

Testing the distinctiveness of intonational tunes: Evidence from imitative productions in American English

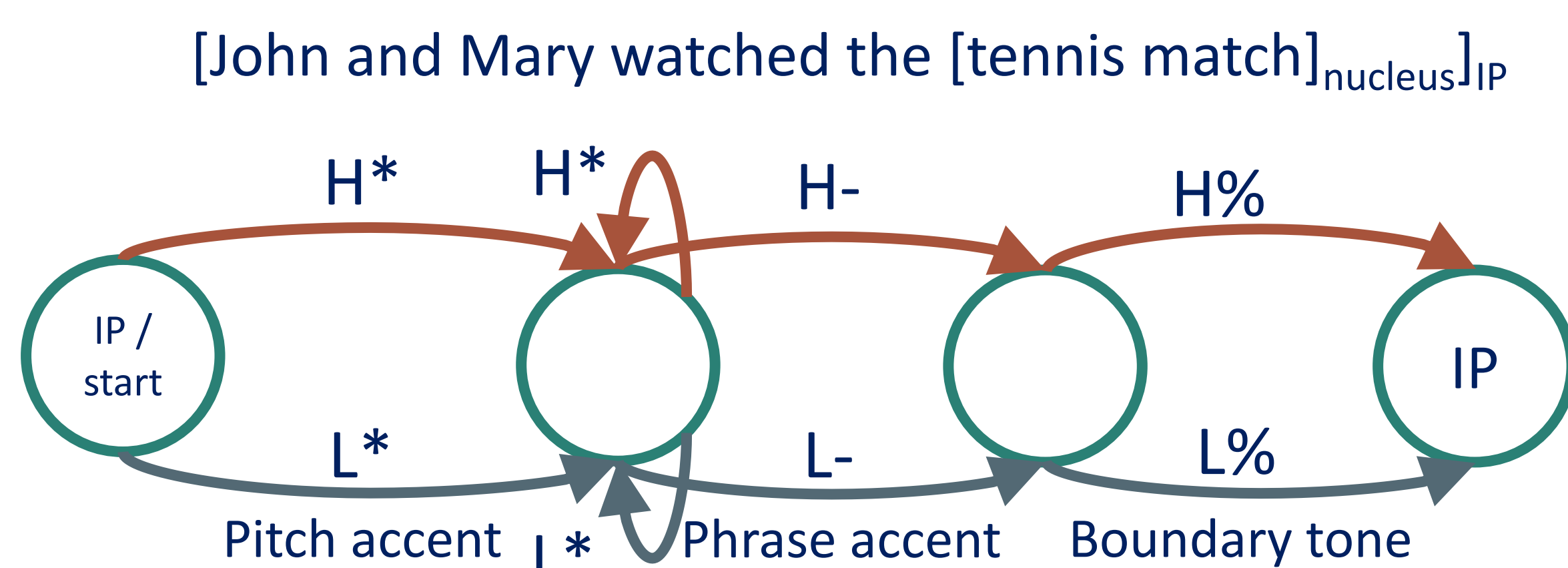
Eleanor Chodroff¹ and Jennifer Cole²

¹University of York, Department of Language and Linguistic Science | ²Northwestern University, Department of Linguistics

