音声研究　第 29 巻　87–102 頁
2025年

Journal of the Phonetic Society of Japan
Vol. 29, pp. 87–102, 2025

特集論文

Differences in Voicing Implementation and Tongue Movements between Japanese Intervocalic Voiced and Voiceless Stops

Weiyu LI*, Ai MIZOGUCHI**, Maho MORIMOTO*** and Takayuki ARAI*

有声実現と舌調音動作から見る日本語の語中有声・無声破裂音の違い

SUMMARY: This study investigated the voicing implementation of Japanese intervocalic stops using ultrasound and acoustic analysis. Acoustic analysis revealed that Japanese voiced stops exhibited a high voicing ratio during closure and a relatively short release burst interval, along with a high occurrence rate of consonant weakening. Japanese voiceless stops showed partial voicing during closure due to passive voicing from the preceding vowel. Articulatory analysis indicated that the tongue position at the maximal constriction was higher for voiceless stops than for voiced stops. Active sustainment of voicing was observed for voiced stops, and consonant weakening may also contribute to the sustainment.

Key words: Japanese stops, tongue position, voicing contrast, ultrasound imaging, intervocalic

1. Introduction

Voice onset time (VOT) has long been considered an effective indicator of the contrast between voiced and voiceless stops (Lisker and Abramson 1964). Japanese stop consonants have voiced and voiceless distinctions, and their contrast is primarily determined by the presence or absence of the vocal fold vibration during closure (Koizumi 2003). This implies that Japanese is considered a true voice language. However, recent studies have revealed that word-initial voiced stops in Japanese are typically devoiced and realized with a positive VOT, especially among younger generations (Tokyo Japanese: Gao and Arai 2019; Kinki Japanese and Tohoku Japanese: Takada 2011). In other languages with two-way contrast in stops, there are so-called aspirating languages that contrast voiceless unaspirated and voiceless aspirated sounds, such as Cantonese (Lisker and Abramson 1964). In aspirating languages, stop consonants are distinguished by short-lag positive VOT and long-lag positive VOT.

1.1 Voicing in Intervocalic Stops

Regarding stops in the intervocalic position, the implementation of voicing can be more complicated, depending on the phonological and prosodic contexts.

VOT alone cannot capture the detailed characteristics of voicing. Davidson (2016, 2018) identified four patterns of voicing implementation during the closure interval of intervocalic stops in English: “bleed”, where voicing continues from the preceding vowel; “hump”, where voicing occurs in the middle of the closure; “trough”, where voicing discontinues in the middle of the closure and reappears before the release; and negative VOT, which refers to phonation beginning partway through the closure and extending after the burst. The proportion of voicing at the end of the closure interval is significantly higher for voiced stops than for voiceless stops. Except for the negative VOT pattern, VOT cannot fully explain the other forms of voicing implementation.

Other attempts to explain the voicing implementation of intervocalic stops have examined whether stops in intervocalic environments undergo passive voicing, because of the voiced context instead of active voicing gestures by speakers in aspirating and true voice languages (Beckman, Jessen and Ringen 2013). Beckman, Jessen and Ringen (2013) suggested that the occurrence of passive voicing can be accounted for by positing privative features of [voice] or [spread glottis] (henceforth [sg]) instead of the binary voicing feature ([±voice]). They further assumed that a numerical

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

value is assigned prior to the phonetics, where a relatively high value is given for a specified feature (such as [9sg]), and a relatively low value for an unspecified feature (such as [1sg] or [2sg]). Within this framework, they attempted to characterize the laryngeal contrasts in true voice languages and aspirating languages. For example, in their explanation, in true voice languages such as Russian, where stops contrast [9voice] and [1voice], voiced sounds with [9voice] actively undergo voicing, whereas voiceless sounds specified as [1voice] should actively exclude passive voicing (because phonetic processes cannot change a numerically specified phonological feature). In aspirating languages, such as Icelandic, where contrasts exist between [9sg] and [5sg], even the stops specified as unaspirated also possess sufficiently large [5sg] to prevent phonetic passive voicing because the glottal spreading would be too great. Stops in aspirating languages, such as German or English, which contrast between [9sg] and [1sg] but are not specified in [voice] (unlike the traditional view), are more susceptible to passive voicing in unaspirated stops because of their lower [sg] compared to Icelandic unaspirated stops. Beckman, Jessen and Ringen (2013)'s theory explained why languages such as English and German do not exhibit voicing in word-initial positions but allow passive voicing in word-medial positions, which cannot be explained well by the binary categoricity of voicing.

The current study investigates the voicing implementation of Japanese intervocalic stops, applying the framework proposed by Beckman, Jessen and Ringen (2013). Additionally, Japanese stops have been shown to weaken into voiced fricatives in certain cases (Maekawa 2018) (velar stops also weaken into velar nasals, which is called "Bidakuon" in Japanese). Thus, this study also investigates how consonant weakening is related to the voicing of Japanese stops.

1.2 Articulatory Movements of Voiced and Voiceless Stops

When producing voiced sounds, transglottal airflow is required to pass through the glottis for the vocal folds to vibrate. However, when a closure is formed in the oral cavity, the air pressure in the supraglottal cavity becomes too high, thus reducing the airflow through the glottis. In this case, maintaining vocal fold vibration is difficult, which makes the production of voiced sounds challenging (Ohala 1983). To avoid this situation, speakers adapt various strategies. For example, by advancing the tongue root, lowering the tongue body or lowering the larynx, they can expand the space in the oral and pharyngeal cavities, thus increasing airflow through the glottis and maintaining vocal fold vibrations (Kent and Moll 1969; Svirsky et al. 1997; Westbury 1983; Westbury and Keating 1986).

Using ultrasound imaging, Ahn (2018) investigated the tongue position immediately preceding release for word-initial stops in Portuguese and English. Portuguese, which is considered a true voice language, always has prevoicing in voiced stops. By contrast, word-initial voiced stops in English are typically associated with a positive voice onset time (+VOT). Nevertheless, tongue root advancement and tongue body lowering to maintain vocal fold vibration during closure were observed in both languages. Given these results, Ahn (2018) suggested that despite the absence of prevoicing in English, the observed tongue root advancement and tongue body lowering may facilitate potential prevoicing in voiced stops or promote aspiration in voiceless stops. Subsequent investigations by Ahn, Kwon and Faytak (2024) revealed consistent differences in tongue position between unaspirated and aspirated stops in aspiration languages, such as Mandarin Chinese, supporting the hypothesis of potential prevoicing facilitation.

As for Japanese, some studies showed that the articulatory gestures of stops differed between voiced and voiceless sounds. Using pressure-sensitive palatography, Wakumoto et al. (1998) demonstrated that the tongue-palatal contact pressure of /t/ in Japanese is greater than that of /d/. Kochetov (2014) used electropalatography (EPG) and found that voiceless sounds tended to have more contact between the tongue and palate than voiced sounds in Japanese stops. Although some speakers exhibited incomplete contact, the presence and extent of consonant weakening remained unclear, because acoustic analysis results were not provided. Fujimoto, Kitamura and Funatsu (2022) investigated the articulatory movements of /k/ and /ɡ/ in Japanese word-medial positions using real-time MRI. They observed that /k/ tended to have a larger closure length along the mid-sagittal plane than /ɡ/ in the maximal constriction frame. However, significant individual variations were observed in the pharyngeal area, and no consistent trend was observed for /k/ and /ɡ/. Additionally, they discovered that some voiced utterances were produced as fricatives and that one speaker produced all voiced utterances as fricatives while still showing the same tendency as mentioned above.

In the current study, we aim to elucidate the articulatory movements of Japanese intervocalic stops in conjunction with acoustic analysis. Therefore, we chose ultrasound imaging to capitalize on advantages such as relatively high temporal resolution compared to MRI and the ability to capture tongue movements that cannot be captured with EPG.

1.3 Aims of the Study

Based on the above, the research questions for this study are summarized as follows:

I. To elucidate the voicing implementation of Japanese intervocalic stops and the corresponding articulatory tongue movements.
II. To reveal the realization of consonant weakening in Japanese stops and its articulatory mechanisms.
III. To verify the results of Kochetov (2014) by ultrasound imaging and expand insights into the articulatory movements for the voiced-voiceless contrast in Japanese stops by comparing them with Ahn's (2018) data for English and Portuguese.

