F0 correlates of perceived speaker surprise in American English: Accents vs. Edge Tones

Rebekah Stanhope[1], *Thomas Sostarics*[1], *Jennifer Cole*[1]

[1]Department of Linguistics, Northwestern University, USA

rstanhope@u.northwestern.edu, tsostarics@gmail.com, jennifer.cole1@northwestern.edu

Abstract

Previous work has proposed that the L+H* pitch accent is the primary encoding of speaker surprise in English. However, studies in other Germanic languages have shown that perceived surprise may also be influenced by the scaling of L* and H* pitch accents, and by the scaling of the edge tones. This study tests whether Mainstream American English listeners' perceptions of speaker surprise are sensitive to variation in L* and H* pitch accents, and whether perceived surprise is driven solely by the pitch accent or also influenced by the edge tones. We find that both the pitch accent and edge tones contribute to perceived speaker surprise in MAE. We explore an alternative analysis in terms of pitch span and pitch level, finding that while both phonetic parameters contribute to perceived surprise, neither fully account for the pattern of ratings in our data.

Index Terms: intonational meaning, speech perception

1. Introduction

Previous work on intonational meaning in Mainstream American English (MAE) has noted that variation in F0 contours can encode *mirativity*: a speaker's surprise regarding the proposition expressed by their utterance [1, 2, 3, 4]. However, proposals vary in which intonational feature(s) they attribute this meaning to, with one approach isolating a single intonational feature within the inventory posited by Autosegmental-Metrical (AM) theory [5] for MAE [2], while others invoke multiple intonational features, suggesting that cues to speaker surprise are distributed across the utterance [4, 3]. A further difference is whether interpretations of speaker surprise are attributed primarily to category-level distinctions in tonal specification [2, 4] or to within-category variation in pitch span (the distance between F0 targets) or pitch level within a tune [6].[1]

Rett and Sturman [2] propose that mirativity is encoded phonologically by the pitch accent type; based on a small-scale production study, they suggest that L+H* acts as a categorical marker of mirativity, while additional gradient dimensions of meaning (e.g., a speaker's **degree** of surprise) might be conveyed by variation in the scaling of the pitch accent peak within the L+H* category. Under this phonological account, other high-tone accents, e.g., H* or L*+H, do not encode mirativity, meaning that variation in their scaling is not expected to affect listeners' interpretations of speaker surprise. Furthermore, the effect of pitch accent in encoding mirativity is expected to be independent of the specification of other intonational features such as the phrase accent and boundary tone. Thus, variation in the scaling of these edge tones should similarly have no effect on interpretations of speaker surprise.

Rett and Sturman's pitch accent account of mirativity faces two challenges. First, the distinction between accents analyzed in AM theory as H* and L+H* has been subject to debate, with an alternative account cast in terms of gradience within a single high pitch accent category [7, 8], the endpoints of which differ primarily in the scaling and alignment of the pitch accent peak. Given the role of peak scaling suggested in [2], we might expect, under the single-category view, that some accents analyzed as H* in AM might also convey speaker surprise, with the perceived degree of surprise increasing with the height of the pitch accent peak. Second, perception studies in other Germanic languages have suggested that speaker surprise can be conveyed by expanded pitch range over an entire tune [9, 10],[2] indicating that in those languages, mirativity is effectively distributed over multiple pitch targets associated with distinct intonational features, which jointly serve to cue speaker surprise. For instance, in a surprise rating study with Dutch rising nuclear tunes (H*H-H% and L*H-H%), the scaling of both the pitch accent and boundary tone targets influenced Dutch listeners' judgments of speaker surprise [9], though the effect differed by tune. For the H*H-H% tune, ratings of speaker surprise were highest when both the pitch accent and boundary tone targets were at their highest scaling—i.e., when the overall pitch level of the tune was highest. However, for L*H-H%, ratings were highest when the pitch accent scaling was at its lowest and boundary tone scaling was at its highest—i.e., when the pitch span was largest. These findings suggest that to obtain a full picture of mirative intonation in MAE, there is a need to consider the potential role of the edge tones and to obtain evidence from multiple tunes.

