Spatio-Temporal Properties of Japanese Coronal Consonants: An Ultrasound Study of /d/ and /r/

Maho Morimoto[1], Takayuki Nagamine[2]

[1]*Sophia University/Japan Society for the Promotion of Science (Japan)*
[2]*Lancaster University (UK)*
maho.morimoto.jp@gmail.com, t.nagamine@lancaster.ac.uk

Abstract

The current study investigates the articulation of coronal consonants /d/ and /r/ in Japanese. Using ultrasound, we obtained midsagittal tongue images for /d/ and /r/ in three phonological contexts from one male Japanese speaker. Based on the tongue shapes, time-varying changes were analyzed quantitatively using the Principal Component Analysis (PCA). Results suggest that /d/ and /r/ may differ in terms of tongue retraction and dorsal stabilization, while also supporting previous results showing the effect of the surrounding environment. The study demonstrates that quantitative articulatory analysis combining ultrasound and PCA is a useful approach to the spatio-temporal characteristics of Japanese coronal consonants, with implications for future research.

Keywords: ultrasound, PCA, Japanese, coronal consonants, dynamic analysis

1. Introduction

The current study examines the articulatory characteristics of the liquid consonant in Japanese by comparing the spatio-temporal properties of the coronal consonants /d/ and /r/. While Japanese /r/ is canonically realized as an alveolar tap or flap [ɾ], it also shows a wide range of phonetic variation, including stop-like realizations such as [d] and [ɖ] (Arai 1999; Arai 2013; Vance 1987). In addition, the degree of similarity between /d/ and /r/ is reported to vary depending on the context (e.g., Arai 2013; Okada 1991). While some consider Japanese /r/ to be a 'weak [d]' (e.g., Kawakami 1977), others argue that Japanese /r/ is articulatorily different from /d/ in that /r/ involves a ballistic gesture (e.g., Akamatsu 1997).

Previous articulatory research seems to indicate that Japanese /r/ is not a 'weak [d]'. For example, previous studies demonstrate that Japanese /r/ shows a retracted place of articulation compared to coronal stops /t/ and /d/, using electropalatography (EPG; Kochetov 2018) and electromagnetic articulography (EMA; Morimoto 2020). Another EPG study also highlights substantial variability of the /r/ realizations across vowel contexts, while no such variability is mentioned for /d/ (Kawahara and Matsui 2017). However, the exact articulatory mechanisms underlying the similarities and differences between /d/ and /r/ are not well understood. This is especially true for the movements of the tongue dorsum, which are not well-captured using EPG or flesh-tracking methods like EMA, due to the limited amount of information available on the shape of the tongue.

The current study aims to complement the previous discussion regarding the similarity between /d/ and /r/ in Japanese. We use ultrasound tongue imaging to capture clear images of the tongue dorsum, whose behavior may differ depending on the vocalic context. Tokens of /d/ and /r/ are produced in three vowel environments to investigate the realizations of Japanese /d/ and /r/. We analyze how tongue shape changes over time, especially the tongue dorsum, which may provide insights into the articulatory differences between /d/ and /r/.

2. Methods

We report results from one 21-year-old male speaker from Tokyo. The participant produced Japanese words containing intervocalic /d/ and /r/ in three different phonological environments: /a_a/, /a_i/ and /aɴ_o/, as shown in **Table 1**. The speaker produced each token five times in random order, resulting in a total of 30 tokens of /d/ and /r/ for analysis.

Table 1: *List of words analyzed in this study.*

Consonant	Context	Word	Gloss
/d/	/a_a/	/ada/	avenge
/r/	/a_a/	/ara/	coarseness
/d/	/a_i/	/badiː/	body/buddy
/r/	/a_i/	/bariː/	Barry
/d/	/aɴ_o/	/kaɴdou/	sensation
/r/	/aɴ_o/	/kaɴro/	honeydew

We obtained audio recordings (at 22,050 Hz) and midsagittal ultrasound tongue images (at approximately 113 fps) using Articulate Assistant Advanced (AAA) version 221.0.0 (Articulate Instruments 2023). The probe was stabilized using an UltraFit headset to minimize undesirable probe movement (Spreafico, Pucher, and Matosova 2018).

