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
[2]Lancaster University
maho.morimoto.jp@gmail.com, t.nagamine@lancaster.ac.uk

Introduction. Previous literature on Japanese consonants has noted the similarity of the liquid consonant /r/ and the alveolar plosive consonant /d/. While researchers generally agree that there are several variants of /r/ including one like a 'weak [d]' (e.g., Kawakami 1977), some argue that /r/ is articulatorily different from /d/ in that /r/ involves a ballistic gesture (e.g., Akamatsu 1997). It has also been pointed out that there is a varying degree of similarity between these two consonants depending on the context, with more similarity exhibited in phrase-initial and post-nasal environments (e.g., Arai 2013). While several studies using electropalatograhy (EPG) have demonstrated that /r/ is not just a short version of /d/ (e.g., Kawahara et al. 2017; Kochetov 2017), the exact articulatory mechanisms underlying the similarities and differences between /r/ and /d/ are not well understood, especially of the tongue dorsum movements that are not well-captured using EPG. In this study, we investigate articulatory differences between /d/ and /r/ in Japanese. We use ultrasound to obtain clear images of tongue dorsum, whose behavior might differ depending on the vocalic context.

Methods. We report results from one 21-year-old male speaker from Tokyo. The participant produced the following Japanese words containing intervocalic /d/ and /r/: /ada/ ("avenge"), /ara/ ("coarseness"), /badiː/ ("body/buddy"), /bariː/ ("Barry"), /kaNdou/ ("sensation"), /kaNro/ ("honeydew"). They were repeated five times in random order, resulting in a total of 30 tokens of /d/ and /r/ for analysis. 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 a headset (Spreafico et al. 2018). Prior to analysis, we 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 was extracted at 11 equidistant time points during the target 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 et al. 2011). To identify primary variation in midsagittal tongue movement in /d/ and /r/, we ran principal components analysis (PCA) using scripts publicly available from Nance and Kirkham (2021). PCA was run based on the *z*-scored 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/.

Results. The left panel in **Figure 1** shows time-varying changes in midsagittal tongue shapes for /d/ (top) and /r/ (bottom). The results of PCA are shown in the right panel in **Figure 1**. Variations explained by PCs 1 and 2 (top right) are superimposed on the tongue midsagittal tongue shape, in which the mean tongue shape is represented with the bold line and the variation captured by each PC with dashed (plus) and dotted (minus) lines by adding and subtracting a standard deviation associated with each PC from the mean tongue curve. Shown in bottom right in **Figure 1** are time-varying changes in each PC dimension during each consonant. The consonant duration is normalized and expressed proportionally between 0% (consonantal onset) and 100% (consonantal offset). Thin lines represent PC changes of each token, with the thick lines smoothing them and the dotted lines showing the 95% confidence interval.

The midsagittal tongue shape in the left panel in **Figure 1** suggests that there are some differences in the dynamic spatial properties of /d/ and /r/. We have found that the variation in the tongue motion of the two consonants can be described in terms of two principal components, PC1 (76.85%) and PC2 (10.97%). In **Figure 1**, PC1 appears to capture the tongue retraction component at the tongue dorsum, correlated with the height of the tongue body. The time-varying changes for PC1 (bottom right) 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 intervocalic position /a_i/, 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 /aN_o/ context are largely comparable with the two trajectories overlapping for the majority of consonantal intervals. PC2, on the other hand, suggests a very slight variation around the tongue body, which is slightly raised for /r/ across vocalic contexts. The difference in PC2 between /d/ and /r/ spans throughout the consonantal intervals.

