Prominence effects on the processing of spectral cues: testing sonority expansion in perception[1]

Jeremy Steffman UCLA jsteffman@ucla.edu

Background

Phrasal prosody finetunes the timing and amplitude of articulations in individual segments

- Boundaries: initial strengthening, pre-boundary lengthening [1,2]
- Prominence: manifestations of "prominence strengthening" - syntagmatic and paradigmatic contrast enhancement [3-5]
  - **Sonority expansion** [6]: expansion of oral cavity in phrasally prominent vowels; jaw lowering, lingual backing

Less studied is how these patterns impact perception and processing – i.e. the extent to which listeners take prosodically driven variation into account – esp. for cues that are used contrastively e.g. VOT, vowel duration

- Recent interest in integration of prosody in perception: "prosodic analysis" [7,8] – focused on prosodic boundaries
- The acoustic structure of vowels varies based on phrasal **prominence** (next slide) [9,10], what are the perceptual consequences?
- The present study tests how perception of spectral cues (F1 and F2) varies based on prominence
  - Glottalization in American English has been argued to have a prominence marking function [e.g. 5,11]; a localized/segmental cue to prominence?

Goals of the present study

(1) Test if prominence-driven patterns in formant structure (= consequences of sonority expansion), are exploited perceptually
(2) Test two types of contextual prominence: (a) prominence at the level of the phrase (b) word-initial glottalization

- Previous perception studies manipulate phasal prosody – the relevance of localized/segmental prominence cues (e.g. glottalization) remains untested

[1]These slides are intended to function like a decomposed poster – see stimuli examples and model summaries in the appendix

The test case: Prominence effects on vowels

Sonority expansion entails changes in formant structure [1] (for nonhigh vowels)[2]

- Raised F1 and lowered F2, correlating with jaw lowering and lingual backing
- Dual-patterning of F1/F2, (1) as a function of prominence (higher F1/ lower F2 when prominent) and (2) segmental category

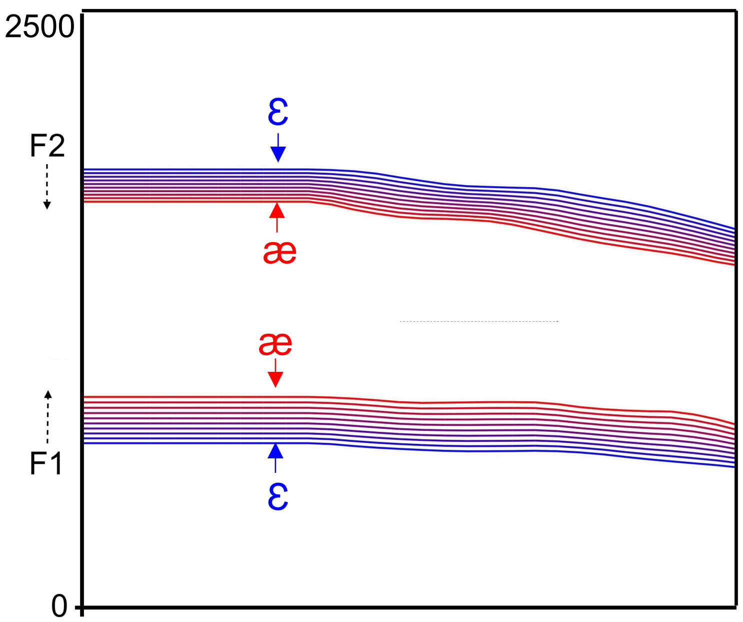
  Will listeners accordingly adjust perception of vowel contrasts based on prominence?

Methods

Forced choice task, 10 step joint F1/F2 continuum ranging from /ɛ/ to /æ/ (see right); categorized as “ebb” or “ab” by listeners.

- Higher F1/ lower F2 when /æ/)
- Two experiments (n = 30 in each); prominent and non-prominent conditions below

Exp 1: Phrasal prominence		Exp 2: V-initial glottalization	
I’ll say [target] now H* H* L-L%	Target has nuclear pitch accent (NPA)	ðəʔɛb	Target preceded by [ðə] with continuous formant transitions, or an intervening [ʔ]
I’ll SAY [target] now L+H* L-L%	Target is post-focus = non-prominent [12,13]	ðəɛb	



F1 and F2 showing the 10-step continuum interpolating from /ɛ/ to /æ/ (from Exp 1). Frequency in Hz on the y axis – time on the x axis.

