To appear in JASA-EL

Articulatory correlates of consonantal length contrasts: the case of Japanese mimetic geminates

Francesco Burroni,[1,a] Shigeto Kawahara,[2] and Jason Shaw[3]

[1] *Institute for Phonetics and Speech Processing, Ludwig Maximilian University, Munich, Bavaria 80799, Germany*

[2] *The Institute of Cultural and Linguistic Studies, Keio University, Minato-ku, Tokyo 108-8345, Japan*

[3] *Department of Linguistics, Yale University, New Haven, Connecticut 06520, USA*

francesco.burroni@phonetik.uni-muenchen.de,

kawahara.research@gmail.com,

jason.shaw@yale.edu

This study investigates the articulatory correlates of consonantal length contrasts in Japanese mimetic words using electromagnetic articulography data. Regression and Dynamic Time Warping analyses applied to intragestural timing, kinematic properties, and intergestural timing reveal that Japanese geminates are characterized by longer closure phases, longer gestural plateaux, higher tongue tip positions, larger movements, and lower stiffness. Geminates also exhibit distinct timing relationships with adjacent vowels, specifically, longer times to target that allow for longer preceding vowels. Our findings shed light on the articulatory mechanisms underlying Japanese geminate production, their relationship to acoustics, and their characterization in a broader cross-linguistic perspective.

---

[a] Author to whom correspondence should be addressed.

1. Introduction

In many of the world's languages, consonant duration can be employed contrastively.[1,2] A well-known example is Japanese. The production of Japanese long consonants, often referred to as "geminates", has been primarily studied through their acoustic manifestations.[3] Acoustically, Japanese geminates are distinguished from singletons mainly by their longer constriction durations.[4] Geminates are also accompanied by slightly longer preceding vowels and by other non-durational cues, for example, larger intensity differences surrounding geminates, larger pitch-accent f0 drops across geminates, vowels with lower F1 after geminates, and more creaky vowels after geminates, *cf.* [3] and references therein.

Unlike their acoustic correlates, the articulatory mechanisms involved in the production of Japanese geminates have not been thoroughly explored.[5–12] To fully understand geminate production in Japanese and contextualize it within a broader linguistic framework, three key areas require further investigation: (i) intragestural timing properties, which encompass the duration and timing characteristics of geminate articulation; (ii) kinematic properties, which are responsible for the shape of the trajectories of articulatory movements during geminate production; (iii) intergestural timing properties, *i.e.*, the timing relationships between geminate articulation and the articulation of surrounding sounds.

With respect to intragestural timing, one important understudied aspect is how the longer closure duration of geminates is implemented kinematically. Two main strategies have been reported based on Japanese kinematic data. The first strategy is the realization of geminates by holding constrictions targets for longer periods of time;

in effect, lengthening the plateau of kinematic trajectories once the target has been achieved.[5] A second strategy is the realization of geminates by slowing down articulator movements, specifically, tongue movements for lingual consonants. Building upon these reports, the current study further explored the intrinsic timing mechanisms of Japanese geminates.

With respect to kinematic features, work on this aspect to our knowledge is limited to a single study of Japanese bilabial geminate plosive,[7] which uncovered (i) higher articulator positions compared to singletons, (ii) larger movement amplitude, (iii) similar peak velocities during closure, and (iv) lower stiffness. These findings point to distinct kinematic specifications for articulator control during geminate production, beyond simply longer durations, *cf.* also [3,13]. More extended articulator contacts, compatible with more constricted targets, have also been reported for lingual consonants using electropalatography.[13,14] However, more investigations of kinematic features in various manners and places of articulation is necessary to fully characterize geminate articulation in Japanese.

Finally, with respect to intergestural timing, pioneering comparative work on Japanese and Italian geminates by Smith [5,6] uncovered a potential kinematic basis for the longer duration of vowels before geminates observed in Japanese. This longer duration may be in part due to intrinsically longer kinematic movements, as also suggested in subsequent work.[9] Crucially, however, Japanese geminates also display longer times to target compared to singletons, which may allow preceding vowels to be acoustically manifested for a longer period of time. This timing feature sets Japanese gemination apart from other languages, for example, from Italian. In Italian, geminates are realized by anticipating closure onset during the preceding vowel, thus, in effect shortening the acoustic manifestation of the vowel.[5,6,15] The current study compared geminates and singleton onsets relative to the preceding vowel in Japanese, thereby explicating the differences between Japanese and those languages that shorten pre-geminate vowels, like Italian.

Smith[5,6] also reported longer trans–consonantal lags between preceding and following vowels across geminates than singleton. This finding suggests a more sequential organization for vowels across a geminate in Japanese than in Italian, in which the trans–consonantal vocalic lag is unaffected or perhaps even shortened, at least in the presence of bilabial geminates.[5,6,15] Smith's [5,6] works offered an intriguing basis for the intergestural timing organization of geminates; however, subsequent work on Japanese geminates did not show clear intergestural timing differences [8,9] and has also called into question the relationship between a longer time to target and duration of the preceding vowel.[11]

With these issues in mind, we investigated the articulatory correlates of geminate production in Japanese using electromagnetic articulography (EMA). We analyzed geminate production along the three main lines discussed above, namely in terms of their intragestural timing, kinematic properties, and intergestural timing. The questions we addressed are: (1) what are the kinematic correlates that underlie the longer duration of Japanese geminates? (2) what kinematic features differentiate the articulatory trajectories of singletons and geminates? (3) How are geminates timed with respect to surrounding segments, and, more specifically, is there a kinematic basis for the longer duration of preceding vowels?

