Separable effects of speech input statistics on phonetic and talker processing

Sahil Luthra[1] & Lori L. Holt[2]
[1]Carnegie Mellon University
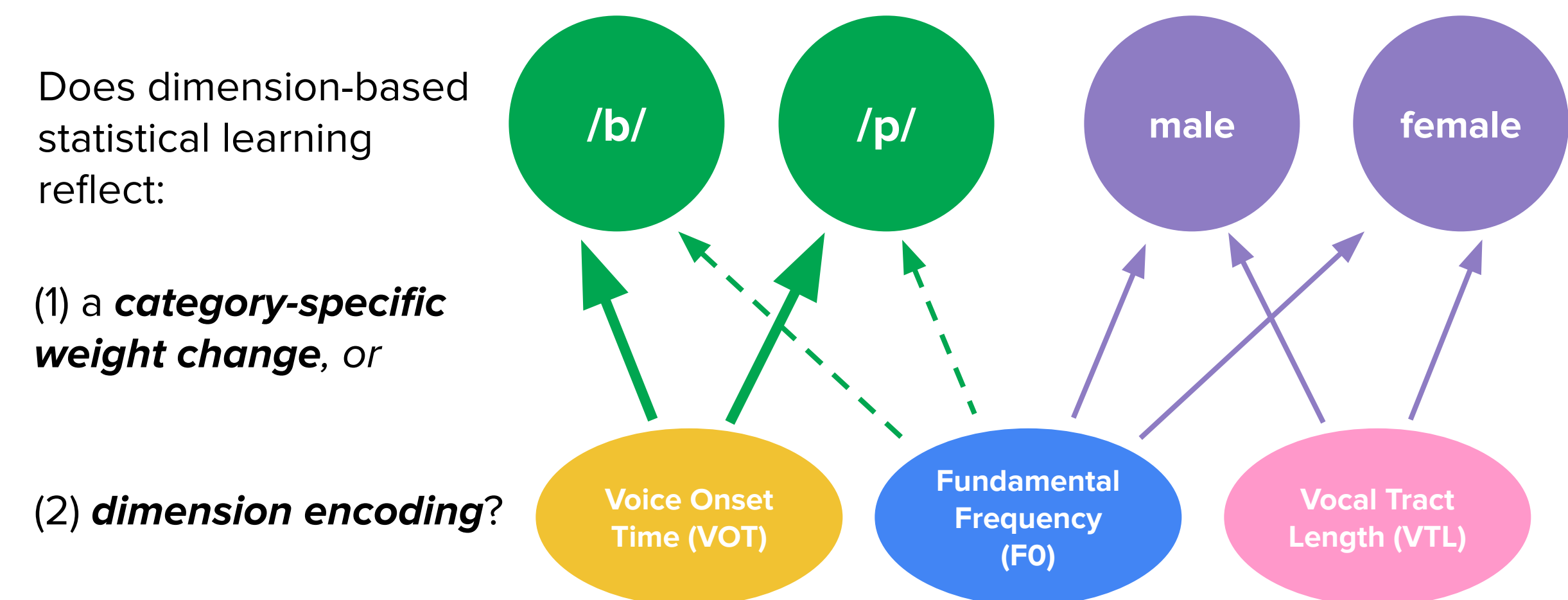[2]The University of Texas at Austin

Dimension-based statistical learning of speech involves adjustments to the weights across which acoustic dimensions activate categories, not auditory encoding

Introduction

Statistical learning dynamically tunes *perceptual weights*, the informativeness of an acoustic dimension in signaling category membership. When short-term speech regularities run counter to long-term experience, listeners *down-weight* the secondary dimension (e.g., Idemaru & Holt, 2011; Jasmin et al., 2023; Liu & Holt, 2015). For example, an artificial accent that reverses the usual English F0xVOT correlation conveying /b/ vs. /p/ leads to reduced reliance on F0.