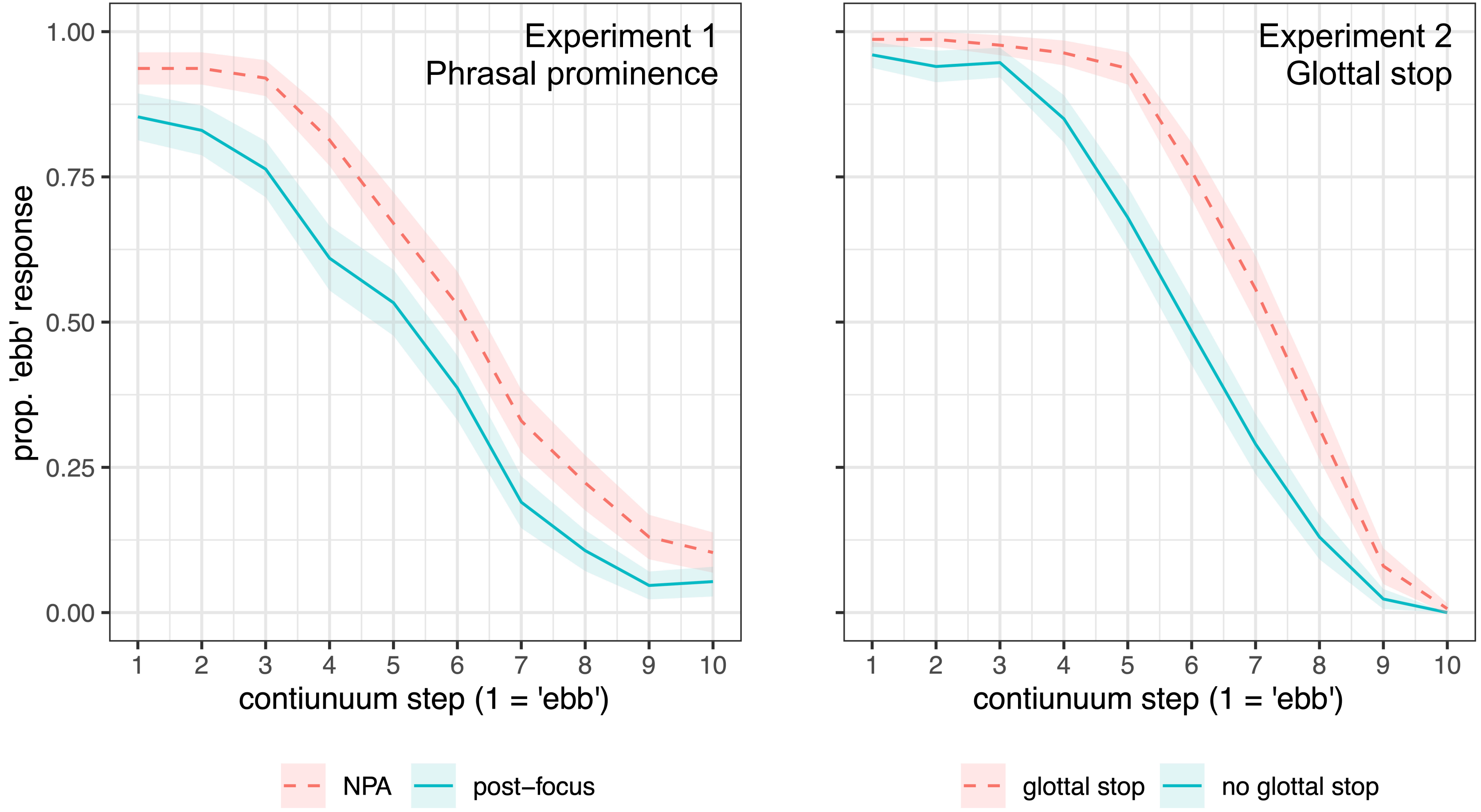
Prediction: If prominence mediates perception of the vowel contrast, higher F1/lower F2 should be categorized as /ɛ/ when marked as prominent:

- i.e. attribution of formant changes to prominence instead of a segmental contrast: **increased /ɛ/ percepts in prominent contexts** – Open questions: will phrasal prominence versus glottalization pattern differently? Is a glottal stop enough?

