Testing the Locus of Speech-Act Meaning in English Intonation

Thomas Sostarics[1] and Jennifer Cole[1]

[1]Northwestern University, Department of Linguistics
tsostarics@u.northwestern.edu, Jennifer.cole1@northwestern.edu

ABSTRACT

Recent work on rising declaratives proposes a distinction between steep inquisitive rising declaratives and shallow assertive rising declaratives. Yet, it is unclear whether this contrast arises from a phonological distinction of the pitch accent used or a phonetic distinction in the scaling of the boundary tone target. In two perception experiments, we evaluate the contributions of pitch accent and boundary tone in the interpretation of assertive force. In Exp. 1, we find a counterintuitive result for the weighting of pitch accent, which is better understood from the perspective of the Tonal Center of Gravity. This perspective provides a path forward for Exp. 2, which shows no evidence of a contribution from the pitch accent in the interpretation of assertive force. Results speak against a phonological contrast in subtypes of rising declaratives and suggest a need for more narrow investigation in the phonetic domain.

Keywords: intonational meaning, prosody, speech perception, rising declaratives, compositionality

1. INTRODUCTION

The pitch contour from the final stressed syllable to the end of an utterance, the **nuclear tune**, conveys pragmatic meaning in Mainstream U.S. English (MUSE) [1]. In the dominant Autosegmental-Metrical theory of intonational phonology, tunes are made up of high- and low-tone building blocks: **pitch accents** and **edge-tones**, the latter of which are made up of configurations of **phrasal accents** and **boundary tones**. The phonological form of a tune is thus the sum of its parts: typical falling intonation can be described as H*L-L%, a fall (L-L%) from high (H*), and rising intonation can be described as L*H-H%, a rise (H-H%) from low (L*). Tune meaning can be described holistically (e.g., the meaning of "fall" vs. "rise") or compositionally, where pitch accent meaning is typically framed in terms of information structure while edge tones are framed in terms of speech acts or commitment [2,3]. Yet, studies on what pitch accents mean often do not make reference to the edge tones they co-occur with (and vice versa for edge tones). Validating a compositional approach to tune meaning thus requires targeted empirical investigation on whether the meaning conveyed by an individual tone is consistent in the context of other tones, which in turn comprise different tunes.

The contrast between rising and falling nuclear tunes is of particular interest for the phenomenon of **rising declaratives**. While declarative sentences with "default" falling intonation are interpreted as assertions [4], rising intonation changes the interpretation to that of a polar question. However, recent work [5] has proposed a further distinction within rising declaratives, where a steep-rising tune is interpreted as **inquisitive** while a shallow-rising tune is heard as **assertive**. While the two rising tunes differ in the speech act they convey, which would suggest a difference in the edge tones, [5] attributed the contrast to the pitch accent, with L* for the steep rise (in L*H-H%) and H* for the shallow rise (in H*H-H%).

Considering that pitch accents are typically taken to convey referential meaning related to information structure (i.e., focus, givenness), linking the pitch accent to the speech act contrast between the two rises would suggest that the locus of this contrast does not reside solely in the edge-tone configuration. However, despite the claimed contrast in the pitch accent (H* vs. L*), the pitch contours used for both the shallow- and steep-rising stimuli in [5] start from the same initial pitch value. This manipulation suggests that the difference in listeners' interpretation of shallow- and steep-rises was instead due to the phonetic scaling of the edge tones, with lower or higher final pitch. Viewed in this light, the evidence from [5] does not support the claim that the distinction between inquisitive vs. assertive is phonologically encoded through the contrastive intonational feature of the pitch accent. Yet, [5] opens the door to using rising declaratives to more critically consider whether the pitch accent has any bearing on the speech act conveyed by the nuclear tune, or alternatively, whether the locus of this meaning dimension resides solely in the edge tones.

The present paper revisits the form-function mapping between falling vs. rising intonation and assertion vs. question interpretations. We expand the materials from [5] in a two-alternative forced choice task to investigate how the phonetic implementation of rising and falling nuclear tunes relates to proposals regarding the phonological specification of these tunes. Our focus is not narrowly on shallow vs. steep rises, but also on variation in falling tunes. We investigate the form of falling pitch contours with

"gradual" linear slopes (Exp. 1) versus "early" falls that occur immediately after the pitch accent's peak (Exp. 2), a difference we analyze in terms of the Tonal Center of Gravity [6], which provides a common currency for shallow and steep falls as well as rises.