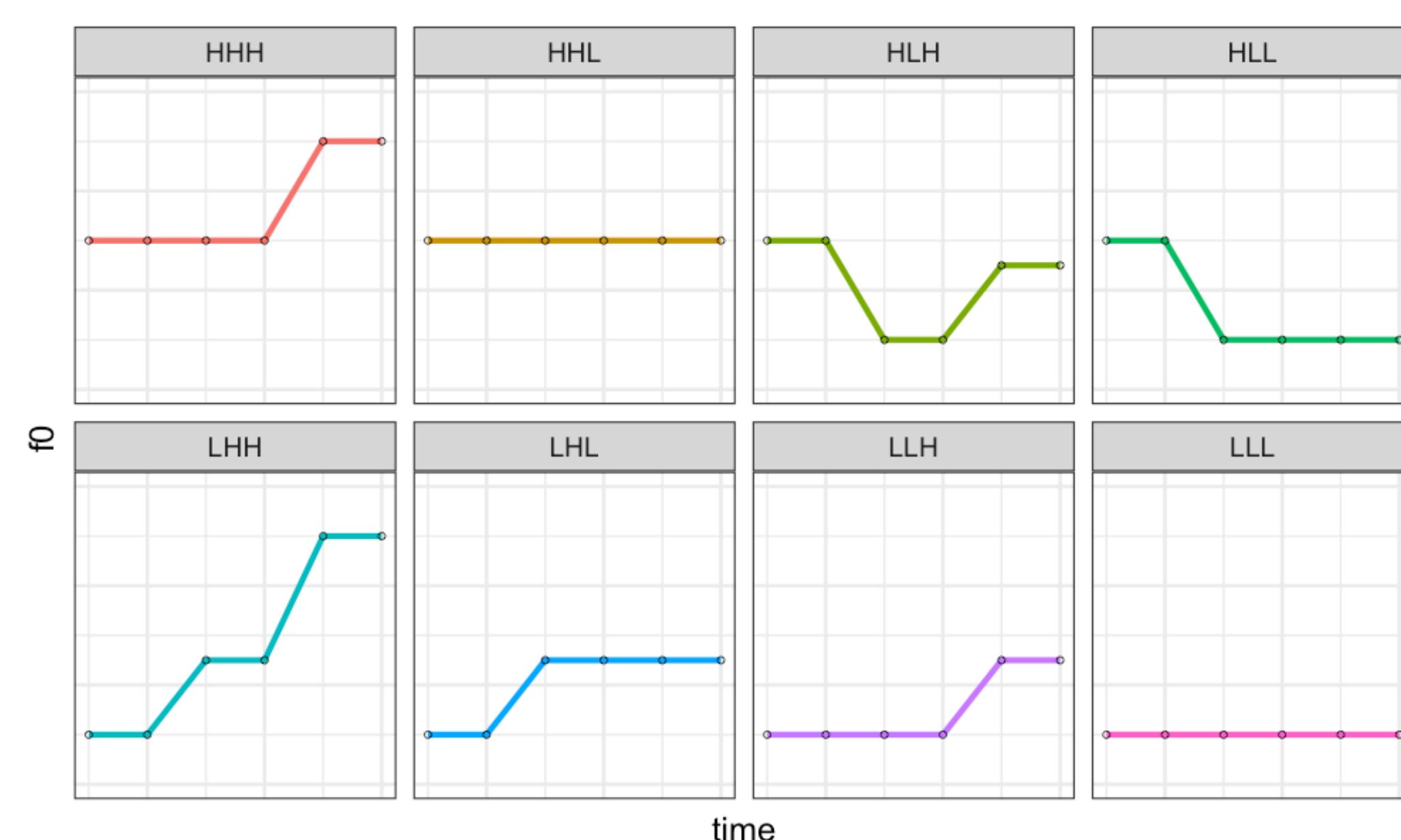
INTRODUCTION

Understanding the structure of intonational variation is a longstanding issue in prosodic research. A given utterance can be realized with countless intonational contours, and while variation in prosodic meaning is also large, listeners nevertheless converge on relatively consistent form-function mappings. This consistency suggests the existence of abstract intonational representations, but it has been unclear how exactly these are defined.

Pierrehumbert 1980: For American English, every intonational phrase (IP) ends in a sequence of three tonal components: a pitch accent, a phrase accent, and a boundary tone.



Simplified picture: additional bitonal pitch accents in the full proposed inventory



Do American English speakers maintain the eight-way distinction among **nuclear** intonational contours posited to exist in this representational system?

We thank Stefanie Shattuck-Hufnagel, Alejna Brugos, and members of the Prosody and Speech Dynamics Lab at Northwestern for helpful discussion, questions and commentary. We also thank Amelia Kimball and Tommy Denby for their help in stimulus creation, as well as Alaina Arthurs, Priya Kurian, and Jonah Pazol for data collection and processing.

