

Pitch Accent Variation and the Interpretation of Rising and Falling Intonation in American English

Thomas Sostarics¹, Jennifer Cole¹

¹Northwestern University, Department of Linguistics

tsostarics@u.northwestern.edu, jennifer.cole1@northwestern.edu

Abstract

This study tests the division of labor in the meaning conveyed by pitch accents and edge tones in English intonation. In three perception studies, we investigate where the locus of the contrast between an assertive vs inquisitive interpretation resides. By doing so, we also gain insight into the role of potentially meaningful within- and between-category variation in the phonetic implementation of discrete intonational tunes. We find that the pitch accent does not contribute to assertive interpretation. Rather, the distinction between assertive and inquisitive interpretation is cued primarily by the final F0 of the pitch contour regardless of the pitch accent, but that increased overall pitch prominence may trigger a salient focus interpretation that interferes with judging assertiveness.

Index Terms: rising declaratives, speech perception, compositionality, intonational meaning, prosody

1. Introduction

A well-known property of English is the encoding of pragmatic speech act meaning in the pitch pattern at the end of a prosodic phrase. A declarative sentence with final falling pitch generally conveys an assertion, while a final rising pitch trajectory typically conveys a question. The critical region for these meaning distinctions extends from the rightmost pitch accent, located on the word with phrasal (nuclear) stress, to the end of the phrase. The pitch contour across this region of the phrase is the phonetic implementation of an abstract, phonologically specified representation called the **nuclear tune**.

The dominant Autosegmental-Metrical (AM) theory [1,2] analyzes these tunes as comprising three intonational components based on the atomic tone features H(igh) and L(ow): a pitch accent (H*, L*, L+H* among others) which associates with the stressed syllable of the accented word, and the edge tones, comprised of a phrase accent (H-/L-) and boundary tone (H%/L%), which associate with the right edges of lower- and higher-level prosodic phrases. Specifying the nuclear tune as the concatenation of smaller atomic units raises the question of whether meaning distinctions between tunes are related to the tunes as a whole [3] or are more narrowly associated with any of the individual intonational components that make up the tune [4]. While compositional accounts vary, pitch accents are generally taken to contribute referential meaning related to information status (new vs given) and focus [4,5], while edge tones convey speech act meaning, analyzed, e.g., in terms of speaker commitment towards a proposition within a broader discourse [6,7]. But this division of labor breaks down in recent work on the intonational encoding of the question/assertion (Q/A) distinction in rising declaratives. In a study of intonational meaning, [8] finds that a rising declarative implemented with a shallow rise is more likely to yield assertive

interpretations compared to a steeper rise. This difference was attributed to a pitch accent contrast, suggesting that pitch accents contribute to the Q/A meaning distinction.

In this paper we test competing accounts of the locus of intonational encoding of the Q/A contrast in Mainstream US English (MUSE). We examine both rising and falling tunes to determine which part of the pitch contour encodes this aspect of meaning: is it the region spanning the pitch accent or the region of the edge tones? Furthermore, in light of the pervasive variation in intonational form [9], we also investigate the degree to which variation in the phonetic implementation of the pitch accent and/or edge tones influences listener interpretation. We present three experiments probing these questions about the locus of tune meaning. Exp. 1 tests the putative roles of the monotonal H* and L* pitch accents in rises and falls while manipulating the phonetic scaling of both pitch accent and edge tones individually and in relation to one another. Exp. 2 looks beyond the monotonal pitch accents to investigate the role of variation in F0 peak scaling of the bitonal L+H* pitch accent. Finally, Exp. 3 investigates the role of F0 peak alignment spanning rising accents with an early peak (L+H*) and late peak (L*+H). While we reference labels from the ToBI annotation system [10], the findings from this study speak to the larger question of whether pragmatic meaning distinctions are encoded by contrastive phonological categories (e.g., H vs. L edge tones), or by continuous variation of F0.

