LOGARITHMIC DURATION TO PREDICT AND DISCRIMINATE SINGLETON AND GEMINATE CONSONANTS IN JAPANESE

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

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

ABSTRACT

Appropriate acoustic cues to predict and discriminate singleton and geminate consonants in Japanese were identified by analyzing the durations of the consonants and their anteroposterior segments. Twelve minimal pairs of words with singleton and geminate of /k/ and /s/ in a carrier sentence were pronounced by 20 native Japanese speakers at various speaking rates. Regression and discriminant analyses revealed that logarithmic duration was better than raw duration to predict and discriminate the distributions of singleton and geminate consonants. The best acoustic cues were a combination of the logarithmic durations of closure and frication of singleton and geminate consonants, and the logarithmic average duration of morae in the preceding phrase. It is suggested that speech planning of singleton and geminate consonants is based on their logarithmic durations and the speaking rate associated with the mora unit.

Keywords: geminate, speaking rate, logarithmic duration, regression analysis, discriminant analysis.

1. INTRODUCTION

Japanese geminate consonants consist of voiceless obstruents, and they have long segment duration. Geminate consonants are phonologically represented as /Q/: e.g. *sakka* “novelist” /saQka/ [sakːa], *setchi* “installation” /seQti/ [setːɕi] and *kassai* “applause” /kaQsai/ [kasːai]. The main acoustic cue to distinguish a geminate consonant from a singleton consonant is either the closure duration for a plosive and affricate or the frication duration for a fricative. These durations covariate with speaking rate. That is, the duration is long for a low speaking rate and short for a high speaking rate [5, 6].

To convert the rate-covariated duration into an invariant cue for singleton and geminate consonants, previous studies have used various segment durations as a secondary cue. Examples of these secondary cues are the preceding vowel duration [5, 7], preceding mora duration [6], following vowel duration [3, 7], duration of a subword ranging from the start of the preceding consonant to the end of the following vowel [1], and duration of successive vowels ranging from the start of the preceding vowel to the end of the following vowel [2]. These secondary cues were used as a denominator of ratio [4] or as a variable of a linear function [1, 2].

However, it is not clear which of these is the most appropriate secondary cue. In addition, the previous studies used only raw duration for the main and secondary cues and did not examine the suitability of these cues of logarithmic duration. Hence, this study examined the raw and logarithmic durations to identify the most appropriate secondary cue for the prediction and discrimination of singleton and geminate consonants at various speaking rates.

2. RECORDINGS

2.1. Speakers

Twenty native Japanese speakers (9 males and 11 females) participated in the recordings. Their average age was 22.8 years (Min. = 18, Max. = 34, SD = 3.6). They spoke standard Japanese and were from the Tokyo metropolitan area and its suburbs.

2.2. Test words

Twelve minimal pairs of Japanese onomatopoeias (Table 1) were used as test words. The minimal pairs

Table 1: Test words: pairs of Japanese onomatopoeias contrasting underlined singleton and geminate consonants.

Singleton	Geminate
kusukusu	kussukusu
pukupuku	pukkupuku
sukusuku	sukkusuku
kusakusa	kussakusa
pukapuka	pukkapuka
sukasuka	sukkasuka
kasukasu	kassukasu
pakupaku	pakkupaku
sakusaku	sakkusaku
kasakasa	kassakasa
pakapaka	pakkapaka
sekaseka	sekkaseka

contrast singleton and geminate consonants (/k/ or /s/) following the first mora, however, other phonemes in a word are identical. The test word was embedded in a carrier sentence /korewa _ desu/ ("This is _") in the recordings.

2.3. Procedure

Recordings were conducted in a soundproof room at Waseda University. Using a personal computer (Panasonic, Let's Note CR-RZ4; or Sony, VAIO, PRO13), a carrier sentence with a test word in Hiragana characters was presented at the top of a liquid crystal display (Sony, LMD-1530W), and a silent video of a digital quartz metronome (Seiko, SQ200) was presented at the center of the display.

