Acoust. Sci. & Tech.

PAPER

Articulation strategies for English liquids used by Japanese speakers

Jeff Moore[1*], Jason Shaw[2], Shigeto Kawahara[3], and Takayuki Arai[1†]

1 Sophia University
7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

2 Yale University
7New Haven, CT 06520, USA

3 Keio University
2-15-45, Mita, Minato-ku, Tokyo, 108-8345 Japan

(Received ~, Accepted for publication ~)

Abstract: This study examines the tongue shapes used by Japanese speakers to produce the English liquids /ɹ/ and /l/. Four native Japanese speakers of varying levels of English acquisition and one North American English speaker were recorded both acoustically and with Electromagnetic Articulography. Seven distinct articulation strategies were identified. Results indicate that the least advanced speaker uses a single articulation strategy for both sounds. Intermediate speakers used a wide range of articulations, while the most advanced non-native speaker relied on a single strategy for each sound.

Keywords: articulatory phonetics, second language acquisition, foreign accent, speech production

PACS number: 43.70.Kv

1. INTRODUCTION

1.1 Japanese /r/ and /l/ distinction

English liquids present a unique challenge for the adult Japanese English learner. The Japanese phonological system includes neither an /ɹ/ nor an /l/ phoneme, but has a phoneme which is often considered to be a tap (/ɾ/), distinct from both [1], and frequently used by less advanced Japanese English speakers in place of both English liquids. Learning to perceive and produce these two different sounds is one of the major hurdles a Japanese learner of English faces.

Best's Perceptual Assimilation Model (PAM) [2] argues that learners categorize non-native sounds according to their proximity to their first language, and that differentiating two sounds placed into the same category is among the most difficult tasks for a second language learner. This model has been applied to Japanese speakers' difficulties with /ɹ/-/l/ distinctions [3], with results indicating that Japanese speakers do indeed perceive /ɹ/ and /l/ as a single sound category.

1.2 Variable pronunciations of / ɹ / and /l/

Acoustically, the /ɹ/ phoneme differs from /l/ both temporally and spectrally [4] in English, but it can be most readily identified acoustically by a low F3 [e.g. 5, 6, 7]. Speakers employ a wide variety of articulations that result in lowered F3 [e.g. 8, 9, 10, 11]. Variation in articulation strategies extends even to Japanese English speakers, as Masaki *et. al* [12] discovered in an MRI study, finding 4 distinct strategies for differentiating /ɹ/ and /l/ employed by both native English speakers and Japanese learners.

Other evidence on English liquids suggests that speakers use a variety of articulatory strategies to achieve similar acoustic targets [e.g. 13, 14, 15], producing constrictions at the tongue back or tongue front, adjusting front and back cavities to reach the targeted low F3 typical of an /ɹ/ sound. However, these studies do not report 3D articulatory data. Arguably, it is the formation of the lateral channel that is the goal of /l/ production [16]. Recent evidence suggests that variation in mid-sagittal gestures may support the consistent formation of a lateral channel across varying contexts [17].

1.3 The relationship between perception and production

Perception and production of non-native speech sounds appear to be intimately related. Flege's Speech Learning Model [18] asserts that perception precedes production. If a learner cannot learn to perceive two sounds as separate, they will be unable to adjust their own articulation in order to produce the sounds accurately. Flege argues further that no critical period exists after which a sound distinction cannot in principle be learned, and that the mechanisms of first language acquisition remain active throughout life. Although children lose the ability to distinguish non-native sounds after the first few months of life [19], and age of acquisition seems to be one of the most important factors in determining the degree of foreign accent a speaker will eventually retain [20], it is clear that adult speakers do learn new sound distinctions. The degree to which they become truly native-like remains an empirical question.

* jeffmoore@eagle.sophia.ac.jp
† arai@sophia.ac.jp

A series of studies conducted by Lively, Bradlow and colleagues [21] followed Japanese English learners as they developed the ability to discriminate between /ɹ/ and /l/. Their results indicate that not all phonetic environments are equally challenging. Initial consonant clusters, such as the /pl/ in "play", or the /pɹ/ in "pray", were found to be particularly troubling. This result indicates that the problem of Japanese /ɹ/-/l/ distinction is not merely a segmental issue, operating at the level of the phoneme, but we also need to take their phonological environments into account.

