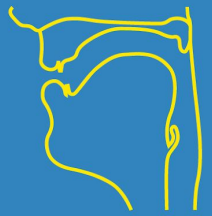
Intonational structure influences perception of contrastive vowel length: the case of phrase-final lengthening in Tokyo Japanese

Hironori Katsuda & Jeremy Steffman
Speech Prosody 2020

The UCLA Phonetics Lab



Background

- When listeners process speech, they need to extract both **segmental** and **prosodic** information from the speech signal

- Sometimes the acoustic information which specifies these segmental and prosodic structures may be the same, e.g.
  - Segmental: VOT duration[1] for /p/ vs. /b/
  - Prosodic: VOT duration is longer phrase-initially (initial strengthening[2,3]**)**
  - “Phonetic encoding of prosodic structure" - a factor in speech perception?

[1]Abramson & Lisker, 1970, [2]Keating et al., 2003, [3]Cho 2016, [4]Cho & Keating, 2009

Background

- Are listeners sensitive to such prosodically driven variation when they perceive segmental categories?

- Recent research on domain-initial prosodic effects (initial strengthening) in speech perception[5,6,7] suggests this may be the case

[6]Mitterer, Cho & Kim, 2016, [7]Mitterer, Kim & Cho, 2019

Past study: Kim & Cho (2013)

- Listeners categorized a VOT continuum as /p/ or /b/ (“pa” or “ba”).
- Prosodic position was manipulated (boundary tone + final lengthening)

  Medial: [*Let’s hear **X** again*]L-L%

  Initial: [*Let’s hear*]L-L% [***X** again*]L-L%

- Listeners required longer VOT for a /p/ response in the initial condition.
- Evidence listeners compensate perceptually for initial strengthening.

Past study: Kim & Cho (2013)

- Changing duration (i.e. manipulating final lengthening) in the context of *durational* cues may influence their perception
  - i.e., speech rate normalization offers another possible explanation the shift Kim & Cho find
  - Longer VOT required for /p/ following lengthened pre-target material

- One proposed solution: manipulate prosodic structure without manipulating duration, we accordingly manipulate only pitch
  - Psychoacoustic pitch/duration effects are also not an issue (see our paper for more information)

The present study

- This study tests the influence of **final lengthening**[8,9,10] on perception of **vowel length contrasts** in Tokyo Japanese.
  - Segmental perception in phrase-final position remains relatively less studied.
  - Only pitch cues (i.e., without temporal cues) are manipulated to influence perceived prosodic structure.

[8]Cho, 2015, [9]Cho, 2016, [10]Turk & Shattuck-Hufnagel, 2007

Tokyo Japanese

- (Lexical) pitch accent: (e.g., *a'me* "rain" vs. *ame* "candy")
- Vowel length contrast (e.g., *shi'sho* "librarian" vs. *shi'shoo* "master")
- Final lengthening[11,12,13,14]
  - Unaccented disyllabic words exhibit greater final lengthening than disyllabic words with the initial pitch accent[14] → prominence suppresses final lengthening[15]

[11]Takeda,Sagisaka & Kuwabara, 1989, [12]Ueyama, 1999, [13]Shepherd, 2008, [14]Seo, et al., 2019, [19]Turk & Shattuck-Hufnagel, 2007

Japanese intonation

We adopt the Autosegmental-Metrical (AM) model of Japanese intonational phonology[16,17,18,19]

- Accentual phrase (AP) and Intonational phrase (IP): both marked by L%
- Pitch accent: “A” in XJ-ToBI[19], at most one pitch accent in AP
- %L: IP-initial, H-: second mora of AP (if no pitch accent)

[16]Beckman & Pierrehumbert, 1986, [17]Pierrehumbert & Beckman, 1988, [18]Venditti, 1995, [19]Maekawa, Kikuchi, Igarashi, & Venditti, 2002

Experimental design

- 2AFC identification task

- Frame sentence:

Wata'shitachi-wa	***X***	*de'sukara*	*shinraideki-ma'su*
We-TOP	**X**	because/therefore	reliable-be

- Prosodic conditions (AP boundary {...}, IP boundary [...]):
  (IP-) Medial: [{*Wata'shitachi-wa*} {***X*** *de(')sukara*} {*shinraideki-ma'su*}]
  "Because we are X, (we are) reliable"
  (IP-) Final: [{*Wata'shitachi-wa*} {***X***}] [{*de'sukara*} {*shinraideki-ma'su*}]
  "We are X. Therefore (we are) reliable."

