ARTICULATORY TIMING OF THE JAPANESE SINGLETON AND GEMINATE /t/ PRODUCED BY SPEAKERS OF STANDARD CHINESE

Maho Morimoto[1, 2], Ai Mizoguchi[3, 4], Li Weiyu[1], Takayuki Arai[1]

[1]Sophia University, [2]JSPS, [3]Maebashi Institute of Technology, [4]NINJAL
maho.morimoto.jp@gmail.com, aimizoguchi@maebashi-it.ac.jp, w-li-5x4@eagle.sophia.ac.jp, arai@sophia.ac.jp

ABSTRACT

Consonant length contrast is known to be problematic to acquire for L2 learners of Japanese. Previous studies pointed out that the difficulty in producing the contrast can largely be attributed to the incomplete mastery of the timing control in the target language. In this study, we conducted an ultrasound experiment to investigate the realization of the consonant length contrast by L2 learners of Japanese with an L1 Standard Chinese background.

We report the results from acoustic and articulatory analyses of the contrast between /t/ and /t:/ with special focus on the gestural duration and articulatory timing. While certain acoustic durational aspects of the contrast were successfully learned by the L2 speakers examined in this study, the difference in the tongue tip gesture duration observed in the native speakers was not present for some of the learners. We discuss the relationship between the articulatory and acoustic timing control.

Keywords: Japanese geminates, Standard Chinese, L2, articulation, ultrasound.

1. INTRODUCTION

For L2 learners of Japanese, learning the quantity contrast in the language can be problematic, especially if their L1 does not make use of such contrasts [1, 2, 3]. Previous acoustic analyses of L2 productions of Japanese singletons (short consonants, as in /kata/ 'shoulder') and geminates (long consonants, as in /kat:a/ 'win-PAST') have revealed that issues in the realization of the contrast largely arise due to the difficulty in controlling the duration of the segments, especially the constriction duration which is known to be the primary perceptual cue to the contrast [1, 4, 5]. L2 learners are often reported to underdifferentiate singletons and geminates by producing either geminates that are too short in constriction duration, or singletons that are too long in duration [4, 6].

Issues in manipulating the segment duration extend to the duration of the vowels surrounding the singletons and geminates [4, 6]. Especially, in Japanese, vowels preceding geminates (V_1) are phonetically longer than vowels preceding singletons [7, 8]. This can be a notable challenge for learners with various L1 backgrounds, as it goes against the cross-linguistic tendency whereby the nucleus of a closed syllable shortens [9].

Previous studies on the articulation of geminate stops by native speakers of Japanese have discussed the duration of the closure gesture and its timing as a possible cause for the phonetic lengthening of V_1 [8, 10]. Several studies have shown that the tongue raising gesture at the offset of V_1 is longer in duration for geminates than for singletons [11, 12, 13, 14]. In addition, the onset of the tongue raising gesture relative to the acoustic onset of the consonant closure occurs earlier in geminates than in singletons [11]. Based on these findings, it has been pointed out that slower consonantal gestures allow for a more articulated and prolonged production of V_1 [8, 10].

While various acoustic studies have investigated the timing control in the realization of the consonant length contrast by L2 learners of Japanese, how the manipulation of segmental duration is implemented articulatorily has not been addressed directly. In this study, we report on a production experiment using ultrasound in order to investigate the lingual gestures and their relation to the acoustic durations in the singleton and geminate /t/ as produced by native speakers and Chinese-speaking learners of Japanese.

2. METHODS

2.1. Participants

We recruited 10 native speakers of Standard and Mandarin Chinese (CH) and 10 native speakers of Japanese (JP). In this study, we report the results from six of these speakers (three from each group), as shown in Table 1. We include three advanced learners (N1 level in the Japanese Language Proficiency Test) identifying themselves as native speakers of Standard Chinese (BCF01 and BCF04) or Mandarin Chinese (BCF05), and three native speakers of Japanese from Tokyo (BJF03 and BJF06) or an adjacent area (BJF07). All of them are female speakers in their 20s with normal hearing and speaking ability, residing in Japan at the time of the experiment. The Standard and Mandarin Chinese speakers were chosen for two reasons. First, their L1 phonology does not allow for a non-nasal obstruent coda, thus lacking obstruent

geminates; second, Chinese speakers constitute one of the largest Japanese learner populations [15, 16]. Advanced learners were chosen to ensure that they learned the consonant length distinction to some extent.

Instructions were provided in written and spoken Japanese and Chinese as necessary, and the participants consented to the experimental procedure approved by the Sophia University Committee for Ethics. A language background questionnaire was administered prior to participation. They were compensated for their participation.

