Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Functional modeling of F0 variation across speakers and between phonological categories: Rising pitch accents in American English

Jennifer Cole[1], *Jeremy Steffman*[2], *Aya Awwad*[2]

[1]Northwestern University
[2]The University of Edinburgh
jennifer.cole1@northwestern.edu, jeremy.steffman@ed.ac.uk, a.awwad@sms.ed.ac.uk

Abstract

The Autosegmental-Metrical model of American English distinguishes three pitch accents with rising F0 trajectories (H*, L+H*, L*+H), differing in peak alignment and presence vs. absence of a low pitch marking the rise onset. Empirical studies report additional distinctions in the dynamics and scaling of the F0 rise, raising the question of which properties best capture variation among accents. We use functional principal components analysis (FPCA) to examine dynamic properties of accentual F0 trajectories in data from an intonation imitation experiment. F0 trajectories from 70 speakers producing rising accents on the phrase-final (nuclear) accented word were submitted to FPCA. The first three PCs account for 95% of variation in F0 trajectories and each shows significant differences between the three rising accents. Variation in PC1 primarily relates to differences in the overall F0 level of the trajectory, PC2 captures differences in rise shape (scooped vs. domed rise) and PC3 captures fine variation from a following Low phrase accent. Alignment distinctions are distributed across all three PCs. Examination of individual speakers shows all use PC1 and PC2 to some degree to distinguish rising accents, with no trading relations. Rises are variously implemented through level or shape distinctions, to varying degrees across individuals.

Index Terms: American English intonation, pitch accent, pitch contours, FPCA, individual differences, trading relations

1. Introduction

In the Autosegmental-Metrical (AM) theory of intonation, patterns of F0 variation that encode pragmatic meaning at the phrase level are generated from an underlying sequence of discrete tone features, which in American English (AE) includes tonally specified pitch accents that associate with the stressed syllable of words with phrasal prominence. The inventory of pitch accents proposed by [1,2] includes three accents that define high or rising F0 trajectories. The monotonal H* accent specifies a high F0 target on the stressed syllable of the accented word. The bitonal L+H* similarly defines a high F0 target on the stressed syllable (the starred tone), but with the addition of a preceding Low tone and a corresponding low F0 target at the onset of the rise, generally located at the beginning of the stressed syllable. The other bitonal accent is L*+H, which specifies a low F0 target on the stressed syllable followed by a high F0 target realized on the following syllable if there is one. Notably, the L*+H accent has a late F0 peak, compared to the relatively earlier F0 peaks of H* and L+H*. Schematic F0 trajectories for these three accents are shown in Figure 1.

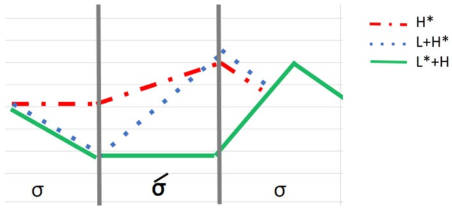

Figure 1: *Schematic F0 trajectories for three high/rising pitch accents: H* (red, dashed), L+H* (blue, dotted), L*+H (green, solid). Vertical lines mark boundaries at beginning and end of the stressed syllable (center).*

The perceptual distinction between accents with an early vs. late peak can also be conveyed by differences in the shape of the F0 trajectory related to the slope [3, 4] or curvature of the F0 rise [5], scaling of F0 extrema, or by similar differences in the post-peak F0 fall [5, 6]. F0 peak alignment, slope, scaling and curvature function independently to determine the acoustic 'tonal center of gravity' (TCoG) of the F0 movement over the stressed syllable [5, 6]. These acoustic parameters may be variously recruited for encoding pitch accent contrasts across languages [5]. Moreover, evidence from German and Italian indicates similar dimensions of variation characterize differences among individual speakers, with some using alignment and other using rise/fall slope as the primary dimension of contrast between rising accents [4].