**Does statistically driven down-weighting of a dimension reflect an influence on
(1) category-specific perceptual weight or
(2) dimension encoding?**

Speech conveys both linguistic and socio-indexical information, with some acoustic dimensions supporting both

Does dimension-based statistical learning reflect:

/b/   /p/   male   female

(1) a *category-specific weight change*, or

(2) *dimension encoding*?

Voice Onset Time (VOT)   Fundamental Frequency (F0)   Vocal Tract Length (VTL)

Exposure to speech with an atypical F0xVOT correlation leads F0 to less effectively convey /b/-/p/ speech category identity (Idemaru & Holt, 2011). If **(1)**, then this should have no impact on male/female voice categorization. But if **(2)**, then F0 may less effectively convey male/female voice categories.

Across two experiments, we exposed listeners to stimuli with **canonical** (**orange squares**; aligned with long-term English experience) or **reverse** (**red squares**) F0xVOT correlation.

F0: 250, 230, 210, 190, 170, 150, 130
beer, *pier*
VOT: -10, 0, 10, 20, 30, 40, 50
VTL

Test stimuli had constant, ambiguous VOT, making categorization F0-reliant. Thus, **test stimulus categorization** reveals the influence of F0 in b/p and male/female categorization.
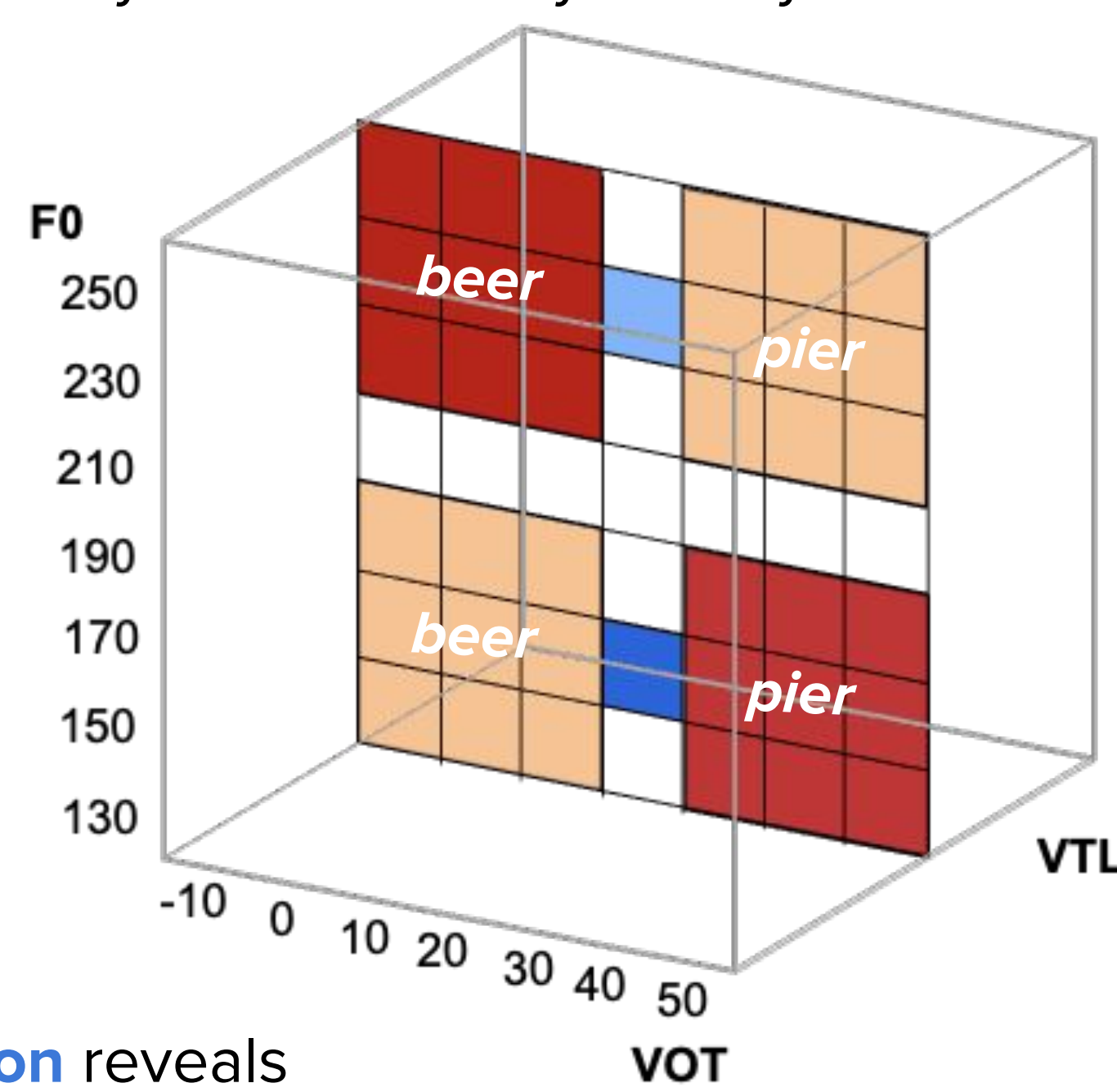
References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388-407.

