WCCFL 38, March 8th 2020

Prosodic prominence in speech perception: the influence of focus structure on the perception of durational and spectral cues

Jeremy Steffman and Sun-Ah Jun

jsteffman@ucla.edu   jun@humnet.ucla.edu

The UCLA Phonetics Lab

Background

- The sound system of a language can be described in terms of…

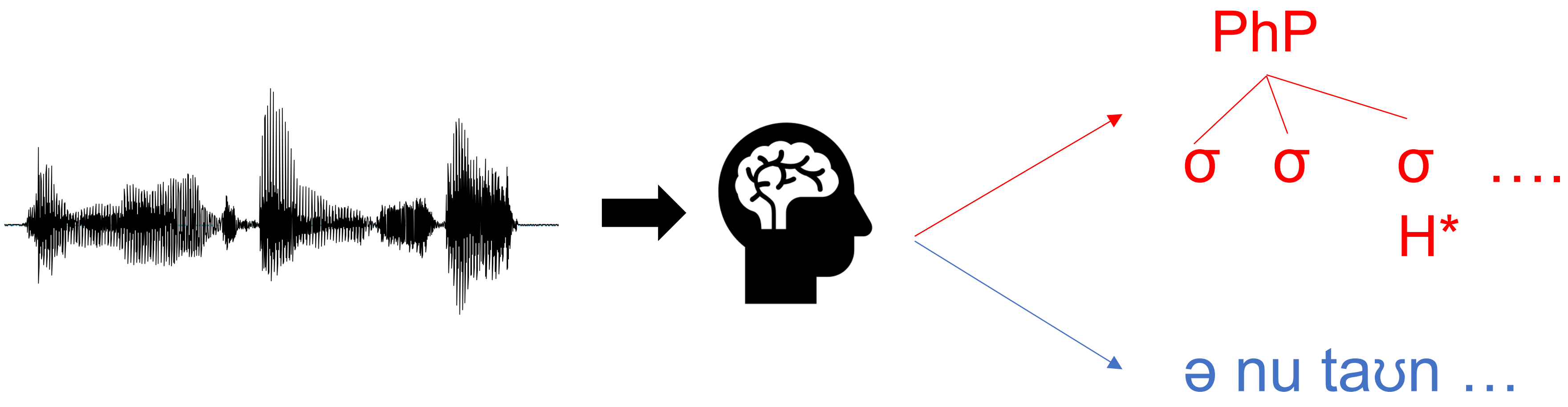
(1) **Segmental structure**: contrastive phonetic content

- represented by features, etc.

(2) **Prosodic structure**: organization of segments into syllables, words, phrases…

Background

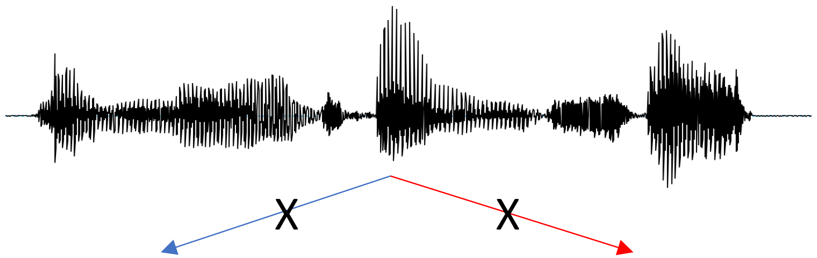
- Listeners evidently extract both from the signal
- Mapping to both types of phonological structure is traditionally assumed to be fairly independent[1]



[1]e.g. Cutler et al. 1997

Background

- One logical possibility: parsing segment and prosody is independent because acoustic cues that specify each in a given language are non-overlapping
- However, this is not the case