2. Experiment 1: Monotonal Accents

We use a two-alternative forced choice task similar to that used in [8] but we extend the phonetic continuum of the test materials to investigate how variation in different parts of the nuclear tune pitch contour affect listeners' interpretation of the Q/A contrast. Participants are instructed that they will be listening to sentences said in different ways and judging whether the speaker is **telling** them something or **asking** them something. On each trial, a declarative sentence such as *Molly's from Branning* with varied intonation is played auditorily over headphones. Based on the pitch contour, participants respond with either Asking or Telling using the F and J keys on their keyboard. Between trials, participants are tasked to count aloud by 2s for ~4 seconds, which was included to avoid order effects [11] from participants comparing the slopes of pitch contours between trials. Participants listen to five repetitions of stimuli from a 5x5 continuum, described below, where each repetition is instantiated by one of five declarative sentences of the form *Name's [preposition/determiner] Noun*, with a final disyllabic noun with initial stress that contains only voiced segments. We recruited 52 participants via Prolific for Exp. 1.

Our hypothesis follows the leading compositional approaches like [4,6] wherein the Q/A contrast is encoded solely by the edge-tone configuration of the nuclear tune. This hypothesis

predicts a higher % Telling responses for pitch contours that end in a lower pitch value (canonical falling contours, with edge tones L-L%) and a lower % Telling responses for contours that end in a higher pitch (canonical rising contours, with edge tones H-H%). This hypothesis does not predict any meaningful change in response behaviour as the pitch accent target changes. An alternative account like [8] posits that the pitch accent plays a role in interpretation such that a higher pitch in the stressed syllable (H*) contributes to assertive force beyond the contribution of the ending pitch in both rising **and** falling tunes. This account would predict higher % Telling responses for contours with higher accentual peak F0, such that high-rising (H*H-H%) contours will be more likely to receive an assertive interpretation compared to the more canonical rising contours (L*H-H%) typically used for inquisitive rising declaratives, which rise from a low pitch accent target. Similarly, % Telling responses would be highest for canonically falling intonation (H*L-L%) that falls from a high accentual peak F0.

We create a 5x5 phonetic continuum crossing the peak F0 of the pitch accent, spanning from L* to H* F0 targets (henceforth **accentual pitch**) and the final F0 of the edge tones, spanning from L-L% to H-H% targets (henceforth **ending pitch**). The pitch accent target is temporally aligned at the end of the stressed syllable and varies along the accentual pitch continuum in equally spaced steps from 70Hz (L*) to 110Hz (H*), using values based on the first author’s natural productions of H*L-L% and L*H-H%. The ending pitch continuum uses equally spaced ERB-scale differentials from -0.25 ERBs to +2 ERBs based on production data from [12]. These differentials are added to the lowest accentual pitch value to obtain target values between 61Hz and 149Hz, which are crossed with the accentual pitch targets for a 25-step continuum. Throughout this paper, steps are defined as **falling** when ending pitch < accentual pitch and **rising** when ending pitch > accentual pitch.

To make the falling steps sound more natural, and in line with descriptions from [13], an additional target equal to the ending pitch target value is added at 30% of the second (unstressed) syllable, which yields a more abrupt fall from the accentual peak. The syllable durations of the final word for each of the five stimulus utterances were normalized based on the averages of the first author’s original recordings. To compensate for phrase-final lengthening, the second syllable was reduced to 70% of the average duration, which helped to make the resynthesized pitch contours sound more natural to our ears. All duration and pitch resynthesis was done in Praat [14] and ensured that the resynthesized pitch contours, shown in Fig. 1, were comparable across items.

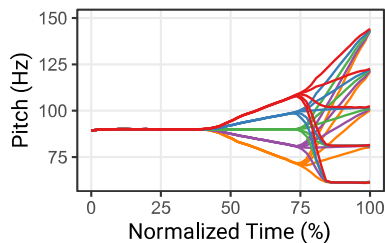


Figure 1: Pitch contours averaged across all utterances. The prenuclear region is held flat at 90Hz.

