Linguistic Research 41(3): 589-609
DOI: 10.17250/khisli.41.3.202412.011

Ultrasound production training of American English retroflex /ɹ/: A case of Korean L2 learners*

Joo-Kyeong Lee
(University of Seoul)

Lee, Joo-Kyeong. 2024. Ultrasound production training of American English retroflex /ɹ/: A case of Korean L2 learners. *Linguistic Research* 41(3): 589-609. This study investigates the efficacy of ultrasound tongue imaging as a tool for training Korean learners of English in producing the American English retroflex /ɹ/. Due to the absence of a direct equivalent of /ɹ/ in Korean and the complex phonetic relationship between Korean and English liquid sounds, Korean learners often struggle with accurate production. In the experiment, nine Korean participants of intermediate-level proficiency were trained to produce American English retroflex /ɹ/. Ultrasound imaging was employed to provide real-time visual feedback on tongue movements, aiming to enhance the articulatory accuracy of both onset and coda variants. Participants underwent six training sessions over three weeks, and results revealed significant improvements in articulatory accuracy after the training, particularly in tongue tip elevation and dorsum lowering. Also, these improvements were sustained a month later, indicating effective generalization. This suggests that ultrasound feedback can successfully facilitate the learning of complex L2 sounds like the English retroflex /ɹ/. In addition, there was a notable asymmetry in the learning process between onset and coda retroflex allophones, with the coda variant ([ɹ]) being acquired more readily than the onset variant ([ɹʷ]). This discrepancy is attributed to the greater phonetic dissimilarity between the Korean L1 [l] and the English L2 [ɹ] in coda positions, compared to the dissimilarity between [ɾ] and [ɹʷ] in onset positions. This finding suggests that Korean learners are likely to acquire the coda variant earlier than the onset variant in their learning development of the English L2 retroflex. **(University of Seoul)**

Keywords ultrasound, production training, retroflex, Korean L2 learners of English

* This work was supported by the Basic Study and Interdisciplinary R&D Foundation Fund of the University of Seoul (2021).

1. Introduction

The acquisition of the English retroflex /ɹ/ by Korean learners of English has proven to be particularly challenging, likely due to the absence of this retroflex sound in the Korean phonemic and phonetic inventories. Additionally, the two-to-one mapping between the liquid phonemes in L2 English and L1 Korean, as explained by the Perceptual Assimilation Model (Best 1994), further complicates this process.[1] When evaluating the difficulty associated with learning the L2 retroflex, it is crucial to assess the degree of similarity between the corresponding L1 and L2 sounds (as posited by the Speech Learning Model, Flege 1995, 2003), a task made even more complex when considering positional variants. There is one liquid /l/ in Korean while there are two liquids /l/ and /ɹ/ in English.

It is widely acknowledged that the Korean liquid /l/ is realized as [ɾ] in intervocalic positions and as [l] in coda positions. Furthermore, it is typically excluded from occurring in onset positions, except in the case of borrowed words. In contrast, English liquids /l/ and /ɹ/ have distinct positional allophones: the lateral liquid /l/ exhibits clear-[l] in onset positions and dark-[ɫ] in coda positions, while the retroflex liquid /ɹ/ appears as rounded-[ɹʷ] in onset positions and unrounded-[ɹ] in coda positions within a syllable (Hayes 2009: 43). Focusing specifically on the acquisition of the English L2 retroflex sound, the degree of dissimilarity or similarity between the English retroflex allophone [ɹw] and the Korean [ɾ] will likely determine the ease or difficulty of learning the L2 onset variant [ɹw]. Similarly,the (dis)similarity between the English retroflex variant [ɹ] and the Korean [l] will influence the learning associated with the English coda [ɹ].

