



Partial devoicing of voiced geminate stops in Tokyo Japanese

Qandeel Hussain, Shigeko Shinohara

► To cite this version:

Qandeel Hussain, Shigeko Shinohara. Partial devoicing of voiced geminate stops in Tokyo Japanese. Journal of the Acoustical Society of America, 2019, 145 (1), pp.149-163. 10.1121/1.5078605 . hal-01975893v2

HAL Id: hal-01975893

<https://hal.science/hal-01975893v2>

Submitted on 7 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Partial devoicing of voiced geminate stops in Tokyo Japanese

Qandeel Hussain^{1,a)} and Shigeko Shinohara²

¹*Department of Linguistics, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, 16 University Avenue, Australian Hearing Hub, Balaclava Road, North Ryde, NSW 2109, Australia*

²*Laboratoire de Phonétique et Phonologie (UMR7018, Centre National de la Recherche Scientifique - Sorbonne Nouvelle), 19 rue des Bernardins, Paris, 75005, France*

(Received 25 August 2017; revised 12 October 2018; accepted 25 October 2018; published online 8 January 2019)

Tokyo Japanese has a constraint against voiced geminate stops in its native lexicon. The present study investigates whether recently introduced word-medial voiced geminate stops [C1V1C(C)2V2] are differentiated from voiceless geminates and voiced singletons in terms of duration, voicing during closure, and spectral moments of stop release bursts. The findings suggest that the voiceless and voiced singleton stops were clearly differentiated by C2 duration. In contrast, C2 duration of the voiceless and voiced geminate stops was not significantly different. The devoicing of the word-medial stops was not only observed in voiced geminates, but voiced singletons also showed devoicing. The duration of the preceding vowel (V1) distinguished the voicing contrast in both singleton and geminate stops. The first four spectral moments of C2 stop release bursts did not distinguish the length and voicing contrasts in stops. These results indicate that, although word-medial voiced geminate stops are fully or partially devoiced, the Tokyo Japanese speakers lengthen the preceding vowels (V1) to maintain a voicing contrast. Production patterns of the voiced geminates are considered in relation to marginal or intermediate phonological contrast.

© 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5078605>

[LK]

Pages: 149–163

I. INTRODUCTION

Tokyo Japanese differentiates word-medial voiceless singleton and geminate stops in its native lexicon (e.g., /kate/ “resource” vs /katte/ “selfishness”). Voiced geminates, on the other hand, are found only in recent loanwords (e.g., /sunobbu/ “snob,” /goddo/ “God,” and /eggu/ “egg”). Ito and Mester (1995) analyzed this phenomenon as a general ban on voiced geminates in Japanese being loosened at the marginal part of the lexicon. However, other forms that avoid voiced geminates, such as /betto/ “bed” (the variant of /beddo/), /bakku/ “bag” (the variant of /baggu/), and /robu/ “Rob” (without gemination), are also currently used. Together with some phonetic evidence implying neutralization of voiced geminates with voiceless geminates reported in previous studies (Fujimoto and Kataoka, 2016; Hirose and Ashby, 2007; Kawahara, 2005, 2006; details of these studies are introduced below), the existence of variant forms avoiding voiced geminates leaves us in doubt whether voiced geminate stops are an established phonological category in Tokyo Japanese. Although voiced geminates are banned in the native vocabulary, there seems to be a driving force to level voiceless and voiced geminate stops in loanwords. Gemination is systematic in case of loanwords from English ending in a voiceless stop: /toppu/ “top,” /kjatto/ “cat,” and /pikunikku/ “picnic.” Among various theories on gemination in Tokyo Japanese (Ito *et al.*, 2017; Katayama, 1997; Kawagoe, 1995; Lovins, 1975; Ohso, 1971; among others),

Tsuchida (1995) explains this phenomenon as an alignment between the word boundary in the source word and the syllable boundary in the loanword form. Since a word cannot end in a consonant in Tokyo Japanese, this can be a way to satisfy this alignment between morphological and prosodic structures. As a result of leveling, the newly introduced voiced geminates appear in a limited number of lexical items. Hall (2013) considers phonological contrast occurring only in certain positions or certain type of lexicons to be an “intermediate” status between contrast and allophony. As shown below, voiced geminate stops are typologically disfavored by articulatory constraints and, hence, the newly created sound sequences introduce a marked structure in Tokyo Japanese.

Several studies have shown that voiced geminates are typologically rarer than voiceless geminates (Blevins, 2008; Thurgood, 1993). For example, Makasar (Austronesian) has a word-medial singleton vs geminate contrast but only in voiceless stops (/rapo/ “unlucky” vs /rappo/ “fruit”: Tabain and Jukes, 2016). In some languages, like Punjabi (Indo-Aryan), voicing is contrastive in word-medial singleton and geminate stops (/bæʈʃa/ “save,” /bæʈʃʃa/ “child,” /sæɖʒa/ “decorate,” and /sæɖʒʒa/ “right”: Bhatia, 1993). The absence of voiced geminates in some languages (Makasar) and their ambiguous status in others (Tokyo Japanese) is generally attributed to the aerodynamic conditions that arise from supraglottal articulations, as maintaining voicing and long closure duration in voiced geminates is challenging (Ohala, 1983). Kirchner (1998) claimed that the production of voiced geminates requires more articulatory effort than voiceless geminates. Fujimoto and Kataoka (2016) provided oral airflow data indicating that only the voiceless and

^{a)}Current address: Department of English (Linguistics Program), North Carolina State University, Campus Box 8105, Raleigh, NC 27695, USA. Electronic mail: qandeel_hussain@ncsu.edu

voiced singletons (/t/ vs /d/) of Tokyo Japanese are clearly differentiated, but voiceless and voiced geminates (/tt/ vs /dd/) are not. It is still unclear how voicing contrast is manifested in the acoustics of voiced geminate stops in Tokyo Japanese. In order to better understand the partial devoicing of voiced geminates, the current study presents a detailed acoustic analysis of the Tokyo Japanese voicing distinction in both singleton and geminate stops.

Kawahara (2005, 2006) examined the devoicing of non-sense voiced geminates in Japanese speakers from Hiroshima, Shizuoka, and Tokyo. The measurements included closure voicing, closure duration, preceding vowel duration, and *F0* and *F1* of the preceding and following vowels. The percentage of closure voicing in voiced singleton stops was almost 100%. For the voiced geminate stops, the Japanese speakers maintained the closure voicing only around 30%–40% of the total closure duration. These findings were confirmed by Hirose and Ashby (2007) who also noted partial devoicing of voiced geminates (47% of the total closure duration) and singletons (87% of the total closure duration). However, Kawahara (2006) noted that the voicing contrast in geminates is differentiated by the duration of preceding vowels.

Cross-linguistic studies have demonstrated that the most salient acoustic correlate of geminate stops is closure duration (Lahiri and Hankamer, 1988). The effect of voicing on closure duration has also been reported. The voiceless singleton and geminate stops of Tokyo Japanese are longer than their voiced counterparts (Homma, 1981; Idemaru and Guion, 2008). The average closure durations of /b/ and /bb/ were 55 ms and 159 ms, respectively; on the other hand, the average closure durations of /p/ and /pp/ were 77 ms and 183 ms, respectively (Homma, 1981). These acoustic differences in the duration of singleton and geminate stops are also reflected in their articulation. The tongue tip and body movements for the long (geminate) consonants in Japanese are much slower than singletons (Fujimoto *et al.*, 2015; Löfqvist, 2007).

Apart from the closure duration of singleton and geminate stops, some studies have investigated the effect of C2 duration on C1 duration in a C1V1C(C)2V2 sequence, where C(C)2 is either a singleton or a geminate. However, there is no consensus in the reported results. For instance, Han (1994) noted that /k/ (74.5 ms) in /kite/ “postal stamps” was longer than /k/ (63 ms) in /kite/ “wearing” in Tokyo Japanese. C1 lengthening by a following geminate consonant in Tokyo Japanese was supported by a recent study by Kamiyama and Turco (2017). Italian also showed a lengthening effect of C2 geminates on C1. The average duration of C1 was 89.1 ms in the vicinity of C2 singletons, but C1 was 97.8 ms long in geminates (Turco and Braun, 2016). The effect of C2 duration on C1 in Tokyo Japanese, if confirmed along with other language data, may provide insights into speech planning (the adjustment of the articulators for the upcoming singleton and geminate consonants) and long-distance anticipatory effects.

Voice Onset Time (VOT) or release duration is another acoustic property of the singleton vs geminate contrast, but compared to closure duration, there is no consensus on how robust VOT is in differentiating consonantal length

contrasts. Hirata and Whiton (2005) found no VOT differences in singletons /t k/ and geminates /tt kk/ (labial stops were not included in the study). See also Beckman (1982) for similar results in Japanese and Ham (2001) for Bernese (a dialect of Swiss German spoken in the Bern city and neighboring regions). A difference was found in initial singleton /t/ and geminate /tt/ of the Ikema dialect of Ryukyuan, where singletons had a significantly longer VOT than their geminate counterparts (Shinohara and Fujimoto, 2018). Other cross-linguistic studies showed no clear differences in singleton and geminate stops in terms of VOT (Bengali: Lahiri and Hankamer, 1988).

In recent years, a number of studies have identified areal and generational differences in the percentage of voicing in voiced stops (singletons and geminates) across Japanese dialects. VOT of the voiced stops is variably realized with either voicing lead or short lag (Takada, 2011; cited in Takada, 2012). For instance, Kinki (a regional dialect of Japanese) speakers born in 1910 had a *true* voicing contrast in word-initial stops, whereby the voiced stops were realized with a voicing lead VOT and the voiceless unaspirated stops with a short lag VOT (Takada, 2012). Kinki speakers born in 1980, on the other hand, merged the VOT contrast of voiceless vs voiced unaspirated stops. Similarly, in Tokyo Japanese, a *true* voicing contrast (voiceless vs voiced) is changing to a voiceless vs devoiced (long lag vs short lag) contrast (Takada, 2012; Riney *et al.*, 2007). Thus, the phenomenon of partial devoicing of voiced stops is active in many Japanese dialects even within the singleton category.

There is also an intrinsic consonant duration difference between voiceless and voiced stops (C2). Voiceless stops are known to take more time to produce than voiced stops (Stevens and Hajek, 2004; among others). This is thought to be related to the duration of the preceding vowel in a syllable closed by a stop consonant. The lengthening of preceding vowels is a well-established acoustic correlate of the voicing contrast in many languages [see Beguš (2017) for a detailed review; Chen, 1970; Lisker, 1974]. Some studies have reported the effect of voicing on the duration of following vowels (longer after voiced singleton/geminate than voiceless singleton/geminate stops: Idemaru and Guion, 2008). A number of studies on Japanese have reported the lengthening of vowels before geminates (Campbell, 1999; Fujimoto and Maekawa, 2014; Han, 1994; Hirose and Ashby, 2007; Idemaru and Guion, 2008; Kawahara, 2006). In contrast, vowels after geminates are characterized by shorter durations than after singletons (Homma, 1981; Idemaru and Guion, 2008).

