

Byron Ahn (*Princeton U.*), Alejna Brugos (*Simmons U.*), Sunwoo Jeong (*Seoul National U.*),
Stefanie Shattuck-Hufnagel (*MIT*), Nanette Veilleux (*Simmons U.*)

Cues to a Speaker's Previous Beliefs in English Intonation

Background. A shared discourse model (“Common Ground”; Stalnaker 1978, Farkas & Bruce 2010) gets updated by each new proposition (p), and the intonation used to communicate p influences how the CG is updated (Pierrehumbert & Hirschberg 1990, Gulongson 2004, a.o.): e.g., final-rises attribute commitment to p to the hearer, and final-falls to the speaker. Despite views like this being well-established, relatively few aspects of intonation have been studied, and often only as (purposefully) idealized contours (e.g., the binary distinction between a final-rise and final-fall). This paper adds to the growing research into intonational meaning, in an underexplored dimension of semantics / pragmatics: how a speaker's previous beliefs can be signalled intonationally (A vs. B in [Table 1](#)).

Hypotheses. H1: Holding a prior belief of $\neg p$ (condition B) will yield a rising pitch accent & final-fall (“surprise-redundancy contour”; Sag & Liberman 1975). **H2:** The hearer having better knowledge about p (greater epistemic authority; condition A & B) will yield a steeper final-rise (Jeong 2018). Notably, H1 and H2 make contradictory predictions for B.

Methods. Native speakers of American English participated in an internet-based production task. To convey the necessarily rich pragmatic contexts, each target sentence is couched in a comic strip discourse ([Fig.1](#)), with response particles like “okay so” or “oh really” preceding it. Phonologically-informed phonetic measures that have been previously used to identify sem/prag meaning differences from intonation (TCoG, *cf.* Barnes *et al.* 2012; PoLaR, *cf.* Ahn *et al.* 2022) were extracted and submitted to machine learning techniques (Random Forests, *cf.* Arnhold & Kyröläinen 2017). To date, over 600 files have consensus labels from 2+ labellers (annotation still ongoing); 485 are from conditions A and B. Files were excluded from analysis if the target word did not bear the final pitch, yielding 299 files.

Results. Random Forest models ([Table 2](#)) revealed that key measures in yielding accurate results included final lengthening, edge tone TCoG, nuclear accent TCoG, and local f0 range. Investigating these features more closely, the B-condition measures appear more strongly bimodal than the A-condition measures, and expert listeners perceived qualitatively different types of rises/falls. The data was divided by condition and rise/fall, and submitted to Wilcoxon T-tests, revealing statistically significant differences in the phonetics of these rises and falls, across the sem/prag conditions (A vs. B). Condition B's rises had a higher normed final f0 that lasted longer ([Fig.2a](#)), its final accents had later TCoGs with lower values (suggesting a low accent going into a rise; [Fig.2b](#)), and its normed local f0 ranges were wider by virtue of going higher ([Fig.2c](#))

Discussion. First, this work identifies a new way that different types of rises/falls can mark different meanings. Second, the bimodality in condition B may result from the context underdetermining which of these contours is appropriate; perhaps this is the source of H1 & H2's conflicting predictions. However, though condition A was more uniform, it was in the opposite direction as predicted by H2. This requires further analysis in intonation (e.g., phonemic labels) and meaning (e.g., a comprehension task). Finally, and most broadly, this methodology (*internet-sourced production data, systematic tracking of subphonemic phonetics with PoLaR, using TCoG measures on both accents and edge tones, and using ML techniques*) was able to reveal some of the ways in which these two conditions differ intonationally. This work adds to the types of meaningful differences that intonation can signal, opening up several avenues for future research.