Given that both acoustic and articulatory properties can be influenced by a host of lexical factors [16–19] that are hard to control, we studied the properties of

Japanese geminates by comparing them to singletons with the same lexical item. This is possible because Japanese mimetic (*i.e.*, sound symbolic) words can be produced with gemination of a medial consonant to express emphasis.[20] Japanese mimetic words thus offer a controlled testing ground for geminate articulation in a situation that is largely independent of lexical influences.

2. Methods

2.1 Participants

Seven native speakers (3 F, 4 M) of Tokyo Japanese took part in the experiment. They did not disclose any speech or hearing impairments. All of them reported speaking the Tokyo variety on a daily basis.

2.2 Materials

The speech material consisted of twenty-one existing Japanese mimetic words. The stimuli included three items for each of seven consonantal types, Table 1. Each target consonant was produced as either a singleton or a geminate (for emphasis).[20]

Table 1. Experimental Stimuli

Target	**Item 1**	**Item 2**	**Item 3**
[t]	kɑt(ː)ɑ–kɑtɑ	gɑt(ː)ɑ–gɑtɑ	pet(ː)ɑ–petɑ
[d]	kud(ː)o–kudo	gud(ː)a–guda	od(ː)o–odo
[ɾ]	paɾ(ː)a–paɾa	peɾ(ː)a–peɾa	doɾ(ː)o–doɾo
[z]–[ɟ]	giz(ː)a–giza	oz(ː)u–ozu	uɟ(ː)i–uɟi
[s]	kos(ː)o–koso	kas(ː)a–kasa	ɸus(ː)a–ɸusa

[ts]	ɸut͡(ː)su– ɸut͡su	kat͡(ː)su–kat͡su	gut͡(ː)su–gut͡su
[t͡ɕ]	net͡(ː)ɕi–net͡ɕi	kat͡(ː)ɕa–kat͡ɕa	gut͡(ː)ɕa–gut͡ɕa

Given that previous work has focused on bilabial consonants[5–8,12] and our interest also lies in the interaction between geminates and preceding vowels, we focused on consonants produced with the same articulator as vowels, the tongue. Specifically, we examined apical/laminal consonants, as they are produced with the front of the tongue, which is straightforward to track with EMA.

In each trial, participants produced singleton mimetic words in isolation, followed by their emphatic geminate variant also produced in isolation. Items were elicited in isolation to minimize total experiment duration and confounds that could arise from different prosodic realizations of a carrier phrase. The target consonants were never adjacent to word boundaries, thus also limiting the need for a carrier phrase. The order of lexical items was fully randomized. In total, 21 unique words × 2 realizations (singleton/geminate) × 10 repetitions × 7 speakers yielded a total of 2940 tokens for analysis.

2.3 Experimental Procedure

We used an NDI Wave system sampling at 100 Hz (3 speakers, S1-S3) and a Carstens AG501 system sampling at 1250 Hz (4 speakers, S4-S7) to capture articulatory movements. The two datasets were first analyzed separately. Since no substantial differences were found, we present combined analyses of the two datasets. Our supplementary material contains scripts demonstrating our analyses for the datasets, both separately and combined, for the interested readers. The sensors' setup was virtually identical for both systems. Five sensors were adhered to tongue locations using high viscosity dental glue and dental cement (KETAC™): three on the tongue sagittal midline and two parasagittally. Only the most anterior sensor on the tongue, attached less than one cm from the tongue apex, entered into the analyses reported here. A figure showing representative lingual sensor placement is included in the supplementary materials, Figure 5. Three additional sensors were placed on the vermilion borders of the upper and lower lips and on the jaw, below the lower incisor, and three references sensors, used for head movement correction, were placed on the nasion and the left/right mastoid processes. For the AG501 sessions, an additional reference sensor was placed on the maxilla, above the left upper incisor.

Acoustic data were simultaneously recorded at 22 kHz with a Schoeps MK 41S supercardioid microphone connected to a Schoeps CMC 6 Ug power module (for the WAVE sessions) and at 25.6 kHz using a Sennheiser ME66 connected to an NI data acquisition device (for the AG501 sessions).

For the experimental sessions, participants sat in a sound-attenuated room. Words were displayed in Japanese orthography on a monitor positioned ~25cm outside of the EMA magnetic field. Stimulus display was controlled manually using an Eprime® script (for the NDI Wave sessions) and using MATLAB ([21] for the Carstens sessions). This procedure allowed for online monitoring of hesitations, mispronunciations, and disfluencies. These occurred very rarely, but when they did, items were marked for repeat presentation by the experimenter. These items were then re-inserted into the random presentation of remaining items, which ensured that we recorded ten fluent tokens of each target item.

Following the main recording session, we also recorded the bite plane of each participant by having them hold a rigid object, with three (for the Wave sessions) and two (for the AG501 sessions) 5DoF sensors attached to it, between their teeth. Head movements were corrected computationally after data collection with reference to three sensors on the head and the three sensors on the bite plane. The head corrected data was rotated so that the origin of the spatial coordinates corresponds to the occlusal plane at the front teeth.

2.4 Data processing and analyses

The acoustic signals were manually segmented at the word and segmental level in Praat[22] using standard criteria based on waveform and the spectrogram characteristics by two research assistants naïve to the purpose of the study, see Figure 4. The segmentation was later double-checked by the first author. The acoustic segmentation was primarily used as a starting point for articulatory landmarking of target singleton and geminate consonants. The only acoustic boundaries used in the analyses involve the acoustic duration of the target consonants and preceding and following vowels. Given the wide variety of vocalic contexts, and the impossibility to consistently identify vocalic landmarks from the kinematics alone, we adopt acoustic landmarking as events that reasonably correlate with kinematic events in gestural unfolding, especially target achievement, following previous work, e.g.,[23] .