2. Methods

2.1 Participants

Ten native Japanese speakers without speech or hearing disorders participated in the experiment. The details of the speakers are listed in Table 1. "Region" refers to the place where participants had lived. All participants resided in the vicinity of Tokyo, Japan, at the time of the experiment. CJF02, who is proficient in Japanese, English, and Chinese, showed different tendencies in acoustic analysis compared with other Japanese speakers. And we found that the probe position shifted during the recording of CJF04 in articulatory analysis. Therefore, data from these two speakers were excluded from the analysis. In total, data from eight speakers were analyzed in this study.

Table 1 Participant details.

Speaker	Gender	Age	Region
CJF01	F	18	Kanagawa
CJF02	F	18	Tokyo/Beijing/ Maryland
CJF03	F	20	Fukushima/Tokyo
CJF04	F	19	Aichi/Chiba/ Melbourne
CJF05	F	21	Shizuoka/Tokyo
CJF06	F	22	Tokyo
CJF07	F	21	Kanagawa
CJM01	M	25	Tokyo
CJM02	M	20	Tokyo
CJM03	M	20	Tokyo/Singapore

2.2 Materials

The stimuli were "aCa" sequences displayed in hiragana, which were randomized and embedded in a carrier phrase (/korewa ___ to iːmasu/ "this is called ___"). C were /p, b, t, d, k, g/; however, in this study, we primarily focused on the results for alveolar and velar consonants, where tongue gestures are involved. Participants were prompted to produce these stimuli ten times each with high-low (HL) and low-high (LH) pitch accent patterns. The HL and LH patterns were recorded in two separate blocks. The order of the accent blocks was counterbalanced across speakers.

2.3 Instrumentation and Procedure

Real-time mid-sagittal images of the oral cavity were recorded using an ultrasound system (MicrUS, EXT-1H) equipped with a microconvex probe (MC10-5R10S-3). The probe was placed under the participant's chin, and an UltraFit Headset (Spreafico, Pucher and Matosova 2018) was used to stabilize the relative positions of the probe and the participant's head. The frame rate of the obtained ultrasound video was 113 fps. The audio was recorded at a sampling rate of 22,050 Hz with a 16-bit resolution using a microphone (RODE-NT2-A). The audio and ultrasound video signals were recorded and synchronized using AAA software (Articulate Instruments Ltd. 2022).

3. Analysis

The total number of trials was 640, with eight speakers, four consonants, two accent patterns, and ten repetitions each. Instances with synchronization errors that prevented us from exporting video files with audio were excluded from the analysis. Images that were too unclear for tongue contour tracking were also excluded from the analysis. The data for velar stops from two speakers (CJF01, CJM01) were excluded from the acoustic and articulatory analyses because of the unclear ultrasound images (depending on the size of the participant's oral cavity, the tongue moved to a higher position resulting in less clear images). A total of 535 utterances were analyzed.

3.1 Acoustic Analysis

For the acoustic analysis, we performed segmentation using Praat (Boersma and Weenink 2023). For with-burst utterances, we segmented the recordings and measured the durations of the preceding vowel (V1), voiced closure (VDCLO), voiceless closure (VLCLO), release/burst duration (REL), and following vowel

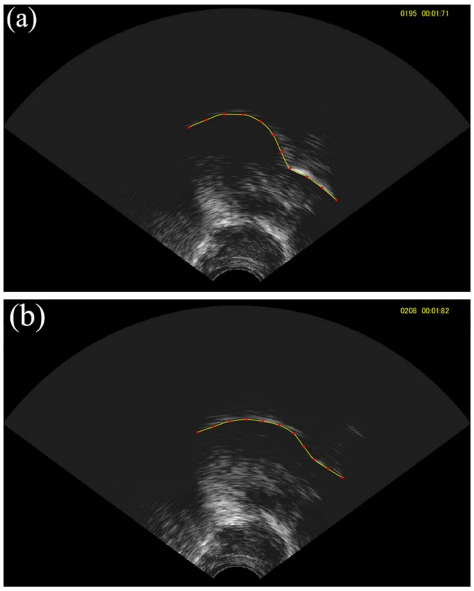

Figure 1 An example of tracking— (a): the onset of articulatory movements of /d/; (b): the maximal constriction point of /d/. This figure shows the mid-sagittal images of the tongue. The right side represents the anterior portion of the tongue, while the left represents the posterior portion.

(V2). Voicing was identified by visual inspection of the waveform and the spectrogram. A period during closure was labeled as VDCLO if a visible concentration of energy at a frequency close to the speaker's fundamental frequency was detected; otherwise, the period was labeled as VLCLO. For the without-burst (i.e., weakened) voiced stops, recordings were segmented into V1, consonant (C), and V2, and then the weakening pattern was classified. The details of the classification method are described in Section 4.1.2. When no pause was inserted between /wa/ in the carrier sentence and V1 /a/, which rendered it difficult to determine the onset of V1, a label "~V1" was assigned from the point where V1 was perceived auditorily to have begun. Data from these cases were not included in the V1 duration measurements.

3.2 Articulatory Analysis

For the articulatory analysis, we used GetContours (Tiede 2022) to track the tongue contours of two ultrasound image frames based on articulatory gestures. The first frame was the onset of articulatory movements of the stops, and the second was the maximal constriction point (the frame where the tongue ceases to rise). Figure 1 shows an example of tracking. The timing information of both frames was extracted and compared

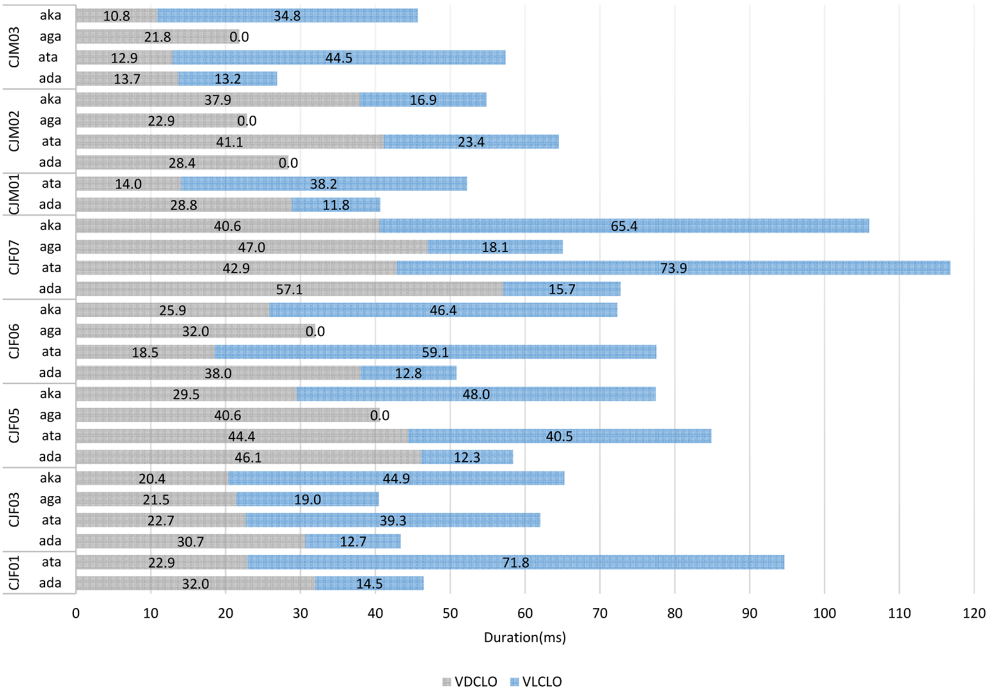

Figure 2 Mean of voiced closure duration and voiceless closure duration.

for voiceless stops and voiced stops with burst, along with the timing information of consonant onset and release in terms of acoustics.

Tongue contour data at the maximal constriction point were divided among speakers into voiceless stops, voiced stops with burst, and weakening patterns of voiced stops distinguished in the acoustic analysis. All tongue contour data for each speaker were modeled via a generalized additive model (GAM) (Wood 2006) using R (R Core Team 2021). Subsequently, GAM plots with 95% confidence intervals were generated using MATLAB (The MathWorks Inc. 2021) to compare voiceless stops against voiced stops with burst, and voiced stops with burst against weakened voiced stops. This process was conducted to investigate whether differences existed in the constriction caused by the tongue among voiceless stops, voiced stop consonants with burst, and weakened voiced stops.