METHODS

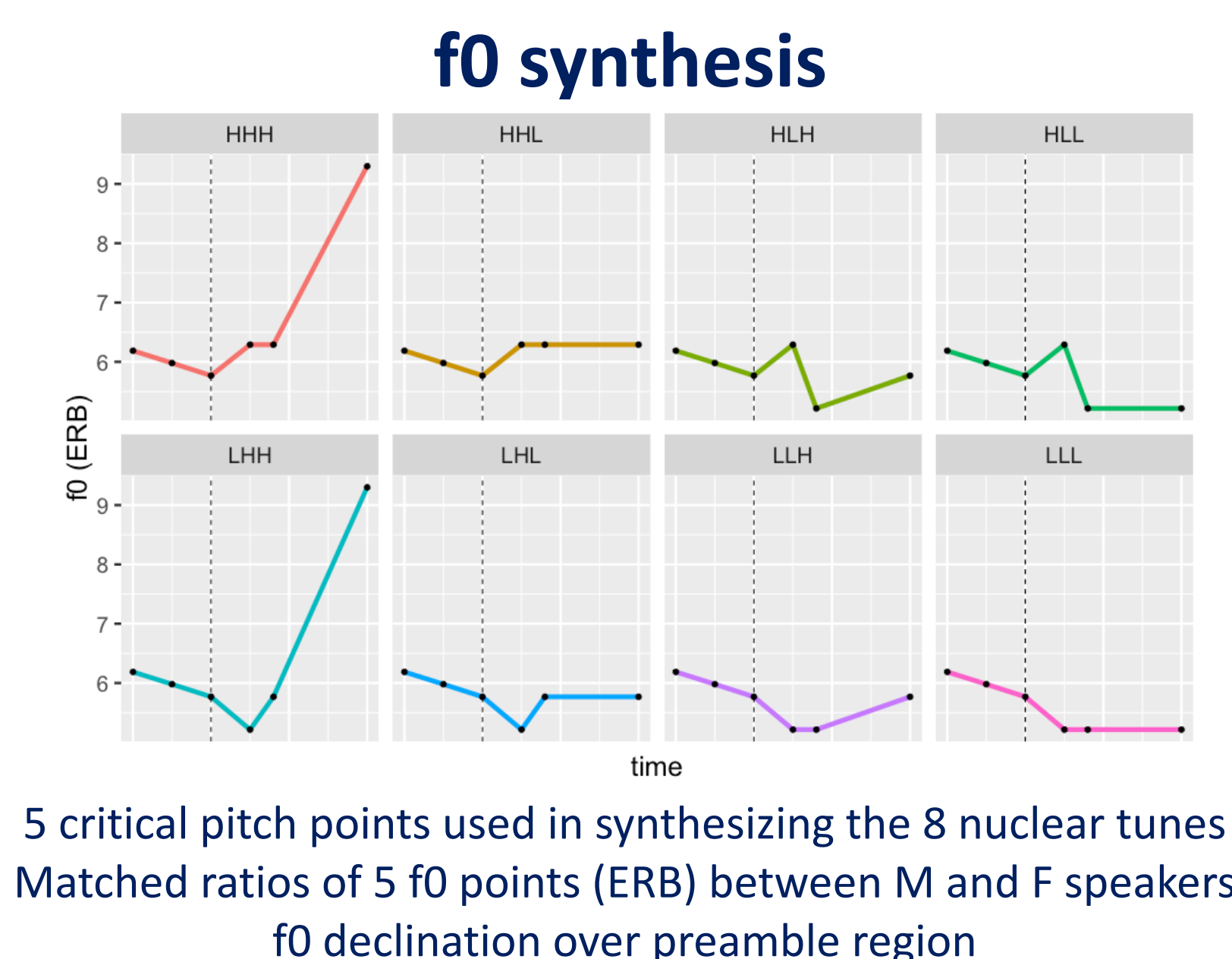
Natural productions + **f0 synthesis**

Male and female model speaker
3 model sentences x 8 contours

“She quoted Helena”
“Her name is Marilyn”
“He answered Jeremy”

Summary statistics of speaker-specific f0 used in synthesis parameterization

Selected flat exemplars (LLH, LLL) to use as the base for the synthesized f0 overlay



Experiment

32 participants | native speakers of American English | 13M, 19F

HEAR THREE:

Three examples of the same tune
Counterbalanced M/F exposure within and across trials

