ARTICULATORY CHARACTERISTICS OF THE JAPANESE /r/: A REAL-TIME MRI STUDY

Kikuo Maekawa
National Institute for Japanese Language and Linguistics
kikuo@ninjal.ac.jp

ABSTRACT

Japanese /r/ is described as a voiced post-alveolar flap in the Handbook of the IPA. This description raises two questions. The validities of the exclusion of the alveolar articulation, and the exclusion of the "tap" articulation as distinguished from the "flap". To clarify these issues, articulation of the Japanese /r/ was analysed using the Real-time MRI Articulatory Movement Database. Regarding the manner of articulation, no "flap" articulation was observed in all 167 samples of the /r/ covering both word-initial and word-medial positions. Regarding the place of articulation, it turned out that the place of articulation of the Japanese /r/ did not cover the post-alveolar region. Moreover, leave-one-out cross-validation revealed that the variation of the place of articulation of the Japanese /r/ could be predicted fairly well by knowing the categories of the adjacent vowels.

Keywords: Japanese /r/, Real-time MRI Articulatory Movement Database, place and manner of articulation

1. INTRODUCTION

In the Handbook of the IPA [1], the Japanese /r/ is described as a "voiced post-alveolar flap" and represented by the phonetic symbol [ɽ] [2]. This description by Okada is problematic at least in two respects ignoring the curious choice of the phonetic symbol. First, concerning the place of articulation (or PoA), its exclusion of the alveolar / dental articulation. Second, concerning the manner of articulation (or MoA), its exclusion of the "tap" articulation as distinguished from the "flap" articulation. Although Okada's description doesn't have wide support from the specialists of Japanese, it is accepted by some non-specialists on the authority of the IPA. This paper examines the PoA and MoA of the Japanese /r/ objectively by using the data of real-time MRI movies.

2. DATA AND METHOD

The data analysed below are all taken from the Real-time MRI Articulatory Movement Database, or rtMRIDB (https://rtmridb.ninjal.ac.jp). This database is a collection of about 26,000 Japanese utterances spoken by 22 native Japanese speakers; it covers materials like mora-unigram, mora-bigram, words containing mora-phonemes, reading aloud of short passages, and so forth. Articulatory movements in the vocal tract are observed in the mid-sagittal plane with a frame rate of about 14 or 27 fps (frames per second). Each frame consists of 256 × 256 pixels (or voxels) with a pixel resolution of 1 mm and a slice thickness of 10 mm. The recording was done using the 3 T MRI scanner (Siemens MAGNETOM Prisma) at the Brain Activity Imaging Center of ATR Promotion Inc., in Kyoto, Japan, in the years 2017–2021.

167 samples of /r/ in the word- (or mora-) initial and word-medial positions were extracted from the rtMRIDB. The initial samples include /ra/, /ri/, /ru/, /re/, /ro/, /rja/, /rju/, /rjo/, and /rje/, while the medial samples include /ara/, /ari/, /aru/, /are/, /aro/, /ira/, /iri/, /iru/, /ire/, /iro/, /ora/, /ori/, /oru/, /ore/, and /oro/. These samples were uttered by seven male speakers of Standard Japanese and recorded in 27 fps. Note that the symbol /rj/ stands for the palatalized version of the /r/.

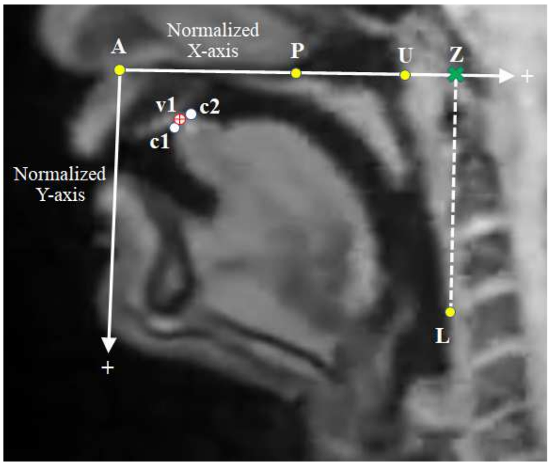

Figure 1: Measurement points and normalized X- and Y-axes.

