Articulation of North American English /r/ by Japanese and Mandarin Speakers

Daichi Ishii, Ian Wilson

CLR Phonetics Lab, University of Aizu, Japan

m5281016@u-aizu.ac.jp, wilson@u-aizu.ac.jp

Abstract

Generally, North American English speakers pronounce /r/ as retroflex or bunched, often depending on context and biomechanics. Mandarin speakers also use both articulations when pronouncing Mandarin /r/, but it seems to be speaker-dependent, not context-dependent. This ultrasound study focused on 14 Japanese and 6 Mandarin L2 English speakers pronouncing 138 words containing /r/ in almost all contexts. While Mandarin speakers showed both tongue configurations, all but three Japanese speakers used predominantly retroflex. Those Japanese participants used bunched in almost all contexts and were judged to have near native-like English /r/.

Index Terms: ultrasound, English /r/ articulation, Japanese, Mandarin, retroflex, bunched

1. Introduction

Generally, native North American English speakers pronounce English /r/ in two ways: *retroflex* ([ɻ]) with the tongue tip raised and *bunched* ([ɹ]) with the tongue tip lowered. The characteristic of retroflex [ɻ] is that "the apex is pointed toward the hard palate", while that of bunched [ɹ] is that "the mid-dorsum of the tongue is raised toward the palate while the lowered apex is retracted from the lower incisors" [1]. Approximately 7% of native English speakers use only retroflex [ɻ], 60% only bunched [ɹ], and 33% use both [2].

Even though there are different ways to produce a North American English /r/, researchers generally agree that listeners do not hear any difference between them [3] [4], even though each seems to use a different pattern of 4th and 5th formants [5].

Although the traditional categorization of North American English /r/ is retroflex versus bunched, other researchers have classified tongue shapes into more categories: front up, tip up, front bunched, mid bunched, and curled up [6] [7]. To a second-language learner, though, such detailed classifications would be confusing. As one goal of our research is to help learners pronounce North American English /r/ better, we focus on whether the tongue tip is raised (retroflex) or lowered (bunched).

Previous studies found that factors affecting the native articulation of /r/ include whether the /r/ is before or after a vowel and the type of adjacent consonant or vowel. Specifically, retroflexion occurs more often in pre-vocalic contexts (Table 1), more often in word-initial or post-labial contexts (Table 2), and more often in back vowel and low vowel contexts (Table 3). Our recent research using all possible contexts (229 words, 5 native North American speakers) supported all those results except for the low vowel preference [8].

One reason for the existence of more than one articulatory strategy for a single acoustic output could be that it is physiologically easier to produce a tip-up or tip-down /r/ depending on the sounds that surround the /r/. Indeed, an articulatory modelling study has shown that reducing the tissue displacement, relative strain, and relative muscle stress does result in widely seen preferences for /r/ articulation by native English speakers [16]. So, if non-native speakers of English could be taught these context-dependent, advantageous /r/ articulations, doing so might help them reduce the effort to articulate English smoothly.

To our knowledge, there are no previous studies focusing on Japanese speakers' tongue shapes and their distribution during the production of North American English /r/. Japanese has no /r/ or /l/, but has /ɾ/ (a sound with a relatively high tongue tip), and Japanese students are taught to pronounce English /r/ in a retroflex way [17], so we would expect that Japanese speakers of English would use retroflex articulation for English /r/.

Unlike Japanese, the /r/ sound exists in Mandarin Chinese, and Mandarin speakers use three kinds of tongue shapes: retroflex, bunched and post-alveolar [18]. It is not surprising then that Mandarin-English bilingual speakers use both bunched /r/ and retroflex /r/ when speaking English, but they apparently do so in free variation, unlike the pattern of usage employed by native English speakers [19]. In Taiwan Mandarin, people also use both retroflex and bunched /r/ when they speak their native language [20].