Table 1: Linguistic background of the six speakers reported in this study.

Speaker	Age	Region	JP level
BCF01	24	Shandong/Tianjin	N1
BCF04	21	Shanghai	N1
BCF05	22	Beijing	N1
BJF03	20	Tokyo	Native
BJF06	22	Tokyo	Native
BJF07	21	Kanagawa	Native

2.2. Speech materials

The speech materials consisted of 16 Japanese bisyllabic words including /t/, /k/, /s/, and /ʃ/ as the target consonant (C_2 in $C_1V_1C_2V_2$). The words were varied in terms of the length of the target consonants and lexical pitch accent, and were embedded in a carrier phrase (/korewa ___ to i:masu/ 'this is called ___'). They were manually presented on a screen one at a time in semi-randomized order, using Japanese characters (both Hiragana and Kanji). We obtained 10 repetitions for each word. For the current study, we report the results for two words, namely, /hata/ 'field (as a surname)' for singletons (S), and /hat:a/ 'crawl-PAST' for geminates (G), both of which have a falling pitch accent.

2.3. Data collection

Audio and ultrasound recordings were made simultaneously for each speaker in a soundproof room, as they read the presented phrases aloud.

The acoustic signal was digitally recorded monaurally at 22,050 Hz to a laptop computer through an audio interface (Focusrite, Scarlett Solo 2nd Gen), using a RODE-NT2-A microphone.

Real-time mid-sagittal images of the oral cavity were recorded with an ultrasound system (MicrUS, EXT-1H) using a microconvex probe (MC10-5R10S-3). The probe was placed under the participant's chin, and an UltraFit Headset [17] was used to stabilize the relative position of the probe and the participant's head. The tongue images were recorded during the utterances.

The audio and ultrasound video were recorded and synchronized using AAA software [18]. The frame rate of the obtained ultrasound videos was 113 fps. Four tokens were excluded from the analysis due to synchronization error. Altogether, 116 tokens (2 consonant length × 10 repetitions × 6 speakers, minus the 4 excluded tokens) were subjected to acoustic and articulatory analyses.

2.4. Analysis

Data analysis was implemented using R [19], based on the acoustic and articulatory landmarks obtained through the following annotation procedures.

The audio recordings were annotated using Praat [20]. The segmental boundaries were identified in the waveform and spectrogram display based on the periodic cycles of the vowels.

Ultrasound images were annotated using the GetContours program [21]. Two articulatory landmarks were identified based on the tongue tip (TT; the most anterior point of the visible tongue contour): *GestOns* (onset of the TT raising gesture, or the frame in which the TT starts to rise) and *GestMax* (onset of the TT maximal constriction, or the frame in which the TT ceases to rise). In addition, an acoustic landmark, *CloOns* (acoustic onset of the consonant closure), was imported from the Praat TextGrids. The time difference between *GestOns* and *GestMax* was calculated to be analyzed as the duration of the TT raising gesture, or *TTGest*. The duration between *GestOns* and *CloOns*, as well as the duration between *CloOns* and *GestMax* were calculated to examine the timing of the TT raising gesture.

In order to evaluate the effect of consonant length on the durational variables, we performed a linear mixed effects model analysis [22] for each speaker group, with consonant length (*hata* or *hatta*) as the fixed effect and speakers as the random effect.

3. RESULTS AND DISCUSSION

First, we confirmed through auditory impression that the learners' productions of /t:/ were perceived as geminates (see 3.3. for an acoustic analysis).

3.1. TT gestural duration

Table 2 summarizes the durational results related to the TT raising gesture. Overall, the TT raising gesture was longer for geminates than for singletons among the Japanese speakers (β=24.95, t=4.53, p<.001), in accordance with previous literature. However, this trend was not observed among the Chinese speakers (β=4.12, t=0.9, p=0.37), suggesting that the learners did not vary the duration of the TT raising gesture across consonant length. Based on the view that V_1

lengthening is a by-product of the slower TT gesture duration, these results tentatively suggest that the L2 learners in this study have not mastered the timing control for Japanese geminates, including V_1 lengthening (however, see 3.3).

In the meantime, Fig. 1 shows that the TT gesture duration in singletons and geminates vary among the learners. Specifically, BCF05 showed the same tendency as the Japanese speakers.

Table 2: Average duration in milliseconds (sd) and geminate-to-singleton ratios (GSR) for *TTGest*, *GestOns~CloOns*, and *CloOns~GestMax*, per native language (NL) and consonant length (L).