In this paper we test the predictions of the Rett and Sturman account for MAE, investigating MAE listeners' perceptions of speaker surprise in relation to variation in the phrase-final (nuclear) intonation of the utterance, which extends from the rightmost pitch accent to the end of the intonational phrase. We examine both rising and falling tunes to compare the relative contributions of pitch accent and edge tone scaling, asking whether F0 targets which cue these intonational features influence perceived surprise. We investigate monotonal pitch accents with F0 targets along a continuum spanning L* and H*, addressing the question of whether variation in pitch accents other than L+H*

[1]Specifically, [6] suggests that different degrees of surprise are conveyed by "the gradient steepness of the [phrase-final] rise". Differences in steepness can manifest as differences in both pitch span and pitch level. Holding the time over which the movement is realized constant, increasing steepness can be achieved either by lowering the starting point or raising the ending point. Both of these manipulations increase the distance between the starting and ending F0 targets, yielding higher pitch span, and raising the ending point yields higher overall pitch level.

[2]The use of pitch range to convey surprise has also been investigated in Mandarin Chinese by [11].

10.21437/TAI.2025-25

contributes to interpretations of speaker surprise. Additionally, we attempt to replicate previous findings that pitch level and pitch span influence perceived speaker surprise [9, 10].

2. Methods

We use a rating task comparable to that used in [9] (Exp. 2). On each trial, participants listened to a declarative sentence such as *Gavin's on Broadway*, with the nuclear tune on the final word, resynthesized to vary scaling of the pitch accent and edge tone targets across trials. Participants were asked to indicate how surprised the speaker sounded based on the way the sentence was said, using a 6-point scale from "not surprised" (1) to "very surprised" (6). To avoid order effects resulting from participants comparing pitch contours across trials [12], participants were instructed to count aloud by twos after each trial for a interval of 3-5 seconds. Each participant rated 25 resynthesized F0 contours (described below). Each contour was repeated five times, with each repetition instantiated by one of five different sentences,[3] resulting in a total of 125 trials.

Our stimuli (repurposed from [13], Exp. 2c) comprise a 5x5 phonetic continuum crossing the F0 of the pitch accent target (accentual pitch) with the utterance-final F0 target corresponding to the edge tones (ending pitch). The resulting pitch contours are defined in this paper as **falling** when the F0 of the ending pitch was less than that of the accentual pitch, and **rising** when the F0 of the ending pitch was greater than that of the accentual pitch. The accentual pitch continuum spans five equally-spaced F0 values from 70 Hz (L*) to 110 Hz (H*). The ending pitch continuum spans five equally-spaced ERB-scale differentials from -0.25 ERBs to +2 ERBs added to the lowest accentual pitch value, yielding F0 values ranging from 61 Hz to 149 Hz. The wider range of values for the ending pitch continuum compared to the accentual pitch continuum allows for a falling pitch excursion from the lowest accentual pitch step and a rising pitch excursion from the highest accentual pitch step. The alignment of the accentual pitch target is adjusted depending on its height in line with the covariation of these parameters in natural speech; alignment varies in equidistant steps from 30% of the duration of the stressed syllable for the lowest target (L*) to 100% of the stressed syllable (i.e., the end) for the highest target (H*). The ending pitch target is aligned with the end of the phrase-final word. To make the falling contours sound more natural, an additional pitch target equal to the ending pitch value was added at 30% of the second (unstressed) syllable in the phrase-final word. The F0 of the prenuclear region (i.e., *Gavin's on*) was held flat at the midpoint of the accentual pitch continuum (90 Hz) for all contours. The resulting continuum of 25 acoustically distinct pitch contours is shown in Fig. 1; see [13] (Ch. 2) for further information about these materials.

We recruited 55 adults aged 18-65 (mean age 41; 38 female, 17 male) from the Prolific crowdsourcing platform to participate in the experiment. Of the 55 participants, 38 identified their race as White, 10 as Black, 5 as Mixed, 1 as Asian, and 1 as Other. All participants reported that English was their first language and that they primarily resided in the United States prior to age 18, indicating familiarity with MAE; four participants reported knowing a second language and 22 reported musical experience. All participants reported having normal hearing, corrected-to-normal vision, and no speech or reading impairments.

