Tongue position in Japanese intervocalic voiced and voiceless stops

Weiyu Li[1], Ai Mizoguchi[2,3], Maho Morimoto[1,4], Takayuki Arai[1]

[1]Sophia University, [2]Maebashi Institute of Technology, [3]NINJAL, [4]Chuo University
w-li-5x4@eagle.sophia.ac.jp, aimizoguchi@maebashi-it.ac.jp, maho.morimoto.jp@gmail.com, arai@sophia.ac.jp

Keywords: Japanese stops, voicing contrasts, tongue position, intervocalic, oral pressure

Introduction

Japanese stops have a two-way contrast between voiced and voiceless consonants [1], where voiced stops may lack pre-voicing at the word-initial position [2, 3] or have incomplete pre-voicing at the word-medial position. Similarly, English voiced stops at the word-initial position often lack pre-voicing; however, studies have shown tongue root advancement (TRA) and tongue body lowering (TBL) to maintain vocal fold vibration during closure [4]. Therefore, this study aimed to investigate whether similar phenomena are observed in Japanese stops as in English.

Experiment

Recordings were made of native Japanese speakers, while simultaneously recording tongue movements using an ultrasound imaging device. Stimuli were nonce-words "aCa" (C=p,b,t,d,k,g), randomized and inserted into carrier sentences (/korewa __ to i:masu/'this is called __'), spoken with HL and LH accents, with each stimulus repeated 10 times.

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 [5] was used to stabilize the relative positions of the probe and the participant's head. The frame rate of the obtained ultrasound videos 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 the AAA software [6].

Recordings with synchronization errors, unclear visibility of the tongue root, or inability to track the tongue were excluded from the analysis. Data from five speakers (bilabial data from one participant were excluded due to tongue root invisibility) were included in the analysis, totaling 536 utterances. Details of the five speakers are presented in Table 1.

Table 1: Participants

Speaker	Gender	Age	Region
CJF05	F	21	Shizuoka/Tokyo
CJF06	F	22	Tokyo
CJF07	F	21	Kanagawa
CJM02	M	20	Tokyo
CJM03	M	20	Singapore/Tokyo

Analysis

For the acoustic analysis, we segmented the recordings using Praat [7], measuring the durations of voiced closure (VDCLO) and voiceless closure (VLCLO), as well as the release/burst interval (REL). Voiced stop sounds sometimes occurred without burst, in which case only VDCLO and VLCLO were measured.

Using GetContours [8], tongue contours were tracked in frames immediately prior to the release of the closure or to the following vowel (in cases of lenition). Tongue contours for voiced and voiceless stops were plotted and compared using the GAM model [9] for each speaker and accent. If TRA and TBL were observed in voiced stops, a "+" was given; conversely, if TRA and TBL were observed in voiceless sounds, a "-" was given; if no difference was observed between voiced and voiceless stops, "0" was given. Figure 1 shows samples of TRA and TBL judgments. The results are presented in the following section.

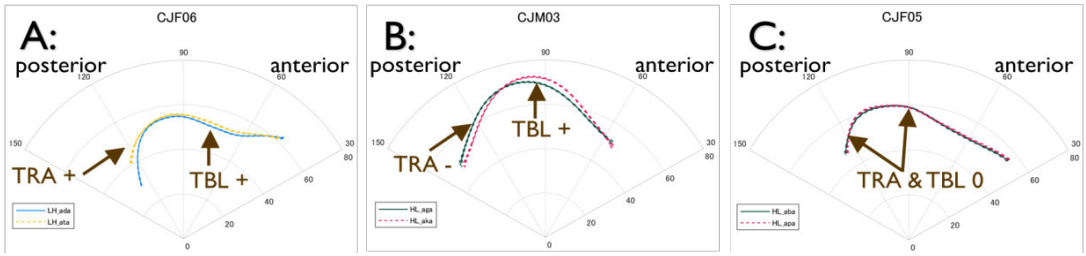

Figure 1: Examples of "+," "-," and "0." A: an example of TRA & TBL "+"; B: an example of TRA "-" and TBL "+"; C: an example of "0".