NL	L	*n*	*TTGest*	*GestOns ~CloOns*	*CloOns~ GestMax*
CH	S	30	80.21 (19.55)	54.35 (12.27)	25.85 (12.66)
	G	29	83.89 (21.66)	56.53 (16.23)	27.36 (15.69)
GSR			1.05	1.04	1.06
JP	S	29	85.41 (24.72)	66.57 (18.34)	18.85 (13.06)
	G	28	110.81 (36.66)	77.44 (26.21)	33.37 (18.64)
GSR			1.30	1.16	1.77

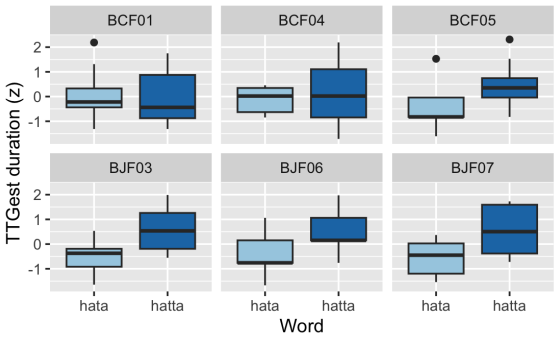

Figure 1: *TTGest* duration per word and speaker.

3.2. TT gestural timing

A similar trend is observed in the timing of the TT raising gesture relative to the acoustic onset of the closure. While the duration for *GestOns~CloOns* differed significantly for singletons and geminates among the Japanese speakers (β=10.52, t=2.34, p<.05), no such difference was observed among the Chinese speakers (β=2.2, t=0.6, p=0.55). The trend among the Japanese speakers conforms to the findings in [11], in that there was a greater overlap between V_1 and the TT raising gesture for geminates than for singletons. On the contrary, the timing with which the Chinese speakers initiated the TT raising gesture was the same across consonant length.

Fig. 2 again shows some interspeaker variability, but this time for both speaker groups. Early gestural onset for geminates was observed in learner BCF05, but not for the native speaker BJF03. The exceptional behavior of BCF05 is consistent with the distinction in the TT gestural duration, and suggests that she has acquired the gestural timing for Japanese geminates.

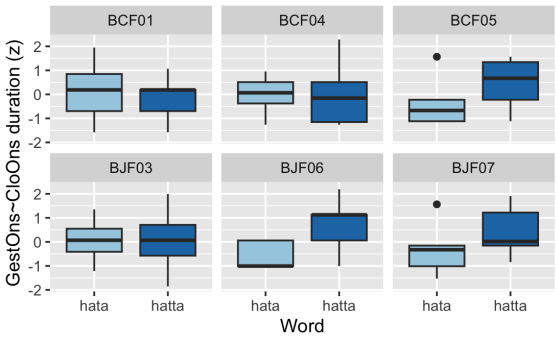

Figure 2: *GestOns~CloOns*, per word and speaker.

Comparison of the *CloOns~GestMax* durations reveals that overall, the TT maximal height was achieved later for geminates among the Japanese speakers (β=14.45, t=4, p<.001), but not among the Chinese speakers (β=1.83, t=0.62, p=0.53). Fig. 3 shows that the covert TT raising gesture continued after the acoustic onset of the consonant, and its duration was slightly longer for geminates than for singletons in native speakers.

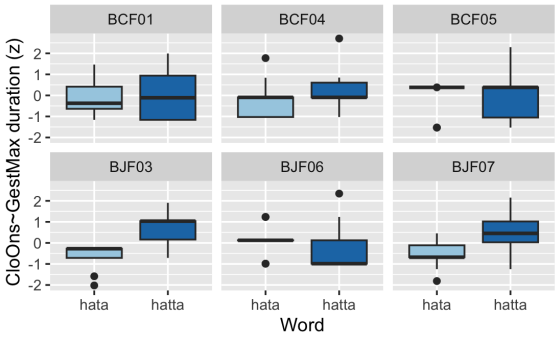

Figure 3: *CloOns~GestMax*, per word and speaker.

The difference was the largest for BJF03 (3.94 times longer for geminates on average), suggesting that the timing of the TT raising gesture onset in her geminate productions was no different from that of her singletons, but that completing the gesture took much longer. While BJF03's tendency deviates from those of the other Japanese speakers reported in this study, some variability in the timing control for geminates among native speakers has been previously reported [10]. Once a fuller analysis with more Japanese speakers is implemented, it may provide an insight into the typology of L1 articulatory strategies for the consonant length.