4. Results

4.1 Acoustic Results

The closure voicing rate and duration of release burst intervals were compared between voiceless stops (all with burst) and voiced stops with burst, and the findings are summarized in Section 4.1.1. In Section 4.1.2, we classify the weakening patterns of voiced stops without burst based on spectrograms, waveforms, and perceptual impressions.

4.1.1 Voicing Rate during Closure and Release Burst Duration

Figure 2 shows the mean voiced and voiceless closure durations for all the speakers. If we refer to Davidson's (2016) categorization, all intervocalic stops in Japanese would belong to the "bleed" pattern. In other words, voicing continues from the preceding vowel. Four speakers out of eight speakers (CJF05, CJF06, CJM02 and CJM03) did not have voiceless closure intervals in voiced velar stops, and CJM02 did not have voiceless closure intervals in voiced alveolar stops

As expected, voiced stops had longer voiced closure intervals than voiceless stops, except for CJM02, whereas voiceless stops had longer voiceless closure intervals than voiced stops with voiceless closure. The total closure duration was significantly longer for voiceless stops than for voiced stops (Welch's two sample *t*-test, $p < 0.05$). Based on the 50% criterion proposed by Abramson and Whalen (2017), intervocalic stops have a negative voice onset time (−VOT) that is equal to the closure duration if the closure is voiced for at least half of its duration; otherwise, they have a positive voice onset time (+VOT) equal to the release burst duration. Table 2[1)] presents the number and ratio of utterances in which each speaker's closure voicing rate exceeds 50%. All voiced stop utterances had voiced closure intervals that exceeded half the closure duration. In other words, all voiced stop utterances exhibited a negative VOT. However, voiceless stop utterances from five speakers (CJF03, CJF05, CJF07, CJM01, and CJM02) also had voiced closure intervals that exceeded half of the closure duration to some extent, with a particularly high probability of being observed in voiceless stop utterances produced by CJM02 (83% for /ata/ and 95% for /aka/). Consequently, based on the 50% criterion, voiceless stops exhibiting a negative VOT existed, which is in contrary to the speakers' speech intentions and researchers' perceptual impressions. Thus, the 50% closure voicing rate criterion may

Table 2 The number and ratio of utterances where the closure voicing rate exceeds 50% for each speaker.

Word	Speaker	Ratio of VDCLO over 50%	VDCLO over 50%/total
ada	CJF01	100%	17/17
	CJF03	100%	11/11
	CJF05	100%	15/15
	CJF06	100%	15/15
	CJF07	100%	16/16
	CJM01	100%	20/20
	CJM02	100%	16/16
	CJM03	100%	11/11
ata	CJF01	0%	0/18
	CJF03	26%	5/19
	CJF05	16%	3/19
	CJF06	0%	0/19
	CJF07	0%	0/20
	CJM01	5%	1/19
	CJM02	83%	15/18
	CJM03	0%	0/19
aga	CJF03	100%	1/1
	CJF05	100%	13/13
	CJF06	100%	14/14
	CJF07	100%	14/14
	CJM02	100%	4/4
	CJM03	100%	2/2
aka	CJF03	5%	1/19
	CJF05	11%	2/19
	CJF06	0%	0/18
	CJF07	12%	2/17
	CJM02	95%	19/20
	CJM03	0%	0/20

not be suitable for Japanese intervocalic stops.

Notably, the duration of voiceless closure for voiceless stops was considerably longer than that for voiced stops (Figure 2). Even if the closure voicing rate for voiceless stops exceeded 50%, the absolute duration of the voiceless closure interval was longer, which may help maintaining the distinction between voiceless and voiced stops.

The mean and standard deviation of the release burst duration for all speakers are presented in Table 3, along with the results of the Welch's two sample *t*-test. The number of data points for voiced stops was fewer than that for voiceless stops, owing to the exclusion of without-burst utterances. For alveolar stops, the mean release burst interval for voiceless sounds (16.33 ms) was slightly longer than that for voiced sounds (14.99 ms); however, this difference barely reached significance ($p = 0.052$). For velar stops, the mean release burst interval for voiceless sounds (23.54 ms) was significantly longer than that for voiced sounds (19.44 ms; $p < 0.01$).

In conclusion, when native Japanese speakers produce voiced and voiceless intervocalic stops, they may adjust both the closure and release burst intervals.

4.1.2 Weakening Patterns in Voiced Stops

We observed three primary patterns of consonant weakening in voiced stops: (1) the voiced glide pattern, (2) the voiced frication pattern, and (3) the voiced frication pattern with halted vocal fold vibration and weak aspiration. For convenience, we categorized utterances without consonant weakening as (0) with burst pattern. Examples of the waveforms and spectrograms for each pattern are shown in Figure 3. These three weakening patterns are determined by the degree of burst weakening and the persistence of vocal fold vibration. Pattern (1) involves the complete absence of burst with sustained vocal fold vibration, pattern (2) includes frication weakening of burst with sustained vocal fold vibration; and pattern (3) arises from frication weakening of burst with incomplete vocal fold vibration.

While a significant portion of utterances with consonant weakening are still perceptually identified as voiced plosives ((1) the voiced glide pattern in Figure 3 is an example) according to the first author's impressions, some instances resemble glides or fricatives. Utterances exhibiting glide-like perceptions retained vowel-like waveforms in the acoustic closure interval (Figure 4), thus suggesting the potential incompleteness of articulatory closure formation. This possibility is discussed in Section 4.2.3.

Differences in the frequency of weakening patterns were observed in alveolar and velar consonants.

Table 3 Mean and standard deviation of release burst duration for all speakers.

Place of articulation	Voiced			Voiceless					
	Mean (ms)	*SD* (ms)	n	Mean (ms)	*SD* (ms)	*n*	*df*	*t*	*p*
Alveolar	14.99	6.51	121	16.33	4.24	151	197.18	−1.955	0.052
Velar	19.44	7.91	48	23.54	5.03	113	63.724	−3.311	< 0.01

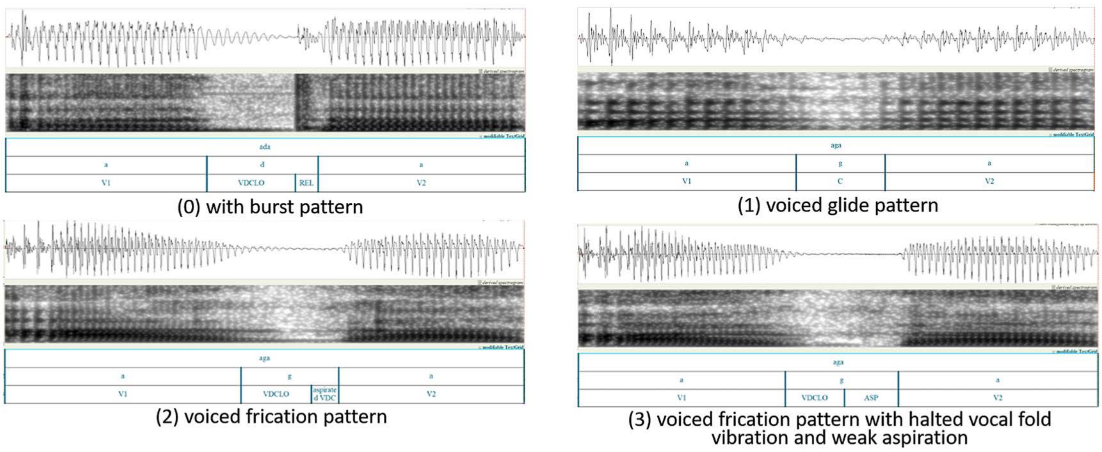

Figure 3 Examples of waveforms and spectrograms for each weakening pattern.

Table 4 shows the number and ratio of weakening pattern utterances for each accent and articulation point. Alveolar voiced stops indicated a higher ratio of with burst pattern (0) than velar stops, which is different from the tendency observed in English, where /d/ tends to weaken the most and /g/ weakens less (Warner and Tucker 2011). However, this finding confirms the results of Maekawa (2018). Maekawa attributes this sequence in Japanese to the complexity of the phonemic contrast between articulation places within the phonological system. In Japanese, alveolar phonemes include: plosive /d/, fricative /z/ (which can also be pronounced as an affricate allophone /dz/), and nasal /n/, whereas velar phonemes only involve the plosive /g/. Thus, finer articulatory details are required when a single articulation point accommodates multiple articulation patterns.

Differences in weakening rates based on accent patterns were observed (see /HL/ and /LH/ in /d/, /g/, /All/ of “0. With burst” in Table 4). The LH patterns exhibited lower rates of weakening than the HL patterns, with a tendency toward increased susceptibility to weakening when a pitch fall occurred.