In this study, we use ultrasound, a non-invasive method of looking at the distribution of tip-up and tip-down tongue shapes when Japanese and Mandarin speakers of L2 English pronounce North American /r/. We investigate the contexts in which they use one or the other, and whether perceived pronunciation proficiency roughly correlates with that.

2. Method

2.1. Participants

There were 14 native speakers of Japanese (J1–J14; 4 male and 10 female), 3 native speakers of Taiwanese Mandarin (T1–T3; 2 male and 1 female), and 3 native speakers of mainland Chinese Mandarin (C1–C3; 1 male, 2 female). All 20 participants were living in Japan at the time of data collection, 17 computer science students and 3 professors (1 Taiwanese and 2 Japanese). The tongue images for speaker C2 were unclear and so her data were not included.

To evaluate participants' pronunciation of English /r/, 8 native North-American-English listeners living outside Japan completed an online evaluation task using a Google form. They listened to 84 sound files (4 words × 21 speakers including 2 native North-American-English speakers) including "room" (word-initial /r/), "word" (post-V /r/), "year" (word-final /r/) and "strong" (pre-V /r/) and evaluated participants' fluency from 1 ("completely non-native") to 5 ("native-like"). The 84 sound files were arranged in a random order and were presented in

Table 1: *Previous research showing prevocalic /r/ favours retroflexion; including number of participants and words.*

STUDY	TYPE	PARTIC.	WORDS	FAVOURS RETROFLEXION
Delattre & Freeman 1968 [1]	x-ray	46	32	pre-V > post-V
Uldall 1958 [9]	palatography	1	N/A	pre-V > post-V > syl
Hagiwara 1995 [10]	probe-contact	15	6	pre-V > post-V (blade)

Table 2: *Previous research showing contexts favouring retroflexion when /r/ follows a consonant.*

STUDY	TYPE	PARTIC.	WORDS	FAVOURS RETROFLEXION
Delattre & Freeman 1968 [1]	x-ray	46	32	lab > cor > dor
Westbury et al. 1998 [11]	microbeam	53	5	# > lab > dor > /stri/
Guenther et al. 1999 [12]	EMMA	7	5	#, lab > cor > dor
Espy-Wilson & Boyce 1994 [13]	EMMA	1	N/A	other > dor
Tiede et al. 2010 [14]	MRI	4	3-5	cor > other contexts
Uldall 1958 [9]	palatography	1	N/A	cor > other Cs

Table 3: *Previous research showing contexts favouring retroflexion when /r/ is followed by a vowel.*

STUDY	TYPE	PARTIC.	WORDS	FAVOURS RETROFLEXION
Ong & Stone 1998 [15]	ultrasound	1	11	back > front
Tiede et al. 2010 [14]	MRI	4	3-5	low > high

that same order for each listener. The listeners were told that it was North-American English and to specifically rate the "r" sound.

2.2. Stimuli

A list of 138 words containing possible vowel and consonant combinations [(C)rV, Vr(V), and Vr(C)] was created by searching for ARPABET characters in the Carnegie Mellon University (CMU) Pronouncing Dictionary [21]. The number of words for each context is shown in Table 4.

Table 4: *The number of stimuli used for each context. The numbers do not sum to 138 (total words used in experiment) because some words contained more than one context.*

rV		rC	
Context	Words	Context	Words
R + high V	24	R + lab C	15
R + low V	17	R + cor C	33
R + front V	40	R + dor C	11
R + back V	30	R + other C	10
R + high-mid V	9		
R + low-mid V	20		
Vr		**Cr**	
Context	Words	Context	Words
high V + R	14	lab C + R	18
low V + R	15	cor C + R	19
front V + R	18	dor C + R	8
back V + R	37	other C + R	0
high-mid V + R	0		
low-mid V + R	26		

Because Japanese high school students learn about 3,000 English words [22], stimuli were chosen from the Corpus of Contemporary American English (COCA) [23] such that their frequency of occurrence was in the top 3,000. However, some contexts had insufficient words with high enough frequency, so a word commonly known by Japanese was chosen, despite not being in the top 3,000. In one Vr and one Cr context, no words were found meeting the preceding criteria. When choosing stimuli, contexts were balanced in terms of vowels' place of articulation, and types of neighbouring consonants.