Experimental design

- Target words:
  Exp. 1: Accented pair: *shi'sh**o*** “librarian” 司書 or *shi'sh**oo*** “master” 師匠
  Exp. 2: Unaccented pair: *dooky**o*** “housemate” 同居 or *dooky**oo*** “townmate” 同郷

  Vowel duration continuum: 60-180 ms (7 steps) = /o/ or /oo/

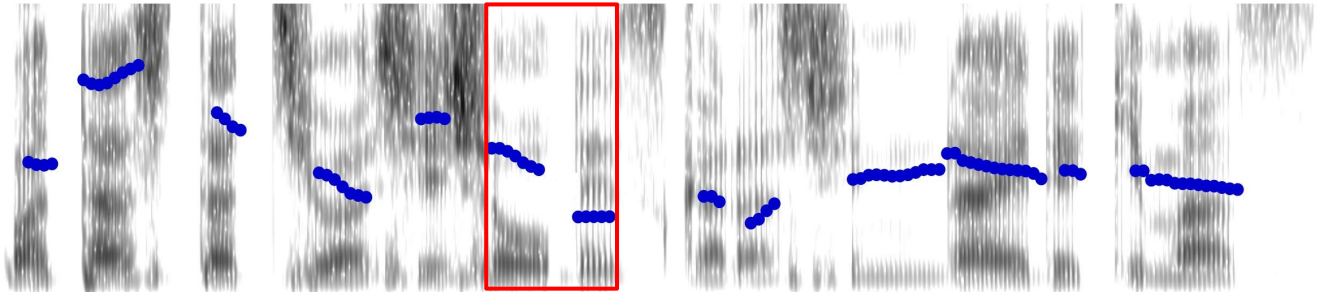
- Participants: 26 speakers of Tokyo Japanese for each experiment

- Results assessed with mixed-effects logistic regression

Stimuli - Experiment 1: accented target

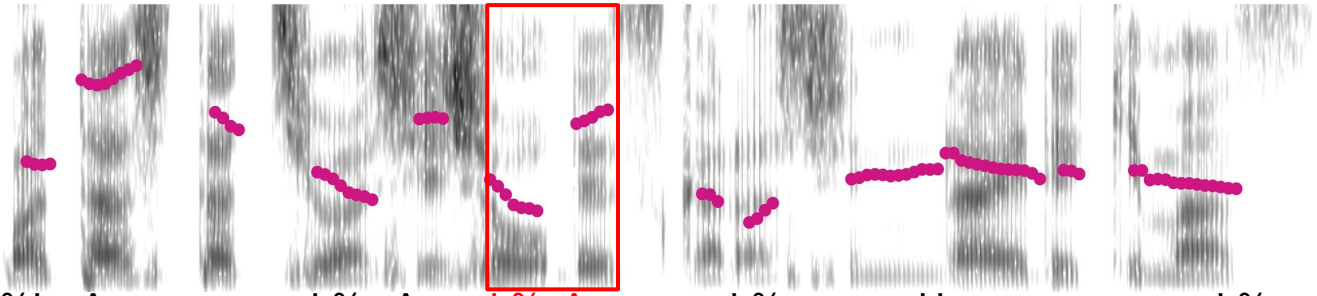
Medial (1 IP)

- Higher pitch on the target
- Accent on /de/ is reduced[20,21,22]



Final (2 IPs)

