音声研究　第 29 巻　103–113 頁
2025年

Journal of the Phonetic Society of Japan
Vol. 29, pp. 103–113, 2025

特集ノート

Tongue Contours during Pre-Geminate Vowel Lengthening in Japanese: A Case Study with Voiceless Alveolar Plosive /t/

Maho Morimoto*, Ai Mizoguchi**, Weiyu Li*** and Takayuki Arai***

促音に先行する母音の長化時の舌輪郭
―無声歯茎音/t/のケーススタディ―

SUMMARY: This study investigates the tongue contours during the production of the vowel /a/ preceding geminate and singleton /t/ in Japanese. We report acoustic and articulatory results from a production experiment using ultrasound imaging. Our results generally support the previous literature predicting a relationship between pre-geminate vowel duration and extreme formant values, as well as more extreme articulatory gestures. We also observed interspeaker variability, some speakers not behaving as expected. Our results suggest the potential role of gestural timing and the need to consider the gestural coordination between the pre-geminate vowel and the surrounding segments.

Key words: Japanese, vowel articulation, ultrasound, sokuon, geminates, pre-geminate vowel lengthening

1. Introduction

One of the acoustic correlates of Japanese consonant length contrast is the duration of the preceding vowels: the phonetic duration of vowels tends to be longer before geminates compared to singletons. This pre-geminate vowel lengthening has been shown to be robust in both laboratory settings (e.g., Campbell 1999; Fukui 1978; Han 1994; Hirata 2007; Idemaru and Guion 2008)[1)] and in spontaneous speech (Fujimoto and Maekawa 2014). Several perceptual studies also suggest that Japanese speakers are sensitive to the duration of preceding vowels when discriminating geminates from singletons (Arai and Kawagoe 1998; Kingston et al. 2009; Ofuka 2003; Ofuka, Mori and Kiritani 2005; Takeyasu 2012).

However, the articulatory mechanisms underlying the pre-geminate lengthening process are not fully understood. Previous articulatory studies have shown that the consonantal gesture towards geminate constriction is slower and takes longer than that for singleton constriction (Fujimoto 2013; Fujimura and Williams 2008; Ishii 1999; Löfqvist 2007; Takada 1985). Fujimoto and Maekawa (2014) point out that this delay in consonantal gesture may be related to the longer vowel duration, as the articulators linger at the articulatory position of the vowel for an extended time for geminates compared to singletons. Furthermore, they consider the possibility that vowels with pre-geminate lengthening have more extreme formant values (e.g., higher F1 values for /a/), given the more extreme jaw and tongue positions for pre-geminate vowels reported in previous articulatory studies (Fujimoto 2013; Takada 1985). However, experimental studies observing the actual behaviors of the articulators are still scarce, and their relationship to the acoustic signals has not been sufficiently investigated. Pioneering studies often employ a very limited number of speakers, and the findings are not always consistent. For instance, Takada (1985) showed that the pre-geminate vowel /a/ is realized with a larger jaw opening for /p/~/p:/, /t/~/t:/, and /h/~/h:/ but not for /k/~/k:/ by using X-ray motion picture data from a native speaker of Tokyo Japanese. On the other hand, Fujimoto (2013) showed that the pre-geminate vowel /a/ involves lower tongue position for /k/~/k:/ by using cine-MRI data from five speakers of Tokyo Japanese.

* Chuo University（中央大学）
** Maebashi Institute of Technology（前橋工科大学）／National Institute for Japanese Language and Linguistics（国立国語研究所）
*** Sophia University（上智大学）

Both studies used nonce-word stimuli. The current study seeks to contribute new data to the previous studies in order to better understand the articulatory nature of pre-geminate vowel lengthening. We aim to examine the hypothesis that pre-geminate vowels feature more exaggerated lingual constriction (e.g., lower tongue body for low vowels) and formant values compared to pre-singleton vowels.

This hypothesis is also backed up by previous literature on phonological length contrasts in Japanese, which points to the relationship between segment duration and the 'strength' of articulatory gestures. For example, EPG studies on Japanese geminates have shown that geminate stops are accompanied with longer and more extended linguopalatal contact than singletons (Kawahara and Matsui 2017; Kochetov 2012). These results are consistent with the view that gemination involves articulatory strengthening (e.g., Kochetov and Kang 2017; Payne 2006). For vowels, acoustic analyses suggest that phonologically long vowels in Japanese have more extreme formant values (i.e., occupy more peripheral positions in the acoustic vowel space) than short vowels (e.g., Hirata and Tsukada 2009; Nishi et al. 2008; Yazawa and Kondo 2019; Yazawa et al. 2023). This is also supported by an EMA study showing that phonologically long vowels are produced with more extreme (but less stable) gestures than with short vowels (Shirose, Kitamura and Nota 2021)[2]. However, pre-geminate vowel lengthening is a phonetic phenomenon as the increased duration does not lead to a change in word meaning, and it is unclear whether we should assume the same strengthening process as in the case of phonological lengthening.

In this study, we investigate the effect of the phonological length of the consonant on the lingual articulation of the preceding vowel and its acoustic consequence. By comparing the durational, spectral, and articulatory properties of the vowel /a/ preceding singleton /t/ and geminate /t:/, we examine whether pre-geminate vowels are realized with a lower tongue body position than pre-singleton vowels.

2. Method

2.1 Participants

We report data from 8 native speakers of Japanese (5 female, 3 male)[3], shown in Table 1. They reported normal hearing and speaking abilities and consented to the experimental procedure approved by the ethics committee at Sophia University. A language background questionnaire was administered prior to participation. They were compensated for their participation.

Table 1 Linguistic background of the speakers reported in this study (years in parentheses indicate the duration of residence outside the greater Kanto area of Japan).