1.4 Research questions

Against this theoretical background, this study examines the following questions:

1. How do Japanese speakers articulate English /ɹ/-/l/ distinctions? What variations are present?
2. How do the articulations of intermediate and advanced adult second language learners differ? How are they different from those of a native speaker?
3. To what degree do speakers' articulation strategies affect the acoustic signal they produce?

2. METHOD

2.1 Participants

Six native speakers of Tokyo Japanese participated. Four participants were students, aged 19 to 23, while two participants were instructors in their late 30s. Participants were balanced on gender, three male and three female, and roughly balanced on level of acquisition, according to time spent abroad, English education level, and a native speaker's judgment of their accents. Only the two most advanced speakers, both instructors, had spent significant time overseas. Other Japanese participants were graduate and undergraduate students at universities in Tokyo.

Participants were chosen from the authors social networks in order to ensure at least three tiers of acquisition: native speakers, highly advanced speakers, and less advanced speakers. Instructors were chosen to represent the most advanced non-native speakers, and students were chosen to represent intermediate, developing learners. Participating students could also be roughly divided by gender, and also between upper intermediate speakers who had spent a few months abroad and had more extensive opportunities to use English in their education, and lower intermediate speakers who had spent no significant time abroad and had not used English in their studies.

Two male native speakers of English also participated as controls, one from Eastern Australia, the other from Washington State in the US, both in their early 30s. .

Procedures were explained to participants in their native language.

Participant information is summarized in Table 1.

Six Japanese Participants	
(JMA)* Japanese Male, Advanced	(JFA) Japanese Female, Advanced
(JMUI)* Japanese Male, Upper Intermediate	(JFUI) Japanese Female, Upper Intermediate
(JMLI) Japanese Male, Lower Intermediate	(JFLI) Japanese Female, Lower Intermediate
Two Native English Speaking Participants	
(AmMN) American Male, Native	(AuMN)* Australian Male, Native

Table 1 Participants

Unfortunately, technical errors in data acquisition rendered some samples unusable. Due to head correction issues, JMA's data could not be included in the current analysis. JMUI and AuMN were excluded for erratic positioning data. These participants were removed from the study and are marked with an asterisk in Table 1.

2.2 Materials

Participants were asked to read sentences from a monitor. Target words were presented in a carrier phrase using ePrime software. A total of fifteen words, shown in Table 2, were presented ten times in random order. The carrier phrase was "Okay, <word>". The word of the carrier phrase preceding the target word ends in [e:], which ensures that the tongue is in a mid-level position at the start of each target word. The word list included /ɹ/ and /l/ in initial singletons and in initial clusters.

Due to an error in stimulus presentation, we recorded only 9 tokens of 'clay' per participant.

Set	Members
Initial Singleton /l/ and /r/	laid
	lay
	raid
Initial Clusters with /l/ and /r/	played
	blade
	clay
	prayed
	braid
	crane

Table 2 Stimuli

2.3 Data Collection

Speech movements were recorded using an NDI Wave Electromagnetic Articulograph (EMA), recording at 100 Hz [22].

Five sensors were affixed to the tongue. Three were affixed along the median sulcus, or sagittal midline. The most anterior sagittal sensor (hereafter T, tongue Tip) was one centimeter from the tip of the tongue. The most posterior sensor (hereafter D, tongue Dorsum) was affixed as far back as was comfortable for the participant, ranging from 5 to 6 centimeters from the tongue tip at furthest extension. A mid-tongue sensor (hereafter B, tongue Blade) was affixed at the midpoint between these two. Two lateral sensors were affixed in line with B, along a coronal plane perpendicular to the sagittal midline. These were placed one centimeter from the right edge of the tongue (hereafter R, tongue Right) and one centimeter from the left edge of the tongue (hereafter L, tongue Left).

Additional movement tracking sensors were placed at the vermillion border of the upper (UL) and lower lips (LL), as well as on the outside of the lower gums, beneath the lower incisors (JAW). Finally, three sensors were affixed to the nasion (N), between the eyes, and behind the ears at the right (RM) and left (LM) mastoid processes. These last three were necessary in order to computationally correct the data for the movement of the participant's head.

Sensor placements are depicted in Figure 1 and summarized in Table 3.

Figure 1 Participant with sensors attached. Photograph used with permission.