While the studies cited above provide converging evidence that multiple dimensions of F0 modulation serve to distinguish phonologically distinct high/rising pitch accents across languages and speakers, there has not yet been a comprehensive production study examining the F0 dynamics of high/rising accents in AE that includes all three of the high/rising accents, or which examines pitch accent production in all contexts of the following phrase accent and boundary tone. The present study aims to fill this knowledge gap by identifying the characteristic dynamic F0 properties that distinguish the high/rising accents in AE. Distinctions in the global shape of accentual F0 trajectories are analyzed using functional principal component analysis (hereafter, FPCA), to test the hypothesis from the AM model that variation in F0 dynamics is structured in terms of a low target preceding the accentual peak, and in the temporal location of the peak. We also look for evidence of accent distinctions conveyed through 'shape' differences manifest in the slope and curvature of the F0 rise. FPCA is applied to data aggregated over speakers, with results analyzed to identify parameters of variation in terms of F0 peak alignment and in parameters that contribute to F0 shape, including slope and rise curvature. In addition, we examine FPCA results for individual participants to identify parameters of individual speaker variation along the same acoustic dimensions.

10.21437/SpeechProsody.2024-206

2. Methods

2.1. Materials

The F0 trajectories analyzed in this study are drawn from an intonation production experiment [7] using an imitation paradigm in which on each trial an intonation pattern is presented auditorily via two model sentences, and participants reproduce the heard melody on a new sentence presented in text format. The stimuli were short sentences ending in a 3-syllable name with initial stress (e.g., "He answered Jeremy"), recorded from two native speakers of AE, one male and one female. These base recordings underwent pitch resynthesis using the PSOLA algorithm in Praat [8, 9] to generate one of 12 intonation patterns over the final word, specified in terms of the 'nuclear' (phrase-final) pitch accent (H*, L+H*, L*+H) on the stressed syllable, followed by a phrase accent (H- or L-) and boundary tone (H% or L%). The resynthesized F0 patterns of the final word were based on straight-line approximations from [10], which were in turn modeled after empirical data in [1], as shown in Figure 2.[1] 70 participants, monolingual speakers of AE, were recruited from Prolific (12F, 14M, 2 non-binary, mean age = 23.7) and from the undergraduate Linguistics subject pool at Northwestern University (17F, 17M, 1 non-binary, mean age = 19.7). Participants reproduced the 12 tunes (12 repetitions each) on three target sentences (e.g., "He modeled Harmony.") with syllable and stress patterns similar to the stimuli.[2] We removed likely F0 tracking errors [11], which resulted in exclusion of approximately 12% of the data, retaining 8,914 files for analysis.

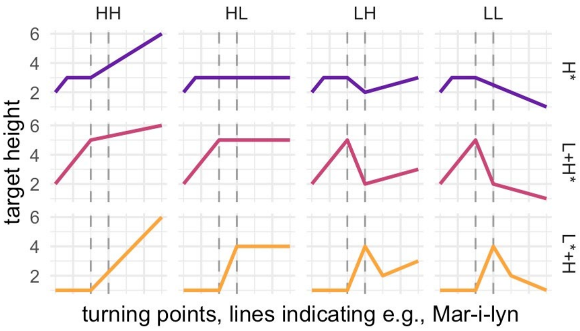

Figure 2: *Schematic trajectories for the model tunes in the study, showing the three pitch accents (rows) and four boundary tones (columns). Dashed lines indicate syllable boundaries in the nuclear-accented word.*

F0 trajectories were extracted from the first two syllables of the sentence-final word of each analyzed trial. This portion of the phrase-final F0 trajectory represents the phonetic implementation of the pitch accent, though the extracted interval also reflects the transition from the accentual peak to the high- or low-tone phrase accent and the following boundary tone, especially for productions where the accentual peak falls in the first syllable (as expected for imitations of H* and L+H*). The analyzed portions of the time-normalized speaker-mean trajectories for each pitch accent are shown in Figure 3A. Based on visual inspection, these empirical findings already suggest a pattern of co-variation, with F0 rise slope, peak alignment and peak scaling jointly distinguishing the three accents.

2.2. Analysis

We use FPCA to quantify the dynamic properties of the F0 trajectories implementing high/rising pitch accents. FPCA is a data-driven method that identifies the distinct global shape characteristics in time-series data [12, 13]. As discussed in [12],