Each speaker was asked to pronounce the presented sentences at different speaking rates (Table 2). The speaking rates were indicated by the silent video of the metronome equipped with lined LED lights that flashed from side to side. The speaker synchronized the start of each sentence with the flash of the endpoint light on one side of the LED line (it can be the left or the right side to start) and the end of the sentence with the flash of the endpoint light on the other side of the line. Each sentence was pronounced in one breath without a pause. The test word embedded in the sentence was pronounced with no accent (i.e., low-high-high-high for singleton and low-high-high-high-high for geminate). The speakers pronounced each sentence at each speaking rate at least three times. Sometimes speakers did not pronounce a sentence at the correct speaking rate, so they repeated those sentences until they produced three attempts at the correct speaking rate. The order of the speaking rates and test words were counterbalanced between speakers. The pronunciation was digitally recorded with a microphone (SONY, F-780) and a digital recorder (ZOOM, R8) in a 16-bit quantization and 48-kHz sampling frequency.

Table 2: Beats per minute (BPM) set by an electric metronome. Speaking rate was calculated with BPM and the number of morae in a carrier sentence.

Singleton		Geminate	
Beats per minute (BPM)	Speaking rate (mora/s)	Beats per minute (BPM)	Speaking rate (mora/s)
101	15.2	91	15.2
81	12.2	73	12.2
64	9.6	57	9.5
51	7.7	46	7.7
40	6.0	36	6.0
32	4.8	29	4.8

3. ANALYSIS

3.1. Data selection and preparation

The middle item of the three recordings with a correct speaking rate was selected. Some of the speakers devoiced some /u/ in some test words. These test words were not used for the analysis. In the test words without devoiced vowels, there were 881 items of geminate /k/, 448 of geminate /s/, 661 of singleton /k/, and 326 of singleton /s/. Professional labelers used a waveform and spectrogram to mark the start and end times of each phoneme segment in each sentence item. Durations of the closure and frication of singleton and geminate consonants were obtained using these timings. The duration was also obtained for the preceding vowel, following vowel, and preceding mora of the respective single and geminate consonants. The duration of successive vowels was obtained as the time span from the start of the preceding vowel to the end of the following vowel. The duration of a subword was obtained as the time span from the start of the preceding consonant to the end of the following vowel. The closure part of the preceding consonant was not included in the subword. The average mora duration in the phrase (/korewa/) preceding the test word was also obtained. These raw durations and their logarithmic-converted durations were used in the following analyses.

3.2. Regression analysis

Regression analyses for singleton and geminate consonants were separately conducted for /k/ and /s/. The dependent variable was the raw closure or frication duration. The independent variables were the raw average mora duration in the preceding phrase, the raw successive vowel duration, the raw subword duration, the raw preceding mora duration, the raw preceding vowel duration, and the raw following vowel duration. Regression analyses were also conducted with logarithmic-converted value of these raw durations. Figures 1 and 2 show coefficient of determination (R^2) for /k/ and /s/, respectively. In almost all cases, R^2 was larger with logarithmic durations than raw durations. This result indicates that logarithmic durations predict singleton and geminate distributions better than raw durations. Also, durations of subword and successive vowels could predict the distributions at least as well as average mora duration.

Figure 1: Coefficient of determination (R^2) for singleton and geminate /k/.

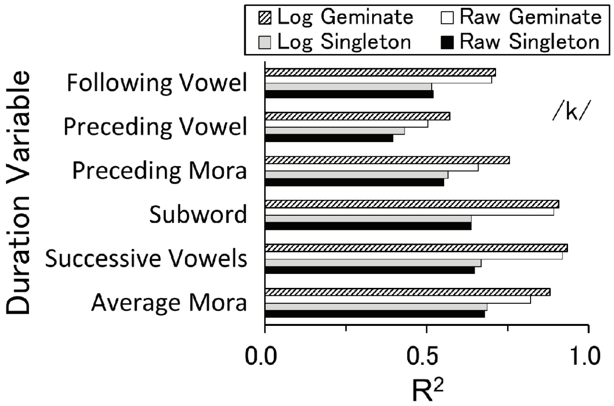


Figure 3: Discriminant error (%) of singleton and geminate /k/.

