DYNAMIC TONGUE MOVEMENTS IN L1 JAPANESE AND L2 ENGLISH LIQUIDS

Takayuki Nagamine

Lancaster University
t.nagamine@lancaster.ac.uk

ABSTRACT

Dynamic tongue movements of intervocalic English and Japanese liquids are analysed based on ultrasound data obtained from 17 L1 Japanese and 12 L1 North American English speakers. Principal Component Analysis (PCA) identified three key articulatory properties, including variation in tongue retraction and height as well as tongue tip movement. The time-varying changes showed different magnitude and timing for tongue retraction for Japanese speakers compared to English speakers. Substitution of English liquids with Japanese /r/ was not clearly observed in the tongue movement pattern. The findings highlight the complexity involved in the articulation of English liquids in L2 speech and the usefulness of the finer-grained articulatory analysis in understanding the particular challenges L2 learners face in producing English liquids in L2 speech.

Keywords: L2 speech production, liquids, ultrasound, Principal Component Analysis

1. INTRODUCTION

The acquisition of English liquids /l/ and /ɹ/ presents particular challenges to L2 learners, notably to L1 Japanese speakers [1]. English /l/ and /ɹ/ can be considered gesturally complex sounds because multiple articulatory gestures need to be coordinated spatially and temporally [2, 3]. English /l/ requires two lingual gestures, tongue coronal and dorsal gestures, which are patterned differently depending on the syllable position. Similarly, English /ɹ/ requires coordination of labial, tongue anterior and posterior gestures, and the tongue shape shows substantial cross-speaker variation [4, 5].

While it is generally agreed that L2 learners' difficulty in L2 speech production is rooted in the perception of L2 sounds [1], adult L2 learners' established articulatory routines that are optimised for L1 production could also constrain the accurate production of L2 sounds, especially gesturally complex sounds such as English liquids [6]. L2 learners struggle to produce articulatory gestures that are absent in their L1 [7]. Previous articulatory studies suggest that Japanese speakers' tongue dorsum movement may be the key for them to produce L2 English liquids accurately. Less advanced L1 Japanese learners of English exhibit little movement in the tongue posterior compared to the more experienced learner or the L1 English speaker [8]. Also, the substitution of English liquids with Japanese /r/ was not clearly observed in articulatory data even for less advanced Japanese learners of English [9]. These studies illustrate a learning scenario of L1 Japanese learners of English; while they attempt to differentiate English liquids from Japanese /r/, they struggle to realise the dorsal gesture in producing English liquids.

Japanese has one liquid category /r/, canonically realised as alveolar taps or flaps [ɾ] [10]. In previous research, whether a dorsal gesture is actively involved in Japanese /r/ remains unclear. The tongue dorsum in plain taps [ɾ] shows a stronger coarticulatory effect with the vowels than the palatalised taps [ɾʲ] [11]. Compared to alveolar stops, however, the tongue dorsum in alveolar taps/flaps is more retracted and stabilised [12, 13]. It would, therefore, be useful to compare the tongue movement between English and Japanese liquids in discussing the articulatory L1 influence in L2 speech, especially with regard to the tongue dorsum gesture.

Identifying specific articulatory difficulties in L2 speech also has theoretical implications. The Speech Learning Model (SLM) [14] hypothesises that articulatory realisation rules for L2 sounds are specified in the phonological representation. The Perceptual Assimilation Model for L2 learning (PAM-L2) [15] posits that L2 learners perceive articulatory gestures directly, and they assimilate L2 sounds into the L1 phonological categories based on the gestural information. The case of Japanese speakers' acquisition of English liquids could be a good testing ground as to whether and what articulatory information is important for successful L2 speech learning.

Building on the previous research, the current study aims to identify and compare key articulatory properties involved in English liquids produced by L1 English and L1 Japanese speakers. I particularly focus on the tongue dorsum movement for which I expect to observe the L1 influence carried over from Japanese /r/. Methodologically, the use of ultrasound tongue imaging would complement the findings of previous qualitative articulatory work.

2. METHOD

2.1. Participants

Data from 29 speakers are analysed in the study, including 17 L1 Japanese (eight females and nine males, M_{age} = 19.76 years, SD_{age} = 0.97) and 12 L1 North American English speakers (ten females and two males, M_{age} = 29.08 years, SD_{age} = 6.30).