Numerous studies have investigated the effects of L2 speech production and/or perception training on the accuracy of L2 learners' production and perception of target sounds (e.g., Catford et al. 1970; Rvachew 1994; Lively et al. 2002; Iverson et al. 2005, 2012; Aliaga-Garcia and Mora 2009; Kondaurova and Francis, 2010; Huensch and Tremblay 2015; Linebaugh and Roche 2015) These studies primarily focused on assessing whether the training modality employed would have transfer effects on learning L2 sounds. A key area of inquiry was whether production training could enhance not only the accuracy of L2 sound production but also perceptual accuracy,

1 There is one liquid /l/ in Korean while there are two liquids /l/ and /ɹ/ in English.

and conversely, whether perception training could lead to improved articulation of sounds in addition to better perceptual performance. The present study aims to investigate the extent to which production training on the English retroflex /ɹ/ enhances the production accuracy of Korean learners. Additionally, this study seeks to scrutinize individual differences among participants in response to the training.

In the instruction of complex motor skills, the provision of visual information regarding model actions, as well as visual feedback on learners' own actions, has been demonstrated to be effective (Carroll and Bandura 1982). This approach has been integrated into speech production training through the use of visualization technology. A computer-animated talking head designed for production training was found to significantly enhance Japanese participants' production and perception accuracy of the English /l/ and /ɹ/ sounds (Massaro and Light 2003). Similarly, another production training that utilized spectrographic visual cues, such as F3 frequency and duration, also demonstrated a substantial impact on the articulatory accuracy of Japanese learners for the English /l/ and /ɹ/ sounds (Hattori 2009).

Ultrasound tongue imaging was initially employed in speech pathology and clinical settings before its application in L2 sound training. It has proven successful in enabling participants, both children and adults, with speech sound disorders of various origins to acquire new lingual articulations. Moreover, these newly acquired articulations have been effectively generalized to untrained items and retained over the long term (Bacsfalvi 2010; Preston and Leaman 2014; Cleland et al. 2015). Ultrasound imaging and video have recently emerged as highly effective tools for providing visual biofeedback during speech production (Wilson 2014; Bird and Gick 2018; Antolik 2020). They are particularly valuable for real-time visual feedback in second language (L2) speech sound acquisition, as it allows learners to observe tongue movements during speech, thereby enhancing motor control and the execution of target articulatory gestures. As noted by Antolik (2020), the tongue is the most dynamic articulator involved in the production of nearly all speech sounds, yet it remains largely invisible during speech. Thus, ultrasound imaging and video technology represent significant advancements in the ability to capture and analyze tongue shape and movement in real-time. This study examines the effects of real-time ultrasound feedback on the accuracy of Korean L2 learners' production of the English retroflex /ɹ/, including both onset and coda variants.

Previous studies have investigated the use of ultrasound for L2 speech production training (Gick et al. 2008; Tsui 2012; Antolik et al. 2013; Tateish 2013; Sisinni et al. 2016; Antolik and Volín 2019), with the majority focusing on Japanese learners acquiring English sounds. These studies consistently reported successful outcomes, with participants demonstrating increased accuracy in producing target sounds following training that incorporated ultrasound biofeedback. However, to the best of our knowledge, there have been no articulatory studies employing ultrasound to examine the production accuracy of the English retroflex by Korean learners through production training.

Gick et al. (2008) conducted one of the pioneering studies utilizing ultrasound technology for L2 speech training. Their pilot stu: dy focused on accent modification in three Japanese learners of English, who received one hour of training each on the production of the English /l/ and /ɹ/ sounds using ultrasound tongue imaging. The results demonstrated increased articulatory accuracy for both /l/ and /ɹ/, indicating that ultrasound feedback is an effective and beneficial methodology for L2 sound acquisition. The rapid improvement observed in participants may be attributed to the clear visualization of tongue position and shape provided by the ultrasound images.

Tsui (2012) conducted a study on the effectiveness of ultrasound training for the production of English L2 consonants /l/ and /ɹ/, aiming to assess its impact on both production and perception accuracy. Six Japanese learners of English participated in four training sessions, each lasting approximately 45 minutes, where they practiced producing /l/ and /ɹ/ in word-initial, medial, and final positions while receiving real-time ultrasound feedback on their tongue movements. Participants' production accuracy was recorded both before and after the training, with native English listeners evaluating the target sounds /l/ and /ɹ/ through word identification tasks to calculate accuracy.