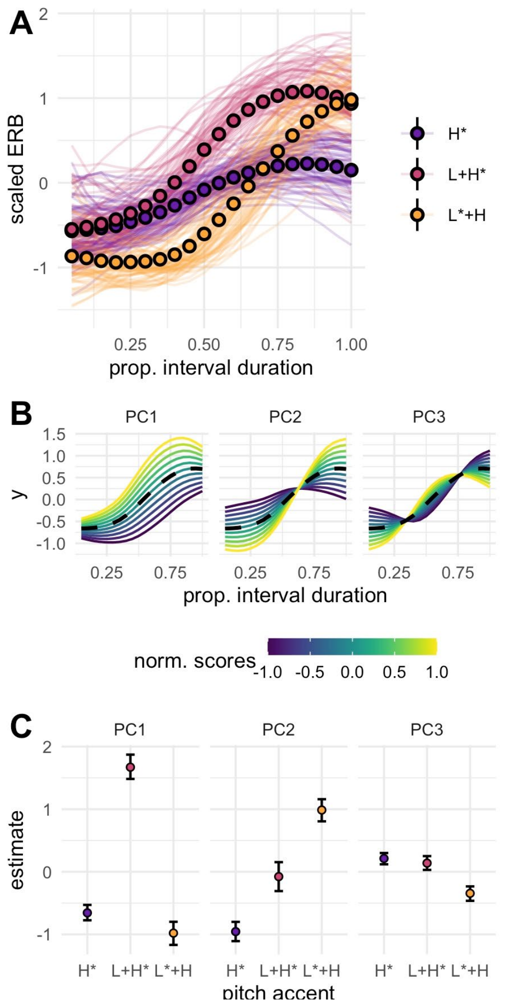

Figure 3: *Empirical F0 trajectories of imitated pitch accents. Thin lines are mean F0 trajectories by speaker (N=70). Dark dotted lines represent the mean trajectory of each accent types over all speakers* (Panel A). *The first three PCs shown as trajectories that deform the mean F0 curve fit to the aggregated data (dashed black line) as determined by the normalized PC score, coded by color* (Panel B). *Bayesian regression model estimates of mean PC scores (and 95% credible intervals) for empirical data grouped by accent type* (Panel C).

[1] Audio files of stimuli are available on the OSF at https://osf.io/b3su6/.

[2] Reproducing a heard melody on a new sentence requires encoding and retrieving a representation of that melody. While gross shape distinctions were reproduced, pitch level, range and some finer shape distinctions were not. See [7] for details.

the advantage of FPCA is twofold. First, it captures global shape characteristics jointly determined through multiple acoustic parameters, e.g., F0 peak alignment, slope and curvature, in terms of the principal components (PCs) from a set of continuous functions that when applied, deform the mean curve fit to the entire dataset. The shape of a PC, generated as a time series over the dependent variable (here, F0), represents the deformation it exerts on the mean curve at each normalized time step. Individual F0 trajectories input to the analysis can then be identified in terms of the associated weight of each PC (the PC 'score'), which represents the relative contribution of that PC's deformation to the mean curve, in modeling the shape of the input along its whole trajectory. The second benefit of FPCA is that the individual PC scores associated with each input F0 trajectory can be submitted to statistical analysis using regression or other methods to test hypotheses about class distinctions, in our case, corresponding to the three high/rising accent types proposed as phonological contrasts in the AM model of AE. We adopted the workflow in [12]: 8,914 (unlabeled) F0 trajectories from the first two syllables of the nuclear word were scaled within speaker, normalizing for speaker differences in F0 height and range, and time-normalized with 20 temporal samples, then submitted to FPCA using the fda package in R [14]. Scores from the first three PCs were subsequently submitted to separate Bayesian mixed-effects linear regressions [15], with weakly-informative student t priors, predicting variation in the PC score as function of pitch accent, boundary tone, and their interaction. Models included random intercepts for speaker, and by-speaker slopes for both fixed effects and interactions. Here we focus on just the marginal effect of pitch accent which we computed using [16]. In reporting the effects, we give posterior median and 95% credible intervals (CrI) for the contrasts comparing each of the three pitch accents. When the 95% CrI excludes the value of zero, we take this as credible evidence for an effect (i.e. a reliably estimated non-zero difference).