Hodson, A. J., Shinn-Cunningham, B. G., & Holt, L. L. (2023). Statistical learning across passive listening adjusts perceptual weights of speech input dimensions. *Cognition, 238*, 105473.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1939-1956.

Jasmin, K., Tierney, A., Obasih, C., & Holt, L. (2023). Short-term perceptual reweighting in suprasegmental categorization. *Psychonomic Bulletin & Review, 30*(1), 373-382.

Lehet, M., & Holt, L. L. (2020). Nevertheless, it persists: Dimension-based statistical learning and normalization of speech impact different levels of perceptual processing. *Cognition, 202*, 104328.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1783-1798.

Wu, Y. C., & Holt, L. L. (2022). Phonetic category activation predicts the direction and magnitude of perceptual adaptation to speech. *Journal of Experimental Psychology: Human Perception and Performance, 48*(9), 913-925.

Zhang, X., & Holt, L. L. (2018). Simultaneous tracking of coevolving distributional regularities in speech. *Journal of Experimental Psychology: Human Perception and Performance, 44*(11), 1760-1779.
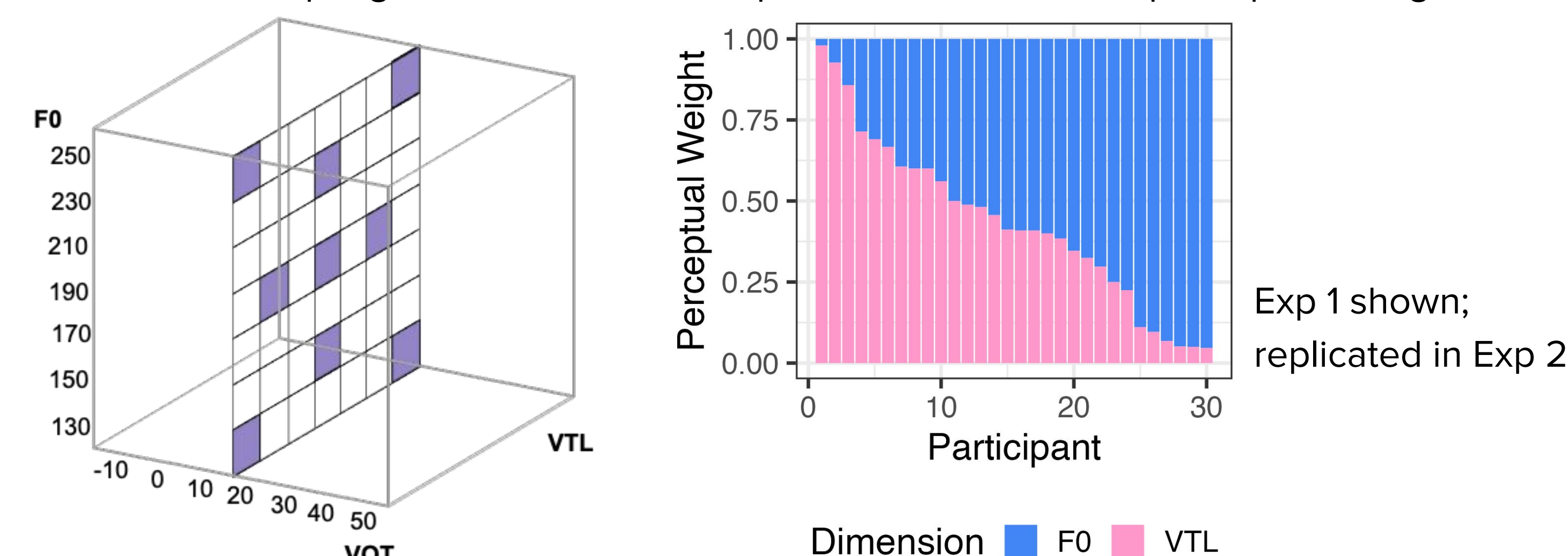
Methods and Results