[3]The five sentences were *Gavin's on Broadway; Molly's from Branning; Ryan's in Greenview; Megan's a grandma;* and *Joey's from Bronville.*

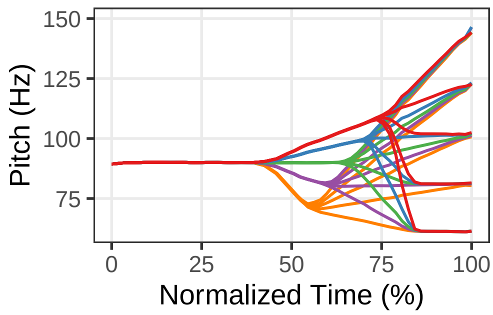

Figure 1: *25-step continuum crossing 5 accentual pitch targets with 5 ending pitch targets. Shown here are the F0 contours across the entire utterance (averaged across the five stimulus sentences). Contours plotted with the same line color share the same accentual pitch value.*

3. Results

We modeled variation in participants' surprise ratings using Bayesian mixed effects cumulative link models with a logit link function, implemented with the `brms` package in R [14, 15]. We first present our main analysis in terms of accentual and ending pitch, then a second analysis in terms of pitch span and pitch level.

3.1. Contribution of the pitch accent vs. edge tones

For our first analysis, our statistical model included predictors of accentual pitch, ending pitch, and their interaction. The random effect structure included random slopes of accentual pitch, ending pitch, and their interaction by participant, random intercepts by participant, and random intercepts by item.[4] The accentual pitch and ending pitch predictors were transformed to semitones from the midpoint of the accentual pitch continuum (90 Hz). We additionally included a discrimination (disc) parameter with a predictor of contour shape (rising or falling) which allowed the model to estimate unequal variances for rising and falling contours [16].

If the L+H* pitch accent is the only intonational feature that encodes mirativity in MAE, we would expect neither accentual pitch nor ending pitch to have a credible effect on surprise ratings since none of the contours in our stimuli included a pitch accent with both L and H targets (as in L+H*). Under the alternative hypothesis that mirativity is actually encoded by high F0 targets for pitch accent peaks, we predict a positive effect of accentual pitch. In light of the positive effect of boundary tone scaling on perceived speaker surprise in Dutch [9], we might also expect a positive effect of ending pitch.

We plot our results as a 5x5 heatmap showing the empirical mean surprise rating. In the heatmap, accentual pitch steps correspond to columns and ending pitch steps correspond to rows. Fig. 2 shows schematic depictions of how positive effects of accentual pitch and ending pitch would appear in the heatmap. Evidence for an effect of accentual pitch displays as horizontal gradation in cell color (darker shade for stronger effect), while evidence for an effect of ending pitch displays as vertical gradation. The empirical results are shown in Fig. 3. Qualitatively, we can observe that ratings are highest when accentual pitch is low and ending pitch is high (upper left cell), and lowest when both accentual and ending pitch are low (lower left cell).

[4]Where *item* refers to the sentence encountered in a trial.

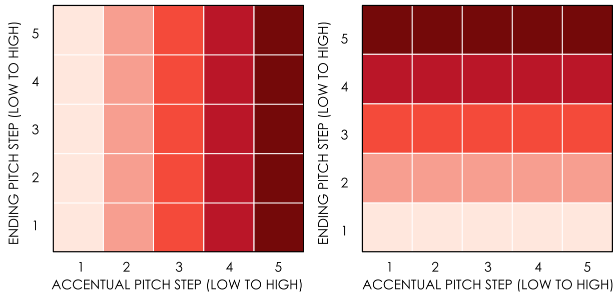

Figure 2: *Schematics showing a positive effect of accentual pitch (left) vs. a positive effect of ending pitch (right). Darker colors represent higher mean surprise ratings.*

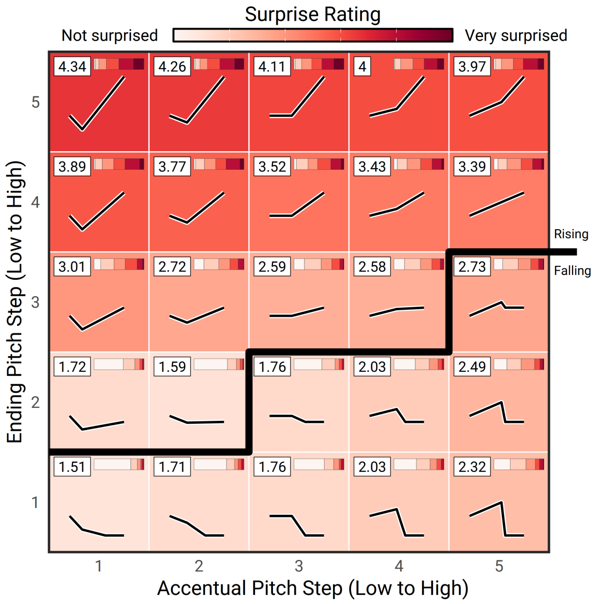

Figure 3: *Empirical mean surprise rating for each continuum step shown in color with numeric values in the top left corner. Stacked bars show the proportions of 1-6 ratings in each cell. Schematic depictions of the pitch contour over the phrase-final word are shown in each cell.*