3. Results

Figure 3B shows the time-series generated from each of the first three PCs, each plotted across a range of PC scores (proportions of the s.d. of the PC) from -1 to 1. The black dashed line superimposed on each PC series is the mean curve fit to the aggregated data (same curve projected on each PC). Notice that functions with the highest (positive) and lowest (negative) PC scores generate shapes with greater positive or negative magnitude, which when applied to the mean curve will have a stronger deforming effect. Also notice that the PCs differ in the location, direction and magnitude of their effect on shaping the mean trajectory. These PCs together account for 95% of the variation in the combined input F0 trajectories: 57% for PC1, 28% for PC2, 10% for PC3. **PC1 primarily captures variation in the overall level of the F0 trajectory, with secondary effects on the scaling and alignment of the F0 peak and in the F0 range.** Higher positive PC1 scores (e.g., the yellow trajectory of PC1) shift the mean curve upward and have a larger F0 range, and an earlier and higher F0 peak. By comparison, trajectories associated with lower PC1 scores (e.g., the purple PC1 trajectory) are shifted downward in F0 and span a smaller F0 range with later and lower peaks. **PC2 primarily captures variation in F0 trajectory shape through combined effects of peak alignment, rise slope, and F0 range.** Higher PC2 scores exert a deformation resulting in a distinct scooped shape and a correspondingly steep rise slope, late rise onset and late peak, compared to trajectories with low PC2 scores, which have a flatter shape. **PC3 has a uniform effect of imposing a scooped rise shape and captures variation primarily in peak alignment.** Higher PC3 scores yield trajectories that begin lower and have an earlier and lower F0 peak, compared to trajectories with higher PC3 scores. Note that all three PCs also capture variation in the final quarter of the interval related to the variable influence of a following phrase accent (H- or L-).

PC scores for each input F0 trajectory were determined as those whose combined effects on the mean curve generate the mostly closely matching trajectories. Scores for each PC, aggregated over speakers, were submitted to separate mixed effects models. Figure 2C plots the model-estimated PC scores for each accent type and each PC. For **PC1** (left panel**)**, large positive scores are associated with productions of L+H*, indicating F0 trajectories with large excursions that are overall higher in the F0 range and have an early, high peak. In comparison, the negative PC1 scores for H* and L*+H productions indicate F0 trajectories with smaller excursions, placed lower in the F0 range, with later, lower peaks. Comparison of marginal effects estimates for PC1 across pitch accents confirms a credible difference for each pair. L+H* shows credibly higher PC1 values as compared to both H* (b = -2.33, 95CrI[-2.57,-2.08]) and L*+H (b = -2.65, 95CrI [-2.97,-2.33]). H* also showed credibly higher PC1 compared to L+H* (b = 0.32, 95CrI [0.05,0.66]). **PC2** scores show a more graded distinction, with positive scores for L*+H indicating trajectories with a scooped rise, later peak, and a steeper slope spanning a larger F0 range. PC2 scores for L+H* indicate a trajectory with shape characteristics like L*+H, but milder—less scooped, earlier peak, shallower slope. The negative PC2 scores for H* indicate little deformation of the mean curve. Comparisons of PC2 across pitch accents also show a credible difference for each pair. H* shows credibly lower PC2 values than L*+H (b = -1.94, 95CrI[-2.09,-1.80]) and lower than L+H* (b = -0.88, 95CrI[-1.03,-0.73]). L*+H also shows credibly higher PC2 scores in comparison to L+H* (b = 1.07, 95CrI[0.87,1.26]). **PC3** shows much weaker differentiation of accents, promoting a scooped accent shape with a late F0 peak for L*+H, and a milder but similarly scooped shape, with an earlier peak, for H* and L+H*. A credible difference is confirmed for each pair of comparisons across pitch accents in PC3, though the magnitude of differences is much smaller. L*+H shows credibly lower PC3 values in comparison to H* (b = 0.56, 95CrI[0.47, 0.65]) and L+H* (b = -0.48, 95CrI[-0.58,-0.40]). Credibly lower PC3 scores are also found for L+H* as compared to H* (b = 0.07, 95CrI[0.005,0.14]), though the magnitude of this difference is small and credible intervals only narrowly exclude 0. Individually and combined, these three PCs succeed in characterizing a three-way distinction among productions of these accents (Table 1).

Table 1: *Model estimated PC scores, by sign (negative, positive, zero) for each pitch accent. Marginal non-zero value in parentheses.*

Accent	PC1	PC2	PC3
H*	-	-	(+)
L+H*	+	0	0
L*+H	-	+	-

Jointly, PC1 and PC2 capture roughly 85% of the variation in the data, with primary effects on overall F0 level and F0 trajectory shape. We next examine the PC1 and PC2 scores for individual speakers to assess variation in the degree to which a