The Japanese speakers represent the typical English-as-a-foreign-language (EFL) learner population in Japan. They are university students from two universities located in Western and Central Japan who studied English mostly through the school and the university curriculum. They had little experience of a long-term stay outside Japan, except for six who had stayed in an English-speaking country for up to four months.

The North American English speakers were recruited in London and Lancaster in the UK, eight of whom came from the US and four from Canada. Although one of the Canadian speakers was originally born in Poland, she moved to Canada early in her childhood and thus considered herself an L1 Canadian English speaker. No participants reported any history of speech or hearing impairments.

2.2. Materials

Intervocalic English and Japanese liquids are analysed, elicited with the words *believe* /biˈliːv/, *bereave* /biˈɹiːv/, and *biribiri* /biɾibiɾi/ (びりびり). These words are a subset of a larger data collection session. The flanking vowel /i/ is chosen because the vowel quality is similar in Japanese and English and is appropriate for cross-linguistic comparisons [10]. *biribiri* is a Japanese mimetic word that describes the sound and/or the situation of paper being torn. The intervocalic environment most likely yields the canonical tap/flap realisations of Japanese /r/ as it is subject to allophonic variations in other environments [16]. The Japanese speakers were asked to read the Japanese *biribiri* in the LHHH accent so that the liquids in both languages appeared as an onset consonant in an accented syllable.

2.3. Data collection

Participants wore an ultrasound stabilisation headset to stabilise the ultrasound probe under their lower jaw [17]. At the beginning of the recording, they were asked to bite a plastic plate to measure their occlusal plane [18]. Then, the participants read aloud the target words in isolation, resulting in 271 tokens for analysis. The Japanese participants' language modes were controlled by changing the language of instructions and including an English conversation activity between the Japanese and English recording blocks.

Ultrasound data were obtained using a Telemed MicrUs system, with a 64-element probe of 20 mm radius, recorded with the Articulate Assistant Advanced (AAA) software version 220.4.1 [19]. Midsagittal tongue views were imaged with a fixed probe frequency between 2-4 MHz, 80 mm depth, 100% field of view and 64 scan lines, resulting in a framerate of ca. 80 per second. Simultaneous acoustic signals were collected with the signal pre-amplified and digitised using a USBPre2 audio interface, and then recorded onto a laptop computer at 44.1 kHz with 16-bit quantisation. Side-profile lip images and participants' perceptual identification accuracy data for English /l/ and /ɹ/ were also collected but will not be presented here.

2.4. Data analysis

Tongue spline data were exported from AAA using the DeepLabCut (DLC) plug-in [20]. DLC estimates the tongue shape based on 11 key points along the tongue contours in the ultrasound videos based on the trained models [21].

The current study takes the dynamic measurements throughout the vowel-liquid-vowel(V_1LV_2) interval in the target words (i.e., *b**elie**ve*, *b**erea**ve*, and *b**iri**biri*) under the assumption that English liquids exhibit dynamic changes in the acoustic and articulatory realisations, and it is often difficult to specify a particular time point that best represents the liquid quality [22, 23]. Segmentation was carried out based on the acoustic signals using Montreal Forced Aligner [24], with the V_1 onset and the V_2 offset marked at the point where periodic cycles began or ended in waveforms and where the formant structures were clearly visible.

Principal Component Analysis (PCA) has been performed in order to summarise the tongue spline data into a manageable number of key articulatory dimensions that allow for cross-speaker comparisons [25]. Prior to running PCA, data from each speaker were normalised into *z*-scores to allow for comparisons across speakers. Then, PCA was run based on all the x and y coordinates of the 11 points along the tongue surface at 11 equidistant time points during the V_1LV_2 interval produced by all speakers. PCA was performed using the *princomp* function in R [26]. The code and data for analysis are available online at https://osf.io/29tac/.