In addition to English stimuli, Japanese and Mandarin Chinese stimuli were prepared as their native language stimuli. Japanese stimuli were ら (/ra/), り (/ri/), る (/ru/), れ (/re/) and ろ (/ro/). Mandarin Chinese stimuli were prevocalic rhotic and syllabic rhotic. They were the same stimuli used in [19].

2.3. Apparatus

A Shure Beta 87A microphone and a Steinberg UR22mkII USB Audio Interface were used to record 24 bit, 192 kHz audio. A Famio 8 SSA-530A ultrasound machine with a 3.75 MHz probe was used to record tongue movement. Video was captured and mixed with the audio using a Canopus ADVC-700 Advanced DV Converter and Final Cut Pro on a late 2014 Mac mini computer running macOS 12.7.4. The older Mac was used because it had a built-in FireWire connector compatible with the DV converter. Participants wore a helmet with a 3D-printed probe holder attachment to keep the probe fixed relative to the head.

2.4. Data collection

Firstly, participants filled out their personal background name, age, etc., and signed an agreement allowing us to use their data anonymously. In addition, they filled out a payment form to be paid for their participation. After that, they tried on and adjusted the helmet so that it was snug.

Participants were seated about 2 meters from the laptop screen displaying the stimuli. A microphone test was done to adjust the input volume.

A PowerPoint file containing one of the 138 stimuli per slide, was displayed to participants. The slideshow advanced automatically every two seconds, and after the 46th slide and the 92nd slide, there was a 30-second rest break. The order of the slides was randomized for each participant using a VBA macro. Participants read the 138 English stimuli first and then some stimuli with /r/ or a tap/flap in their native language.

2.5. Data Analysis

Each /r/ frame was extracted from ultrasound movies (.mov) manually. When tongue shapes were somewhat unclear, frames before and after were checked. The frame in which the constriction was the narrowest (highest tongue position) was selected as the /r/ frame and the frame number was noted.

Next, the software "GetContours", which can track tongue contours automatically and also help manual tracking, was used [24] [25]. We set GetContours to provide 100 points along the tongue contours. Firstly, at least three red dots, one of them at the tongue tip, were roughly chosen by clicking along the tongue contour. "Image Forces" helped to make tongue contours clearer for marking those dots by hand. Secondly, "Apply Tracking" was used to make 100 points fit each tongue contour. Sometimes "Apply Tracking" did not work precisely, so manual adjustments were needed. Finally, the points' (x,y) coordinates were extracted into a .tsv file, which was then converted into an .xls file for analysing with Microsoft Excel.

One of the most important analysis steps was deciding how to distinguish categorically between retroflex /r/ (tongue tip up) and bunched /r/ (tongue tip down). We used the following definition: if the slope of the tongue tip for /r/ is higher than the slope of the tongue tip for tap/flap (/r/), the articulation is retroflex and if lower, it is bunched.

The reason for choosing /r/ articulation as the border between retroflex and bunched was that the tongue tip cannot be higher than the alveolar ridge unless it is curled back behind the ridge. Also, /r/ is found in both English and Japanese, so participants' samples were readily available. Each /r/ frame was extracted the same way as extracting the /r/ frames. The sound /r/ was collected from Japanese participants by having them say /ru/, and from Mandarin speakers when they said the second "t" of the English word "strategy" (a tap). The exception were for J10 and J13, so the /r/ in "strategy" was used for them instead.

The tongue tip slopes were calculated based on points 95–100 that had been obtained from GetContours. The 5 slopes between each pair of those 6 points were averaged together for a mean tongue tip slope.