The articulatory analyses reported in this paper focus on apical/laminal consonants; thus, the main sensor of interest is the tongue tip, henceforth TT. The wide variety of vocalic contexts in which target consonants appears has a strong influence on horizontal tongue position, thus, our landmarking was based on the dimension that we found to be most reliable to identify consonantal articulation, namely the TT vertical movement (henceforth TTy); for a similar choice for apical/laminal consonants in Japanese see [24]. Support in favor of the decision of using TTy as the main landmarking dimension also comes from a Principal Component Analysis (PCA). We entered all three dimensions of TT movement in a PCA analysis, and found that the direction of maximum movement, the 1st PC (explaining 95.7% of variance) correlates almost perfectly with vertical movement (median r = .97 over all tokens). These results suggest that tongue movement during target consonants production takes place robustly in TTy and that this dimension is reliable to track consonantal kinematic movements.

All articulatory signals were smoothed and interpolated using a MATLAB implementation of the algorithm in [25]. For landmarking, TTy trajectories were further upsampled to 1000 Hz for the Wave sessions and kept at the original sampling rate of 1250 Hz for the AG501 sessions. All trajectories were smoothed using a Savitzky–Golay finite impulse response filter of polynomial order 4 and a frame length of 21 samples for landmarking. Consonantal gestures were automatically landmarked based on the TTy trajectory using a custom MATLAB routine based on peak velocity thresholding. We first use the acoustic boundaries to locate the midpoint of each consonant in the acoustic signal. The timestamp of the midpoint was then used to define a symmetric window spanning the acoustic duration of the segment to the left and the right of the midpoint. Within this window, we located velocity extrema of appropriate sign for the upward movement denoting closure and the downward movement denoting release. We then defined the gestural onset as the first timepoint surpassing 20% of peak velocity during closure and the gestural target as the last timepoint surpassing 20% of peak velocity during closure. The gesture release and offset were similarly defined as the first timepoint surpassing 20% and the last timepoint surpassing 20% of peak velocity during release. Landmarks were visually inspected and retracked in a small number of cases (~50 tokens) where it was found to be necessary. Additionally, 30 tokens could not be landmarked as no clear gestural boundaries could be detected in the TTy trajectory. Accordingly, these 30 tokens were excluded from the final analysis together with their associated singleton or geminate. As a result, a total of 60 tokens were excluded (~2% of the originally collected data). Excluding the paired token was necessary because some of our analyses require pairing singleton and geminate obtained from the same trial, as described below.

From the acoustic and articulatory signals, we derived a set of intragestural timing, kinematic, and intergestural timing measures. For intragestural timing, Figure 1, we investigated:

(i) the closure phase duration, defined as the lag between gestural onset and target

(ii) the plateau duration, defined as the duration of the lag between gestural target and release

(iii) the release phase duration, defined as the lag between gestural release and offset

We also investigated the relationship between plateau duration and acoustic consonant duration using the Pearson correlation coefficient.

Additionally, we also developed more holistic analyses of the kinematic trajectories that rely on Dynamic Time Warping (DTW). Specifically, we took each singleton/geminate and used DTW to derive a pairwise warping function that identifies which portions of a singleton articulatory trajectory need to be stretched to "derive" a geminate in each trial. Specifically, we inspected the shape of the warping functions and obtained localized average warps over normalized duration, for a similar approach *cf.* [26].

For kinematic properties, Figure 2, we investigated:

(iv) maximum tongue position, as a proxy for constriction target, in both the vertical and horizontal dimensions

(v) peak velocity during closure

(vi) movement amplitude from onset to maximum constriction during closure

(vii) kinematic stiffness, defined as the ratio of peak velocity and movement amplitude (e.g.,[27]). This is an empirical measure of movement speed. Peak velocity is normalized for movement amplitude because of the well-known observation that these two measures are positively related. Amplitude normalization gives us a measure of speed that is more independent of amplitude. The higher the stiffness of an articulatory gesture, the shorter its time to peak velocity and target attainment.

For intergestural timing properties, Figure 4, we investigated:

(viii) the duration of the vowel preceding the target consonants estimated from acoustics (V_1 Dur.)

(ix) the duration of the lag between the preceding vowel acoustic onset and the consonantal gesture onset (V_1 Ons. – C Ons.)

(x) the duration of the lag between the preceding vowel acoustic onset and the consonantal gesture target (V_1 Ons. – C Targ.)

(xi) the duration of the trans–consonantal lag between the preceding vowel acoustic onset and following vowel acoustic onset (V_1 Ons – V_2 Ons.)

We also investigated the relationship between consonantal and preceding vowel duration using the Pearson correlation coefficient, following [11].

All the properties (i)–(xi) were analyzed by fitting linear mixed effect regression models, using the MATLAB *fitlme*() function. The effect of gemination was analyzed by comparing a baseline model without the singleton vs. geminate difference to an alternative model with a fixed effect for geminate *via* loglikelihood ratio tests, using the MATLAB *compare*() function. Both models had identical random effect structures with by-subject and by-item (collapsed for gemination) random intercepts and a random slope for gemination. The alternative model had, in addition, a fixed effect for gemination (with reference coded as singleton), *i.e.*, DV ~ 1 + (gemination|subject) + (gemination |item).

With respect to data exclusion procedures, we identified as outliers all values whose z-score was below -4 or greater or above 4 and excluded them from analyses. In this way, between 0% and 1.46% of the data was excluded for different variables. Finally, given the well-known sensitivity of Pearson correlation to outliers, we excluded values below the 1% percentile and above the 99% percentile in our correlation analyses, as a small number of extreme datapoints can artificially inflate or deflate correlations.