3. RESULTS

3.1. Identifying key articulatory properties using PCA

PCA identified three key dimensions that account for 88.33% of the data, with each dimension exceeding the 5% threshold suggested in the literature [27]: PC1: 58.03%, PC2: 22.65%, and PC3: 7.65%. In order to make the PCs interpretable, the loadings of each PC are plotted against the mean tongue shape using the standard deviation information [25]. Fig 1 shows key dimensions involved in the production of the V_1LV_2 sequence concerning the degrees of tongue retraction (PC1: left), tongue height (PC2: middle), and the variation in the tongue tip (PC3: right). Given the focus of the current study being the tongue posterior movement, I will focus on PC1 in the following sections.

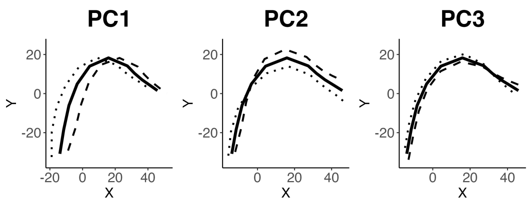

Figure 1: Variation captured in PCs 1 to 3. The thick line represents mean tongue for all of the data for all speakers, with the dimension of variation captured by that PC shown in the dashed and dotted lines.

3.2. Dynamic changes in tongue retraction

Fig 2 presents the time-varying changes of the PC1 values (i.e., tongue retraction) during the V_1LV_2 intervals aggregated for the L1 English and Japanese speakers, where larger PC values indicate more tongue fronting. The overall shape of the PC1 trajectories shows a somewhat similar front-back tongue movement for English liquids for both English and Japanese speakers. The variation associated with the degree of tongue retraction, however, is smaller for the Japanese speakers than the English speakers during the production of *believe* (blue). Regarding *bereave* (yellow), despite a similar degree of tongue retraction between the two speaker populations, the Japanese speakers seem to show a different timing in which they achieve tongue retraction later during the interval than the English speakers do. Finally, changes in PC1 for Japanese /r/ in *biribiri* (grey) were different from either English /l/ or /ɹ/.

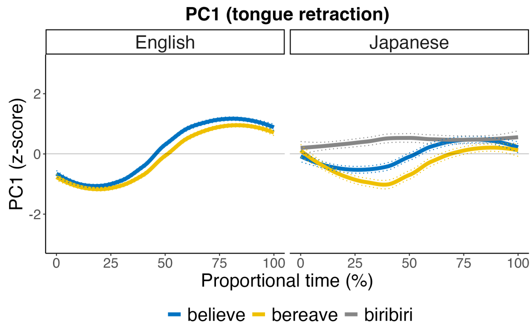

Figure 2: Time-varying changes of the tongue retraction (PC1).

In light of the articulatory patterns identified by PCA, three speakers have been selected to compare the articulatory differences in detail: a female speaker of Canadian English ('English A'), a male and a female EFL Japanese learner of English ('Japanese A' and 'Japanese B', respectively). The midsagittal tongue splines extracted at 11 equidistant points during the V_1LV_2 intervals are presented in Fig 3 and the time-varying changes in PC1 in Fig 4. The two Japanese speakers are chosen based on the auditory impression of the author.

Japanese A maintains an auditory three-way contrast among the three liquids, which is also obvious in the midsagittal tongue shapes. Changes in the PC1 values, however, suggest that the front-back tongue movement would be different from that of English A. Similarly, despite her clear substitution in the auditory analysis and the similarity between English and Japanese liquids in the tongue shapes, the PC1 movement for Japanese B's English /l/ and /ɹ/ is different from that for Japanese /r/.

4. DISCUSSION

The main findings of the current study include that English and Japanese speakers differed in the magnitude of tongue retraction for *believe* and timing for *bereave*. The results could be taken as evidence that the tongue posterior movement may impose difficulty on Japanese EFL learners in articulating English /l/ and /ɹ/ accurately.

While only one vowel environment has been investigated in this study, the current results could

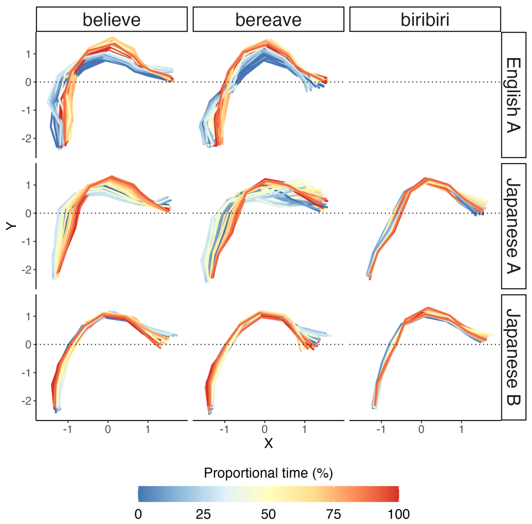

Figure 3: Midsagittal tongue shapes during the V_1LV_2 intervals for a Canadian English speaker (top) and two Japanese speakers (middle and bottom). Tongue tip to the right.