We model the % Telling responses, indicating an assertive interpretation, using a Bayesian logistic mixed effect model with the brms package in R [15,16] using predictors of accentual pitch, ending pitch, and their interaction. We used a

random effects structure of random intercepts by participant and utterance and random slopes of accentual pitch, ending pitch, and their interaction by participant. Predictors are transformed to semitones from the midpoint of the accentual pitch continuum (90Hz). Fig. 2 shows the % Telling response for each step of the continuum as a bivariate heatmap. Evidence for an effect of ending pitch would be shown by vertical variation of cell color across the rows of the heatmap, while evidence for an effect of accentual pitch would be shown by horizontal variation of color, across the columns. All materials, data, and analyses are available at <https://osf.io/8hrfv/>.

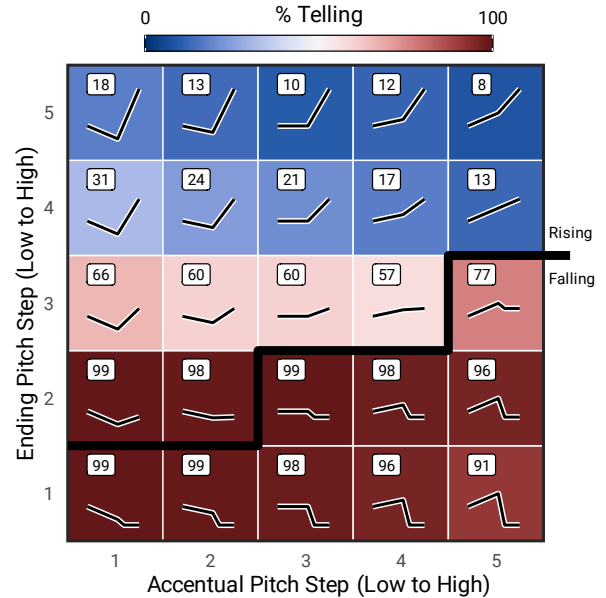


Figure 2: Heatmap of % Telling responses for Exp. 1. Schematic depictions of the nuclear pitch contours (e.g., F0 across “Branning”) are shown in each cell. The accentual pitch target occurs at roughly the midpoint of the schematized contours (cf., Fig. 1).

Our statistical model reveals the predicted credible negative main effect of ending pitch ($\hat{\beta} = -0.76$, 95% CI [-0.87,-0.65]) but no credible main effect of accentual pitch ($\hat{\beta} = -0.05$, CI [-0.13, 0.03]) nor their interaction ($\hat{\beta} = 0$, CI [-0.02, 0.03]). These results suggest that interpretation of the Q/A contrast is driven primarily by variation in ending pitch and pitch accent does not contribute to assertive interpretations. Yet, there are a few cells that merit more specific discussion.

Denoting cells in Fig. 2 by column-row indices, cells 1-2 and 2-2 are rising steps (ending pitch > accentual pitch), but likely do not rise *enough* to be perceived as a rising pitch movement (indeed these movements are below the ~4st threshold reported by [11]). Cells in row 3 are closest to chance. Cell 5-1 is closest to canonical falling intonation (H*L-L%) yet this pitch contour elicits a numerically lower % Telling responses than other steps of the same ending pitch (same row); we’ll return to this point in the general discussion based on results from Exp. 2. Finally, comparing the cells in rows 5 and 4 from left to right shows that contours that rise from a high accentual pitch target (i.e., are **globally** higher) were more likely to receive an inquisitive interpretation than contours that rise from a lower target. This finding is counter to the proposal [8] that H*H-H% is more assertive than L*H-H% and will be revisited in the general discussion.

