LOGARITHMIC DURATIONS FOR CLASSIFYING AND PREDICTING JAPANESE SHORT AND LONG VOWELS

Shigeaki Amano[1], Yukari Hirata[2], Kimiko Yamakawa[3]

[1]Aichi Shukutoku University, [2]Colgate University, [3]Shokei University
[1]psy@asu.aasa.ac.jp, [2]yhirata@colgate.edu, [3]jin@shokei-gakuen.ac.jp

ABSTRACT

Raw duration has conventionally been used as a time-related acoustic property to represent phonemic length distinction. However, a previous study reported evidence that logarithmic duration is better than raw duration for classifying Japanese singleton and geminate consonants accurately and predicting their distributions concisely across various speaking rates. To obtain further evidence for the effectiveness of logarithmic duration, this study analyzed the durational characteristics of Japanese short and long vowels. The results indicate that the logarithmic duration is better than the raw duration for the classification and prediction of these vowels. This suggests that logarithmic duration provides a relational invariant acoustic parameter that can cope with durational variations caused by speaking rates.

Keywords: Japanese, short and long vowels, category boundary, logarithmic duration

1. INTRODUCTION

Japanese has a phoneme distinction between short and long vowels [1]. For example, the word [kɯɾo] (meaning 'black') with a short vowel [ɯ] contrasts with the word [kɯːɾo] (meaning 'air route') with a long vowel [ɯː]. The duration of the periodic waveform of a vowel determines whether the vowel is short or long.

The Japanese language also has a phoneme distinction between singleton and geminate consonants [2]. For example, the word [ika] (meaning 'less than') with a singleton consonant [k] contrasts with the word [ikːa] (meaning 'family') with a geminate consonant [kː]. In this example, the closure duration of [k] determines whether the consonant is a singleton or a geminate.

However, vowel and closure durations covary with speaking rate, and they become shorter at faster speaking rates and longer at slower speaking rates [e.g., 3-4]. Therefore, their duration is not an invariant acoustic parameter corresponding to the phonological features of length. This is an instance of "the lack of invariance problem in phonemes."

Relational acoustic invariance theory [5] was proposed to address this problem. It claims that a relational measure, such as the ratio of consonant to vowel duration, provides an invariant property for phonological length distinction, and that this property is used to perceive of this distinction.

Following relational acoustic invariance theory, Amano and Hirata [4] analyzed the closure duration of singleton and geminate consonants in relation to the duration of other speech parts that varied with speaking rate. One of the parts they used was a subword defined as a mora sequence consisting of a preceding mora, closure of a stop, and following mora. Amano and Hirata [4] found that a linear function with closure and subword durations could classify singleton and geminate consonants with a small error. Based on these findings, they concluded that these two durational units provide invariant acoustic properties, which aligns with relational acoustic invariance theory.

Amano and Hirata [4] and other previous studies have conventionally used a linear duration (hereafter "raw duration") to represent the acoustic features of duration-sensitive phonemes. However, as a new approach, Amano et al. [6] introduced logarithmic duration to characterize the acoustic properties of singleton and geminate consonants. They compared the effectiveness of raw versus logarithmic durations for classifying and predicting singleton and geminate consonants across a wide range of speaking rates. Their results revealed that logarithmic duration performed better than raw duration in classifying and predicting the two consonants, which suggests that logarithmic duration provides relational invariant acoustic properties [e.g., 5] of singleton and geminate consonants.

Although Amano et al. [6] provided evidence of the effectiveness of logarithmic duration, it has only been used for singleton and geminate consonants, and their findings on logarithmic duration might be an exceptional case for singleton and geminate consonants. Subsequently, it is necessary to examine the general applicability of logarithmic duration to other duration-sensitive phonemes. Therefore, this study examines the effectiveness of logarithmic duration using Japanese short and long vowels pronounced at various speaking rates.