SAY ONE:

“He modeled Harmony”
“They honored Melanie”
“She remained with Madelyn”

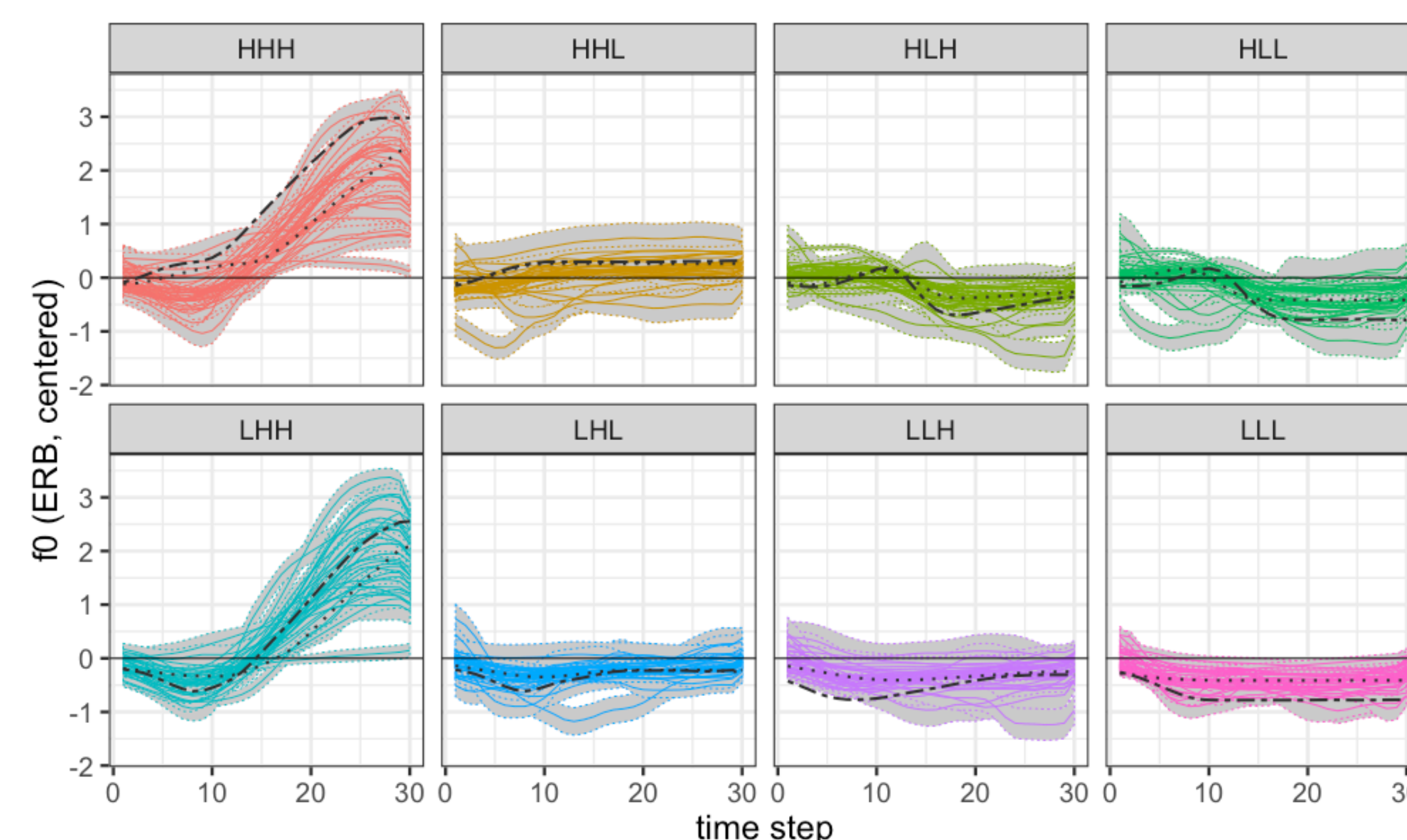
144 trials / participant

“Listen to the melody of the 3 sentences, and produce a new sentence with the same melodic pattern, but said the way you think it should sound if it were spoken by a human English speaker, in a manner that is familiar to you.”