4.2 Articulatory Results

In this section, we explain the results of the articulatory analysis. First, we compare the tongue contours at the maximal constriction between voiced and voiceless

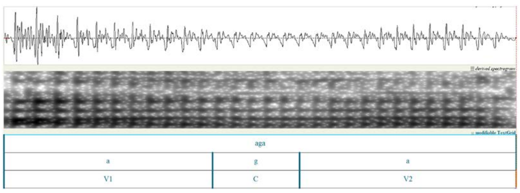

Figure 4 An example of waveform and spectrogram for voiced glide pattern with glide-like perception.

stops (4.2.1), then among different weakening patterns classified via acoustic analysis (4.2.2), and finally between weakened utterances that were perceived as stops and those that were not (4.2.3). For alveolar stops, we focus on the tongue tip, and for velar stops, we focus on the tongue body for comparison. The results for the velar stops are fewer than those for the alveolar stops owing to the reasons mentioned at the beginning of Chapter 3.

4.2.1 Comparison between Voiced and Voiceless Stops

In Figures 5 and 6, we compare the maximal constriction points of the tongue for alveolar and velar stops between voiced and voiceless consonants for each speaker. In the case of alveolar stops (Figure 5), the tongue tip position of the voiceless consonants was higher than that of the voiced consonants for six speakers (CJF01, CJF03, CJF06, CJM01, CJM02, and CJM03) among the eight speakers. For one speaker (CJF07), the tongue blade was slightly higher for voiceless consonants than for voiced consonants. For one speaker (CJF05), no significant difference was indicated in the tongue tip position between voiceless and voiced stops. In the case of velar stops (Figure 6), the tongue body position of voiceless consonants was higher than that of voiced consonants for four speakers (CJF06, CJF07, CJM02, and CJM03) among the six speakers. For two speakers (CJF03 and CJF05), no significant difference was indicated in the tongue body position between voiceless and voiced stops. In general, tongue position at the maximal constriction of voiceless stops tended to be higher than that of voiced stops, thus supporting previous research findings that voiced stops tend to show less linguopalatal contact than voiceless stops (Kochetov 2014; Wakumoto et al. 1998).

Additionally, for the four speakers who exhibited a higher tongue body position for voiceless consonants

Table 4 Number and ratio of weakening pattern utterances for each accent and place of articulation.

Place of articulation	Accent	(0) With burst		(1) Voiced glide		(2) Voiced frication		(3) With halted vocal fold vibration and ASP	
/d/	HL	56	70%	9	11%	13	16%	2	3%
	LH	68	85%	4	5%	7	9%	1	1%
	All	124	78%	13	8%	20	13%	3	2%
/g/	HL	20	25%	27	34%	25	31%	8	10%
	LH	38	48%	19	24%	22	28%	1	1%
	All	58	36%	46	29%	47	29%	9	6%
All	HL	76	48%	36	23%	38	24%	10	6%
	LH	106	66%	23	14%	29	18%	2	1%

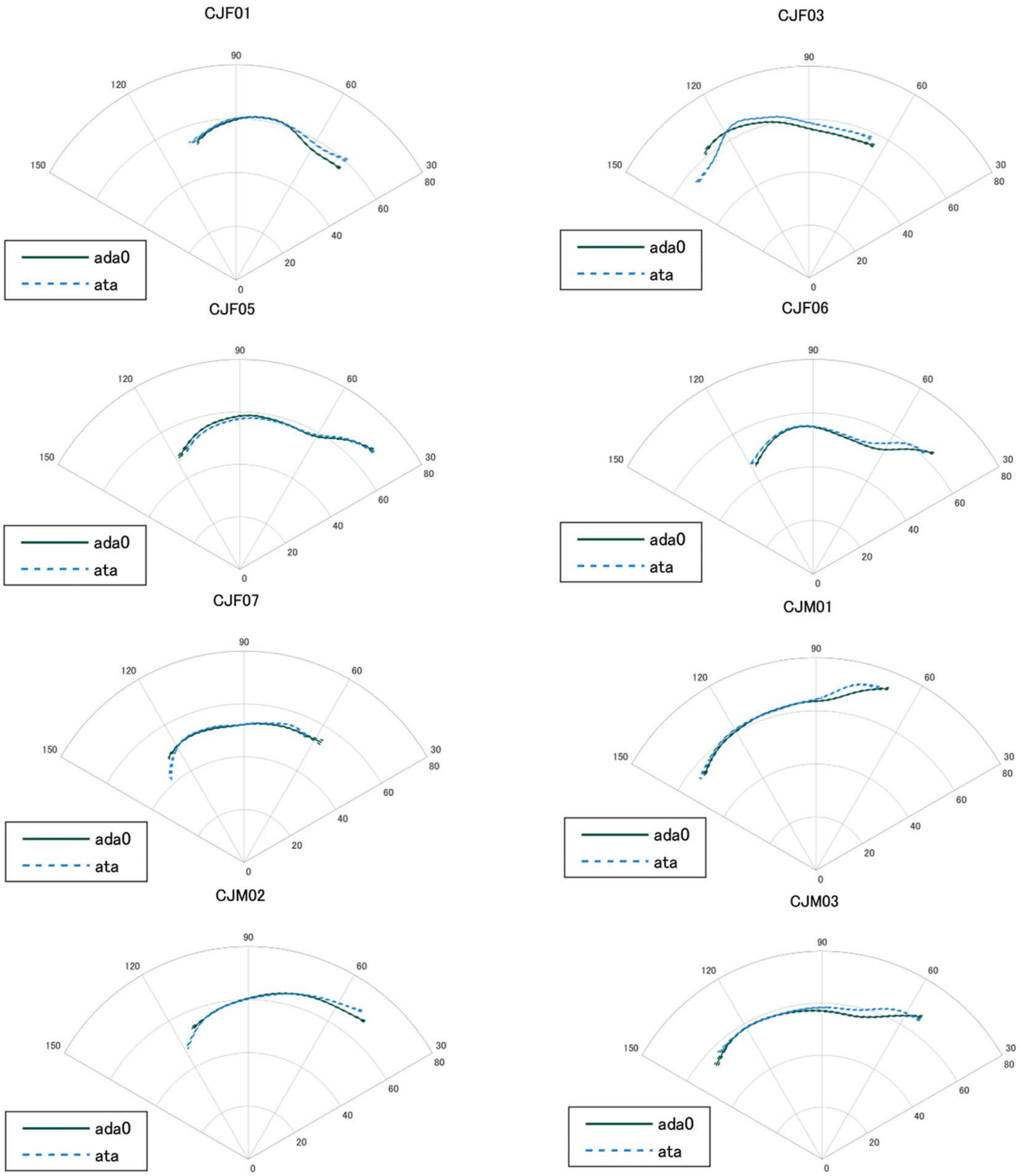

Figure 5 With-burst /d/ vs. /t/ generalized additive model (GAM) plots of eight Japanese speakers. This figure shows the mid-sagittal plane of the tongue. The right represents the anterior portion of the tongue, and the left side represents the posterior portion. 95% confidence intervals are plotted as dashed lines.

compared to voiced consonants in velar stops, their tongue roots were more advanced for voiced stops compared to voiceless stops. In a previous study (Li et al. 2024) using the same dataset as in this study, they analyzed the tongue position in frames immediately prior to the release of the closure, and found that tongue body lowering (TBL) and tongue root advancement (TRA) in voiced stops were observed more frequently in alveolar and velar stops than in bilabial stops, although the degree of TRA and TBL observed in Japanese voiced stops were lower than those observed in native English speakers. These results provide