2. EXPERIMENT 1

We adopt the two-alternative forced choice paradigm in [5] to investigate how the accentual and ending pitch of the nuclear pitch contour modulate the probability of a question or assertion interpretation. On each trial, participants listen to a declarative sentence such as *Molly's from Branning*, where *Branning* bears a resynthesized nuclear pitch contour (details below). After listening, participants judge whether the speaker was **telling** them something (=assertion interpretation) or **asking** them something (=question) using the F and J keys on their keyboard.

After making their response, participants count aloud by 2s for 3-5 seconds, starting from a random number presented on the screen. The counting task is added to each trial to prevent participants from comparing pitch contours from trial to trial, as order effects have been previously found for the perception of prominence in rising and falling intonation [7].

Stimuli were created from naturally produced utterances through pitch resynthesis. As base recordings, we use five declarative sentences of the form {Name}'s {determiner/preposition} {noun} such as *Molly's from Branning*. The final noun is disyllabic with word-initial stress and bears the nuclear pitch contour. The first author recorded all utterances in a sound-attenuated booth using H*L-L% (falling) intonation while avoiding phrase-final creak to ensure successful pitch resynthesis.

The resynthesized nuclear pitch contours are linear falls or rises from an **accentual pitch** target (as the cue to pitch accent) to an **ending pitch** target (cueing the edge-tone configuration). Accentual pitch varies from 70Hz to 110Hz in five steps of 10 Hz and is aligned with the end of the stressed syllable of the nuclear word, following [1]. The onglide of the accentual rise/fall begins 50ms before the onset of the nuclear word, while all prenuclear material is held constant at the continuum midpoint (90Hz). Ending pitch varies by ERB-scale differentials based on production data from [8], with five steps between endpoints of -.25 and +2.5 ERBs. These differentials are then added to the lowest accentual pitch (70Hz) to yield five targets between 61Hz and 149Hz. This ending pitch continuum is then crossed with the accentual pitch continuum, shown in Fig. 1. We manipulated the duration of the nuclear word in Praat [9] to maintain a constant duration (thus ensuring equivalent pitch contours when resynthesized) for all utterances.

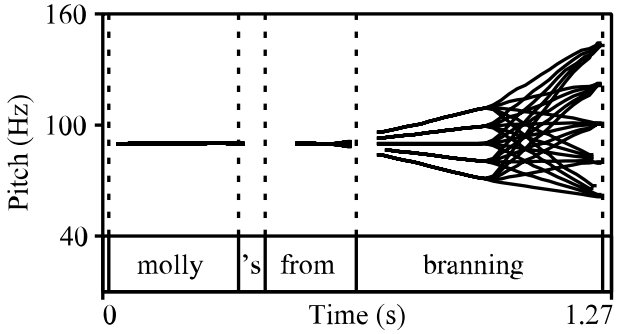

Figure 1: Final 25-step continuum crossing 5 accentual pitch targets and 5 ending pitch targets.

2.1. Experiment 1 Results

We recruited 56 online participants from the Prolific crowdsourcing platform. While participants were instructed to make their decisions as quickly as possible, we omitted trials with reaction times over eight seconds (~3% of the data) from the analysis based on participants' reaction time distributions.

We use Bayesian logistic mixed effects regression to test how accentual pitch, ending pitch, and their interaction (henceforth: the *scaling model*) affect the probability of a *telling* response. We include random intercepts by utterance and by participant and random slopes of accentual pitch, ending pitch, and their interaction by participant. We transformed pitch target values to semitones from 90Hz, which centers the predictors and allows the effects to be interpreted on the semitone scale. The intercept is thus the average log odds of a *telling* response at hypothetical flat pitch at 90Hz. All data, materials, and analyses are available online at osf.io/8hrfv.

