ARTICULATION OF GEMINATED LIQUIDS IN JAPANESE

Maho Morimoto[a] and Tatsuya Kitamura[b]

[a]University of California Santa Cruz, USA; [b]Konan University, Japan
mamorimo@ucsc.edu

ABSTRACT

The study explores the articulatory strategies Japanese speakers draw on to produce geminated liquids, focusing on the movements and shapes of tongue tip and tongue blade. Native speakers of standard Japanese ($n = 8$) produced pairs of reduplicative mimetics with and without emphatic gemination on the liquid consonant. Tongue movements were recorded using EMA (electromagnetic articulography). Results suggest variable strategies within and across speakers, including lateral productions.

Keywords: EMA, Japanese, geminate, liquid, tap, lateral

1. INTRODUCTION

While the phonetics of Japanese geminates have been extensively surveyed, the consequences of geminating the liquid phoneme /r/ have not been documented in detail. It has often been claimed that Japanese, a language with contrastive consonant length, lacks geminated liquids. However, they are attested in certain contexts: in interjection phrases such as /a**rr**e, maa/ 'oh dear' [12]; in emphatic reduplicative mimetics such as /ki**rr**akira/ 'very shiny' derived from /kirakira/ 'shiny' [9, 15, 21, 26]; and in loanwords from languages with contrastive liquid geminates [9, 30], as can be seen in the adaptation of the Italian word *tagliatelle* (a kind of pasta) as /tariate**rr**e/. Representing a geminated liquid in Japanese orthography is straightforward, and geminating a liquid consonant comes with a clear and robust durational difference. In the data presented below, a reliable difference in acoustic duration was observed, geminate duration tripling that of singleton: $t(102.92) = -18.076, p < 0.01$.

Yet, the phonetic details of geminated liquids are unclear, and not easy to predict. Given that apico-alveolar tap has been considered to be the prototypical realization of Japanese /r/ [1, 12, 16, 33], one may say that a viable solution is to simply prolong its brief closure, so as to produce [dd] [18]. However, there is an intuition that it is not an appropriate realization of /rr/ [12]. It has also been pointed out that there is an inherent conflict between the momentary nature of a tap and gemination [23], and that it seems impossible to prolong a tap without turning it into a trill [8]. Trill is a socially marked, stereotypically vulgar realization of Japanese /r/ and is unlikely to be a normal candidate for geminated liquid in Japanese. In the meantime, many have highlighted the remarkable range of subphonemic variability with which Japanese /r/ is realized [1, 2, 7, 11, 16, 22, 32]. While intervocalic variants are reported to be some flavor of taps and flaps, laterals are also reported in post-pausal and post-nasal contexts [1, 32]. Hence, gemination of /r/ does not necessarily involve prolonging or repeating a tap. In fact, some researchers have shared their impression that /rr/ is realized as [ll] [11, 12], and results from an EPG study back up the claim with at least two out of five speakers implementing tight constriction in the alveolar region and weak lateral constriction [11].

In order to investigate the phonetic realization of geminated liquids in Japanese, we conducted a production experiment using three-dimentional electromagnetic articulography (EMA). In this paper, we address the following questions: What are the tongue tip and tongue blade activities involved in the realization of geminated liquids in Japanese? Are production patterns consistent within and across speaker and vowel environments? Do we see evidence of lateral productions?

2. METHOD

2.1. Speakers

Eight native speakers of Tokyo Japanese (female = 5) were recruited in Japan (referred to as S1 ∼ S8). Age of the speakers ranged from 19 to 28. Speakers were compensated for their participation. Instructions were provided in written and spoken Japanese, and a consent form written in Japanese was provided. All of the speakers self-reported as having normal speaking ability at the time of the experiment. A language background questionnaire was administered.

2.2. Speech material

The speech material consisted of 39 Japanese reduplicative mimetics in regular form and emphatic form, embedded in a carrier phrase "konnani <mimetic> nanowa hajimeteda (I've never experienced something so <mimetic>)." The carrier phrase allowed the sentences to be natural for either form. All non-emphatic mimetics had the structure of CVCVCVCV. Emphasizing these mimetics involved geminating the onset of the second syllable. A summary of the material is provided in Table 1.[1] While the experiment included various target consonants, only results pertaining to liquid consonants in the environment of a_a, e_e, and o_o (n = 144; /garagara/ 'empty', /deredere/ 'lovestruck', and /dorodoro/ 'muddy') are reported here due to space.

Table 1: Speech material.