RESULTS

Nuclear f0 contours

Averaged by participant and tune

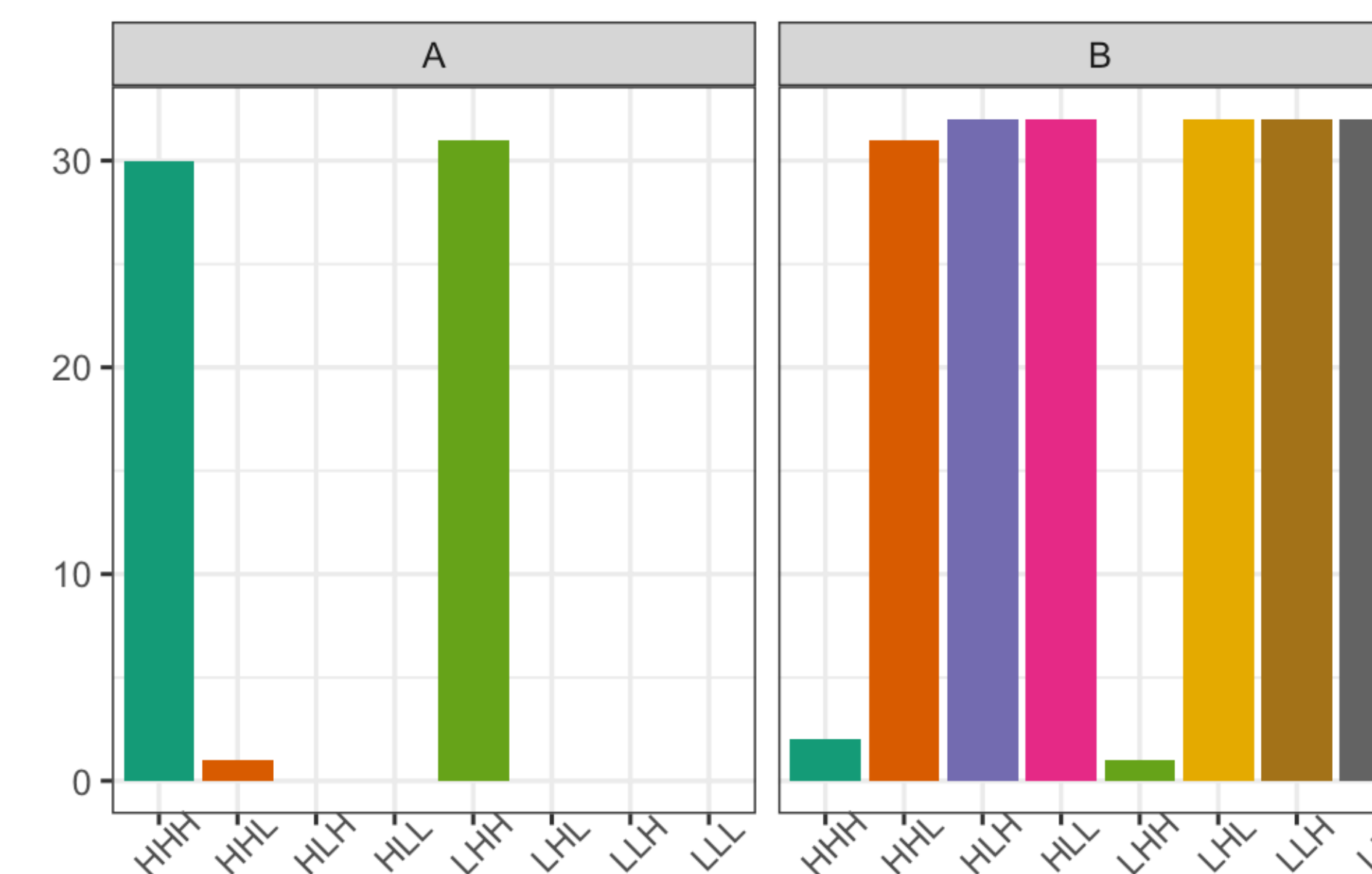


Deviation Analysis

Deviation between raw f0 contour and model speaker average