- Lower pitch on the target (L%)
- Accent on /de/ is realized

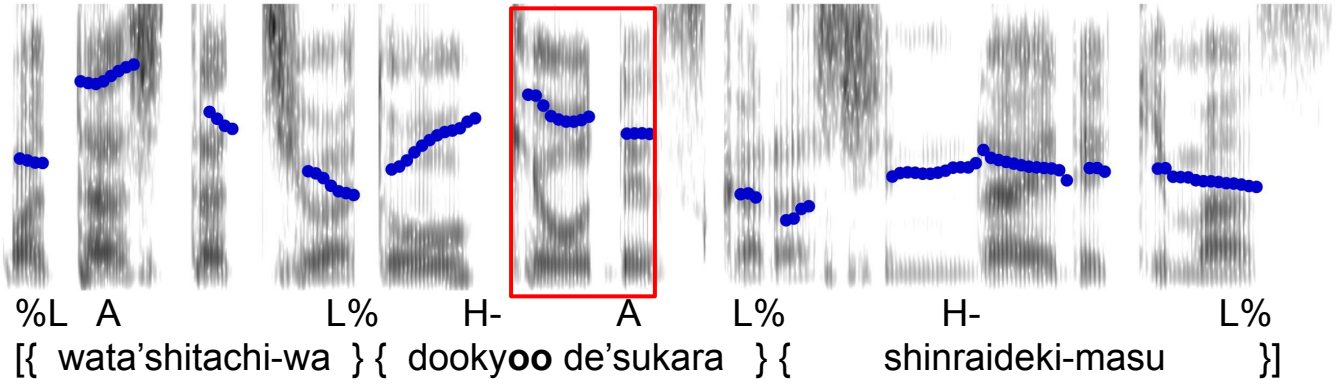


[20]Poser, 1984,[21]Kubozono, 1987, [22]Maekawa, 1994

Stimuli - Experiment 2: unaccented target

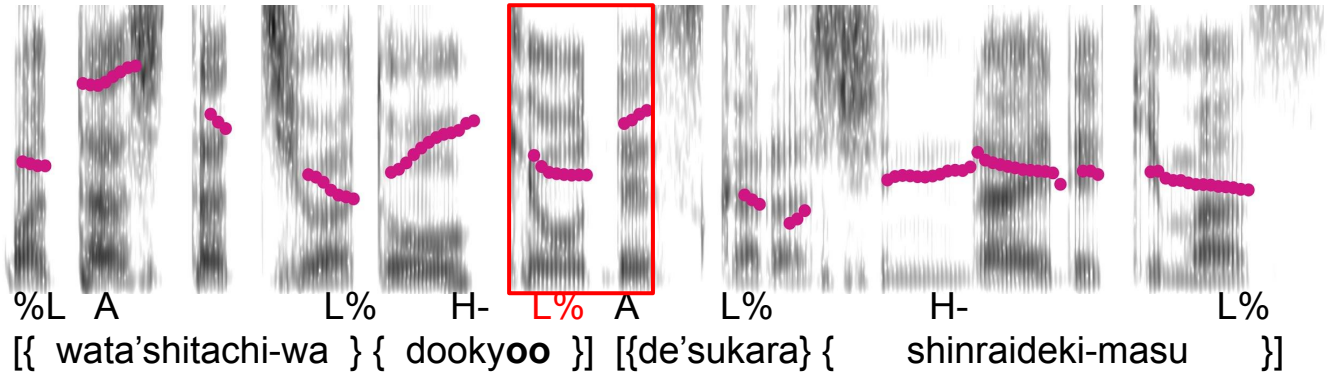
Medial (1 IP)

- Higher pitch on the target
- Lower pitch on /de/



Final (2 IPs)

- Lower pitch on the target (L%)
- Higher pitch on /de/ (pitch reset)



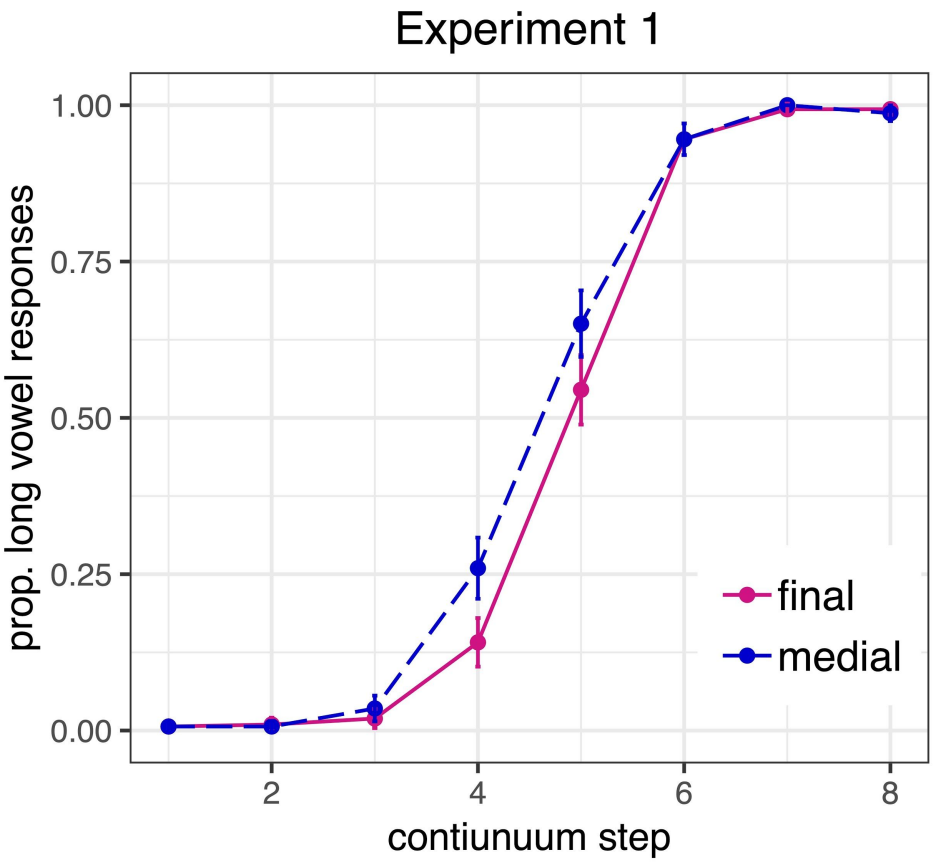
Predictions

1. A phrase-final target should overall require ***longer*** vowel duration for a long vowel percept
   → ***decreased*** long vowel responses in the final condition