Label	Placement
T	1cm from tip; Tongue Tip
B	Between TT and TD; Tongue Blade
D	5-6 cm from tip; Tongue Dorsum
L	1 cm from left edge; Tongue Left
R	1 cm from right edge; Tongue Right
JAW	Gum of lower incisor; Jaw
UL	Vermillion border of Upper Lip
LL	Vermillion border of Lower Lip
N	Nasion
RM	Right Mastoid Process
LM	Left Mastoid Process

Table 3 Sensor Placement

Participants' speech was recorded acoustically at 22kHz, using a Schoeps MK 415 supercardioid microphone with Schoeps CMC 6 Ug power module.

2.4 Articulatory Analysis

Participants' articulatory data was head corrected using Donald Derrick's 2012 head correction scripts and analyzed using Mark Tiede's MView package for MATLAB [23].

2.5 Shapes

Certain consistent patterns emerged in the course of analyzing participants' articulation data. Speakers seem to draw from a set of possible strategies in attempting to produce /ɹ/ and /l/. These strategies are described below in terms of the movements of five tongue sensors, and given names for simplicity. Based on the first author's visual inspection of the data in MView, each token of a target word was categorized as one of seven shapes, shown in figures 2-8. These figures are traces of tongue shapes used by our speakers. Dashed lines do not indicate the tongue surface, as this was not directly measured. They are included to give a rough approximation of the tongue's surface as an aid to the reader.

Figure 2, Retroflex: The tongue tip rises, the blade remains fairly steady on the vertical axis, and the whole tongue body retracts, indicating a pharyngeal gesture at the tongue root. This figure traces JFA's tongue shape.

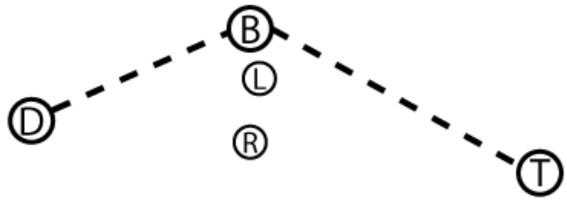

Figure 3, Bunch: The tongue tip remains low, the dorsum rises, and the tongue body retracts. Note also that the left and right lateral sensors are below the blade, indicating a domed shape in the tongue. This strategy appeared most often after /k/ for the intermediate speakers, and was used in all environments by native English speaking participant AmMN. This figure traces AmMN's tongue shape.

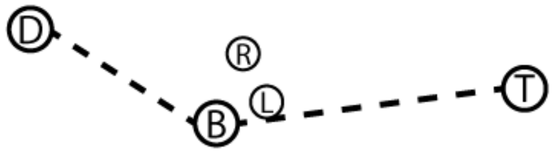

Figure 4, Cup: The tongue tip rises, the mid-sagittal blade sensor drops down between the two lateral sensors, and the tongue does not retract from the resting position horizontally. This figure traces JFA's tongue shape.

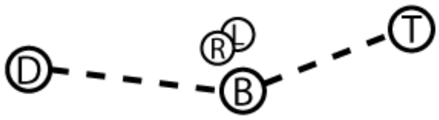

Figure 5, Cup in Retroflex: The tongue tip rises, the mid-sagittal blade sensor drops down between the two lateral sensors, and the tongue retracts. This was the tongue shape overwhelmingly preferred by Speaker JMLI, and the figure traces his tongue shape.

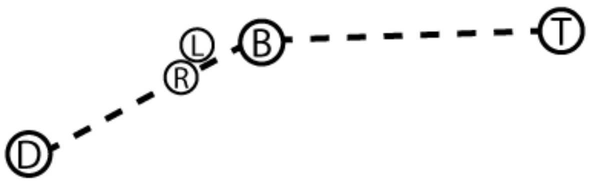

Figure 6, Flat: The blade and tongue tip rise together, while the dorsum remains relatively low. These gestures probably produce a post-alveolar, lamino-apical closure, typical of Japanese productions of /t/ and /d/ sounds. EMA data does not directly indicate closure, however, so it is not possible to determine the precise location or of closure. This figure traces JFUI's tongue shape.

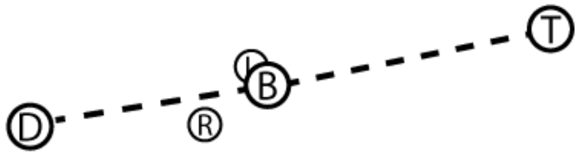

Figure 7, Reach: The tongue tip juts forward, stretching the body of the tongue in a straight line. The blade does not drop, as it does for the Cup shape, nor does it rise with the tongue tip. Rather, the tongue tip shoots forward and pulls the blade and dorsum with it. Where the Retroflex tongue shape is characterized by movement toward the back of the mouth, Reach is characterized by forward motion at the front, a lengthening of the tongue body, and therefore constriction in the front of the oral cavity. This figure traces AmMN's tongue shape.