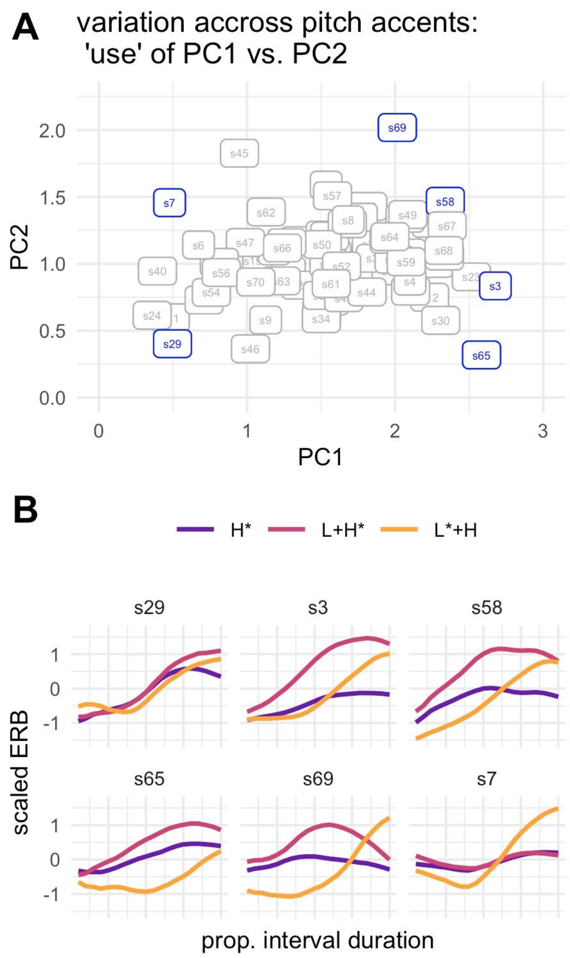

Figure 4: *Mean PC1 and PC2 'use' scores (s.d. of mean PC scores across accents) by speaker* (Panel A)*. Mean F0 trajectories for six speakers* (Panel B).

speaker implements high/rising accent distinctions through variation in F0 level (PC1) or shape (PC2) as the primary dimension, or through co-variation along both dimensions. Figure 4A displays how much PC1 and PC2 varied as a function of pitch accent for each speaker. We operationalized this in terms of standard deviation, taking the mean PC1 and PC2 of each accent for each speaker, and computing the standard deviation in terms of the difference between the mean PC for a given accent and the mean PC over the three accents, for each speaker. Larger values on the x and y axes correspond to larger variation in PC1 and PC2 *as a function of pitch accent*, for a given speaker. We take this to represent **PC 'use'**: a speaker with a higher standard deviation for PC1 uses variation in F0 level (the primary dimension of variation for PC1) in differentiating among the high/rising accents more heavily than a speaker who has a low value for PC1 'use'. Notably, there is nearly continuous individual variation along both PC 'use' dimensions, with individuals filling every quadrant of the plot—all speakers implement pitch accents with variation in both F0 level and shape. There is also not a clear negative or positive correlation between PC1 and PC2 'use', indicating that level and shape are not in a trading relationship.

To better understand how the variable patterns of PC 'use' relate to the F0 trajectories speakers produce for different high/rising accents, we examine the mean F0 trajectories of each accent for six individual speakers who represent relatively extreme values of PC1 and PC2 use (Figure 3B). Speaker 29 makes minimal distinctions among accent types, with very low 'use' scores for PC1 and PC2. This speaker produces very similar F0 trajectories for all three accent types. On the other extreme we have speakers 69 and 58 with high 'use' scores for both PC1 and PC2. These speakers differentiate accents by overall F0 level (high for L+H*, low for L*+H), and shape (later peaks and more scooped shape for L*+H compared to L+H* and H*). Speakers 65 and 3 use level distinctions (high PC1) as the major dimension of contrast, while speaker 7 uses shape, with a scooped rise for L*+H, and nearly identical, minimally rising trajectories for H* and L+H*.

4. Discussion

F0 trajectories of three high/rising pitch accents in phrase-final (nuclear) position were examined in data from an intonation imitation experiment with speakers of AE. F0 trajectories were analyzed using FPCA to identify the F0 parameters associated with the first three PCs, representing the primary dimensions of variation in the data aggregated over speakers. Following the tonal specification of these accents in the AM model and their schematic illustrations in ToBI training materials, we hypothesized that a low pitch target at the onset of the rise and the alignment of the F0 peak would be the primary dimensions of variation. FPCA analysis identified three PCs that together account for 95% of the variation in F0 trajectories in our data. There are three important findings from the FPCA results from data aggregated over speakers. First, **F0 level is the primary dimension of variation among the high/rising accents:** L*+H is produced with an overall higher F0 level than L+H* and H*. Second**, variation in peak alignment is distributed across all three PCs:** For each PC, the location of the F0 peak varies from earlier to later across PC score variation. Third, **the second most substantial dimension of variation, PC2, is best described in terms of TCoG** [5], conditioning co-variation of several acoustic parameters related to rise shape: peak alignment, curvature (scooped shape with low pitch at rise onset), and rise slope. In addition, PC scores from individual speakers show that all speakers implement distinctions among high/rising accents through co-variation of F0 level (PC1) and shape (PC2). And while there is no evidence for trading relations between these two dimensions of variation, some speakers favor level distinctions, while others favor shape distinctions, similar to findings from German and Italian [4].