3.3. Acoustic duration

Table 3 summarizes the acoustic durations. The first point to note is the considerable difference in the overall speaking rate between the two groups, as

exemplified in the word duration (β=−280.48, t=−16, p<.001). Second, the acoustic results reveal that the Chinese speakers in this study were able to differentiate singletons and geminates in their productions at the segmental level. The geminate-to-singleton ratios for each segment were largely in line with those of the native speakers, and their C_2/Word ratios, known to have high classification accuracy at the boundary ratio of 0.35 [23], were 0.33 for singletons and 0.46 for geminates.

Table 3: Duration in milliseconds (sd) and geminate-to-singleton ratios (GSR) for word, V_1, C_2, and V_2 per native language (NL) and consonant length (L).

NL	L	*n*	Word	V_1	C_2	V_2
CH	S	30	489.53 (49.79)	51.3 (9.67)	159.83 (22.26)	141.96 (29.68)
	G	29	663.86 (62.43)	64.77 (14.2)	309.72 (66.52)	171.88 (35.84)
GSR			1.36	1.26	1.94	1.21
JP	S	29	249.29 (30.09)	47.92 (9.79)	82.05 (15.56)	60.1 (9.33)
	G	28	341.21 (45.39)	67.42 (10.4)	161.13 (27.24)	50.74 (8.13)
GSR			1.37	1.40	1.96	0.84

In addition, the learners have acquired the pre-geminate V_1 lengthening (CH: β=13.12, t=5.54, p<.001; JP: β=19.24, t=7.32, p<.001). This was unexpected given our findings on the TT gesture duration. However, not only the native speakers and BCF05, the exceptional learner, but also the other two Chinese speakers with a shorter TT gesture duration have mastered pre-geminate V_1 lengthening (Fig. 4).

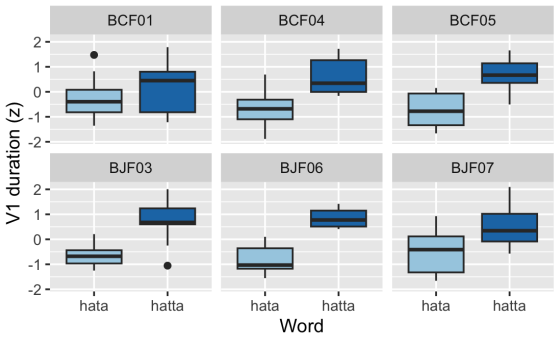

Figure 4: V_1 duration per word and speaker.

Meanwhile, although clear segmental distinctions were made by the learners, some of the durational correlates were not adequately acquired, as illustrated in Fig. 5. For example, the slight shortening of V_2 among the Japanese speakers (β=−9.58, t=−5.2, p<.001) was not shown by the Chinese speakers. Rather, those learners lengthened V_2 after geminates (β=29.87, t=4.31, p<.001). Furthermore, the ratio of the acoustic closure duration and V_1 duration was strikingly different between the two groups. This pattern is reminiscent of the over-exaggerated geminates in L2 speech reported in [1]. The slower speaking rate and the V_1 lengthening effect suggest that TT gesture duration was short for Chinese speakers.

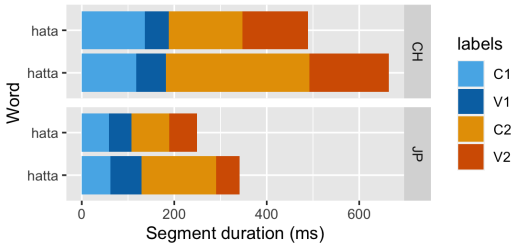

Figure 5: Bar chart for average segment duration.

4. CONCLUSION

The duration and timing of the TT raising gesture in the singleton and geminate /t/ were investigated, in relation to the acoustic segmental durations. Results showed that while native speakers of Japanese employed a longer TT raising gesture for geminates than for singletons, such an articulatory distinction was not made by the Chinese speakers except for one (BCF05). In addition, while the TT raising gesture started early relative to the acoustic consonantal onset in geminates for two out of the three Japanese speakers and the learner BCF05, that was not the case for the other two Chinese speakers and the other Japanese speaker (BJF03). This suggests that it is possible for the learners to acquire the gestural timing control for Japanese geminates, and that there may be some interspeaker variability among native speakers.

In the meantime, the acoustic V_1 lengthening in geminates was seen robustly across speaker groups. While this can be accounted for by the high proficiency of the speakers included in the current analysis, it also shows that it is possible for advanced learners of Japanese with an L1 Chinese background to manipulate the segmental duration to realize the perceptually relevant acoustic distinction, even without acquiring a native-like gestural timing. This finding is in support of the view that learners employ different strategies from that of native speakers to achieve timing control, possibly governed by a different rhythmic unit [4, 6, 8].