Although ultrasound tongue images and acoustic measurements of F3 were provided as samples, the data were not subjected to quantitative analysis. The findings indicated that Japanese participants showed greater improvement in the production of /l/ compared to /ɹ/ across all positions, except in word-final clusters. Specifically, the accuracy of the retroflex /ɹ/ in word-initial and medial positions increased by 58% and 42%, respectively, post-training, while no improvement was observed in word-final singletons or clusters. Despite these gains in production, participants' perceptual accuracy did not improve following the ultrasound training, suggesting that the transfer effect of training modality was not supported in this study.

Tateishi (2013) conducted an ultrasound production training study similar to that of Tsui (2012). In this study, ten Japanese learners of English underwent training focused on the production of the English word-initial /l/ and /ɹ/ sounds while receiving ultrasound feedback. Participants were recorded both before and after the training, and production accuracy was assessed through native English listeners' perceptions of intelligibility, as well as acoustic measurements of F2 and F3. Additionally, participants completed perception tasks before and after training to evaluate changes in their perception accuracy. The statistical analysis revealed that the articulatory distinction between /l/ and /ɹ/ became more pronounced following the training; however, perceptual discrimination did not show significant improvement over time. The study concluded that production and perception may develop independently in the process of L2 sound acquisition.

In the study by Antolik et al. (2013), ultrasound feedback was employed to train four intermediate-level Japanese learners in the production of two L2 French vowels, /u/ and /y/. The participants underwent three 45-minute training sessions, during which ultrasound was utilized as a visual aid to help them achieve and control the tongue positions necessary for the target vowels. Ultrasound recordings were conducted at three distinct time points: (1) pre-training, one week prior to the first session; (2) post-training, one week after the final session; and (3) two months following the post-training recording. Analysis of the ultrasound splining data revealed that the tongue positions for the two target vowels became more distinct after training, with further improvement observed in the follow-up session. These findings suggested that ultrasound was an effective tool for facilitating the acquisition of L2 vowel sounds.

Sisinni et al. (2016) explored the role of ultrasound tongue imaging in the training of L2 vowel production. Their study focused on the American vowel contrast between /ɑ/ and /ʌ/, involving six native Italian speakers. Three participants received a one-hour ultrasound training session, while the remaining three served as a control group without such training. During the ultrasound sessions, the experimental group observed tongue images for the target vowels, with particular attention to the movement of the tongue dorsum and root. They learned that /ɑ/ is characterized by a more retracted tongue root and a lower tongue dorsum compared to /ʌ/. Both groups were recorded before and after the training sessions, and the vowels /ɑ/ and /ʌ/ were analyzed acoustically using F1 to assess tongue height and F2 to assess tongue backness. The results indicated that the F1 and F2 measurements in the ultrasound group more closely approximated those of native English speakers, demonstrating that ultrasound training positively impacted the pronunciation of the English L2 vowels /ɑ/ and /ʌ/ for Italian learners.

Similarly Antolik and Volín (2019) conducted a study to train Czech learners in the production of English L2 vowels /ɛ/ and /æ/ using ultrasound technology. They utilized real-time ultrasound tongue imaging as visual feedback during the training process. Eight adult Czech speakers participated in three 40-minute ultrasound training sessions, during which they practiced articulatory differences between 12 minimal pair words containing the target vowels while observing their tongue movements via ultrasound imaging. Participants' perceptual accuracy was assessed both before and after the training. The results indicated that, with the exception of one participant, all learners reliably produced the articulatory contrast between /ɛ/ and /æ/ as shown in F1 and F2 trajectories, and three participants demonstrated a lowered front tongue position for /æ/ in the ultrasound images. Perceptual accuracy improved following the training, suggesting that the ultrasound training modality effectively facilitated the learning of L2 vowels. A noted limitation of the study was the inadequate collection of smooth spline data for participants' tongue images, due to issues related to tongue elevation.

Previous studies have demonstrated that production training using ultrasound feedback is generally effective for learning L2 sounds, although improvements in perception have not been consistently observed. This sheds light on Korean native speakers' learning of English /ɹ/ in ultrasound training, because it has attracted a particular attention in L2 sound learning due to its notable allophonic mappings to Korean L1 liquid sounds. The current study explores the effect of ultrasound feedback on the production accuracy of Korean learners with a focus on retroflexion tongue gestures visualized through ultrasound imaging. It particularly investigates whether their L2 sound learning remains extensively effective even a month after the training.

