Acquisition of articulatory dynamics in second language speech: Japanese speakers' production of English and Japanese liquids

Takayuki Nagamine[1]

[1]*Department of Linguistics and English Language, Lancaster University, UK.*
t.nagamine@lancaseter.ac.uk

Introduction. First-language (L1) Japanese speakers have difficulty in producing English liquids. It is widely hypothesised that this is a consequence of L1 Japanese speakers classifying English liquids as poor instances of the L1 liquid category (Japanese /r/; Bradlow 2008). The precise nature of this influence in articulation, however, is not widely understood, due to a lack of direct comprehensive research on English versus Japanese liquids. Given that alveolar taps, the canonical realisation of Japanese /r/, show a greater vocalic coarticulation than other members of liquids, I hypothesise that (1) L1 Japanese speakers show distinct liquid-vowel coarticulatory patterns in English liquids according to vowel contexts due to a carry-over effect from articulatory strategies for Japanese /r/ and (2) proficient L2 learners are more capable of adjusting coarticulation patterns than less proficient L2 learners (Beristain 2022).

Methods. Midsagittal ultrasound tongue images and audio recordings were collected from 29 L1 Japanese speakers and 14 L1 North American English speakers. L1 Japanese speakers are classified into 'intermediate' (n = 9) and 'advanced' (n = 20) groups based on perceptual identification accuracy of English /l ɹ/ using Gaussian mixture models. Target words for the production study include 16 English words (eight minimal pairs) contrasted by word-initial /l/ and /ɹ/ and five Japanese words with word-initial Japanese /r/ preceding /a/, /i/ and /u/. In total, there are 1,309 tokens of English /l/, 1,321 tokens of English /ɹ/ and 445 tokens of Japanese /r/ for analysis. Tongue splines were tracked based on a set of x/y Cartesian coordinates for 11 reference points along the tongue surface using the DeepLabCut plug-in via Articulate Assistant Advanced (AAA), which were then within-speaker z-scored for cross-speaker comparison. Tongue splines were extracted in the analysis window consisting of acoustically delimited word-initial liquid-vowel intervals with an additional 350ms interval padded before the liquid onset (as articulatory onset can precede acoustic onset).

Three statistical analyses were used to evaluate liquid-vowel coarticulatory patterns. First, I run Principal Component Analysis (PCA) to identify key dimensions in midsagittal tongue shape throughout the interval and summarise them into numeric values (PC scores). I then conduct *functional* PCA (FPCA) to further convert into numeric values (FPC scores) time-varying changes of the PC scores from 350ms prior to the onset through to the offset of the word-initial liquid-vowel interval. Finally, I fit Bayesian hierarchical regression models to predict the PC trajectory patterns, expressed numerically by FPC scores, by vowel contexts and groups as fixed effects and the interaction between them. The random effects include by-speaker varying intercepts and slopes for vowel contexts and by-item varying intercepts and slopes for groups. I use weakly informed priors to allow for a wide range of possible values. I report the estimated coefficients (β) and the 95% credible intervals (i.e., [*min.*, *max.*]) calculated from the model (cf. Roettger, Mahrt, and Cole 2019).

Results. The PCA analysis is shown in the left panel in Figure 1. The first two PCs explain the largest variation in the data: 39.28% for PC1 and 30.59% for PC2. PC1 is associated with tongue body raising and lowering and PC2 with tongue advancement. In the subsequent analyses, I focus on PC1 given its proportion of variance being the largest.

Next, the results of the FPCA analysis is shown in the middle panel in Figure 1 for English /l/ (top), English /ɹ/ (middle) and Japanese /r/ (bottom). Here, to highlight the overall tendency, time-varying changes of PC1 scores are reconstructed directly from FPC1 that accounts for the largest variation in trajectory pattern (58.10%). The x-axis represents proportional time from the onset of the 350ms window (0%) to the vowel offset (100%), and the two vertical dotted lines represent mean liquid interval. The trajectories show that the speaker's tongue position is close to the mean tongue shape prior to the liquid onset, representing their pre-speech posture (Wilson and Kanada 2014). The tongue is then slightly retracted immediately before transitioning into the liquid and vowel portion. During the liquid and vowel interval, the trajectory for the /i/ context (red) is associated with higher FPC1 values, indicating a greater degree of tongue body raising and fronting (i.e., higher PC1 values), followed by in the /u/ (green) and /a/ (blue) contexts. Here, a clear clustering of dynamic PC1 contours is shown for L1 Japanese speakers (i.e., intermediate and advanced groups), suggesting that tongue body movement across vowel contexts is more variable for L1 Japanese speakers than for L1 English speakers.

