Contrast enhancement and the distribution of vowel duration in Japanese

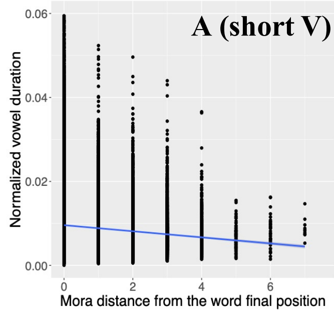
Shinichiro Sano[1], Céleste Guillemot[2]

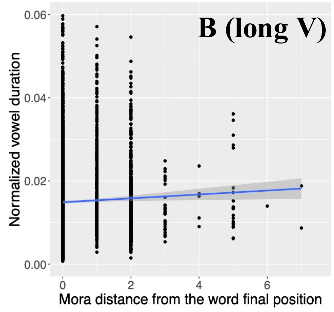
[1, 2]Keio University (Japan)

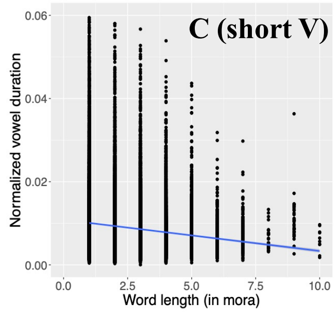
shinichirosano@gmail.com, celesteguillemot@gmail.com

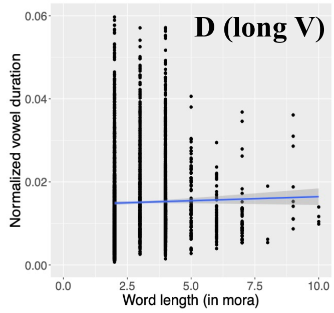
Background: Previous research has shown that patterns in phonetic implementation of segments and subsegmental features are controlled by information-related factors [1-5]. Phonetic cues that contribute less information are more prone to undergo reduction or neutralization [1,2,5]. This is illustrated by the crosslinguistic tendency for phonological processes involving neutralization to exhibit a preference for word-ends over beginnings [5,6]. On the other hand, words with low predictability tend to be longer [7], and their segments need to convey more disambiguating information [8].

Vowel duration provides useful test cases to consider the role of information: [9,10] demonstrated that, in English, more predictable or less informative vowels are shorter. In Japanese, in which vowels contrast in length (short vs. long, [11,12]), [13] show that preceding consonant and information-related measures (Surprisal and Entropy) play a role in variations of vowel duration at the sub-phonemic level. Building upon previous research, this paper focuses on the vowel length contrast to examine (i) how the positional bias and difference in intonation phrase (IP)/word length is reflected in the distribution of vowel duration in Japanese, (ii) how the distribution differs depending on the type of linguistic unit, and (iii) how these are related to the vowel length contrast.









Method: Data were retrieved from the CSJ-RDB (Corpus of Spontaneous Japanese – Relational Database, National Institute for Japanese Language and Linguistics 2012), among which the present study targeted 12 speech samples. An exhaustive search of the data in the CSJ- RDB resulted in 44,219 tokens, of which 40,703 (92%) were short vowels and 3,516 (8%) were long vowels, where tokens with filled pauses, word fragments, and other non-linguistic events (e.g., laughter) were excluded. The duration of each token was analyzed in terms of position in IP and word, and length of IP and word. All distributional skews discussed below were tested by the linear mixed-effects model using *lmer* of the lmerTest package in R [14]. We fit separate models for short and long vowels. In the models, response variable was duration of short/long vowels normalized by speech rate (duration of IP divided by the number of moras); we included factors of interests (position in IP/word and length of IP/word) and other control variables (e.g., kinds of vowels, accented or not); random intercepts for speaker and item (lemma) and by-speaker and by-item random slopes were also included in the model.