Target	Number	Example
T (ch, ts)	5	gatagata ∼ gattagata
D	6	hidahida ∼ hiddahida
R	19	garagara ∼ garragara
N (ny)	4	uneune ∼ unneune
S (sh)	4	kasakasa ∼ kassakasa
Z	1	mazemaze ∼ mazzemaze

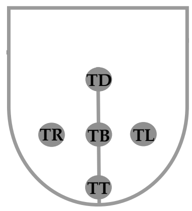
The sentences were presented manually by the experimenter on a screen, in Japanese orthography. Each sentence appeared three times in randomized order. In case of mispronunciations, speakers were asked to read the sentence aloud again.

2.3. Procedure and post-procescsing

Recordings for each speaker were made individually in a sound-proof room[2], using the NDI Wave Speech Research System at 400 Hz. Speech sound was simultaneously recorded through the M-Audio interface with a Sony ECM-77B microphone, at 22.05 kHz. Prior to the recordings, five sensors were placed on the tongue of the speaker in order to track tongue movement: on the sagittal midline, tongue tip (TT; 5 mm from the tip of the tongue), tongue dorsum (TD; as far back as was comfortable for the participant), and tongue blade (TB; mid-point between TT and TD). There were also two lateral sensors, 1 cm to the left (TL) and to the right (TR) of TB. An additional sensor on the gums beneath the lower incisors (LI) was placed to track jaw movement, and a reference sensor on the nasion area (N). The sensor configuration is exemplified in Fig. 1.

Speakers were also asked to hold a rigid plate with sensors attached between their teeth in order to identify the occlusal plane, and palate shape was traced on a palate impression (data not reported here due to space) [14]. Speakers were seated comfortably on a chair, and the field generator was placed so that the articulator is included in a cube of 30 cm^2 from the device. After the sensors were affixed, speakers were asked to read the experiment instruction aloud to get used to articulating with the sensors. The first six trials were practice trials. Speakers could take a break anytime during the experiment.

Figure 1: Tongue sensor configuration



The articulatory data was head-corrected by rotating and transposing the data based on the reference position. Extreme outliers in the articulatory signals were removed and filled using linear interpolation. Garcias's robust smoothing algorithm was applied to all articulatory signals [6, 27]. We used the Mview package [31] for visualizing trajectories and calculating articulatory landmarks [5]. The acoustic signals were annotated using Praat [3]. The acoustic onset and offset of the consonantal constriction were identified in the waveform and spectrogram display, based on the periodic cycles of the adjacent vowels. Data analyses were implemented using R [25].

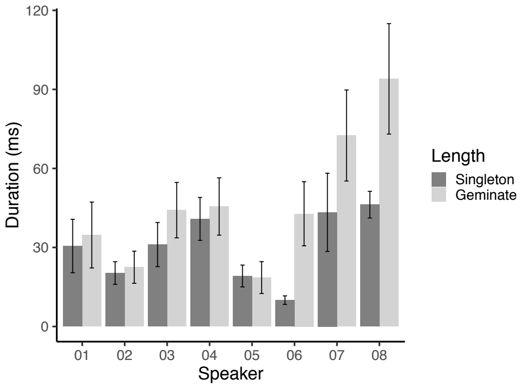
3. RESULTS AND DISCUSSION

3.1. Gestural duration

Overall, no multiple lingual occlusions (i.e. trills) were observed based on spectrographic and auditory information. The duration of the gestural plateau of /r/ formed with TT (determined at 20 % of peak velocity) was longer for geminates than singletons overall: $t(111.58) = -2.86$, $p < 0.01$. There was no significant difference in the overall plateau duration of TB: $t(142) = -0.014$, $p = 0.99$. In the meantime, TT plateau duration varied considerably among speakers, as shown in Fig. 2.

While the acoustically determined constriction duration was at least 2.4 times longer in geminates than in singletons for all speakers, not all speakers had longer TT gestural plateaus for geminated /r/. In fact, TT plateau seems to be only responsible for the acoustic difference for S6, S8, and tendentially for S7. The lack of difference in S5 can be attributed to S5's propensity to insert an audible glottal stop or laryngeal constriction at the beginning of a geminated liquid. As for S1-S4, the acoustic contrast cannot be accounted for solely by TT plateau duration due to intra-speaker variability, some relying on prolonged TB constriction. Results suggest that prolonged TT constriction is not the only way to achieve the acoustic contrast needed for distinguishing length.

Figure 2: Mean TT Plateau duration (ms) per speaker (collapsing across three vowel environments; error bars indicate standard error).