2. SPEECH DATA

The original speech data obtained by Hirata [3] were used for the analysis in this study. In Hirata [3], four native Japanese speakers (two males and two females) pronounced Japanese word and nonword sets three times at fast, normal, and slow speaking rates, each including contrasting short and long vowels.

The nonword set consisted of three types of /m/ containing sequence: /CV.CV/, /CVV.CV/, and /CV.CVV/, where the period "." represents a syllable boundary; C is /m/; V is one of /i/, /e/, /a/, /o/, and /u/; and the underline indicates a long vowel. Each sequence type contained 180 tokens. Pairing /CV.CV/ and /CVV.CV/ produced 180 short-long vowel contrasts in the first syllable. Pairing /CV.CV/ and /CV.CVV/ also produced 180 contrasts in the second syllable. These 360 pairs of contrasting short and long vowels (a total of 720 items) were used to analyze nonwords.

The word set consisted of 10 minimal pairs of Japanese disyllabic words containing short and long vowels in the first or second syllable, that is, the /CV.CV/-/CVV.CV/ pair or the /CV.CV/-/CV.CVV/ pair. C included various consonants; V included /i/, /e/, /a/, /o/ and /u/; and the underlines indicate contrasting short and long vowels. In total, there were 720-word tokens (20 words × three repetitions × three speaking rates × four speakers).

3. ANALYSES

The duration of each phoneme in the speech data was measured in milliseconds by inspecting the waveform and spectrogram.

3.1. Classification of vowels

Discriminant analyses were conducted to obtain a category boundary of short and long vowels and classification errors for words and nonwords. These analyses used the following discriminant model:

$$f = a_0 + a_1 v_1 + a_2 v_2 \qquad (1)$$

$$f = a_0 + a_1 \log v_1 + a_2 \log v_2 \qquad (2)$$

where the dependent variable f is the label for short and long vowels, a_0 to a_2 are discriminant coefficients, v_1 and v_2 are independent variables: v_2 is the vowel duration (ms), and v_1 is the average mora duration (ms) calculated by dividing the sentence duration by the number of morae in the sentence. In this calculation of average mora duration, the words with target short and long vowels were excluded from the sentence. Eq. 1 was based on the formulation of Amano and Hirata [4], in which an intercept and a variable related to the speaking rate were introduced to the discriminant model. To examine the effects of the logarithmic duration, Eq. 2 was introduced by replacing the raw durations in Eq. 1 with logarithmic durations.

3.2. Prediction of vowels

Regression analyses were conducted to obtain regression lines for short and long vowels using raw durations (dependent variable v_2 and independent variable v_1) and logarithmic durations (dependent variable $\log v_2$ and independent variable $\log v_1$), respectively. The coefficient of determination (R^2) was calculated to examine the goodness-of-fit of each regression. To investigate the extent of short and long vowel distributions as a function of speaking rate, root mean square errors (RMSE) from their regression lines were calculated as a function of the average mora duration. For further distribution examination, the distance from the category boundary to each data point was calculated, and the coefficient of variation (CV) was obtained for each distribution.

4. RESULTS

4.1. Classification of vowels

Fig. 1 illustrates scattergrams of short and long vowels on a coordinate plane of vowel duration and average mora duration, with category boundary and regression lines shown as solid and broken lines, respectively. Short and long vowels had a V-shaped distribution for the raw duration (Figs 1a and 1c), whereas they had an almost parallel distribution for the logarithmic duration (Figs 1b and 1d).

For nonwords, the error in short and long vowel classification by category boundary was significantly smaller ($z = 2.04$, $p < .05$) for the logarithmic duration (1.2% in Fig. 1d) than for the raw duration (3.5% in Fig. 1c). However, for words, no significant difference was observed in the classification error between logarithmic (5.1% in Fig. 1b) and raw (5.6% in Fig. 1a) durations. These results indicate that the logarithmic duration is equal to or better than the raw duration for short and long vowel classifications.