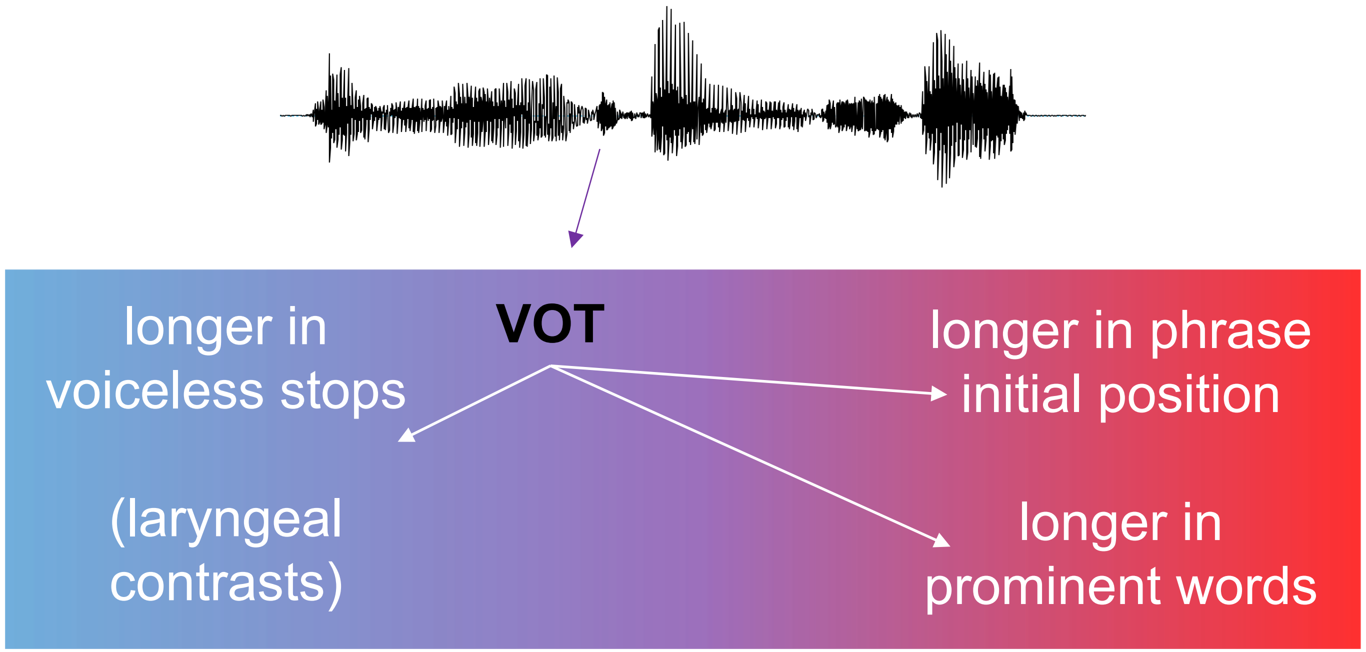


Segmental cues, e.g.
formant structure
VOT
closure duration

Prosodic cues, e.g.
pitch
duration
intensity
voice quality

Background

- A body of phonetic research[e.g. 1-5] suggests…
  - “segmental” cues also encode various prosodic properties
  - “prosodic” cues also encode various segmental contrasts



[1]Cho 2015 [2]2016 [3]Fougeron 1999 [4]Keating 2006 [5]Keating et al. 2003

Background

- Listeners would accordingly benefit from reconciling a cue value with the prosodic context in which it occurs
  - i.e. compensating for prosodic structuring of the signal

- Prosodic boundaries affect segmental categorization in this way[1,2]
  - e.g. longer VOT is required for a voiceless percept…
  - but **even longer VOT** is required when a sound is phrase-initial

  → accounting for prosodic changes in a cue value

- What about prosodic prominence?

[1]Kim & Cho 2013 [2]Steffman 2019

Today's talk

- Today we present evidence that phrasal prominence mediates perception of segmental contrasts in American English, testing
    - a contrast that is cued by formants - vowel categories
    - a contrast that is cued by duration - coda stop voicing

Today's talk

- We manipulate phrasal prominence as cued by the realization of focus in American English
  - the test case: post-focus compression[1-3]

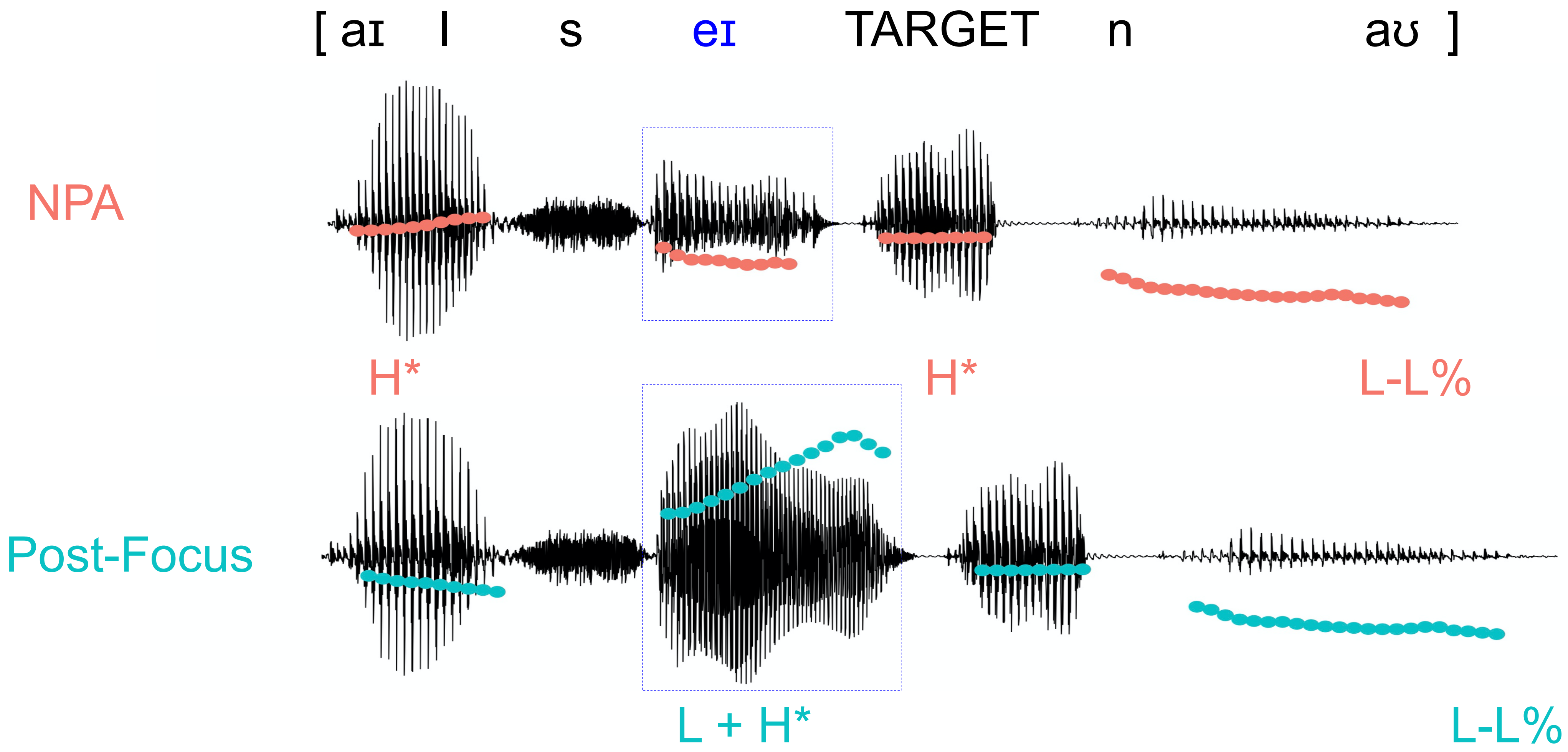
- Words that are focused are:
  - phonologically accented
  - expanded in pitch and duration
  - more sonorous in formant structure (more on this later)
- Words that follow focused material within a phrase are:
  - phonologically de-accented
  - compressed in pitch and duration
  - less sonorous in formant structure