For each experiment, N = 30 adult speakers of English (ages 18-35) with no language or hearing impairments were recruited online via Prolific.ac and completed the experiment on Gorilla (Anywl-Irvine et al., 2020). Stimuli were adapted from Zhang and Holt (2018) using the Praat Vocal Toolkit.

Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Male/Female	*Beer/Pier*	Male/Female	*Beer/Pier*	Male/Female	*Beer/Pier*
Baseline	**Baseline**	**Canonical**	**Canonical**	**Reverse**	**Reverse**
F0 x VTL	F0 x VOT	F0 x VOT	F0 x VOT	F0 x VOT	F0 x VOT

F0=fundamental frequency; VOT=voice onset time; VTL=vocal tract length
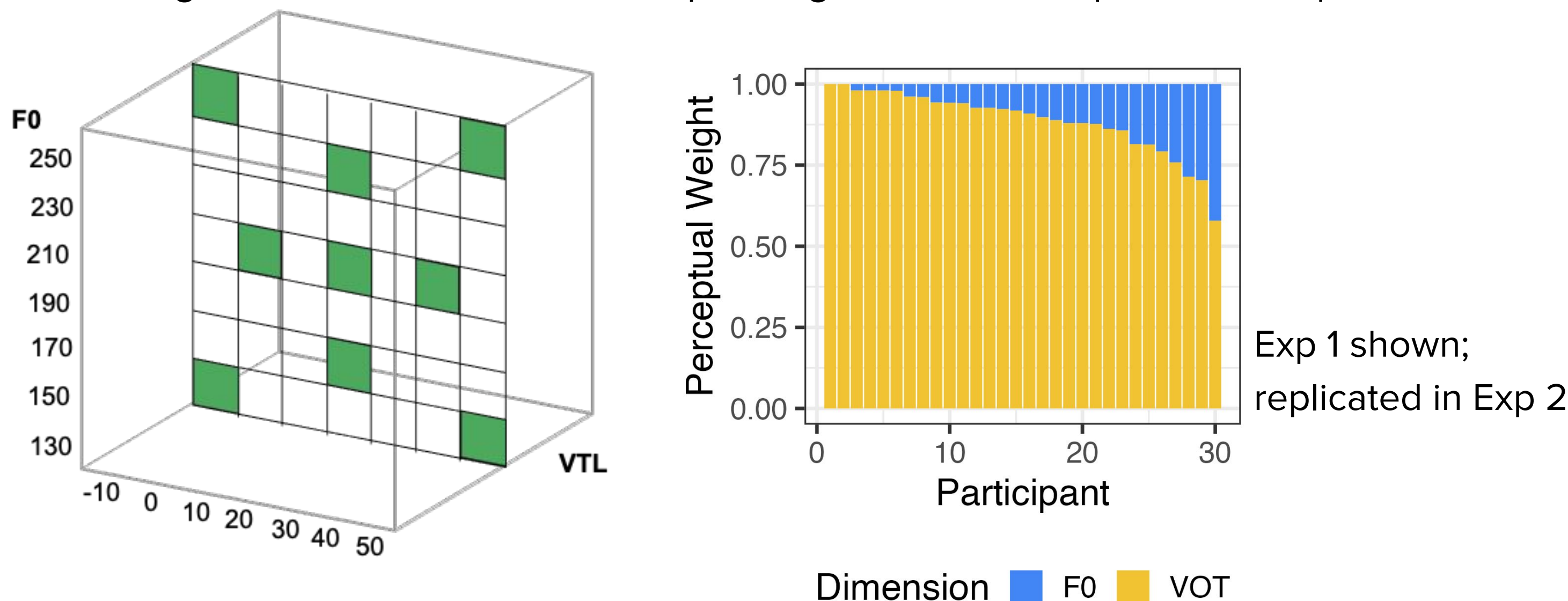
**Listeners vary in perceptual weight of F0 vs. VTL
for male/female voice categorization**