4.2. Prediction of vowels

As for the coefficient of determination (R^2), there was no significant difference between the raw and logarithmic durations for both nonwords (R^2 = .71–.77) and words (R^2 = .62–.77). The results indicate that logarithmic duration has equal performance to raw duration in the prediction of short and long vowels.

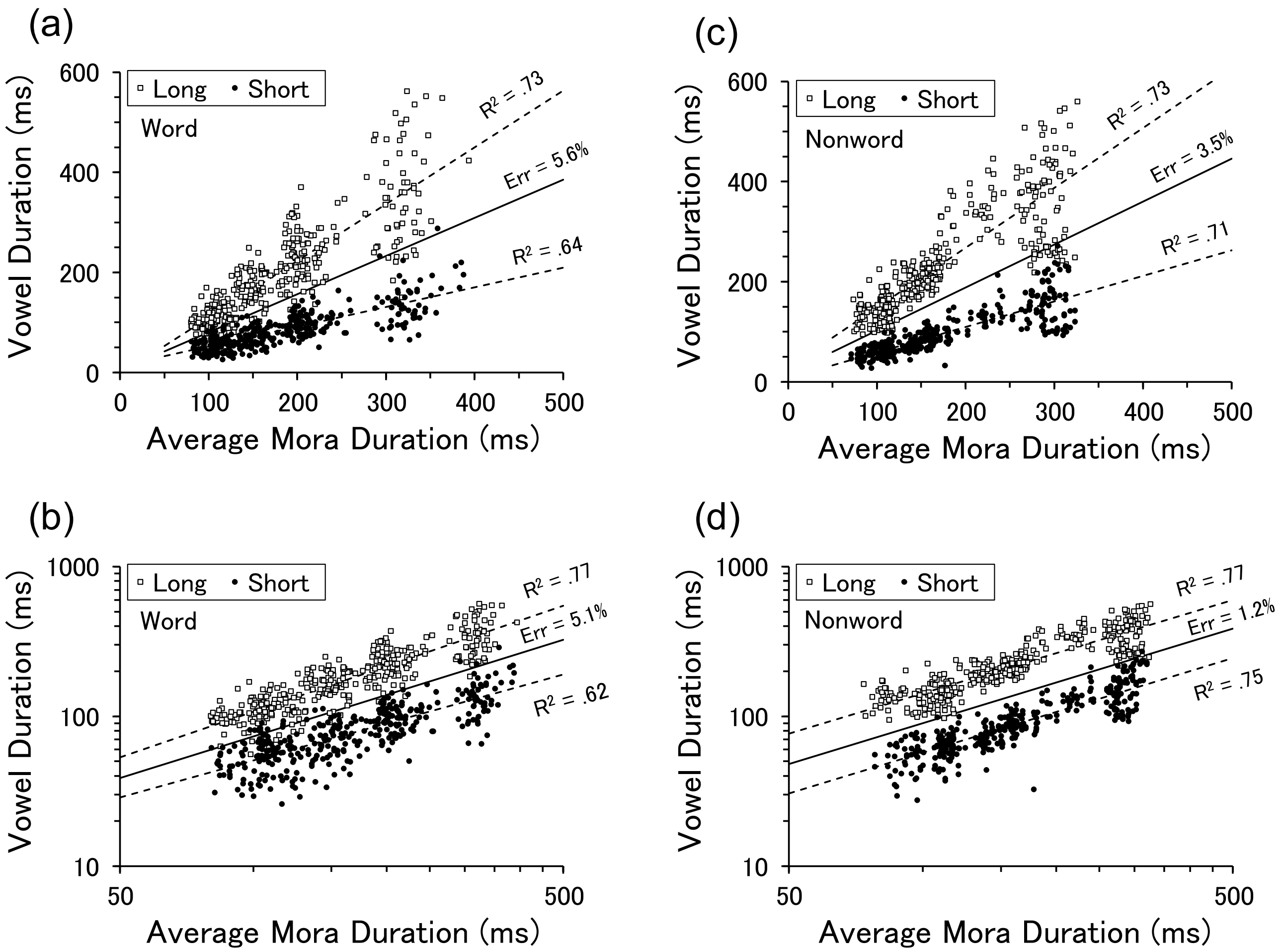

Figure 1: Scattergrams of short and long vowels in (a) words using raw duration, (b) words using logarithmic duration, (c) nonwords using raw duration, and (d) nonwords using logarithmic duration. Solid and broken lines represent category boundaries and regression lines, respectively.