[1] de Jong 2004 [2]Van Summers 1987 [3]Xu & Xu 2005

Manipulating prominence

- Nuclear pitch accent (NPA) condition: I'll say [TARGET] now
  H* H* L-L%
- Post-focus condition: I'll **SAY** [TARGET] now
  L+H* L-L%

Manipulating prominence



Experiment 1: spectral cues

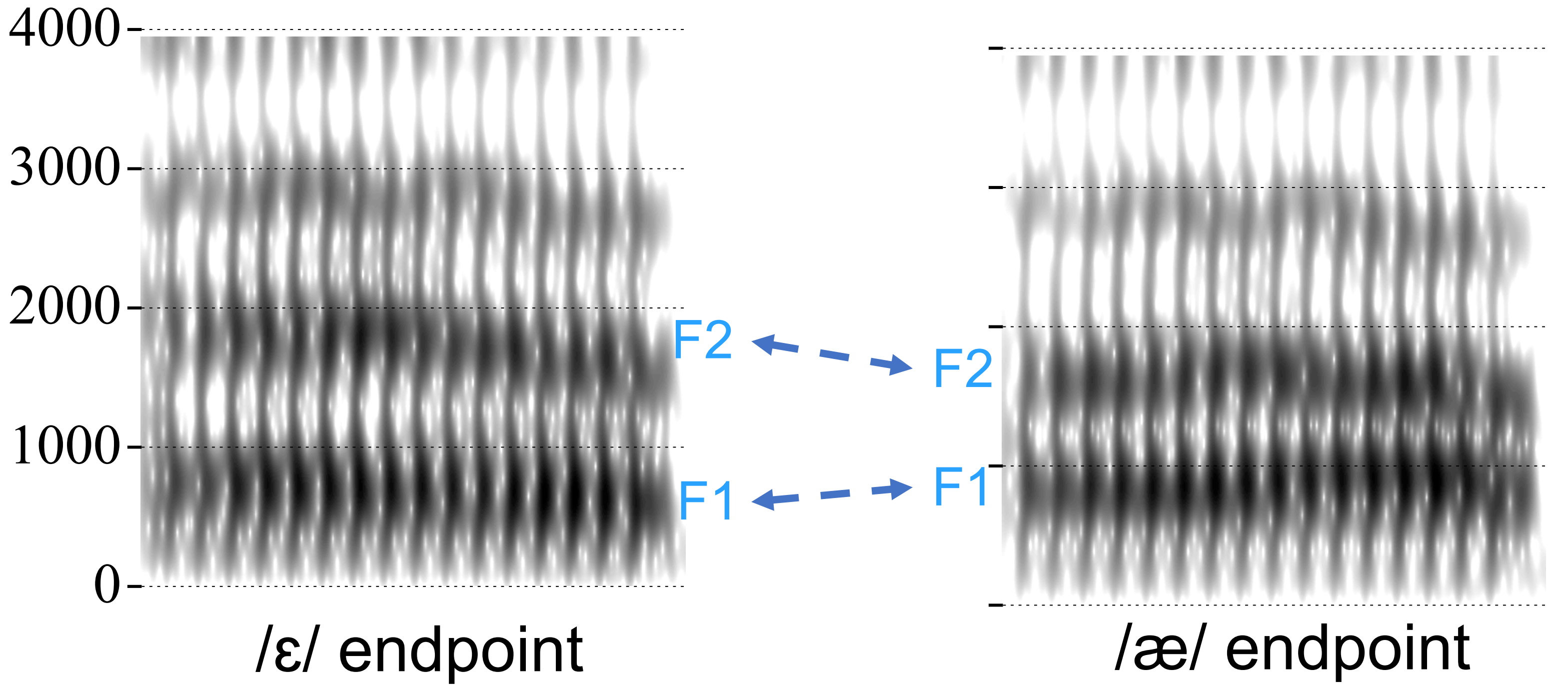
- Phrasal prominence on vowels is marked by phonetic **sonority expansion**[1-3]
  - increased amplitude of jaw movements
  - lowered and back lingual articulations (in non-high vowels)

- An acoustic consequence
  - lower tongue position → raised first formant (F1)
  - more backed tongue position → lowered second formant (F2)