	Speaker's previous belief: p	Speaker's previous belief: $\neg p$
Epistemic authority: Hearer	A	B
Epistemic authority: Speaker	C	D

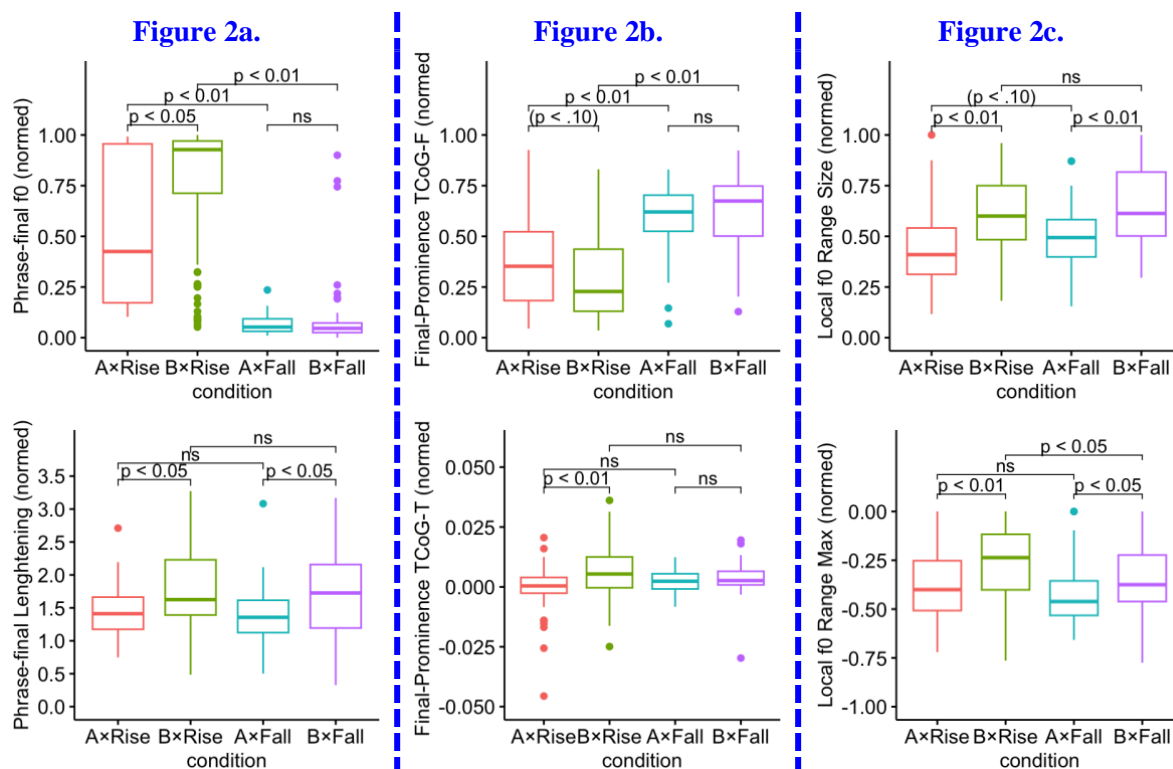
Table 1. Four experimental elicitation conditions. There are 24 target sentences ($\times 4$ conditions = 96 stimuli), with a targeted 25 speakers per stimulus. This present work analyzes conditions A and B.



Figure 1. A comic strip context for the target sentence “There’s only an alligator in there”.

speaker	response particles	target sentence
age, gender	local pitch range (min, max, size), final pitch accent (TCoG), final edge tone (TCoG, slope, final f0 value), phrase final lengthening	local pitch range (min, max, size), penultimate pitch accent (TCoG), final pitch accent (TCoG), final edge tone (TCoG, slope, final f0 value), phrase final lengthening

Table 2. Measures submitted to Random Forests. The model with all of these measures had accuracy of ~80%; the model lacking response particles measures had accuracy of ~65%.



Selected Refs: Ahn et al. 2022. *How meaningful these intonational contours are!* Presented at the HSP 2022. • Arnhold & Kyröläinen. 2017. Modelling the interplay of multiple cues in prosodic focus marking. *LabPhon* 8. • Barnes et al. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *LabPhon* 3. • Gunlogson. 2004. *True to form: Rising and falling declaratives as questions in English*. • Jeong. 2018. Intonation and sentence-type conventions: Two types of rising declaratives. *Journal of Semantics* 35(2). • Sag & Liberman. 1975. The intonational disambiguation of indirect speech acts. *CLS* 11.