An acoustic and articulatory study on the acquisition of Japanese plosive consonants by native Standard Chinese-speaking learners of Japanese

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

[1]*Sophia University,* [2]*Maebashi Institute of Technology,* [3]*NINJAL,* [4]*Chuo University*

w-li-3g4@sophia.ac.jp, aimizoguchi@maebashi-it.ac.jp, maho.morimoto.jp@gmail.com, arai@sophia.ac.jp

This study was supported by Sophia University Special Grant for Academic Research (Research in Priority Areas) and JSPS KAKENHI Grant Numbers JP19K13254, JP22J01381/22KJ2752, and JP23K12212.

2pSCb25

Background

- Word-initial voiced plosives in Japanese are reportedly becoming devoiced (Takada, 2011; Gao & Arai, 2019; Byun, 2021).
- While, traditionally, **Japanese is considered a voiced language**, with tongue root advancement and tongue body lowering observed in word-medial voiced plosives (Li et al., 2024, 2025).
- **Standard Chinese is an aspirating language** (Sato, 2024).
- Chinese learners of Japanese are said to substitute the voiced-voiceless contrast with an unaspirated-aspirated one (Hu, 2019).
- However, the articulatory mechanisms underlying this tendency remain unclear.

Research Questions

- **Do native Standard Chinese-speaking learners of Japanese exhibit tongue root advancement and tongue body lowering when producing Japanese plosives?**
- **If such tongue root advancement and tongue body lowering are observed, do they differ depending on the learners' proficiency level?**

Methods

Participants & Stimuli

- A total of 4 native speakers of Standard Chinese and 3 native speakers of Japanese were included in this analysis (see **Table 1**).
- Native speakers of Standard Chinese (CN) were instructed to produce 10 repetitions of each stimulus word shown in **Table 2**.
- Native speakers of Japanese (JP) were instructed to produce Japanese stimuli only.

Table 1: Participants

Speaker	Gender	Age	From	Japanese Proficiency
ACF02	F	23	Dalian, China	Advanced (JLPT: N1)
ACF03	F	23	Beijing, China	Beginner
ACM02	M	26	Inner Mongolia, China	Advanced (JLPT: N1)
ACM03	M	23	Beijing, China	Beginner
AJF01	F	20	Chiba, Japan	Native
AJF02	F	20	Kanagawa, Japan	Native
AJM01	M	20	Tokyo, Japan	Native

Table 2: Stimuli

Language	Target	Stimuli		Gloss
Chinese	/d/	/a.ta/	[ā.dà]	阿大 'eldest son'
	/t/	/k^{h}a.t^{h}a.ə/	[kǎ.tǎ.ěr]	卡塔尔 'Qatar'
	/g/	/a.kaŋ/	[ā.gāng]	阿刚 person's name
	medial /k/	/k^{h}a. k^{h}a/	[kā.kā]	咔咔 onomatopoeia
Japanese	/d, k/	/hadaka/	[hadaka]	裸 'naked'
	/g, t/	/hagata/	[hagata]	歯型 'tooth mold'

Note: As shown here, because we selected meaningful words as stimuli as much as possible, we were not able to fully control the acoustic environments surrounding the vowels. We examined the vowels' F1 and F2 values as well as the tongue position, but no significant effects were found for the reasons mentioned above. Due to space limitations, these results are not included in this poster.

Audio & Ultrasound Recording

- The audio signal was digitally recorded at 22,050 Hz, 16-bit resolution, using a RODE-NT2-A microphone.
- Mid-sagittal images of the oral cavity were recorded with an ultrasound system (MicrUS, EXT-1H) using a microconvex probe (MC10-5R10S-3). Video frame rate was 113fps.
- The audio signal and ultrasound video were recorded simultaneously and synchronized using AAA software (Articulate Instruments Ltd. 2022).

Analysis

Acoustic analysis:

- A total of 235 tokens from CN speakers and 60 tokens from JP speakers were segmented using Praat (Boersma & Weenink, 2023) for duration analysis (see **Figure 1** for an example).
- Instances with synchronization errors were excluded from the analysis.

Ultrasound image analysis:

- The mid-sagittal plane of the tongue contour of the maximal constriction point were tracked using GetContours (Tiede, 2022).
- The predicted tongue contours at the maximal constriction point from the repetitions were plotted by speaker and phoneme using a generalized additive model (GAM) (Wood, 2006) (see **Figure 2** for an example).
- Instances with synchronization errors or images too unclear for tracking were excluded from the analysis.

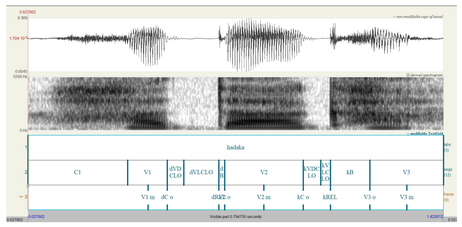

Figure 1 An example of annotation, /hadaka/ from ACF03

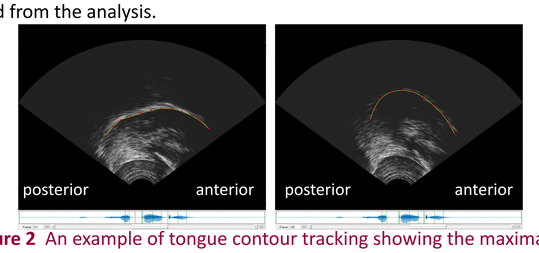

Figure 2 An example of tongue contour tracking showing the maximal constriction point frame of /d/(left) and /k/(right) in /hadaka/ from the same tokens as Figure 1

