PERCEPTUAL STUDIES OF JAPANESE GEMINATE INSERTION PHENOMENA BASED ON TIMING CONTROL CHARACTERISTICS

Mee Sonu[a,d], *Keiichi Tajima*[b], *Hiroaki Kato*[c] *& Yoshinori Sagisaka*[d]

[a]Faculty of Science and Technology, Sophia University, Japan;
[b]Department of Psychology, Hosei University, Japan;
[c]NICT, Japan;
[d]GITI and LASS, Waseda University, Japan
sonumee@toki.waseda.jp; tajima@hosei.ac.jp; kato.hiroaki@nict.go.jp; sagisaka@giti.waseda.jp

ABSTRACT

The present study investigated the identification of geminate and singleton consonants in Japanese by native Korean learners of Japanese. Moreover, the present study analyzed the relationship between the learners' identification errors and timing control factors. Results revealed a *geminate insertion* phenomenon, along with a systematic phonetic and perceptual explanation for the phenomenon. The analyses showed that native Korean listeners use multiple perceptual cues that depend on timing control factors such as speaking rate and segment duration.

Keywords: geminate & singleton stop, perceptual characteristics, L2, geminate insertion, time-loudness model

1. INTRODUCTION

Japanese words can be distinguished by the length of vowels and the length of consonants, e.g., singleton and geminate consonants. It has been shown that segmental duration is a primarily cue to the phonemic length distinction in Japanese [1]. However, it is known that the length contrast is difficult to perceive for non-native speakers. To assist the learning of length contrasts in Japanese, attempts have been made to build up an effective perceptual training paradigm for second language (L2) learners [7-9]. However, the learning effect was only to a limited extent. Results suggested that the learners were affected by timing control factors such as speaking rate, presentation context, and segmental content. To develop a more effective perceptual training method, it is necessary to understand learners' identification errors that depend on their L1, and to analyze the cause of the identification errors. It is necessary to confirm whether the identification errors are related to the timing control factors or not.

In the present study, the following timing control factors were considered: 1) duration of neighboring segments, and 2) segment duration difference weighted by time-loudness model.

Regarding the first timing control factor, segment duration is variable depending on intrinsic duration and neighboring phonemes [6]. However, it is possible that the learners are not familiar with such temporal variability in L2 speech. This may be a possible cause of the identification error by learners.

As a second timing control factor, the segment duration difference weighted by time-loudness model was considered. The time-loudness model has been proposed to evaluate the naturalness in timing control [2, 3], as well as naturalness in L2 speech [5]. By applying the time-loudness model to the singleton/geminate distinction in Japanese, it might become possible to provide an explanation for the learners' perceptual characteristics.

The present paper is organized as follows. As a first step, in section 2, native Korean learners participated in an identification test involving singleton and geminate consonants in Japanese, in order to find out the pattern of identification errors made by Korean learners. In section 3, possible cause of the identification errors was examined by analyzing their degree of correlation with timing control factors. In section4, the findings are interpreted in terms of native Korean learners' perceptual characteristics for Japanese.

2. PERCEPTUAL IDENTIFICATION TEST

2.1. Participants

73 Korean speakers who have studied Japanese participated (170 hours of Japanese learning on average). No participant reported a history of hearing or speech disorders.

2.2. Stimuli and procedure

Before taking the identification test, all participants were given a brief description of Japanese length contrasts, given in Korean. The test stimuli consisted of 15 real word pairs that contrasted in the singleton-geminate consonant distinction, e.g., *haken* "dispatch" vs. *hakken* "discovery". Each member of a pair differed from its counter part only in whether one of the consonants was singleton or geminate. The type of consonant in question was either plosive (6 pairs in total), fricative (6 pairs), or affricate (3 pairs). All word pairs were selected based on a Japanese lexical database [9]. All the stimuli were taken from the same speech database that was used in [9] (see [9] for details). The talker was a professionally trained female native Japanese talker who had been trained as a voice actor and spoke standard Tokyo Japanese comfortably. The talker produced each test word in isolation at three speaking rates: slow, normal, and fast.