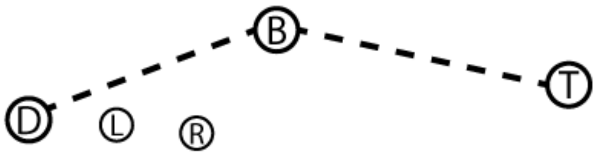

Figure 8, Hunch: The center of the tongue blade takes the lead, rising up above the dorsum, tip, and lateral sensors. These gestures probably indicate a closure at the hard palate. The Hunch shape is distinct from the Bunch in that the tongue does not retract toward the back of the mouth. The front oral cavity is constricted in a Hunch, while Bunching features a larger anterior cavity and constriction behind the tongue. It is also distinct from the Flat tongue shape, in that the tongue tip does not raise. It is apparently a lamino-palatal closure, although again, EMA data cannot directly provide evidence of a closure. This figure traces AmMN's tongue shape.

2.6 Acoustic Analysis

Acoustic data was analyzed in Praat [24], taking intensity, duration, fundamental frequency, and formant frequencies for each segment. Formant frequencies reported here were taken at the segment's midpoint.

2.7 Statistical Analysis

Cohen's *d* [25] is used here as an estimate of the amount of variance between two groups, the productions of /ɹ/ and /l/. Cohen's *d* subtracts the mean of one group from the mean of another and divides by the standard deviation of the set of both groups.

$$d = \frac{\text{M2-M1}}{\text{pooled standard deviation}}$$

In this case, a speaker's /ɹ/ segments constituted one group, and the same speaker's /l/ segments constituted a second group. Since /ɹ/ and /l/ are distinguished mainly by the third formant frequency, calculating Cohen's *d* for F3 gives an estimate of how effectively speakers differentiated /ɹ/ and /l/.

3. RESULTS

3.1. Variable Strategies in Articulation

3.1.1 JFA—Japanese Female, Advanced

JFA has a single, consistent strategy for producing /ɹ/ segments, exclusively using the Retroflex configuration (Figure 2). Her /l/s are nearly as consistent, as she uses a Cup (Figure 4) for all but two of the tokens.

	retro	bunch	cup	cup in retro	flat	reach	hunch
blade			10				
played			10				
clay			8				1
laid			10				
lay			9				1
braid	10						
prayed	10						
crane	10						
raid	10						
Total	40	0	47	0	0	0	2

Table 4 Articulation strategies, JFA

3.1.2 JFUI—Japanese Female, Upper Intermediate

JFUI primarily uses the Flat tongue shape for producing /l/, and primarily uses Retroflex for /ɹ/. She shows some variation, however, particularly in her productions of 'clay' and 'crane'.

She does not use Retroflex or Bunch to produce /l/, and only once did she use Flat to produce an /ɹ/. She seems to have clear-cut categories of the sorts of tongue shapes that she uses for an /l/ and the sorts of shapes that she uses for an /ɹ/. Although there is some variation within categories, rarely uses the same tongue shape to produce both /l/ and /ɹ/.

	retro	bunch	cup	cup in retro	flat	reach	hunch
blade			1		8	1	
played			1		8	1	
clay			3			5	1
laid					8	1	1
lay					9	1	
braid	8	2					
prayed	7	3					
crane	4	6					
raid	7	2			1		
Total	26	13	5	0	34	9	2

Table 5 Articulation strategies, JFUI

3.1.3 JFLI—Japanese Female, Lower Intermediate

JFLI uses a variety of strategies to attempt to produce the English liquid sounds. In fact, she is the only participant who uses each strategy at least once. Unlike JFUI, she uses some tongue shapes for both /l/ and /ɹ/, using Retroflex and Bunch in attempting to produce an /l/ after a bilabial plosive. She also uses the Cup strategy seven times in attempting to produce /ɹ/, as well as the Cup in Retroflex and Hunch strategies once each.