3. Results and Discussion

In the left column of Table 5, the numbers in parentheses indicate the mean native-listener judged proficiency of /r/ production (henceforth, "R-score"), from 1 (completely non-native) to 5 (native-like). Native speakers' R-scores were 4.81 and 4.83. The highest Japanese R-score was 3.94, and 4 participants' scores were higher than 3.00, but the other 10 participants were rated lower than 3.00, meaning their /r/ pronunciation was closer to "completely non-native" than to "native-like". All three Taiwan Mandarin speakers' R-scores were above 3.00, but not as high as the top 3 Japanese speakers. The two mainland Chinese Mandarin speakers' R-scores were both below 3.00. The participants within each native language are listed in R-score order.

Table 5 shows each participant's bunched /r/ rate in each context. Blue indicates contexts in which /r/ was bunched more often than retroflex; pink indicates the opposite. Overall (in "All contexts"), J12, J1, J14, J8, J3, T1, T2, and C1 all use bunched more than retroflex. J1 was the only speaker to use bunched /r/ more in every single context. It was very interesting that J1, who had higher R-score, showed /r/ articulation closest to the tendencies of native speakers (from Tables 1, 2 and 3). Also, J12 and J14 used bunched more in all contexts except for one. This seems to indicate that more native-like pronunciation results from following a native-speaker distribution of bunched/retroflex tongue shapes. An exception, though, was J13, who had the 3rd-highest R-score but used retroflex more in every single context.

The following participants had no blue cells in any context (indicating they used retroflex at least as often as bunched in every context): J13, J11, J10, J5, J6, J9, and T3. Considering the fact that most Japanese participants mainly used retroflex /r/, it is natural to think that they were explicitly taught to pronounce /r/ in a retroflex way [17]. Even without such explicit instructions, it is possible that they chose to use retroflex /r/ naturally on their own because Japanese /r/ is a tap, so the tongue tip is raised.

In contrast to the fact that Japanese mainly used retroflex /r/, Mandarin Chinese speakers were more variable. As shown in [19], low-proficiency Mandarin L2 speakers use retroflex /r/ more in Post-V position than in Pre-V position, and high-proficiency Mandarin L2-English speakers show almost same retroflexion rate in both Pre-V and Post-V position (36% and 30%) . Our results were similar only for high-proficiency Mandarin L2-English speakers. However, one thing that should be noted is that while [19] used standardized English test scores to evaluate fluency, our standard of proficiency was based on R-scores (subjective evaluations from native listeners).

The context in which J1 used bunched /r/ the least in was word-initial and post-labial, but still used it two-thirds of the time. All other Japanese participants except J12 and J14 used retroflex equally or more than bunched /r/ in those contexts. In the post-dorsal consonant context, when native speakers tend to use bunched /r/, J1 used bunched all the time, but all other Japanese speakers except J14 used retroflex most of the time. Japanese participants including J1 (but excluding J12, J13, J14) used bunched /r/ less in Pre-V context than Post-V context, the same as native speakers.

Focusing on the type of vowel following /r/, Table 6 shows that J1 used bunched /r/ 87.5% of the time before high vowels, which have the tongue blade raised but the tip lowered, making a bunched articulation more natural for a neighbouring /r/, as in [14]. Overall, all participants except J13, J14 and C1 used bunched /r/ equally or less often in pre-low than pre-high context. On the other hand, there was not much difference between the rate of bunching for pre-front vowel versus pre-back.