[2]In high vowels, sonority expansion can be suppressed as it conflicts with attainment of a high vowel target– findings shows various patterns [3,14]

Results[3]

In both experiments a prominent context shows a credible effect: increased ‘ebb’ responses



Categorization functions in both experiments, showing the proportion of ‘ebb’ responses on the y axis and continuum step on the x axis. Shading around lines shows 95% CI

[2]Results are assessed by mixed effect Bayesian logistic regression [15] – see appendix slides for model summaries.

Conclusions

Take home message: Listeners' perception of vowels is mediated by prominence – compensatory adjustments for sonority expansion driven formant modulations

- Some new evidence for the involvement of prominence in segmental processing – a departure from previous studies that focus on prosodic boundaries
- Both prominence at the phrasal level and a glottal stop generated similar adjustments in categorization – n.b. in natural speech phrasal prominence and glottalization often co-occur [5,11] dissociating them in this study helped test an independent contribution, though future work might test additive effects

Further directions

(1) Do high vowels show analogous effects? Sonority expanding gestures are suppressed for high vowels and some studies document the opposite: high vowels are *hyperarticulated* when prominent [3,15]

- Perceptual consequences of contrast-specific prominence enhancement strategies is a next step in understanding these effects (this is work in progress)

(2) Processing implications: how do these effects play out online? Is prosody integrated with formant cues immediately, or does it show a delayed influence as suggested by recent findings for prosodic boundaries? [8,16]

- Possible reasons to expect prominence processing to be different?
    - Perceived prominence can also be intrinsic [17] – conveyed by pitch/duration for a segment, i.e. properties that might not necessitate referencing contextual phrasal prosodic structure
    - Though n.b. the manipulation in Experiment 1 was purely contextual (identical target across conditions)
- Prominence marking in e.g. vowel-initial glottalization can be *localized* – quicker processing for local cues?
    - To test: do the effects in Experiment 1 and 2 play out similarly online? (this is work in progress)

(3) Prominence effects relation to *vowel-intrinsic pitch and duration* [18] – when do pitch and duration cue prominence as compared to cuing a contrast directly? Use of prominence cues in this regard has been shown to be flexible [19]