Figure 1 shows the measurement points and normalized XY-axes used in this study. Points c1 and c2 are the edges (beginning and end) of the vocal tract closure for /r/ (or the constriction of /s/, see 3.1). Point v1 (with a red cross) is the tip of the tongue. A and P stand respectively for the anterior and posterior nasal spines. U is the point where the line connecting A and P reaches the pharyngeal wall. L is the point on the pharyngeal wall corresponding to the boundary between the third and fourth cervical vertebrae. These points were measured for all frames corresponding to the timing of the vocal tract closure by the /r/. In addition, c1, c2, and v1 are also measured for the frame that precedes the /r/ (the "preceding frame"). Note that the relative location among the c1, c2, and v1 differs depending on the MoA of the /r/ (See 3.2).

The XY-values of the c1, c2, and v1 were normalized as follows. The origin of the normalized coordinate is on the A. The normalized x-axis is the line connecting the A, P, and U. Y-axis is perpendicular to the X-axis. The unit of measurement is the distance between the A and U for the X-axis, and the distance between the L and Z for the Y-axis. Z is the point where the line starting from point L (broken line in Figure 1) intersects perpendicularly with the normalized X-axis. This method is nearly identical to that proposed in [3]; see the literature for the effectiveness of the proposed method.

3. ANALYSIS

3.1. Place of articulation (PoA)

To evaluate the PoA of the /r/, the distributions of the c1 and c2 were compared to those of the word-initial /s/. Japanese /s/ is phonetically realized as a voiceless alveolar fricative [s] before the vowels /a, u, e, o/, but is realized as an alveolo-palatal fricative [ɕ] (or post-alveolar [ʃ]) before the /i/. Phonologically palatalized /s/, represented as /sj/ and appearing before /a/, /o/, and /u/ is also realized as [ɕ] (or [ʃ]). Figure 2 compares the distributions of the c1 and c2 of the plain and palatalized /s/ (the latter including /si/) and /r/. The distribution range of the plain /r/ turns out to be almost identical to that of plain /s/ and more frontal than palatalized /sj/. One more finding about the PoA is that the ranges of distribution are nearly the same in the plain /r/ and palatalized /rj/. This issue will be discussed in section 5.

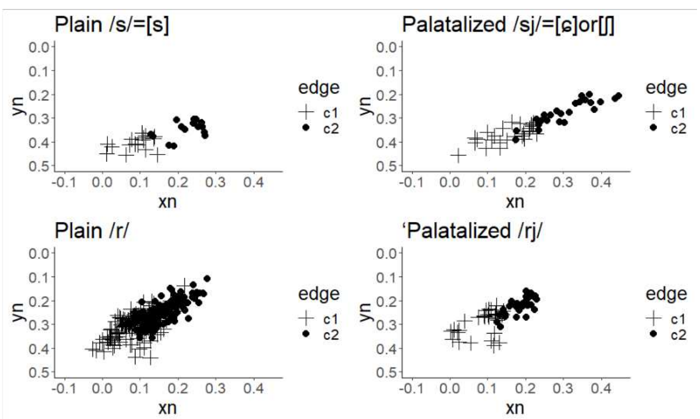

Figure 2: Comparison of the constriction/closure locations between /s/ and /r/. Separate panels are prepared for plain and palatalized /s/ and /r/. "xn" and "yn" stand respectively for "normalized x" and "normalized y" values.

3.2. Manner of articulation (MoA)

The analysis of the MoA is concerned with Okada's distinction between the "tap" and "flap" articulations, which are not distinguished in the IPA chart of pulmonic consonants. In this paper, following the proposals by [4] and [5], the "flap" is defined as the articulation where the vocal tract closure occurs between the palate and the "under-side of the tongue". On the other hand, in the "tap" articulation, the upper surface of the tongue is used for the closure.

Figure 3 compares a typical "flap" /r/ as uttered by a phonetician (panel A) and two variants of "tap" articulation (B and C). Points c1, c2, and v1 are shown in the same manner as in Figure 1. Both the c1 and c2 are to the left of the v1 in the "flap" articulation. In the "tap" articulation, c2 locates always to the right of the v1, but the relative location of the v1 and c1 is variable. Sometimes, c1 is to the right of the v1 (panel B), but there is also the case where the c1 locates to the left of the v1 (panel C).