Speaker	Age	Gender	Region
BJF01	18	F	Kanagawa
BJF03	20	F	Tokyo/Fukushima (1 year)
BJF05	21	F	Shizuoka
BJF06	22	F	Tokyo
BJF07	21	F	Kanagawa
BJM01	25	M	Tokyo
BJM02	20	M	Tokyo
BJM03	20	M	Tokyo/Singapore (4 years)

Table 2 List of words reported in this study.

C2 Length	Accent Pattern	Word	Gloss
Singleton	accented	/hata/	(surname)
Singleton	unaccented	/hata/	flag
Geminate	accented	/hat:a/	crawl-PAST
Geminate	unaccented	/hat:a/	paste-PAST

2.2 Speech Materials

The speech materials consisted of 16 real Japanese bisyllabic words with a C1V1C2V2 structure. Among these, singleton /t/ and geminate /t:/ were analyzed as the target consonant (C2) for this study. V1s with and without pitch accent were included in the analysis to control for the effect of accent (Table 2). The speech materials were presented on a screen one at a time in a semi-randomized order using Japanese characters (both Hiragana and Kanji), with manual control by the experimenter. Participants were instructed to read the words aloud, embedded in a carrier phrase (/korewa __ to i:masu/ 'this is called __'). We obtained 10 repetitions for each item. In the case of mispronunciations, speakers were asked to read the sentence aloud again.

The recordings were made in separate blocks according to the pitch accent pattern. The order of the blocks was counterbalanced across speakers.

2.3 Data Collection

We made simultaneous audio and ultrasound recordings 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 midsagittal 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 speaker's chin, and an UltraFit Headset (Spreafico, Pucher and Matosova 2018) was used to stabilize the relative position of the probe and the speaker's head.

The audio and ultrasound video were recorded and synchronized using the AAA software (Articulate Instruments Ltd. 2022). The frame rate of the obtained ultrasound videos was 113 FPS. 16 tokens were excluded from the analysis due to synchronization error. Altogether, 304 tokens (2 consonant length × 2 pitch accent patterns × 10 repetitions × 8 speakers, minus the 16 excluded tokens) were subjected to acoustic and articulatory analyses.

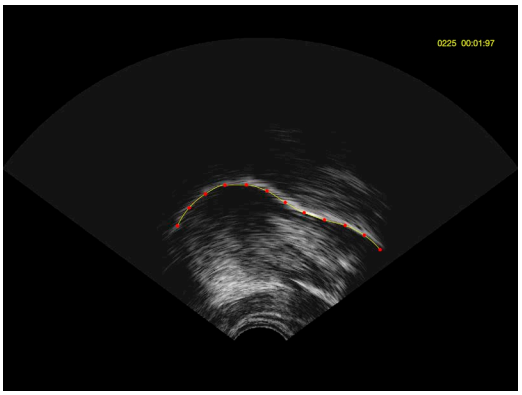

Figure 1 An example of midsagittal tongue contour tracking using GetContours. Tongue tip points to the right.

2.4 Analysis

For the acoustic analysis, the audio recordings were annotated using Praat (Boersma and Weenink 2024). The segmental boundaries were identified in the waveform and spectrogram display based on the periodic cycles of the vowels. The offset of the periodic cycle of V1 was used as the acoustic onset of C2 (i.e., the onset of the /t/ or /t:/ closure). We then measured the acoustic duration of V1 and C2 as well as the first formant (F1) value at the midpoint of the V1 acoustic interval.

For the articulatory analysis, ultrasound images were annotated using the GetContours program (Tiede 2021). First, to examine the tongue body position during the vowel articulation, the midsagittal tongue contours at the midpoint of the V1 acoustic interval were manually traced (Figure 1) and extracted for analysis. Second, to explore the gestural characteristics of V1, we identified the V1 maximum constriction (henceforth, *V1 max*). Among the frames where the tongue body was at its lowest position, we selected the frame immediately before it began to rise towards the alveolar closure[4] (i.e., the offset of V1 constriction). For the frame selection, two researchers (the first and the second authors) independently identified the target frames. If there were discrepancies of more than two frames (which occurred in no more than 5% of the overall data), the frames were re-examined to reach a consensus. The timestamp of the *V1 max* frame was extracted, and the duration between the *V1 max* frame and C2 acoustic onset (*V1 max to C2 onset duration*) was used for the gestural timing analysis. Figure 2 provides a schematic representation of the acoustic and articulatory landmarks used in this study.

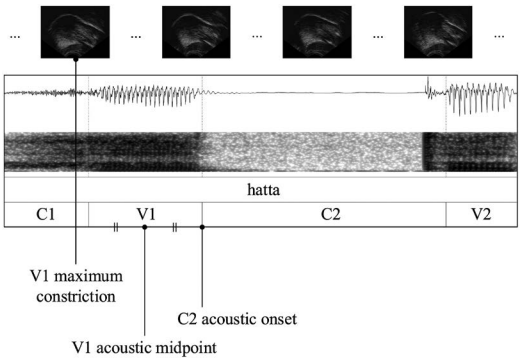

Figure 2 Schematic representation of the acoustic and articulatory landmarks examined in this study.

For each acoustic and gestural timing measure, we calculated the geminate-to-singleton ratio (GSR). Values beyond two standard deviations from the mean for each speaker were excluded from the analysis. For statistical analyses, we fitted linear mixed-effects models using the *lme4* package (Bates et al. 2015) and *lmerTest* package (Kuznetsova, Brockhoff and Christensen 2017) in R (v4.3.3, R Core Team 2021) with V1 duration (ms), F1 values (Hz) of V1, and *V1 max to C2 onset duration* (ms) as response variables. Consonant length and pitch accent pattern were included as fixed effects, along with their interaction. Speaker was included as a random effect to account for individual differences. Random slopes for consonant length and pitch accent pattern were included to capture variation in how each speaker responds to these factors. Additionally, as part of a post-hoc analysis, we constructed a model with F1 (Hz) as a response variable, and V1 duration (ms) and pitch accent pattern as fixed effects. We

also conducted a Pearson correlation test to examine the relationship between the degree of V1 lengthening and F1 raising.