JFLI prefers some tongue shapes for /ɹ/ and some for /l/, but these categories slightly overlap. She uses a Bunch shape once for 'blade', a Retroflex shape once for 'played'—both shapes that she uses more often to produce /ɹ/. Raid shows even more variation, with examples of Cup, Cup in Retroflex, and Hunch, all tongue shapes that she more often applies to /l/. She is not as consistent in differentiating /ɹ/ and /l/ as JFUI.

	retro	bunch	cup	cup in retro	flat	reach	hunch
blade		1	6		3		
played	1		7		1	1	
clay			8				1
laid			10				
lay			9				1
braid	8		2				
prayed	7	1	2				
crane	6	4					
raid	5		3	1			1
Total	27	6	47	1	4	1	3

Table 6 Articulation strategies, JFLI

3.1.4 JMLI—Japanese Male, Lower Intermediate

Unlike the previous three participants, JMLI does not have preferred strategies that differentiate /l/ and /ɹ/. Rather, he has a single strategy, Cup in Retroflex, that he applies to almost three fourths of all tokens. 'Clay' is the only word for which he does not predominantly use the Cup in Retroflex tongue shape, and there he uses a normal Retroflex shape.

	retro	bunch	cup	cup in retro	flat	reach	hunch
blade			1	9			
played			2	7		1	
clay	5			2	1	1	1
laid	2		1	6	1		
lay				6	3	1	1
braid				10			
prayed			1	9			
crane	2			8			
raid				9	1		1
Total	9	0	5	66	6	3	3

Table 7 Articulation strategies, JMLI

3.1.5 AmMN—American Male, Native

AmMN is a Bunch user, applying it to every /ɹ/ articulation. He does have some variation in his /l/ production, sometimes using a Reach, sometimes a Hunch. Recall that these are fairly distinct movements. A Reach involves the tongue tip stretching up and forward toward the alveolar ridge from a lower, resting position, pulling the rest of the tongue forward. In contrast, the tongue blade leads in the Hunch shape, most likely making contact a bit behind the alveolar ridge.

AmMN is entirely consistent in his /ɹ/ production, but his /l/ has variation, more even than our advanced Japanese speaker.

	retro	bunch	cup	cup in retro	flat	reach	hunch
blade						8	2
played						7	3
clay			2			6	1
laid						6	4
lay						9	1
braid		10					
prayed		10					
crane		10					
raid		10					
Total	0	40	2	0	0	36	11

Table 8 Articulation strategies, AmMN

3.2 Results of the Acoustic Analysis

Participant	Mean /l/ F3	Mean /r/ F3	Mean difference	Cohen's *d*
JFA	3026	2489	537	2.67
JFUI	2939	2309	630	3.38
JFLI	3125	2748	377	1.48
JMLI	2372	2174	198	0.52
AmMN	3039	2304	735	1.56

Table 5 F3 values in Hz, all participants

Table 9 shows the mean F3 calculations for each speaker. AmMN has the largest mean difference between /l/ and /ɹ/, which should suggest greater perceptual salience in his samples. However, his data had a much wider variance than either JFA or JFUI, with a standard deviation in his /ɹ/ samples of 674.49 Hz—nearly as large as the mean difference between categories. Looking at the speakers by Cohen's *d*, we see that JFA and JFUI had much more consistently different /l/ and /ɹ/ productions, although the overall differences between their /l/ and /ɹ/ productions were lower.

JMLI can distinguish the sounds, but he applies the distinction inconsistently. His use of the Cup in Retroflex strategy for nearly three quarters of all utterances was the major contributing factor. Other strategies are more successful, as shown in Table 10:

Tongue shape	Mean F3 (Hz)
cup	2753
flat	2249
reach	2587
cup in retro	2148
retroflex	2136

Table 6 F3 Values by tongue shape, JMLI

JMLI is very capable of producing segments with a low F3, using his Cup in Retroflex and Retroflex strategies, and he can produce segments with high F3 by using either Cup or Reach. JMLI's relatively small difference in F3 between his /ɹ/ and /l/ productions seems to come down to an inability to consistently apply the Cup and Reach strategies when producing /l/.

3.3. Acoustic Effect of Articulation Strategies

Generally speaking, participants were able to achieve large distinctions in F3 by using appropriate tongue shapes. Cup, Flat, Reach, and Hunch tongue shapes mostly resulted in segments with F3 values in the high 2000s to low 3000s. Table 11 presents the mean F3 values in Hz for each speaker's articulation strategies.