Results

Tongue Position at the Maximal Constriction Point of Plosive Consonants

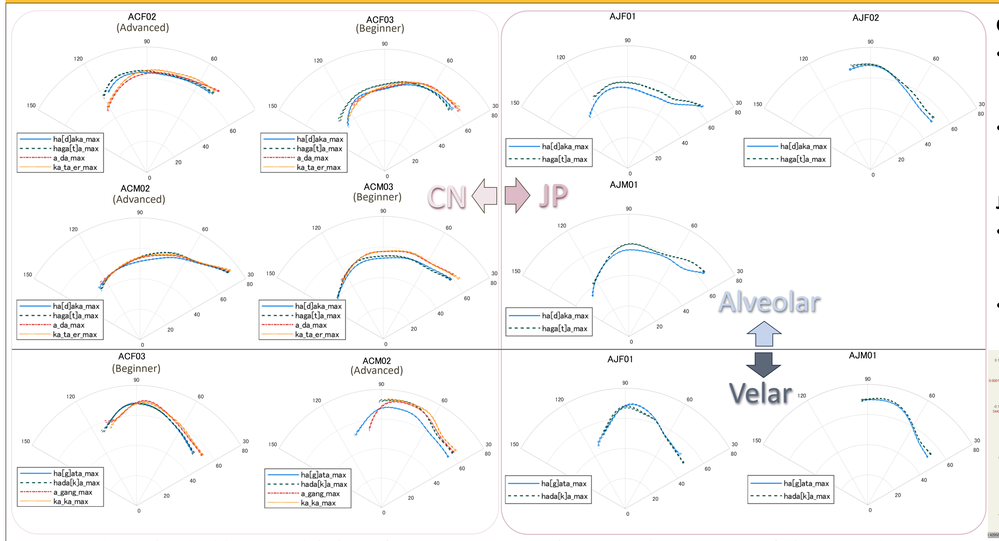

Figure 3 Generalized additive model plots of tongue position at the maximal constriction of Chinese and Japanese plosives. 95% confidence intervals are plotted as dashed lines. The legends are written in pinyin or romaji.

Chinese L1 speakers:

- Tongue body lowering / tongue root advancement observed in most Japanese plosives (except ACF03 velars).
- ACM02: lower tongue position for Japanese voiced velars; the only learner showing lenition in Japanese utterances.

Japanese L1 speakers:

- Alveolars: tongue lowering + root advancement observed as Li et al.(2024, 2025).
- Velars: smaller effects than previous reports, likely due to nasalized /g/.

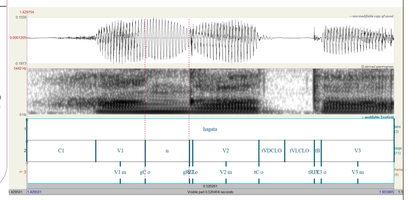

Figure 4 An example of /g/ with nasalized closure from AJF01.

Voicing Rate during Closure

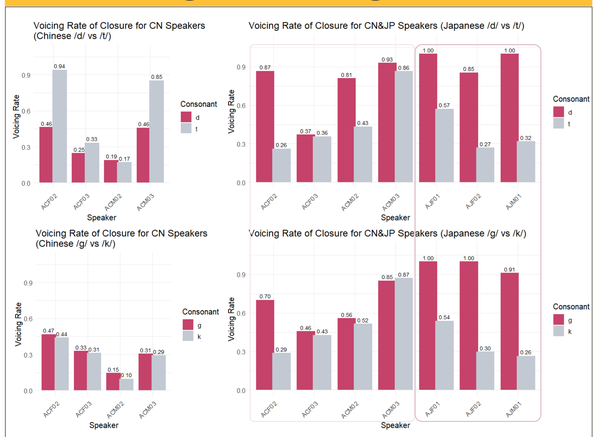

Figure 5 Voicing rate during closure for each speaker. Left: Chinese utterances; right: Japanese utterances.

Voicing or Aspirating?

Figure 6 Mean durations of voiced closure (VDCLO), voiceless closure (VLCLO), and burst interval (B) for each speaker. Left: Chinese utterances; right: Japanese utterances.

Conclusion

Summary:

- Japanese voiced plosive productions by Chinese-speaking learners showed some tongue body lowering and tongue root advancement regardless of proficiency level.
- However, only advanced learners (ACF02 & ACM02) used these articulatory adjustments to contribute to a voicing contrast.
- Beginner learners (ACF03 & ACM03) still rely on the presence or absence of aspiration in Chinese to produce Japanese plosives.

Limitations of this study:

- Small number of speakers.
- The acoustic environment surrounding the vowels was not fully control.

References:

[1] 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.

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

[3] 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.

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

[5] Li, W., Mizoguchi, A., Morimoto, M., & Arai, T. (2025). "Differences in Voicing Implementation and Tongue Movements between Japanese Intervocalic Voiced and Voiceless Stops." *Journal of the Phonetic Society of Japan* (Onsei Kenkyu), 29, 87–102.

[6] Sato, A. (2024). *Chūgokugo Onseigaku: Nihongo / Eigo to no Hatsuon Hikaku o Jūshi shite*[Chinese Phonetics: With a focus on comparison with Japanese and English pronunciation]. Asahi Press.

[7] Hu, W. (2019). Acquisition of Japanese Plosive by Chinese Learners: Focusing on Learners' Consciousness and Learning Resources. *Waseda studies in Japanese language education*, 26, 87–105.

[8] P. Boersma, D. Weenink, "Praat: Doing phonetics by computer. Version 6.3.16, retrieved 2023 from http://www.praat.org/," 2023.

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

[10] S. N. Wood, Generalized additive models: an introduction with R. chapman and hall/CRC, 2006.

Sixth Joint Meeting
Acoustical Society of America
And Acoustical Society of Japan
@Honolulu 2025.12.1~12.5