3. Results

3.1 Acoustic Characteristics of the Vowels

Table 3 shows the durational characteristics of /hata/ and /hat:a/. Overall, V1 duration in the pre-geminate condition is significantly longer than in the pre-singleton condition, with a raw mean difference of 20.5 ms ($\beta = 18.36$, $SE = 2.85$, $t = 6.44$, $p < 0.001$, $n = 304$). Table 4 summarizes the mean values and GSR of each measure for each speaker. Together with Figure 3, it shows that while the tendency for pre-geminate lengthening is largely consistent across speakers, there is some interspeaker variability such that some speakers (BJF07, BJM01, and BJM03) lengthen V1 less than others.

Table 3 Mean acoustic duration (ms) and GSR (geminate-to-singleton ratio) of V1 and C2.

C2 Length	*n*	V1 Duration (SD)	C2 Duration (SD)
Singleton	152	42.07 (10.21)	83.63 (14.96)
Geminate	152	62.57 (11.18)	178.25 (28.72)
GSR		1.49	2.13

On average, F1 in the pre-geminate condition is slightly but significantly higher than in the pre-singleton condition, with a raw mean difference of 64.98 Hz ($\beta = 62.88$, $SE = 16.52$, $t = 3.81$, $p < 0.01$, $n = 299$), suggesting a lower tongue position for pre-geminate vowels than for pre-singleton vowels (overall GSR = 1.1)[5]. However, Figure 4 and Table 4 indicate interspeaker variation. Notably, F1 raising

Table 4 Average values (SD) and GSR per speaker.

Speaker	C2 Length	V1 Duration (ms)	F1 (Hz)	V1 Max to C2 Onset Duration (ms)
BJF01	Singleton	37.59 (8.78)	700.94 (51.17)	89.49 (12.33)
	Geminate	66.62 (8.59)	763.64 (50.2)	93.02 (14.79)
	GSR	1.77	1.09	1.04
BJF03	Singleton	44.53 (6.45)	626.92 (53.01)	76.28 (15.74)
	Geminate	63.06 (7.71)	704.75 (60.09)	88.35 (12.59)
	GSR	1.42	1.12	1.16
BJF05	Singleton	38.76 (12.51)	671.04 (90.86)	60.58 (9.35)
	Geminate	63.02 (11.01)	811.75 (69.89)	61.13 (12.06)
	GSR	1.63	1.21	1.01
BJF06	Singleton	38 (11.08)	699.76 (114.55)	54.45 (4.14)
	Geminate	72.25 (8.25)	816.06 (71.32)	61.5 (5.85)
	GSR	1.9	1.17	1.13
BJF07	Singleton	51.54 (10.59)	680.83 (71.23)	79.64 (17.43)
	Geminate	71.11 (11.45)	702.91 (50.09)	81.5 (20.6)
	GSR	1.38	1.03	1.02
BJM01	Singleton	41.83 (11.19)	630.1 (29.54)	67.16 (9.11)
	Geminate	52.21 (8.15)	635.74 (36.55)	88.02 (11.89)
	GSR	1.25	1.01	1.31
BJM02	Singleton	41.15 (6.95)	526.07 (38.42)	62.33 (8.39)
	Geminate	57.2 (9.13)	567.67 (25.51)	67.51 (9.26)
	GSR	1.39	1.08	1.08
BJM03	Singleton	42.7 (6.36)	650.18 (44.17)	67.2 (10.78)
	Geminate	54.96 (7.08)	683.12 (32.77)	61.41 (8.7)
	GSR	1.29	1.05	0.91

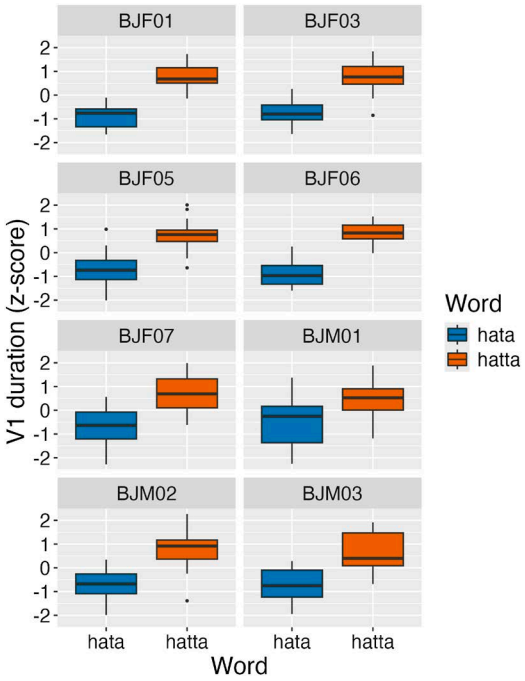

Figure 3 V1 duration (normalized) per speaker.