One of the interesting aspects of Tokyo Japanese is its moraic isochrony (Beckman, 1982; Han, 1962; Homma, 1981). While the exact isochrony of each mora is denied, there is a consensus that moraic timing control is active at the word level (Han, 1994; Port *et al.*, 1987; Sagisaka and Tohkura, 1984). Hence, word durations are proportional to their mora counts at a given tempo. We may expect a systematic duration difference between words containing a singleton and those containing a geminate consonant even though the difference does not have any impact on the syllable count (e.g., /ata/ = two syllables, two moras; /atta/ = two syllables, three moras).

There are non-durational acoustic measures of stop release bursts, like the first four spectral moments [spectral center of gravity (CoG), spectral standard deviation (SD), spectral skewness, and spectral kurtosis], that have been widely used to classify stops produced at different places of articulation (Hussain *et al.*, 2017; Tabain, 2012; Tabain and Butcher, 2015; Tabain *et al.*, 2016) but are not generally reported in the literature on the singleton vs geminate contrast. Spectral moments are correlated with the articulatory configurations of the vocal tract and therefore can be used to investigate the length of the front cavity, tongue posture differences, and the voicing distinction in singleton and geminate stops (Chodroff and Wilson, 2014; Hussain, 2017; Hussain *et al.*, 2017; Li *et al.*, 2009; Sundara, 2005). More precisely, spectral CoG represents the mean distribution of acoustic energy in the burst spectrum and is correlated with the length of the front cavity (Forrest *et al.*, 1988; Li *et al.*, 2009; Nittrouer, 1995). Spectral SD is a measure of dispersion on both sides of the mean (spectral CoG) and is generally used to differentiate between the tongue postures of consonants. Spectral skewness shows whether the acoustic energy is skewed toward higher (negative skewness) or lower (positive skewness) frequencies and also correlated with the length of the front cavity (Jongman *et al.*, 2000). Spectral kurtosis indicates the degree to which the spectral envelope is focused in a particular region, whether it is *compact* or *diffuse*, which is an indicator of the tongue postures of consonants produced at different places of articulation. One of the reasons spectral moments have not been investigated is because these measures play little role in the classification of consonantal length contrasts. For instance, the singleton and geminate fricatives of Lebanese Arabic are not clearly distinguished by spectral SD, spectral skewness, or spectral kurtosis (Al-Tamimi and Khattab, 2015). However, spectral moments may be useful to characterize voicing contrasts (Sundara, 2005). Moreover, some studies have noted that geminates have stronger constriction and more extended tongue contact than singletons (Kawahara and Matsui, 2017; Kochetov and Kang, 2017). This suggests that the burst spectrum for geminates would be different from their singleton counterparts. We tested the relevance of spectral moments for the characterization of voicing and length contrasts in Tokyo Japanese.

The aim of the current study is to investigate whether voiced geminates are realized as fully or partially voiced in Tokyo Japanese. We compared a number of temporal correlates of voicing and length contrasts in word-medial stops. In particular, we investigated whether (a) acoustic correlates of the voicing distinction are reflected in the duration of voiceless and voiced singleton and geminate stops, as well as in the durations of flanking vowels; and whether (b) spectral moments of stop release bursts reliably differentiate the voiceless and voiced singleton and geminate stops of Tokyo Japanese. If voiced geminates (as well as singletons) are partially devoiced, their release bursts may be similar to voiceless singletons and geminates. In addition, we examined the effects of C2 voicing on C1 duration.

First, we predicted that C1 duration would be longer when C2 is a geminate but C1 would be shorter when C2 is a singleton, as reported for Tokyo Japanese (Kamiyama and

Turco, 2017) and Italian (Turco and Braun, 2016). C2 duration is predicted to be longer in geminates than in singletons (Han, 1994; Hirata and Amano, 2012; Idemaru and Guion, 2008; among others). Moreover, C2 duration would be longer in voiceless singleton stops than in voiced singleton stops (Idemaru and Guion, 2008; Stevens and Hajek, 2004). The percentage of voicing during the closure intervals of C2 would be higher in voiced singletons than in their voiced geminate counterparts (Kawahara, 2006; Hirose and Ashby, 2007). We predicted that vowel (V1) duration would be longer before geminates than before singletons (Campbell, 1999; Han, 1994; Fujimoto and Maekawa, 2014; Kawahara, 2005, 2006). On the other hand, vowel (V2) duration following the geminate would show the reverse pattern, with shorter vowel durations after geminates than after singletons (Campbell, 1999; Han, 1994). We also predicted that if voiced geminates in Tokyo Japanese are partially devoiced, then their duration and spectral moments would be similar to the voiceless geminates (Kawahara, 2015, 2016; among others). Hence, we predicted no clear differences in the duration and spectral moments of C2 voiceless and voiced geminates.

II. METHODS

A. Participants

Five female participants were recruited for the experiment (mean age 38.8 yr). All the participants were native speakers of Tokyo Japanese, born and raised in Tokyo or in neighboring Kanagawa, and were residing or working in Tokyo during the recording session.

B. Speech material

Table I presents the list of nonsense words presented to the participants. All the words had C1V1C(C)2V2 structure where C1 was a velar /k/, C2 was either a singleton /p b t d k g/ or a geminate /pp bb tt dd kk gg/ (see Homma, 1981, and Kawahara, 2006, for a similar set of words). The following and preceding vowels (V1 and V2) were /a/.¹ As the neighboring segments can have an impact on the phonetic properties of the target sounds, a controlled set of stimuli was needed in the current study. The target words were elicited in a carrier phrase *Dakara _____ desu* “Thus, it is _____.” As mentioned also in Kawahara (2006), it is hard to completely avoid the intrusion of lexical words in the word-list. /kappa/ “an imaginary creature living in rivers,” /kaba/ “hippopotamus,” /kata/ “shoulder,” /kakka/ “excellency” and /kaga/ “a place name” could be recognized as lexical words. /katta/ could be recognized as a phrase (“to win. PAST”).

TABLE I. Target nonsense words used in the current study.

Place	± Voicing	Singleton	Geminate
Labial	–	/kapa/	/kappa/
	+	/kaba/	/kabba/
Alveolar	–	/kata/	/katta/
	+	/kada/	/kadda/
Velar	–	/kaka/	/kakka/
	+	/kaga/	/kagga/

Since Tokyo Japanese is a pitch accented language and the accent is not indicated in the orthography, when a real word is recognized in the test wordlist, a participant might use the lexical accent of that word instead of the regular default initial accent pattern on nonsense words [high-low (HL) contour]. In the wordlist, the five out of six possible lexical words have the same initial accent. The only word that can carry a different pitch accent, if recognized, is /kappa/, “an imaginary creature living in rivers,” with a low-high (LH) contour. However, the test words were all written in *katakana*, which is used to represent sound sequences without reference to meaning. The second author (S.S., a native speaker of Japanese) checked the pitch accent patterns of the produced tokens. All the target words were produced with HL contour, except some items by one speaker (S4) who produced the following tokens with LH contour: /kaka/ (two tokens), /kaga/ (five tokens), /kata/ (three tokens), and /kada/ (four tokens). /kata/ with LH contour could correspond to the word meaning “form.” The others do not correspond to any lexical items.

C. Procedure

The participants were recorded in a sound-attenuated room at Sophia University, Tokyo, Japan. A list of nonsense words written in Japanese (*katakana* syllabary) was presented to the participants. All the participants were instructed to produce the target words in a normal speech tempo with a HL accent, which is the default pitch pattern in nonsense word pronunciation (McCawley, 1968). Note that a prosodic resetting of F_0 occurs at the onset of the target word so that there is no “downstep” (Kubozono, 1993) or “catathesis” (Pierrehumbert and Beckman, 1988) that would flatten the initial accent after the accented word *dakara* “thus.” The participants were given one training example before the recording. A linear PCM recorder (SONY PCM-D10, Sony Corporation, Tokyo, Japan) with an external Sony ECM-959DT microphone was used to make the audio recordings. All speech was recorded in stereo mode, at 44.1 kHz and encoded in 16 bit, uncompressed WAV format. The stereo recordings were converted to mono by averaging the two channels, using the default mono conversion algorithm in Praat (Boersma and Weenink, 2014).

D. Acoustic analysis

A total of 300 tokens (12 words \times 5 repetitions \times 5 participants) were recorded and segmented in Praat (Boersma and Weenink, 2014). Five tokens were excluded (four due to spirantization of the voiced velar /g/ to the voiced velar fricative [ɣ] in /kaga/ and one token due to devoicing of V1 in /kapa/, which made it difficult to mark the boundaries between C1 and V1). Figures 1(a) and 1(b) present examples of the segmented tokens. Closure durations (Clo1 and Clo2), and C1 and C2 release bursts were collapsed to get C1 and C2 durations. Both vowels (V1 and V2) and consonants (C1 and C2) were segmented by visual inspection of the waveforms and spectrograms, making reference to periodicity and formant energy bands (Hussain, 2015; Hussain et al., 2017; Ridouane, 2007). Whole word duration was calculated by

collapsing C1, C2, V1, and V2. Closure voicing, which can only be found in voiced stops, was identified by observing the voicing bar (as well as waveform) during C2 closure [/kaba/ in Fig. 1(b); see Kawahara, 2005]. The first four spectral moments (CoG, SD, skewness, and kurtosis) of C2 stop release bursts [the interval of *Rel2* in Figs. 1(a) and 1(b)] were also measured (Forrest et al., 1988). The fast Fourier transform (FFT) spectra were generated over a single 5 ms Hamming analysis window centered at the beginning of C2 release, using the power value of $p = 2$ (Boersma and Weenink, 2014). The duration of the shortest C2 release burst was 4 ms, therefore, a narrow analysis window was used to reduce the effect of following vowel (V2) on spectral moments (Hussain et al., 2017). To calculate the first four spectral moments, the acoustic signal was down-sampled from 44.1 kHz to 22.05 kHz (Bundgaard-Nielsen et al., 2016). In order to compare the spectral moments of release bursts of voiced stops with voiceless stops, a high pass filter with a cut-off frequency of 200 Hz was used to filter out the voicing components during the whole C2 interval of each voiced stop (Jongman et al., 1985; Sundara, 2005).

E. Statistical analysis

A series of linear mixed effects (LME) models was conducted in R (R Core Team, 2012), using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages. For all the temporal [C1, V1, C2, C2 release, V2, WW (whole word) durations] and spectral (CoG, SD, skewness, and kurtosis) measures, Speaker (five speakers) was used as a random factor, Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) were included as fixed factors. Separate LME models for singleton and geminate stops were also performed, with Speaker as a random factor, Place (labial, alveolar, and velar), and Voicing (voiceless and voiced) as fixed factors. Random slopes of Speaker for main effects of all fixed factors were included in the models (Barr et al., 2013). Our interest is in the two fixed factors of Voicing and Length. The main effect of Place is not of interest in the current study and, therefore, will not be discussed in the text (though presented in the tables). We will only discuss the main effects of Voicing and Length and their interactions.

III. RESULTS

A. C1 and C2 durations

Figure 2 summarizes the durations of C1 and C2 in voiceless and voiced singleton and geminate stops, across five speakers. Absolute mean durations are presented in the Appendix. To test the effects of C2’s place, voicing, and length on C1 duration, an LME model was conducted, with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) as fixed factors. Table II presents the statistical results of the LME model, which showed that there was a significant effect of Voicing on C1 duration (voiced > voiceless). However, no effect of Length on C1 duration was observed. The two-way interaction of Voicing \times Length was also significant.