[1] de Jong 1995 [2]de Jong et al. 1993 [3]Van Summers 1987

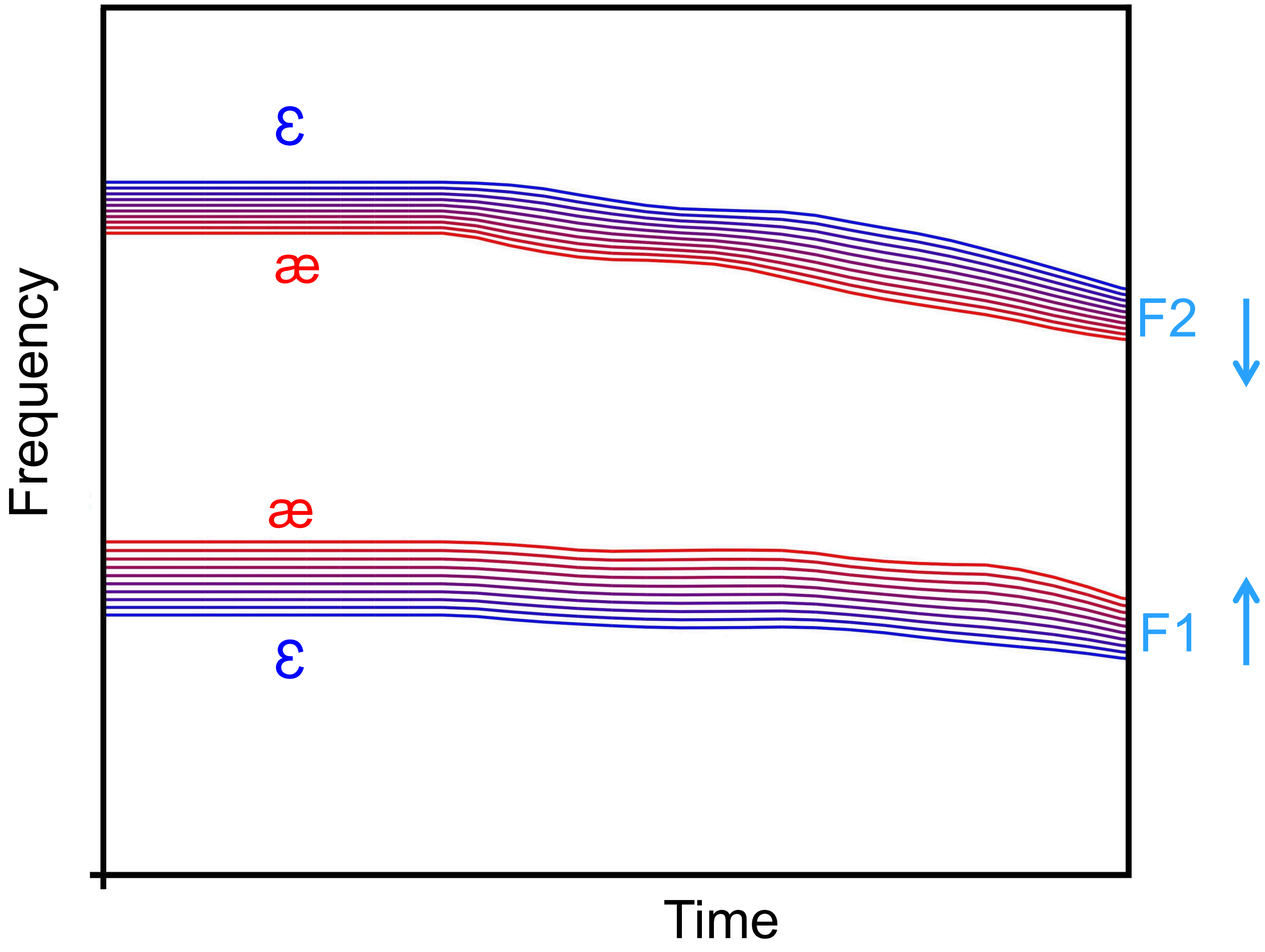
Experiment 1: method

- 2AFC task: participants categorized a target as “ebb” or “ab”
- /ɛ/ - /æ/ varying only in the first and second formant - 10- step continuum
  - ***/ɛ/*** (‘ebb’) has lower F1 & higher F2 than ***/æ/*** (‘ab’)



Experiment 1: continuum

- The continuum varies along…
  - a segmental dimension: vowel height and backness
  - a prosodic dimension: prominence, phonetic sonority in F1/F2



Experiment 1: Predictions

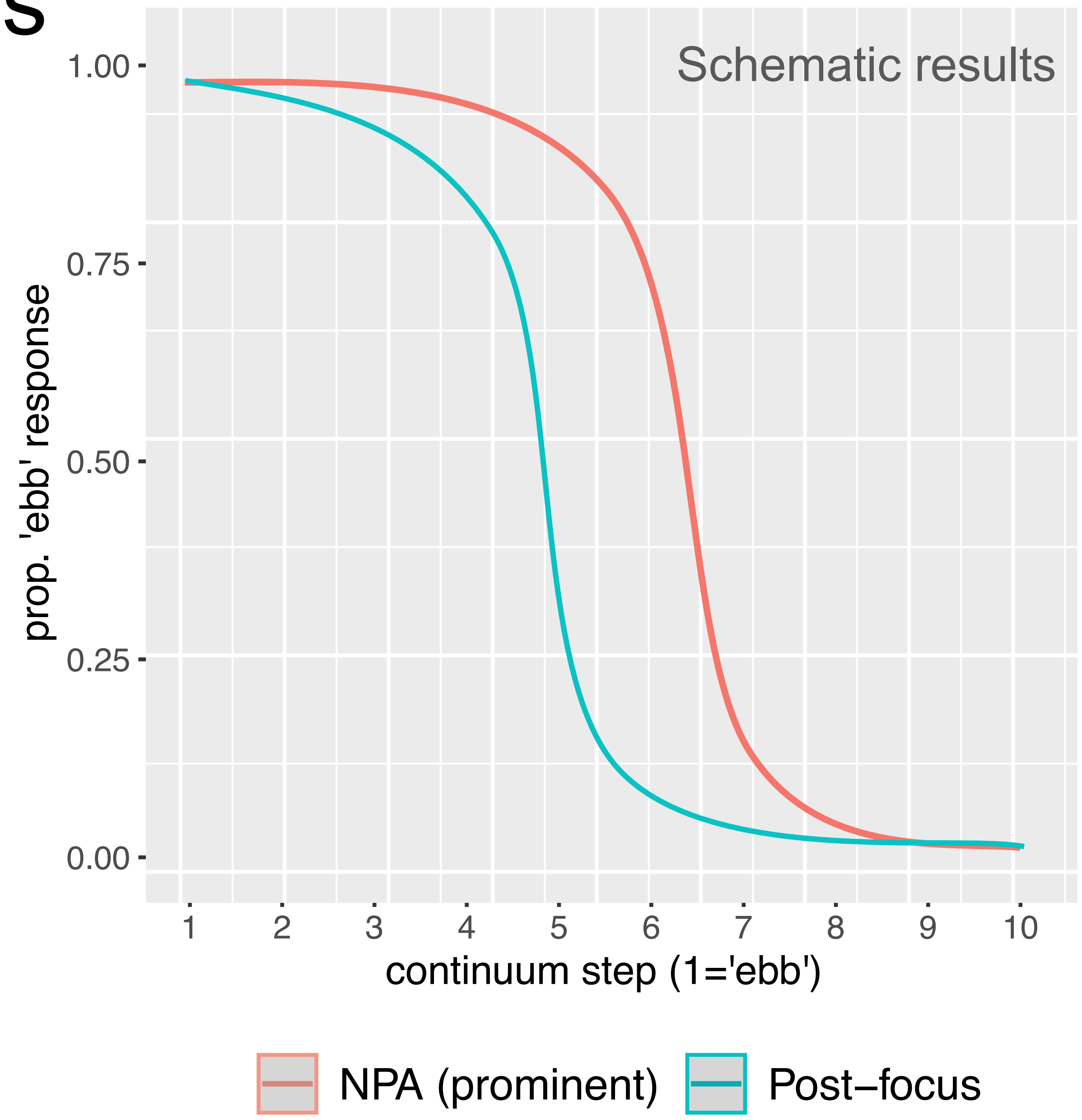
- Accordingly, in prominent contexts, higher F1 and lower F2 could be interpreted as an effect of prominence, not as cuing segmental contrast
- If listeners compensate accordingly, they would categorize more sounds as /ɛ/ in prominent contexts (= NPA condition)
  - i.e. attributing **high** F1 and **low** F2 to prominence, not segment