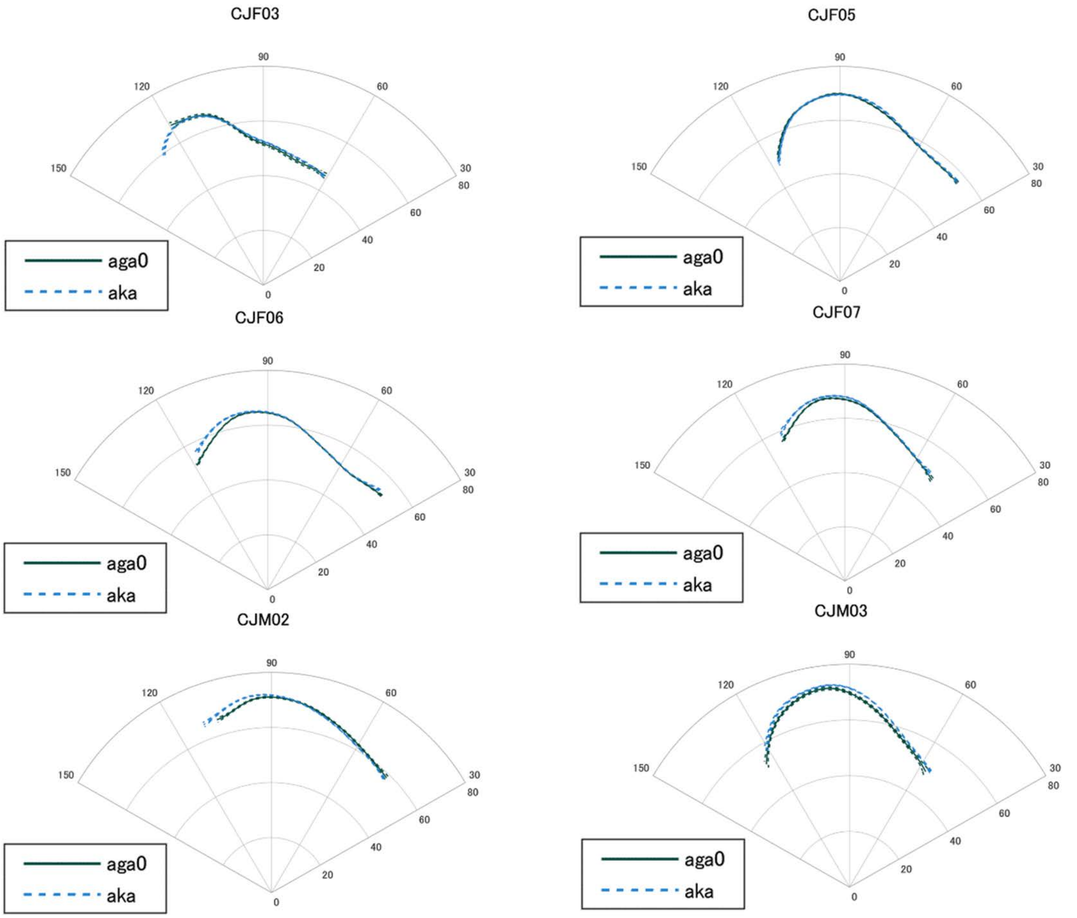

Figure 6 With-burst /g/ vs. /k/ generalized additive model (GAM) plots of six Japanese speakers. This figure shows the mid-sagittal plane of the tongue. The right side represents the anterior portion of the tongue, while the left side represents the posterior portion. 95% confidence intervals are plotted as dashed lines.

evidence that tongue root advancement and tongue body lowering, which were observed in English and Portuguese to maintain voicing, as confirmed by Ahn (2018), also exist in Japanese stops to some extent.

4.2.2 Comparison among Different Weakening Patterns

In Figures 7 and 8, we compare the tongue contours at the maximal constriction between with-burst and weakened voiced stops for each speaker. Here, weakening patterns (2) and (3), both characterized by frication weakening, were unified into a single category. We did not conduct this comparison for CJM01's voiced alveolar stop consonants owing to the absence of weakened utterances.

In alveolar stops (Figure 7), no noticeable difference was observed in tongue tip position between with-burst and weakened voiced stops, although slight variations were observed. Among the seven speakers, two (CJF01 and CJM02) exhibited the lowest tongue tip position in voiced glide weakening (1) utterances, whereas no difference was indicated between the other two patterns. The remaining speakers exhibited no differences among the three patterns.

The tongue tip position of with-burst utterances was expected to be the highest, followed by frication weakening (2, 3) utterances and voiced glide weakening (1). However, the results contradicted this expectation. One conceivable explanation is that the phenomenon observed by Kochetov (2014), in which voiced fricatives showed more linguopalatal contact than voiceless fricatives, also manifests in frication weakening (2, 3) utterances of voiced stops. Although tongue tip position

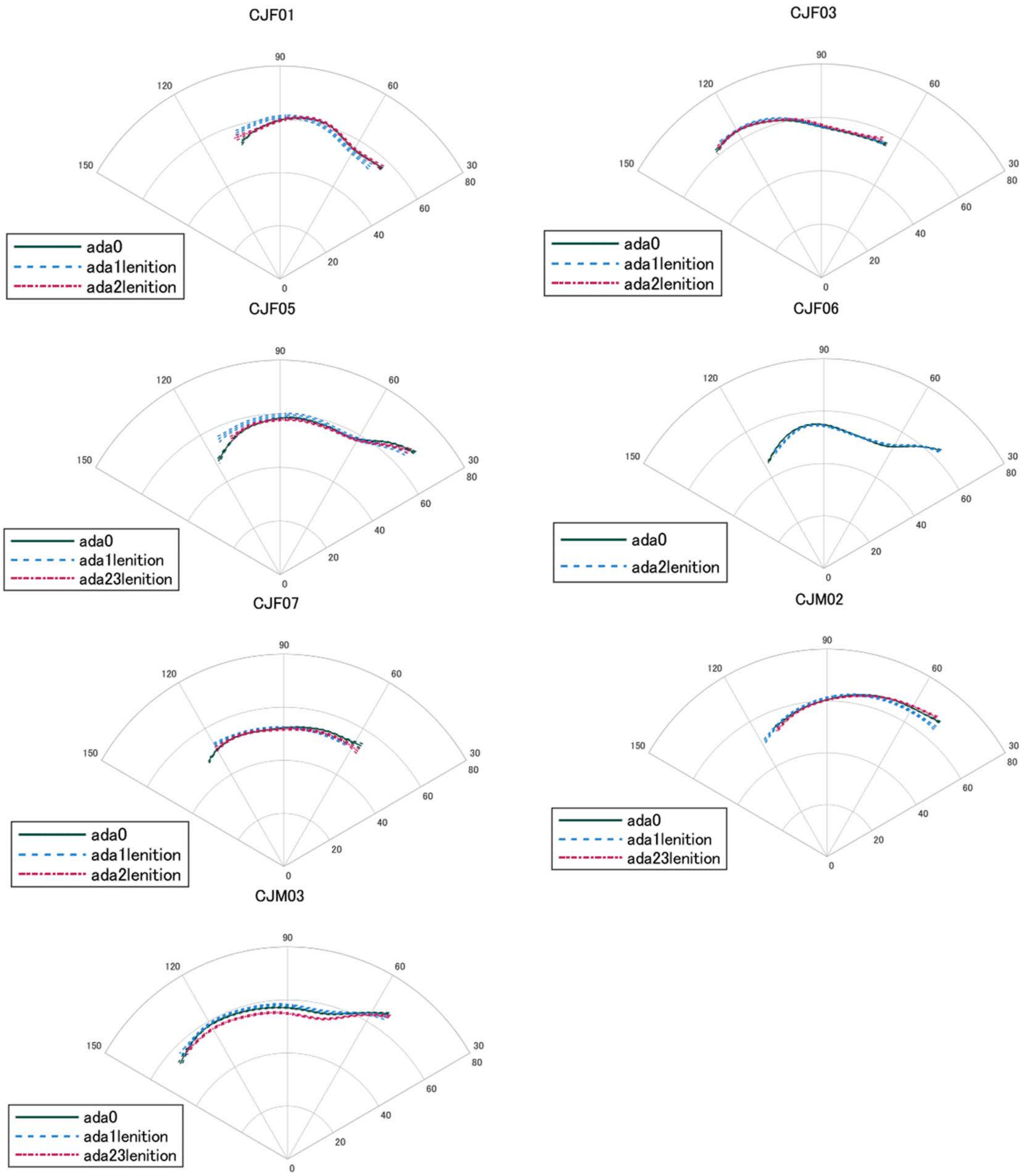

Figure 7 With-burst /d/ vs. weakening /d/ generalized additive model (GAM) plots of seven Japanese speakers. This figure shows the mid-sagittal plane of the tongue. The right side represents the anterior portion of the tongue, while the left side represents the posterior portion. 95% confidence intervals are plotted as dashed lines. "ada0" for with-burst utterances; "ada1lenition" for (1) voiced glide weakening utterances; (2) and (3) frication weakening utterances are plotted together as "ada2(3)lenition".

