EL13–EL19.

Veilleux, N., Shattuck-Hufnagel, S., Brugos, A. (2006). *6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI.* January IAP 2006. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

Wennerstrom, A. (2001). *The music of everyday speech.* Oxford: Oxford University Press.

Xie, X., Buxó-Lugo, A., & Kurumada, C. (2021). Encoding and decoding of meaning through structured variability in intonational speech prosody. *Cognition*, *211*, 104619.