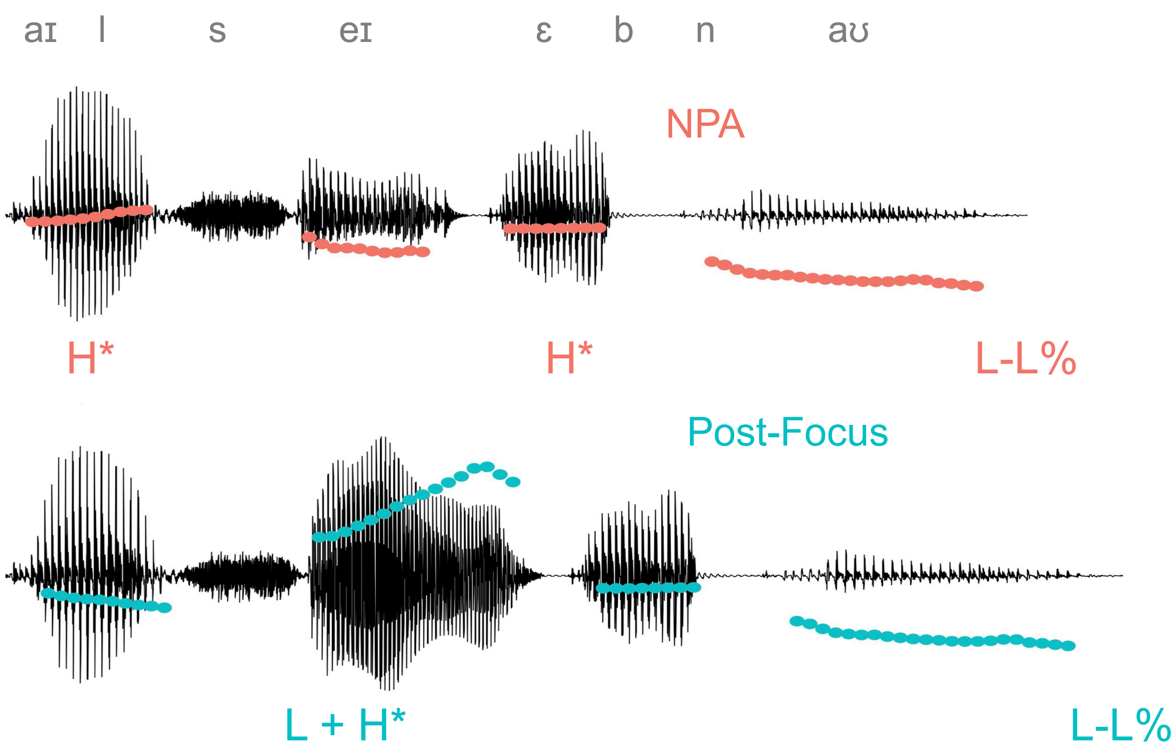
Acknowledgements

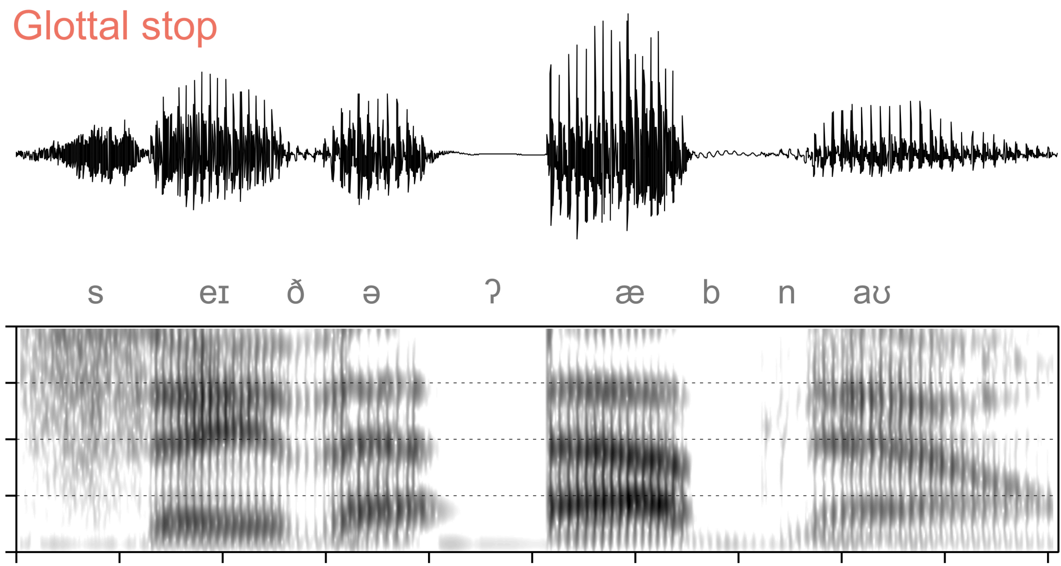
Many thanks are due Adam Royer for recording material for the stimuli, and to Danielle Bagnas, Qingxia Guo and Jae Weller for help with data collection. I am further grateful to Sun-Ah Jun, Pat Keating, Megha Sundara and Taehong Cho, and audiences at the UCLA Phonetics seminar for helpful feedback.