Mandarin speakers who have relatively low R-scores (C1 and C3) used bunched /r/ less in pre-V context than post-V context (like native speakers do). On the other hand, the bunched /r/ rate of T1, T2 and T3 (all with higher R-scores) was almost the same comparing pre-V and post-V. Thus, Mandarin speakers seem to allow for more free variation, unlike many of the Japanese speakers who followed native speaker norms in pre-V versus post-V contexts. Also, although the bunched /r/ rate of T1, T2 and T3 is almost the same, C1 and C3 used bunched /r/ less in Post-C context than Pre-C. The post-C context which

Table 5: *Rate at which /r/ was pronounced with a bunched (tip-down) articulation in various contexts by participants. The "R" number in parentheses is the mean proficiency of /r/ pronunciation from 1 (completely non-native) to 5 (native-like), judged by 8 native North-American-English listeners. Cell background colours are blue if bunched /r/ prevails and pink if retroflex /r/ prevails. Cell shading colours: dodger-blue* $\geqslant$ *90, deep-sky-blue* $\geqslant$ *70, light-blue* $>$ *50, white* $=$ *50, light-pink* $<$ *50, hot pink* $<$ *30, deep pink* $<$ *10.*

Partic. (R) Sex	All contexts (%)	Pre-V (%)	Post-V (%)	Pre-C (%)	Post-C (%)	Post-# (%)	Post-Lab. (%)	Post-Cor. (%)	Post-Dor. (%)
J12 (3.94) F	61.4	65.7	60.7	59.5	64.4	66.7	55.6	84.2	37.5
J1 (3.88) M	86.9	78.6	93.4	93.2	77.8	66.7	61.1	84.2	100
J13 (3.69) M	24.2	25.7	21.3	25.7	28.9	16.7	27.8	26.3	37.5
J14 (3.31) F	56.9	65.7	59.0	47.3	57.8	83.3	61.1	52.6	62.5
J11 (2.59) F	24.2	11.4	32.8	31.1	8.9	8.3	0.0	15.8	12.5
J2 (2.53) F	37.9	40.0	47.5	36.5	35.6	16.7	0.0	68.4	37.5
J10 (2.50) F	2.0	1.4	3.3	2.7	2.2	0.0	0.0	5.3	0.0
J7 (2.44) F	36.6	12.9	50.8	58.1	13.3	0.0	5.6	21.1	12.5
J8 (2.34) F	58.8	44.3	73.8	74.3	37.8	41.7	22.2	52.6	37.5
J3 (2.31) F	51.0	34.3	62.3	66.2	33.3	25.0	22.2	52.6	12.5
J5 (2.22) M	7.8	5.7	14.8	9.5	2.2	16.7	0.0	5.3	0.0
J6 (2.22) F	20.9	28.6	27.9	14.9	26.7	8.3	16.7	36.8	25.0
J9 (2.06) F	32.7	15.7	45.9	50.0	15.6	16.7	11.1	15.8	25.0
J4 (1.78) M	46.4	44.3	54.1	50.0	37.8	50.0	50.0	26.3	37.5
T1 (3.63) M	66.0	68.6	65.6	62.2	62.2	83.3	27.8	84.2	87.5
T3 (3.56) F	23.5	22.9	26.2	20.3	17.8	25.0	5.6	31.6	12.5
T2 (3.03) M	51.0	52.9	52.5	47.3	53.3	25.0	38.9	73.7	37.5
C1 (2.81) F	71.9	58.6	83.6	81.1	51.1	91.7	22.2	78.9	50.0
C3 (2.53) M	37.3	17.1	52.5	55.4	13.3	33.3	0.0	21.1	25.0

Table 6: *Rate at which /r/ was pronounced with a bunched articulation in various pre-V contexts by participant. For vowel types, H = high, L = low, F = front, B = back. The R-score numbers in parentheses and the colour coding of cells is the same as in Table 5.*