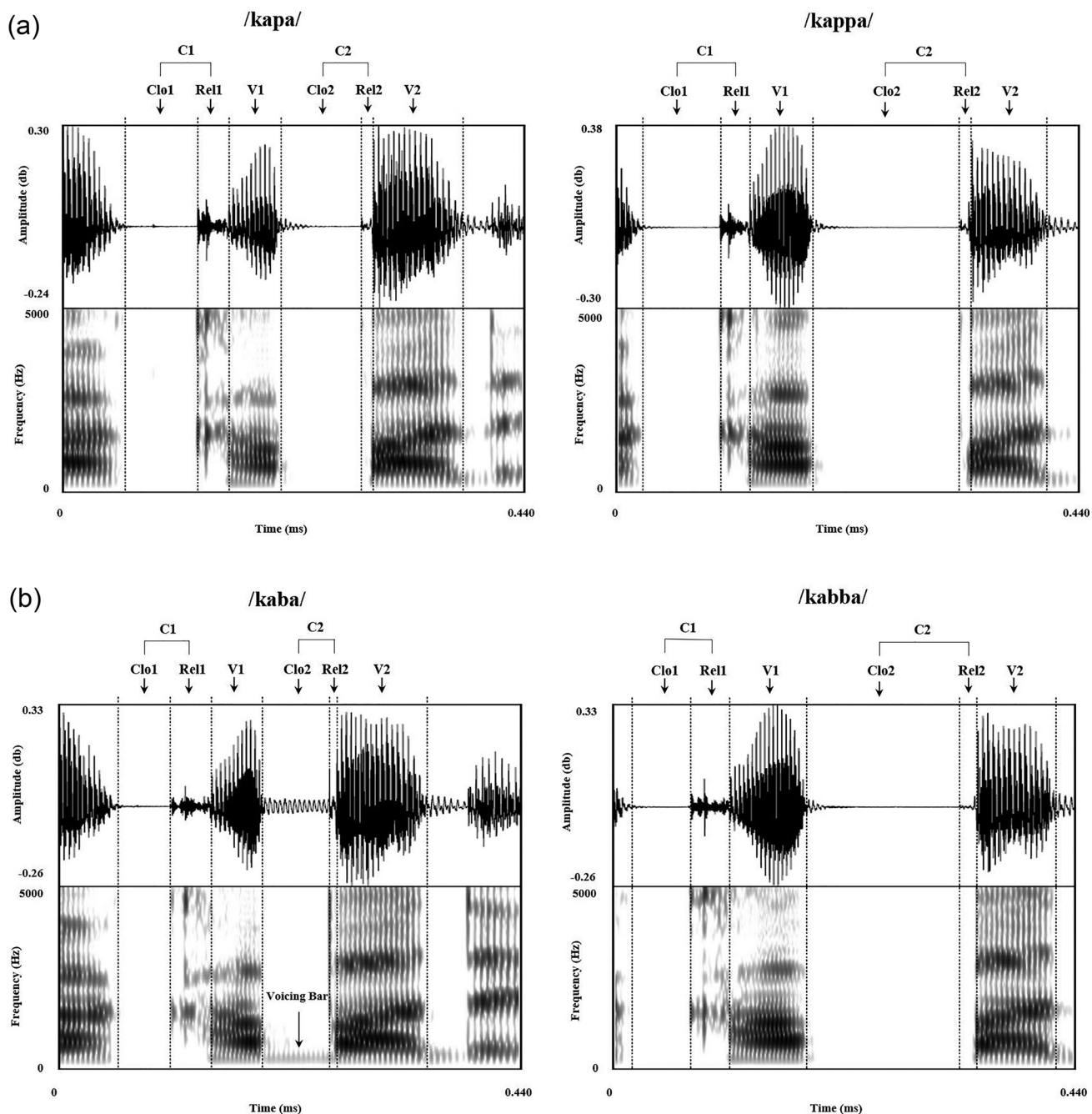


FIG. 1. (a) Spectrograms and waveforms of /kapa/ (left) and /kappa/ (right) produced by a female Tokyo Japanese speaker. C1, V1, C2, and V2 intervals are indicated with dashed lines. (b) Spectrograms and waveforms of /kaba/ (left) and /kabba/ (right, with devoiced /bb/) produced by a female Tokyo Japanese speaker. C1, V1, C2, and V2 intervals are indicated with dashed lines. Voicing bar during the C2 closure interval is shown on the left spectrogram.

Separate LME models on C1 duration in singleton and geminate contexts were performed (Table III). The results showed that, for the singleton context, the factor of Voicing (voiced > voiceless) was statistically significant. In the geminate context, none of the factors were statistically significant. Although C1 duration varied as a function of Voicing, it can be noted that the mean durations of C1 (Fig. 2) do not show very large differences across speakers.

Another LME model was conducted on C2 duration, with the random factor of Speaker (five speakers) and the three fixed factors of Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and

geminate). The results indicated that the two main fixed factors of Voicing (voiceless > voiced stops) and Length (geminate > singleton stops) were significant (see Fig. 2, the right two columns). The two-way interaction of Voicing \times Length was also significant. Although geminate stops were consistently longer than their singleton counterparts, the magnitude of difference was larger for the voiced pairs. Separate LME models in singleton and geminate contexts (Table III) showed a main effect of Voicing in the singleton context (voiceless singleton > voiced singleton). In contrast, none of the factors achieved statistical significance in the geminate context. This is in line with our predictions, where C2

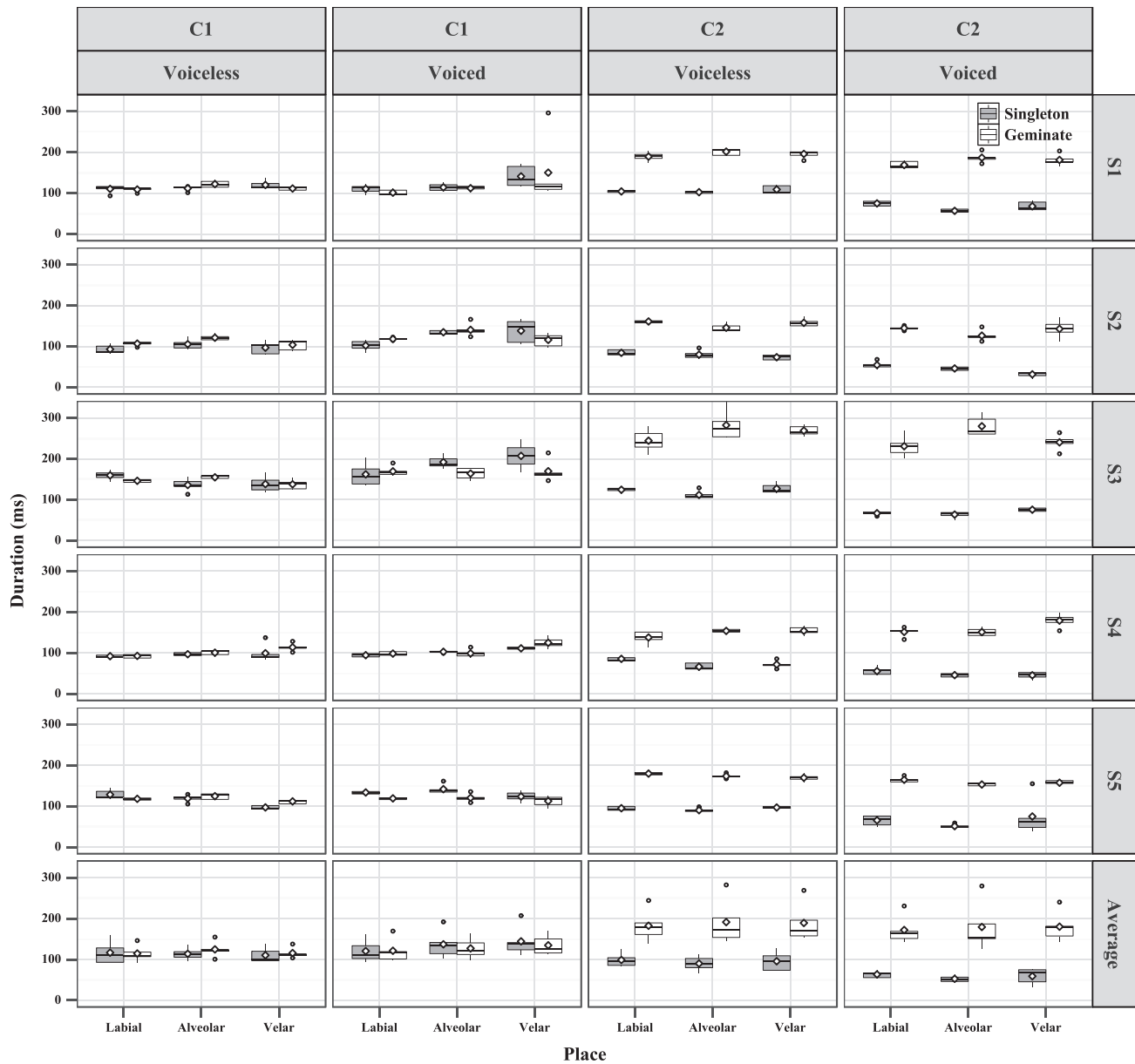


FIG. 2. Boxplots of C1 (left two columns) and C2 (right two columns) durations in voiceless and voiced singleton and geminate stops of Tokyo Japanese. The figure indicates the mean [diamonds (\diamond)], median (black horizontal line), the highest 25% of the data points (upper quartile), the lowest 25% of the data points (lower quartile), 50% of the data points (between the upper and lower quartiles of the boxes), maximum value (upper whisker/outlier), and minimum value (lower whisker/outlier). The black dots (\bullet) denote outliers. The raw data of all the five speakers (S1, S2, S3, S4, and S5: the first five rows) are presented, followed by the average data across five speakers (last row).

TABLE II. Results of LME models of C1, V1, C2, C2 release, V2, and WW (whole word) durations (ms), with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) as fixed factors. Asterisks denote the level of statistical significance ($p < 0.05^*$; $p < 0.01^{**}$; $p < 0.001^{***}$) (footnote 2).

Temporal measures													
Factors	df	C1		V1		C2		C2 release		V2		WW	
		<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Place (P)	2	1.66	=0.327	9.55	=0.050	0.40	=0.701	7.89	=0.064	0.74	=0.548	2.18	=0.261
Voicing (V)	1	8.67	*	52.31	**	15.91	*	6.21	=0.068	40.38	**	16.32	*
Length (L)	1	0.01	=0.910	13.64	*	54.26	**	0.00	=0.953	15.67	*	135.13	***
$P \times V$	2	9.56	***	3.56	*	0.52	=0.596	6.97	**	2.70	=0.069	4.09	*
$P \times L$	2	0.04	=0.960	1.18	=0.309	7.51	***	4.92	**	2.19	=0.114	2.03	=0.133
$V \times L$	1	8.08	**	6.65	*	34.93	***	10.12	**	41.00	***	0.36	=0.548
$P \times V \times L$	2	3.34	*	2.21	=0.112	0.23	=0.792	12.04	***	1.06	=0.347	2.45	=0.088

TABLE III. Results of separate LME models in singleton and geminate contexts, with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), and Voicing (voiceless and voiced) as fixed factors. Asterisks denote the level of statistical significance ($p < 0.05^*$; $p < 0.01^{**}$; $p < 0.001^{***}$) (footnote 3).