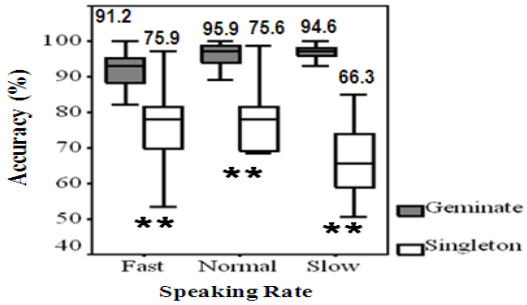
The test was a single-stimulus, two-alternative forced-choice identification test. On each trial, alphabetical transcriptions of two Japanese words comprising a minimal pair appeared as clickable buttons in the computer program window, and one of the two words was presented through headphones. Listeners' task was to select the word they heard by clicking the appropriate button. The trials were self-paced. The test consisted of 90 trials (30 words in isolation ×3 rates). The stimuli were presented once each in a random order.

2.3. Results and discussion

Figure 1 shows the results of the perceptual identification test. As shown by the low accuracy scores for singleton stimuli at all speaking rates, many singleton segments were mis-identified as geminate consonants. The native Korean listeners' accuracies were submitted to a univariate ANOVA with speaking rate (fast, normal and slow) and consonant type (geminate and singleton stops) as within-subjects variables and accuracy as the dependent variable.

Results revealed a no significant (n.s.) main effect of speaking rate [$F(2,84)=1.35$, n.s.], but a significant main effect of consonant type [$F(1, 84)=106.25$, $p<0.001$], and a significant speaking rate-by consonant type interaction [$F(2,84)=4.35$, $p<0.05$].

Figure 1: Results of the identification test for geminate and singleton consonants in Japanese by Korean speakers. (** Correlation is significant at the 0.01 level, * Correlation is signification at the 0.05 level).



Further analyses of the speaking rate-by-consonant type interaction using simple effects tests and multiple comparisons were conducted with Bonferroni correction. Accuracy for the geminate consonants was not significantly different among the speaking rates. In contrast, accuracy for the singleton consonants at a slow speaking rate was significantly lower than accuracies at other speaking rates ($p<0.05$). Moreover, there were significant differences between consonant types for all speaking rates ($p<0.001$). These results can be summarized as follows.

(1) Identification errors in which singleton consonants were mis-identified as geminates occurred much more frequently than errors in the opposite direction
(2) The identification errors were affected by speaking rate to some amount.

These error characteristics coincide with previous works by Min [4]. Min [4] found that Korean learners of Japanese produce the Japanese voiceless stops in intervocalic position as geminates. He called this phenomenon *geminate insertion*. Our observations confirmed that the geminate insertion phenomenon is not only plosive consonant contrasts but also affricative and fricative consonants contrasts by Korean speakers.

To further understand the cause of geminate insertion, we analyzed the singleton and geminate perception characteristics based on timing control characteristics.

3. FACTORS INFLUENCING PERCEPTUAL IDENTICFICATION

In this section, the geminate insertion phenomenon is analyzed in relation to timing control factors.

3.1. Correlation with duration of neighboring segments

One possible cause of the geminate insertion phenomenon may be that the learners concentrated on identifying the words based on segment duration alone, without considering other temporal characteristics.

To examine the above possibility, correlations between identification accuracy and the duration of the following segments were calculated: preceding mora, preceding vowel, closure (including voice onset time) of the target singleton or geminate consonant, and following vowel. Table 1 shows the correlations separately for the singleton and geminate consonants and for the three speaking rates.

Table 1: Correlation between the identification accuracy and stimuli of original duration categorized by speech unit (** Correlation is significant at the 0.01 level, * Correlation is signification at the 0.05 level).