Following work from [2,3,4] hypothesizing that differences in the commitment towards a proposition is primarily encoded by the edge-tone configuration and evidence from [5] showing that steeper rises sound more inquisitive than shallow rises, i.e., a higher ending pitch predicts lower likelihood of *telling* responses, *a priori* we predict a negative effect of ending pitch. Under a hypothesis that the pitch accent contributes to the question/assertion contrast [5], where L*H-H% is more inquisitive than H*H-H%, we predict a positive effect of accentual pitch. Alternatively, under a strict compositional account like [2], we would predict no effect of accentual pitch.

Given our bivariate continuum, the proportion of *telling* responses at each combination of accentual and ending pitch are plotted as a heatmap in Fig. 2, with variation in accentual pitch (for the pitch accent) on the horizontal dimension and variation in ending pitch (for the edge tones) on the vertical dimension. An effect of pitch accent would be seen by horizontal

gradation while an effect of edge tone would be seen by vertical gradation.

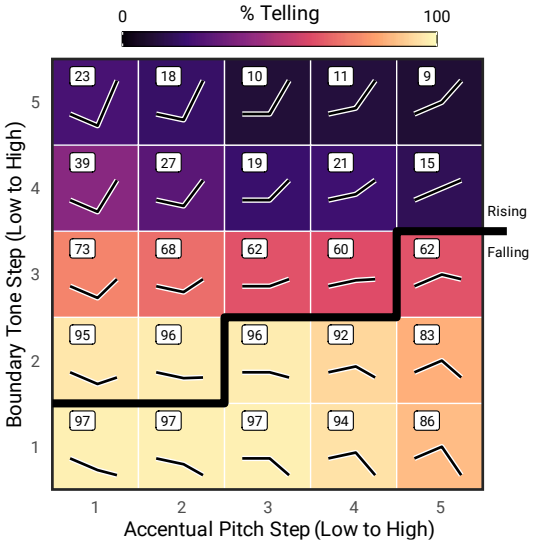

Figure 2: Heatmap of proportion *telling* responses for Exp 1. Schematic F0 contours of continuum steps are shown in each cell, depicting the onglide to the accentual peak across the first syllable and the final rise/fall across the second syllable. Proportions are labelled within each cell.

Our statistical model of the probability of telling responses for Exp. 1 shows a **negative** effect of ending pitch ($\hat{\beta}$ = -0.61, 95% CI [-0.69,-0.53]) and a **weaker negative** effect of accentual pitch ($\hat{\beta}$ = -0.15, CI [-0.23,-0.06]). The model shows no credible evidence of an interaction (CI [-0.01, 0.02]). Overall, there is a bias for *telling* responses, shown most clearly in the middle row of Fig. 2, where responses are closest to chance but still lean towards *telling* responses. This bias is also reflected by a **positive** intercept ($\hat{\beta}$ = 1.98, CI [1.26,2.73]).

2.2. Experiment 1 Discussion

Overall, we find that the accentual pitch and ending pitch are differently weighted as cues to a *telling* interpretation, with the ending pitch cue being four times larger in magnitude than the accentual pitch cue. We also replicate the results of [5], shown by our results in the third column of Fig. 2: steeper rising slopes are more likely to receive inquisitive interpretations. However, rising steps with near-equal pitch excursions, and therefore equal slopes (referencing cells by column-row in Fig. 2, cells 1-3, 3-4, 5-5), differ substantially in the proportion of *telling* responses. Therefore, slope alone, as defined from the pitch accent to the boundary tone, does not capture the range of variation in responses. This same observation suggests that there **is** an effect of pitch accent—yet not in the direction predicted by [5]. This result is counterintuitive: we do not expect **more** *asking* responses as the accentual pitch becomes more like H*, nor do we expect **fewer** *asking* responses with the steepest rise.

The top and bottom rows of Fig. 2 highlight an interesting property: pitch contours that are **overall** "more high" are more likely to receive inquisitive interpretations. This notion of overall highness is captured by the Tonal Center of Gravity (TCoG, [7]), a metric originally proposed to model variation in onglide shapes of bitonal pitch accents. Broadly, TCoG is a 2-dimensional point in the **time** domain (TCoG-T), showing where in time the bulk of high F0 lies, and the **frequency** domain (TCoG-F), showing the overall pitch across a stretch of time. TCoG is calculated by the weighted average of values in one domain weighted by their respective values in the complementary domain (e.g., TCoG-F = Hz values weighted by their timestamps).