How the distinction is achieved articulatorily should be addressed in more details in future research. In particular, this study focused on the TT gesture of a limited number of advanced learners. Further analysis should include the articulation of the tongue as a whole, in more tokens produced by learners with various proficiency levels. Of special interest is the tongue positions during V_1, to address the magnitude of articulatory gestures before geminates and their relation to gestural duration.

5. ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP19K13254, JP22J01381, JP23K12212, and Sophia University Special Grant for Academic Research (Research in Priority Areas). We thank the anonymous reviewers for their comments.

6. REFERENCES

[1] Toda, T. 2003. Acquisition of special morae in Japanese as a second language. *Journal of the Phonetic Society of Japan* 7(2), 70–83.

[2] Tsukada, K., Yurong. 2022. Non-native perception of the Japanese singleton/geminate contrast: Comparison of Mandarin and Mongolian speakers differing in Japanese experience. *Interspeech 2022* 3068–3072.

[3] Lee, A., Mok, P. 2018. Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research* 34(4), 419–448.

[4] Toda, T. 1994. Interlanguage phonology: Acquisition of timing control in Japanese. *Australian Review of Applied Linguistics* 17(2), 51–76.

[5] Han, M. 1992. The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica* 49, 102–127.

[6] Yamakawa, K., Amano, S., Kondo, M. 2021. Mispronunciation of Japanese singleton and geminate stops by Korean and Taiwanese Mandarin speakers. *Acoust. Sci. & Tech.* 42(2), 73–82.

[7] Idemaru, K., Guion, S. 2008. Acoustic covariants of length contrast in Japanese stops. *J. Int. Phon. Assoc.* 38, 167–186.

[8] Fujimoto, M., Maekawa, K. 2014. Effects of sokuon on adjacent vowel duration: An analysis of the Corpus of Spontaneous Japanese. *Journal of the Phonetic Society of Japan* 18(2), 10–22.

[9] Maddieson, I. 1985. Phonetic cues to syllabification. In: Fromkin, V. A. (ed.), *Phonetic Linguistics*. New York: Academic Press, 203–221.

[10] Fujimoto, M., Funatsu, S., Hoole, P. 2015. Articulation of single and geminate consonants and its relation to the duration of the preceding vowel in Japanese. *Proc. of the ICPhS 2015*.

[11] Takada, M. 1985. Sokuon no chouon jo no tokucho ni tsuite [Articulatory characteristics of sokuon]. *Kenkyu Hokokushu* 6, 17–40.

[12] Ishii, T. 1999. A study of the movement of the articulatory organs in Japanese geminate production: An X-ray microbeam analysis. *J. Otolaryngol. Jpn.* 102, 622–634.

[13] Löfqvist, A. 2007. Tongue movement kinematics in long and short Japanese consonants. *J. Acoust. Soc. Am.* 122(1), 512–518.

[14] Fujimoto, M. 2013. Timing differences in articulation between single and geminate voiceless stop consonant: An analysis of cine-MRI data. Paper presented at the 3rd International Conference on Phonetics and Phonology, Tachikawa, Japan.

[15] Agency for Cultural Affairs, Government of Japan. 2021. Kokunai no nihongo kyoiku no gaiyo [Summary of domestic Japanese language education]. *Nihongo Kyoiku Jittai Chosa Hokokusho* https://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/nihongokyoiku_jittai/r03/pdf/93791201_01.pdf.

[16] The Japan Foundation. 2018. *Survey Report on Japanese-Language Education Abroad 2018*. https://www.jpf.go.jp/j/project/japanese/survey/result/dl/survey2018/Report_all_e.pdf.

[17] Spreafico, L., Pucher, M., Matosova, A. 2018. UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. *Interspeech 2018* 1517–1520.

[18] Articulate Instruments Ltd. 2022. Articulate Assistant Advanced (AAA). Version 2.19.08, http://www.articulateinstruments.com/.

[19] R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

[20] Boersma, P., Weenink, D. 2022. Praat: Doing phonetics by computer. Version 6.2.19, retrieved September 2022 from http://www.praat.org/.

[21] Tiede, M. K. 2022. *GetContours.* GitHub repository. https://github.com/mktiede/GetContours.

[22] Bates, D., Mächler, M., Bolker, B., Walker, S. 2014. Fitting linear mixed-effects models using lme4. *arXiv Prepr. arXiv1406.5823*.

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