Figure 4 consists of two histograms showing the distributions of the differences of the X-values between the v1 and c1 (abb. v1-c1), and those between the v1 and c2 (v1-c2), across the initial and medial /r/ positions. The v1-c2 values are positive in

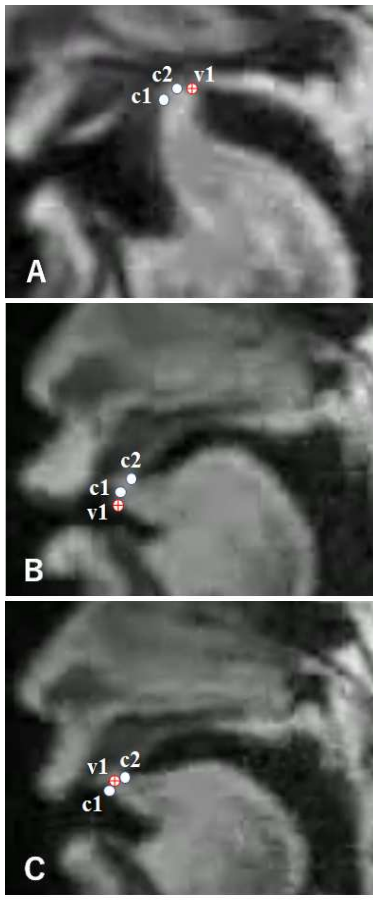

Figure 3: A typical "flap" (A) and two variants of "tap" articulation (B, C).

all samples, indicating that c2 is always to the right of the v1. And the v1-c1 values distribute on both sides of the zero. These properties are the ones expected only in the "tap" articulation like panels B and C of Figure 3.

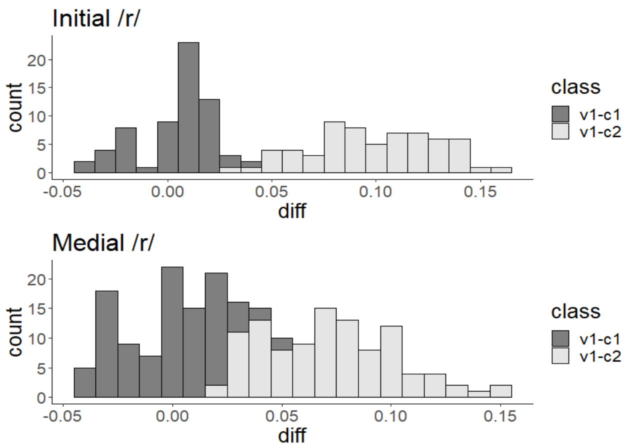

Figure 4: Histograms of v1 – c1 and v1-c1 in initial- and medial-/r/.

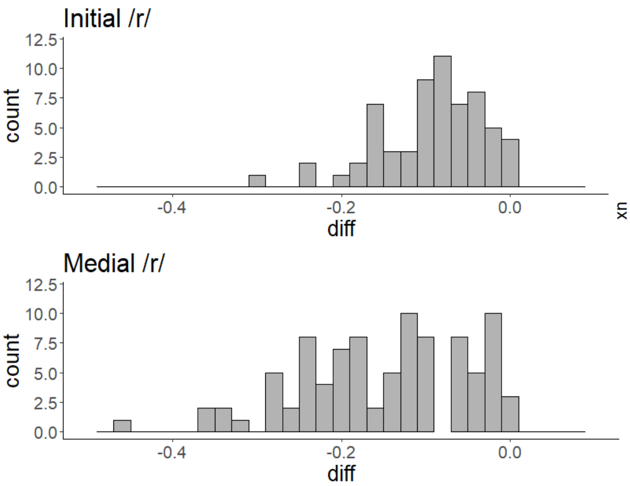

Figure 5: Histograms of the vertical displacement of the tongue tip between the timing of /r/ closure and the preceding frame.

Additional analysis of MoA was conducted concerning the vertical displacement of the tongue tip between the timing of /r/ closure and the preceding frame (see section 2). Figure 5 compares the distributions of the difference between the Y-values of the tongue tip in the frame corresponding to the /r/ closure (v1) and that of the preceding frame (pv1), i.e., v1 – pv1. As can be seen from the figure, all sample values have negative values, meaning that the tongue tip moved upward from the preceding frame to the /r/ in all samples (Note that lower and higher articulator locations correspond respectively to larger and smaller Y-values). Here again, this is what is expected in the "tap" rather than "flap" articulation.

Variables	c1x			c2x		
	Estimate	p		Estimate	p	
Intercept	0.050	0.013	***	0.125	0.000	***
pvwl_a	0.038	0.008	**	0.040	0.003	**
pvwl_o	0.050	0.008	***	0.050	0.000	***
fvwl_e	0.016	0.008		0.019	0.141	
fvwl_u	0.018	0.202		0.014	0.262	
fvwl_a	0.014	0.313		0.017	0.194	
fvwl_o	0.026	0.065	.	0.026	0.049	*
pvwl_a : fvwl_e	-0.023	0.244		-0.022	0.230	
pvwl_o : fvwl_e	-0.014	0.473		-0.015	0.418	
pvwl_a : fvwl_u	-0.007	0.728		0.004	0.812	
pvwl_o : fvwl_u	-0.013	0.500		-0.017	0.350	
pvwl_a : fvwl_a	-0.015	0.457		-0.009	0.628	
pvwl_o : fvwl_a	-0.007	0.727		-0.020	0.277	
pvwl_a : fvwl_o	-0.008	0.668		-0.020	0.280	
pvwl_o : fvwl_o	0.001	0.963		-0.020	0.267	