In the context of our data and materials shown in Fig. 2, TCoG-F increases both as accentual pitch increases (from left to right in Fig. 2) and as ending pitch increases (from bottom to top). Accordingly, the likelihood of *telling* responses decreases in both directions, hence TCoG-F may provide a univariate lens to better explain the data from Exp. 1. A TCoG perspective also provides an opportunity to return to the simplifying assumption we made in our materials for Exp. 1 where our falling steps are implemented linearly between the accentual/ending pitch targets.

3. EXPERIMENT 2

When making the materials for Exp. 1, we used linear rises **and** falls (henceforth, "gradual" falls), building on [5] to allow for a shared interpretation of slope across both types of contours. However, in the natural recordings using H*L-L% intonation, pitch fell rapidly from the accentual peak, reaching the pitch floor at, on average, 30% of the duration of the second syllable (henceforth, "early" falls). Crucially, TCoG-F for early falls is always lower than that of their gradual counterparts. Based on the results from Exp. 1, the proportion of *telling* responses increases as TCoG-F decreases, and so early falls should be more likely than their gradual counterparts to receive assertive interpretations.

Source files for resynthesis and the continua pitch targets are the same as those described in Exp. 1. However, for falling steps (where ending pitch target < accentual pitch target) we add an additional target at 30% of the duration of the second syllable of the nuclear word. This target uses the same pitch value as the ending pitch target. Rising steps are unchanged. These differences can be seen in the falling steps of the continuum shown in Fig. 3. The paradigm is the same as in Exp. 1, but we added a trial time-out after 8 seconds to improve participant engagement.

3.2. Experiment 2 Results

Fig. 3 shows the heatmap of responses from 54 new participants and the schematic contours for Exp. 2.

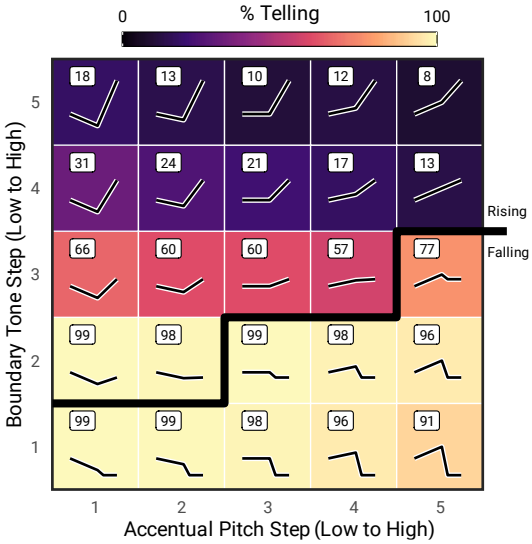

Figure 3: Proportion *telling* responses for Exp. 2 with schematic continuum steps shown in each cell.

We model the results of Exp. 2 using the same statistical model (the scaling model) as in Exp. 1. Again, we find a negative effect of ending pitch ($\hat{\beta}$ = -0.77, CI [-0.88,-0.66]). Importantly, we now find **no credible effect of accentual pitch** ($\hat{\beta}$ = -0.05, CI [-0.13, 0.03]). As predicted, this reduction in the magnitude of the accentual pitch effect is most evident for the falling steps in the lower right quadrant of Fig. 3. Again, we find no credible interaction of accentual and ending pitch. We turn now to modelling our results in terms of TCoG; univariate identification curves are shown in Fig. 4.

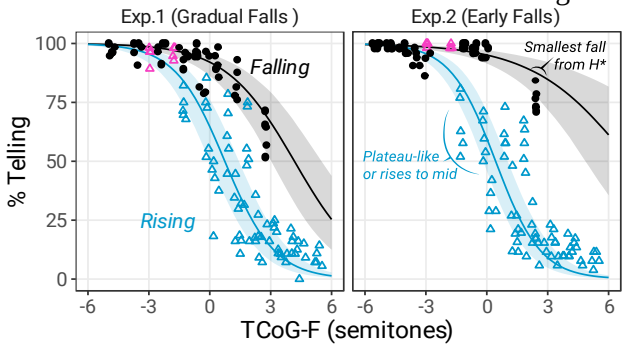

Figure 4: Identification curves with 95% credible interval for TCoG-F, in semitones from 90Hz. Each point is a single utterance from a single rising (blue triangles) or falling (black circles) contour of the 5x5 continuum. The shallowest rises are highlighted in pink.