Partic. (R-score)	H (%)	H-mid (%)	L (%)	L-mid (%)	F (%)	B (%)
J12 (3.94)	70.8	44.4	52.9	80.0	62.5	70.0
J1 (3.88)	87.5	55.6	64.7	90.0	77.5	80.0
J13 (3.69)	16.7	33.3	23.5	35.0	20.0	33.3
J14 (3.31)	58.3	66.7	58.8	80.0	62.5	70.0
J11 (2.59)	16.7	11.1	11.8	5.0	12.5	10.0
J2 (2.53)	50.0	22.2	29.4	45.0	45.0	33.3
J10 (2.50)	4.2	0.0	0.0	0.0	2.5	0.0
J7 (2.44)	25.0	0.0	0.0	15.0	10.0	16.7
J8 (2.34)	54.2	44.4	23.5	50.0	50.0	36.7
J3 (2.31)	29.2	22.2	29.4	50.0	35.0	33.3
J5 (2.22)	4.2	22.2	5.9	0.0	2.5	10.0
J6 (2.22)	45.8	22.2	11.8	25.0	32.5	23.3
J9 (2.06)	29.2	22.2	5.9	5.0	22.5	6.7
J4 (1.78)	58.3	22.2	35.3	45.0	55.0	30.0
T1 (3.63)	83.3	66.6	58.8	60.0	75.0	60.0
T3 (3.56)	45.8	11.1	0.0	20.0	30.0	13.3
T2 (3.03)	70.8	22.2	35.3	60.0	57.5	46.7
C1 (2.81)	41.7	55.6	76.5	65.0	57.5	60.0
C3 (2.53)	20.8	11.2	11.8	20.0	12.5	23.3

favored retroflex /r/ the most among Mandarin speakers was post-labial, similar to some past findings with native speakers (Table 2).

Besides high proficiency Japanese, other Japanese participants also showed similar tendencies to previous research on native speakers, like having a following low or back vowel trigger retroflex /r/. However, their bunched /r/ rate was less than 50% in almost all contexts and the articulatory difference between pre-high versus pre-low or pre-front versus pre-back was not substantial, so it cannot be said that many Japanese speakers use retroflex /r/ and bunched /r/ in the same way as native speakers do.

Based on the results of this research — specifically the fact that participants J12, J1, J14, and T1, who were perceived to be pronounce closer to native North American English /r/, used bunched articulations than retroflex in more contexts — it is tempting to believe that Japanese learners of English should be taught to produce a bunched articulation for /r/. However, research has shown that exclusively teaching bunched articulation to American native English-speaking children (in intervention situations) does not help [26]. What that research *did* find helpful is making sure that learners know *both* varieties so that they can choose the one that best suits their own abilities.

Recall that participants J13 and T3, whose R-scores were higher than many, used retroflex /r/ more often in every context, and that 7% of native speakers use exclusively retroflex /r/ [2], indicating that although bunched articulation may help an L2 speaker's R-score, using bunched /r/ is not a *necessary* condition for pronunciation proficiency.

4. Acknowledgements

We thank the 20 participants for their ultrasound tongue images and the 8 native listeners who judged the /r/ productions. Research was supported by JSPS Kakenhi grant JP20K00606 to the second author.

5. References

[1] P. Delattre and D. C. Freeman, "A dialect study of American r's by x-ray motion picture," *Linguistics*, vol. 6, no. 44, pp. 29–68, 1968.

[2] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/," *Language*, pp. 101–140, 2016.

[3] S. Chen, D. H. Whalen, and P. P. K. Mok, "What R Mandarin Chinese /ɹ/s? – acoustic and articulatory features of Mandarin Chinese rhotics," *Phonetica*, 2024. [Online]. Available: https://doi.org/10.1515/phon-2023-0023

[4] M. Gunji, I. Wilson, and J. Perkins, "Reaction time of Japanese listeners to retroflex and bunched /r/ pronunciation by native English speakers," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3334–3334, 2016.

[5] X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4466–4481, 2008.

[6] E. Lawson, J. M. Scobbie, and J. Stuart-Smith, "Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation," *Journal of Phonetics*, vol. 41, no. 3-4, pp. 198–210, 2013.

[7] H. King and E. Ferragne, "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," *Journal of Phonetics*, vol. 80, p. 100978, 2020.

