Three Dimensions of Sentence Prosody and their (Non-)Interactions

Michael Wagner & Michael McAuliffe. McGill University

PaPE
June 12-14 2017

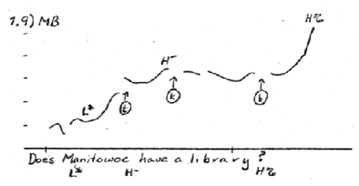
Three dimensions of Prosody:

Constituent structure ~ Phrasing
[AB]C vs. A[BC]

Focus ~ Prominence
ABC vs. AB**C**

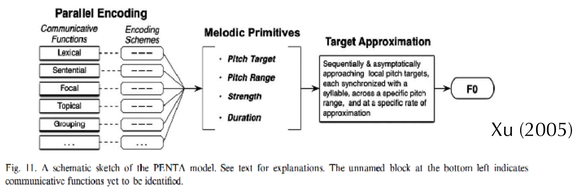
Speech Act ~ Tune
ABC! vs. ABC?

(How) Do the three dimensions interact?



★ Questions with initial focus consist of single Intonational Phrase (e.g., Pierrehumbert 1980)
★ Prominence should neutralize postfocal phrasing
★ "[..] at least one pitch accent somewhere in every (prosodic) phrase [..]" (Beckman 1996)
★ CF: Norcliffe & Jaeger 2005

Current AM Models: Yes



Xu (2005)

Fig. 11. A schematic sketch of the PENTA model. See text for explanations. The unnamed block at the bottom left indicates communicative functions yet to be identified.

★ Separate F0 function of Tune, Prominence, Phrasing
★ They could affect F0 without interacting
★ Direct phonetic correlates of semantic/syntactic functions (but see Ladd 2008 why phonological mediation is needed)
★ Only crucial here: independent representation of dimensions

Overlay Models: No

Test Case: F0 Scaling for Phrasing

★ Are there scaling effects in postfocally?

★ Are scaling effects different in questions?

Do prominence and tune interact with scaling?

Production Experiment

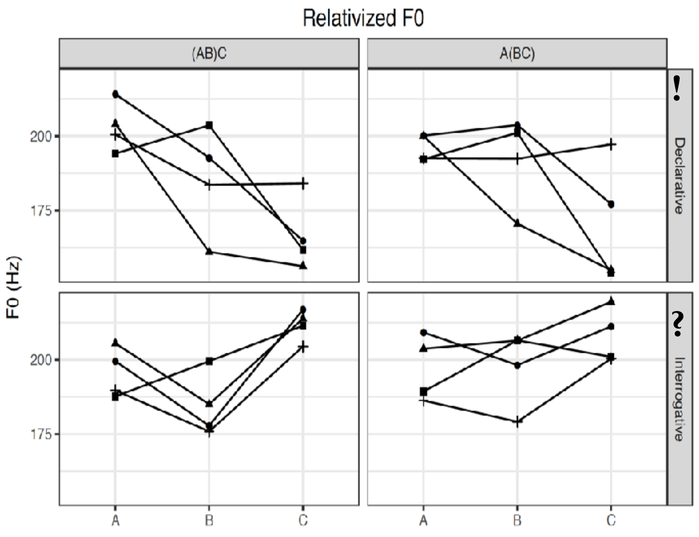
[AB]C
ABC I thought they said Sarah or Marvin, and Nolan arrived.
A**B**C I thought they said Marion or Sarah, and Nolan arrived.
AB**C** I thought they said Marion or Marvin, and Sarah arrived.
ABC I thought they said Sarah arrived.

A[BC]
ABC I thought they said Sarah, or Marvin and Nolan arrived.
A**B**C I thought they said Marion, or Sarah and Nolan arrived.
AB**C** I thought they said Marion, or Marvin and Sarah arrived.
ABC I thought they said Sarah arrived.

! But in fact they said that **Marion or Marvin, and Nolan** arrived.
? But you say that **Marion or Marvin, and Nolan** arrived?

! But in fact they said that **Marion, or Marvin and Nolan** arrived.
? But you say that **Marion, or Marvin and Nolan** arrived?

Scaling Effect No consistent phrasing pattern across different tunes and foci:

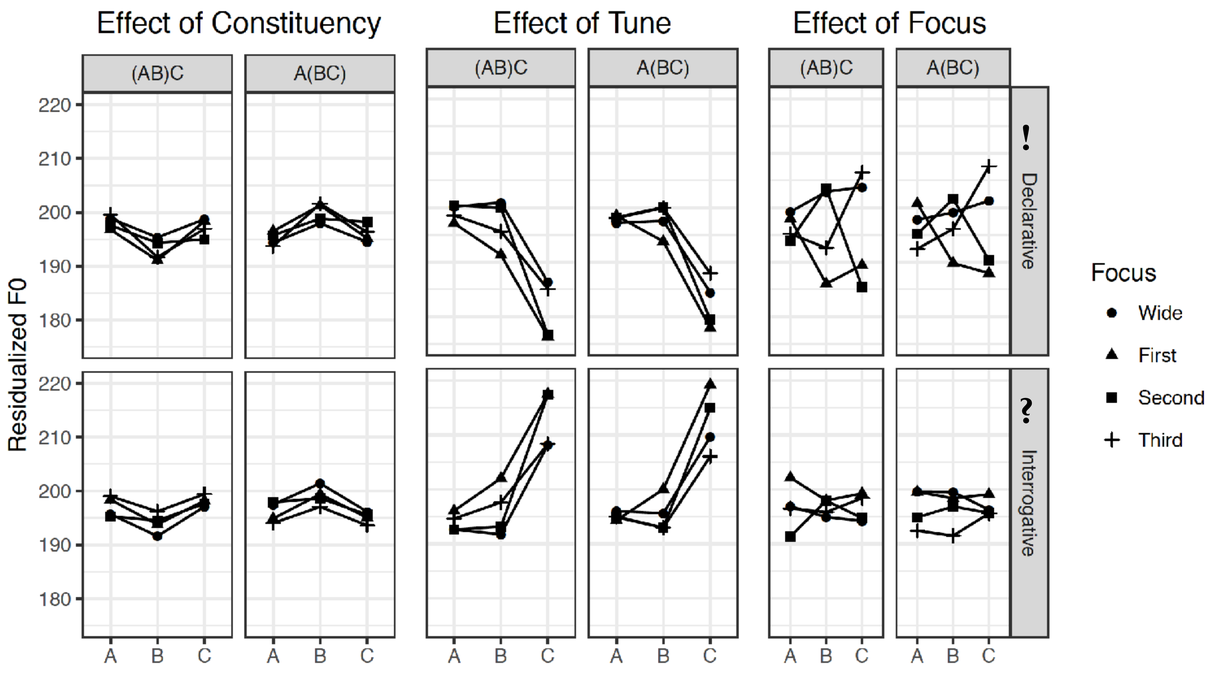


★ Aligned with Montreal Forced Aligner (McAuliffe et al. 2017a), normalized with SCT/Polyglot DB (McAuliffe et al. 2017b)
★ F0 based on z-scores for expected F0 based on speaker and segmental content, converted back to Hertz for display
★ Average measures over entire words
★ Disfluent & incorrect utterances were excluded (25%)

Focus
● Wide
▲ First
■ Second
+ Third

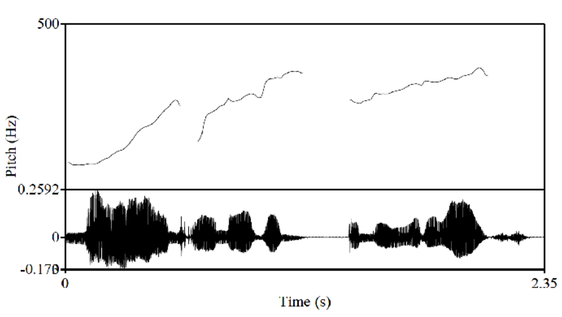
Residualization

★ Residualization for each dimension with LMER with other dimensions * position as fixed effects
★ … and item and participant random effects
★ Residualization is used here as a tool to visualize to what extent the contribution of 3 dimensions to pitch is consistent/interactive



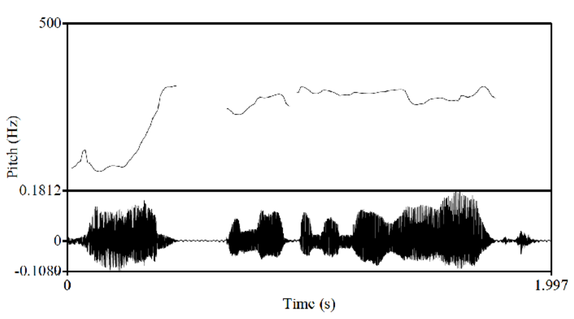
What does this look like? Scaling example with initial focus (focus on first conjunct; question; left vs. right branching):

[**A**B]C



Marion or Marvin, and Nolan

A[BC]