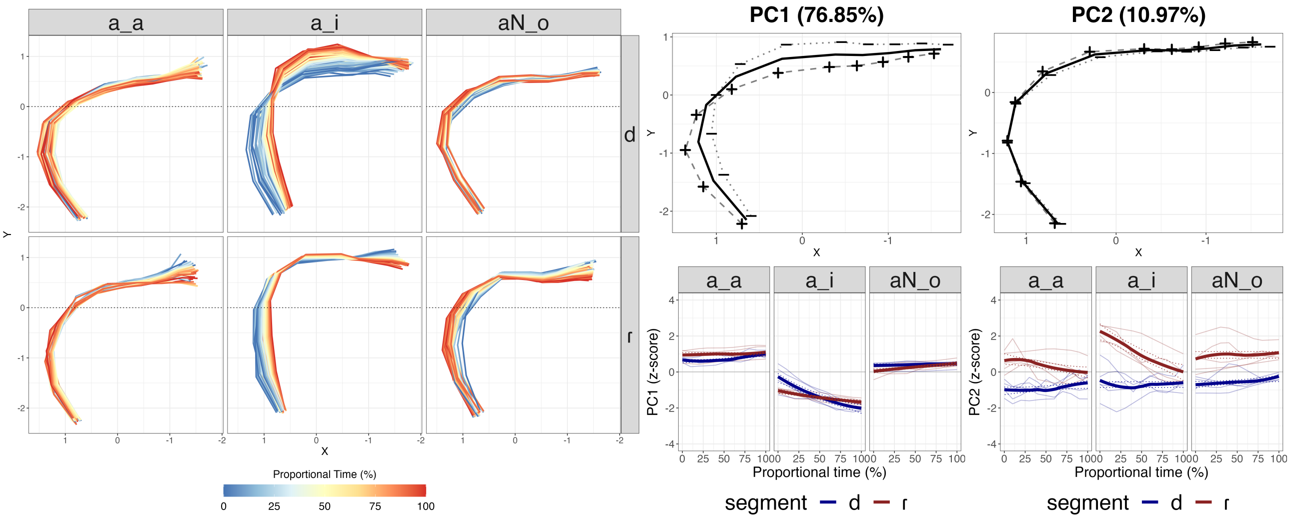

Figure 1: *Left: midsagittal tongue shapes during the consonant intervals in each vowel context for /d/ and /r/. Tongue tip points to the right; Right: variation captured in PCs 1 and 2 (top) and time-varying changes of each PC (bottom).*

Discussion. The current study highlights the 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 /a_a/ context. In addition, as seen in the midsagittal tongue shape and the dynamic changes in PC1 in **Figure 1**, the relative stability in the tongue dorsum position for /r/ in /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 /aN_o/ context is consistent with previous literature pointing out that the two consonants are especially confusable after coda nasals. It is noteworthy that the duration of the two consonants was also largely comparable in this environment, while /d/ was generally longer than /r/ in other contexts. 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 might result from different manner requirements for /d/ and /r/.

Although it only considers a small number of tokens, the current study demonstrates that ultrasound paired with PCA allows us to better investigate articulatory mechanisms of Japanese coronal consonants. Future research will incorporate a larger number of speakers as the current study is based on the productions of a single speaker. In addition, it is necessary to examine the productions of /d/ and /r/ in a wider variety of contexts, as articulation of /d/ and /r/ may be influenced by prosodic positions and adjacent vowels (Yamane et al. 2015; Maekawa 2023). Note also that the current methodology may not fully account for tongue tip movement or jaw displacement. These considerations will help to better characterize the articulation of Japanese coronal consonants.

References

Akamatsu, T. (1997). “Japanese phonetics: Theory and practice (Vol. 3)”. Lincom Europa.

Arai, T. (2013). “On why Japanese /r/ sounds are difficult for children to acquire”. In: *Interspeech 2013.* Lyon, France: International Speech Communication Association, pp. 2445–2449.

Articulate Instruments. (2023). “Articulate Assistant Advanced version 221.0.0”. [Computer software]. Edinburgh, Articulate Instruments.

Boersma, P., & Weenink, D. (2022). “Praat: Doing phonetics by computer version 6.2.19”. [Computer software].

Kawahara, S., Matsui, M., & Shaw, J. (2017). “Some aspects of Japanese consonant articulation: A preliminary EPG study”. In: *ICU Working Papers in Linguistics II*, pp. 1–12.

Kawakami, S. (1977). “Nihongo onsei gaisetsu [Outline of Japanese phonetics]”. Tokyo: Ōfūsha.

Kochetov, A. (2017). “Linguopalatal contact contrasts in the production of Japanese consonants: Electropalatographic data from five speakers”. In: *Acoustical Science and Technology* 39.2, pp. 84–91.

Maekawa, K. (2023). “Articulatory characteristics of the Japanese /r/: A real-time MRI study”. In: *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague: Guarant International, pp. 992–996.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). “Montreal Forced Aligner: Trainable text-speech alignment using Kaldi”. In: *Interspeech 2017*. Stockholm, Sweden: International Speech Communication Association, pp. 498–502.

Nance, C., & Kirkham, S. (2022). “Phonetic typology and articulatory constraints: The realization of secondary articulations in Scottish Gaelic rhotics”. In: *Language* 98.3, pp. 419–460.

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

Spreafico, L., Pucher, M., & Matosova, A. (2018). “UltraFit: A speaker-friendly headset for ultrasound recordings in speech science.” In: *Interspeech 2018*. Hyderabad, India: International Speech Communication Association, pp. 1517–1520.

Yamane, N., Howson, P., & Wei, P.-C. G. (2015). “An ultrasound examination of taps in Japanese”. In: *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The International Phonetic Association, pp. 1–5.