Fig. 2 illustrates the RMSE from the regression line. The RMSE of short and long vowels for the logarithmic duration were almost constant, whereas the RMSE for the raw duration increased as the average mora duration increased. The results indicate that the logarithmic duration provides a less spreading and more compact distribution than the raw duration.

This tendency was also confirmed by the coefficient of variation in Table 1. The coefficient was significantly smaller for logarithmic duration than raw duration of short vowels in words [$F(359, 359) = 30.45$, $p = 6.12\times10^{-162}$], long vowels in words [$F(359, 359) = 34.45$, $p = 5.10\times10^{-174}$], short vowels in nonwords [$F(359, 359) = 46.91$, $p = 5.00\times10^{-197}$], and long vowels in nonwords [$F(359, 359) = 49.06$, $p = 2.09\times10^{-200}$]. These results indicate that logarithmic duration provides less variation than raw duration in terms of distance from the category boundary for both words and nonwords.

Item	Vowel	Duration	
		Raw	Logarithm
Word	Short	0.394	0.071
	Long	0.394	0.067
Nonword	Short	0.310	0.045
	Long	0.310	0.044

Table 1: Coefficient of variation of vowel's distance from the category boundary.

5. DISCUSSION

The results of the discriminant and regression analyses indicated that, compared to raw duration, logarithmic duration provided equal or better classification and prediction of short and long vowels (Fig. 1), and provided a simpler and more compact representation of vowels in terms of the RMSE from the regression line (Fig. 2) and distance from the category boundary (Table 1).

In some cases, the logarithmic and raw durations did not significantly differ in classification error and R^2. One possible reason for the non-significant difference is that this study reanalyzed Hirata [3]'s speech data pronounced by only four speakers at a not-so-wide range of speaking rates (2.5–14.1 mora/s), which might have unexpectedly resulted in a