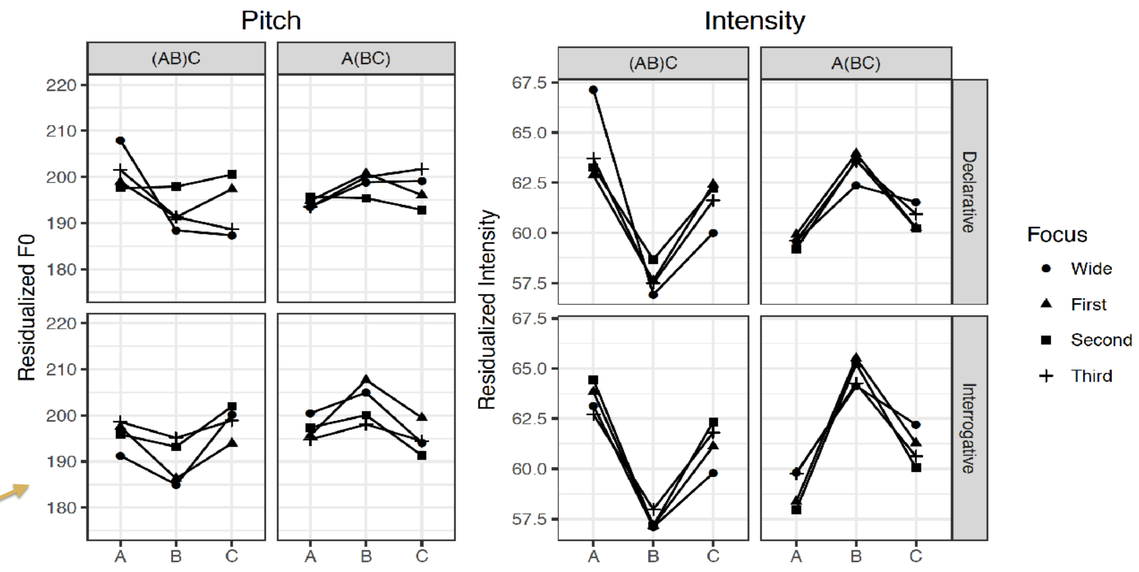
Marion, or Marvin and Nolan

★ No apparent obliteration of phrasing effect scaling in postfocal domain
★ No apparent difference in phrasing effect between questions and declaratives

★ Focus marked differently in questions (interaction between Focus*Tune)
★ Pitch accents lower in questions (L* vs. H*)

★ Statistical analysis as based not on residuals but complete model; model reveals significant contribution of phrasing with no significant interactions
★ Phrasing contributes to pitch in subset of cases where it is expected not do

But is scaling really about adjusting F0? Pitch correlates highly with intensity; intensity effect is more consistent:



Did people actually say these according to the manipulation?

★ Participants varied in how natural (vs. read) their speech sounded (all utterances are included)
★ For each dimension, 2 RAs annotated which tune/focus/phrasing their heard (or annotated 'unclear').
★ Inter-annotator agreement was 'almost perfect' for falling vs. rising (Cohen's kappa: 0.96),
★ … 'substantial' for constituency (Cohen's kappa: 0.73) and for prominence (Cohen's kappa: 0.63).
★ Based on one annotator:
- tune (rising vs. falling) 96.3% of the time as expected
- bracketing 61% (with about one third of sound files marked as 'unclear')
- prominence 38% (high rate of confusion between wide vs. focus on C; 21% 'unclear').
★ We based our pitch vs. intensity plot on the subset of data for which the correct levels for all three dimensions were annotated.

★ Gamming et al. 1988: Speakers raise F0 when aiming to talk louder for about a half semi-tone per dB
★ Could it be that F0 scaling is a passive reflex of speakers adjusting intensity for rhythmic reasons?

Selected References:
van den Berg, Gussenhoven, & Rietveld, "Downstep in Dutch: implications for a model," in Papers in Laboratory Phonology, vol. II: Gesture, segment, prosody, G. Docherty and R. Ladd, Eds. Cambridge: Cambridge University Press, 1992, pp. 335–58.
Féry & Truckenbrodt, "Sisterhood and Tonal Scaling," Studia Linguistica, vol. 59, no. 2-3, pp. 223–243, 2005.
Féry & Kentner, "The prosody of embedded coordinations in German and Hindi," in Proceedings of Speech Prosody, vol. 5, 2010.
Gramming, Sundberg, Ternström, Leanderson, and Perkins, "Relationship between changes in voice pitch and loudness," Journal of Voice, vol. 2, no. 2, pp. 118–126, 1988.
Kentner & Féry , "A new approach to prosodic grouping," The Linguistic Review, vol. 30, no. 2, pp. 277–311, 2013.
Ladd, "Declination and 'reset' and the hierarchical organziation of utterances," JASA, vol. 84, pp. 530–544, 1988.
Öhmann, "Word and sentence intonation: A quantitative model," KTH Department of Speech, Music and Hearing, Quarterly Status Report, 1967
Petrone, Truckenbrodt, Wellmann, Holzgrefe-Lang, Wartenburger, & Höhle, "Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists," Journal of Phonetics, vol. 61, pp. 71–92, 2017.
E. Norcliffe and T. Jaeger, "Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain," in Proceedings of Interspeech, 2005.
Truckenbrodt & Féry,"Hierarchical organisation and tonal scaling," Phonology, vol. 32, no. 01, pp. 19–47, 2015.
Xu, "Speech melody as articulatorily implemented communicative functions," Speech Communication, vol. 46, pp. 220–251, 2005.

Montreal Forced Aligner: McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger (forthcoming at Interspeech 2017)
Speech Corpus Tools & Polyglot DB: McAuliffe, Stengel-Eskin, Socolof, & Sonderegger. (forthcoming at Interspeech 2017)
Related full paper on this data: Wagner & McAuliffe (forthcoming at Interspeech 2017)

Thanks to Francisco Torreira and the McGill Prosody & Meaning Reading group for comments, and Erin Olson, Thea Knowles, and other members of ProsodyLab for help analyzing the data.