'***' p<.001 '**' p<.01 '*' p<.05 '.' p<.1

Table 1: Summary of the GLMM analyses

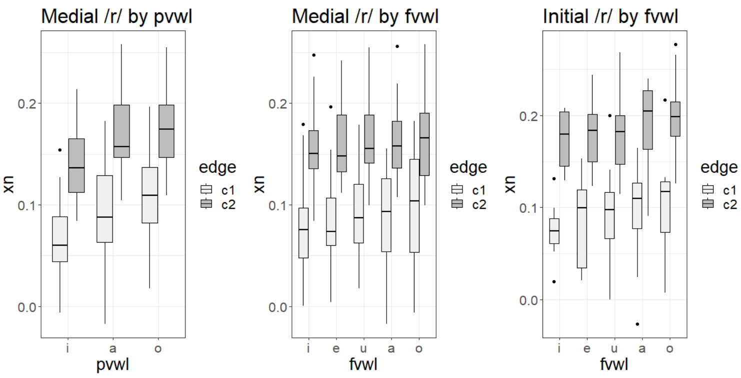

Figure 6: Effects of the preceding and following vowels on the X-values of c1 and c2.

3.3. Variation of the PoA

To examine the effects of adjacent segments on the variation of the PoA (Figure 2), Generalized Linear Mixed-effect Model (GLMM) regression analyses were conducted using the lme4 [6] and lmerTest [7] in R [8]. The X-values of the c1 and c2 of medial /r/ were predicted using the preceding vowels (/i/, /a/, or /o/, abb. "pvwl"), the following vowels (/a/, /i/, /u/, /e/, or /o/, abb. "fvwl"), and all interactions between the pvwl and fvwl. In addition, the subject was used as a random effect on the intercept. The result is summarized in Table 1. The effects of the pvwl are all significant (p<.05), while the effect of the fvwl and all interactions were not significant with the sole exception of the fvwl /o/. Note here the significances were computed between the first member of the factor level (i.e., /i/) and each of the remaining members (i.e., /i/ vs. /a/ and /i/ vs. /o/ in the case of the pvwl). Similar analysis of the effect of fvwl on the initial /r/ revealed that following /o/ and /a/ had significant effects on both c1x and c2x at the levels of 0.05 or less. Figure 6 is a box-whisker representation of the effects of the pvwl and fvwl on the x values of the medial and initial /r/.

4. MODELING

Analyses in sections 3.1 and 3.3 revealed that the PoA of the Japanese /r/ has a more restricted distribution than Japanese /s/, but the distribution varies by the influence of the adjacent vowels. To know how well it is possible to predict the variation of the PoA by the adjacent vowels, GLMM modelling was conducted. X-values of the c1 and c2 were predicted by four GLMM models and the predictive performance of the models was evaluated by leave-one-out cross-validation (LOOCV).

Model	/r/ positon	LMER models	Corr. coef.	Mean Error
1	Medial	~pvwl + fvwl + edge + (1|subject)	0.898	0.021
2	Medial	~pvwl + edge + (1|subject)	0.897	0.020
3	Medial	~fvwl + edge + (1|subject)	0.845	0.025
4	Initial	~fvwl + edge + (1|subject)	0.933	0.019

Table 2: Summary of the LOOCV

Table 2 summarizes the results. The five columns of the table show respectively the model ID, the position of the target /r/ (either medial or initial), the model used for the prediction (formulae notation is in the lmer format), coefficient of product-moment correlation between the observed and predicted values, and the mean (root-mean-square) prediction error. Note, in the third column, the notation "(1|subject)" means that the subject is used as a random effect, and "edge" is a variable that takes binary values of either 'c1' or 'c2'. This variable was introduced to estimate the c1 and c2 simultaneously.

All models worked well. Especially, with models 1, 2, and 4 it was possible to explain more than 80% of the variance of the c1 and c2 on the X-axes. Figure 7 is the scatter plots of the observed and predicted values in cases of models 2 and 4.