3.2. Movement trajectories

Tables 2 and 3 show the movement directions of TT relative to the positional extrema of the surrounding vowels. Table 2 summarizes the TT movement patterns based on the vertical position. It shows that most of the liquids start with TT in low position, which moves up and then moves back down (*Returning-down*). This pattern is common regardless of the vowel environment or consonant length for most speakers. For singletons, a few mid-vowel productions involved the opposite pattern, starting in high position, moving down and then moving back up (*Returning-up*). S7 was mostly responsible for this auditorily glide-like production. The *Upward* pattern, used by S6, suggests that TT starts in a lower position, hits the alveolar ridge and keeps rising so as to produce a flap [4].

Table 3 summarizes the TT movement patterns based on the horizontal position, along the length of the tongue. NA indicates that signals were too noisy for classification. For singletons, the popular pattern for mid-vowel environments is for TT to start in the front, achieve a target at a more posterior position, and then move back to the front (*Returning-front*). For low-vowel environments, however, the popular patterns are either for TT to start in the back, move to the anterior position, and return to the back (*Returning-back*), or to start in back and move forward (*Forward*). For this, we suspect that the onset of the first and third syllables (/d/ for mid vowel environments and /g/ for low vowel) affected the movement pattern to some degree. It is also compatible with a retroflex tap realization previously reported to be a common option when the surrounding vowels are identical [16, 32].

Table 2: Frequency by vertical TT movement.

Length	Vowel	Returning-down	Returning-up	Upward	Downward
Singleton	a_a	25	0	0	0
	e_e	21	3	0	0
	o_o	21	2	1	0
	total	67	5	1	0
Geminate	a_a	25	0	0	0
	e_e	22	0	0	0
	o_o	24	0	0	0
	total	71	0	0	0

Table 3: Frequency by horizontal TT movement.

Length	Vowel	Returning-front	Returning-back	Forward	Backward	NA
Singleton	a_a	0	12	8	0	4
	e_e	22	0	0	2	0
	o_o	24	0	0	0	0
	total	46	12	8	2	4
Geminate	a_a	1	14	6	1	3
	e_e	15	6	0	1	0
	o_o	11	11	1	1	0
	total	27	31	7	3	3

For geminates, the dominant patterns for low-vowel environments remain *Returning-back* and *Forward*. Unlike for singletons, multiple speakers produced *Returning-back* in e_e, and all speakers except S6 produced *Returning-back* in o_o. The increased number of *Returning-back* productions, especially for o_o, may be attributed to (i) longer durations of the consonant and preceding vowel [10] which allow the tongue tip to move back after the alveolar onset, or (ii) tongue body retraction often associated with liquid consonants cross-linguistically [1, 24, 29].

3.3. Para-sagittal curvature

We adapt a lateralization index from Ying et al. [34], taking the difference of the vertical position of TB and the two lateral sensors, TL and TR, at the point of gestural maxima (minimum velocity point) of TB. A positive value indicates that the side of the tongue is lower than TB; a negative value indicates that the side of the tongue is higher than TB. Fig. 3 and Fig. 4 show the shaping of coronal plane per speaker.

Table 4 summarizes the coronal plane shape per speaker, based on the combination of lateralization indices. We identified four patterns: *Concave* (index is negative for TL and TR), *Convex* (index is positive for both), *Right-lowering* (index is negative for TL and positive for TR), and *Left-lowering* (index is positive for TL and negative for TR).

The shaping of the coronal plane was fairly consistent within speakers across consonant length, and speakers were fairly consistent within themselves as to which side of the tongue, if any, is lowered. For example, S6 was consistently *Right-lowering* and S7 was consistently *Convex*; S1, S5 and S8 preferred *Concave* across consonant length; S3 employed multiple shapes, preferring *Convex* for the o_o. S2 and S4 were not as consistent, using two preferred shapes for singletons, but converging into one for geminates.

Figure 3: Left-side lateralization (TB−TL).

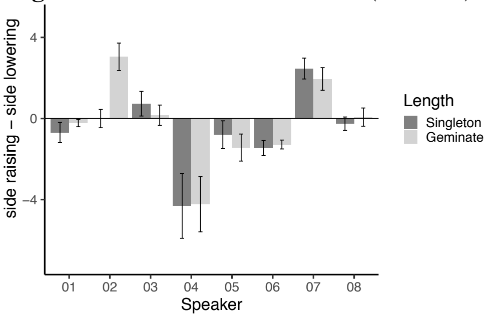


Figure 4: Right-side lateralization (TB−TR).

