LOGARITHMIC DURATIONS FOR CLASSIFYING AND PREDICTING JAPANESE VOICED AND VOICELESS PLOSIVES

Kimiko Yamakawa[1], Shigeaki Amano[2], Mariko Kondo [3]

[1]Shokei University, [2]Aichi Shukutoku University, [3]Waseda University
[1]jin@shokei-gakuen.ac.jp, [2]psy@asu.aasa.ac.jp, [3]mkondo@waseda.jp

ABSTRACT

Conventionally, the duration of a phoneme is represented to be linear (raw duration). However, studies have reported that as an acoustic parameter of Japanese singleton/geminate consonants or short/long vowels, the logarithmic duration is better than the raw duration. The durational characteristics of Japanese voiced and voiceless plosives were analyzed to investigate the effectiveness of the logarithmic duration. The results indicated that the logarithmic duration is better than or comparable to the raw duration as an acoustic parameter for the classification and prediction of plosives at various speaking rates. This phenomenon suggests that the logarithmic duration provides a relational invariant acoustic parameter that can cope with durational variations caused by speaking rates.

Keywords: Japanese, voiced and voiceless plosives, category boundary, logarithmic duration

1. INTRODUCTION

Phoneme sounds exhibit three fundamental acoustic properties, namely intensity, frequency, and duration. Intensity is typically represented in logarithms (i.e., decibels, ISO 226:2003). Logarithm (e.g., the Mel scale [1]) or linear representations are used for frequency. Duration is conventionally represented as linear (raw duration). For example, Amano and Hirata [2–3] used the raw durations of speech segments to classify Japanese singleton and geminate stops. Pickett et al. [4] used a raw duration to contrast Italian singleton and geminate consonants. Miller et al. [5] used a raw duration for voice onset time (VOT; i.e., time from burst to voicing segment) as an acoustic parameter of English /p/ and /b/.

In contrast to the previously mentioned studies, Amano et al. [6] used the logarithmic duration as an acoustic parameter for Japanese singleton and geminate consonants. Specifically, they used the logarithmic duration of /k/'s closure and /s/'s frication as primary variables. Furthermore, they used the logarithmic average mora duration as a secondary variable related to the speaking rate. They found that compared with the raw duration, the combination of logarithmic durations provided a higher coefficient of determination (R^2) in the prediction of singleton and geminate consonants and almost the same discriminant error in the classification of these consonants. They concluded that the logarithmic duration is a better acoustic parameter for consonants than the raw duration.

Amano et al. [7] analyzed the effectiveness of the logarithmic duration in the classification and prediction of Japanese short and long vowels in words and non-words. The results of the study revealed that compared with the raw duration, the logarithmic duration provided a smaller discriminant error in the classification of vowels in non-words. Moreover, logarithmic durations provided a smaller coefficient of variation for the distance of the vowel from the category boundary. Furthermore, they revealed that logarithmic durations provided a simple and compact representation of the vowel distribution in terms of the root mean square error (RMSE) from the regression line. The results indicated that the logarithmic duration exhibits considerable advantage over the raw duration as an acoustic parameter for short and long vowels.

Although the effectiveness of the logarithmic duration has been studied previously [6–7], their effectiveness is to be confirmed for other phonemes. The positive findings for the logarithmic duration could be an exceptional case only for singleton and geminate consonants or short and long vowels. Therefore, confirming the general applicability of the logarithmic duration to other phonemes is necessary. Thus, this study examined the effectiveness of the logarithmic duration in Japanese voiceless and voiced plosives.

Voiceless and voiced plosives can be contrasted with VOT [e.g., 5]. However, VOT is not always measurable because a burst does not exist in some cases, such as a plosive at a very high speaking rate. Instead of uncertain VOT, this study used the entire duration of the plosives, namely the duration from the end of the preceding vowel to the beginning of the following vowel because the duration is always measurable. Consistent with previous studies [6–7], this study examined target phonemes across various speaking rates. The logarithmic durations can cope with the speaking rate and provide better classification and prediction of voiceless and voiced plosives compared to raw durations.

2. SPEECH RECORDINGS

2.1 Speakers

Nine native Japanese speakers (three men and six women) participated in the recording. The average age was 21.0 years (Min. = 19, Max. = 25, SD = 1.8 years). They spoke standard Japanese and were from the Tokyo metropolitan area and its suburbs. The speakers had no symptoms of speech disorders and were paid to participate in the recording.