Exp. 1 tested whether interpretation of the Q/A contrast is driven by variation in the pitch accent or edge tones in falling and rising intonation. We found a strong effect of edge tones (high ending pitch: H-H%, or low ending pitch: L-L%) but no evidence for an effect of pitch accent (scaling of H* or L*). Moreover, rising contours that were globally higher were more likely to receive an inquisitive interpretation. Yet, this does not exhaust the full inventory of pitch accents in MUSE. The rising bitonal pitch accents L+H* and L*+H are frequently discussed in relation to prosodic focus [17,18] and rise-fall-rise intonation [19], but it is unclear whether the focus-related contributions they make impact interpretation of the Q/A contrast. If the focus-related meaning encoded by the bitonal pitch accents is orthogonal to the Q/A contrast encoded by the edge-tones, then tunes using a bitonal pitch accent should yield results comparable to those of Exp. 1. Alternatively, it may be that semantic alternatives invoked by prosodic focus may affect participants’ interpretation of rising and falling intonation. Exp. 2 investigates whether, and to what degree, the L+H* pitch accent affects interpretation of the Q/A contrast.

3. Experiment 2: L+H* Scaling

We use the same task setup from Exp. 1 but change the accentual pitch targets used in the continuum to define pitch contours that rise from an initial accentual low pitch target. The ending pitch targets remain the same as in Exp. 1. We add a low pitch target at 70 Hz aligned with the start of the first syllable in the nuclear accented word and shift the previous accentual pitch continuum up from 70-110Hz to 80-120Hz to ensure that there is a rising movement from the initial low target to the accentual peak F0 for all steps on the accent continuum. As the L+H* pitch accent has been described as having a “domed” onglide shape [20], we use Bezier curves to create curved onglides for the pitch resynthesis, shown in Fig. 3. We recruited 55 new participants from Prolific for Exp. 2.

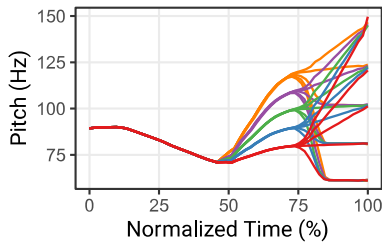


Figure 3: Exp. 2 stimuli with curved onglides.

There are two main observations from the results shown in Fig. 4 compared to the results of Exp. 1. First, like in Exp. 1, the % Telling responses for the falling steps of the continuum show a similar decreasing pattern when moving rightward and downward towards cell 5-1. Second, the % Telling responses for rows 4 and 5 for Exp. 2 are overall higher than rows 4 and 5 for Exp. 1, though still favor an Asking response. As in Exp. 1, rising steps in row 3 are closest to chance.

We present two perspectives of this data. First, we fit the same model used in Exp. 1 to the data from Exp. 2 but omit the interaction term as it did not return a credible effect for Exp. 1 and did not improve model fit for Exp. 2 when included. Second, we compare between the two experiments by pooling the data from both experiments and adding a predictor of experiment and its interactions with the other predictors. Experiment is a categorical predictor with Exp.2 (coded +.5) compared to Exp.1 (-.5). We are interested in whether the effect

of ending pitch is lower in Exp. 2 compared to Exp. 1 as shown by a positive interaction of experiment and ending pitch.

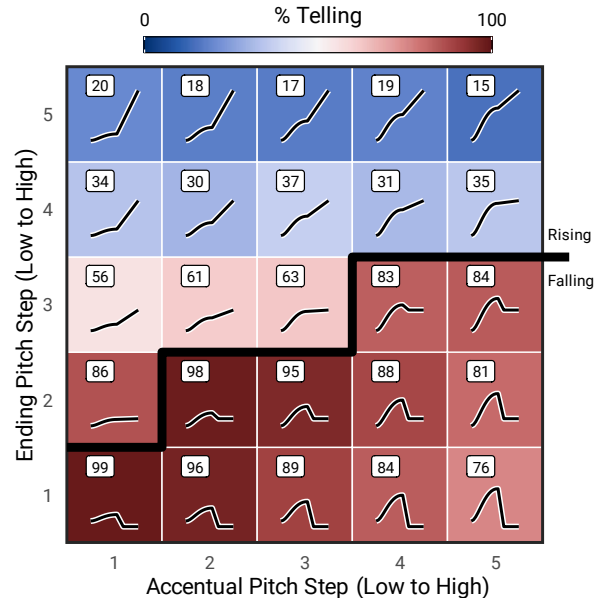


Figure 4: Heatmap of % Telling responses for Exp. 2.