Data analysis is based on acoustically-delimited intervals. We first automatically segmented /d/ and /r/ using Montreal Forced Aligner (McAuliffe et al. 2017), and then manually adjusted the boundaries wherever necessary using Praat (Boersma and Weenink 2022). Tongue splines were automatically fitted using the DeepLabCut (DLC) plug-in on AAA based on the acoustic consonantal intervals. DLC estimates tongue splines based on 11 *x*/*y* coordinates in each ultrasound frame. The tongue contour data were extracted at 11 equidistant time points during the target intervals of the consonants /d/ and /r/. The tongue splines were rotated and offset using the speaker's occlusal plane that we measured by having the speaker bite a thin plastic plate (Scobbie, Lawson, et al. 2011).

To identify the primary variation in midsagittal tongue movement in /d/ and /r/, we conducted a principal component analysis (PCA) using scripts publicly available from Nance and Kirkham (2022). PCA was run based on the *z*-normalized *x*/*y* coordinates from all tongue splines extracted for /d/ and /r/, and we tracked the time-varying changes of the first two PCs that accounted for the largest proportion of variance to visually inspect how tongue movement differs between /d/ and /r/.

3. Results

Figure 1 shows time-varying changes in the midsagittal tongue shapes for /d/ (left) and /r/ (right) across the three vowel environments during the consonantal intervals. While the tongue dorsum movement seems similar in the /a_a/ context, we observe a slight difference in the shape of the tongue body between the two consonants. In addition, some qualitative differences can be found in the /a_i/ and /aN_o/ contexts. Overall, tongue dorsum movement is smaller for /r/ than for /d/ in the /a_i/ context. Minor differences can also be found around the tongue dorsum in the /aN_o/ context. Finally, there is a difference in tongue tip variation in the /a_a/ and /aN_o/ contexts (note, however, that our methodology does not allow for a clear visualization of the tongue tip).

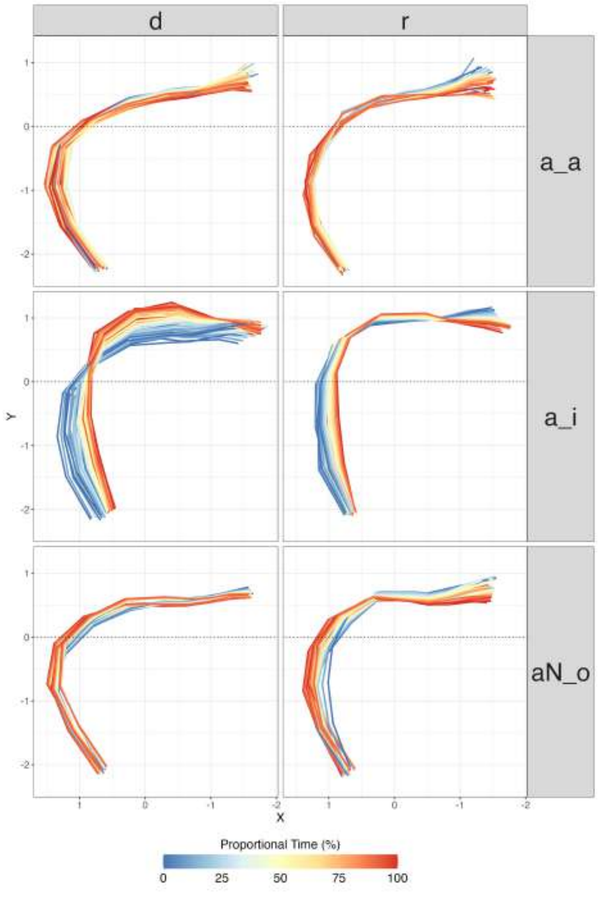

Figure 1: *Midsagittal tongue shapes in each frame during the consonant intervals in each vowel context for /d/ and /r/. Tongue tip points to the right.*

In order to explore the articulatory differences quantitatively, the results of PCA are shown in **Figure 2**. Variations explained by each principal component (PC) are superimposed on the midsagittal tongue shape, in which the mean tongue shape is represented with the bold line and the variation captured by each PC with the dashed (plus) and dotted (minus) lines by adding and subtracting a standard deviation associated with each PC from the mean tongue curve. We have found that the variation in the tongue motion of the two consonants can be described primarily in terms of two principal components, PC1 (76.85%) and PC2 (10.97%). In **Figure 2**, PC1 appears to capture the tongue retraction component at the tongue dorsum, correlated with the height of the tongue body. PC2 suggests a very subtle variation around the tongue body.