2.2 Equipment

Recordings were conducted in a soundproof room at Waseda University in Tokyo, Japan, using a microphone (Rode, NT-2A) with a pop screen (Stedman, PROSCREEN 101) connected to a computer (Panasonic, CR-RZ4, or CF-FV1SRCQP) through an audio interface (Roland, Rubix 24).

2.3 Target consonants

Target consonants were voiceless (/p/, /t/, /k/) and voiced (/b/, /d/, /g/) plosives in three-mora-long non-words (Table 1). The plosives were located at the beginning of the second mora in the non-words. The vowel of the second mora was /a/. /e/, or /o/. The first and third moras were /na/. With these phoneme sequences, vowels could not be devoiced.

2.3 Procedure

The item was embedded in a carrier sentence /koreka __ tosuru/ (Let it this or __) and presented at the top of the computer screen in hiragana (Japanese phonological script). A silent video of flashing lights on a digital metronome (Seiko, SQ200) was displayed at the center of the computer screen to indicate the speaking rate.

Voiceless	Voiced
/napana/	/nabana/
/napena/	/nabena/
/napona/	/nabona/
/natana/	/nadana/
/natena/	/nadena/
/natana/	/nadana/
/nakana/	/nagana/
/nakena/	/nagena/
/nakana/	/nagana/

Table 1: Non-word items having a target plosive at the second mora.

Beats per minute (BPM)	Speaking rate (mora/s)
81	12.2
51	7.7
32	4.8
25	3.8

Table 2: Beats per minute (BPM) set on a digital metronome and the corresponding speaking rate.

Speakers were asked to pronounce the sentence with a non-word with a flat accent pattern (i.e., low-high-high pitch pattern) three times at the speaking rate indicated by flashing lights. Speaking rates are presented in Table 2. Half of the speakers recorded voiceless items first, and subsequently voiced items. The other half of the speakers were recorded in the reverse order.

3. ANALYSES

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

3.1. Classification of plosives

Discriminant analyses were conducted to obtain the category boundary of voiced and voiceless plosives and their classification errors. The following discriminant model was used in the analyses:

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

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

where the dependent variable f is the label for voiced and voiceless plosives, a_0 to a_2 are discriminant coefficients, v_1 and v_2 are independent variables, v_2 is the plosive 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 the average mora duration, non-words with target voiced and voiceless plosives were excluded from the sentence. Equation (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, (2) was introduced by replacing the raw durations in (1) with logarithmic durations. The category boundaries of the voiced and voiceless plosives were obtained by substituting 0 for f in (1) and (2).

3.2. Prediction of plosives

Regression analyses were conducted for voiced and voiceless plosives using the following models:

(3) $$v_2 = b_0 + b_1 v_1$$

(4) $$\log v_2 = b_0 + b_1 \log v_1$$

where v_2 is the plosive duration (ms), v_1 is the average mora duration (ms), and b_0 and b_1 are the regression coefficients. The coefficient of determination (R^2) was calculated to examine the goodness-of-fit of each regression. To investigate the extent of the plosive distribution, the RMSE from the regression line was calculated as a function of the average mora duration. For further distribution examination, the distance from the category boundary to the plosive was calculated and its coefficient of variation (CV) was obtained as an index of the distance variance.

4. RESULTS

Fig. 1 displays the scattergrams of voiced and voiceless plosives on a coordinate plane of plosive duration and average mora duration, with the category boundary and regression lines displayed as solid and broken lines, respectively.

4.1. Classification of plosives

The raw and logarithmic durations did not exhibit any significant difference in the classification error of the voiced and voiceless plosives (raw 12.4% vs. logarithm 11.2%, $z = 0.82$, *ns*), /p/ and /b/ (raw 11.4% vs. logarithm 11.1%, $z = 0.12$, *ns*), /t/ and /d/ (raw 9.9% vs. logarithm 10.3%, $z = 0.16$, *ns*), and /k/ and /g/ (raw 14.0% vs. logarithm 10.8%, $z = 1.24$, *ns*). These results indicate that the logarithmic duration performs as well as raw duration in voiced and voiceless plosive classifications. The gradient of the category boundary ($-a_1/a_2$) was 0.19 and 0.57 for the raw and logarithmic duration, respectively.