[8] D. Ishii and I. Wilson, "Contexts and biomechanics of native English speakers' /r/ tongue shapes," in *Proceedings of the 2024 Spring Meeting of the Acoustical Society of Japan*. Acoustic Society of Japan, 2024, pp. 1023–1026.

[9] E. Uldall, "American 'molar' R and 'flapped' T," *Revista do Laboratório de Fonética Experimental da Faculdade de Letras da Universidade de Coimbra*, vol. 4, pp. 103–106, 1958.

[10] R. E. Hagiwara, *Acoustic realizations of American /r/ as produced by women and men*. UCLA Working Papers in Phonetics, vol. 90, 1995.

[11] J. R. Westbury, M. Hashi, and M. J. Lindstrom, "Differences among speakers in lingual articulation for American English /ɹ/," *Speech Communication*, vol. 26, no. 3, pp. 203–226, 1998.

[12] F. H. Guenther, C. Y. Espy-Wilson, S. E. Boyce, M. L. Matthies, M. Zandipour, and J. S. Perkell, "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2854–2865, 1999.

[13] C. Espy-Wilson and S. Boyce, "Acoustic differences between "bunched" and "retroflex" variants of American English /r/," *The Journal of the Acoustical Society of America*, vol. 95, no. 5, pp. 2823–2823, 1994.

[14] M. K. Tiede, S. E. Boyce, C. Y. Espy-Wilson, and V. L. Gracco, "Variability of North American English /r/ production in response to palatal perturbation," in *Speech Motor Control: New developments in basic and applied research*, B. Maassen and P. van Lieshout, Eds. Oxford University Press, 2010, pp. 53–67.

[15] D. Ong and M. Stone, "Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics," *Phonoscope*, vol. 1, no. 1, pp. 1–13, 1998.

[16] I. Stavness, B. Gick, D. Derrick, and S. Fels, "Biomechanical modeling of English /r/ variants," *The Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. EL355–EL360, 2012.

[17] N. Shusse, "English Sounds: Learning and Teaching," *Kokugakuin Zasshi*, vol. 118, no. 6, pp. 1–12, 2017. [Online]. Available: https://k-rain.repo.nii.ac.jp/records/288

[18] K. Xing, "Phonetic and phonological perspectives on rhoticity in Mandarin," Ph.D. dissertation, The University of Manchester, 2021.

[19] S. Chen, D. H. Whalen, and P. P. K. Mok, "Production of the English /ɹ/ by Mandarin–English bilingual speakers," *Language and Speech*, vol. 0, no. 0, p. 00238309241230895, 2024. [Online]. Available: https://doi.org/10.1177/00238309241230895

[20] Y.-h. S. Chang, "Articulation of the rhotic /ɹ/ in Taiwan Mandarin," in *Ultrafest XI: Extended Abstracts*, I. Wilson, A. Mizoguchi, J. Perkins, J. Villegas, and N. Yamane, Eds. University of Aizu, 2024, pp. 118–119. [Online]. Available: https://doi.org/10.5281/zenodo.12578650

[21] K. Lanzo, "The CMU Pronouncing Dictionary," accessed Jan., 18, 2024. [Online]. Available: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[22] Ministry of Education, Culture, Sports, Science and Technology (MEXT), "Explanation of Courses of Study for Senior High Schools Foreign Language Edition English Edition," 2010, https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2010/01/29/1282000_9.pdf.

[23] M. Davies, "One-billion word Corpus of Contemporary American English (COCA), 1990–2019," 2020, https://www.english-corpora.org/coca/.

[24] M. Tiede and D. Whalen, "GetContours: an interactive tongue surface extraction tool," *Proceedings of Ultrafest VII*, 2015.

[25] M. Tiede, "GetContours: Tongue contour fitting software; v3.5. [Computer program]," 2021, https://github.com/mktiede/GetContours.

[26] T. M. Byun, E. R. Hitchcock, and M. T. Swartz, "Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 6, pp. 2116–2130, 2014.