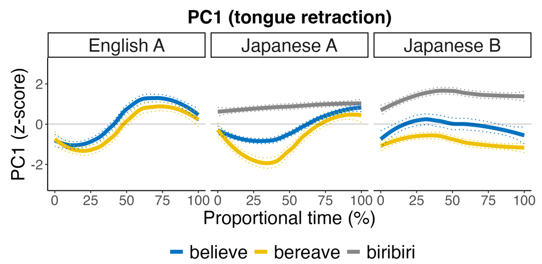

Figure 4: Time-varying changes of the tongue retraction (PC1) during the V_1LV_2 interval for a L1 English and two L1 Japanese speakers.

nevertheless characterise the front-back dimension of the tongue movement for Japanese /r/ compared to the English liquids. The changes in PC1 in Fig 2 suggest minor movements of the tongue posterior over the course of the V_1LV_2 sequence. This agrees with the previous claims that the tongue dorsum for the alveolar taps/flaps shows a substantial coarticulatory effect with the flanking vowels [11] and the dorsal 'stabilisation' strategy [12].

The time-varying trajectory of the front-back movement for *believe* in Fig 2 is 'flatter' for the Japanese speakers than for the English speakers. This may indicate the dorsal stabilisation carried over from Japanese /r/ to the English liquids. In addition, the trajectories for *bereave* (yellow in Fig 2) suggest different gestural timing patterns between Japanese and English speakers' production, such that the Japanese speakers achieved tongue retraction later than the English speakers. Overall, these results could provide evidence for the previous claim that L1 influence is observed in the tongue posterior movement in L1 Japanese speakers' production of English liquids [8].

The individual data replicate the previous findings [9] that a less advanced EFL learner (i.e., Japanese B) does not substitute English liquids with Japanese /r/ completely given the clear differences in the trajectory height (see Fig 4). SLM might explain that she forms separate articulatory realisation rules for English /l/, /ɹ/ and Japanese /r/. Under PAM's account, she might not yet be fully capable of using the dorsal gesture in learning the L1-L2 contrast of liquids. Given that dynamic information of segments might need to be part of phonological representation [28], the dynamic approach in this research is worth pursuing, especially for English liquids that show dynamic changes in articulation.

Finally, the interpretation of the liquid quality in the dynamic approach used in the current study may be subject to the articulatory realisations of vowels. This could be true of V_1; The preliminary acoustic analysis suggested that the second formant frequencies for V_1 seem to differ by 500 Hz between the two participant populations, meaning that the tongue retraction effect could not only be due to the articulatory realisations of English liquids but also the carry-over coarticulatory effects from V_1.

5. CONCLUSIONS

The study analysed dynamic tongue movements in L2 English liquids produced by L1 Japanese speakers. Dynamic analyses of the principal components show a smaller tongue dorsum movement for English liquids produced by L1 Japanese speakers compared to L1 English speakers. Future research could analyse liquids in other vowel environments and the intergestural timing to better generalise this assumed dorsal coarticulatory effect. The participants' perceptual identification accuracy data will also provide further theoretical insights into the nature of L2 speech learning.

6. ACKNOWLEDGEMENTS

Thank you to all the participants for their time and efforts. I thank Dr Claire Nance and Dr Sam Kirkham for their comments and support. Prof. Noriko Nakanishi, Prof. Yuri Nishio, and Dr Bronwen Evans helped me with data collection. The research is financially supported by Graduate Scholarship for Degree-Seeking Students by Japan Student Services Organization and the 2022 Research Grant by the Murata Science Foundation.

7. REFERENCES

[1] J. E. Flege, K. Aoyama, and O.-S. Bohn, "The Revised Speech Learning Model (SLM-r) Applied," in *Second Language Speech Learning*, 1st ed., R. Wayland, Ed. Cambridge University Press, Feb. 2021, pp. 84–118.