All data and scripts necessary to replicate our analyses and figures and to run separate analyses on the NDI wave and AG501 datasets are publicly available in an OSF repository at https://osf.io/27nyz/?view_only=4f93c383e24642e48d027c58fd945a27 .

3. Results

3.1 Intragestural timing

For intragestural timing, we investigated: (i) the closure phase duration; (ii) the plateau duration; and (iii) the release phase duration, Figure 1 top left.

We found a not very large but yet statistically reliable difference in closure phase duration ($\chi^2_{(1)} = 8.73$, $p = 0.003$). Specifically, the singleton closure phase is 86 ms (95% CI [73–100] ms), while the geminate closure phase is +11 ms longer (95% CI [6–17] ms), Figure 1 top right.

A much more robust difference was observed for plateau duration ($\chi^2_{(1)} = 12.61$, $p = 0.0004$). Singleton plateau duration was 41 ms (95% CI [26–56] ms), while geminate plateau duration was substantially longer, with an effect estimated at +108 ms (95% CI [72–144] ms), Figure 1 bottom left.

Finally, no reliable differences were found in terms of the release phase duration between singletons and geminates ($\chi^2_{(1)} = 0.04$, $p = 0.84$), Figure 1 bottom right.

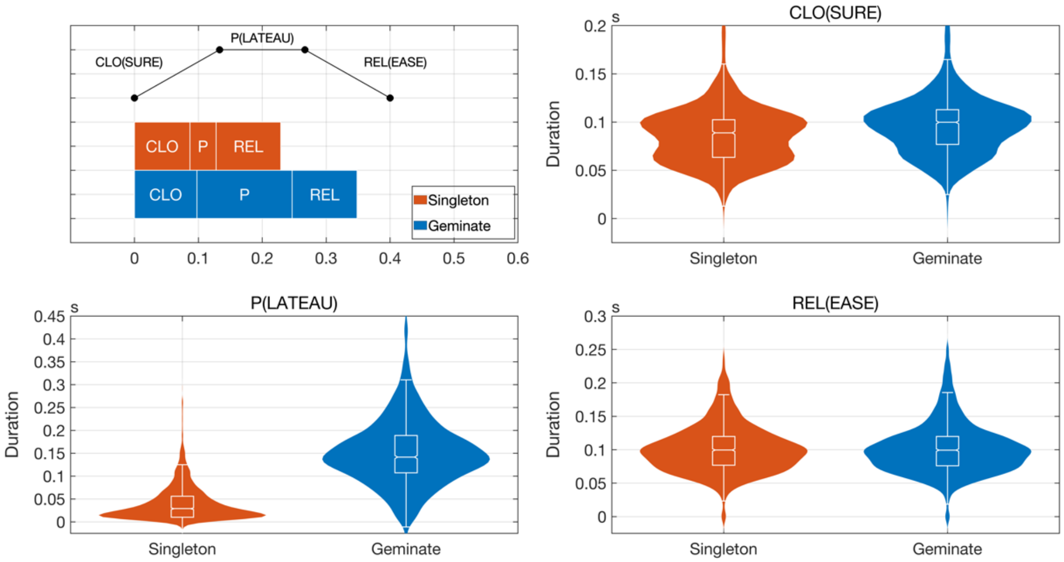

Figure 1 **Top left**: schematic illustration of intragestural phases and their mean durations. **Top right**: Boxplot with superimposed gaussian kernel density estimate (kde) of CLO duration. **Bottom left**: Boxplot of P duration. **Bottom right**: Boxplot of REL duration.

The fact that singletons and geminates differ primarily in terms of their plateau duration is also evident from DTW analyses. When each mimetic singleton is stretched to its geminate counterpart, a strong distortion of time is observed around the midpoint of the consonant, 0.2 to 0.6 of its proportional duration, Figure 2F. Each singleton sample between 33% and 66% of the singleton trajectory is repeated between two and almost 3.5 times to derive a geminate, indicating a stretching of the plateau region, Figure 2F.