Our statistical model revealed credible positive main effects of both accentual pitch ($\hat{\beta}$ = 0.11, 95% CrI [0.08, 0.14]) and ending pitch ($\hat{\beta}$ = 0.28, 95% CrI [0.22, 0.34]), reflecting that, overall, perceived speaker surprise tended to increase as both F0 targets increased—with the ending pitch effect having a greater magnitude. This greater effect of ending pitch is reflected in Fig. 3 by the more substantial vertical gradation in cell background color compared to horizontal gradation. However, there was also a credible interaction between these factors ($\hat{\beta}$ = -0.03, 95% CrI [-0.04, -0.02]), reflecting that, at higher ending pitch steps, increases in accentual pitch are associated with decreases in the mean surprise rating (Fig. 4).

3.2. Pitch level and pitch span analyses

A remaining question from the previous analysis is whether, and to what extent, our results can be characterized in terms of pitch level or pitch span. Recall that these measures have been previously suggested to convey a speaker's degree of surprise

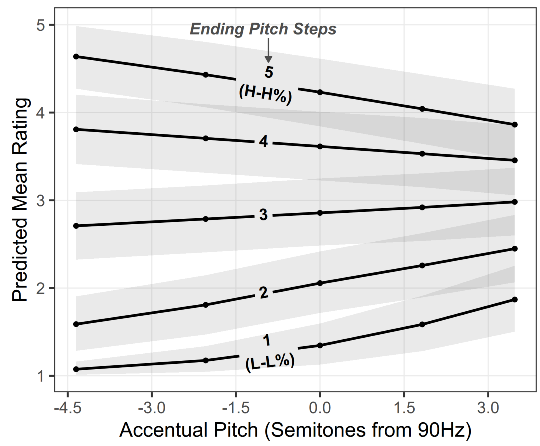

Figure 4: *Model-predicted mean ratings for each ending pitch step as accentual pitch increases. The change in slopes across ending pitch steps shows the change in the direction of the accentual pitch effect from positive at low ending pitches to negative at high ending pitches.*

in MAE [6] and have been shown to correlate with perceived speaker surprise in Dutch, with the nature of this correlation depending on the tune. To explore this question with our data, we fit separate models by operationalizing pitch level as the mean F0 in the phrase-final word and pitch span as the size of the pitch excursion from the accentual pitch target to the ending pitch target. Descriptions of each model are shown in Table 1. Note that each model contains a predictor of contour shape and its interaction with the main predictor of interest, allowing the magnitude of that predictor's effect to vary between rising and falling contours, which improves model performance (see [13] Ch. 3). In addition to the predictors and random effects listed in Table 1, each model included a disc parameter with a predictor of contour shape, random slopes of contour shape by participant, and random intercepts by participant and utterance.

Table 1: *Model descriptions for pitch span (excursion size) and level (mean F0) analyses. Continuous predictors were transformed to semitones from 90 Hz, and the excursion size predictor was centered. Contour shape was treatment coded with falling as the reference level.*

Model name	**Predictors and random effects**
Excursion size	rating ∼excursion size * contour shape + (1 + excursion size * contour shape | participant) + (1 | utterance)
Mean F0	rating ∼mean F0 (in nuclear region) * contour shape + (1 + mean F0 * contour shape | participant) + (1 | utterance)

These models revealed credible positive effects of pitch excursion size ($\hat{\beta}$ = 0.09, 95% CrI [0.07, 0.11]) and mean F0 ($\hat{\beta}$ = 0.35, 95% CrI [0.28, 0.44]), indicating that surprise ratings vary in relation to these cues: listeners perceive a higher degree of surprise when the phrase-final word has a greater pitch excursion and when the phrase-final word has a greater mean F0. Both models also showed credible effects of contour shape (**Excursion size model:** $\hat{\beta}$ = 1.05, 95% CrI [0.84, 1.27]; **Mean F0 model:** $\hat{\beta}$ = 0.61, 95% CrI [0.43, 0.81]), with rising contours eliciting higher surprise ratings than falling contours, and credible positive interactions between contour shape and the main predictor of interest (**Excursion size model:** $\hat{\beta}$ = 0.15, 95% CrI [0.12, 0.18]; **Mean F0 model:** $\hat{\beta}$ = 0.13, 95% CrI [0.09, 0.17]). These interactions reflect that the magnitude of the excursion size and mean F0 effects are estimated to be greater in rising contours than falling contours.