A balanced sampling across the F0xVTL space reveals baseline perceptual weights.

F0: 250, 230, 210, 190, 170, 150, 130
VOT: -10, 0, 10, 20, 30, 40, 50
VTL

Perceptual Weight / Participant

Exp 1 shown; replicated in Exp 2

Dimension: F0   VTL

**Listeners perceptually weight VOT more than F0
for *beer/pier* phonetic categorization**

Listeners weight VOT more than F0 in /b/-/p/ categorization, as expected from previous work.

F0: 250, 230, 210, 190, 170, 150, 130
VOT: -10, 0, 10, 20, 30, 40, 50
VTL

Perceptual Weight / Participant

Exp 1 shown; replicated in Exp 2

Dimension: F0   VOT

**Subsequent blocks examined influence of
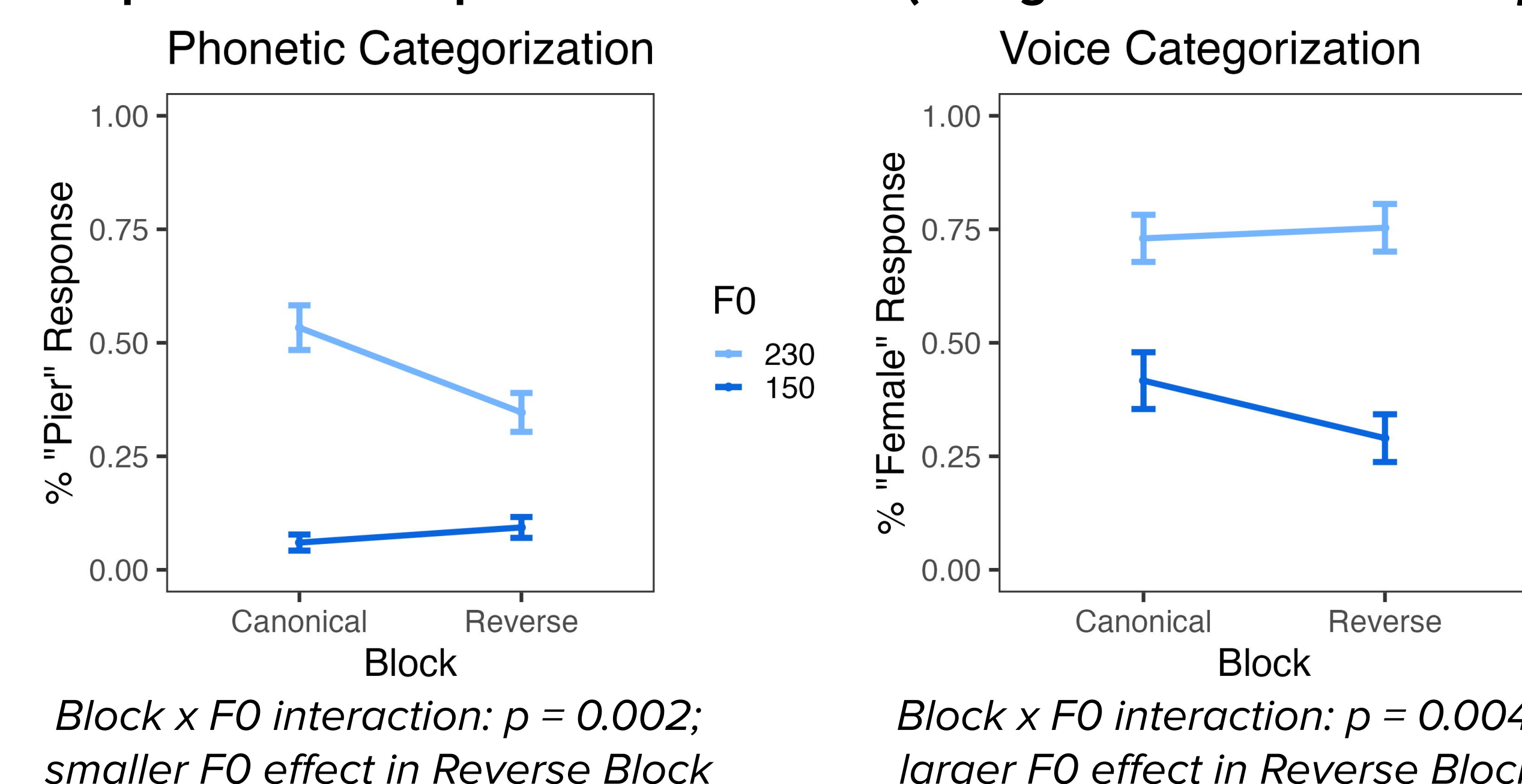statistical learning across short-term phonetic F0xVOT regularities**