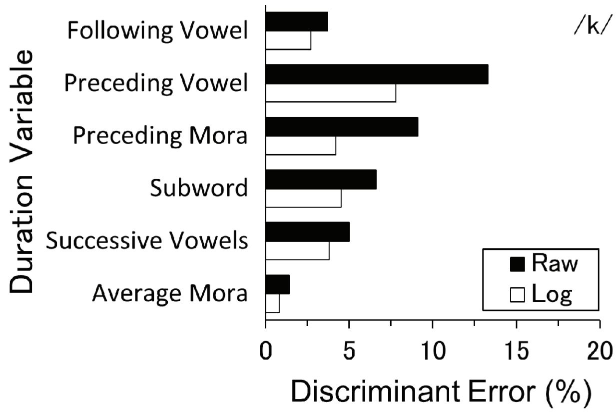


Figure 2: Coefficient of determination (R^2) for singleton and geminate /s/

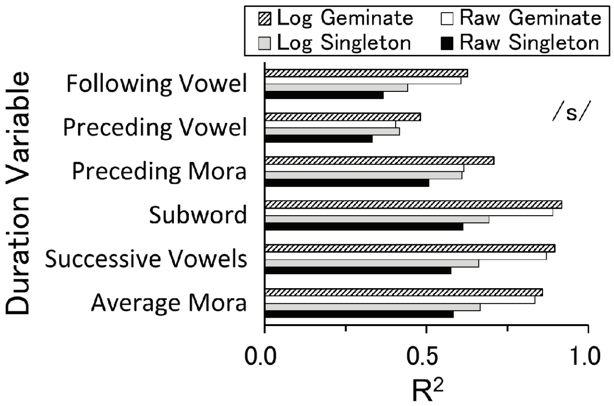
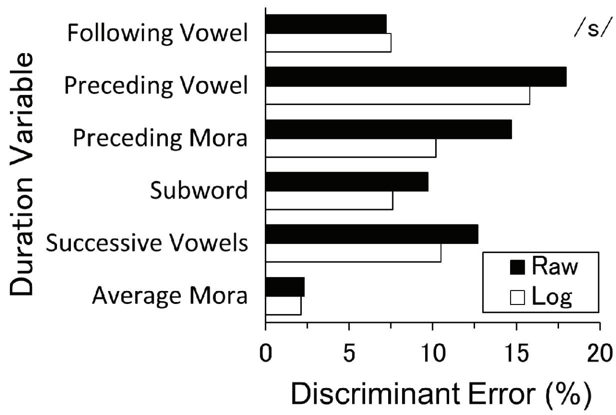


Figure 4: Discriminant error (%) of singleton and geminate /s/.



3.3. Discriminant analysis

Discriminant analyses of singleton and geminate consonants were separately conducted for /k/ and /s/. The dependent variable was a label of singleton and geminate consonants. The independent variables were a pair of the raw closure or frication duration and one of the following durations: the raw average mora duration in the preceding phrase, the raw successive vowel duration, the raw subword duration, the raw preceding mora duration, the raw preceding vowel duration, and the raw following vowel duration. Discriminant analyses were also conducted with logarithmic-converted value of these raw durations.

Figures 3 and 4 show discriminant errors for /k/ and /s/, respectively. For /k/, significant difference occurred between the logarithmic and raw preceding vowel durations ($z = 4.971$, $p < .001$, two-tailed), the logarithmic and raw preceding mora durations ($z = 5.461$, $p < .001$, two-tailed), and the logarithmic and raw subword durations ($z = 2.547$, $p < .05$, two-tailed). For /s/, significant difference occurred between the logarithmic and raw preceding mora durations ($z = 2.681$, $p < .01$, two-tailed). Although other logarithmic and raw durations did not differ significantly, the discriminant error tended to be smaller for the logarithmic duration than the raw duration for almost all cases for both /k/ and /s/. The smallest discriminant error occurred with logarithmic average mora durations for /k/ and /s/. They differed significantly from every other logarithmic or raw duration ($p < .01$, two-tailed). Average mora duration produced the least discriminant error, followed by the durations of following vowel, subword and successive vowels.

Regression and discriminant analyses were conducted using combined data of /k/ and /s/. Figure 5 shows the distribution of singleton and geminate consonants on a coordinate plane of logarithmic

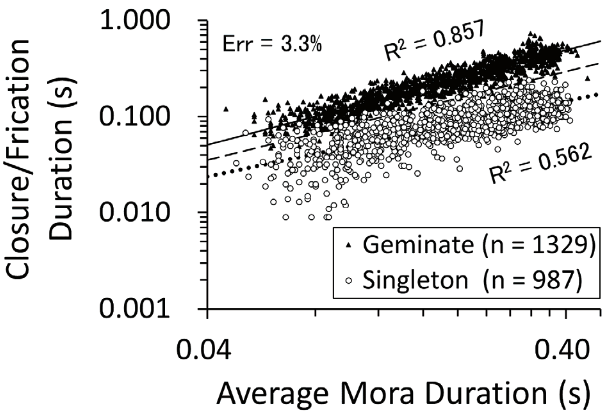


Figure 5: Scattergram of singleton and geminate consonants (/k/ and /s/). The break line represents a discriminant equation [8.246 log(y) - 7.592 log(x) + 1.352 = 0], the solid line represents a regression equation for geminate consonants [log(y) = 0.982 log(x) + 0.787], and the dotted line represents a regression equation for singleton consonants [log(y) = 0.783 log(x) - 0.527], where y is closure/frication duration, and x is average mora duration.