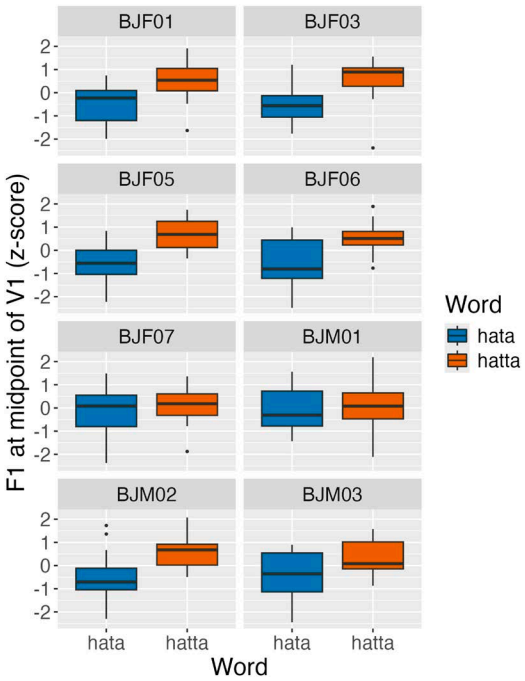

Figure 4 F1 at midpoint of V1 (normalized) per speaker.

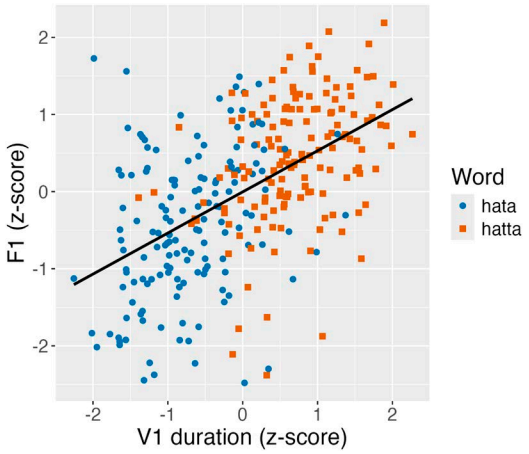

Figure 5 The relationship between V1 duration and F1 (normalized).

is not observed in speakers with a lesser degree of V1 lengthening (BJF07, BJM01, and BJM03). In fact, there was a significant effect of V1 duration on F1 ($\beta = 2.09$, $SE = 0.41$, $t = 5.09$, $p < 0.001$, $n = 299$), such that a 1 ms increase in V1 duration increased F1 for about 2 Hz. The relationship between V1 duration and F1 is also visualized in Figure 5. The regression line indicates that F1 tends to increase as V1 duration gets longer ($r = 0.53$), showing an overall moderate correlation. To summarize, speakers employ various degrees of pre-geminate V1 lengthening, and those with less lengthening tend to show less or no F1 raising.

3.2 Tongue Contours of the Vowels

Figure 6 shows the midsagittal tongue contours at the midpoint of the acoustic interval of V1, predicted from the repetitions with 95% confidence intervals using the generalized additive model (GAM; Wood 2006)[6]. Most speakers show tongue body lowering in the pre-geminate condition, although the effect is more subtle in some compared to others. A notable exception is BJF07, whose tongue contours show no evidence of tongue lowering associated with consonant length. This is consistent with the acoustic results, as BJF07 shows less pre-geminate V1 lengthening than other speakers and shows no F1 raising. BJM03, the other speaker with less V1 lengthening, shows a slight tongue lowering effect, but it is not as robust as some other speakers with more V1 lengthening. Finally, BJM01 shows a unique gestural pattern in which the tongue tip raising for the consonantal articulation starts earlier than others. This may explain the lack of F1 raising despite the