Speaker	cup	flat	reach	hunch	cup in retro	retro	bunch
JFA	3022	3035	—	3108	—	2488	—
JFUI	2980	2942	2917	2853	—	2280	2357
JFLI	3148	3030	3151	2935	2761	2581	3018
JMLI	2753	2249	2587	—	2148	2136	—
AmMN	2825	—	3053	3029	—	—	2304

Table 7 F3 values in Hz by speaker and tongue shape

JMLI does not achieve quite so large a distinction with his Flat, and his Reach still only achieves an F3 of 2587 Hz, but his Cup shape is very nearly in the range that other speakers achieved, and his Retroflex and Cup in Retroflex shapes achieve F3 values much lower than other participants'.

JFLI's Bunching strategy does not seem to produce a low F3. In the six times that she employs it, the lowest F3 she ever achieves is 2772 Hz, when saying the word 'prayed'. Further, her Retroflex tongue shape produces a higher F3 than any other speaker's /ɹ/ shapes.

4. DISCUSSION

4.1 Patterns of Articulation

The articulation strategies employed by these speakers of varying levels may allow us to entertain a specific hypothesis regarding a path by which Japanese English speakers may develop their /ɹ/ and /l/ pronunciations. This study does not track the development of any individual, and a longitudinal will be necessary to make any strong statements on the process of accent development. With only snapshots of a few speakers' articulations on a given day, it is impossible to say that any of these speakers is likely to develop in any particular way. Nonetheless, some interesting patterns do emerge.

In JMLI, our least advanced speaker, we see a second language learner taking the first steps away from his native Japanese tapped [ɾ] pronunciation. While a Japanese [ɾ] is usually produced forward in the mouth at the alveolar ridge[1], JMLI produces a retroflexed tongue shape, which is here labeled Cup in Retroflex. He does not maintain separate shapes for /ɹ/ and /l/, suggesting that his interlanguage may not yet have separate phonological categories for the two consonants.

JFLI shows some distinction between /l/ and /ɹ/ categories. She mostly focuses on a single strategy for producing each sound, but she still varies her production. Unlike JMLI, she uses a wide repertoire of articulations. And although the majority of the variation in her production is within category—that is, using a separate set of productions for /l/ and a separate set for /ɹ/—she does cross categories from time to time. It appears that she is developing two categories, but has not completely differentiated them articulatorily. Acoustically, she produces an F3 distinction between the /l/ and /ɹ/ categories, but not nearly as strongly as the more advanced speakers do.

The patterns in JFUI's strategy use are superficially similar to JFLI's strategies, but there is a clear difference. JFUI almost never uses a single tongue shape to produce both an /ɹ/ and an /l/. With the exception of a single token, pronouncing the word 'raid' with a Flat tongue shape, she uses Cup, Flat, Reach and Hunch for /l/, and uses Retroflex and Bunch for /ɹ/. She strongly prefers the Flat shape for /l/ sounds, indicating that she has more or less established an articulatory strategy for /l/ that differentiates it from /ɹ/. For /ɹ/, however, she varies Bunch and Retroflex strategies.

JFA shows a very clear, unambiguous distinction between two different categories. She uses a single strategy for /l/, a single strategy for /ɹ/, and never mixes the two. Her articulations produce a stark distinction between her high F3 /l/ and her low F3 /ɹ/.

4.2 Snapshots of Acquisition

We should be cautious, and remember that these are different individuals, and that there is no evidence to suggest that JFA once used tongue shapes like JMLI, JFLI, or JFUI. But the data allows us to form a specific hypothesis as to how Japanese speakers who are developing the /l/-/ɹ/ distinction manage to produce those sounds.

JMLI shows one picture of a speaker beginning to acquire English sounds. His Cup in Retroflex strategy produces an /ɹ/ with a lower F3, so he is approaching an

English-like sound. He uses a single strategy, but it is not precisely a Japanese sound, rather, it is an interlanguage sound. JFLI uses a variety of articulations, and although her /l/ and /ɹ/ categories are not wholly distinct, she is capable of distinguishing the two fairly frequently. JFUI restricts the variation in her articulations, and does not cross categories. JFA, at the most advanced level, uses one articulation strategy to produce her /l/, and another strategy to produce the /ɹ/ phoneme.