[2] D. Recasens, "A cross-language acoustic study of initial and final allophones of /l/," *Speech Communication*, vol. 54, no. 3, pp. 368–383, 2012.

[3] M. Proctor, R. Walker, C. Smith, T. Szalay, L. Goldstein, and S. Narayanan, "Articulatory characterization of English liquid-final rimes," *Journal of Phonetics*, vol. 77, p. 100921, 2019.

[4] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/," *Language*, vol. 92, no. 1, pp. 101–140, 2016.

[5] 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.

[6] B. Gick, P. Bacsfalvi, B. M. Bernhardt, S. Oh, S. Stolar, and I. Wilson, "A motor differentiation model for liquid substitutions in children's speech," in *Proceedings of the Meeting Acoustics*, vol. 1, Salt Lake City, Utah, 2008, p. 060003.

[7] S. Harper, L. Goldstein, and S. S. Narayanan, "L2 Acquisition and Production of the English Rhotic Pharyngeal Gesture," in *Interspeech 2016*. ISCA, Sep. 2016, pp. 208–212.

[8] G. N. Zimmermann, P. Price, and T. Ayusawa, "The production of English /r/ and /l/ by two Japanese speakers differing in experience with English," *Journal of Phonetics*, vol. 12, no. 3, pp. 187–193, 1984.

[9] J. Moore, J. Shaw, S. Kawahara, and T. Arai, "Articulation strategies for English liquids used by Japanese speakers," *Acoustical Science and Technology*, vol. 39, no. 2, pp. 75–83, 2018.

[10] T. J. Vance, *The Sounds of Japanese*. Cambridge University Press, 2008.

[11] N. Yamane, P. Howson, and W. Po-Chun (Grace), "An ultrasound examination of taps in Japanese," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Aug. 2015.

[12] M. Proctor, "Towards a gestural characterization of liquids: Evidence from Spanish and Russian," *Laboratory Phonology*, vol. 2, no. 2, pp. 451–485, 2011.

[13] M. Morimoto, "Geminated Liquids in Japanese: A Production Study," Ph.D. dissertation, University of California Santa Cruz, Mar. 2020.

[14] J. E. Flege, "The intelligibility of English vowels spoken by British and Dutch talkers," in *Studies in Speech Pathology and Clinical Linguistics*, R. D. Kent, Ed. Amsterdam: John Benjamins Publishing Company, 1992, vol. 1, pp. 157–232.

[15] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. Amsterdam: John Benjamins Publishing Company, 2007, pp. 13–34.

[16] T. Arai, "On Why Japanese /r/ Sounds are Difficult for Children to Acquire," in *Interspeech 2013*, Lyon, France, 2013, pp. 2445–2449.

[17] L. Spreafico, M. Pucher, and A. Matosova, "UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science," in *Interspeech 2018*. ISCA, Sep. 2018, pp. 1517–1520.

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

[19] Articulate Instruments, "Articulate Assistant Advanced version 220," Articulate Instruments, Edinburgh, 2022.

[20] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[21] A. Wrench and J. Balch-Tomes, "Beyond the Edge: Markerless Pose Estimation of Speech Articulators from Ultrasound and Camera Images Using DeepLabCut," *Sensors*, vol. 22, no. 3, p. 1133, 2022.

[22] J. Ying, J. A. Shaw, C. Kroos, and C. T. Best, "Relations Between Acoustic and Articulatory Measurements of /l/," in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Dec. 2012, pp. 109–112.

[23] S. Kirkham, C. Nance, B. Littlewood, K. Lightfoot, and E. Groarke, "Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English," *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 784–794, 2019.

[24] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Interspeech 2017*. ISCA, Aug. 2017, pp. 498–502.

[25] C. Nance and S. Kirkham, "Phonetic typology and articulatory constraints: The realisation of secondary articulations in Scottish Gaelic rhotics," *Language*, pp. 419–460, 2022.

[26] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2022.

[27] R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press, 2008.

[28] G. Schwartz and K. Kaźmierski, "Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners," *Language Acquisition*, vol. 27, no. 3, pp. 227–254, Jul. 2020.