Consonant	Factors	df	Temporal measures											
			C1		V1		C2		C2 release		V2		WW	
			<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Singleton	Place (P)	2	0.72	=0.555	6.84	=0.077	3.09	=0.189	2.80	=0.208	0.25	=0.791	2.13	=0.269
	Voicing (V)	1	8.96	*	53.42	**	23.61	**	7.28	=0.055	52.49	**	13.99	*
	P × V	2	12.80	***	0.10	=0.901	0.41	=0.666	12.35	***	1.82	=0.165	4.33	*
Geminate	Place (P)	2	1.67	=0.326	10.13	*	0.55	=0.628	12.45	*	1.77	=0.311	1.46	=0.362
	Voicing (V)	1	4.51	=0.101	31.47	**	3.91	=0.119	1.70	=0.263	3.60	=0.131	5.25	=0.084
	P × V	2	2.61	=0.077	5.54	**	0.07	=0.935	0.97	=0.383	2.17	=0.118	1.99	=0.142

duration was expected to differ between voiceless and voiced singleton stops but no difference in C2 duration was predicted for voiceless and voiced geminate stops, indicating partial neutralization of the voicing contrast in geminates. There were some durational differences across speakers. S3 showed the largest difference in C2 durations between singletons and geminates.

Figure 3 presents the C2 release durations in voiceless and voiced singleton and geminate stops. An LME model was performed on C2 release duration, with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) as fixed factors (Table II). The results indicated that there were no main effects of Voicing and Length on C2 release durations. However, the two-way interaction of Voicing × Length was significant. Separate LME models on C2 release duration in singleton and geminate contexts also showed no significant effect of Voicing in both contexts (Table III). These findings suggest that C2 release duration is not a reliable descriptor of voicing and length of word-medial stops in Tokyo Japanese. There were no clear inter-speaker differences in C2 release durations.

B. C2 voicing

Figure 4 presents the percentage of voicing (proportion of voicing over the entire C2 duration) in voiced singleton and voiced geminate stops. Representative examples of complete or partial devoicing can be found in supplementary materials.⁴ To test whether there are differences in the voicing of both type of stops, an LME model was performed on voicing (%), with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), and Length (voiced singleton and voiced geminate) as fixed factors. The results showed that Length had a significant effect on voicing ($df = 1$, $F = 14.89$, $p = 0.018$: voiced singleton > voiced geminate). It was noted that both voiced singletons and voiced geminates showed voicing at the beginning of closure. As illustrated in Fig. 4, the average percentage of closure voicing in voiced singleton stops was around 70%–80%. In contrast, for voiced geminate stops, the Tokyo Japanese speakers maintained closure voicing for only around 20%–25% of the total closure duration. Some individual differences can also be noticed. S1 showed the lowest percentage of voicing for both voiced singleton and geminate stops. S2 produced the voiced singleton

stops with 100% voicing throughout the closure duration and around 70%–90% voicing for the geminate stops.

C. V1 and V2 durations

Boxplots of V1 and V2 durations preceded/followed by voiceless and voiced singleton and geminate stops are presented in Fig. 5. To test the effects of place, voicing, and length on V1 duration, an LME model was conducted. The random factor of Speaker (five speakers) and three fixed factors of Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) were included in the model. The results are presented in Table II, which showed that Voicing (voiced > voiceless stops) and Length (geminate > singleton stops) had a significant effect on V1 duration. The two-way interaction of Voicing × Length was also significant. Separate LME models (Table III) indicated the effect of Voicing on V1 duration in the singleton (voiced singleton > voiceless singleton) and geminate (voiced geminate > voiceless geminate) contexts. These results suggest that Tokyo Japanese speakers may use V1 duration to maintain the voicing contrast in singletons and geminates. Some minor inter-speaker differences in V1 durations were also observed. S1 had a larger difference in V1 durations, compared to other speakers.

Another LME model on V2 duration was performed, using the random factor of Speaker (five speakers) and the three fixed factors of Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate). The results showed the effects of Voicing (voiced > voiceless) and Length (singleton > geminate) on V2 duration, and a significant interaction between Voicing × Length. Separate LME models in the singleton and geminate contexts (Table III) indicated a significant effect of Voicing in the singleton context (voiced singleton > voiceless singleton). In the geminate context, none of the factors were statistically significant. V2 durations produced by S3 were slightly longer than other speakers.

D. Whole word durations

Figure 6 presents the whole word durations of words containing voiceless and voiced singleton and geminate stops. An LME model was performed using the random factor of Speaker (five speakers) and fixed factors of Place (labial, alveolar, and velar), Voicing (voiceless and voiced),

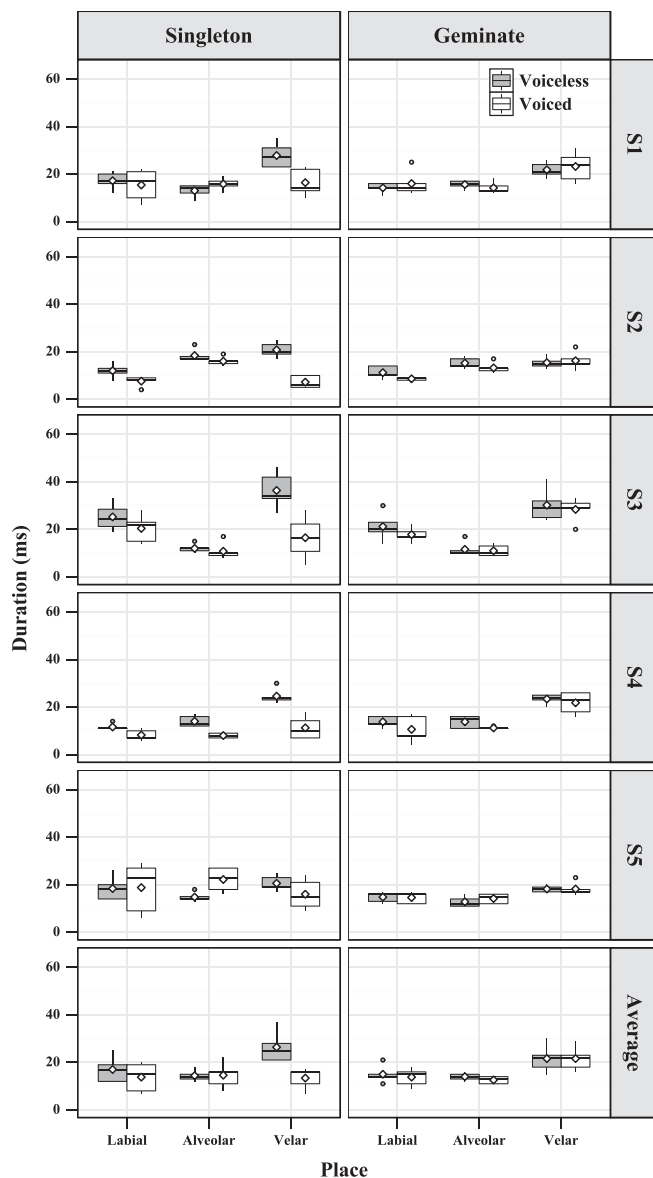


FIG. 3. Boxplots of C2 release durations in singleton (left column) and geminate (right column) stops of Tokyo Japanese.

and Length (singleton and geminate) (Table II). The results indicated that Voicing (voiced > voiceless) and Length (geminate > singleton) had a significant effect on whole word duration. Separate LME models in singleton and

geminate contexts (Table III) showed differences in the whole word duration of voiceless and voiced singleton stops (voiced singleton > voiceless singleton). In contrast, none of the factors in the geminate context were statistically significant, which further confirms partial neutralization of the voicing contrast in geminates. As noted earlier, S3 had comparatively longer C2 and V2 durations than other speakers. These durations are also reflected in the whole word durations of S3.

E. Spectral moments of C2 stop release bursts

Boxplots of the first four spectral moments (spectral CoG, spectral SD, spectral skewness, and spectral kurtosis) of C2 stop release bursts across voiceless and voiced singleton and geminate stops are presented in Figs. 7 and 8. Table VI in the Appendix shows the absolute mean values of all four spectral moments. To test the effect of place, voicing, and length on the spectral moments of C2 stop release bursts, an LME model with the random factor of Speaker (five speakers) and the three fixed factors of Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) was performed. The results are presented in Table IV, which indicated that Voicing and Length had no significant effect on the first four spectral moments. There were no clear inter-speaker differences in spectral moments of C2 stop release bursts.

Separate LME models in singleton and geminate contexts are presented in Table V. In both contexts, the main factor of Voicing was not significant for any of the four spectral moments. The overall results suggest that the first four spectral moments did not reliably differentiate the voicing and length contrasts of word-medial stops of Tokyo Japanese.

IV. SUMMARY AND DISCUSSION

The current paper presented a detailed acoustic analysis of the voicing contrast in word-medial singleton and geminate stops of Tokyo Japanese. In particular, since voiced geminate stops are reported to be devoiced, we investigated whether voiced geminate stops are acoustically similar to voiceless ones. Both temporal and spectral measures were examined, which provided information about the secondary

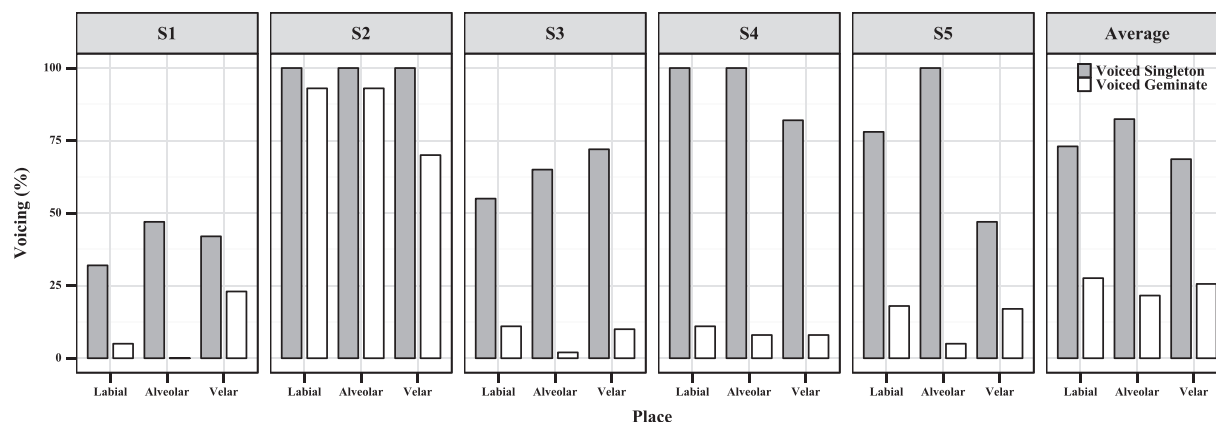


FIG. 4. Bar plots of proportion of voicing (%) over the entire closure duration in C2 voiced singleton and voiced geminate stops of Tokyo Japanese.