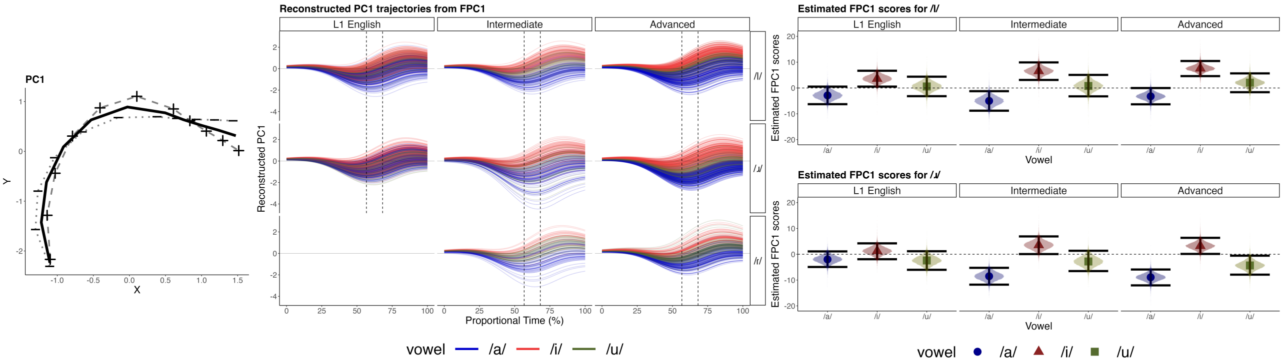

Figure 1: *Left: Variation in midsagittal tongue shape captured by PC1. Middle: Time-varying PC1 changes reconstructed by FPC1. Right: Posterior distributions of FPC1 values for English /l/ (top) and /ɹ/ (bottom).*

Finally, the FPC1 values estimated from the Bayesian hierarchical regression models are shown in the right panel in Figure 1. FPC1 values are overall higher in the /i/ context, followed by the /u/ and /a/ contexts. English /ɹ/ shows a group-vowel interaction for the estimated FPC1 scores. Credible intervals for L1 English speakers show greater overlap with zero (β = 1.26 [−1.93, 4.22] for /i/, β = −1.99 [−4.93, 1.10] for /a/ and β = −2.40 [−6.03, 1.17] for /u/). This indicates that their productions are more similar across vowel contexts compared to that of L1 Japanese speakers (β = 3.55 [0.07, 6.91] for /i/, β = −8.51 [−11.78, 6.91] for /a/ and β = −2.79 [−6.51, 1.35] for /u/ for the intermediate group; β = 3.22 [0.13, 6.36] for /i/, β = −8.92 [−12.05, −5.91] for /a/ and β = −4.34 [−7.90, −0.55] for /u/ for the advanced group).
English /l/, on the other hand, does not show clear group-vowel interactions. Credible intervals show a greater degree of overlap with zero for /u/ (β = 0.59 [−3.17, 4.38] for L1 English speakers, β = 0.79 [−3.24, 5.07] for intermediate L1 Japanese speakers, and β = 2.05 [−1.63, 5.66] for advanced L1 Japanese speakers) compared to /i/ (β = 3.59 [0.53, 6.68] for L1 English speakers, β = 6.57 [3.12, 9.91] for intermediate L1 Japanese speakers and β = 7.59 [4.60, 10.45] for advanced L1 Japanese speakers) or /a/ (β = −2.87 [−6.27, 0.53] for L1 English speakers, β = −4.99 [−8.79, −1.23] for intermediate L1 Japanese speakers and β = −3.22 [−6.32, −0.02] for advanced L1 Japanese speakers).

Discussion and Conclusion. The analysis demonstrates that L1 Japanese speakers show a greater liquid-vowel coarticulation than L1 English speakers for English /ɹ/, suggesting L1 transfer of liquid-vowel coarticulation into L2 (cf. Yamane, Howson, and Po-Chun (Grace) 2015). Vowel contexts influence the production of English /l/ in a similar manner across groups, which may reflect a weaker coarticulatory resistance for English /l/ compared to English /ɹ/ (Proctor et al. 2019). No notable effects of L2 proficiency, investigated through classifying L1 Japanese speakers into ‘intermediate’ and ‘advanced’ groups using perceptual accuracy, are found. This could be because perceptual accuracy does not well capture differences in coarticulatory adjustability among L1 Japanese speakers who are relatively homogeneous late bilinguals (Beristain 2022). Future research will explore the articulatory dimensions not included in this study.

References.

Beristain, Ander Murillo (2022). “The Acquisition of Acoustic and Aerodynamic Patterns of Coarticulation in Second and Heritage Languages”. PhD thesis. University of Illinois Urbana-Champaign.

Bradlow, Ann R (2008). “Training Non-Native Language Sound Patterns”. In: *Phonology and Second Language Acquisition*. Ed. by Jette G. Hansen Edwards and Mary Zampini L. John Benjamins Publishing Company, pp. 287–308.

Proctor, Michael, Rachel Walker, Caitlin Smith, Tünde Szalay, Louis Goldstein, and Shrikanth Narayanan (2019). “Articulatory Characterization of English Liquid-Final Rimes”. In: *Journal of Phonetics* 77, p. 100921. DOI: 10.1016/j.wocn.2019.100921.

Roettger, Timo B., Tim Mahrt, and Jennifer Cole (2019). “Mapping Prosody onto Meaning – the Case of Information Structure in American English”. In: *Language, Cognition and Neuroscience* 34.7, pp. 841–860. DOI: 10.1080/23273798.2019.1587482.

Wilson, Ian and Sunao Kanada (2014). “Pre-Speech Postures of Second-Language versus First-Language Speakers”. In: *Journal of the Phonetic Society of Japan* 18.2, pp. 106–109. DOI: 10.24467/onseikenkyu.18.2_106.

Yamane, Noriko, Phil Howson, and Wei Po-Chun (Grace) (2015). “An Ultrasound Examination of Taps in Japanese”. In: *Proceedings of the 18th International Congress of Phonetic Sciences*. Ed. by The Scottish Consortium for ICPhS 2015. Glasgow, UK: The International Phonetic Association, pp. 1–5.