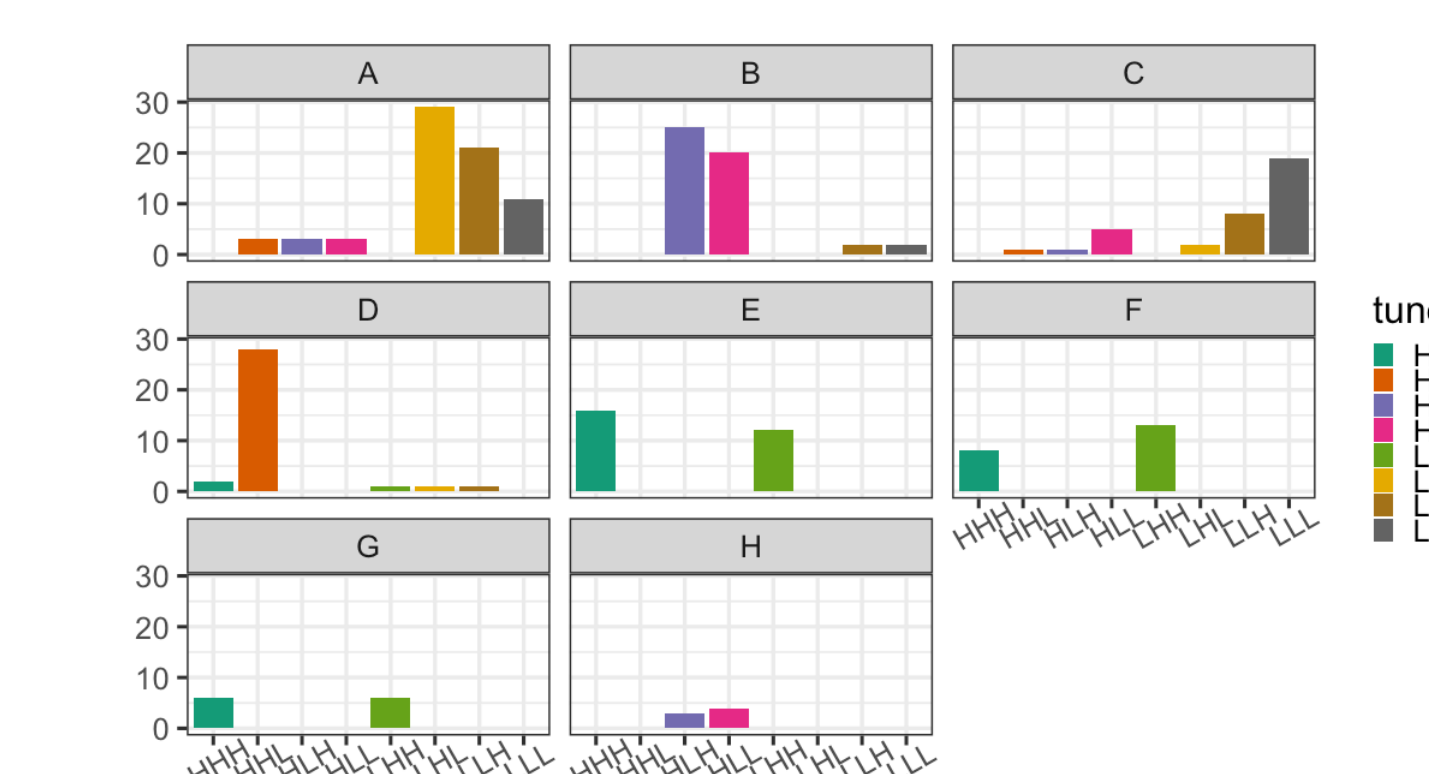
Tune	1. HHH	2. LHH	3. HLL	4. HLH	5. LLH	6. HHL	7. LLL	8. LHL
RMSE	0.90	0.79	0.48	0.46	0.44	0.43	0.43	0.37