average mora duration and logarithmic closure/frication duration. With the two logarithmic durations, the singleton and geminate consonants were well separated with a very small discriminant error (3.3%). Regression analyses well predicted the distribution of geminate consonants (R^2 = .857) and singleton consonants (R^2 = .562), but some deviance was observed for singleton consonants in the high speaking rates (i.e., short average mora durations). These results indicate that logarithmic average mora duration combined with either logarithmic closure or frication duration are the best acoustic cues to predict and discriminate singleton and geminate consonants.

4. DISCUSSION

The results of regression and discriminant analyses indicate that logarithmic duration is better than raw duration to predict and discriminate singleton and geminate consonants in a wide range of speaking rates. Human sound perception such as loudness or pitch is often related to logarithmic values of physical quantities such as intensity or frequency. Speech perception is also probably related to logarithmic values of physical quantities in a speech waveform. With regard to singleton and geminate consonants, speech perception would follow the logarithmic value of duration. Because speech perception and production are closely related [1], it is probable that speech production of singleton and geminate consonants is also based on the logarithmic value of duration. This notion is supported by the results of this study, at least for Japanese. However, the universality of this notion still needs to be confirmed by conducting similar studies on other languages.

This study revealed for the first time that logarithmic duration is better than raw duration at predicting and discriminating singleton and geminate consonants in Japanese. Previous studies [1-7] had focused only on raw duration and did not consider logarithmic duration, probably because raw duration exhibits passable performance for prediction and discrimination as shown by the results of this study. However, the results of the previous studies should be re-examined using logarithmic duration.

The results of this study indicate that logarithmic average mora duration in combination with logarithmic closure or fricative duration can predict and discriminate singleton and geminate consonants well. It suggests that speech planning depends on the mora unit to control the durational features of singleton and geminate in Japanese. In other words, the speaking rate is planned using the mora unit, and based on this speaking rate, the closure or frication duration is determined for singleton and geminate consonants. Therefore, our current results support the notion that the mora is the processing unit of the Japanese language [8].

A word item can be pronounced without a preceding phrase. In such a case, the average mora duration in the preceding phrase would not be available. However, as shown in Figures 1-4, overall the durations of subwords or successive vowels have approximately equal predictive and discriminatory performance to the average mora duration. So, it is likely that one of these durations is an appropriate acoustic cue that could be used for speech planning when a preceding phrase does not exist.

This study used only the plosive /k/ and the fricative /s/ for contrasting singleton and geminate consonants. However, affricates should also be examined in a future study. Devoicing, which was not treated in this study, is another factor that should be examined because the devoicing of vowels affects the durations of singleton and geminate consonants and their anteroposterior speech segments. Therefore, other new acoustic cues may be necessary for devoicing items. However, even if that is the case, logarithmic closure or fricative duration and average duration of the mora unit may be effective in the prediction and distinction of singleton and geminate consonants.

5. ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP25284080, JP15H03207, JP17K02705.

6. REFERENCES

[1] 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.

[2] Amano, S., Yamakawa, K. 2018. Variables of production boundary between singleton and geminate with plosive /t/, fricative /s/, and affricate /ch/. *Proceedings of Fall meeting of Acoustical Society of Japan*, 803-804 (in Japanese).

[3] Hirata, Y. 2007. Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels. *J. Phon. Soc. Jpn.* 11, 9-22.

[4] Hirata, Y., Amano, S. 2012. Production of single and geminate stops in Japanese three- and four-mora words. *J. Acoust. Soc. Am.* 132, 1614-1625.

[5] Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *J. Acoust. Soc. Am.* 118, 1647-1660.

[6] Idemaru, K., Guion-Anderson, S. 2010. Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica* 67, 25-46.

[7] Ofuka, E., Mori, Y., Kiritani, S. 2005. Perception of a Japanese geminate stop: the effect of the duration of the preceding/following vowel. *J. Phon. Soc. Jpn.* 9, 59-65 (in Japanese).

[8] Vance, T. J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.