data for alveolar fricative utterances were not acquired in this study, we suspect that the results for fricatives would align with those of previous study, as the results for stops support those of previous studies. Therefore, if voiced alveolar fricatives have a higher tongue tip position than voiceless alveolar fricatives, then the tongue tip position in voiced (whether fully or partially voiced) frication weakening utterances may be higher,

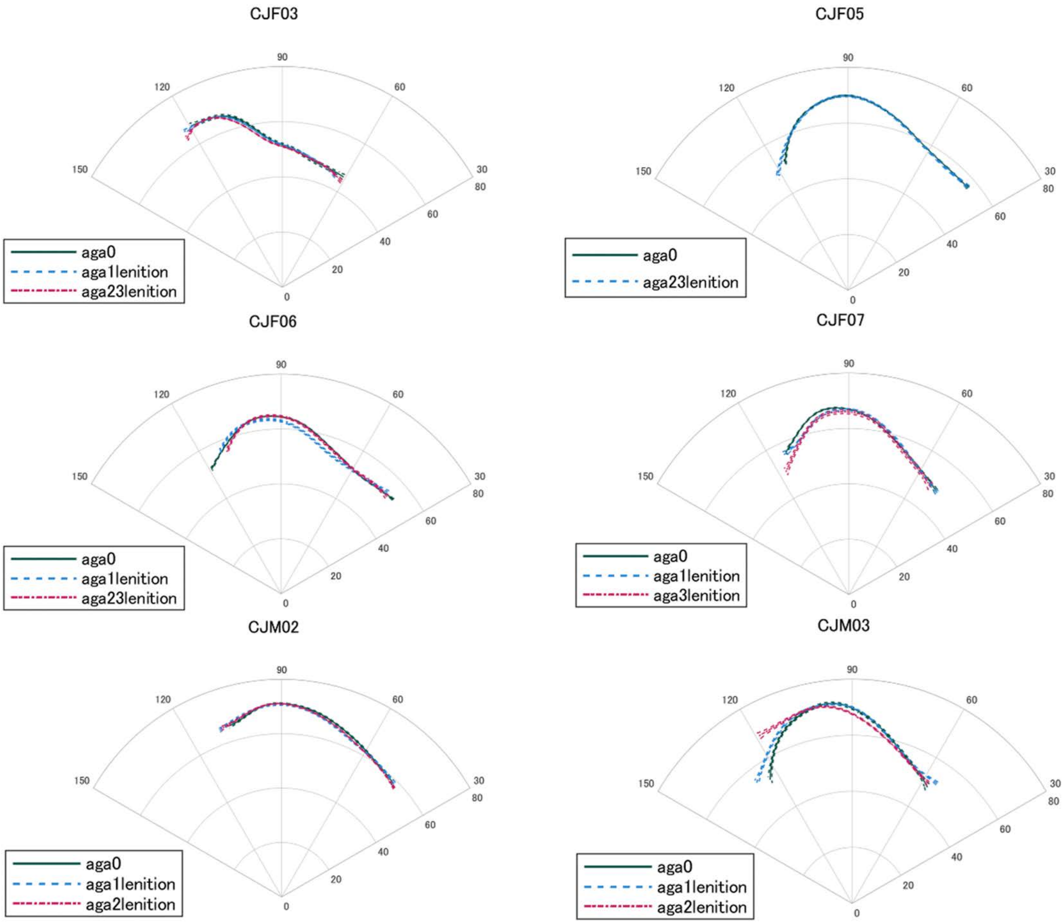

Figure 8 **With-burst /g/ vs. weakening /g/ generalized additive model (GAM) plots of six Japanese speakers. This figure shows the mid-sagittal plane of the tongue. The right side represents the anterior portion of the tongue, while the left side represents the posterior portion. 95% confidence intervals are plotted as dashed lines. "aga0" for with-burst utterances; "aga1lenition" for (1) voiced glide weakening utterances; (2) and (3) frication weakening utterances are plotted together as "aga2(3)lenition".**

thus causing them to be higher than for the case of with-burst voiced stops. If this explanation is valid, then the phenomenon of the tongue approaching the palate more in alveolar voiced fricatives may be physiologically determined, regardless of the speaker's speech intention.

In the case of velar stops (Figure 8), no differences were observed in the position of the tongue body, creating a constriction against the palate in two speakers (CJF05 and CJM02). In other two speakers (CJF06 and CJF07), the tongue body position in the glide weakening (1) utterances was lower than that in the with-burst utterances. CJF06 had frication weakening (2, 3) utterances at the same height as the with-burst utterances, whereas CJF07 had the frication weakening (2, 3) utterances at the same height as the glide weakening (1) utterances. CJF03 and CJM03 had the glide weakening (1) and with-burst utterances at the same position, with frication weakening (2, 3) utterances being similar to the two aforementioned patterns in the area of constriction, but the tongue body shifted slightly posteriorly.

Only one speaker exhibited the lowest tongue position in glide weakening (1) utterances, which implies that no clear trend similar to that of alveolar stops was observed in velar stops. This may be related to the softness of the velar articulator. When the tongue body presses firmly against the soft palate, a firm closure is formed, allowing for complete obstruction of airflow from the lungs. However, even with a soft touch, closure can be achieved, while it may not completely obstruct airflow from the lungs, leaving a slight gap. Additionally, given the high rate of weakening of voiced velar stops, the overall tongue position may be lower for voiced than for voiceless velar stops. In cases of frication weakening (2, 3) utterances, the tongue body shifted slightly posteriorly for CJM03, suggesting a different place of articulation for voiced velar fricatives than voiced velar stops in some speakers. However, since velar fricatives exist only as allophones of velar stops in the Japanese phonological system, it is difficult to verify this difference in the place of articulation in Japanese speech; thus, validation with data from languages that have velar fricative sounds is required.

4.2.3 Comparison between Weakened Utterances with/without Stop Perception

In Section 4.1.2, we mentioned the possibility that weakened utterances with glide or fricative perception may result in incomplete closure formation compared with burst utterances. In this section, we verify this possibility using ultrasound images. Figure 9 shows a generalized additive model comparison of the tongue contours at the maximal constriction for three speakers who had a sufficient number of weakened utterances without stop perception (based on the first author's impression).

CJF03 showed little difference, whereas CJM02 exhibited more tongue root advancement in with-burst utterances; however, no difference was observed at the highest point of the tongue body. CJM03 had almost the same tongue position for burst utterances and glide weakening (1) utterances with stop perception and a lower position for utterances with glide perception. In alveolar stops, oral constriction is produced by complete contact between the tongue tip and the superior alveolar ridge; however, the positioning of the ultrasound probe in this study only allowed the observation of mid-sagittal images of the oral cavity. If, hypothetically, only the middle of the tongue tip contacts the superior alveolar ridge in weakened utterances, while the sides do not, this distinction would not be discernible on ultrasound images. Speakers who produce glide weakening (1) voiced alveolar stops in a way similar to alveolar tap [ɾ], may fit this scenario. As mentioned in the previous section, the tongue position may be lower in general for voiced velar stops. Therefore, the tongue position can be presumed to remain relatively unchanged regardless of the perceptual impression. However, in the case of CJM03, the tongue position in the glide perception utterances was significantly lower. This suggests individual differences in the realization of weakened stops and the diversity of articulatory