The first model shows a credible main effect of ending pitch ($\hat{\beta} = -0.44$, CI [-0.51,-0.37]) and no credible effect of accentual pitch ($\hat{\beta} = 0.05$, CI [-0.02,0.12]). Yet, there still seems to be substantial horizontal gradation within the falling steps of the continuum more broadly. A post-hoc analysis modeling the effect of accentual pitch within the rising and falling groups does reveal a significant conditional effect of accentual pitch for falling steps ($\hat{\beta} = -0.31$, CI [-0.39,-0.23]), reflecting the decrease in % Telling as the pitch accent is scaled higher. Our final model, comparing Exp. 1 and 2, shows a credible positive interaction between ending pitch and experiment ($\hat{\beta} = 0.20$, CI [0.08,0.33]), in line with the observation that some steps moved closer towards chance compared to Exp. 1.

In Exp. 2 we found that using the bitonal L+H* pitch accent made rising steps (rows 4 & 5) more likely to receive an assertive interpretation compared to the rising steps of Exp. 1. Additionally, higher scaling of the accentual pitch decreased % Telling responses within the falling steps of the continuum. Given the link between the L+H* pitch accent and prosodic focus, these findings suggest an interaction between focus marking and rising/falling intonation, with an interpretation that is perhaps distinct from the asking/telling responses we provide our participants with. For our final experiment, we build on this finding and test whether an earlier-aligned L+H* affects interpretation differently from the later-aligned L*+H.

4. Experiment 3: Bitonal Accent Alignment

Exp. 3 manipulates accentual peak alignment while holding the accentual pitch target at the highest value from Exp. 2 (120Hz). We again use Bezier curves to create domed onglides for earlier steps and scooped onglides for later steps, following [20]. Peak alignment varied in equally spaced steps from 80% of the stressed syllable’s duration for the early-aligned L+H* accent to 115% (slightly into the second syllable) for the late-aligned L*+H accent, shown in Fig. 5. The task is the same as in the previous experiments. The results are shown in Fig. 6 and

analyzed with a model like that used for Exp. 1, but now the predictor of accentual pitch is replaced by a predictor of peak alignment centered at 100% of the stressed syllable. We recruited 58 new participants from Prolific for Exp. 3.

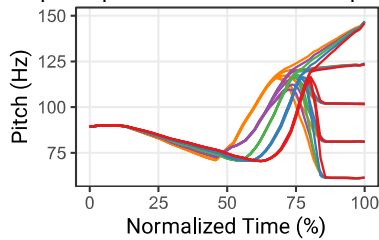


Figure 5: *Bitonal alignment continuum.*

We again find a credible main effect of ending pitch ($\hat{\beta} = -0.54$, CI [-0.64,-0.45]) and no main effect of alignment ($\hat{\beta} = -0.32$, CI [-0.97,0.36]) but there is a credible interaction between the two ($\hat{\beta} = -0.67$, CI [-0.88,-0.47]), which is most strongly shown by the diagonal gradation from cell 1-4 towards cell 5-5 in Fig. 6.