- (results assess by mixed-effects logistic regression with maximal by subject random slopes)

Experiment 1: Predictions

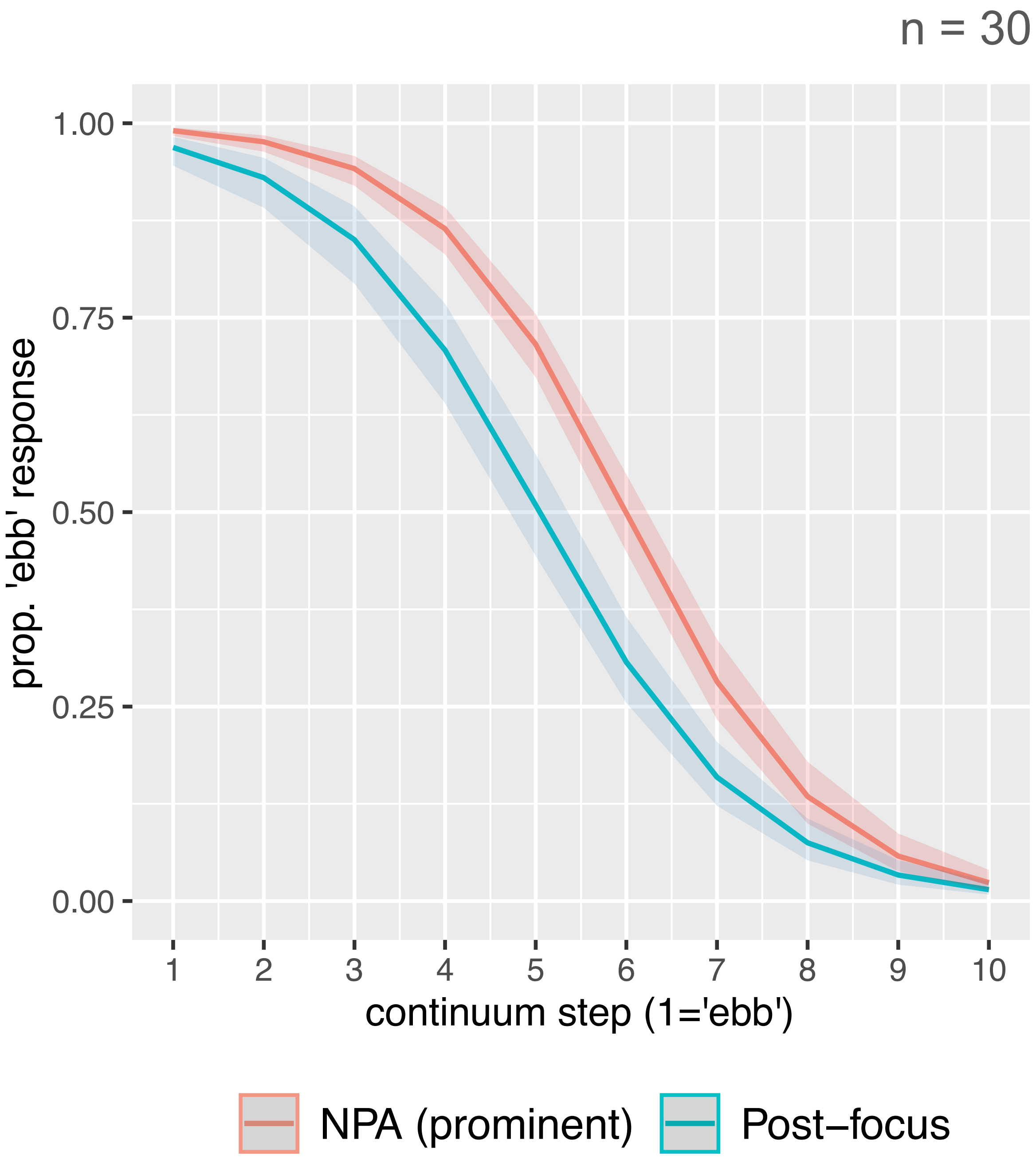
Prediction: **increased** “ebb” responses in the NPA condition

- Visually: the NPA line is above/ right of the Post-focus line



Experiment 1 results

- Model estimates plotted with CI
- As predicted, a prominent (NPA) context shows **increased /ε/ responses** (β = 0.42 z = 3.26)



Interim

- Experiment 1:
  - novel evidence for the involvement of prominence in perception of segmental material

- Experiment 2 goals:
  - replicate the pattern in Experiment 1 with a durational contrast
  - test possible involvement of domain-general effects relevant in the perception of duration

Experiment 2: method

- Recall: post-focus words are **temporally compressed**[1,2]
  - will listeners' perception of duration be modulated accordingly?