None of the speakers behaves precisely like the native speaker. JFA actually demonstrates less variation in her articulation than native speaker AmMN. Where AmMN uses two strategies for producing /l/, JFA uses just one. JFA also uses different strategies than AmMN. She Cups her /l/; he Reaches and Hunches. She Retroflexes her /ɹ/; he Bunches. She is nonetheless very capable of producing a clear acoustic distinction between a high F3 /l/ and a low F3 /ɹ/.

4.3 Theoretical concerns

The Direct Realist theory of perception adopted by Best in her Perceptual Assimilation Model [2] asserts that listeners directly perceive articulatory gestures and categorize them according to their similarity to native gestures. If this is the case, then we surmise that learners will show consistency in production that mirrors their perceptual categorization of gestures. Compared to the available acoustic data on L2 learners, there is very little articulatory data on L2 learners that can serve as the basis for evaluating this claim. The current study provides a first descriptive step. Although the analysis is coarse, we have observed that some speakers, particularly JFLI and JFUI, show a variety of articulatory strategies resulting in common acoustic effects, i.e., lowering F3. This may indicate that these speakers are exploring ways to achieve an acoustic target [e.g. 26]. Although evidence from a variety of sources supports direct perception of gestures [e.g. 27], this study complicates that picture somewhat, at least in regard to non-native speakers. The freedom with which JFUI varies her articulation of /l/ sounds between Cup, Flat, Reach and Hunch suggests that these tongue shapes group together according to similar acoustic consequences, rather than being treated separately according to their very distinct tongue shapes. More detailed quantitative analysis of the tongue shapes, including the time course of constriction formation would be needed to evaluate this claim.

4.4 Limitations

Regrettably, participants were not recruited wholly on the basis of their language skills, gender, age, or other relevant attributes. While we would have preferred to sample participants randomly on the basis of, e.g., English language test scores or time spent overseas, our participants were all sampled from the social networks of the authors. A larger sample of participants, and one that is randomly sampled according to the above criteria, would support generalizability of the findings, an aim for future work on the acquisition of articulatory strategies for non-native phonemes. A second limitation involves the data analysis. Sensor configurations were categorized based on visual inspection of the signal by the first author. Quantitative analysis of the data is needed to substantiate both the number of distinct categories and the membership of each token in a given category.

5. CONCLUSION

This study provides insight into the strategies Japanese speakers use in articulating English /ɹ/ and /l/ sounds. Our Lower Intermediate speaker uses a single articulation primarily, making no distinction between the two consonants, while experimenting with alternative articulatory strategies. The Upper Intermediate speaker focuses on a small set of strategies. The Advanced speaker uses a single, consistent strategy for each consonant—more consistent, in fact, than the native English speaker, although her strategy is not one that the native English speaker uses.

In teaching pronunciation, we should be aware of these patterns and understand that becoming more advanced is a process for non-native speakers. Understanding the path that learners take will require a wider-ranging, longitudinal study.

Inconsistency seems to be the rule. We should expect developing students to be inconsistent, to pronounce a sound differently at different times, even in the same linguistic context. If an intermediate student produces a beautiful /l/ on Tuesday, we should not be surprised when her /l/ is less clear on Friday. She may simply be pushing her vocal tract to try new things, experimenting to find her best articulation strategy. And we should avoid defining ‘advanced’ as ‘native-like’, because an advanced second-language speaker may look very different from a native upon closer inspection.

Future work on this data set will examine the timing of gestures in initial consonant clusters for Japanese English speakers, in the hopes of determining how overlapping of gestures develops. Additionally, we will examine native speakers’ judgments of our speakers’ utterances to assess how gestural strategies and timing differences impact comprehensibility perceptions of accentedness. It is hoped that this work will contribute not only to theoretical understanding of the development of foreign accent and its underpinnings, but also provide practical, practicable information for improving second language accent instruction.

6. ACKNOWLEDGEMENTS

Parts of this experiment were previously reported at the 2016 Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan in Honolulu. We wish to thank all our speakers, and would also like to thank Hinako Masuda and Jihyon Yun for their participation in the data collection process.

REFERENCES

[1] T. J. Vance, *The Sounds of Japanese.* (Cambridge: Cambridge University Press, 2008), p. 89.

[2] C. T. Best, "A direct realist view of cross-language speech perception," In W. Strange, *Speech perception and linguistic experience: issues in cross-language research* (Timonium, MD: York Press, 1995), pp. 171-204.