4. Discussion and Conclusions

Overall, we do not find support for the Rett and Sturman account of intonational mirative marking in MAE. Our results show that both pitch accents and edge tones contribute to interpretations of speaker surprise in MAE. In our experiment, we found that participants' interpretation of speaker surprise was sensitive to variation in the scaling of both L* and H* pitch accents, indicating that the encoding of mirativity is not restricted solely to the use of L+H*. In the introduction, we suggested the alternative hypothesis that H* pitch accents, in addition to L+H*, can convey speaker surprise, with the degree of perceived surprise increasing with the height of the pitch accent F0 target. While we find evidence that H* pitch accents convey speaker surprise, we only find partial support for the hypothesized relationship between the height of the pitch accent target and the degree of perceived surprise. At low ending pitch steps, higher accentual pitch targets were associated with with higher surprise ratings, as predicted, but at high ending pitch steps, lower accentual pitch targets were associated with higher surprise ratings.[5] The effect of ending pitch was shown to be greater than that of accentual pitch, which was unexpected based on proposals that emphasize the role of pitch accents in conveying mirativity in MAE [2, 4]. While this may suggest that the boundary tone plays a greater role than the nuclear pitch accent in encoding mirativity, this result could also reflect a limitation imposed by the set of tunes investigated in this study; variation in ending pitch comprises a larger portion of the acoustic variation in the stimuli than variation in accentual pitch, as the ending pitch continuum spans a wider range of F0 values than the accentual pitch continuum (a necessary design feature in order to create a single continuum that includes rising trajectories from high pitch accents and falling trajectories from low pitch accents). It is possible that a similar experiment that included higher pitch accent target values than those investigated here would find a larger effect of accentual pitch; this possibility is being explored in ongoing work. However, note that, for H*H-H%, H% by definition must be higher than H*; hence, any experiment using H*H-H% would be subject to the constraint that there needs to be an ending pitch value higher than the maximum value of the accentual pitch continuum. There may also be variation in perceived surprise that correlates with pitch accent shape, which is not investigated here.

Our pitch level and pitch span analyses found variation in surprise ratings correlating with both of these phonetic parameters, in line with [9]. However, Gussenhoven and Rietveld's study in Dutch showed that perceived surprise for L*H-H% contours was highest when pitch **span** was high (i.e., accentual pitch was low and ending pitch was high), whereas perceived surprise for H*H-H% contours was highest when pitch **level** was high (i.e., both accentual and ending pitch were high), meaning that the observed variation in perceived surprise could not be captured solely by either parameter. This difference in rating behavior between L*H-H% and H*H-H% tunes was not replicated for MAE; among the rising contours, mean surprise ratings monotonically decreased as accentual pitch increased. Along with the credible interaction between accentual and ending pitch, this could suggest that in MAE, mirativity is conveyed solely by large pitch span regardless of tune, a conclusion that is consistent with evidence from additional tunes not investigated here. For instance, [18] found that the rise-fall-rise tune (L*+HL-H%) offers an incredulous interpretation[6] when expressed with a higher pitch span. However, the hypothesis that mirativity is conveyed entirely by pitch span makes predictions that are not supported by the empirical mean ratings for some contours. For example, for contours with accentual pitch at the highest step (i.e., the rightmost column in Fig. 3), the largest pitch spans occur at ending pitch steps 1 and 2, suggesting that these contours should have higher ratings than ending pitch steps 3, 4, and 5. Similarly, an account based solely on F0 level would capture some generalizations about the data, e.g., the overall higher mean ratings for rising contours than falling contours, but would make incorrect predictions within the rising contours, i.e., predicting that lower accentual pitch leads to lower surprise ratings rather than higher ones.

Together, our results suggest both that intonational mirative marking is distributed over multiple pitch targets across the nuclear tune,[7] and that the contributions of these distinct pitch targets cannot be explained by a single phonetic parameter such as pitch level or pitch span.