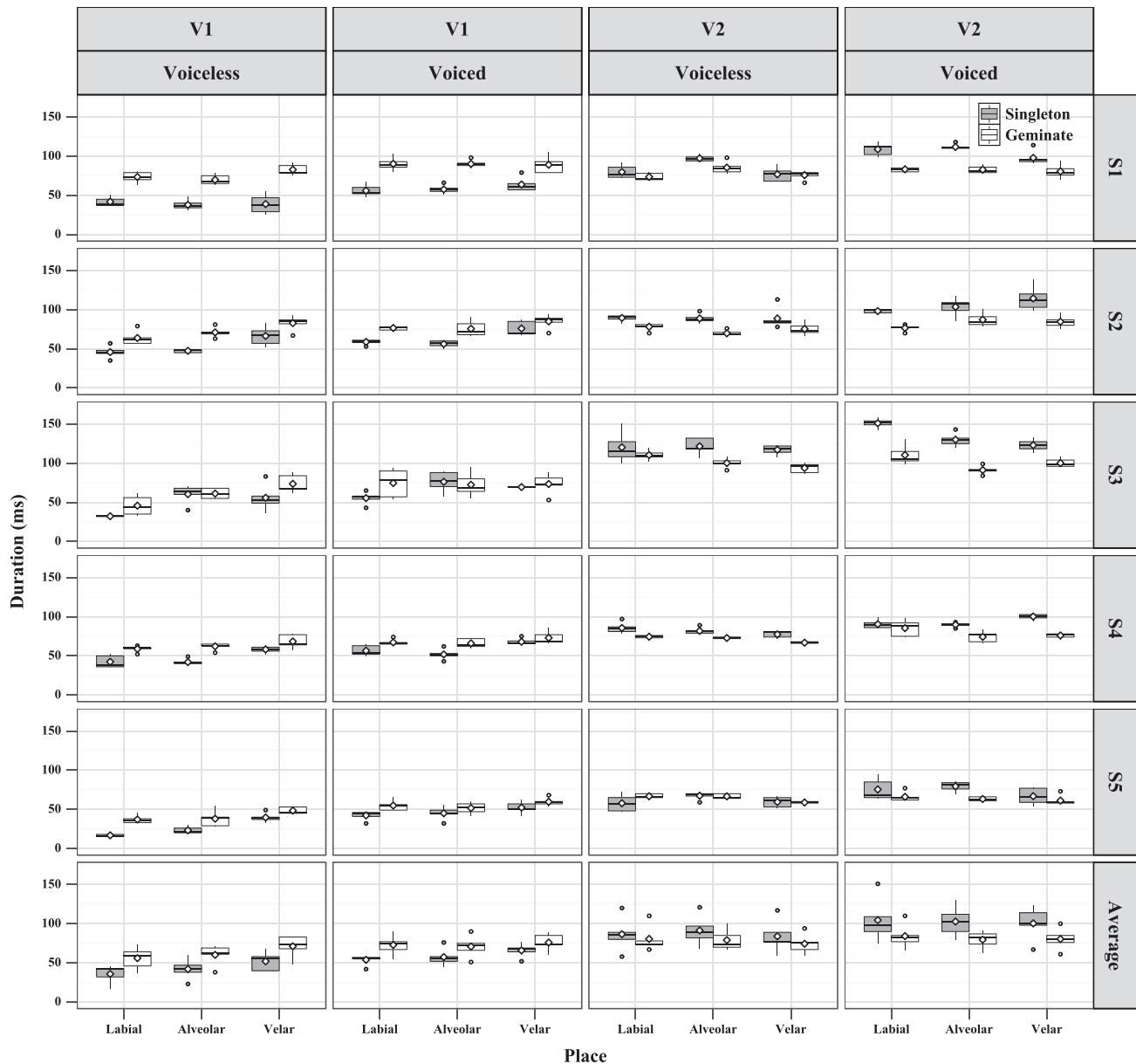


FIG. 5. Boxplots of V1 (left two columns) and V2 (right two columns) durations preceded/followed by the voiceless and voiced singleton and geminate stops of Tokyo Japanese.

(C1, V1, and V2 durations) and primary (C2 duration, voicing during C2 closure, and spectral moments of C2 stop release bursts) acoustic correlates of voicing and length in stop consonants. The findings indicated that the phonetic realization of C2 closure voicing in Tokyo Japanese is much more variable than has been previously reported (Kawahara, 2006; Homma, 1981). The temporal correlates of the voicing contrast of C2 are robust, but only for singletons. Moreover, while the expected devoicing was found in voiced geminates, devoicing was also observed in voiced singletons. The sole duration of the preceding vowel (V1) clearly differentiated the voicing contrast in geminate stops. This result implies that the category of voiced geminate is phonetically manifested.

Our results of C1 duration showed that the voicing of C2 singletons had a significant effect on the duration of C1 singletons. However, the voicing of C2 geminates had no effect on the duration of C1 geminates (Table III). From the

perspective of moraic timing, we can assume that C1 duration compensates for C2 duration variation due to voicing. Could this be an additional correlate of voicing for C2? Given that C2 geminates were not fully voiced, the control on C1 duration could be thought of as a preplanning for the phonologically voiced geminates. Overall, our results are in line with previous studies that also found no significant effect of C2 duration on C1 duration (Punjabi: Hussain, 2015). However, our findings differ from other studies of Japanese (Kamiyama and Turco, 2017) and Italian (Turco and Braun, 2016) that reported differences in C1 duration before C2 singletons and geminates. These studies only investigated the effect of C2 duration on C1 duration. There are no detailed studies investigating the effect of C2 voicing on C1 duration.

C2 duration was consistently longer in geminates than in singletons as reported in previous studies of Japanese

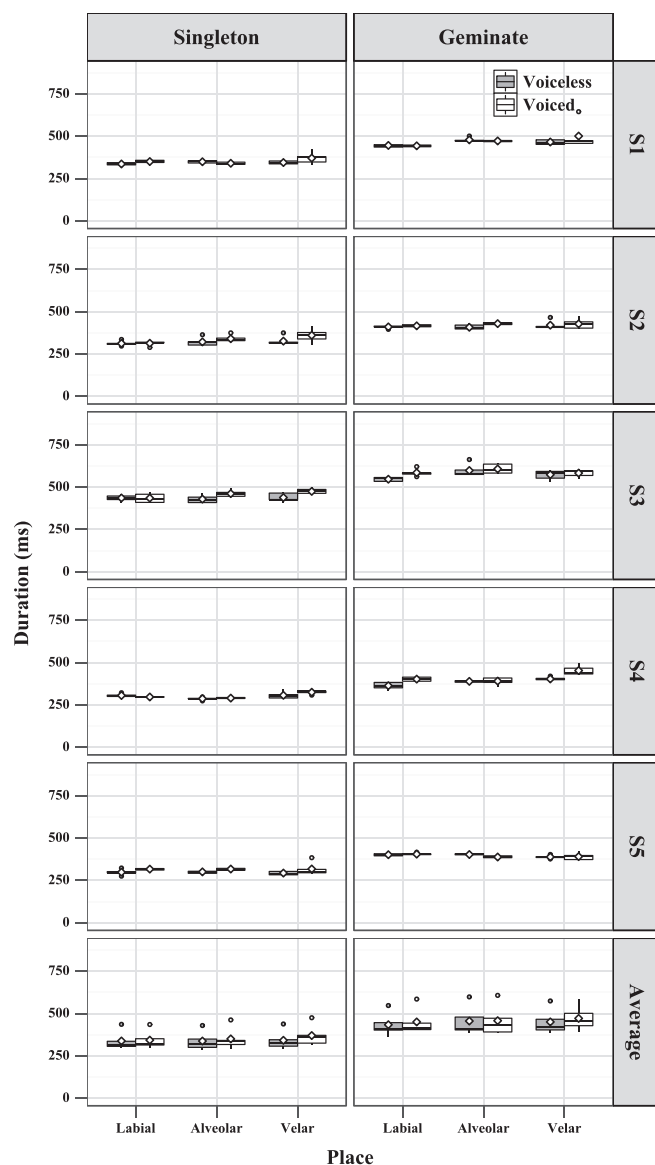


FIG. 6. Boxplots of whole word (C1 + V1 + C2 + V2) durations of the singleton (left column) and geminate (right column) stops of Tokyo Japanese.

(Campbell, 1999; Fujimoto and Maekawa, 2014; Han, 1994; Hirose and Ashby, 2007; Hirata and Whiton, 2005; Homma, 1981; Idemaru and Guion, 2008; Idemaru and Holt, 2007; Kawahara, 2015, 2016; among others). This pattern is also observed in a wide range of typologically unrelated languages (Arabic: Al-Tamimi and Khattab, 2015; Persian: Hansen, 2012; Punjabi: Hussain, 2015; Bengali and Turkish: Lahiri and Hankamer, 1988; Italian: Payne, 2005; Turco and Braun, 2016). It has also been shown that cross-linguistically voiced consonants have short durations compared to voiceless consonants (Ohala, 1983). In our data, voiceless and voiced singletons were distinct in terms of C2 duration. In contrast, C2 duration of the voiced geminates was similar to that of voiceless geminates (as also noted by Kawahara, 2005). These acoustic results imply neutralization between voiceless and voiced geminate stops.

The release burst duration of C2 was not significantly different in terms of voicing and length. Some recent studies have identified a loss of voicing contrast (voiced vs voiceless)

in word-initial singleton stops in Tokyo Japanese (Takada, 2012). The categorical division of VOT of word-initial voiceless vs voiced singleton stops is variably realized in Tokyo Japanese. In the current study, voiced singleton and geminate stops were voiced at the beginning of closure but devoiced at the offset of closure (near the release onset), which may have contributed to the lack of a significant difference in C2 release durations. Our results on the voicing contrast in word-medial singleton (and geminate) stops complement recent studies that report a collapse of the voicing distinction in word-initial singleton stops of Tokyo Japanese (Takada, 2012). The comparison of C2 closure voicing (%) in voiced singleton and geminate stops also indicated differences. Not only voiced geminates, but voiced singletons also showed devoicing. This was not surprising, since devoicing of voiced singletons was also noted by Kawahara (2006) and Hirose and Ashby (2007). However, the average percentage of voicing during closure reported in these studies was slightly higher (singletons: 87%–100%; geminates: 30%–47%) than in the current study (singletons: 70%–80%; geminates: 20%–25%).

Vowel duration was longer before geminates than before singletons, as reported by several studies on Japanese (Campbell, 1999; Fujimoto and Maekawa, 2014; Han, 1994; Hirose and Ashby, 2007; Kawahara, 2006). The articulatory patterns noted by Fujimoto *et al.* (2015) correspond to these acoustic results: the tongue movement for the production of an upcoming geminate consonant /kk/ was much slower during preceding vowel than for a singleton /k/. Hussain (2015) discussed various languages that differ in the phonetic realization of preceding vowels (V1): there are languages that shorten vowels before a geminate consonant (Italian: Esposito and Benedetto, 1999; Tashlhiyt Berber: Ridouane, 2007), while others show no difference in vowel durations before singletons and geminates (Punjabi: Hussain, 2015). Vowels in a closed syllable are known to shorten their duration (Maddieson, 1985). When the first half of the geminate closes the syllable formed with the preceding vowel, it is natural to suppose that syllable isochrony regulates the segmental duration. Japanese is a rare example where vowels before geminates are lengthened. This might be associated with moraic timing. Lengthening of preceding vowels is also a well-attested acoustic correlate of voicing in a wide range of languages. Vowels preceding voiced stops are longer than those before voiceless stops (Beguš, 2017; Lisker, 1974). In German, where the voicing in syllable final stops is banned, the duration of the preceding vowel is adjusted to the extent that it compensates for the loss of voicing (Braunschweiler, 1997). Although these voicing correlates in word-medial geminate (and singleton) stops in Tokyo Japanese are not robust, duration of the vowel preceding the geminate provided consistent differentiation of the voicing contrast.