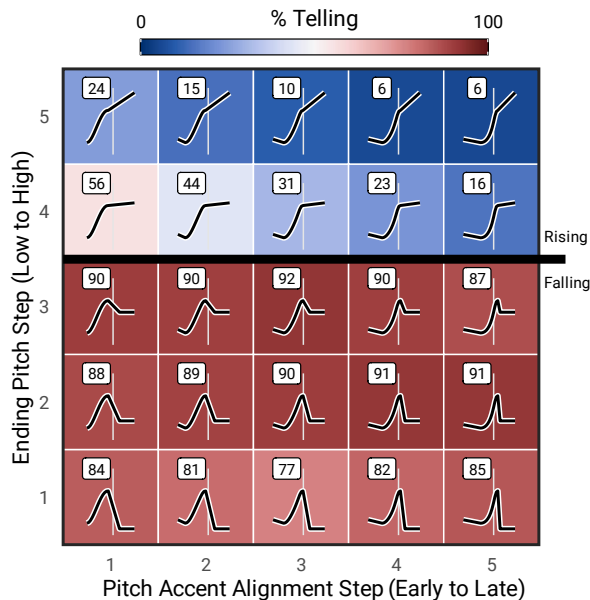


Figure 6: *Heatmap of % Telling responses for Exp. 3; vertical line shows accent peak alignment relative to the end of the stressed syllable.*

Beyond the broad pattern relating % Telling responses to variation in ending pitch, the pattern of variation within the rising steps merits further discussion. First, as in the previous experiments, the contours that are closest to chance are those whose ending pitch trajectory is more of a plateau than a rise—here, in row 4. However, % Telling responses decreases in row 4 as alignment of the pitch accent becomes later, suggesting an interaction with the **duration** of this plateau. That is, the plateau, and the canonical “listing” function it conveys [21], may only be perceptible when sustained for a long enough time. The flat portion of cell 5-4 may not be long enough to be distinguishable from the contour’s overall rising trajectory. Second, the top-right quadrant shows the lowest % Telling responses across all three experiments—far lower than even our canonical step rises in Exp. 1 (Fig.2 cell 5-1): 6% vs 18%. This comparison suggests either (1) that the alignment of the low pitch accent target in canonical rises for inquisitive rising

declaratives should be aligned at a point earlier than the end of the stressed syllable (c.f. as in Exp. 1) or (2) that L* needs more of a sustained low region to be perceptible. This latter account is supported by the lower onglide into the low accentual pitch target in Exp. 3 and in line with findings that L*, unlike H*, requires compensation in other suprasegmental cues to boost its perceptibility [22].

5. General Discussion

We hypothesized that the locus of the Q/A contrast resides solely in the edge-tone configuration of the nuclear tune. This hypothesis was supported by the robust effect of ending pitch in all of our experiments, which however differed in the pitch accent context that the edge-tone configuration co-occurred with. The results of Exp. 1 & 2 suggest that while the pitch accent does not **contribute** to assertive force, as % Telling decreased rather than increased with higher scaled accentual pitch, it may nonetheless **interfere** with participants’ interpretation of the Q/A contrast. One account for this is that prosodic focus is conveyed not by the categorical use of the L+H* pitch accent rather than H*, but by higher overall pitch **prominence** [23]. This account would predict that as prominence increases, the likelihood of a contrastive or corrective focus interpretation would also increase for the falling steps, and participants may interpret such contours as distinct from Telling. For instance, we did not give participants a “correction” response option, but that interpretation may have interfered with their selection of a Telling response in such cases. This same focus account also helps to explain why rising steps in Exp. 2, which had overall higher pitch prominence, were less likely to receive an inquisitive interpretation compared to Exp. 1. A full pragmatic account of focus with rising declaratives, however, is beyond the scope of this paper.

We presented three experiments probing the interpretation of inquisitiveness/assertiveness in rising and falling declaratives amidst a greater degree of phonetic and phonological variation than has been previously investigated by crossing phonetic continua between and within pitch accent and edge tone categories. Across all experiments, a robust effect of ending pitch was found, supporting the hypothesis that the locus of the question/assertion contrast resides solely in the edge tones, with a primary dichotomous contrast between tunes with a rising vs. falling trajectory, and secondary gradient distinctions dependent on the vertical scaling of the ending F0. However, we find curious results suggesting that focus, as cued by accentual prominence, may be salient to participants and affects their capacity to judge along the inquisitive/assertive dimension of meaning. Moreover, variation in the shape of rises and how they interact with duration (as in Exp.3) gives new insight into the cues listeners use when interpreting intonation, though more work on the effects of duration, loudness, and voice quality is needed. One limitation is that our stimuli are presented without contexts that would license the use of rising declaratives as well as focus. Thus, we cannot disentangle whether participants infer meaning **beyond** the asking/telling options we provide them or whether the interpretation of some signals is truly **ambiguous**. Understanding the role of focus in our findings and the extent of the meaning space is left to future work.