2. Experiment

2.1 Participants

Thirty-eight Korean college students were recruited for an ultrasound study conducted at a university in Seoul, Korea. An announcement seeking participants for the ultrasound training experiment was posted on the university's community SNS. Participants were required to record and email a reading of an English passage in their own voice.[2] To identify nine participants with intermediate English proficiency, their recordings were evaluated for foreign accent (FA, Munro 1998). The FA evaluation involved programming 38 recordings into Quizziz, where three native English speakers rated each recording on a scale from 1 to 9, with 1 indicating minimal foreign accent and 9 indicating a strong foreign accent. The average FA score was calculated for each participant. Those scorings 1-3 were classified as high proficiency, 4-6 as intermediate, and 7-9 as low proficiency. Nine participants with intermediate proficiency were selected for the study. All Korean participants reported no history of speech or hearing disorders and were compensated for their participation in the ultrasound recordings and training sessions.

For collecting native speakers' ultrasound images for training, one native English speaker, who was one of the raters in the FA task, participated in the ultrasound recordings. American native speakers largely produce two different rhotics: retroflex /ɹ/ and bunched /ɹ/. Their articulatory configurations are distinct; that is, the retroflex one is articulated with the tongue tip raised and the tongue dorsum lowered while the bunched one conversely involves tongue dorsum raising and tongue tip lowering (Espy-Wilson et. al. 2000). The native speaker was confirmed to produce retroflex sounds as opposed to bunched ones in a pilot study in advance. He is from the United States and employed as an English instructor at a college in Seoul. He reported no speech or hearing disorders and was similarly compensated. The study was approved by the Institutional Review Board of the University.

2.2 Stimuli

The study utilized three sets of stimulus words containing the English retroflex /ɹ/, with each set comprising 20 words—ten featuring the retroflex in the word-initial onset position and ten in the word-final coda position.[3] The onset retroflex [ɹw]

---

2 See Appendix I for the English passage. It was an excerpt from a children's story book. We tried to choose an English passage containing as many /ɹ/ sounds as possible with its balanced positions of onset and coda.

3 See Appendix II for the stimuli.

involves an additional lip rounding gesture, whereas the coda retroflex [ɹ] is produced solely through retroflexion in American English. The study separately examined these two distinct variants of /ɹ/ to assess their respective improvements following training. One set of stimuli was employed for both the pre-test and training, another set for the post-test, and the third set for the generalization test. In addition to 10 stimuli words, there were 10 filler words differing in target sounds in each set. They were embedded in a carrier sentence, 'Please say _____ for me.' and randomized in order. Nine Korean participants conducted ultrasound recordings of the tongue for thee set of 20 sentences three times in each test. In total, 540 ultrasound images of /ɹ/ were collected (9 participants * (10 word-initial +10 word-final /ɹ/s) * 3 times repetitions).

2.3 Procedure

Three production tests were processed in a consistent manner; participants read a randomized list of sentences 'Please say ______ for me' containing target words. More specifically, they were seated in a comfortable position. They wore a specialized headset from Articulate Instruments Ltd. that held the transducer or probe steady under their jaw (Articulate Instruments Ltd. 2008). The ultrasound data was collected using the software 'Articulate Assistant Advanced (AAA, Articulate Instruments Ltd. 2012)', which simultaneously recorded audio signals. The ultrasound and audio recordings were automatically synchronized within the AAA software. They took the pre-test right before the first training session and the post-test immediately after they finished the last training session. The generalization test was given a month after the last training.