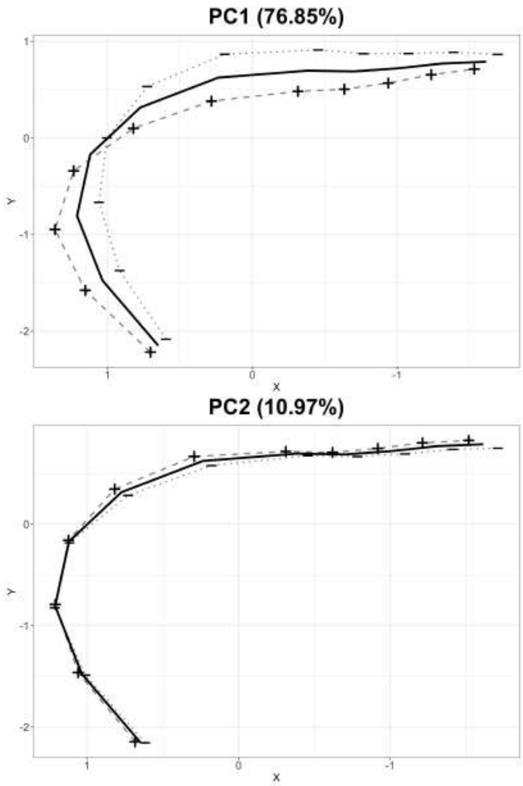

Figure 2: *Variation captured in PCs 1 and 2.*

Finally, **Figure 3** shows the changes in PC scores tracked during the consonantal intervals, allowing us to infer the articulatory movements along the PC dimensions. The consonant duration is normalized and expressed proportionally between 0% (consonantal onset) and 100% (consonantal offset). The thin lines represent PC changes of each token, with the thick lines smoothing them and the dotted lines showing the 95% confidence interval. The time-varying changes for PC1 (top three panels in **Figure 3**) show that the tongue dorsum for /r/ maintains a retracted tongue position when flanked by low vowels, while /d/ transitions from an anterior tongue dorsum position to one comparable to /r/ at the offset. In the /a_i/ context, we observe that the PC1 changes were relatively small for /r/ compared to /d/, which might suggest a dorsal stabilization mechanism for /r/. The PC1 changes for /d/ and /r/ in the /aN_o/ context are largely comparable with the two trajectories overlapping for the majority of consonantal intervals. Turning to PC2 (bottom three panels in **Figure 3**), the results suggest that the tongue body is slightly raised for /r/ across vocalic contexts. The difference in PC2 between /d/ and /r/ spans throughout the consonantal intervals.

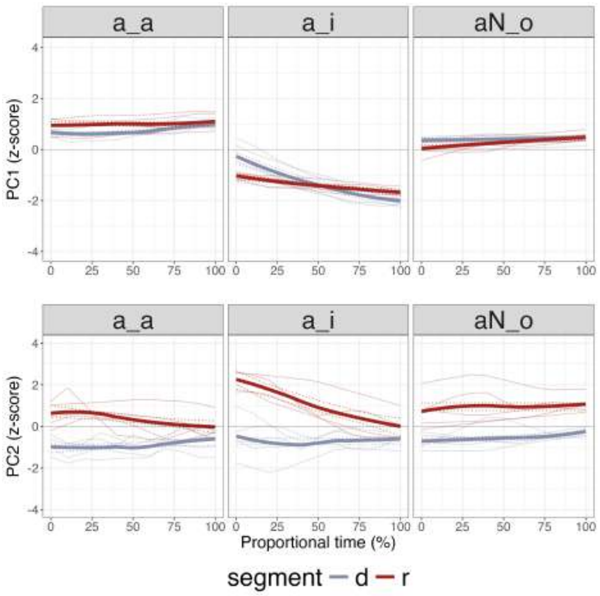

Figure 3: *Time-varying changes of each PC.*

4. Discussion and conclusion

The current study highlights some possible differences in the articulation of Japanese /d/ and /r/. First, we suggest that one of the key articulatory differences between /d/ and /r/ lies in tongue retraction and stabilization. The tongue retraction in /r/ is evident in the overall posterior tongue dorsum in the /a_a/ context. In addition, as seen in the midsagittal tongue shape and the dynamic changes in PC1 in **Figure 3**, the relative stability in the tongue dorsum position for /r/ in the /a_i/ context points to some dorsal stabilization mechanism of /r/, while /d/ is more susceptible to vowel coarticulation.