5. Acknowledgements

We are grateful to the members of the ProSD Lab and Phonatics group at Northwestern University for their feedback, and to Chun Chan for his help with experiment implementation.

6. References

[1] S. DeLancey, "The mirative and evidentiality," *Journal of Pragmatics*, vol. 33, no. 3, pp. 369–382, 2001.

[2] J. Rett and B. Sturman, "Prosodically marked mirativity," in *Proceedings of the 37th West Coast Conference on Formal Linguistics*, D. K. E. Reisinger and M. Huijsmans, Eds., 2021, pp. 1–20.

[3] K. Kraus, "Great intonations," Doctoral dissertation, University of California, Santa Cruz, 2018.

[4] I. Sag and M. Liberman, "The intonational disambiguation of indirect speech acts," in *Papers from the Eleventh Regional Meeting: Chicago Linguistic Society*, R. E. Grossman, L. J. San, and T. J. Vance, Eds., 1975, pp. 487–497.

[5] J. Pierrehumbert, "The phonetics and phonology of English intonation," Doctoral dissertation, Massachusetts Institute of Technology, 1980.

[6] D. R. Ladd, *The Structure of Intonational Meaning: Evidence from English.* Indiana University Press, 1980, ch. Paralanguage and Gradience.

[7] ——, "The trouble with ToBI," in *Prosodic Theory and Practice*, J. Barnes and S. Shattuck-Hufnagel, Eds. MIT Press, 2022, pp. 247–257.

[5] This is consistent with the view that expanded pitch range across an utterance conveys speaker surprise; [17] show that L* targets are lowered when the speaker increases their pitch range, while H* targets are raised. However, it should nonetheless be noted that the degree to which L* can be lowered is markedly reduced compared to the degree to which H* can be raised.

[6] I.e., an interpretation that involves speaker surprise, but also additional nuances such as disbelief or skepticism. Whether these additional nuances correlate with intonational differences is left to future work.

[7] Other work has suggested that prenuclear pitch accents may also contribute to mirative marking in MAE [4, 3]; see also [19] for German. Since our study was restricted to the nuclear tune, our results cannot speak to these claims; this remains an avenue for future research.

[8] K. Iskarous, J. Cole, and J. Steffman, "A minimal dynamical model of intonation: Tone contrast, alignment, and scaling of American English pitch accents as emergent properties," *Journal of Phonetics*, vol. 104, p. 101309, 2024.

[9] C. Gussenhoven and T. Rietveld, "The behavior of H* and L* under variations in pitch range in Dutch rising contours," *Language and Speech*, vol. 43, no. 2, pp. 183–203, 2000, place: United Kingdom Publisher: Kingston Press.

[10] D. R. Ladd, K. Silverman, F. Tolkmitt, G. Bergmann, and K. Scherer, "Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect," *Journal of The Acoustical Society of America*, vol. 78, pp. 435–444, 1985.

[11] X. Liu, Y. Xu, W. Zhang, and X. Tian, "Multiple prosodic meanings are conveyed through separate pitch ranges: Evidence from perception of focus and surprise in Mandarin Chinese," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 21, no. 6, pp. 1164–1175, 2021.

[12] L. Schiefer and A. Batliner, "A ramble round the order effect." in *Forschungsberichte des Instituts fürPhonetik und Sprachliche Kommunikation derUniversität München*, 1991.

[13] T. Sostarics, "Structured variation in intonational form and interpretation in American English," Doctoral dissertation, Northwestern University, 2025.

[14] P.-C. Bürkner, "Bayesian item response modeling in R with brms and Stan," *Journal of Statistical Software*, vol. 100, no. 5, pp. 1–54, 2021.

[15] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2025. [Online]. Available: https://www.R-project.org/

[16] P.-C. Bürkner and M. Vuorre, "Ordinal regression models in psychology: A tutorial," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 1, pp. 77–101, 2019.

[17] M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language sound structure: studies in phonology presented to Morris Halle*, M. Aronoff and R. Oehrle, Eds. MIT Press, 1984, pp. 157–233.

[18] J. Hirschberg and G. Ward, "The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English," *Journal of Phonetics*, vol. 20, no. 2, pp. 241–251, 1992.

[19] H. Seeliger and S. Repp, "Information-structural surprises? contrast, givenness, and (the lack of) accent shift and deaccentuation in non-assertive speech acts," *Laboratory Phonology*, vol. 14, no. 1, 2023.