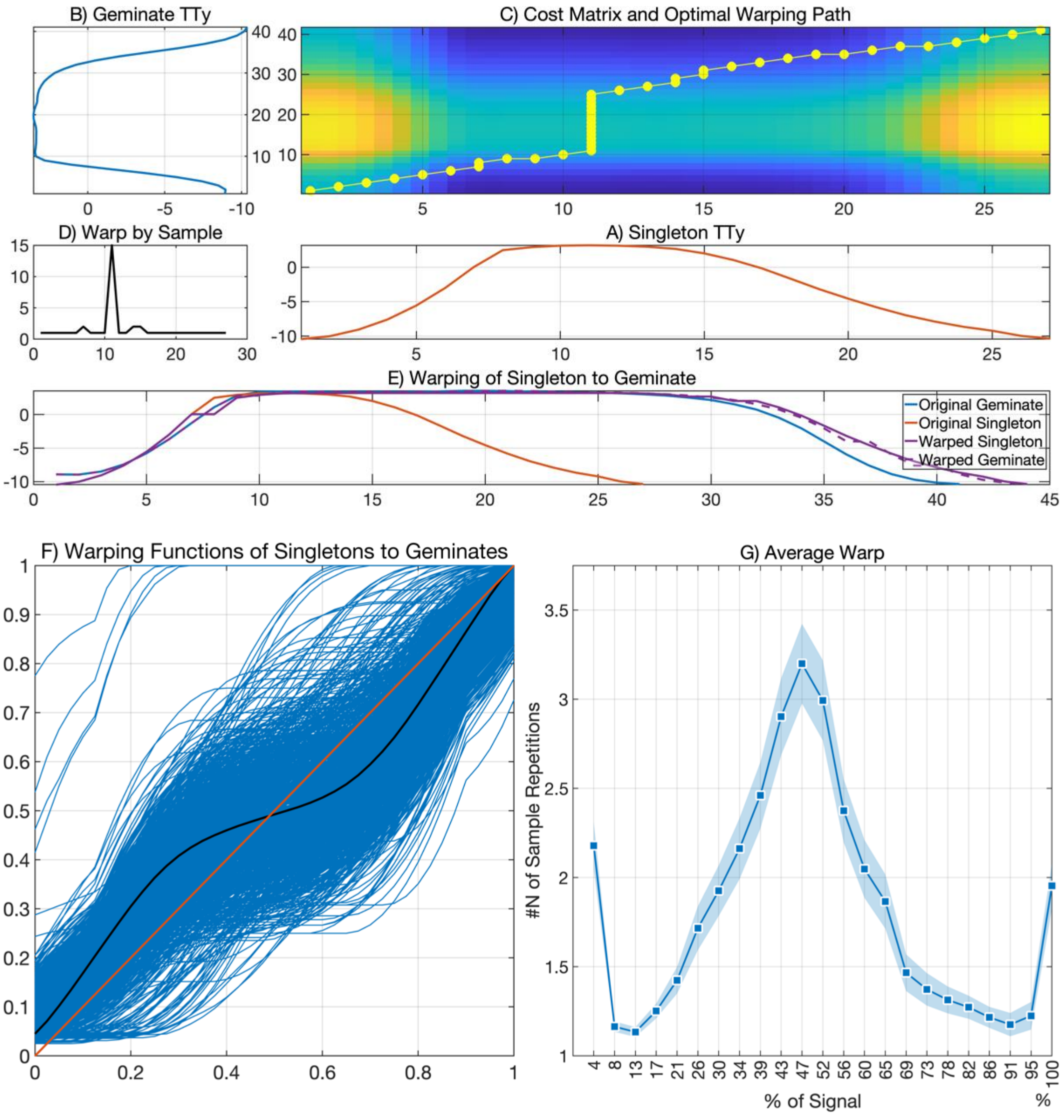

Figure 2 **A**: Example of TT vertical movement for singleton. **B**: Example of TT vertical movement for geminate. **C**: Cost matrix and optimal warping path for singleton to geminate alignment. **D**: Warp function showing the number of repetitions each sample undergoes to stretch singleton to geminate. **E**: Alignment of singleton and geminate TT vertical movement trajectories. **F**: Warping function of singleton to geminates (blue lines), with average warping functions (solid black line) of singleton to geminate showing a strong distortion of linear time (orange line) in the plateau region (0.2-0.6). Time is normalized between 0 and 1. **G**: Average warp at different % of the trajectory.

The observed longer acoustic duration for geminates is closely related to the longer plateau, as we observed a robust correlation between target consonant acoustic duration and plateau duration ($r = 0.84$, $p < 1e\text{-}215$), see Supplementary Material Figure 6.

3.2 Kinematic properties

For kinematic properties, we investigated (i) maximum tongue height*;* (ii) peak velocity during closure*;* (iii) movement amplitude*;* and (iv) stiffness, Figure 3 top panel.

We found higher TTy position for geminates than singletons ($\chi^2_{(1)} = 8.76$, $p = 0.003$). Specifically, compared to singletons (6.6 mm, 95% CI [1.6–11.6]), geminates are produced with higher or more constricted targets +0.80 mm (95% CI [0.39–1.19] mm), Figure 3 mid left. In the horizontal position, the difference was not significant ($\chi^2_{(1)} = 3.25$, $p = 0.07$), however there is a tendency for geminates to be more fronted +0.23 mm (95% CI [0–0.47] mm), the lack of significance is possibly due to high variability in horizontal position due to different vowel combinations.

The difference in movement amplitude was significant ($\chi^2_{(1)} = 8.5$, $p=0.003$). Compared to singletons (8.16 mm, 95% CI [6.7–9.7]), geminates are produced with greater closure movement amplitude +0.99 mm (95% CI [0.45–1.53] mm), Figure 3 mid right.

We found no significant differences in peak velocity during closure $\chi^2_{(1)} = 1.45$, $p = 0.22$), Figure 3 bottom left.

Finally, we found a significantly lower stiffness for geminates ($\chi^2_{(1)} = 5.13$, $p = 0.02$). Compared to singletons (19.92 s^{-1}, 95% CI [16.17–23.67 s^{-1}]), geminates are produced with lower stiffness −2.61 s^{-1} (95% CI [−4.54 – −0.69] s^{-1}), indicating a slower movement and time to target, Figure 3 bottom right.