Results and Discussion

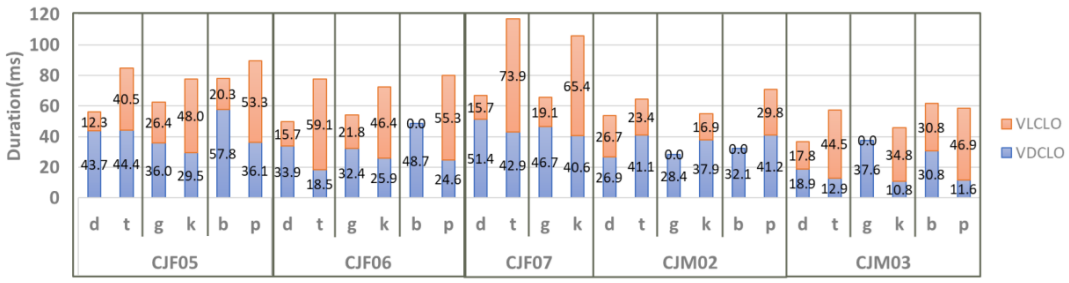

Figure 2: Mean of voiced and voiceless intervals (ms) during closure for each speaker.

The findings from the acoustic analysis are presented hereby: Regarding the results of voicing during closure (Figure 2), both voiced and voiceless stops had voiced closure intervals, but voiced stops had longer intervals (except for alveolar stops in CJM02). Voiceless stops' VDCLO may originate from passive voicing from the preceding vowel. Voiced stops also had short voiceless closure intervals, while voiceless stops had much longer intervals (except for alveolar stops in CJM02). Moreover, the total length of the closed interval was typically longer for voiceless stops than for voiced stops. Regarding the duration of the release/burst interval, Table 2 shows the mean and standard deviation of the REL duration for each speaker. Voiced stops without release were excluded from this analysis. Overall, the release burst interval for voiceless stops tended to be slightly longer, but the difference was significant only under four conditions.

Speaker	Place of articulation	Voiced			Voiceless					
		Mean (ms)	SD (ms)	n	Mean (ms)	SD (ms)	n	df	t	p
CJF05	Bilabial	10.13	4.11	19	12.91	3.18	20	33.937	-2.353	0.02456*
	Alveolar	19.50	4.26	15	17.93	5.07	19	31.846	0.98219	0.3334
	Velar	17.96	5.61	12	25.04	5.62	19	25.992	-3.5079	0.001663**
CJF06	Bilabial	9.24	2.90	12	12.87	2.88	19	23.418	-3.4037	0.002393**
	Alveolar	12.77	5.08	15	12.91	2.06	19	17.649	-0.099574	0.9218
	Velar	20.97	7.15	14	20.19	4.20	18	19.812	0.35918	0.7233
CJF07	Alveolar	15.71	4.70	16	14.99	2.79	20	20.257	0.56855	0.5759
	Velar	19.51	10.79	14	20.74	2.21	17	14.44	-0.41386	0.6851
CJM02	Bilabial	10.85	4.49	7	18.27	5.79	20	13.556	-3.4753	0.003877**
	Alveolar	18.59	8.08	16	20.32	4.55	18	23.014	-0.75519	0.4578
	Velar	21.70	7.29	4	26.58	3.97	20	3.364	-1.3002	0.2754
CJM03	Bilabial	12.73	5.04	9	12.07	3.36	18	11.672	0.35344	0.7301
	Alveolar	16.25	10.77	11	17.27	4.53	19	12.083	-0.2985	0.7704
	Velar	17.22	8.91	2	24.86	5.97	20	1.0917	-1.1858	0.433

Table 2: Mean and standard deviation of REL duration for each speaker with T-Test results

Speaker	Accent pattern	Bilabial		Alveolar		Velar	
		TRA	TBL	TRA	TBL	TRA	TBL
CJF05	HL	0	0	0	0	+	0
	LH	0	0	0	-	+	0
CJF06	HL	0	0	+	+	+	0
	LH	0	0	+	+	+	-
CJF07	HL	NA	NA	0	+	-	+
	LH	NA	NA	+	0	0	+
CJM02	HL	0	0	0	+	0	+
	LH	0	0	0	+	0	0
CJM03	HL	0	0	+	+	-	+
	LH	-	+	+	+	-	+
Number of "+"		1		12		9	
Number of "-"		1		1		4	

Table 3: Tongue position compared by 0 (no difference between voiced and voiceless), + (TRA or TBL observed in voiced stops), - (TRA or TBL observed in voiceless stops).