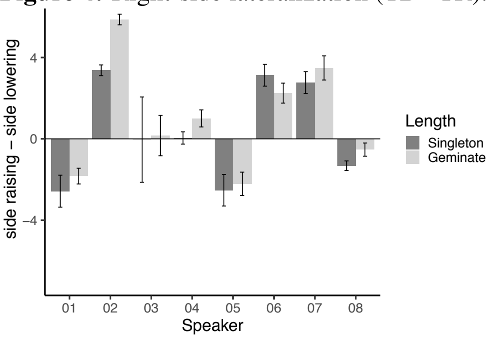


Table 4: Frequency of coronal shape pattern per speaker, based on the lateral indices.

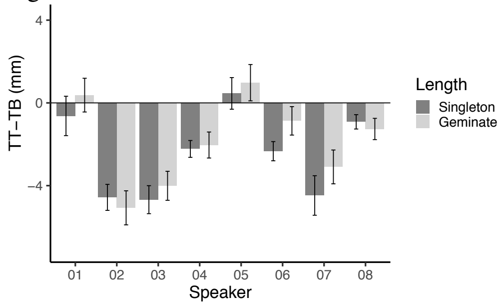
Length	Pattern	S1	S2	S3	S4	S5	S6	S7	S8	total
Singleton	Concave	8	0	2	4	6	0	0	6	26
	Convex	1	5	2	1	2	1	9	0	21
	Right-low	0	4	0	3	0	8	0	0	15
	Left-low	0	0	5	1	1	0	0	4	11
Geminate	Concave	6	0	2	1	5	1	0	4	19
	Convex	1	9	2	0	0	0	8	2	22
	Right-low	0	0	2	7	0	8	1	1	19
	Left-low	2	0	3	1	2	0	0	3	11

Overall, 68% of the productions had either or both sides of the tongue blade lowered, possibly allowing airflow through the lowered side(s) of the tongue.

3.4. Mid-sagittal curvature

Fig. 5 shows the difference in height between TT and TB at the gestural extrema of TB, as an indirect index of tongue curling [28, 34]. A positive value suggests tongue curling (TB is lower than TT), while a negative value suggests otherwise (TB is higher than TT). For most speakers, the values were largely negative. S2 and S7, with a *Convex* coronal shape, may be producing laminal laterals. In contrast, the TB height of S1, S5 and S8 is almost at level with TT. Taken together with their preference for *Concave* coronal shape, it is possible that their realization of liquids is apical rather than laminal. The general lack of difference across length ($t(142) = -0.99, p = 0.33$) may suggest that tongue curling is not one of the gestural goals for Japanese liquids.

Figure 5: Relative height (mm) of TT and TB at TB gestural maxima.



4. CONCLUSION

The present paper examined the articulation of geminated liquids through the activities of the tongue tip and tongue blade. While the results are preliminary and do not allow us to draw conclusions as to the precise strategies speakers employ, we obtained useful pointers for further analyses of the data. First, prolonged TT constriction is not the gestural target for geminated liquids for all speakers. We need to look into inter- and intra-speaker variability, taking into account the effect of surrounding vowels [13, 32] and the role of other gestures. Second, it is possible that some liquids are laterals, as pointed out previously, but not necessarily apical laterals with tongue curling. More detailed quantitative analyses taking into account the palate shape and acoustic consequences are in order, as well as further analyses with more temporal and spatial resolution to see what caused the relative consistency across length. It is also possible that /rr/ is realized as an apico-alveolar lateral tap, with prolonged tongue tip constriction and a tap-like release, described as a post-pausal variant of /r/ in [1]. The place of geminated liquids is still marginal in Japanese phonology, and the status may be responsible for the variability observed here. We hope that this line of investigation contributes to the understanding of liquids in general [19, 20, 24], the featural content of Japanese /r/ [16, 9, 23, 17], as well as L2 acquisition and pedagogy of liquids for Japanese speakers [11].

5. REFERENCES

[1] Akamatsu, T. 1997. *Japanese phonetics: Theory and practice* volume 3. Lincom Europa.

[2] Arai, T. 2013. On why Japanese /r/ sounds are difficult for children to acquire. *INTERSPEECH* 2445–2449.

[3] Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. *Glot International 5:9/10* 341–345.

[4] Derrick, D., Gick, B. 2011. Individual variation in English flaps and taps: A case of categorical phonetics. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 56(3), 307–319.