Type	Plosive	Duration		Difference
		Raw	Log	
Voiceless	All	0.45	0.68	$p < .001$
	/p/	0.66	0.72	*ns*
	/t/	0.64	0.72	$p < .05$
	/k/	0.27	0.62	$p < .001$
Voiced	All	0.46	0.47	*ns*
	/b/	0.52	0.53	*ns*
	/d/	0.45	0.45	*ns*
	/g/	0.47	0.51	*ns*

Table 3: Coefficient of determination (R^2) for voiced and voiceless plosives in raw and logarithmic durations.

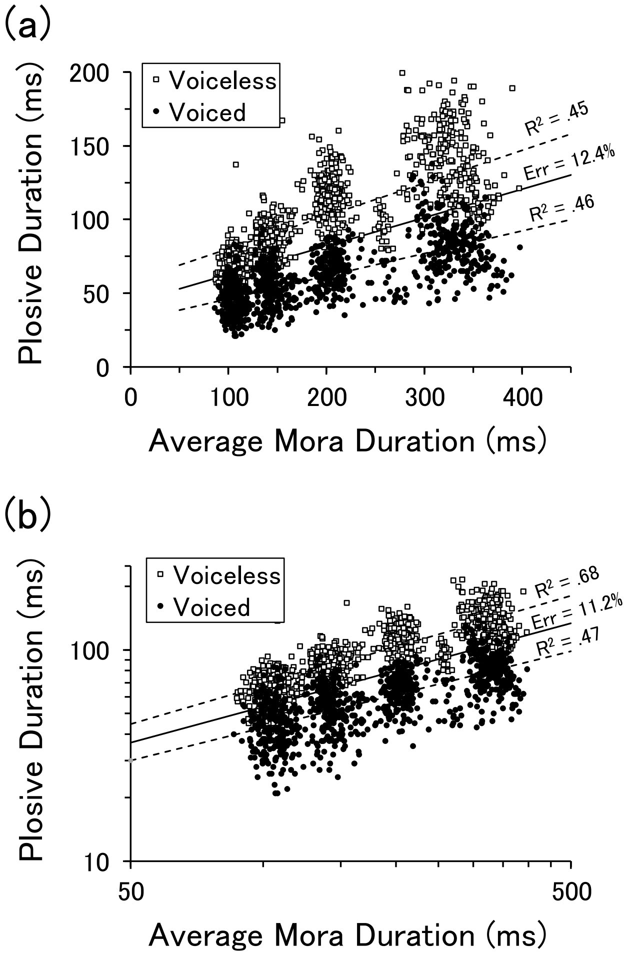

Figure 1: Scattergrams of voiced and voiceless plosives in (a) raw duration and (b) logarithmic duration. Solid and broken lines represent category boundaries and regression lines, respectively.

4.2. Prediction of plosives

The gradients of the regression line (b_1) in raw duration were 0.15 and 0.22 for voiced and voiceless plosives, respectively. The b_1 in the logarithmic duration was 0.52 and 0.61 for voiced and voiceless plosives, respectively. These results indicated that the regression lines for both the raw and logarithmic durations may not be parallel to the category boundary.

However, the distribution in the logarithmic duration (Fig. 1b) appears to be more compact than the distribution in the raw duration (Fig. 1a). This tendency is confirmed by R^2 in Table 3. Logarithmic duration provided a significantly higher R^2 than raw duration for all voiceless plosives, /t/, and /k/. No

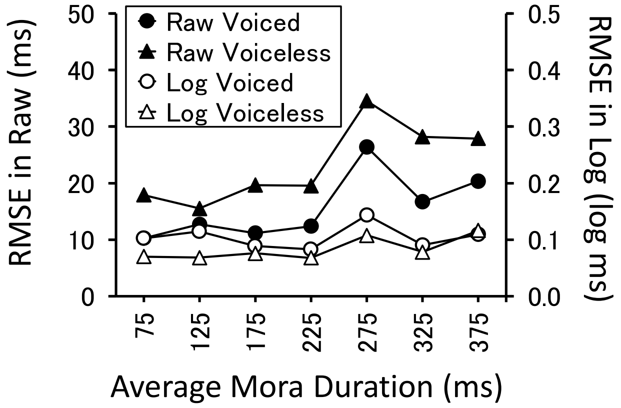

Figure 2: Root mean square errors (RMSE) from the regression lines for voiced and voiceless plosives. The left vertical axis is for RMSE of voiced plosives (●) and voiceless plosives (▲) for raw duration. The right vertical axis is for RMSE of voiced plosives (○) and voiceless plosives (△) for logarithmic 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.

significant R^2 difference was observed between the logarithmic and raw durations for the /p/ and voiced plosives. However, R^2 tended to be slightly higher in the logarithmic duration than in the raw duration in these cases. These results indicate that the logarithmic duration is better than or equal to the raw duration in predicting plosives.