Results and discussion: [Position] We measured positions of vowels in IP by word distance from the IP-final position and in word by mora distance from the word-final position. At the IP level, short and long vowels showed the same pattern: duration is longer at more back positions than more front positions (short V: $t = -11.309$, $p < 0.01$, long V: $t = -3.7$, $p < 0.01$). This can be attributed to the effect of final lengthening, which occurs in utterance-final and phrase-final position, but almost never in word-final position in Japanese [15,16]. At the word level, however, short and long vowels showed different patterns. At more front positions (rightward in A and B), short vowels become shorter ($t = -11.348$, $p < 0.01$), while long vowels become longer ($t = 2.954$, $p < 0.01$), resulting in a larger durational gap between them that provides enhanced cues for short vs. long contrast. This suggests contrastive hyperarticulation at informationally salient positions. At more back positions (leftward), short vowels are longer, while long vowels are shorter, making the durational difference between them closer. As a result, the durational distinction is more likely to be neutralized, which is consistent with the fact that back positions are informationally non-salient [5]. **[Length]** We measured IP length by word count and word length by mora count. At the IP level, short and long vowels showed the same pattern. The duration is longer when the IP is shorter (short V: $t = -3.352$, $p < 0.01$, long V: $t = -2.566$, $p <$

0.01). This may be due to physiological reasons: a limitation of breath entails a limited IP duration; hence when an IP is longer, each segment becomes shorter (cf. Respiratory Code for f0; [17,18]). On the contrary, at the word level, distinct patterns were again observed in short and long vowels, as C and D illustrate. In longer words, short vowels become shorter (t = -4.900, $p < 0.01$), while long vowels become longer (t = -3.011, $p < 0.01$), making the durational distance between short and long vowels greater, that is, enhanced cues for short vs. long contrast in longer words. This may be due to the lexical distribution of shorter and longer words. In token frequency, shorter words are more frequent than longer words (59.8% and 59.2% of all words are less than two (for short V) and three moras (for long V)). Since shorter words are more frequent and predictable, phonetic signal in these words tend to be phonetically reduced (probabilistic reduction, [4,9]), while longer words are less frequent and less predictable, and therefore phonetic signal in these words should be enhanced ([8]). In type frequency, however, longer words are more frequent than shorter words (75.8% and 61.4% of all words are more than three moras (for short V) and four moras (for long V)). With more lexical competitors, in longer words the predictability with which a target segment is identified becomes lower, and thus requires the phonetic signal to be more informative or salient to differentiate the target from other competitors.

The results suggest that, at the word level, duration is effectively controlled (enhanced cues for salient positions and words with less predictability or more competitors, and reduced cues for non-salient positions and words with more predictability or less competitors) to give appropriate degree of speech signal to balance the successful transmission of lexical information and the cost for phonetic implementation. However, this is not the case with IP, which contributes to sentence-level information (e.g., intonation). In addition, hyperarticulation (reduction or enhancement) not only targets a particular linguistic unit independently, making it shorter or longer, but also a contrast in such a way as to increase the durational distance between contrasting segments.

References

[1] Hall K. C., Hume E., Jeager F. & Wedel A. (2016). *The message shapes phonology*. ms.

[2] Hall K. C., Hume E., Jeager F. & Wedel A. (2018). The role of predictability in shaping phonological patterns. Linguistics Vanguard, 4(S2), 20170027.

[3] Cohen-Priva U. (2015). Informativity affects consonant duration and deletion rates. *LabPhon*, 6(2), 243-278.

[4] Turnbull R. (2018). Effects of lexical predictability on patterns of phoneme deletion/reduction in conversational speech in English and Japanese. Linguistics Vanguard, 4(S2), 20170033.

[5] Wedel A., Ussishkin A. & King A. (2019). Crosslinguistic evidence for a strong statistical universal: Phonological neutralization targets words-ends over beginnings. Language, 95(4), e428-e446.

[6] Houlihan K. (1975). *The role of word boundary in phonological processes*. Doctoral dissertation. The University of Texas at Austin.

[7] Zipf G. K. (1935). The psycho-biology of language. Boston: Houghton Mifflin.

[8] King A., & Wedel, A. (2020). Greater Early Disambiguating Information for Less-Probable Words: The Lexicon Is Shaped by Incremental Processing. *Open Mind: Discoveries in Cognitive Science*, 1-12.

[9] Aylett M. & Turk A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31-56.

[10] Aylett M. & Truk A. (2006). Language redundancy predicts syllable duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America,* 119(5), 3048-3059.

[11] Vance T. (1987). *An introduction to Japanese phonology*. New York: SUNY Press.

[12] Vance T. (2008). *The sounds of Japanese.* Cambridge: Cambridge University Press.

[13] Shaw J. & Kawahara S. (2017). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*, 62, 80-114.

[14] R Core team (2019). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

[15] Hoequist C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, 40(3), 203-237.

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

[17] Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Proceedings of Speech Prosody* 2002, 47-57.

[18] Gussenhoven C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, 8(2), 425-434.