2. The unaccented pair (Exp. 2) exhibits a greater effect of position
   - Following Seo, et al. (2019): unaccented words exhibit greater final lengthening
   - Listeners may accordingly *shift categorization more* for unaccented words?
   - i.e., more lengthening leading to more of a difference between final/medial conditions in unaccented words → a larger effect

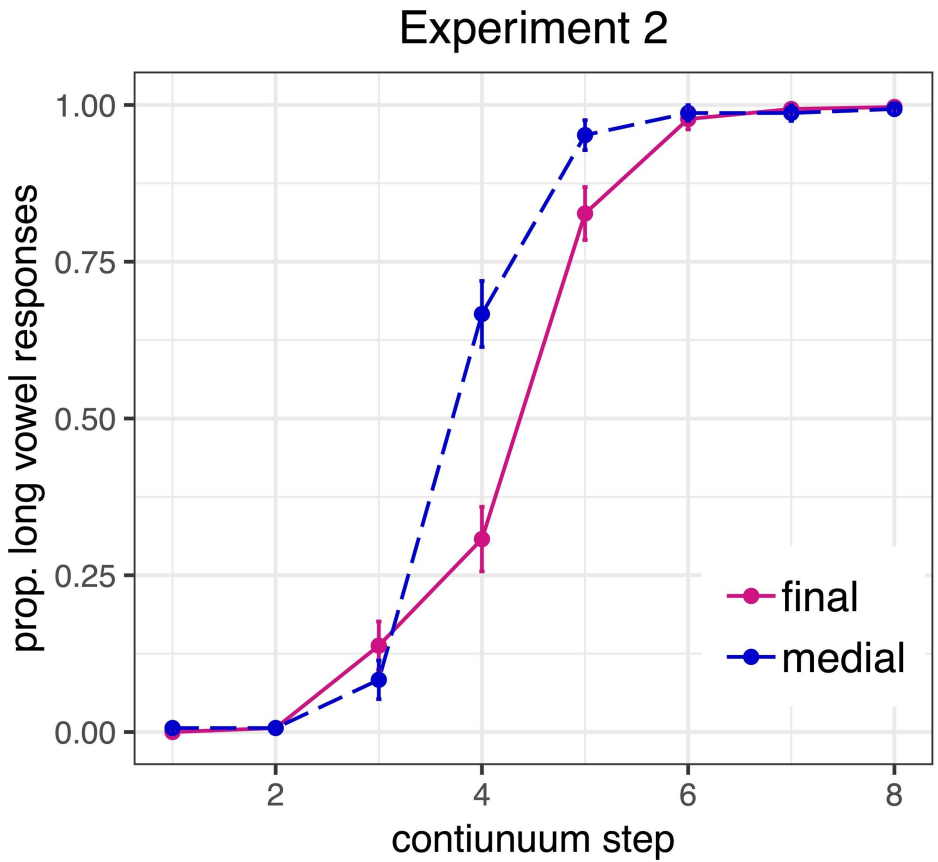
Results - Experiment 1

- Main effect of position
  (β = -0.39, z = -2.84, p < 0.01)
  - Final position *decreases* long vowel responses
  - Supports predictions: a phrase-final target is expected to be relatively long



Results - Experiment 2

- Main effect of position (β = -0.66, z = -6.42, p < 0.001)
  - Replicates the effect in Exp.1
- The effect (β = -0.66, SE = 0.10) is larger than Exp. 1 (β = -0.39, SE = 0.14)
  - Consistent with the observation that unaccented words exhibit greater final lengthening[18]



[18]Seo, et al., 2019

General discussion

- Listeners are sensitive to intonational structure in their perception of contrastive vowel length
  - A phrase-final target required longer vowel duration to be categorized as long
  - Pitch cues alone suffice to influence perceived prosodic structure
  - Consistent with the idea that listeners process segmental/prosodic info in parallel[23,24]