In Fig. 4, the proportion of *telling* responses for all falling steps (except for cell 5-3 in Fig. 3) are close to ceiling. The scaling model fits the data better compared to a model with TCoG-F as the sole predictor, but it also has more predictors with which to capture variation. Adding a categorical predictor of rising vs. falling shape (coded as +/- .5) and its interaction with TCoG-F allows the identification function to change depending on global tune shape (as in Fig. 4), giving the TCoG-F model a comparable number of parameters to the scaling model. This model shows the expected negative effect of TCoG-F in both experiments (Exp.1 CI: [-0.80,-0.61], Exp2: [-0.77,-0.55]) but a larger difference between the rising/falling groups with the early fall manipulation (Exp.1: [-2.03,-1.67], Exp.2: [-2.83,-2.33]).

Notably, the rising steps with the lowest TCoG-F are the shallow rises (cells 1-2 & 2-2), overlapping with the falling steps' distribution of *telling* response probabilities (in pink at the top of Fig. 4). The interpretation of the most ambiguous steps (row 3) is at chance, with a slight bias towards an assertion interpretation. This bias is likely stems from the declarative syntax and the fact that utterances were presented in isolation with no context. The most ambiguous steps are also most like plateau intonation (ToBI: H*H-L%) which is often used for listing rather than making assertions or questions [2].

4. CONCLUSIONS

In this work we investigated what role, if any, the pitch accent may play in the interpretation of falls and rises within the context of rising declaratives. In Exp. 1 we find a counterintuitive effect of pitch accent when using a simplified implementation of falling steps. A TCoG perspective helps make sense of this finding, predicting that more natural early falls would lower TCoG-F and increase assertion interpretations. This manipulation in Exp. 2 nearly eliminates the previous effect of pitch accent. The TCoG-F model shows a sigmoidal response pattern but requires a supplementary parameter distinguishing the rising/falling shape of the contour to approach performance of the scaling model using accentual and ending pitch. Overall, our results do not support a phonological distinction between inquisitive and assertive rising declaratives based on pitch accent. A phonological distinction may perhaps be maintained by ascribing the contrast to different edge-tone configurations, i.e., L-H% vs. H-H%. Yet, the TCoG results also suggest that the distinction may arise from a probabilistic relation to phonetic gradience such that rising intonation that is overall "more low," as captured by TCoG, is more likely to receive an assertion interpretation. Uncertainty in the interpretation of ambiguous steps points towards the role of context in guiding listeners towards one interpretation over another as well as what other interpretations are available.

5. ACKNOWLEDGMENTS

Thanks to the ProSD Lab at Northwestern and Chun Chan for the experiment implementation. JC's work is supported by NSF BCS-1944773.

5. REFERENCES

[1] Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*.
[2] Pierrehumbert, J. and Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, vol. 271.
[3] Rudin, D. 2022. Intonational Commitments. *Journal of Semantics* 39(2), 339-383.
[4] Farkas, D., Roelofsen, F. 2017. Division of Labor in the Interpretation of Declaratives and Interrogatives. *Journal of Semantics* 34(2), 237-289.
[5] Jeong, S. 2018. Intonation and Sentence Type Conventions: Two types of Rising Declaratives. *Journal of Semantics* 35(2), 305-356.
[6] Barnes, J., Veilleux, N., Brugos, A., and Shattuck-Hugnagel, S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2), 337-383.
[7] Schiefer, L., Batliner, A. 1991. A ramble round the order effect. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München.*
[8] Steffman, J., Shattuck-Hufnagel, S., Cole, J. 2022. The rise and fall of American English pitch accents: Evidence from an imitation study of rising nuclear tunes. *Proc. Speech Prosody 2022* Lisbon, 857-861.
[9] Boersma, P., Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.2.