Longitudinal K-means Cluster Analysis

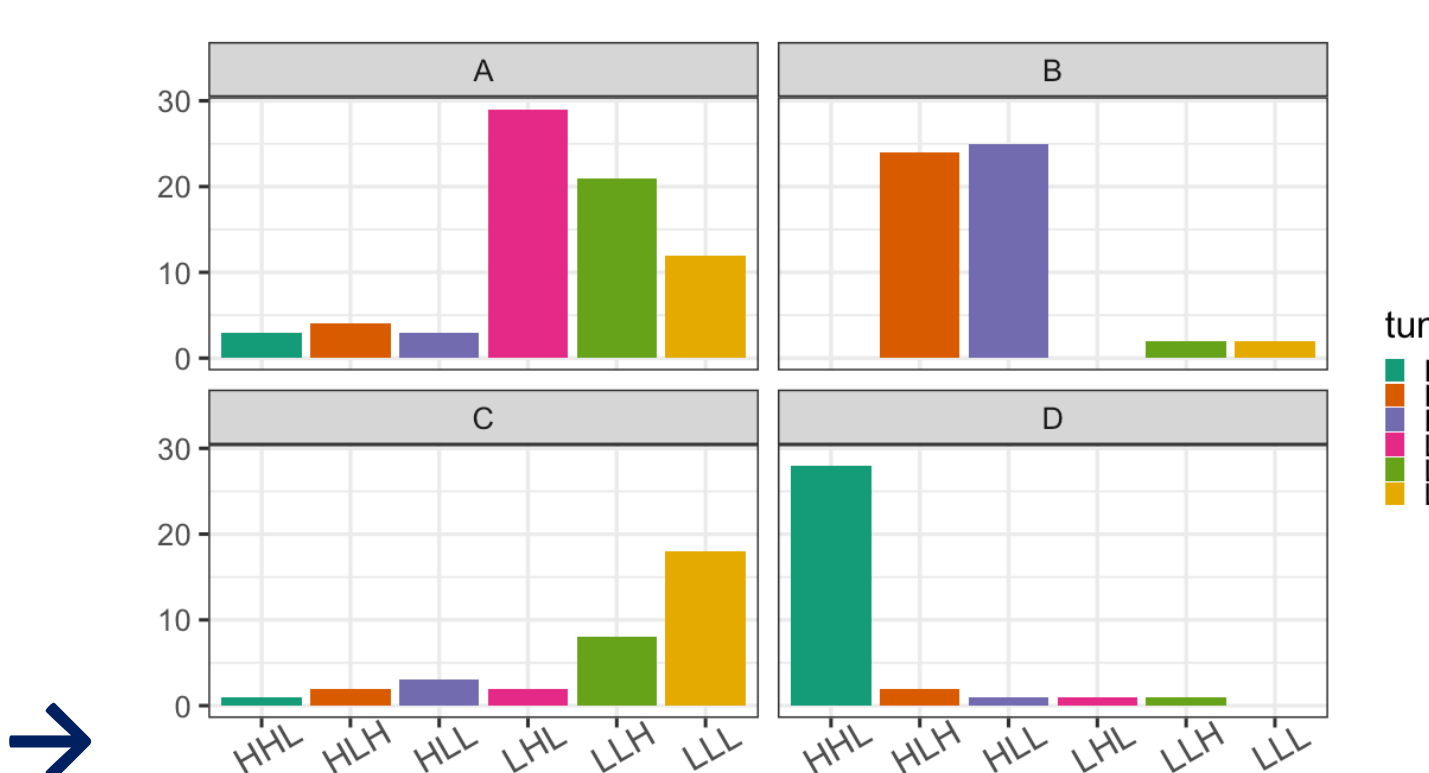


↑ Two optimal clusters from all averaged contours

Four optimal clusters from non-rising averaged contours →



↑ Forced 8-cluster classification



SUMMARY

The distinction between rising and non-rising f0 contours is readily accessible in speech perception and production for American English listeners and speakers

→ Salient acoustic and illocutionary distinction in American English
→ Reflects the cross-linguistic tendency to make use of this prosodic contrast

Among the non-rising contours produced, four contour types emerged:
Flat-mid Mid-falling Low-to-mid Low

Non-rising contours were imitated with less precision. Possible explanations:

- The f0 resynthesis did not succeed in tapping into the intended phonological representations
- Difficulty accessing an appropriate meaning hindered the ability to reproduce a target intonation