- Unaccented target words (Exp. 2) exhibited a larger positional effect.
  - Consistent with differences found between accented and unaccented words in speech production[18]
  - May suggest that listeners incorporate prominence-boundary interactions in their sensitivity to phrasal positional effects

[23]Kim, Mitterer & Cho, 2018, [24]Mitterer, Kim & Cho, 2019, [18]Seo, et al. 2019

Further directions

- Directly compare accented and unaccented pairs in terms of prosodic modulation

- Recent studies[23,24] suggest that the effects of prosodic structure come relatively late in processing (Later-stage prosodic modulation of lexical competition)
- A promising extension would be to test the time-course of the observed effects with eye-tracking

[23]Kim, Mitterer & Cho, 2018, [24]Mitterer, Kim & Cho, 2019

Acknowledgement

- This research was funded by the UCLA Ladefoged Scholarship.

- Many thanks are due to Sun-Ah Jun, Megha Sundara, Pat Keating and members of the UCLA Phonetics lab for valuable feedback on this project.

- Further thanks to Shigeto Kawahara, Mami Gosyo, and Naoki Ishikawa for recruitment assistance.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences* (Vol. 196, No. 7, pp. 569-573). Academia Prague.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(01), 255–309.

Brigner, W. L. (1988). Perceived Duration as a Function of Pitch. *Perceptual and Motor Skills*, *67*(1), 301–302.

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., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*(4), 466–485.

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.

Kubozono, H (1987). The organization of Japanese prosody. Ph.D. Dissertation, University of Edinburgh.

Maekawa, K. (1994). Is there 'dephrasing' of the accentual phrase in Japanese? Working Papers in Linguistics: Papers from the Linguistics Laboratory 44, 146-165.

Maekawa, K., Kikuchi, H., Igarashi, Y. & Venditti, J. (2002). X-JToBI: An extended J_ToBI for spontaneous speech. In Proceedings of the 7th International Congress on Spoken Language Processing (pp. 1545-1548).

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.

Pierrehumbert, J. and Beckman, M. E., *Japanese Tone Structure.* Cambridge, MA: MIT Press, 1988.

Poser, W. (1984). The phonetics and phonology of tone and intonation in Japanese. PhD. Dissertation, MIT.

Seo, J., Kim, S., Kubozono, H., & Cho, T. (2019). Preboundary lengthening in Japanese: To what extent do lexical pitch accent and moraic structure matter? *The Journal of the Acoustical Society of America*, *146*(3), 1817.

Shepherd, M. A. (2008). The scope and effects of preboundary prosodic lengthening in Japanese. USC Working Papers in Linguistics 4, 1-14.

Steffman, J., & Jun, S. A. (2020). *Prosodic cues facilitate speech rate normalization: Exploring listener sensitivity to prosody in speech perception*. The 94th annual meeting of Linguistic Society of America, New Orleans, LA.

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.

Takeda, K., Sagisaka, Y. & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. The Journal of Acoustical Society of America 86, 2081-2087.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472.

Ueyama M. (1999). An experimental study of vowel duration in phrase-final contexts in Japanese. UCLA Working papers in Phonetics 97, 174-182.

Van Dommelen, W. (1993). Does Dynamic F0 Increase Perceived Duration? New Light on an Old Issue. *Journal of Phonetics*, *21*(4).

Venditti, J. J. (1995). Japanese ToBI Labelling Guidelines. Ohio State University Working Papers in Linguistics 50, 127-162.

Yu, A. (2010). Tonal effects on percieved vowel duration. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory Phonology 10* (pp. 151–168). Walter de Gruyter.

Appendix

Psychoacoustic effects?

- Pitch can influence perceived duration[25,26]
- However, this has been shown to be restricted to monosyllabic stimuli (i.e. no carrier phrase)[27,28]
- Prosodic factors have also been shown to override these effects[28,29]
- In using a full carrier phrase these effects are unlikely to play a role in our stimuli
    - This could be confirmed by e.g. a duration rating task with these stimuli

[25]Brigner 1988, [26]Yu 2010, [27]Van Dommelan 1993, [28]Steffman & Jun 2020, [29]Steffman & Jun 2019

Model for Experiment 1

	β(SE)	z	p
(Intercept)	-0.65(0.26)	-2.57	0.01
position	-0.39(0.14)	-2.84	< 0.01
step	6.08(0.34)	17.78	< 0.001
position:step	0.43(0.23)	1.86	0.06

Model for Experiment 2

	β(SE)	z	p
(Intercept)	1.25(0.23)	5.54	< 0.001
position	-0.66(0.10)	-6.42	< 0.001
step	6.08(0.37)	16.49	< 0.001
position:step	-0.41(0.20)	-2.01	< 0.05