[3] C. T. Best and W. Strange, "Effects of phonological and phonetic factors on cross-language perception of approximants," *Journal of Phonetics,* **20(3)**, 305-330 (1992).

[4] R. M. Dalston,"Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults," *The Journal of the Acoustical Society of America,* **57(2)**, 462-469 (1975).

[5] S. Boyce and C. Y. Espy-Wilson, "Coarticulary stability in American English /r/," *J. Acoust. Soc. Am.,* **101**, 3741-3753 (1997).

[6] P. Delattre, and D. C. Freeman, "A dialect study of American r's by x-ray motion picture," *Linguistics,* **44**, 29-68 (1968).

[7] J. R. Westbury, M. Hashi, and M. J. Lindstrom, "Differences among speakers in lingual articulation of American English /ɹ/," *Speech Commun.,* **26**, 203-226 (1998).

[8] A. Alwan, S. Narayanan, and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," *J. Acoust. Soc. Am.*, **101**, 1078-1089 (1997).

[9] C. Y. Espy-Wilson and S. Boyce, "Acoustic differences between "bunched" and "retroflex" variants of American English /r/," *J. Acoust. Soc. Am.*, **95**, 2823 (1994).

[10] R. Hagiwara, "Three types of American /r/," *UCLA Working Papers in Phonetics*, **88**, 63-90, (1994).

[11] R. Hagiwara, "Acoustic realizations of American /r/ as produced by women and men." *UCLA Working Papers in Phonetics*, **90**, 1-187 (1995).

[12] S. Masaki, R. Akahane-Yamada, M. Tiede, Y. Shimada, and I. Fujimoto, "An MRI-based analysis of the English/r/and/l/articulations." *In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* **3**, pp. 1581-1584 (1996).

[13] J. S. Perkell, M. L. Matthies, M. A. Svirsky, and M. I. Jordan, "Trading relations between tongue-body raising and lip rounding in production of the vowel [u]: A pilot "motor equivalence" study," *J. Acoust. Soc. Am.*, **93**, 2948-2961 (1993).

[14] J. S. Perkell, M.L., Matthies, and M. A. Svirsky, "Articulatory evidence for acoustic goals for consonants," *J. Acoust. Soc. Am.* **96**, 3326 (1994).

[15] F. H. Guenther, C. Y. Espy-Wilson, S. 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, 105(5), 2854-2865 (1999).

[15] R. Sproat and O. Fujimura, "Allophonic variation in English /l/ and its implications for phonetic implementation" *Journal of Phonetics*, 21, 291-311 (1993).

[16] J. Ying, C. Carignan, J. A. Shaw, M. Proctor, D. Derrick, and C. Best, "Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English", *Interspeech 2017*, [ed] Włodarczak, Marcin, International Speech Communication Association (ISCA), Stokholm, Sweden, (2017).

[17] J. E. Flege, "Assessing constraints on second-language segmental production and perception," In N. O. Schiller, and A. Meyer, *Phonetics and phonology in language comprehension and production: Differences and similarities* (Berlin: Mouton de Gruyter, 2003), pp. 319-355.

[18] P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom, "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science*, **277**, 684–686 (1992).

[19] T. Piske, I. R. MacKay, and J. E. Flege, "Factors affecting degree of foreign accent in an L2: a review," *Journal of Phonetics*, **29**, 191-215, (2001).

[21] A. R. Bradlow, "Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast," In J. G. Edwards and M. L. Zampini (ed.), *Phonology and Second Language Acquisition* (Amsterdam: John Benjamins Publishing Co, 2008), pp. 287-308.

[22] J. J. Berry, "Accuracy of the NDI wave speech research system," *Journal of Speech, Language, and Hearing Research,* 54(5), 1295-1301 (2011).

[23] M. Tiede, "MVIEW: software for visualization and analysis of concurrently recorded movement data," *New Haven, CT: Haskins Laboratories* (2005).

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.28, retrieved 23 March 2017 from http://www.praat.org/ (2017)

[25] J. Cohen, *Statistical power analysis for behavioral sciences* (revised ed.), (New York: Academic Press, 1977).

[26] R. Diehl, and K. Kluenter, "On the objects of speech perception," *Ecological Psychology*, 1, 121-144, (1989).

[27] C.A. Fowler, "Listeners do hear sounds, not tongues," *The Journal of the Acoustical Society of America*, **99(3)**, 1730-1741, (1996).