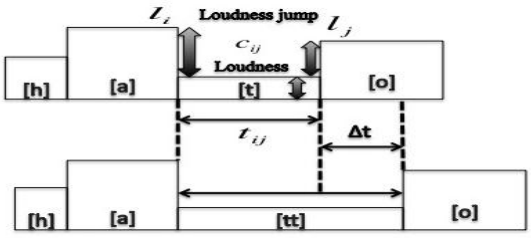
	Correlation coefficient			
	Pre Mora	Pre Vowel	CD with VOT	Post Vowel
Total (N=54)	-0.00	0.10	**0.69****	-0.11
Geminate (N=27)	0.20	0.29	**0.46***	-0.01
Singleton (N=27)	-0.34	**-0.43***	**-0.59****	-0.24
Fast speaking rate				
Geminate (N=9)	0.29	-0.16	-0.03	-0.49
Singleton (N=9)	-0.44	-0.48	-0.13	0.15
Normal speaking rate				
Geminate (N=9)	-0.31	-0.50	-0.20	-0.65
Singleton (N=9)	0.23	0.14	-0.31	-0.24
Slow speaking rate				
Geminate (N=9)	-0.54	-0.17	0.39	-0.13
Singleton (N=9)	-0.54	**-0.70***	**-0.93****	-0.07

Roughly speaking, the table shows that closure duration including voice onset time was highly correlated with identification accuracy of the geminate and singleton consonants by native Korean speakers. However, these correlation scores did not show the same pattern for all speaking rates. In particular, identification accuracy for singleton consonants at a slow speaking rate showed significant negative correlations with the preceding vowel duration and closure duration including voice onset time. These results suggest the possibility that the Korean learners made judgments based on absolute segment duration.

3.2. Correlation with segment duration different weighted by time-loudness marker model

To further examine factors related to learners' identification of singleton and geminate consonants, the time-loudness marker model [2, 3] was introduced. The model can predict a perceived difference between two words based not only on segment durations but also on segment durations with weighting proportional to the loudness of a segment. In a previous study, Nakamura, et al. [5] found that the time-loudness model predicts human perceptual difference between two utterances of the same text by using the duration and loudness of speech segments. By applying this model to the present study, it is possible that the geminate insertion phenomena shown by the Korean learners may be related to perceptual difference between singleton and geminate consonants.

Figure 2: An example showing the phone duration difference weighting using time-loudness marker model.



As a first step, we calculated the weighting factor (W_{ij}). The weighting factor considered loudness jump from the adjacent segments (l_i and l_j) and mean of the segment loudness between two markers (C_{ij}); the constant b was used for scale adjustment (1).

$$W_{ij} = b\frac{(l_i + l_j)}{2} + C_{ij} \quad (1)$$

Next, the perceptual difference (l_{ij}) was calculated using the segment duration difference (Δt) and the segment duration (t_{ij}) (2).

$$l_{ij}(\Delta t) \cong \frac{a \cdot W_{ij} \cdot \Delta t^2}{\sqrt{t_{ij}}} \quad (2)$$

Finally, the obtain values are summed up to obtain the perceptual difference (L).

$$L = \sum_{i=0}^{n-1} \sum_{j=i+1}^{n} l_{ij} \quad (3)$$

By adopting the time-loudness marker model, a value called Segment Duration Difference

Weighted by time-loudness model (SDDW) is obtained, which characterize the perceptual timing differences of the segments by weighting the loudness of the corresponding segments.

Table 2 shows the correlation between SDDW and the difference in accuracy between test words containing geminate and singleton consonants. The results indicated that segment duration difference weighting by time-loudness model could explain singleton and geminate identification error characteristics to some extent. However, a significant correlation was found only for the slow speaking rate. This suggests that it might be necessary to reconsider how to incorporate the time-loudness model in accounting for the singleton-geminate identification data. For example, a new parameter could be employed normalizing the contextual duration differences by factoring out other segmental duration control contributions.