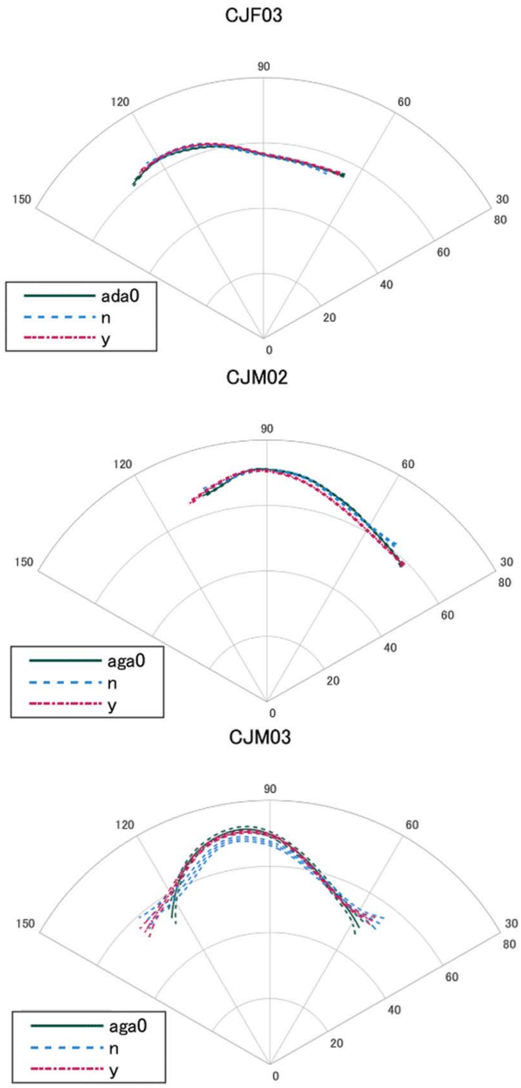

Figure 9 With-burst stops vs. glide weakening stops with glide perception vs. glide weakening stops with stop perception generalized additive model (GAM) plots of three Japanese speakers. This figure shows the mid-sagittal plane of the tongue. The right side represents the anterior portion of the tongue, while the left side represents the posterior portion. 95% confidence intervals are plotted as dashed lines. The figure for CJF03 shows the result of /ada/. The figures for CJM02 and CJM03 show the results of /aga/. "ada0" and "aga0" for with-burst utterances; "n" for glide weakening (1) utterances with glide perception; "y" for glide weakening (1) utterances with stop perception.

movements corresponding to the same phoneme.

5. Discussion

5.1 Efforts of Larynx and Supralaryngeal Articulatory Organs in Japanese Voiced-Voiceless Contrast

In Japanese voiceless stops, vocal fold vibration is passively inherited from the preceding vowel until the difference between intraoral and subglottal air pressure falls low enough for voicing to stop, followed by the abduction of the vocal folds. During voiceless closure, air is trapped, and constriction by the tongue is rapidly released at the burst, followed by the resumption of vocal fold vibration after release. This process is illustrated in Figure 10 (Panel 1). In voiced stops, the degree of tongue root advancement observed in Japanese is lower than that in English, as observed by Ahn (2018) (Li et al. 2024). Instead, the rate of weakening in Japanese is higher than that in English (Warner and Tucker 2011). Japanese native speakers, faced with the conflicting tasks of forming a closure and maintaining vocal fold vibration, may not adopt strategies such as tongue root advancement or tongue body lowering as much as Portuguese or English speakers. Instead, they may attempt to maintain vocal fold vibration more by loosening the closure to create an air passage, which is also an effective method to preserve voicing in stop consonants. (Ohala 1997). In other words, in Japanese voiced stops, efforts are made to maintain phonation, but voicing during closure is not as solid as in true voice languages such as Portuguese or Russian (Figure 10, Panel 2). Thus, Japanese native speakers employ a unique method to maintain voicing in voiced stops through consonant weakening to realize the voiced-voiceless contrast (Figure 10, Panel 3). And as mentioned in 4.1.1, they may also integrate the duration of both the closure and release burst intervals, potentially even adjusting the length of the preceding and following vowels (Li et al. 2023).

5.2 Phonetic and Phonological Classification

In this section, we aim to predict the diachronic changes in the voiced-voiceless contrast of Japanese stops based on the framework of Beckman, Jessen and Ringen (2013). It should be noted that the numerical values specified here are purely theoretical predictions and are not rigorously based on empirical data.

The traditional perspective, which classifies Japanese as a true voice language, is illustrated in the top table of Figure 11, showing a strict contrast between voiced and voiceless stops. This type of contrast is still commonly observed among older generations and in certain regions (Byun 2021; Takada 2011).

The middle table in Figure 11 represents the current phonetic realizations of Japanese stops, interpreted under the assumption that they are specified for both [sg] and [voice]. Word-initial /p, t, k/ are realized as

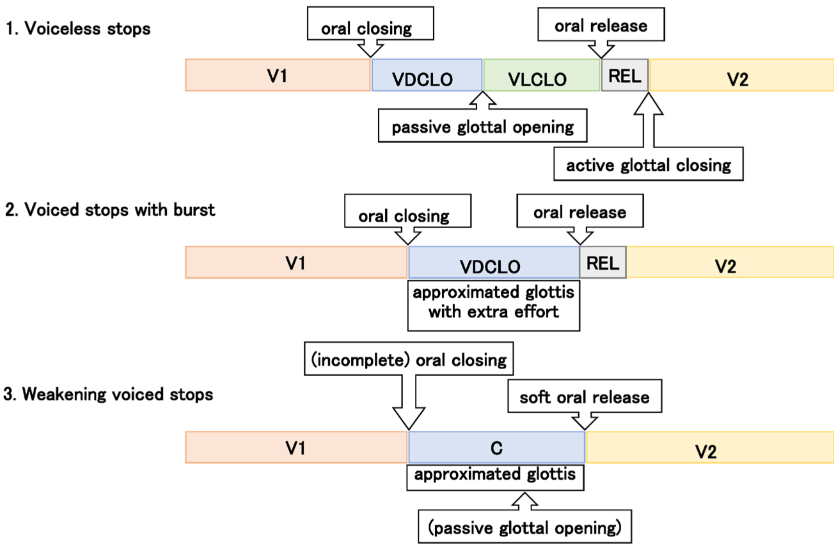

Figure 10 Schematic images of the acoustic implementation and the timing of larynx and supralaryngeal articulatory movements of Japanese stops for 1. Voiceless stops, 2. Voiced stops with burst, 3. Weakened voiced stops.

Numerical values assigned to laryngeal features in Japanese as a true voice language (traditional view)		
	/p, t, k/	/b, d, g/
Word-initial	[1voice]	[9voice]
Intervocalic	[2voice]	[9voice]

⇩

Numerical values assigned to laryngeal features in Japanese as it undergoes changes (based on our experimental results)		
	/p, t, k/	/b, d, g/
Word-initial	[5~8sg] [1voice]	[1~2sg] [5~8voice]
Intervocalic	[2sg] [2voice]	[1sg] [8voice]

⇩

Numerical values assigned to laryngeal features in Japanese as an aspirating language (future prediction)		
	/p, t, k/	/b, d, g/
Word-initial	[5~8sg]	[1~2sg]
Intervocalic	[3sg]	[1sg]

Figure 11 Predicted diachronic changes in the voiced-voiceless contrast of Japanese stops based on Beckman, Jessen and Ringen (2013)'s framework. The top table represents the traditional perspective, the middle table reflects the current interpretation of phonetic realizations, and the bottom table shows a potential future state. A diachronic shift from the top to the bottom can be anticipated.

voiceless aspirated stops, which suggests they may receive a low [voice] specification and a medium-to-high [sg] specification (say [5～8sg] and [1voice]). On the other hand, word-initial /b, d, g/ are typically realized as voiceless unaspirated stops, though some younger speakers from the Tokyo area still produce them with prevoicing. Overall, these sounds are likely specified with a low [sg] and medium-to-high [voice] feature (say [1～2sg] and [5～8voice]). In intervocalic positions, /p, t, k/ are realized as slightly voiced and slightly aspirated stops, indicating low [voice] and low [sg] specifications (say [2sg] and [2voice]). These specifications do not inhibit passive voicing, which allows for the existence of voiced closure intervals in intervocalic /p, t, k/. Intervocalic /b, d, g/, in contrast, are realized as voiced unaspirated stops with prevoicing. While their degree of voicing is stronger than the passive voicing observed in /p, t, k/, it is less robust compared to languages like Russian. These sounds can be specified with low [sg] and slightly higher [voice] features (say [1sg] and [8voice]). As noted in the previous section, intervocalic /b, d, g/ still exhibit remnants of active voicing, reflecting their historical roots as part of a true voice language.

Given the current decline in the voicing contrast of Japanese stops (Byun 2021; Gao and Arai 2019; Takada 2011), it seems that Japanese is transforming into an aspirating language, similar to English or German. The bottom table in Figure 11 presents a prediction of what Japanese might look like as an aspirating language in the future. In this prediction, the privative feature [voice] and their numerical conversions disappear entirely, leaving only [sg] specifications.

There are also studies suggesting the increasing importance of F0 (the fundamental frequency) in addition to VOT for word-initial stops (Byun 2021; Gao and Arai 2019), thereby indicating that the current contrast in Japanese stops may still be in a transitional stage. It still seems too early to conclude whether the voiced-voiceless contrast in Japanese stop consonants will develop into an unaspirated-aspirated contrast, change into a tonal contrast, or remain as a voiced-voiceless contrast. Further longitudinal investigations will be necessary to address this question.