Exp 1, Passive Listening (schematized at right): Eight stimuli provide short-term phonetic regularity in F0xVOT space that is **canonical** or **reverse**. One of two F0-differentiated **test stimuli** follows and participants make an overt response (after Hodson et al., 2023).

Exp 2, Active Categorization: Participants overtly categorize both **canonical** and **reverse** exposure and **test stimuli** (after Idemaru & Holt, 2011)
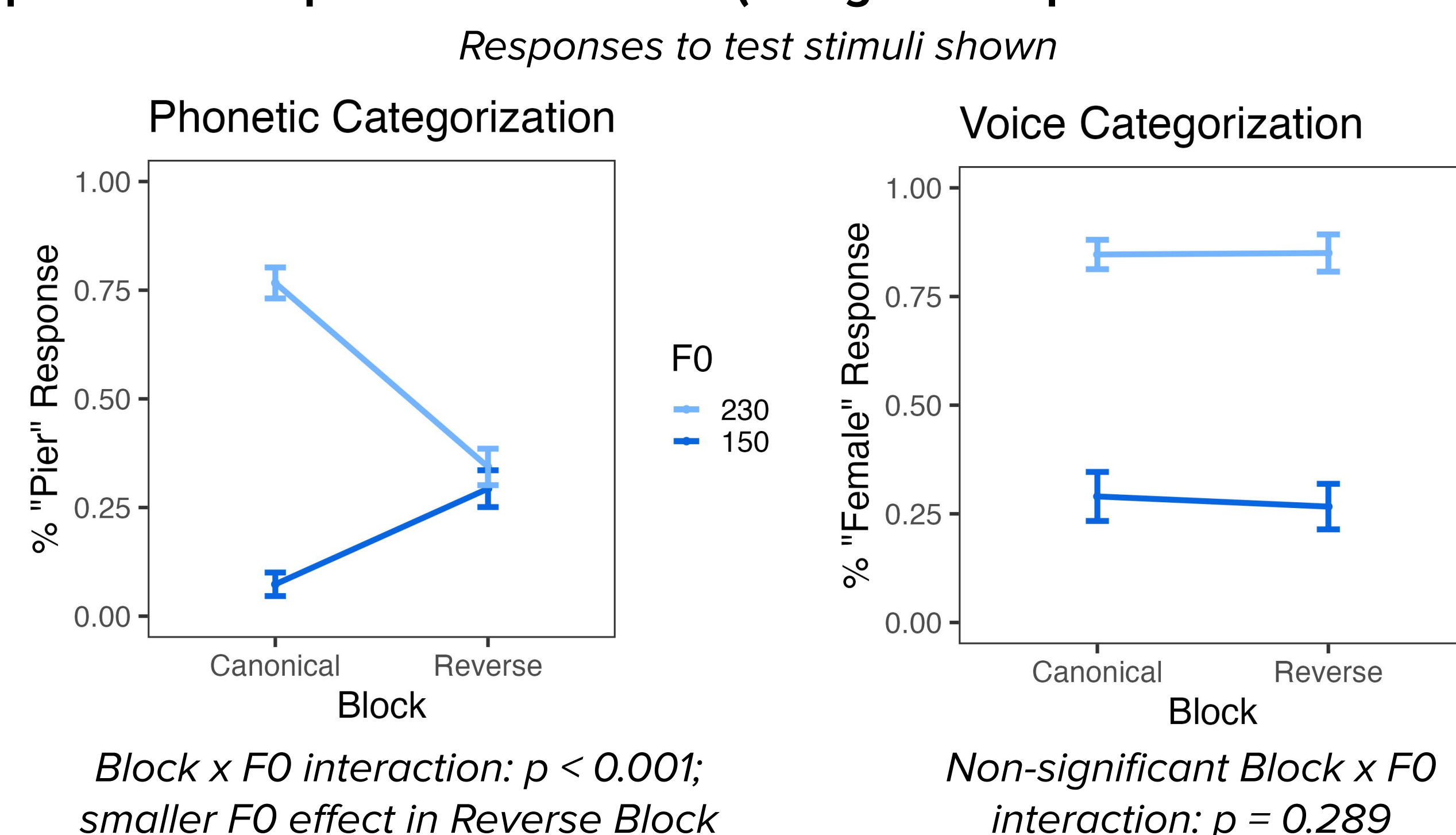
Blocks 3/5: male/female categorization
Blocks 4/6: *beer/pier* categorization