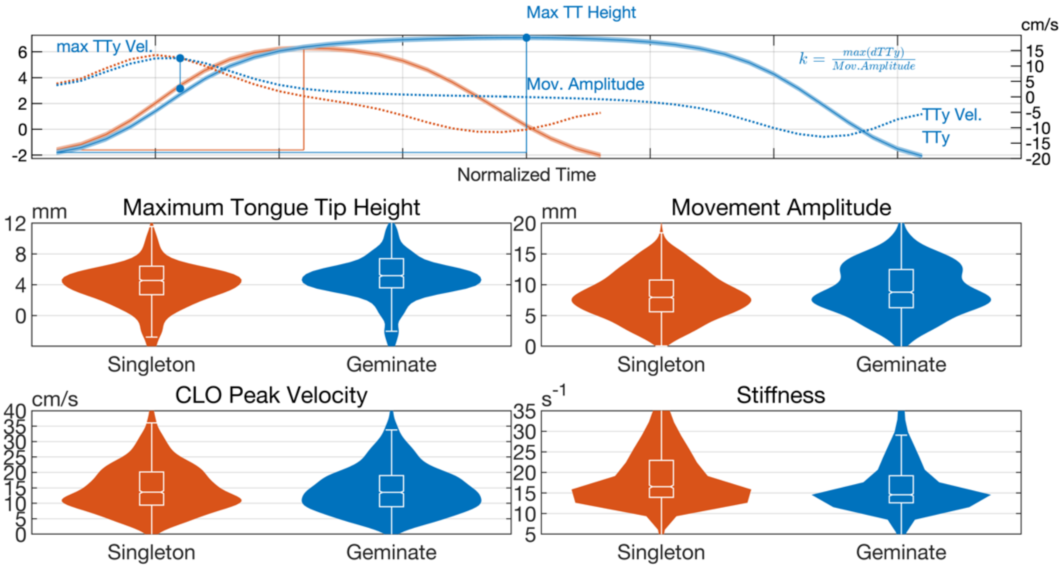

Figure 3 **Top panel**: averaged vertical (y) tongue tip (TT) movement trajectories for singleton and geminate production (solid trajectories), together with their first derivatives over time representing vertical velocity (dashed trajectories). Circular marker on dashed trajectories mark the time of maximum vertical velocity, circular marker on solid trajectories marks the maximum tongue tip TTy value. Vertical line from the maximum TTy value to the value at the trajectory onset marks movement amplitude. Kinematic stiffness (k) calculation is included as equation based on the depicted values. **Mid left**: Boxplot with superimposed gaussian kde of maximum tongue tip height values. **Mid right**: Boxplot of movement amplitude values. **Bottom left**: Boxplot CLO peak velocity values. **Bottom right**: Boxplot of stiffness values.

3.3 Intergestural timing properties

For intergestural timing properties, we investigated (i) preceding vowel duration (V1 Dur.), (ii) the preceding vowel and consonantal gesture onset lag duration (V1 Ons. – C Ons.), (iii) the preceding vowel onset and the consonantal gesture target lag duration (V1 Ons. – C Targ.), (iv) the duration of the trans–consonantal vowel to vowel onset lag (V1 Ons – V2 Ons.), Figure 4 top panel.

We found that vowels preceding geminates had longer duration than vowels preceding singletons ($\chi^2_{(1)}$ = 7.57, p = 0.006). Specifically, compared to vowels preceding singletons (70 ms, 95% CI [60–79]), vowels preceding geminates were +13 ms longer (95% CI [6–20] ms), Figure 4 mid left panel.

The duration of the lag between vowel acoustic onset and consonantal gesture onset (V1 Ons. – C Ons.) was significantly different, but very similar between geminates and singletons ($\chi^2_{(1)}$ = 4.12, p=0.04), Figure 4 mid right panel. Compared to the lag of vowels preceding singletons (-4 ms, 95% CI [-1–7]), the lag of vowels preceding geminates is slightly longer by +7 ms (95% CI [1–13] ms). Note, however, that the 95% CI are very close to an overlap with 0. Moreover, this finding only emerges by pooling our datasets together, suggesting that, if consistently present, the difference between singleton and geminates in V1 Ons. – C Ons. is very small and quite variable. Future work should further investigate the robustness of this finding.

On the other hand, the duration of the lag between vowel acoustic onset and consonantal gesture target (V1 Ons. – C Targ.) was longer in geminates than in singleton production ($\chi^2_{(1)}$ = 12.6, p=0.0004). Specifically, compared to the lag in singleton production (83 ms, 95% CI [69–96] ms), the V1 Ons. – C Targ lag in geminate production was +20 ms longer (95% CI [13–27] ms), Figure 4 bottom left panel.

Finally, we found that the duration of trans–consonantal lag between the preceding vowel acoustic onset and following vowel acoustic onset (V1 Ons – V2 Ons.) was much longer in geminate than in singleton production ($\chi^2_{(1)} = 19.43$, $p < 0.0001$). Specifically, compared to the lag in singleton production (218 ms, 95% CI [158–279] ms), the V1 Ons. – V2 Ons. lag in geminate production was +134 ms longer (95% CI [108–160] ms), Figure 4 bottom right panel.