- The test case: vowel duration as a cue to coda stop voicing in American English[3,4]
  - vowels are longer before voiced coda stops (which are often devoiced)
  - this is a robust cue to voicing for listeners

- We created a **vowel duration** continuum ranging from "coat" (60ms) to "code" (120ms)

[1] de Jong 2004 [2]Xu & Xu 2005 [3]Chen 1970 [4]Raphael 1972

Experiment 2: temporal cues

- Predictions: in the Post-focus condition
  - overall shorter vowel durations required for a “code” percept, given prosodically driven adjustment of duration
  - compensation for compression would allow for mapping fewer target sounds to “coat” → decreased “coat” responses when Post-focus

Extending Exp 1: we synthesized target pitch to **vary** across conditions:

  - higher in the NPA condition (marking prominence)
  - lowered in the Post-focus condition (de-accentuation)

- Pitch patterns were otherwise the same as Exp. 1

Experiment 2: Psychoacoustic effects

- Perception of duration also influenced by…
  - Adjacent segment durations – perception of durational cue is relative[1]
  - Pitch on a segment – higher pitch perceived as longer[2,3]

	target pitch	**pre-target duration**
NPA	higher pitch (accented)	shorter pre-target duration
Post-focus	lower pitch (deaccented)	longer pre-target duration (accented "say")
Comparison	shorter perceived target duration when Post-focus	
Prediction	increased "coat" (short duration) responses when Post-focus	

[1] e.g. Mitterer et al. 2016 [2]Steffman & Jun 2019 [3]Yu 2010

Experiment 2: Predictions

Psycho-acoustic predictions	**Increased** “coat” responses in the Post-focus condition
Prosodic predictions	**Decreased** “coat” responses in the Post-focus condition

- A third possibility: prosodic effects are limited by target vowel duration
  - Post-focus vowels are **short,** typically < 100 ms[1]
  - Previous work[2,3] suggests prosodic context effects are limited by their mapping to typical context durations
    - i.e. longer durations are too long to be interpreted as de-accented

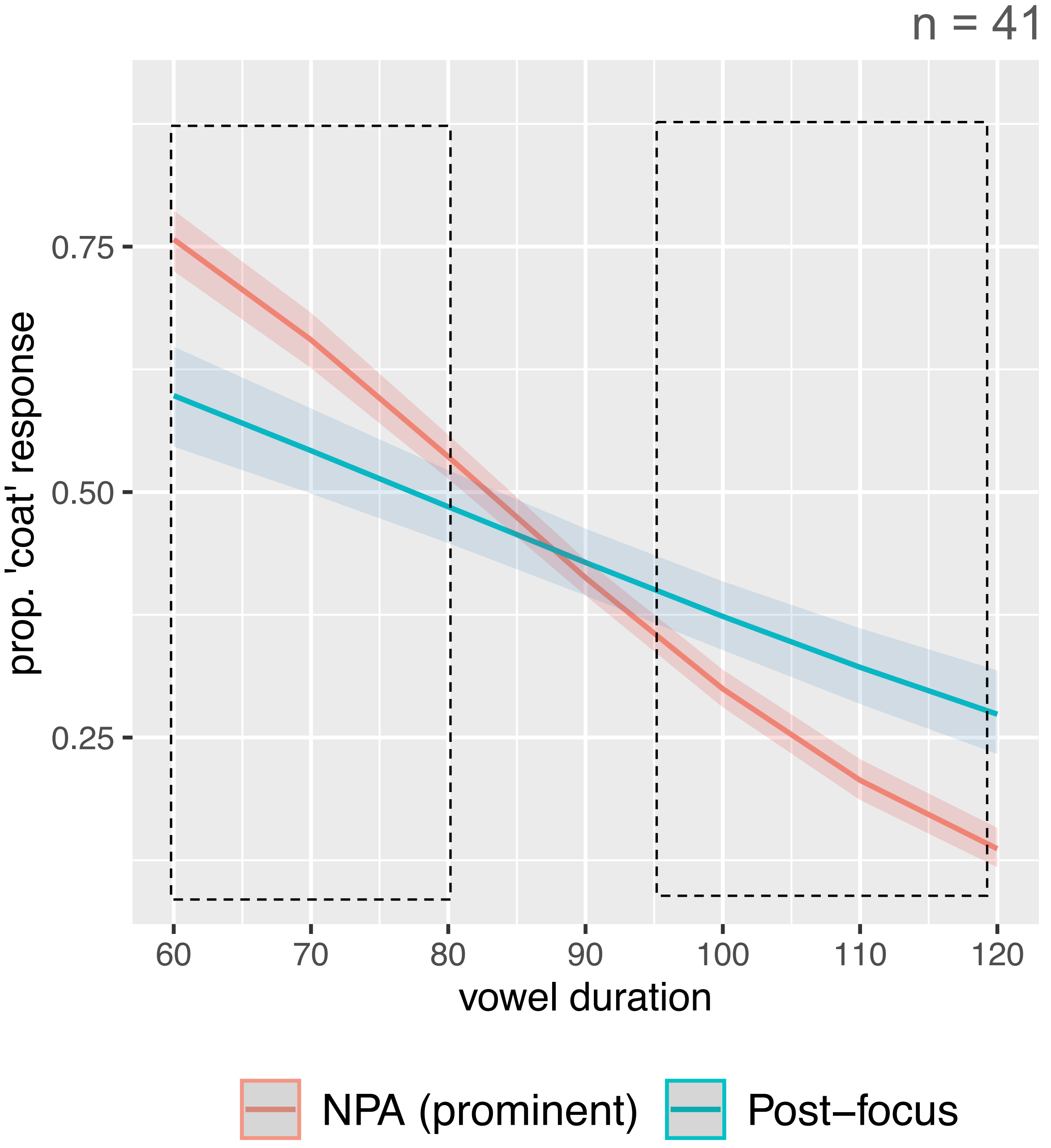
[1] e.g. Greenberg et al. 2003 [2]Steffman 2019 [3]Steffman & Jun 2019