The similarity in the degree of tongue retraction in /d/ and /r/ in the /aN_o/ context seems to be in line with previous findings reporting that liquids are sometimes replaced by plosives after coda nasals in child speech, although the particular instance provided was of post-nasal /r/ replaced by /g/ (Arai 2013). This similarity may be explained by the durational differences among the phonological environments. While /d/ was generally longer than /r/ overall, we find that the duration of /d/ was quite short and thus comparable with /r/ (around 28 ms overall) in the post-nasal environment, as illustrated in **Figure 4**. We also observed this in one token of /d/ in the /a_i/ context. Spectrographic representation of this token suggests that this is an instance of the lenition of /d/, and we intend to explore the relationship between duration, lenition, and the similarity between liquids and plosives in different phonological contexts in future research. Finally, the slight raising of the tongue body in /r/ as suggested by PC2 may be a by-product of tongue body compression as a result of tip retraction in /r/, which could be indicative of the difference in the manner requirements for /d/ and /r/.

While it is based on a small number of tokens, the current study demonstrates that ultrasound paired with PCA allows us to investigate the articulatory mechanisms of coronal consonants. The current results seem to indicate a more stable dorsal movement for /r/ than for /d/, especially in the /a_i/ context. This could reflect dorsal stabilization as a unifying articulatory characteristic of liquid consonants (Proctor 2011), but this possibility needs to be further evaluated in future research as it is also possible that it is a result of specific manner requirements (Recasens 2016). Methodologically, note also that the vowel environment /a_i/ may exhibit a joint effect of the tongue movement and jaw displacement, making this observation inconclusive. Since the probe tracks the movement of the lower jaw, the transition of the tongue position from one vowel to another needs to be evaluated with caution (Scobbie, Wrench, and van der Linden 2008).

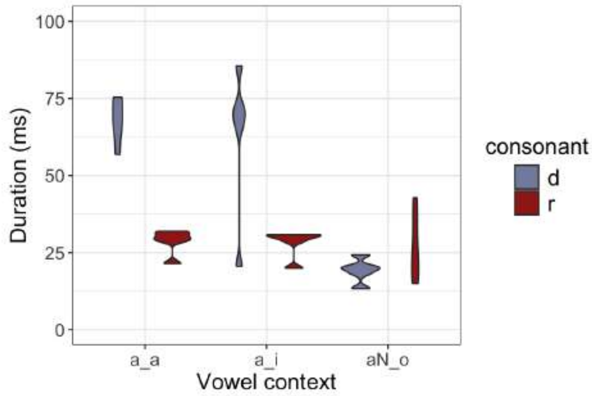

Figure 4: *Duration (ms) of the acoustic constriction of each consonant.*

Nevertheless, we believe that the dynamic analysis on dorsal movement in this study is promising in identifying what articulatory mechanisms could distinguish Japanese /r/ from the coronal consonant /d/. Future research will incorporate a larger number of speakers, as the current study is based on a small number of tokens produced by a single speaker. It would also be necessary to examine the productions of /d/ and /r/ in a wider variety of contexts, as articulation of /r/ is known to be largely influenced by prosodic positions and adjacent vowels (Yamane, Howson, and Wei 2015; Maekawa 2023). Our results also suggest the need to consider the prosodic position and its effect on the duration and lenition of /d/. Furthermore, in controlling the vowel environments, we would need to take into account the dynamic jaw movement mentioned above.