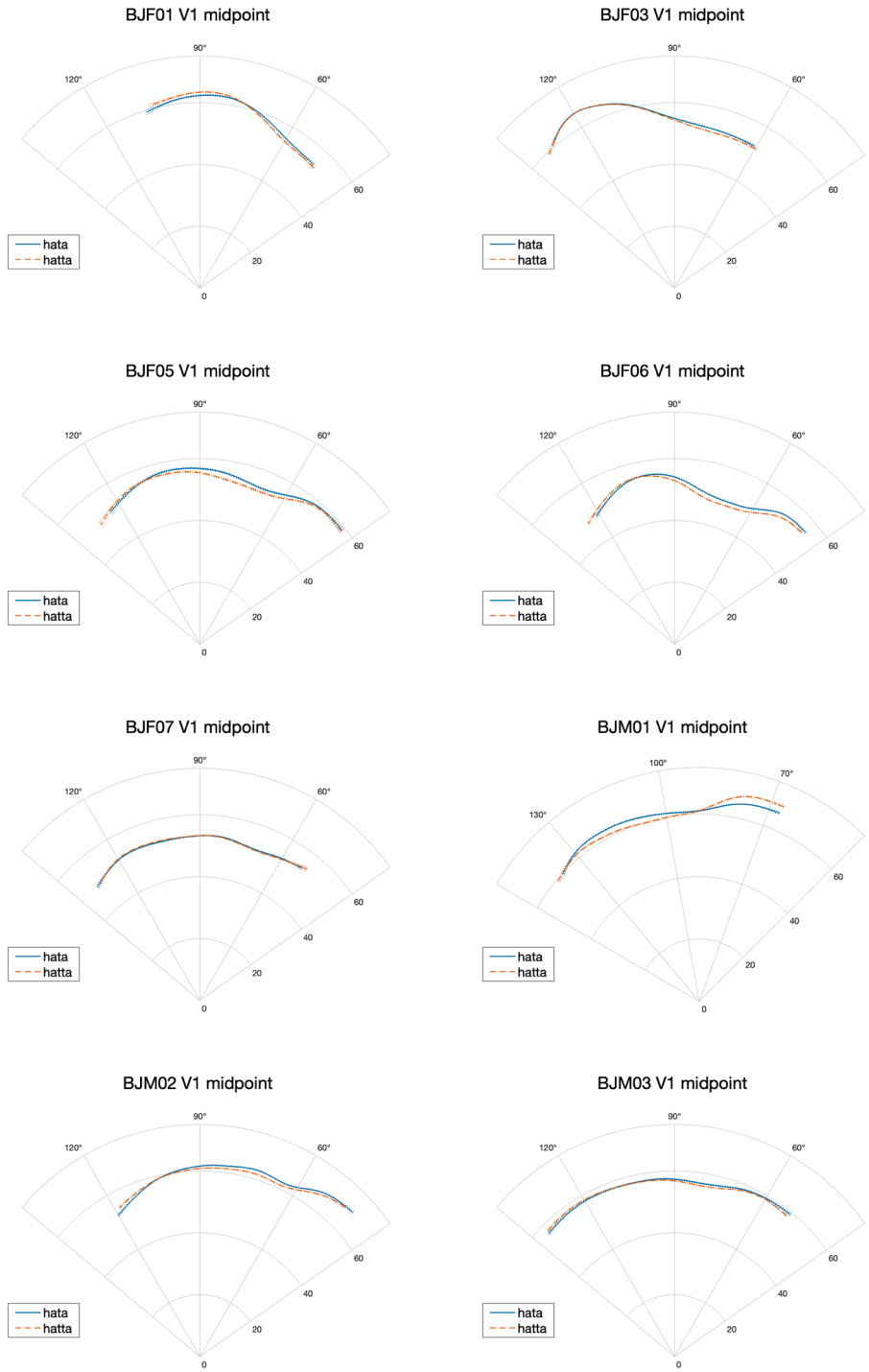

Figure 6 **GAM (Generalized Additive Model) contours for the repetitions of /hata/ (blue solid lines) and /hat:a/ (orange dashed lines) at the acoustic midpoint of V1. The radial axis values are in millimeters. Dotted lines indicate the 95% confidence intervals. Tongue tip points to the right.**

presence of substantial tongue lowering in BJM01.

Overall, five out of the eight speakers (BJF01, BJF03, BJF05, BJF06, and BJM02) fully conform to the patterns predicted from the literature. They consistently show tongue body lowering for pre-geminate vowels, and together with V1 lengthening and F1 raising, their patterns are in support of a relationship between duration and the magnitude of the vocalic gesture. Furthermore, the tongue contour results for the other two speakers, BJF07 and BJM03, do not necessarily contradict the view that there is a relationship between duration and magnitude of the movement. In fact, their insufficient tongue lowering for pre-geminate vowels is consistent with the fact that they do not show a V1 lengthening or F1 raising effect as much as in the aforementioned group. The exceptional behavior of BJM01, who shows a robust tongue lowering effect but no F1 raising, as mentioned earlier, is likely attributable to the gestural coordination of the vocalic (tongue body) and consonantal (tongue tip or blade) gestures. A more dynamic approach would be necessary to further examine interarticulator timing and its acoustic consequences.

3.3 Gestural Timing of the Vowels

So far, our results are in favor of the view that pre-geminate vowel lengthening, when acoustically present, involves more extreme gestures that give rise to a more extreme formant value, albeit with some interspeaker variability. However, the gestural timing results in Figure 7 and Table 5 suggest that a more careful analysis is in order. First, our results are not quite in line with the previous studies showing that the gesture towards the consonantal closure is slower and takes longer for geminates compared to singletons. While the trend is there, with a raw mean difference of 5.45 ms, it is not statistically significant ($\beta = 3.11$, $SE = 3.01$, $t = 1.03$, $p = 0.32$, $n = 296$). In fact, Table 4 and Figure 7 show that *V1 max to C2 onset duration* exhibits considerable interspeaker variability, and that there is no straightforward correspondence between the robustness of the acoustic V1 lengthening and *V1 max to C2 onset duration*. This is possibly attributable to the fact that the gestural timing we measured was based on the offset of V1 maximum constriction relative to the acoustic onset of C2, which may not precisely correspond to the onset of C2 gesture (where the tongue tip/blade begins to rise) and the offset of C2 gesture (where the tongue tip/blade stops rising).

Nonetheless, it is possible to identify three patterns in the data. First, BJF07 and BJM03 show no tongue

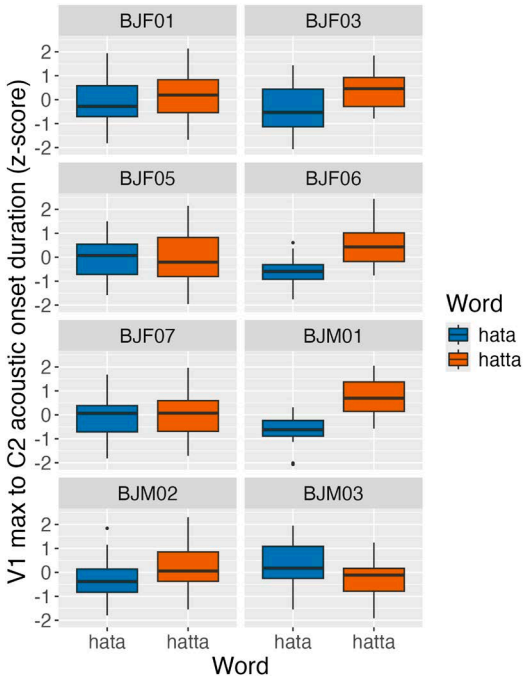

Figure 7 ***V1 max to C2 onset duration*** **(normalized) per speaker.**

Table 5 Mean duration (ms) and GSR between V1 max and C2 acoustic onset.