Experiment 2 Results

- Prominence*vowel duration interaction (β = 0.26 z = 8.13)

- At shorter ends of the continuum: **decreased** "coat" responses in the Post-Focus condition
  - prosodic effect

- At longer ends of the continuum: **increased** "coat" responses in the Post-Focus condition
  - psychoacoustic effect



Summarizing Exp 2

- This effect restricted to vowel durations which **map onto those appropriate for a prosodic context**
  - similar findings for prosodic boundary effects[1]

- In cases where other effects compete (duration perception), prosodic effects are mediated by language-typical durational patterning

[1] Steffman 2019

Summing up

- Two test cases show prosodic prominence mediates perception of segmental categories

- Favors a perception/processing model in which both segmental and prosodic structures are extracted in parallel from the speech signal[1-3]

[1]Cho et al 2007 [2]Kim et al. 2018 [3]Mitterer et al. 2019

Further directions

- Questions remain:
  - Are prominence effects categorical, or more gradient?
  - What makes something prominent to listeners?
    - e.g. segmental correlates of prominence such as glottalization

Further directions

- Crosslinguistic comparison: how do different prominence marking systems engender different perceptual outcomes?
  - In the spectral domain:
    - languages vary in the extent to which prominence impacts formant structure[1]
  - In the temporal domain:
    - some languages (e.g. Mandarin[2]) don't exhibit post-focus compression
    - some languages (e.g. Taiwanese[2], Kyungsang Korean[3]) show post-focus **expansion**
- Do perceptual adjustments mirror these patterns?

[1]e.g. Delattre 1969 [2]Xu et al 2012 [3]Jun et al. 2006

Thank you!

Contact us: jsteffman@g.ucla.edu jun@humnet.ucla.edu

- Additional thanks are due to:
  - Adam Royer for recording speech materials
  - Yang Wang, Qingxia Guo, Danielle Bagnas and James Weller for help with data collection
  - The UCLA Phonetics seminar for helpful feedback

References

Chen, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica*, *22*(3), 129–159.

Cho, T. (2015). Language Effects on Timing at the Segmental and Suprasegmental Levels. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 505–529). John Wiley & Sons, Inc.

Cho, T. (2016). Prosodic Boundary Strengthening in the Phonetics–Prosody Interface. *Language and Linguistics Compass*, *10*(3), 120–141.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*(2), 210–243.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, *32*(4), 493–516.

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The Interplay Between Prosodic Structure and Coarticulation. *Language and Speech*, *36*(2–3), 197–212.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, *97*(1), 491–504.

Delattre, P. (2009). An acoustic and articulatory study of vowel reduction in four languages. *IRAL - International Review of Applied Linguistics in Language Teaching*, *7*(4), 295–326.

.

Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics*, 1–80.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics*, *31*(3), 465–485.

Jun, J., Kim, J., Lee, H., & Jun, S.-A. (2006). The prosodic structure and pitch accent of Northern Kyungsang Korean. *Journal of East Asian Linguistics*, *15*(4), 289–317.

Keating, P. (2006). Phonetic Encoding of Prosodic Structure. In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167–186). Macquarie Monographs in Cognitive Science, Psychology Press.

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.

Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, *134*(1), EL19–EL25.

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.

Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics*, *54*, 68–79.

Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language*, *108*, 104034.

Raphael, L. J. (1972). Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *The Journal of the Acoustical Society of America*, *51*(4B), 1296–1303.

Steffman, J. (2019). Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing. *The Journal of the Acoustical Society of America*, *145*(6), EL560–EL566.

Steffman, J., & Jun, S.-A. (2019). Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception. *The Journal of the Acoustical Society of America*, *146*(3), EL251–EL257.

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, *58*(2), 434–445.

Van Summers, W. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America*, *82*(3), 847–863.

Xu, Y., Chen, S., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *Tlir*, *29*(1), 131–147.

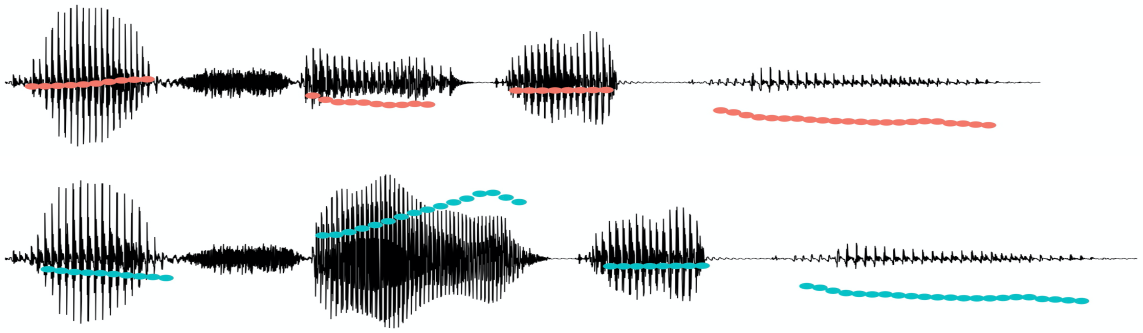
Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.