5. Conclusions

Variation in the F0 trajectories of high/rising pitch accents in AE is structured around two primary dimensions: F0 level and rise shape, the latter a dynamic property conditioned by the acoustic parameters of F0 peak alignment, slope and rise curvature. These findings, along with prior findings from German [3, 17, 18] and Italian [4], support a theory where peak alignment, a key contrastive feature in the AM model of AE [1, 19] and other languages (e.g., [20]), co-varies with other shape parameters to define phonological distinctions among intonational categories.

6. Acknowledgements

This work was supported by NSF BCS-1944773 (Cole, PI).

7. References

[1] Pierrehumbert, J.B. *The Phonology and Phonetics of English Intonation*, PhD thesis, M.I.T. 1980.

[2] Beckman, J. and J.B. Pierrehumbert. 1986. "Intonational structure in Japanese and English. *Phonology, 3,* 255-309.

[3] Niebuhr, O. 2007. The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica, 64*, 174-193.

[4] Niebuhr, O., D'Imperio, M., Gili Fivela, B., and Cangemi, F. 2011. Are there 'shapers' and 'aligners'? Individual differences in signaling pitch accent category. *Proceedings of the 17th International Congress of Phonetic Sciences,* pp, 120 –123.

[5] Barnes, J., Veilleux, N., Brugos, A., and Shattuck-Hufnagel, S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology,* 3(2), 337–383.

[6] Barnes, J., Brugos, A., Veilleux, N., and Shattuck-Hufnagel, S. 2021. On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics, 85,* 101020.

[7] Steffman, J., Cole, J., & Shattuck-Hufnagel, S. (2024). Intonational categories and continua in American English rising nuclear tunes. *Journal of Phonetics*, *104*, 101310.

[8] Boersma, P. and Weenink, D. 2020. Praat: doing phonetics by computer (version 6.1.09). http://www.praat.org.

[9] Moulines, E. and Charpentier, F. 1990, Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9* (5-6), pp. 453–467.

[10] Veilleux, N, Shattuck-Hufnagel, S. and Brugos, A. 2006. Transcribing prosodic structure of spoken utterances with ToBI (version 6.911). Massachusetts Institute of Technology: MIT OpenCourseWare. [Online]. Available: https://ocw.mit.edu

[11] Steffman, J., & Cole, J. (2022). An automated method for detecting F0 measurement jumps based on sample-to-sample differences. *JASA Express Letters*, *2*(11).

[12] Gubian, M., Torreira, F., and Boves, L. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16–40. https://doi.org/10.1016/j.wocn.2014.10.001

[13] Arvaniti, A., Katsika, A. & Hu, N. (in press). Variability, overlap, and cue trading in intonation. *Language*.

[14] Ramsay J. 2023. fda: Functional Data Analysis. R package version 6.1.4. https://CRAN.R-project.org/package=fda.

[15] P.-C. Bürkner, 2017. brms: An R package for Bayesian multilevel models using stan," *Journal of Statistical Software, 80,* 1–28.

[16] R. V. Lenth. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means, 2021, R package version 1.7.1-1. [Online]. Available: https://CRAN.Rproject.org/package=emmeans

[17] Grice, M., Ritter, S., Niemann, H., and Roettger, T. 2017. Integrating the discreteness and continuity of intonational categories. J. Phon. 64, 90–107.

[18] Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., and Braun, B. 2022. Three Kinds of Rising-Falling Contours in German wh-Questions: Evidence From Form and Function. *Frontiers in Communication*, *7*, 838955.

[19] Pierrehumbert, J. and Steele, S. 1989. Categories of tonal alignment in English, Phonetica 46, 181-196.

[20] Prieto, P., D'Imperio, M., & Gili Fivela, B. (2005). Pitch accent alignment in romance: Primary and secondary associations with metrical structure. *Language and Speech*, *48*(4), 359–396.