C2 Length	*n*	V1 Max to C2 Acoustic Onset
Singleton	149	69.83 (15.55)
Geminate	147	75.28 (17.90)
GSR		1.08

lowering, less V1 lengthening, and no substantial gestural lengthening, or lengthening of *V1 max to C2 onset duration* (in fact, BJM03 shows gestural shortening) for geminates. Second, BJF03, BJF06, and BJM01 all show some degree of gestural lengthening as well as tongue lowering. Third, BJF01, BJF05, and BJM02 show almost no gestural lengthening despite the presence of robust V1 lengthening and tongue lowering. Thus, it appears that while acoustic V1 lengthening is a robust effect, its articulatory implementation may not be uniform across speakers. While it is not surprising that we observe interspeaker variability as previous studies have documented intra- and inter-speaker variability in the articulatory gestures related to Japanese geminates (Ishii 1999; Löfqvist 2007; Morimoto et al. 2023), a more dynamic analysis of the lingual gestures is needed

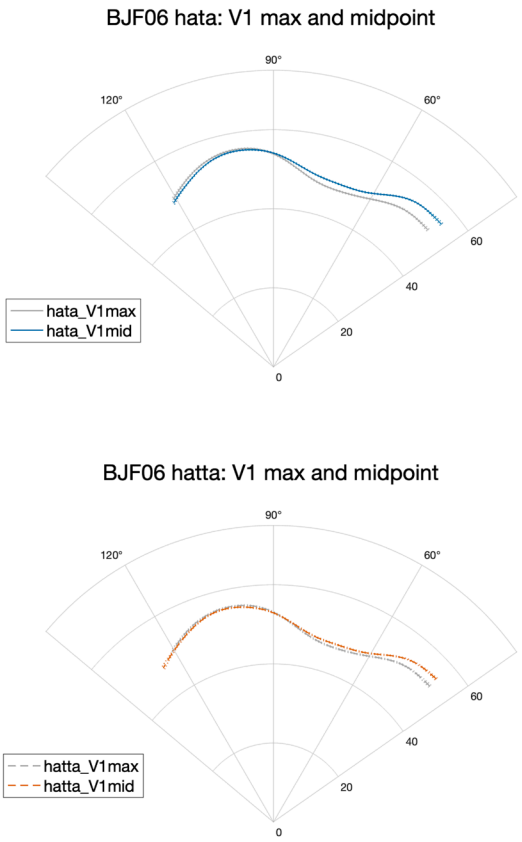

Figure 8 **GAM (Generalized Additive Model) contours for the repetitions of /hata/ (top panel) and /hat:a/ (bottom panel). For each condition, the gray line is based on the *V1 max* frame. The blue solid line (/hata/) and orange dashed line (/hat:a/) is based on the V1 acoustic midpoint. The radial axis values are in millimeters. Dotted lines indicate the 95% confidence intervals. Tongue tip points to the right.**

to better understand each of these patterns.

In the meantime, the values for *V1 max to C2 onset duration* in Table 4 suggest that in our data, the offset of V1 maximum constriction (*V1 max*) generally occurs before the V1 acoustic onset (i.e., during C1 or /h/ in /hata/ and /hat:a/). This means that the tongue contours in Figure 6 (extracted at V1 acoustic midpoint) do not necessarily represent the lowest possible position of the tongue in the utterance of the syllable. As can be seen in Figure 8, showing a representative relationship between the tongue positions in *V1 max* and V1 acoustic midpoint for /hata/ and /hat:a/, the tongue leaves its lowest position at *V1 max*, and the V1 acoustic midpoint captures the tongue position in the middle of its way up towards the consonantal closure. Because /h/ potentially requires a lower and more retracted position of the tongue than /a/, it is not possible for us to determine which of the following better represents the gestural target of V1: the tongue position at *V1 max* or the tongue position of the V1 acoustic midpoint[7]. It would therefore be premature to conclude from our data that pre-geminate vowels have a more extreme gestural target, as is sometimes claimed in the case of phonological vowel lengthening, or that the longer V1 duration allows sufficient time for the achievement of a more extreme gesture. Yet, our results provide insights into the relationship between the V1 formant values and the timing of the tongue body raising gesture accompanying C2. For example, in BJF03 and BJF06, who exhibit higher F1 values in geminates, as well as BJM01 (although coarticulation seems to mask the effect in F1), the tongue reaches a lower point at the V1 acoustic midpoint for geminates even though their tongue body raising gesture starts relatively early. Results for these speakers are consistent with previous findings that the gesture towards consonantal closure is slower and longer for geminates. Similarly, the lack of formant and tongue displacement in BJF07 and BJM03 may be accounted for by the lack of difference in the timing of tongue body raising. However, in order to account for the remaining three speakers (BJF01, BJF05, and BJM02) who show F1 raising and tongue lowering but no difference in gestural timing, further examination including gestural velocity and gestural extrema would be necessary[8].

4. Discussion and Conclusion

In this study, we investigated the acoustic and articulatory realization of vowels in pre-geminate and pre-singleton conditions. Our results are generally in line with the view expressed in the previous literature that pre-geminate vowels are more acoustically peripheral and involve more extreme articulatory gestures than pre-singleton vowels. Durational measurements of V1 as well as spectral and articulatory measurements at the midpoint of V1 suggest that while there is considerable interspeaker variability, F1 raising and tongue lowering tend to concur with V1 lengthening. We discuss three issues and further questions that arise from our results below.

First, it should be noted that since our data is limited to the low vowel /a/, further examination is needed to clarify whether the F1 raising and tongue lowering

effects related to pre-geminate lengthening are part of a vowel space expansion process or a formant displacement process. To address this issue, it would be necessary to investigate the behavior of other vowels in the Japanese vowel system as well as to examine the relationship between F2 values, tongue retraction and fronting.