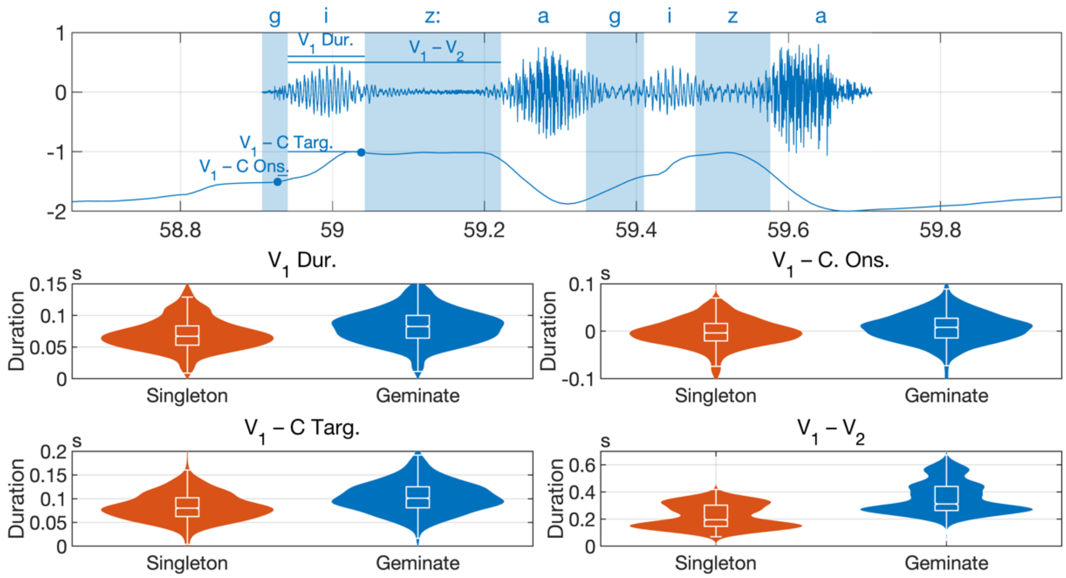

Figure 4 **Top panel:** Illustration of waveform and TTy trajectories with shaded areas marking segmental boundaries for production of [giz:agiza]. Circular markers on TTy trajectory mark onset and target of tongue tip gesture for [z]. Labeled blue segments mark lag durations: V_1 Duration, $V_1 – V_2$, V_1 – C Ons., V_1 – C Targ. **Mid left**: Boxplot with superimposed gaussian kde of V_1 Duration values. **Mid right**: Boxplot of V_1 – C Ons. values. **Bottom left**: Boxplot of V_1 – C Targ. values. **Bottom right**: Boxplot of $V_1 – V_2$ values.

Given the findings presented above, we hypothesized that the longer duration of vowels preceding geminates may be due to a longer time to target compared to singletons, effectively allowing for a longer steady-state period for vowel production. Evidence in favor of this hypothesis is offered by a strong correlation ($r = .70$, $p < 1e-215$) observed between time to target and preceding vowel duration, see Supplementary Material Figure 7. In other words, the longer the time to target, the longer the acoustic duration of the preceding vowel.

4. Discussion

Returning to our research questions, our analyses revealed robust differences between Japanese geminates and their singleton counterparts in terms of intragestural timing, kinematic properties, and intergestural timing organization.

With respect to intragestural timing, geminates were produced with slightly longer closure phases and much longer gestural plateaux, a finding also confirmed by DTW analyses. No differences were observed for release phase duration, in line with the previously-reported acoustic analysis, which reported no difference in VOT.[13] We also found an almost perfect correlation between consonantal acoustic duration and tongue tip vertical plateau duration. Taken together, our findings are in line with previous work,[5,9] indicating that Japanese speakers produce geminates with slightly longer closures movements and much longer gestural plateaux. Findings of longer closures movements and plateaux are not unique to Japanese, as these characteristics have also been reported for geminate production in other languages, like Italian.[15]

Turning to kinematic features, we found that Japanese speakers produce (lingual) geminates with a higher tongue tip, *i.e.*, with a more constricted posture, slightly larger movements, similar peak velocity, and lower stiffness, in line with previous reports based on lexical bilabial geminates in Japanese.[7] In other languages, like Italian, however, higher peak velocity is also observed.[15] Taken together, the differences in kinematic features suggest that Japanese geminates are not just longer versions of singleton consonants, as their articulation is characterized by a different set of kinematic parameters. In this respect, our findings support the view that geminates differ from singletons not only in terms of durational properties but also in terms of more general articulatory strategies, an idea that has been proposed for Tashlhiyt Berber and Italian [2,15] and that also been entertained for Japanese[3,13] and other Japonic languages.[28]

Finally, in terms of intragestural timing, we found that Japanese speakers produced slightly longer vowels before geminate consonants. Geminates also start roughly at the same time as singletons with respect to the preceding vowel. However, geminate targets are reached later compared to singletons. Longer trans–consonantal lags between preceding and following vowels are also observed across geminates than across singletons. These findings suggest that Japanese speakers produce geminates and singletons with relatively similar timing organization with respect to the preceding vowel. This has been noted in previous work [5,6] and is a feature that sets apart Japanese geminate production from languages like Italian, where geminate closure robustly "intrudes" in the preceding vowel resulting in shorter pre–geminate vowel duration.[5,6,15] With respect to intergestural timing, our work is compatible with previous findings by Smith[5,6] in uncovering kinematic correlates that can offer a potential basis for the observed acoustic patterns that have been reported. Specifically, unlike previous work that found only limited evidence for a correlation between consonant time to target and preceding vowel duration,[11] our data suggests a robust positive correlation between the duration of the vowel onset to consonantal target lag and (preceding) vowel duration. This pattern provides a possible kinematic basis for the longer vowel durations preceding Japanese geminates. A longer time to target allows the vocalic gesture to be (acoustically) longer before the acoustic consequences of consonantal articulation “kick in” during the consonantal plateau, as we have demonstrated when discussing the relationship between consonantal acoustic duration and plateau duration. The longer time to target also reinforces other potential bases for longer vocalic acoustic duration. Namely, the slower and longer tongue body movements associated with vowels during geminate production [5,6,9] can contribute to longer acoustic vowel durations in the presence of delayed consonantal target achievements for geminates and their associated acoustic consequences. Such slower tongue body movements are also observed in our data. This is illustrated in the supplementary material, Figure 8, where we present DTW analyses to show that there is a generalized need to warp time. Specifically, to obtain the combined vertical and horizontal tongue movement observed during pre- and post-geminate vowel production from the tongue movement observed during pre- and post-singleton vowel production time needs to be slowed down, especially while the consonant is being produced.