There are a few studies on Japanese reporting vowel durations (V2) following the singleton and geminate stops. In the current study, the duration of vowels following geminates was consistently shorter than for their singleton counterparts, as found in previous studies (Homma, 1981; Idemaru and Guion, 2008). The duration of the following vowel may differ at varying speech rates. As noted by Hirata (2007), it was shorter after geminates in a slow speaking rate, but this difference was not

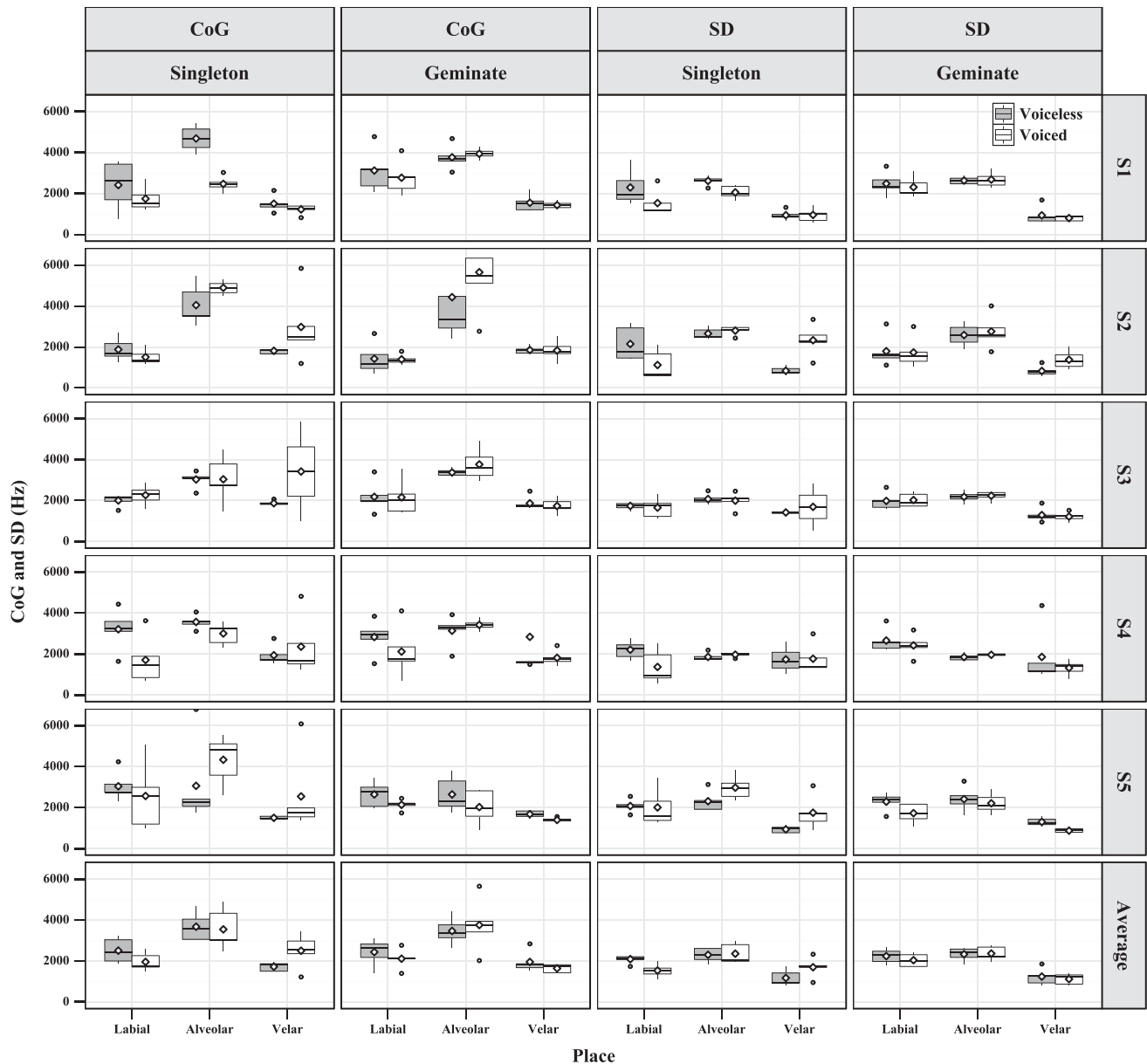


FIG. 7. Boxplots of spectral CoG (left two columns) and spectral SD (right two columns) of C2 stop release bursts in voiceless and voiced singleton and geminate stops of Tokyo Japanese.

significant in faster rates, or when segments were varied. With regard to voicing difference, duration of the vowel following the voiceless and voiced singletons indicated a significant difference in our data; however, no such durational differences were found for voiceless and voiced geminates. These results also point toward partial neutralization of the contrast between voiceless and voiced geminates.

Whole word duration was consistently longer in words with geminates than in singletons. Whole word duration was affected by voicing but only in the singleton context (Table III). The fact that no significant whole word duration difference was found for the geminate pairs further suggests that voiceless and voiced geminate stops are partially neutralized. We expected that the longer duration of voiceless stop segments in these words would be compensated by other segments at the word level to keep word duration of the same mora count constant (Sagisaka and Tohkura, 1984; among others). However, it

could be the case that compensation occurs at the phrase level (Port *et al.*, 1987). Kondo and Shinohara (2003) showed that a phrase with the same number of moras had approximately the same duration despite differences in segmental durations. It should also be noted that the duration of whole words or parts of words might vary depending on the type of word (real or nonsense). Some studies have shown that real words are produced faster than nonsense words. In the current study, some words might have been identified by the speakers as real words (Maxwell *et al.*, 2015).

The results of spectral moments of C2 stop release bursts showed no statistically significant differences in voicing and length contrasts of stops in Tokyo Japanese. Voiced singleton stops are generally characterized by lower spectral CoG than voiceless singletons (Chodroff and Wilson, 2014). Studies have shown that languages might differ in exploiting spectral SD as a cue to the characterization of voiceless and

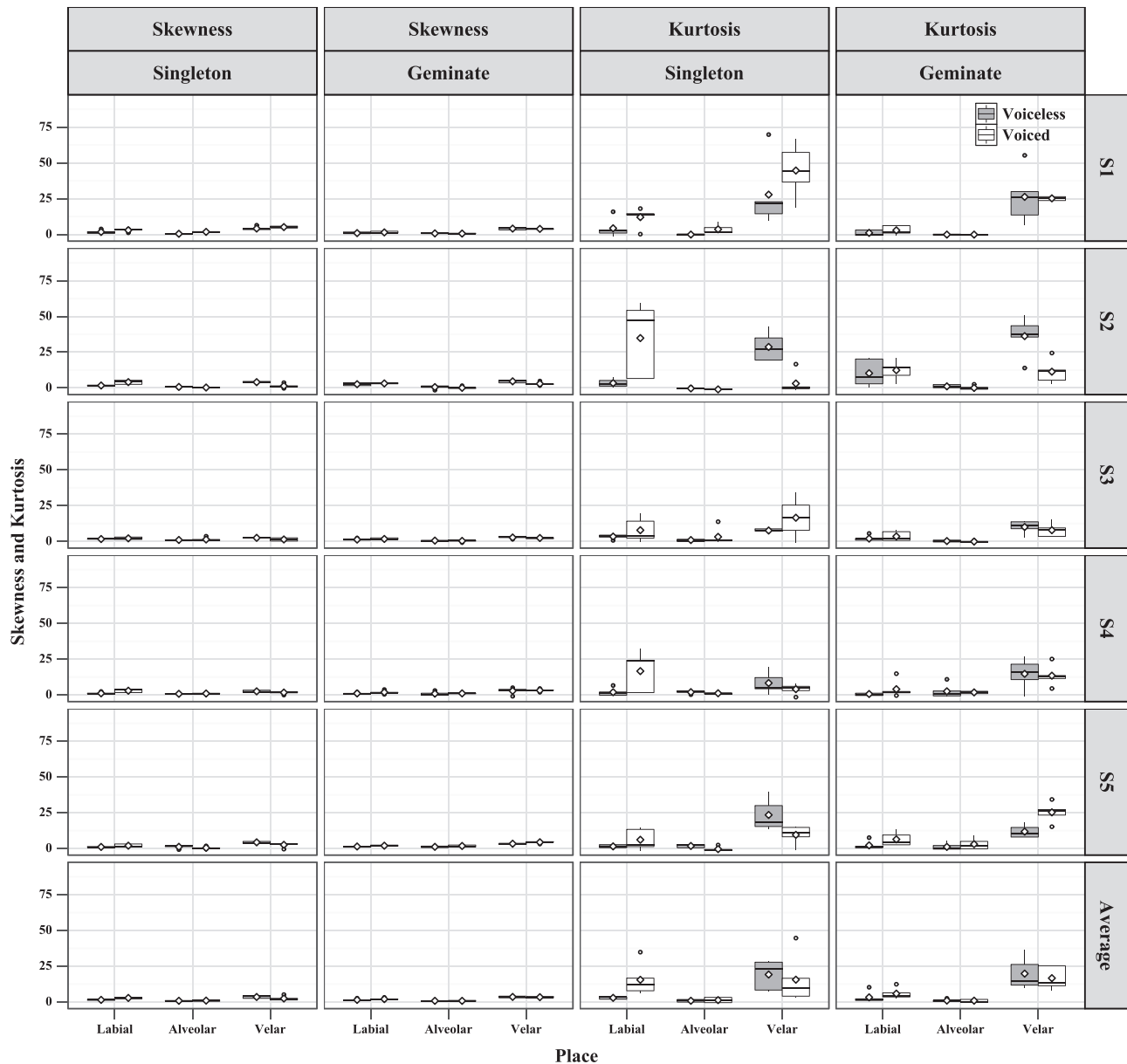


FIG. 8. Boxplots of spectral skewness (left two columns) and spectral kurtosis (right two columns) of C2 stop release bursts in voiceless and voiced singleton and geminate stops of Tokyo Japanese. Both spectral measures are unitless.

TABLE IV. Results of LME models of spectral CoG (Hz), spectral SD (Hz), spectral skewness (unitless), and spectral kurtosis (unitless), with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), Voicing (voiceless and voiced), and Length (singleton and geminate) as fixed factors. Asterisks denote the level of statistical significance ($p < 0.05^*$; $p < 0.01^{**}$; $p < 0.001^{***}$) (footnote 5).