To conclude, the current study provides a preliminary articulatory description of Japanese /d/ and /r/ based on ultrasound data. The results suggest that Japanese /r/ may not involve the same articulatory mechanism as /d/, highlighting key differences in the degree of tongue retraction and stabilization. Based on our observartions, we tentatively argue that Japanese /r/ is not a ‘weak [d]’. The limitations of the study offer important implications for future research, which will help to achieve a better articulatory characterization of Japanese coronal consonants.

5. Acknowledgements

This study was supported by JSPS KAKENHI Grant Number JP20K21979. We thank the participant for taking part in the experiment, and Professor Takayuki Arai and the Speech Communication Lab at Sophia University for their support. We are also grateful to the anonymous reviewers of our conference abstract for pointing out, among other issues and considerations, the dynamic displacement of the probe related to jaw lowering.

6. References

Akamatsu, T. (1997). *Japanese Phonetics: Theory and Practice*. München, Newcastle: Lincom Europa.

Arai, T. (1999). "A case study of spontaneous speech in Japanese". In: *Proceedings of the 14th International Congress of Phonetic Sciences*. Ed. by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey. San Francisco: ICPhS Archive, pp. 615–618.

— (2013). "On why Japanese /r/ sounds are difficult for children to acquire". In: *Interspeech 2013*. Lyon, France, pp. 2445–2449.

Articulate Instruments (2023). *Articulate Assistant Advanced version 221.0.0*. Articulate Instruments. Edinburgh.

Boersma, P. and D. Weenink (2022). *Praat: Doing Phonetics by Computer version 6.2.19*.

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

Kawakami, S. (1977). *Nihongo Onsei Gaisetsu [Outline of Japanese Phonetics]*. Tokyo: Ofusha.

Kochetov, A. (2018). "Linguopalatal contact contrasts in the production of Japanese consonants: Electropalatographic data from five speakers". In: *Acoustical Science and Technology* 39.2, pp. 84–91. DOI: 10.1250/ast.39.84.

Maekawa, K. (2023). "Articulatory characteristics of the Japanese /r/: A real-time MRI study". In: *Proceedings of the 20th International Congress of Phonetic Sciences*. Ed. by R. Skarnitzl and J. Volín. Prague: Guarant International, pp. 992–996.

McAuliffe, M., M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger (2017). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi". In: *Interspeech 2017*. ISCA, pp. 498–502. DOI: 10.21437/Interspeech.2017-1386.

Morimoto, M. (2020). "Geminated liquids in Japanese: A production study". PhD thesis. University of California Santa Cruz.

Nance, C. and S. Kirkham (2022). "Phonetic typology and articulatory constraints: The realization of secondary articulations in Scottish Gaelic rhotics". In: *Language*, pp. 419–460.

Okada, H. (1991). "Japanese". In: *Journal of the International Phonetic Association* 21.2, pp. 94–96.

Proctor, M. (2011). "Towards a gestural characterization of liquids: Evidence from Spanish and Russian". In: *Laboratory Phonology* 2.2, pp. 451–485. DOI: 10.1515/labphon.2011.017.

Recasens, D. (2016). "What is and what is not an articulatory gesture in speech production". In: *Gradus - Revista Brasileira de Fonologia de Laboratório* 1, pp. 23–42. DOI: 10.47627/gradus.v1i1.101.

Scobbie, J., E. Lawson, S. Cowen, J. Cleland, and A. Wrench (2011). "A common co-ordinate system for mid-sagittal articulatory measurement". In: *QMU CASL Working Papers* 20, pp. 1–4.

Scobbie, J., A. A. Wrench, and M. van der Linden (2008). "Head-Probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement". In: *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg, France, pp. 373–376.

Spreafico, L., M. Pucher, and A. Matosova (2018). "UltraFit: A speaker-friendly headset for ultrasound recordings in speech science". In: *Interspeech 2018*. ISCA, pp. 1517–1520. DOI: 10.21437/Interspeech.2018-995.

Vance, T. J. (1987). *An Introduction to Japanese Phonology*. State University of New York Press.

Yamane, N., P. Howson, and P.-C. G. Wei (2015). "An ultrasound examination of taps in Japanese". In: *Proceedings of the 18th International Congress of Phonetic Sciences*. Ed. by The Scottish Consortium for ICPhS 2015. Glasgow, UK: The International Phonetic Association, pp. 1–5.