Table 2: Correlation between difference accuracy and adapted factors.

	Speaking rate		
	Fast (N=9)	Normal (N=9)	Slow (N=9)
SDDW	0.01	0.13	**-0.91****

4. CONCULUSIONS

In the present paper, to understand L2 learners' identification error pattern and analyze the cause of the identification errors, we investigated the characteristics of native Korean speakers' identification of geminate and singleton consonants in Japanese. One important result from the identification test is that a geminate insertion phenomenon was observed. To analyze the cause of this phenomenon, the correlation between learners' identification performance and timing control factors was analyzed. The timing control factors that were considered were: 1) the duration of neighboring segments, and 2) the segment duration different weighted by time-loudness model.

The analyses revealed that native Korean speakers used multiple perceptual cues related to timing control factors such as speaking rate and segment duration. In addition, the introduction of Segment Duration Difference Weighted by time-loudness model (SDDW) suggested a new possibility to interpret perceptual tendencies based on timing perceptual characteristics dependent on loudness. Further analyses are needed to apply this analysis to more general phonetic contexts.

From an educational point of view, these results suggest that it is important to assist the learning of temporal differences rather than knowledge of phonological characteristics. More effective perceptual training methodology has been developed by focusing on temporal characteristics [7, 8].

5. ACKNOWLEDGEMENTS

This work was supported in part by Grant-in-Aid for Scientific Research B, No.20300069 of JSPS and No.2105072 of JSPS. We thank Aki Kunikoshi for cooperating by calculating loudness model.

6. REFERENCES

[1] Fujisaki, H., Nakamura, K., Imoto, T. 1975. Auditory perception of duration of speech and non-speech stimuli. In Fant, G., Tatham, M. (eds.), *Auditory Analysis and Perception of Speech*. 197-219, Academic Press.

[2] Kato, H., Tsuzaki, M., Sagisaka, Y. 1998. Acceptability for temporal modification of single vowel segments in isolated words. *J. Acoust. Soc. Am.*104, 540-549.

[3] Kato, H., Tsuzaki, M., Sagisaka, Y. 2002. Effects of phoneme class and duration on the acceptability of modifications in speech. *J. Acoust. Soc. Am.*111, 387-400.

[4] Min, K. 2007. The cause of the occurrence of geminate insertion: Evidence from Korean learners' production of the Japanese voiceless stops as geminates. *J. Phon. Soc. Jpn* 11(1), 58-70 (in Japanese).

[5] Nakamura, S., Matsuda, S., Kato, H., Tsuzaki, M., Sagisaka, Y. 2009. Objective evaluation of English learners' timing control based on a measure reflecting perceptual characteristics. *Proc. IEEE ICASSP*, 4837-4840.

[6] Sagisaka, Y., Tohkura, Y. 1984. Phoneme duration control for speech synthesis by rule. *Trans. IEICE.* J67-A, No.7, 629-636 (in Japanese).

[7] Sonu, M., Tajima, K., Kato, H., Sagisaka, Y. 2009. Perceptual training of singleton and geminate stops in Japanese language by Korean learners. *Proc. Interspeech*, 1747-1750.

[8] Sonu, M., Tajima, K., Kato, H., Sagisaka, Y. 2010. The effect of a word embedded in a sentence and speaking rate variation on the perceptual training of geminate and singleton consonant distinction. *Proc. Interspeech*, 1850-1853.

[9] Tajima, K., Kato, H., Rothwell, A., A-Yamada, R., Munhall, K. 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *J. Acoust. Soc. Am.*123, 397-413.

[10] Zwicker, E., Fastl, H., Widmann, U., Kurakata, K., Kuwano, S., Namba, S. 1991. Program for calculating loudness according to DIN 45631 (ISO 532B). *J. Acoust. Soc. Am.*12, 39-42.