Factors	df	Spectral moments							
		CoG		SD		Skewness		Kurtosis	
		<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Place (P)	2	8.64	=0.057	6.87	=0.076	12.94	*	5.67	=0.096
Voicing (V)	1	0.10	=0.773	0.27	=0.629	0.77	=0.430	0.92	=0.393
Length (L)	1	0.14	=0.728	0.04	=0.846	0.00	=0.972	0.42	=0.550
P × V	2	1.76	=0.174	7.28	***	12.17	***	9.21	***
P × L	2	1.15	=0.318	8.10	***	4.89	**	2.85	=0.059
V × L	1	0.12	=0.727	0.49	=0.486	0.32	=0.574	2.85	=0.092
P × V × L	2	2.20	=0.113	4.46	*	2.83	=0.061	2.44	=0.089

voiced stops (Sundara, 2005). In the current study, voiceless geminates were characterized by higher spectral SD than voiced geminates, but there were no statistically significant differences. No consistent patterns in spectral skewness or spectral kurtosis were observed across voicing and length. It was noted that both voiced singleton and voiced geminate stops showed 20%–80% voicing during closure duration (Fig. 4). As spectral moments were measured from the release of all stops (close to the end of the closure interval), the release bursts of devoiced singleton and geminate stops may have resulted in non-significant differences in the spectral energy and shape. Our results on spectral moments are in line with Al-Tamimi and Khattab (2015), who reported no clear differences in the spectral moments of singleton and geminate fricatives in Lebanese Arabic.

The current study investigated both temporal and spectral characteristics of voiceless and voiced singleton and

TABLE V. Results of separate LME models in singleton and geminate contexts, with Speaker (five speakers) as a random factor, Place (labial, alveolar, and velar), and Voicing (voiceless and voiced) as fixed factors. Asterisks denote the level of statistical significance ($p < 0.05^*$; $p < 0.01^{**}$; $p < 0.001^{***}$) (footnote 6).

Consonant	Factors	df	Spectral moments							
			CoG		SD		Skewness		Kurtosis	
			<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Singleton	Place (P)	2	7.10	=0.074	4.28	=0.134	8.17	=0.062	4.60	=0.123
	Voicing (V)	1	0.00	=0.958	0.00	=0.993	0.37	=0.574	1.76	=0.256
	P × V	2	2.34	=0.100	8.46	***	10.57	***	6.20	**
Geminate	Place (P)	2	8.37	=0.059	17.20	*	30.69	*	10.99	*
	Voicing (V)	1	0.28	=0.627	0.89	=0.399	0.22	=0.662	0.01	=0.926
	P × V	2	1.63	=0.200	0.85	=0.432	2.06	=0.132	2.20	=0.115

geminate stops of Tokyo Japanese. While the status of voiced geminates is marginal in Tokyo Japanese, and partial neutralization with voiceless geminates has been reported (Fujimoto and Kataoka, 2016; Kawahara, 2006, 2015, 2016), we observed that some acoustic attributes, particularly vowel durations preceding the geminates, still differentiate them. This suggests that the contrast between voiceless and voiced geminate stops in Tokyo Japanese is not completely neutralized. Thus, the voicing contrast in geminates can be maintained, and classified as a marginal or intermediate contrast appearing only in recently integrated loanwords in Tokyo Japanese. The results of this study can be used to further explore the articulatory and perceptive correlates of the acoustic measures presented here (e.g., duration or spectral moments). Our results also invite studies beyond those on Tokyo Japanese. Some Japanese dialects, as well as Ryukyuan dialects, have contrastive voiced geminates in word-initial and medial positions. Depending on the phonological structure of the dialect, the use of acoustic cues and articulatory strategies accompanying them might be distinct. Thus, the results of this study can be used to investigate cross-dialectal differences in Japanese, both in terms of perception and production. We used a small but very controlled dataset. Other studies could use larger datasets to show differences in

length and voicing of stops in Tokyo and other Japanese dialects. This study may also stimulate more studies on marginal contrasts. A structure introduced by any phonological alternation or loanword adaptation might conflict with markedness. Production patterns in this situation may give insight on how the phonological structure of a language evolves.

ACKNOWLEDGMENTS

We thank the Associate Editor, two anonymous reviewers, Jeff Mielke, and Jeff Moore for their valuable comments and suggestions. We are also grateful to all the participants, Tomohiko Ooigawa, and Toshimi Ogawa for help with experiments, and Masako Fujimoto for comments on the data. The remaining errors are our own. The study was supported by an Endeavour Postgraduate Scholarship (Australian Government, Department of Education and Training) to author Q.H. and by the Sophia Linguistic Institute for International Communication.

APPENDIX

Table VI shows the absolute mean values of all the temporal and spectral measures.

TABLE VI. Mean (SD) durations (ms) of C1 (Clo1 + release), V1, C2 (Clo2 + release), C2 release alone, V2, WW (whole word) durations, and spectral moments [CoG (Hz), SD (Hz), skewness, and kurtosis] of C2 stop release bursts across places of articulation (labial, alveolar, and velar), voicing (voiceless and voiced), and length (singleton and geminate).

Place	± Voice	Length	Temporal measures						Spectral moments			
			C1	V1	C2	C2 rel	V2	WW	CoG	SD	Skewness	Kurtosis
Labial	–	/p/	117 (28)	36 (12)	99 (16)	17 (5)	87 (12)	339 (57)	2508 (599)	2093 (215)	1.48 (0.32)	2.90 (1.16)
	–	/pp/	115 (20)	56 (15)	183 (40)	15 (4)	81 (17)	434 (70)	2439 (662)	2240 (354)	1.48 (0.67)	3.25 (3.99)
	+	/b/	120 (28)	54 (7)	64 (9)	14 (6)	105 (29)	342 (56)	1957 (440)	1536 (330)	2.83 (0.84)	15.51 (11.55)
	+	/bb/	121 (29)	73 (13)	172 (34)	13 (4)	84 (16)	450 (77)	2115 (486)	2042 (320)	2.03 (0.66)	5.85 (3.89)
Alveolar	–	/t/	114 (15)	42 (13)	90 (18)	14 (2)	91 (20)	337 (57)	3677 (702)	2303 (344)	0.84 (0.28)	0.87 (1.04)
	–	/tt/	125 (19)	60 (13)	191 (55)	14 (2)	79 (14)	455 (87)	3470 (676)	2332 (332)	0.81 (0.23)	1.00 (1.02)
	+	/d/	137 (34)	57 (12)	53 (8)	15 (5)	103 (20)	350 (66)	3547 (1018)	2359 (487)	0.99 (0.62)	1.32 (2.08)
	+	/dd/	127 (25)	71 (14)	179 (60)	13 (2)	80 (11)	458 (90)	3760 (1299)	2368 (339)	0.80 (0.65)	0.93 (1.42)
Velar	–	/k/	110 (18)	52 (12)	96 (24)	26 (7)	84 (21)	342 (57)	1724 (210)	1173 (387)	3.51 (0.89)	19.18 (10.40)
	–	/kk/	116 (13)	71 (14)	189 (47)	22 (6)	74 (13)	450 (75)	1956 (511)	1241 (403)	3.54 (0.80)	19.83 (11.12)
	+	/g/	144 (37)	66 (9)	59 (20)	14 (4)	100 (21)	370 (63)	2504 (825)	1698 (491)	2.44 (1.59)	15.59 (17.12)
	+	/gg/	135 (24)	76 (12)	180 (37)	22 (5)	80 (14)	471 (74)	1641 (213)	1117 (266)	3.37 (0.77)	16.65 (8.10)