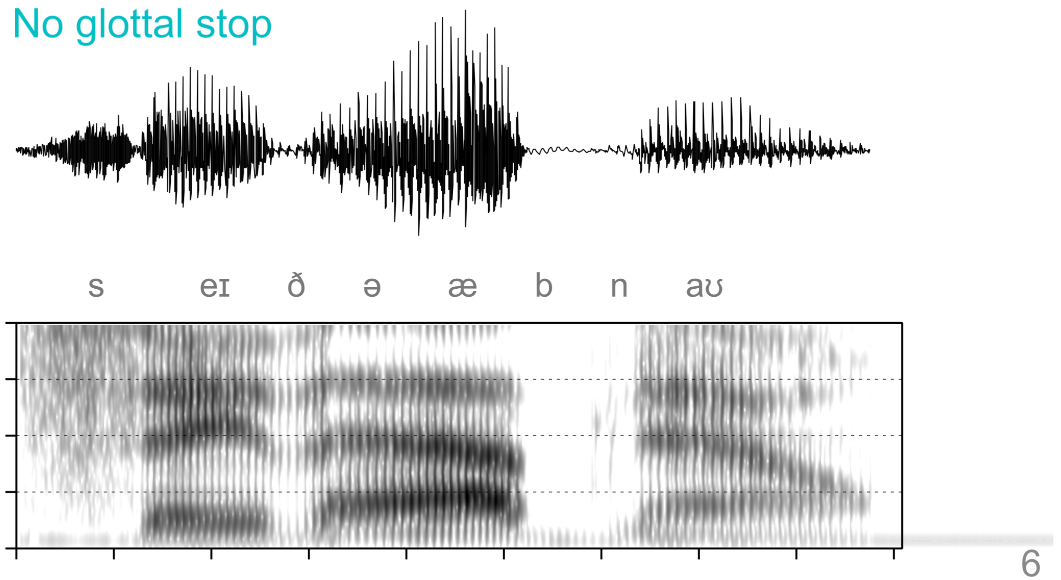
References

[1] Keating, P., Fougeron, C., Hsu, C., & Cho, T. (2003). Domain initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge University Press. **[2]** Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *JPhon*, *35*(4), 445–472. **[3]** Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *JASA*, *117*(6), 3867–3878. **[4]** De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *JASA*, *97*(1), 491–504. **[5]** Garellek, M. (2014). Voice quality strengthening and glottalization. *JPhon*, *45*, 106–113. **[6]** De Jong, K., Beckman, M. E., & Edwards, J. (1993). The Interplay Between Prosodic Structure and Coarticulation. *Language and Speech*, *36*(2–3), 197–212. **[7]** Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *JPhon*, *35*(2), 210–243. **[8]** Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *JML*, *108*, 104034. **[9]** Van Summers, W. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *JASA*, *82*(3), 847–863. **[10]** Mo, Y., Cole, J., & Hasegawa-Johnson, M. (2009). Prosodic effects on vowel production: Evidence from formant structure. *INTERSPEECH*, 2535–2538. **[11]** Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *JPhon*, *24*(4), 423–444.**[12]** De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *JPhon*, *32*(4), 493–516. **[13]** Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *JPhon*, *33*(2), 159–197. **[14]** Kent, R. D., & Netsell, R. (1971). Effects of Stress Contrasts on Certain Articulatory Parameters. *Phonetica*, 24(1), 23–44. **[15]** Bürkner P (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395–411. **[16]** Kim, S., Mitterer, H., & Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLOS ONE*, *13*(8), e0202912. **[17]** Mo, Y. (2011). *Prosody production and perception with conversational speech* [Doctoral Dissertation]. University of Illinois at Urbana-Champaign. **[18]** Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5 Pt 1), 3099–3111. **[19]** Steffman, J., & Jun, S.-A. (2019). Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception. *JASA*, *146*(3), EL251–EL257.

Appendix: stimuli examples







Above Waveforms and pitch tracks for the two conditions in Experiment 1 with ToBI labels.
At right Waveforms and spectrograms for the two conditions used in Experiment 2. Frequency range is 0-4 kHz, ticks on x axis are 100 ms intervals

Appendix: model summaries

Exp 1	β	Error	L95%CI	U95%CI
Intercept	0.05	0.17	-0.29	0.39
prominence	**0.83**	**0.28**	**0.27**	**1.39**
continuum	**-2.57**	**0.28**	**-3.15**	**-2.03**
prom:cont	-0.24	0.13	-0.50	0.01

Exp 2	β	Error	L95%CI	U95%CI
Intercept	**1.18**	**0.16**	**0.87**	**1.50**
prominence	**1.75**	**0.23**	**1.31**	**2.22**
continuum	**-3.35**	**0.17**	**-3.71**	**-3.02**
prom:cont	**-0.79**	**0.20**	**-1.20**	**-0.41**

Above Model summaries for Bayesian mixed effects regression models in both experiments. Random effects in both are by-participant random intercepts and random slopes for all fixed effects. Credible effects (whereby the lower and upper CI in the two rightmost columns exclude zero) are bolded.