The results of the articulatory analysis (Table 3) showed that all five speakers exhibited some degree of TRA or TBL for voiced stops, albeit with interspeaker variability. Bilabial stops showed lower degrees of TRA and TBL compared to alveolar and velar stops, while they had the same high voicing rates during closure. Velar stops had more "-" than bilabial and alveolar stops. In bilabial stops,

protrusion of the lips could make it easier to facilitate vocal fold vibration maintenance compared to alveolar and velar stops. Another reason why alveolar and velar stops might show more TRA and TBL than bilabial stops is that the intraoral air pressure is higher the farther back the stop constriction. Therefore, to obtain sufficient delta-P (i.e., the pressure difference across the glottis) for voicing, TRA and TBL would be more necessary in velar and alveolar stops.

Velar stops showed slightly lower TRA and TBL compared to alveolar stops, yet had higher voicing rates, possibly because of the higher rate of consonant weakening in voiced velar stops, which can facilitate vocal fold vibration maintenance. Table 4 shows the consonant weakening rates in voiced stops for each speaker. Consonant weakening hardly occurred in voiceless stops. In voiced stops, velar stops tended to have consonant weakening the most, at a rate of approximately 51%, followed by bilabial stops at 35%. Alveolar stops tended to exhibit consonant weakening the least, at a rate of approximately 24%.

Speaker	Place of articulation	Weakening rate
CJF05	Bilabial	5%
	Alveolar	20%
	Velar	25%
CJF06	Bilabial	40%
	Alveolar	20%
	Velar	30%
CJF07	Bilabial	10%
	Alveolar	15%
	Velar	30%
CJM02	Bilabial	65%
	Alveolar	20%
	Velar	80%
CJM03	Bilabial	55%
	Alveolar	45%
	Velar	90%

Table 4: Consonant weakening rate in voiced stops for each speaker

For accent patterns, HL pattern had the same number of “+” as LH pattern. The number of "-" in the HL pattern was slightly less than that in the LH pattern, but as there were not that many in either pattern, it can be considered that there was almost no difference. In addition, even if there were speakers who had differences between HL and LH pattern, the difference was only between “0” and “+” or “0” and “-”, and there were no cases where one accent pattern was “+” and the other was “-”; thus, it seems that the degree of TRA and TBL does not vary with accent.

Overall, fewer TRA and TBL were observed in Japanese voiced stops compared to native English speakers.

Conclusions

- In Japanese intervocalic stops, both voiced and voiceless stops had voiced closure.
- TRA and TBL were observed. In voiced stops, voicing is actively extended.
- Voiced stops often occurred with consonant weakening, which might help maintain vocal fold vibration.
- Although this study focuses on Japanese intervocalic stops, the same phenomenon is likely to be observed in word-initial stops, although the voicing rate during closure is expected to be lower. We will verify the word-initial data in the near future.

Acknowledgments

This study was supported by the Sophia University Special Grant for Academic Research (Research in Priority Areas) and JSPS KAKENHI (Grant Numbers JP22KJ2752, JP19K13254, and JP23K12212). We would like to thank the anonymous reviewers for their valuable comments and suggestions, which have significantly improved the quality of this paper.

References

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

[2] Takada, M. 2011. *Nihongo no gotō heisaon no kenkyū: VOT no kyōjiteki bunpu to tsūjiteki henka* [*A study of Japanese word-initial stop consonants: Co-occurrence distribution of VOT and diachronic changes*]. Kurosio Publishers.

[3] Gao, J., Arai, T. 2019. Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *J. Phon.* 77, 100932.

[4] Ahn, S. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *J. Phon.* 71, 451–467.

[5] Spreafico, L., Pucher, M., Matosova, A. 2018. UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. *Interspeech*, 1517–1520.

[6] Articulate Instruments Ltd. 2022. Articulate Assistant Advanced (AAA) (2.19.08). http://www.articulateinstruments.com/downloads/

[7] Boersma, P., Weenink, D. 2021. Praat: Doing phonetics by computer. Version 6.1.42, retrieved 2021 from http://www.praat.org/.

[8] Tiede, M. 2022. GetContours. GitHub repository, https://github.com/mktiede/GetContours.

[9] Wood, S.N. 2006. *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.