Second, the examination of the position and gestural timing of the tongue body using ultrasound imaging revealed that the extrema of the lingual articulation of V1 do not necessarily coincide with the vocalic interval. In our data, the lowest position of the tongue body around V1 predominantly occurred during C1. On the one hand, it is conceivable that this maximum of tongue lowering actually belongs to the articulatory gesture of the glottal fricative /h/; on the other hand, while there were no instances of entirely voiceless V1, it is possible that there was some partial devoicing of V1, which resulted in the achievement of the lingual extrema for V1 before its voicing period. The onset /h/ is thought to have minimal coarticulatory influence on the nucleus (Nishi et al. 2008) and has also been used for articulatory investigation (Löfqvist 2007), but further study is needed to clarify the potential effect of C1 and to draw a definitive conclusion as to whether phonetically lengthened vowels have a more extreme gestural target.

Finally, our gestural timing results did not support the previous findings for all speakers, as the lingual gesture towards the geminate closure did not consistently take longer than that towards the singleton closure. One possible explanation for this discrepancy is our use of the duration between *V1 max* (offset of V1 maximum constriction) and the acoustic onset of C2, where most of the previous studies measured the duration between the onset and offset of the consonantal gesture. We used *V1 max* as we focused on the V1 articulation in this study, but we would need to examine the gestural organization of vowels and consonants more dynamically in the future. In particular, in our data, *V1 max* and C2 gestural onset were usually within a few frames of each other, implying that the lag between these two gestural landmarks is not what influenced our results. Rather, it is likely that the lag between the achievement of the C2 constriction gesture (where the tongue tip/blade ceases to rise) and C2 acoustic onset is accountable for the difference between the gestural duration for singleton and geminates. All in all, further investigation is needed in order to explore the relationship between gestural duration and timing, gestural strength, and their acoustic consequences.

Ultimately, the current study should be viewed as a case study as it examines only a limited number of target segments and phonological environments, making it difficult to draw broad generalization. Future analysis will consider a wider variety of environments, especially taking into account the quality of vowels in V1 position, consonant types in C1 and C2, and the effect of pitch accent pattern, speech rate, as well as individual variability. Furthermore, a dynamic analysis of the tongue contours would be beneficial in achieving a better understanding of the articulatory mechanism underlying pre-geminate vowel lengthening.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Numbers JP19K13254, JP22J01381/22KJ2752, and JP23K12212. We thank the two anonymous reviewers for their insightful comments and suggestions, which helped improve this manuscript. Any remaining errors are our own.

Notes

1) Interspeaker variation regarding pre-geminate vowel lengthening was reported in Fujimoto, Funatsu and Hoole (2015), based on one of the two native speakers of Japanese who did not lengthen the vowel duration before geminates, in an experiment using nonce-word stimuli and a German carrier phrase.
2) Some discrepancies between the results of Shirose, Kitamura and Nota (2021) and Ishii (1999) are reported. Shirose, Kitamura and Nota (2021) discuss that the difference may be potentially attributable to the lexical status of the words (as they used real words, where Ishii (1999) used nonce-words), or the measurement location of the tongue.
3) Ten native speakers of Japanese participated in the experiment. In the current analysis, two speakers were excluded due to their extensive residence history abroad and outside of the greater Kanto area of Japan.
4) It was often impossible to determine from the ultrasound image alone whether the consonantal articulation of /t/ was apical or laminal. For frame selection, we focused on tongue body movements to avoid any potential effect of the individual variability related to the portion of the tongue involved in the articulation of /t/, and to capture the tongue shape in its lowest position within the syllable. Note that *V1 max* is not the same as *GestOns* in Morimoto et al. (2023), which is defined as the "onset of the TT raising gesture, or the frame in which the TT starts to rise" (where TT refers to the tongue tip). *V1 max* is identified based on the tongue body gesture, whereas

GestOns is identified based on the tongue tip gesture. While they often coincide, there are cases where they do not.

5) In this study, we are concerned with the articulatory differences in pre-geminate and pre-singleton conditions and its acoustic consequences. Whether a 60 Hz difference in the formant values is perceptible or available as one of the cues for gemination is debatable, yet outside the scope of the current study.

6) The 95% confidence intervals are generally very small given the consistency of the tongue contours across repetitions and may only be visible around the tongue root and tongue tip.

7) An anonymous reviewer noted that the data for BJF05 and BJF06 deviate from the pattern in which *V1 max* occurs during C1, possibly suggesting that /h/ has a pharyngeal place of articulation for these speakers (Yamane 2013), which could prevent the gestural target of /a/ from being realized during C1.

8) A comparison of the tongue contour at *V1 max* for BJF05 and BJM01 reveals that the tongue body is lower for geminates than for singletons, which may account for the F1 raising (in BJF05) and the lowered tongue body at the V1 acoustic midpoint, despite the lack of difference in gestural timing. As discussed in section 4, further study is necessary to examine the effect of C1 on this result.

References

Arai, M. and I. Kawagoe (1998) "Eigo no onsetsugata to sokuon chikaku: Nansensugo niyoru chikakutesuto hōkoku [Consonant gemination and syllable types in English: a perceptual test of nonsense words]." *Journal of the Phonetic Society of Japan* 2(3), 87–92.

Articulate Instruments Ltd. (2022) Articulate Assistant Advanced (AAA). Version 219.8. Edinburgh, UK: Articulate Instruments Ltd., http://www.articulateinstruments.com/ (accessed May 25, 2022)

Bates, D., M. Mächler, B. Bolker and S. Walker (2015) "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software* 67(1), 1–48.

Boersma, P. and D. Weenink (2024) Praat: Doing phonetics by computer. Version 6.4.06. http://www.praat.org/ (accessed February 26, 2024)

Campbell, N. (1999) "A study of Japanese speech timing from the syllable perspective." *Journal of the Phonetic Society of Japan* 3(2), 29–39.

Fujimoto, M. (2013) "Timing differences in articulation between single and geminate voiceless stop consonants: An analysis of cine-MRI data." *3rd International Conference on Phonetics and Phonology*.