Yu, A. (2010). Tonal effects on perceived vowel duration. In Cécile Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory Phonology 10*. Walter de Gruyter.

Appendix

Duration effects in Exp 1?

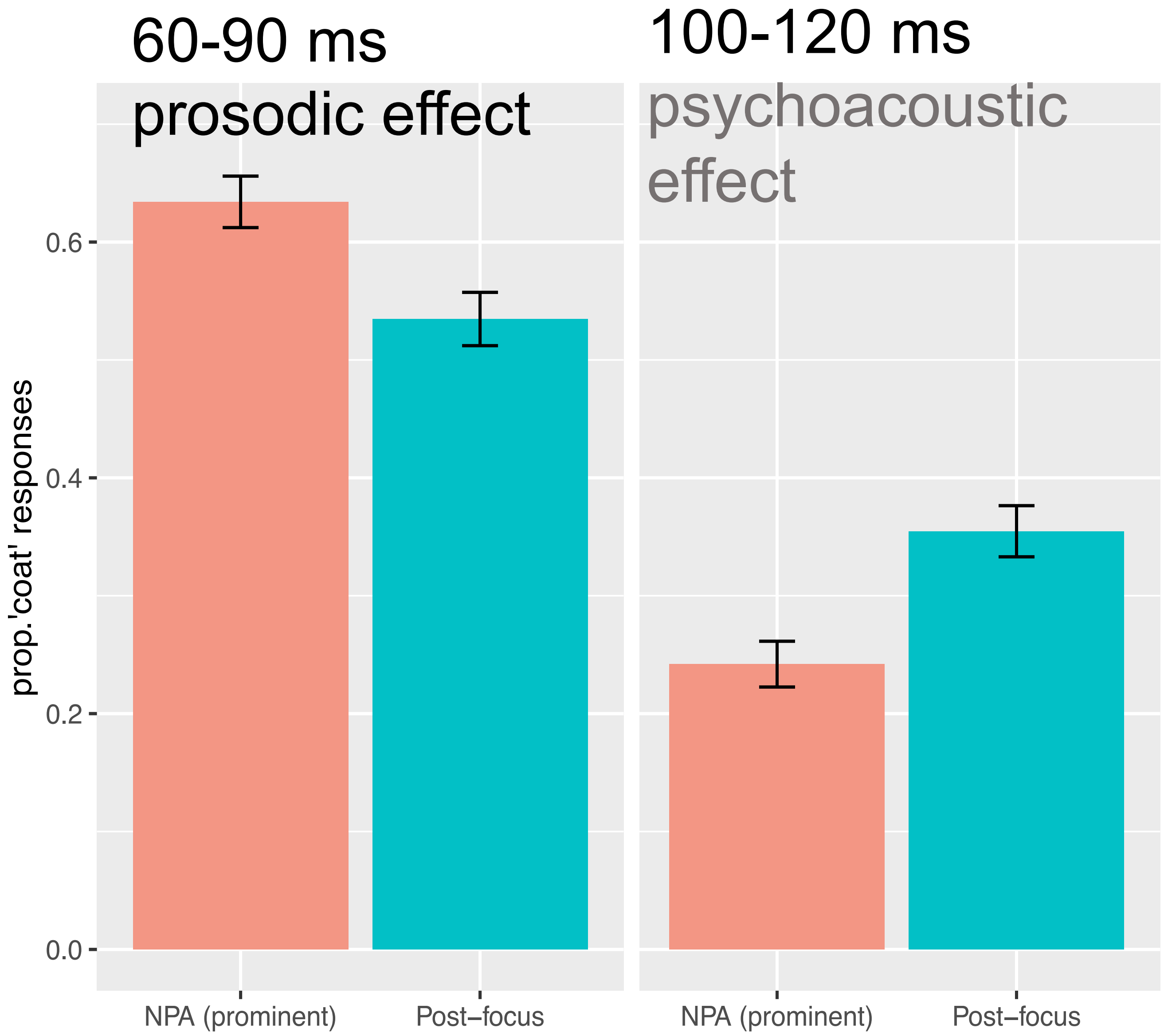
- Note: The /ɛ/ - /æ/ contrast is **also** durational - /æ/ is longer[1]
  - how would this relate to durational contrast perception?
  - recall: longer pre-target duration in Post-focus condition



Psycho-acoustic predictions	**shorter** perceived target sound – **increased** /ɛ/ responses in the Post-focus condition
Prosodic predictions	compensation for sonority expansion - **increased** /ɛ/ responses in the NPA condition – found in Exp. 1

[1]e.g. Umeda 1975

Barplots Exp. 2



Exp 1 model

	β	SE	z	p
Intercept	0.04	0.15	0.235	0.81
continuum	-2.55	0.25	-10.09	< 0.001
prominence	0.42	0.13	3.26	< 0.01
cont : prom	-0.11	0.05	-2.19	< 0.05

Exp 2 model

	β	SE	z	p
Intercept	-0.32	0.06	-5.46	< 0.001
continuum	-0.73	0.08	-9.208	< 0.001
prominence	0.03	0.09	0.34	0.72
cont : prom	0.26	0.03	8.215	< 0.001

Exp 2 interaction (emmeans)

Step (ms)	est.	SE	z-ratio	p
60	0.74	0.20	3.59	<0.01
70	0.47	0.19	2.44	0.01
80	0.20	0.19	1.1	0.27
90	-0.06	0.18	-0.35	0.72
100	-0.33	0.18	-1.78	0.07
110	-0.59	0.19	-3.06	<0.01
120	-0.86	0.21	-4.14	<0.001