6. Conclusion

In this study, we conducted recording experiments on word-medial stops with native Japanese speakers and simultaneously recorded the articulatory movements of the tongue using an ultrasound system. The results obtained for the research questions in the Introduction are as follows:

I. Japanese intervocalic stops have partially or fully voiced closure intervals for voiced sounds and partially voiced closure intervals for voiceless sounds. Voiced stops may also have partially voiceless closure intervals, although their duration is considerably shorter than that of voiceless stops. The difference in the release burst interval varied depending on the place of articulation. No significant differences were indicated between alveolar voiced and voiceless stops. However, for velar stops, voiceless stops were significantly longer than voiced stops. In addition, voiced stops were frequently realized without burst.

II. At the point of maximal constriction, the tongue position was higher for voiceless sounds than for voiced sounds. This supports the acoustic analysis results that voiceless utterances consistently exhibit burst, whereas voiced utterances are often weakened. When comparing weakened utterances with and without bursts for voiced sounds, with-burst utterances tended to have a slightly higher tongue position; however, the difference was not as pronounced as that between voiceless and voiced with burst. Some speakers also showed a tendency for glide weakening (1) to be lower, but there may also be differences in articulatory movements to realize weakened utterances among speakers.

III. In Japanese intervocalic voiced stops, supralaryngeal articulatory movements have been observed to actively maintain voicing in addition to passive voicing from the preceding vowel. While Japanese stops categorized phonologically as "voiced" may not have as solid prevoicing as true voice languages such as Russian or Portuguese, they are realized voiced phonetically in intervocalic environments. Conversely, stops categorized phonologically as "voiceless" in Japanese are in fact not specified for totally voiceless but for a relatively lower voicing degree and a relatively lower aspirating degree. Thus, intervocalic voiceless stops can undergo passive voicing in intervocalic environments but still have much longer voiceless closure durations than voiced stops.

Further investigations into bilabial stops using high-frame-rate cameras are needed, as the articulatory movements of bilabial stops cannot be confirmed via ultrasound imaging. Additionally, an investigation of the articulatory movements of Japanese word-initial stops is required.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Numbers JP19K13254, JP22J01381/22KJ2752, and JP23K12212. The results regarding consonant weakening include revisions of a presentation titled "Weakening patterns of intervocalic voiced plosives in Japanese," which was presented at Acoustics 2023 Sydney. This study adhered to the ethical guidelines of Journal of the Phonetic Society of Japan and was approved by the Sophia University Ethics Committee for Research on Human Subjects (Approval No.: 2022-052). Informed consent regarding participation and the scope of data publication was obtained from all participants.

Notes

1) Each speaker repeated each word 20 times; however, those with synchronization errors or unclear ultrasound images were excluded. Additionally, only utterances with burst were included for voiced stop consonants. Thus, there were cases in which the total number of utterances was less than 20.

References

Abramson, A. S. and D. H. Whalen (2017) "Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions." *Journal of Phonetics* 63, 75–86.

Ahn, S. (2018) "The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese." *Journal of Phonetics* 71, 451–467.

Ahn, S., H. Kwon and M. Faytak (2024) "Tongue position in Mandarin Chinese voiceless stops." *JASA Express Letters* 4(2), 025207.

Articulate Instruments Ltd. (2022) Articulate Assistant Advanced (AAA) (2.19.08). http://www.articulateinstruments.com/downloads/ (accessed May 25, 2022)

Beckman, J., M. Jessen and C. Ringen (2013) "Empirical evidence for laryngeal features: Aspirating vs. true voice languages." *Journal of Linguistics* 49(2), 259–284.

Boersma, P. and D. Weenink (2023) "Praat: Doing phonetics by computer." Version 6.3.16, retrieved 2021 from http://www.praat.org/ (accessed August 30, 2023)

Byun, H. (2021) "Acoustic characteristics for Japanese stops in word-initial position: VOT and post-stop fo." *Journal of the Phonetic Society of Japan* 25, 41–63.

Davidson, L. (2016) "Variability in the implementation of voicing in American English obstruents." *Journal of Phonetics* 54, 35–50.

Davidson, L. (2018) "Phonation and laryngeal specification in American English voiceless obstruents." *Journal of the International Phonetic Association* 48(3), 331–356.

Fujimoto, M., T. Kitamura and S. Funatsu (2022) "Nihongo no haretsuon /g/ to /k/ no chōon-teki to onseiteki chigai—MRI ni yoru kēsusutadī—." [Articulatory and acoustic differences between the Japanese plosives /g/ and /k/: A case study using MRI]. *Memoirs of Konan University. Intelligence & Informatics Series* 14(2), 103–113.

Gao, J. and T. Arai (2019) "Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery." *Journal of Phonetics* 77, 100932.

Kent, R. D. and K. L. Moll (1969) "Vocal-tract characteristics of the stop cognates." *The Journal of the Acoustical Society of America* 46, 1549–1555.

Kochetov, A. (2014) "Voicing and tongue-palate contact differences in Japanese obstruents." *Journal of the Phonetic Society of Japan* 18(2), 63–76.

Koizumi, T. (2003) *Kaitei Onseigaku Nyūmon* [Introduction to Phonetics (Revised ed.)]. Daigaku-Shorin.

Li, W., A. Mizoguchi, M. Morimoto and T. Arai (2024) "Tongue position in Japanese intervocalic voiced and voiceless stops". In I. Wilson, A. Mizoguchi, J. Perkins, J. Villegas and N. Yamane (eds.) *Ultrafest XI: Extended Abstracts*, 58–62. University of Aizu.

Li, W., M. Morimoto, A. Mizoguchi and T. Arai (2023) "Nihongo no gochū haretsuon hatsuwa ni okeru bogowasha ni yoru yūsei-musei tairitsu oyobi chūgokugo o bogo to suru nihongo gakushūsha tono hikaku." [Japanese voicing contrast in word-medial plosives produced by native speakers and comparison with Chinese-native Japanese learners]. *Proceedings of the 2023 Spring Meeting of The Acoustic Society of Japan*, 563–566.

Lisker, L. and A. S. Abramson (1964) "A cross-language study of voicing in initial stops: Acoustical measurements." *Word* 20(3), 384–422.

Maekawa, K. (2018) "Weakening of stop articulation in Japanese voiced plosives." *Journal of the Phonetic Society of Japan* 22(1), 21–34.

Ohala, J. (1983) "The origin of sound patterns in vocal tract constraints." In P. MacNeilage (ed.) *The Production of Speech*, 189–216. New York: Springer.

Ohala, J. (1997) "Aerodynamics of phonology." *Proceedings of the Seoul International Conference on Linguistics*, Seoul: Linguistic Society of Korea, 92, 97.

R Core Team (2021) R: A language and environment for statistical computing (version 4.1.1). R Foundation for Statistical Computing.

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

Svirsky, M. A., K. N. Stevens, M. L. Matthies, J. Manzella, J. S. Perkell and R. Wilhelms-Tricarico (1997) "Tongue surface displacement during bilabial stops." *The Journal of the Acoustical Society of America* 102(1), 562–571.

Takada, M. (2011) *Nihongo no Gotō Heisaon no Kenkyū: VOT no Kyōjiteki Bunpu to Tsūjiteki Henka* [Research on the Word-initial Stops of Japanese: Synchronic Distribution and Diachronic Change in VOT]. Kurosio Publishers.

The MathWorks Inc. (2021) MATLAB version: 9.10.0 (R2021a), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com (accessed July 11, 2022)

Tiede, M. (2022) GetContours. GitHub repository. https://github.com/mktiede/GetContours (accessed October 26, 2022)

Wakumoto, M., S. Masaki, K. Honda and T. Ohue (1998) "A pressure sensitive palatography: Application of the new pressure sensitive sheet for measuring tongue-palatal contact pressure." *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, 7, 3151–3154.

Warner, N. and B. V. Tucker (2011) "Phonetic variability of stops and flaps in spontaneous and careful speech." *Journal of the Acoustical Society of America* 130(3), 1606–1617.

Westbury, J. (1983) "Enlargement of the supraglottal cavity and its relation to stop consonant voicing." *The Journal of the Acoustical Society of America* 73, 1322–1336.

Westbury, J. and P. Keating (1986) "On the naturalness of stop consonant voicing." *Journal of Linguistics* 22, 145–166.

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

(Received May 7, 2024, Accepted Feb. 23, 2025,
e-Published Apr. 30, 2025)