Fig. 2 illustrates the RMSE of the voiced and voiceless plosives from the regression line. The RMSE for the logarithmic duration was almost constant, whereas the RMSE for the raw duration tended to be higher as the average mora duration increased. The results indicated that the logarithmic duration provides a less spreading and more constant-width distribution than the raw duration against the speaking rate.

These characteristics were confirmed by the CV of the plosive distance from the category boundary (Table 4). All CVs were significantly smaller for the logarithmic duration than for the raw duration ($p < .001$). These results indicated that the logarithmic duration exhibited lower variation than the raw duration in terms of distance from the category boundary.

5. DISCUSSION

The results of the discriminant analysis of voiced and voiceless plosives indicate that the performance of the logarithmic duration is superior to that of the raw duration. The results of the regression analysis indicate that the logarithmic duration outperforms or is equal to the raw duration; moreover, the results of the RMSE and CV indicate that compared with the raw duration, the logarithmic duration is less affected by the speaking rate and provides a stable and compact distribution of plosives. The logarithmic duration is superior to the raw duration as an acoustic parameter of voiced and voiceless plosives in classification and prediction across various speaking rates.

However, the results revealed that the effectiveness of the logarithmic duration, classification, and regression were not perfect even if the logarithmic duration was used. The discriminant errors were 10.8%–11.2%, and the R^2s were 0.45–0.72. A possible cause of these not-low errors and not-high R^2s is that duration is not the only acoustic parameter of voiced and voiceless plosives. The plosives can be distinguished and predicted by other parameters, such as VOT and formant transition. Thus, plosives have multiple parameters, and their duration may not be their primary parameter, which may result in imperfect classification and regression. This differs from the case of singleton/geminate consonants [6] and short/long vowels [7]. In these consonants and vowels, duration is the primary parameter, and logarithmic duration outperforms raw duration.

Although the duration is not the primary parameter of plosives, the logarithmic duration is still effective in classifying and predicting them. This result indicates that the logarithmic duration is effective not only for singleton/geminate consonants [6] and short/long vowels [7] but also for plosives. This finding is critical because it is consistent with the claims of previous studies [6-7] regarding the effectiveness of the logarithmic duration.

Another significant aspect is that the findings of this study support relational acoustic invariance theory [4]. Consistent with previous studies [6-7], in this study, two durations, namely plosive duration and average mora duration, were used. The combination of these durations provided excellent classification and prediction. Thus, the acoustic invariance of phonemes against speaking rates was achieved in the relationship of the durations, as the theory claims. Furthermore, the results of this study suggest that relational acoustic invariance is provided by logarithmic rather than raw duration. To obtain further evidence of relational invariance theory and the effectiveness of the logarithmic duration, other phonemes in Japanese and other languages should be considered.

A limitation of this study was the use of read speech. Because read speech has a more regular rhythm than spontaneous speech, such regularity could affect the results of this study and provide an unexpected advantage to the logarithmic duration. Further verification using spontaneous speech is critical.

6. REFERENCES

[1] Stevens, S. S., Volkmann, J. 1940. The relation of pitch to frequency: A revised scale. *Am. J. Psychol.* 53(3), 329–353.

[2] Amano, S., Hirata, Y. 2010. Perception and production boundaries between single and geminate stops in Japanese. *J. Acoust. Soc. Am.* 128(4), 2049-2058.

[3] 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.

[4] 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.

[5] Miller, J. L., Green, K. P., Reeves, A. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast, *Phonetica* 43(1-3), 106–115.

[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] Amano, S., Hirata, Y., Yamakawa, K. (Accepted). Logarithmic durations for classifying and predicting Japanese short and long vowels. *Proc. ICPhS 2023.*