Passive listening

Time

Overt categorization

beer Press J   pier Press K

When listeners encounter regularities that oppose long-term phonetic experience, they down-weight F0 for phonetic categorization but not for voice categorization

Exp 1: Passive exposure to statistics (categorize test stimuli only)

Phonetic Categorization
% "Pier" Response
Block: Canonical, Reverse
F0: 230, 150

Voice Categorization
% "Female" Response
Block: Canonical, Reverse
F0: 230, 150

Block x F0 interaction: p = 0.002; smaller F0 effect in Reverse Block

Block x F0 interaction: p = 0.004; larger F0 effect in Reverse Block

Exp 2: Active exposure to statistics (categorize exposure and test stimuli)

Responses to test stimuli shown

Phonetic Categorization
% "Pier" Response
Block: Canonical, Reverse
F0: 230, 150

Voice Categorization
% "Female" Response
Block: Canonical, Reverse
F0: 230, 150

Block x F0 interaction: p < 0.001; smaller F0 effect in Reverse Block

Non-significant Block x F0 interaction: p = 0.289

Conclusions

Statistically driven down-weighting of a dimension appears to reflect an influence of category-specific perceptual weight

Robust statistical learning driven by phonetic short-term regularities **does not correspondingly impact F0 reliance in male/female categorization,** arguing against an influence on dimension encoding (**Hypothesis 1**).

Instead, convergent with Lehet & Holt (2020), statistical learning appears to impact category-specific weights (**Hypothesis 2**). F0 remained available to differentially affect voice categorization even as it was down-weighted in *beer-pier* categorization.

Wu & Holt (2022) provide evidence that **category activation via the dominant dimension** (e.g., VOT) **predicts the magnitude of down-weighting** (on F0), consistent with a learning mechanism driven by category activation. The current results reveal that at baseline, **individuals differ widely in whether VTL or F0 is the dominant dimension for male/female categorization.** Thus, future work might examine down-weighting in voice categorization via the dominant dimension (VTL, F0).