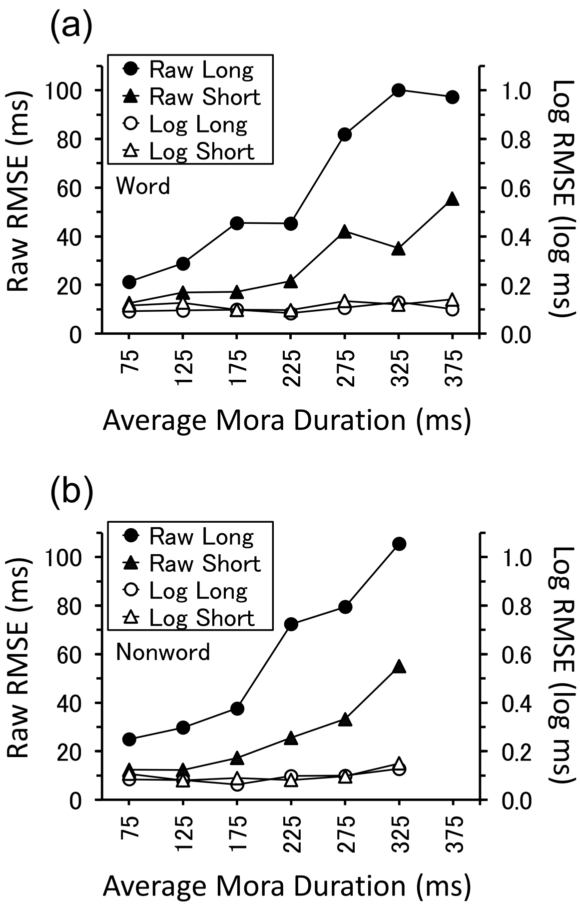

Figure 2: Root mean square errors (RMSE) from the regression lines for short and long vowels in (a) words and (b) nonwords. The left vertical axis is for RMSE of long vowels (●) and short vowels (▲) for raw duration. The right vertical axis is for RMSE of long vowels (○) and short vowels (△) for logarithmic duration. The values of the horizontal axis represent the mid-point of the bin for an interval of the mid-point ± 50ms. The intervals do not include the lower limit.

small variance and a clear separation of short and long vowel distributions. However, the insignificant difference does not indicate negative evidence for the logarithmic duration. Therefore, it was concluded that logarithmic duration has an advantage over raw duration in classifying and predicting short and long vowels.

The findings of this study are consistent with previous studies demonstrating that logarithmic duration is better for the classification and prediction of singleton and geminate consonants [6] and voiced and voiceless stop consonants [7]. Accumulating evidence suggests that logarithmic duration provides better acoustic variables for representing phoneme categories than raw duration which has been conventionally used in speech studies.

However, the evidence was obtained only for Japanese phonemes. Future research is needed in other languages to confirm the generality and universality of logarithmic duration. For example, research should examine languages with short and long vowel contrasts, such as Arabic [8], Estonian [9], and Thai [10], or languages with singleton-geminated stop contrasts, such as Bengali [11], Hungarian [12], and Italian [5].

Amano and Hirata's [4] findings supported relational acoustic invariance theory [5] based on their singleton-geminate study using raw duration. Amano et al. [6] also supported this theory, however, they claimed that logarithmic duration, not raw duration, provides an invariant acoustic parameter. The present study is in line with Amano et al. [6] because the combination of logarithmic durations accurately classified short and long vowels and concisely predicted their durational distributions across various speaking rates. This suggests that the phonological feature of length is well represented by a logarithmic duration relationship.

6. REFERENCES

[1] Vance, T. J. 2008. *The Sounds of Japanese*. Cambridge University Press.
[2] Kawahara, S. 2015. The phonetics of sokuon, or geminate obstruents. In: Kubozono, H. (ed), *Handbook of Japanese Phonetics and Phonology*. Mouton, 43–77.
[3] Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *J. Phon.* 32(4), 565–589.
[4] Amano, S., Hirata, Y. 2015. Perception and production of singleton and geminate stops in Japanese: Implications for the theory of acoustic invariance. *Phonetica* 72, 43–60.
[5] Pickett, E. R., Blumstein, S. E., Burton, M. W. 1999. Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica* 56, 135–157.
[6] Amano, S., Kondo, M., Yamakawa, K. 2021. Predicting and classifying Japanese singleton and geminate consonants using logarithmic duration. *J. Acoust. Soc. Am*. 150(3), 1830–1843.
[7] Yamakawa, K., Amano, S., Kondo, M. (Accepted). Logarithmic durations for classifying and predicting Japanese voiced and voiceless plosives. *Proc. ICPhS 2023*.
[8] Tsukada, K. 2009. An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data. *Int. J. Asian Lang. Process.* 19(4), 127–138.
[9] Harms, R. T. 2006. Estonian. In: Brown, K. (ed), *Encyclopedia of Language & Linguistics* (2nd ed), Elsevier, 228.
[10] Abramson, A. S., Reo, N. 1990. Distinctive vowel length: duration vs. spectrum in Thai. *J. Phon.* 18(2), 79–92.
[11] Lahiri, A., Hankamer, J. 1988. The timing of geminate consonants. *J. Phon.* 16(3), 327–338.
[12] Vago, R. M. 2006. Hungarian: Phonology. In: Brown, K. (ed), *Encyclopedia of Language & Linguistics* (2nd ed), Elsevier, 433–440.