5. DISCUSSION AND CONCLUSION

Sections 3 and 4 provide answers to the questions raised in section 1. As for the MoA, it is mistaken to regard the Japanese /r/ as a "flap" consonant to the exclusion of "tap" articulation. All samples of the /r/ analysed in this study were realized as "taps" (Figures 4, 5). Although there remains some possibility that the /r/ is realized as a "flap" consonant in natural settings, such articulation is unlikely to be observed at least in experimental settings like the recording of the rtMRIDB. It is not completely clear from the literature [2] if Okada really excluded the possibility of "tap" articulation, but his preference to the [ɽ] symbol strongly suggests the flap-only interpretation, since it is widely believed that the lower surface of the tongue tip makes closure in articulating typical retroflex consonants [4,5].

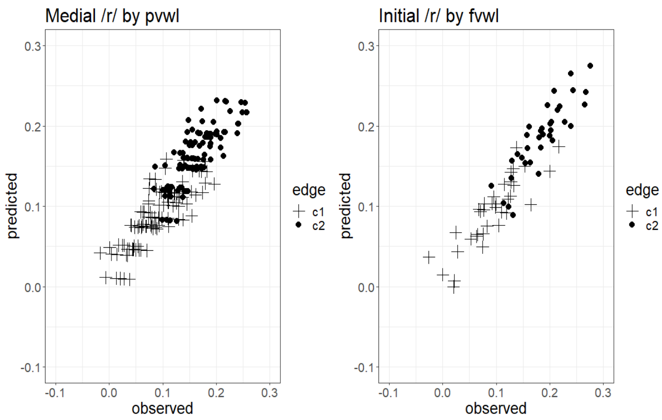

Figure 7: Scatter plots of the predicted and observed X-values of c1 and c2. Models 2 (left) and 4 (right).

As for the PoA, it is clearly mistaken to regard the Japanese /r/ as a post-alveolar sound. /r/ is more frontal than palatalized /s/ (Figure 2). This is contradictory to the description of [2] that regards the Japanese /r/ as a post-alveolar consonant. It is rather an alveolar sound to the exclusion of the post-alveolar articulation. It is, however, unclear if the Japanese /r/ needs to be phonologically specified as an alveolar consonant. As shown by Tables 1, 2, and Figures 6, 7 the precise PoA of the /r/ is almost entirely determined by its adjacent vowels. One possible interpretation is that the /r/ does not have the specification of the PoA as suggested by [9] in somewhat different context. It may be the case that a simple tapping of the tongue tip is superimposed on the global movement of the tongue determined by the adjacent vowels. According to this view, the place of contact between the tongue tip and the palate is determined not by the /r/ but by the adjacent vowel(s). The view that sees consonant articulation as a superimposed event on the articulation of a sequence of vowels is not a novel one; see [10] and its modern version [11], among others.

Moreover, the underspecification view provides a simple explanation for the finding that palatalization does not affect the PoA of the /r/ (Figure 2). Because palatalization is concerned with raising the front of the tongue, it does not affect the movement of the tongue tip and the PoA thereby.

To conclude, putting the issue of phonological underspecification aside, the current description of the Japanese /r/ in the IPA Handbook needs to be revised entirely.

Acknowledgement: This work is supported by the JSPS KAKENHI grants 17H02339 and 20H01265. It is also supported by a grant from the National Institutes for the Humanities.

6. REFERENCES

[1] International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*." Cambridge. Univ. Press.

[2] Okada, H. (1999). "Japanese". In International Phonetic Association (1999).

[3] Maekawa, K. (2021). "Production of the utterance-final moraic nasal in Japanese: A real-time MRI study". *Journal of the International Phonetic Association*, doi:10.1017/S0025100321000050

[4] Catford, J.C. (1977). *Fundamental Problems in Phonetics*. Edinburgh Univ. Press.

[5] Ladefoged, P. & I. Maddieson (1996). *The Sounds of the World's Languages*. Blackwell.

[6] Bates D., Mächler M., Bolker B., & Walker S. (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.

[7] Kuznetsova, A., Brockhoff, P.B., Christensen, R. H. B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13.

[8] R Core Team (2022). R: A language and environment for statistical computing. URL https://www.R-project.org/.

[9] Mester, R.A. & J. Ito (1989). "Feature predictability and underspecification: Palatal prosody in Japanese mimetics". *Language*, 65 (2), 258-193.

[10] Öhman, S. E. G. (1966). "Coarticulation in VCV utterances: Spectrographic Measurements". *JASA,* 39 (1), 151-168.

[11] Fujimura, O. (2000). "The C/D model and prosodic control of articulatory behavior". *Phonetica*, 57, 128-138.