- ¹There are five vowels in Tokyo Japanese (/i e a o u/). The high vowels /i/ and /u/ are reported to induce regular devoicing before another voiceless obstruent. /o/ can also optionally be devoiced such as in the word /kokoro/ “heart,” which is quite often pronounced with the devoicing of the first /o/ [see Vance (1987)]. The remaining two vowels /e a/ will not make an ideal pair to compare the acoustic correlates of voicing and length. Thus, a low vowel /a/, which is the least devoiced vowel in Tokyo Japanese, was used as V1 and V2. There was only one token in the current study where /a/ was devoiced (see Sec. II D) (Vance, 1987).
- ²LME model: (C1/V1/C2/C2 release/V2/WW~Place*Voicing*Length+(1+Place+Voicing+Length|Sp)).
- ³LME model: Singleton/geminate (C1/V1/C2/C2 release/V2/WW~Place*Voicing+(1+Place+Voicing|Sp)).
- ⁴See supplementary material at <https://doi.org/10.1121/1.5078605> for representative examples of partial devoicing in voiced singletons and geminates.
- ⁵LME model: (CoG/SD/skewness/kurtosis~Place*Voicing*Length+(1+Place+Voicing+Length|Sp)).
- ⁶LME model: Singleton/geminate (CoG/SD/skewness/kurtosis~Place*Voicing+(1+Place+Voicing|Sp)).
- Al-Tamimi, J., and Khattab, G. (2015). “Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants,” *J. Acoust. Soc. Am.* **138**, 344–360.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., and Singmann, H. (2015). “lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-15,” available at <https://CRAN.Rproject.org/package=lme4> (Last viewed March 9, 2018).
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *J. Mem. Lang.* **68**, 255–278.
- Beckman, M. (1982). “Segmental duration and the ‘mora’ in Japanese,” *Phonetica* **39**, 113–135.
- Beguš, G. (2017). “Effects of ejective stops on preceding vowel duration,” *J. Acoust. Soc. Am.* **142**, 2168–2184.
- Bhatia, T. K. (1993). *Punjabi: A Cognitive-Descriptive Grammar* (Routledge, London), pp. 1–423.
- Blevins, J. (2008). “Explaining diversity in geminate consonant inventories: An evolutionary approach,” presentation at Leipzig Spring School on Linguistic Diversity, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Available at https://www.eva.mpg.de/lingua/conference/08_springschool/pdf/course_materials/blevins_evening_lecture.pdf (Last viewed February 28, 2017).
- Boersma, P., and Weenink, D. (2014). “Praat: Doing phonetics by computer (version 6.0.17) [computer program],” <http://www.praat.org/> (Last viewed August 5, 2017).
- Braunschweiler, N. (1997). “Integrated cues of voicing and vowel length in German: A production study,” *Lang. Speech* **40**, 353–376.
- Bundgaard-Nielsen, R. L., Kroos, C., Baker, B. J., Best, C. T., and Harvey, M. (2016). “Consonantal timing and release burst acoustics distinguish multiple coronal stop place distinctions in Wubuy (Australia),” *J. Acoust. Soc. Am.* **140**, 2794–2809.
- Campbell, N. (1999). “A study of Japanese speech timing from the syllable perspective,” *J. Phon. Soc. Jpn.* **3**, 29–39.
- Chen, M. (1970). “Vowel length variation as a function of the voicing of the consonant environment,” *Phonetica* **22**, 129–159.
- Chodroff, E., and Wilson, C. (2014). “Burst spectrum as a cue for the stop voicing contrast in American English,” *J. Acoust. Soc. Am.* **136**, 2762–2772.
- Esposito, A., and Benedetto, M. G. D. (1999). “Acoustical and perceptual study of gemination in Italian stops,” *J. Acoust. Soc. Am.* **106**, 2051–2062.
- Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. M. (1988). “Statistical analysis of word-initial voiceless obstruents: Preliminary data,” *J. Acoust. Soc. Am.* **84**, 115–123.
- Fujimoto, M., Funatsu, S., and Hoole, P. (2015). “Articulation of single and geminate consonants and its relation to the duration of the preceding vowel in Japanese,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by the Scottish Consortium for ICPhS 2015, University of Glasgow, UK, Paper No. 0070, pp. 1–5, available at <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPhS0070.pdf> (Last viewed August 14, 2017).
- Fujimoto, M., and Kataoka, R. (2016). “Oral/nasal airflow during Japanese stop consonants,” *J. Acoust. Soc. Am.* **140**, 3108.
- Fujimoto, M., and Maekawa, K. (2014). “Effects of sokuon on adjacent vowel duration: An analysis of the corpus of spontaneous Japanese,” *J. Phon. Soc. Jpn.* **18**, 10–22.
- Hall, K. C. (2013). “A typology of intermediate phonological relationships,” *Linguist. Rev.* **30**, 215–275.
- Ham, W. H. (2001). *Phonetic and Phonological Aspects of Geminate Timing* (Routledge, New York), pp. 1–275.
- Han, M. S. (1962). “The feature of duration in Japanese,” *Onsei no Kenkyuu* **10**, 65–80.
- Han, M. (1994). “Acoustic manifestations of mora timing in Japanese,” *J. Acoust. Soc. Am.* **96**, 73–82.
- Hansen, B. B. (2012). “The perceptibility of duration in the phonetics and phonology of contrastive consonant length,” Doctoral dissertation, University of Texas at Austin, pp. 1–116.
- Homma, Y. (1981). “Durational relationship between Japanese stops and vowels,” *J. Phon.* **9**, 273–281.
- Hirata, Y. (2007). “Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels,” *J. Phon. Soc. Jpn.* **11**, 9–22.
- Hirata, Y., and Amano, S. (2012). “Production of single and geminate stops in Japanese three- and four-mora words,” *J. Acoust. Soc. Am.* **132**, 1614–1625.
- Hirata, Y., and Whiton, J. (2005). “Effects of speech rate on the singleton/geminate distinction in Japanese,” *J. Acoust. Soc. Am.* **118**, 1647–1660.
- Hirose, A., and Ashby, M. (2007). “An acoustic study of devoicing of the geminate obstruents in Japanese,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 909–912.
- Hussain, Q. (2015). “Temporal characteristics of Punjabi word-medial singletons and geminates,” *J. Acoust. Soc. Am.* **138**, EL388–EL391.
- Hussain, Q. (2017). “Phonetic characterization of a complex coronal system: Insights from Punjabi,” Doctoral dissertation, Macquarie University, Sydney, Australia, pp. 1–188.
- Hussain, Q., Proctor, M., Harvey, M., and Demuth, K. (2017). “Acoustic characteristics of Punjabi retroflex and dental stops,” *J. Acoust. Soc. Am.* **141**, 4522–4542.
- Idemaru, K., and Guion, S. (2008). “Acoustic covariants of length contrast in Japanese stops,” *J. Int. Phon. Assoc.* **38**, 167–186.
- Idemaru, K., and Holt, L. L. (2007). “Relational timing or absolute duration? Cue weighting in the perception of Japanese singleton-geminate stops,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 753–756.
- Ito, J., Kubozono, H., and Mester, A. (2017). “A prosodic account of consonant gemination in Japanese loanwords,” in *The Phonetics and Phonology of Geminate Consonants*, edited by H. Kubozono (Oxford University Press, Oxford), pp. 283–320.
- Ito, J., and Mester, A. (1995). “Japanese phonology,” in *The Handbook of Phonological Theory*, edited by J. Goldsmith (Blackwell, Oxford), pp. 817–838.
- Jongman, A., Blumstein, S. E., and Lahiri, A. (1985). “Acoustic properties for dental and alveolar stop consonants: A cross-language study,” *J. Phon.* **13**, 235–251.
- Jongman, A., Wayland, R., and Wong, S. (2000). “Acoustic characteristics of English fricatives,” *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kamiyama, T., and Turco, G. (2017). “Non-local temporal adjustments caused by length contrasts: The case of Japanese,” *Paper presented at the Phonetics and Phonology in Europe*, 12–14 June, Cologne, Germany.
- Katayama, M. (1997). “The markedness of voiced geminates in Japanese,” *Phonology* **15**, 49–56.
- Kawagoe, I. (1995). “Shakuyōgo ni miru sokuonka to rizumu shōtotsu” (“Gemination and rhythm clash in loanwords”), *Gengo Kenkyū* **1198**, 46–73.
- Kawahara, S. (2005). “Voicing and geminacy in Japanese: An acoustic and perceptual study,” *Univ. Mass. Occas. Pap. Linguist.* **31**, 87–120, available at http://www.rci.rutgers.edu/~phonetic/pdf/kawahara_UMOP31.pdf (Last viewed August 8, 2018).
- Kawahara, S. (2006). “A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese,” *Lang.* **82**, 536–574.
- Kawahara, S. (2015). “Geminate devoicing in Japanese loanwords: Theoretical and experimental investigations,” *Lang. Ling. Comp.* **9**, 181–195.
- Kawahara, S. (2016). “Japanese geminate devoicing once again: Insights from information theory,” in *Proceedings of Formal Approaches to Japanese Linguistics*, Vol. 8, pp. 43–62, available at <http://mitwpl.mit.edu/catalog/mwpl79/> (Last viewed July 7, 2016).

- Kawahara, S., and Matsui, F. M. (2017). "Some aspects of Japanese consonant articulation: A preliminary EPG study," *ICU Work. Pap. Linguist.* **II**, 9–20.
- Kirchner, R. (1998). "An effort-based approach to consonant lenition," Doctoral dissertation, University of California, Los Angeles, CA, pp. 1–340.
- Kochetov, A., and Kang, Y. (2017). "Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean," *Can. J. Linguist.* **62**, 18–55.
- Kondo, M., and Shinohara, S. (2003). "Durational evidence of psychological reality of the mora in Japanese speakers' French," in *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 463–466.
- Kubozono, H. (1993). *The Organization of Japanese Prosody* (Kurosio, Tokyo), pp. 1–302.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). "lmerTest package: Tests in linear mixed effects models," R package version 2.0-36, pp. 1–26, available at [10.18637/jss.v082.i13](https://cran.r-project.org/web/packages/lmerTest/index.html) (Last viewed March 9, 2018).
- Lahiri, A., and Hankamer, H. (1988). "The timing of geminate consonants," *J. Phon.* **16**, 327–338.
- Li, F., Edwards, J., and Beckman, M. E. (2009). "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," *J. Phon.* **37**, 111–124.
- Lisker, L. (1974). "On 'explaining' vowel duration variation," *Glossa* **8**, 223–246.
- Löfqvist, A. (2007). "Tongue movement kinematics in long and short Japanese consonants," *J. Acoust. Soc. Am.* **122**, 512–518.
- Lovins, J. B. (1975). "Loanwords and the phonological structure," Doctoral dissertation, University of Chicago, distributed by the Indiana University Linguistics Club, pp. 1–177.
- Maddieson, I. (1985). "Phonetic cues to syllabification," in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, Orlando), pp. 203–221.
- Maxwell, O., Baker, B., Bundgaard-Nielsen, R. L., and Fletcher, J. (2015). "A comparison of the acoustics of nonsense and real word stimuli: Coronal stops in Bengali," in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by the Scottish Consortium for ICPhS 2015, University of Glasgow, UK, Paper No. 900, pp. 1–5.
- McCawley, J. D. (1968). *The Phonological Component of a Grammar of Japanese* (Mouton, The Hague), pp. 1–206.
- Nittrouer, S. (1995). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *J. Acoust. Soc. Am.* **97**, 520–530.
- Ohala, J. J. (1983). "The origin of sound patterns in vocal tract constraints," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 189–216.
- Ohso, M. (1971). "A phonological study of some English loan words in Japanese," M.A. thesis, The Ohio State University, pp. 1–44.
- Payne, E. M. (2005). "Phonetic variation in Italian consonant gemination," *J. Int. Phon. Assoc.* **35**, 153–181.
- Pierrehumbert, J. B., and Beckman, M. E. (1988). *Japanese Tone Structure* (MIT Press, Cambridge, MA), pp. 1–282.
- Port, R. F., Dalby, J., and O'Dell, M. (1987). "Evidence for mora timing in Japanese," *J. Acoust. Soc. Am.* **81**, 1574–1585.
- R Core Team (2012). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, available at <https://cran.r-project.org/bin/windows/base/old/> (Last viewed August 5, 2017).
- Ridouane, R. (2007). "Gemination in Tashlhiyt Berber: An acoustic and articulatory study," *J. Int. Phon. Assoc.* **37**, 119–142.
- Riney, T. J., Takagi, N., Ota, K., and Uchida, Y. (2007). "The intermediate degree of VOT in Japanese initial voiceless stop," *J. Phon.* **35**, 439–443.
- Sagisaka, Y., and Tohkura, Y. (1984). "Phoneme duration control for speech synthesis by rule," *J. Electron. Commun. Eng.* **67-A**, 629–636 (in Japanese).
- Shinohara, S., and Fujimoto, M. (2018). "Acoustic characteristics of the obstruent and nasal geminates in the Ikema dialect of Miyako Ryukyuan," in *Cross-Linguistic Research in Monolingual and Bilingual Speech*, edited by E. Babatsouli, ISMBS, Crete, Greece, pp. 253–270.
- Stevens, M., and Hajek, J. (2004). "Comparing voiced and voiceless geminates in Siamese Italian: What role does preaspiration play?," in *Proceedings of the 10th Australian International Conference on Speech Science and Technology*, pp. 340–345.
- Sundara, M. (2005). "Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French," *J. Acoust. Soc. Am.* **118**, 1026–1037.
- Tabain, M. (2012). "Jaw movement and coronal stop spectra in Central Arrernte," *J. Phon.* **40**, 551–567.
- Tabain, M., Breen, G., Butcher, A., Jukes, A., and Beare, R. (2016). "Stress effects on stop bursts in five languages," *Lab. Phon. J. Assoc. Lab. Phon.* **7**, 1–23.
- Tabain, M., and Butcher, A. (2015). "Stop bursts in Pitjantjatjara," *J. Int. Phon. Assoc.* **45**, 149–176.
- Tabain, M., and Jukes, A. (2016). "Makasar (illustration of the IPA)," *J. Int. Phon. Assoc.* **46**, 99–111.
- Takada, M. (2011). *Nihongo no Goto-heisaon no Kenkyu: VOT no Kyoziteki Bunpu to Tuziteki Henka (Research on the Word-initial Stops of Japanese: Synchronic Distribution and Diachronic Change in VOT)* (Kuroshio, Tokyo), pp. 1–272.
- Takada, M. (2012). "Regional and generational variation of VOT in Japanese," in *Papers from the First International Conference on Asian Geolinguistics*, pp. 273–282.
- Thurgood, G. (1993). "Geminates: A cross-linguistic examination," in *Papers in Honor of Frederick H. Brengelman on the Occasion of the Twenty-Fifth Anniversary of the Department of Linguistics (CSU) Fresno*, edited by J. A. Nevis, G. McMenamin, and G. Thurgood (Department of Linguistics, California State University, Fresno, CA), pp. 129–139.
- Tsuchida, A. (1995). "English loans in Japanese: Constraints in loanword phonology," *Work. Pap. Cornell Phon. Lab.* **10**, 145–164.
- Turco, G., and Braun, B. (2016). "An acoustic study on non-local anticipatory effects of Italian length contrast," *J. Acoust. Soc. Am.* **140**, 2247–2256.
- Vance, T. (1987). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY), pp. 1–226.