5. Conclusion

To conclude, our analyses have revealed that geminates, as produced in Japanese mimetic words, exhibit longer closure phases, extended gestural plateaux, higher tongue tip positions, and more constricted postures. These articulatory profiles accord well with the acoustic properties of Japanese geminates, like their longer closure durations and lengthened preceding vowels.

Additionally, our analyses also situated Japanese geminate production in a wider crosslinguistic context. Japanese geminates seem to be primarily produced by lengthening gestural plateaux compared to singletons, as demonstrated by our DTW analyses. However, even so, they are not simply extended versions of singletons: some of their kinematic parameters are also different. These considerations lend plausibility to the proposal that even "canonical" geminates like those of Japanese and Italian are actually implemented by speakers using dimensions beyond duration, such as tighter constrictions and generally different kinematic profiles that have larger movements and lower stiffness. Additionally, our analysis lends further support to the idea that languages can differ substantially in the timing of geminates.[5] Japanese geminates and singletons start around the same time with respect to the preceding vowels, yet geminates reach their targets later, allowing for longer acoustic vowel durations. This is unlike other languages where geminate production starts earlier with respect to the preceding vowel, in effect, shortening it.[15]

Acknowledgments

We would like to thank Piyapath Srisomyos and Teerawee Sukanchanon for help with data annotation, as well as Jeff Moore, Lia Bučar Shigemori, and Ulrike Rupprecht for help with data acquisition. Data collection was supported by JSPS grants #15F15715 to the second and third authors, #26770147 and #26284059 to the second author.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The authors obtained ethical approval from the IRB boards of Western Sydney University and Keio University (protocol number: HREC 9482) and from the IRB board of the University of Munich. Informed consent was obtained from all participants.

Data Availability

The data that support the findings of this study are openly available in an OSF repository at https://osf.io/27nyz/?view_only=4f93c383e24642e48d027c58fd945a27 .

References and Links

1. Ladefoged, P. & Maddieson, I. *The Sounds of the World's Languages*. vol. 1012 (Blackwell Oxford, 1996).
2. Ridouane, R. Geminates at the junction of phonetics and phonology. *Papers in laboratory phonology* **10**, 61–90 (2010).
3. Kawahara, S. The phonetics of sokuon, or geminate obstruents. *Handbook of Japanese phonetics and phonology* 43–78 (2015).
4. Idemaru, K. & Guion-Anderson, S. Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica* **67**, 25–46 (2010).
5. Smith, C. L. The timing of vowel and consonant gestures. (Yale University, 1992).

6. Smith, C. L. Prosodic patterns in the coordination of vowel and consonant gestures. *Papers in Laboratory Phonology IV, Phonology and phonetic evidence. CUP* 205–222 (1995).

7. Löfqvist, A. Lip kinematics in long and short stop and fricative consonants. *The Journal of the Acoustical Society of America* **117**, 858–878 (2005).

8. Löfqvist, A. Articulatory coordination in long and short consonants: An effect of rhythm class? in *The Phonetics and Phonology of Geminate Consonants* 118–129 (2017).

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

10. Morimoto, M. & Kitamura, T. Articulation of geminated liquids in Japanese. in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* 2811–2815 (2019).

11. Fujimoto, M., Funatsu, S. & Hoole, P. Articulation of single and geminate consonants and its relation to the duration of the preceding vowel in Japanese. in *ICPhS* (2015).

12. Löfqvist, A. Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *The Journal of the Acoustical Society of America* **120**, 2872–2883 (2006).

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

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

15. Burroni, F., Maspong, S., Benker, N., Hoole, P. & Kirby, J. Spatiotemporal Features of Bilabial Geminate and Singleton Consonants in Italian. in *Proceedings of the 13th International Seminar on Speech Production* (2024).

16. Shaw, J. A. & Kawahara, S. Effects of surprisal and entropy on vowel duration in Japanese. *Language and speech* **62**, 80–114 (2019).

17. Tomaschek, F., Tucker, B. V., Wieling, M. & Baayen, R. H. Vowel articulation affected by word frequency. in *10th International Seminar on Speech Production* (2014).

18. Cohen Priva, U. Informativity affects consonant duration and deletion rates. *Laboratory phonology* **6**, 243–278 (2015).

19. Tomaschek, F., Tucker, B. V., Fasiolo, M. & Baayen, R. H. Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard* **4**, 20170018 (2018).

20. Kawahara, S. Emphatic gemination in Japanese mimetic words: A wug-test with auditory stimuli. *Language sciences* **40**, 24–35 (2013).

21. The MathWorks Inc. *MATLAB R2024b*. (2024).

22. Boersma, P. & Weenink, D. Praat: doing phonetics by computer [Computer program]. (2024).

23. Browman, C. P. & Goldstein, L. Some notes on syllable structure in articulatory phonology. *Phonetica* **45**, 140–155 (1988).

24. Shaw, J. A. & Kawahara, S. The lingual articulation of devoiced/u/in Tokyo Japanese. *Journal of Phonetics* **66**, 100–119 (2018).

25. Garcia, D. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis* **54**, 1167–1178 (2010).

26. Burroni, F. & Tilsen, S. The online effect of clash is durational lengthening, not prominence shift: Evidence from Italian. *Journal of Phonetics* **91**, 101124 (2022).

27. Roon, K. D., Hoole, P., Zeroual, C., Du, S. & Gafos, A. I. Stiffness and articulatory overlap in Moroccan Arabic consonant clusters. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* **12**, 1–23 (2021).

28. Burroni, F., Lau-Preechathammarach, R. & Maspong, S. Unifying initial geminates and fortis consonants via laryngeal specification: Three case studies from Dunan, Pattani Malay, and Salentino. in *Proceedings of the Annual Meetings on Phonology* (2021).