6. Acknowledgements

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

7. References

- [1] J. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, Dept. Linguistics, MIT, Cambridge, MA, 1980.
- [2] M. Beckman and J. Pierrehumbert, "Intonational structure in Japanese and English." *Phonology*, vol. 3, pp. 255-309, 1986.
- [3] M. Liberman and I. Sag, "Prosodic form and discourse function," in *Chicago Linguistics Society*, vol. 10, 1974. pp. 416-427.
- [4] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack, Eds. Massachusetts: MIT Press, 1990, pp. 271-311.
- [5] M. Westera, D. Goodhue, and C. Gussenhoven, "Meanings of Tones and Tunes," in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds. Oxford: Oxford Academic Press, 2020, pp. 443-453.
- [6] C. Bartels, *The intonation of English statements and questions: A compositional interpretation*. Abingdon: Routledge, 1997.
- [7] C. Gunlogson, "A question of commitment," *Belgian Journal of Linguistics*, vol. 22(1), pp. 101-136, 2008.
- [8] S. Jeong, "Intonation and sentence type conventions: Two types of rising declaratives," *Journal of Semantics*, vol. 35(2), pp. 305-356, 2018.
- [9] J. Cole, "Prosody in context: A review," *Language, Cognition and Neuroscience*, vol. 30(1-2), pp. 1-31, 2015.
- [10] N. Veilleux, S. Shattuck-Hufnagel, and A. Brugos. *Transcribing Prosodic Structure of Spoken Utterances with ToBI*. January IAP 2006. Massachusetts Institute of Technology: MIT OpenCourseWare, <https://ocw.mit.edu>.
- [11] L. Schiefer and A. Batliner, "A ramble round the order effect," *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*. 1991.
- [12] J. Steffman, S. Shattuck-Hufnagel, J. Cole, "The rise and fall of American English pitch accents: Evidence from an imitation study of rising nuclear tunes," in *Proc. Speech Prosody 2022, Lisbon, Portugal, 2022*.
- [13] J. Pierrehumbert. "Synthesizing intonation," *The Journal of the Acoustical Society of America*, vol. 70(4), pp. 695-701, 1981.
- [14] P. Boersma and D. Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.2. 2022.
- [15] P.C. Bürkner, "brms: An R Package for Bayesian Multilevel Models Using Stan," *Journal of Statistical Software*, vol. 80(1), pp. 1-28. 2017.
- [16] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. URL <https://www.R-project.org/>.
- [17] M. Husband and F. Ferreira. "The role of selection in the comprehension of focus alternatives," *Language, Cognition and Neuroscience*, vol. 31(2), pp. 217-235, 2016.
- [18] B. Braun and L. Tagliapietra. "The role of contrastive intonation contours in the retrieval of contextual alternatives." *Language and Cognitive Processes* vol. 25(7-9), pp. 1024-1043, 2010.
- [19] A. Göbel, "Additives pitching in: L*+ H signals ordered Focus alternatives," in *Semantics and Linguistic Theory*, vol. 29, pp. 279-299, 2019.
- [20] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel. "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology*, vol. 3(2), pp. 337-383, 2012.
- [21] R.S. Burdin and J. Tyler. "Rises inform, and plateaus remind: Exploring the epistemic meanings of 'list intonation' in American English," *Journal of Pragmatics*, vol. 136, pp. 97-114, 2018.
- [22] H. Jeon and A. Heinrich, "Perceptual asymmetry between pitch peaks and valleys," *Speech Communication*, vol. 140, pp. 109-127, 2022.
- [23] D. Watson, M. Tanenhaus, and C. Gunlogson. "Interpreting pitch accents in online comprehension: H* vs. L+ H," *Cognitive science*, vol. 32(7), pp. 1232-1244, 2008.