Fujimoto, M., S. Funatsu and P. Hoole (2015) "Articulation of single and geminate consonants and its relation to the duration of the preceding vowel in Japanese." *Proceedings of the 18th International Congress of Phonetic Sciences*.

Fujimoto, M. and K. Maekawa (2014) "Sokuon ni rinsetsusuru boin no jikanchō no tokuchō ni tsuite: CSJ no bunseki [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.

Fujimura, O. and J. C. Williams (2008) "Prosody and syllables." *Onin Kenkyu [Phonological Studies]* 11, 65–74.

Fukui, S. (1978) "Nihongo heisaon no enchō/tanshuku niyoru sokuon/hisokuon toshite no chōshu [Perception for the Japanese stop consonants with reduced and extended durations]." *Onsei Gakkai Kaihō [The Bulletin of the Phonetic Society of Japan]* 159, 9–12.

Han, M. S. (1994) "Acoustic manifestations of mora timing in Japanese." *Journal of the Acoustical Society of America* 96(1), 73–82.

Hirata, Y. (2007) "Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels." *Journal of the Phonetic Society of Japan* 11(1), 9–22.

Hirata, Y. and K. Tsukada (2009) "Effects of speaking rate and vowel length on formant frequency displacement in Japanese." *Phonetica* 66(3), 129–149.

Idemaru, K. and S. G. Guion (2008) "Acoustic covariants of length contrast in Japanese stops." *Journal of the International Phonetic Association* 38(2), 167–186.

Ishii, T. (1999) "A study of the movement of articulatory organs in Japanese geminate production: An X-ray microbeam analysis." *Journal of Otolaryngology of Japan* 102, 622–634.

Kawahara, S. and F. M. Matsui (2017) "Some aspects of Japanese consonant articulation: A preliminary EPG study." *ICU Working Papers in Linguistics* 2, 9–20.

Kingston, J., S. Kawahara, D. Chambless, D. Mash and E. Brenner-Alsop (2009) "Contextual effects on the perception of duration." *Journal of Phonetics* 37(3), 297–320.

Kochetov, A. (2012) "Linguopalatal contact differences between Japanese geminate and singleton stops." *Canadian Acoustics* 40(3), 28–29.

Kochetov, A. and Y. Kang (2017) "Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean." *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62(1), 18–55.

Kuznetsova, A., P. B. Brockhoff and R. H. B. Christensen (2017) "lmerTest package: Tests in linear mixed effects models." *Journal of Statistical Software* 82(13), 1–26.

Löfqvist, A. (2007) "Tongue movement kinematics in long and short Japanese consonants." *Journal of the Acoustical Society of America* 122(1), 512–518.

Morimoto, M., A. Mizoguchi, W. Li and T. Arai (2023) "Articulatory timing of the Japanese singleton and geminate /t/ produced by speakers of Standard Chinese." *Proceedings of the 20th International Congress of Phonetic Sciences*, 2691–2695.

Nishi, K., W. Strange, R. Akahane-Yamada, R. Kubo and

S. A. Trent-Brown (2008) "Acoustic and perceptual similarity of Japanese and American English vowels." *Journal of the Acoustical Society of America* 124(1), 576–588.

Ofuka, E. (2003) "Sokuon /tt/ no chikaku: Akusentogata to sokuon/hisokuongo no onkyōteki tokuchō niyoru chigai [Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels]." *Journal of the Phonetic Society of Japan* 7(1), 70–76.

Ofuka, E., Y. Mori and S. Kiritani (2005) "Sokuon no chikaku ni taisuru senkō kōzoku boinchō no eikyō [Perception of a Japanese geminate stop: The effect of the duration of the preceding/following vowel]." *Journal of the Phonetic Society of Japan* 9(2), 59–65.

Payne, E. M. (2006) "Non-durational indices in Italian geminate consonants." *Journal of the International Phonetic Association* 36(1), 83–95.

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Shirose, A., T. Kitamura and Y. Nota (2021) "Observation of Japanese short and long vowel productions by EMA." *Proceedings of the Technical Committee on Speech Communication* 1(2), 69–74.

Spreafico, L., M. Pucher and A. Matosova (2018) "Ultrafit: A speaker-friendly headset for ultrasound recordings in speech science." *Interspeech 2018*, 1517–1520.

Takada, S. (1985) "Sokuon no chōonjō no tokuchō ni tsuite [On some articulatory charactersitsics of the mora obstruent (sokuon)]." *NINJAL Occasional Papers* 6, 17–40.

Takeyasu, H. (2012) "Sokuon no chikaku ni taisuru senkō onsetsu shiin boin no jizokujikan no eikyō [Effects of the consonant and vowel durations in the preceding syllable on the perception of geminate stops in Japanese]." *Onin Kenkyu [Phonological Studies]* 15, 67–78.

Tiede, M. K. (2021) GetContours. Version 3.5. GitHub repository, https://github.com/mktiede/GetContours (accessed May 1, 2021)

Wood, S. N. (2006) *Generalized Additive Models: An Introduction with R.* New York: Chapman and Hall/CRC.

Yamane, N. (2013) *'Placeless' Consonants in Japanese: An Ultrasound Investigation*. Doctoral dissertation, University of British Columbia.

Yazawa, K. and M. Kondo (2019) "Acoustic characteristics of Japanese short and long vowels: Formant displacement effect revisited." *Proceedings of the 19th International Congress of Phonetic Sciences*, 671–675.

Yazawa, K., T. Konishi, J. Whang, P. Escudero and M. Kondo (2023) "Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study." *Laboratory Phonology* 14(1), 1–33.

(Received May 9, 2024, Accepted Mar. 5, 2025,
e-Published Apr. 30, 2025)