[5] Gafos, A. I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20(2), 269–337.

[6] Garcia, D. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis* 54(4), 1167–1178.

[7] Katz, W. F., Mehta, S., Wood, M. 2018. Effects of syllable position and vowel context on Japanese /r/: Kinematic and perceptual data. *Acoustical Science and Technology* 39(2), 130–137.

[8] Kawahara, S. 2007. Sonorancy and geminacy. In: *University of Massachusetts Occasional Papers in Linguistics 32: Papaers in Optimality III*. Amherst: GLSA 145–186.

[9] Kawahara, S. 2015. Japanese /r/ is not feature-less: A rejoinder to Labrune (2014). *Open Linguistics* 1(1).

[10] Kawahara, S. 2015. The phonetics of obstruent geminates, sokuon. In: Kubozono, H., (ed), *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*. Mouton 43–73.

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

[12] Kawakami, S. 1977. *Nihongo onsei gaisetsu [Outline of Japanese phonetics]*. Tokyo: Ofusha.

[13] Kiritani, M. M. S. S., Yoshioka, H. 1982. An electro-palatographic study of Japanese intervocalic /r/. *Ann. Bull. RILP* (16), 21–25.

[14] Kitamura, T., Nota, Y., Hashi, M., Hatano, H. 2014. A method of measuring articulatory space using NDI Wave speech research system. *IEICE technical report* 114(303), 89–93.

[15] Kurisu, K. 2014. The phonology of emphatic morphology in Japanese mimetics. *Japanese/Korean Linguistics* 22, 21–36.

[16] Labrune, L. 2014. The phonology of Japanese /r/: a panchronic account. *Journal of East Asian Linguistics* 23(1), 1–25.

[17] Labrune, L. 2017. More on Japanese /r/. *Journal of East Asian Linguistics* 26(3), 301–321.

[18] Ladefoged, P., Maddieson, I. 1996. *The sounds of the world's languages* volume 1012. Blackwell Oxford.

[19] Lindau, M. 1985. The story of /r/. In: Fromkin, V. A., (ed), *Phonetic Linguistics: Essays in honor of Peter Ladefoged*. Academic Press, Orlando.

[20] Magnuson, T. J. 2007. The story of /r/ in two vocal tracts. *Proceedings of the 16th International Congress of the Phonetic Sciences* 6–10.

[21] Nasu, A. 2002. *Nihongo-onomatope no gokeisei to inritu-koozoo [Word formation and prosodic structure of Japanese mimetics]*. PhD thesis University of Tsukuba.

[22] Ohnishi, M. 1987. Dôteki jinkô kôgai ni yoru nihongo shiin no kôsatsu [a study of Japanese consonants by means of an artificial palate]. *Onsei Gakkai Kaihô* 186, 19–24.

[23] Pellard, T. 2016. Why /r/ is not a special, empty consonant in Japanese. *Journal of East Asian Linguistics* 25(4), 351–383.

[24] Proctor, M. 2011. Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Laboratory Phonology* 2(2), 451–485.

[25] R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[26] Schourup, L., Tamori, I. 1992. Japanese palatalization in relation to theories of restricted underspecification. *GENGO KENKYU (Journal of the Linguistic Society of Japan)* 1992(101), 107–145.

[27] Shaw, J. A., Kawahara, S. 2018. The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics* 66, 100–119.

[28] Smith, C. 2014. Complex tongue shaping in lateral liquid production without constriction-based goals. *Proceedings of the International Seminar on Speech Production ISSP* 413–416.

[29] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21(3), 291–311.

[30] Tanaka, S. 2007. Itariago no juushiin to sokuon keisei [Geminate consonants in Italian and *sokuon* in Japanese]. *Proceedings of the 134th Meeting of the Linguistic Society of Japan* 252–257.

[31] Tiede, M. 2005. Mview: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

[32] Tsuzuki, M., Lee, H.-B. 1992. A phonetic study of the Korean and Japanese lateral, flap and nasal. *Proceedings of the 1992 Seoul international conference on linguistics* 761–780.

[33] Vance, T. J. 2008. *The Sounds of Japanese with Audio CD*. Cambridge University Press.

[34] Ying, J., Carignan, C., Shaw, J. A., Proctor, M. I., Derrick, D., Best, C. T. 2017. Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. *INTERSPEECH* 2978–2982.

---

[1] The experiment was followed by another experiment where the same speakers produced /l/ and /r/ in English.

[2] We give special thanks to Mafuyu Kitahara for allowing us to use the lab space at Sophia University.