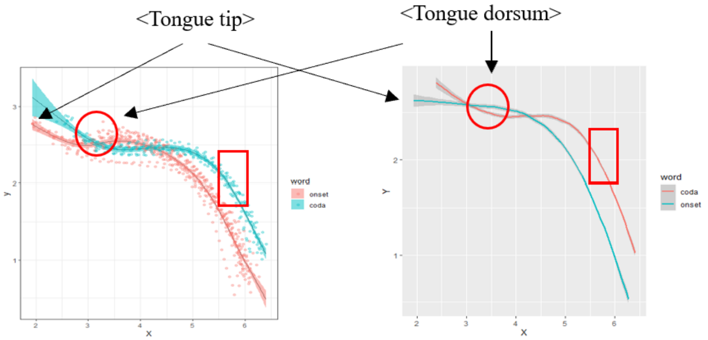
Participants completed six training sessions, held twice a week over three weeks. At the beginning of the first session, they received a 10-minute introduction on the speech organs and the tongue movement of /ɹ/ manifested in ultrasound images. That is, the instructor explained that the tongue tip rose and curled back while playing the native speaker's ultrasound video of /ɹ/ production. In addition, she described that the tongue tip did not touch the roof of the mouth, pointing to the palate line in the image. It was recorded while the native speaker were asked to swallow water. This oral instruction was provided once at the beginning of the first training. Each session lasted 50 minutes, during which participants read a series of sentences in the format, "Please say _____ for me," while viewing ultrasound videos of their own speech production. If the tongue tip was not adequately raised and curled back, the instructor provided corrective feedback by displaying an ultrasound image of the same word produced by a native speaker and inducing the learners to compare it with their tongue images. However, no oral explanation was given to clarify how the participants' production was incorrect.

The target /ɹ/ sounds were annotated based on the aligned information of waveforms and spectrograms. The articulatory data were collected from spline coordinates (over 42 points) at the midpoint of /ɹ/ production. As Espy-Wilson et al. (2000) described, the American English retroflex /ɹ/ is characterized by two primary tongue gestures: the raising of the tongue tip and the lowering of the tongue dorsum. Notably, there are allophonic variations of the retroflex that include additional lip rounding in onset and stressed positions, but not in coda or unstressed positions. Building on this, the current study examines tongue contours using ultrasound image splines, with a focus on the raising of the tongue tip and the lowering of the tongue dorsum. Since lingual gestures are only observed in ultrasound imaging, lip rounding could not be examined in this experiment. The primary aim is to determine whether Korean L2 learners of intermediate proficiency can effectively manipulate the tongue tip and dorsum in both onset and coda positions, and whether their performance improves with ultrasound production training.

Since the sizes of the vocal tract and the tongue were individually different, the spline data should not be collapsed down across participants into one group (Korean or native English group). The individual spline data were submitted to a statistical analysis, that is, SSANOVA (Smoothing Spline ANOVA) and LOESS (Local Estimated Scatter plot Smoothing) in R. SSANOVA shows a midsagittal shape of the tongue, and it generates the best-fit curves from the smoothing splines of repeated stimuli, allowing for visible comparison of curve shapes to identify similarities and differences, which are statistically interpreted by 95% Bayesian confidence intervals (Gu 2002; Davidson 2006; Wang 2011). LOESS is a non-parametric regression method used to fit a smooth curve to data points in a scatter plot. Since it fits multiple local regressions around each point in the data set, it gives a flexible and smooth approximation of the data trend. The onset spline data were aggregated from all onset word recordings, and the coda spline data were similarly aggregated from the coda word recordings.

2.4 Results

Figure 1 shows smoothing splines obtained from SSANOVA in (a) and LOESS in (b) for a native speaker's production of American English /ɹ/. Lips are on the left. Retroflexion gestures are successfully exhibited; the tongue tip on the leftmost part of the splines is raised, and the tongue dorsum marked by a circle is lowered. Moreover, there seems to be additional gestures found for the coda retroflex. The tongue root is retracted as marked by a rectangular, and the degree of tongue tip rising is greater than onset. Therefore, Korean participants' results will be discussed, focusing on three articulatory gestures: tongue tip elevation and tongue dorsum lowering, and tongue root backing. As shown in Figure 1 (a), the spline for coda /ɹ/ lacked data points between x=2 and x=2.5, it resulted in an erroneously wide confidence interval. This occurred because the tip curled back more in coda position, and the remaining tongue (that is, the tongue dorsum and the tongue root) was subsequently retracted. In contrast, LOESS processed the data points as they were, providing a more accurate midsagittal representation of the tongue, as illustrated in Figure 1(b). Therefore, while statistical significance will be discussed based on SSANOVA, the smoothing splines obtained from LOESS will be presented for Korean participants' results.



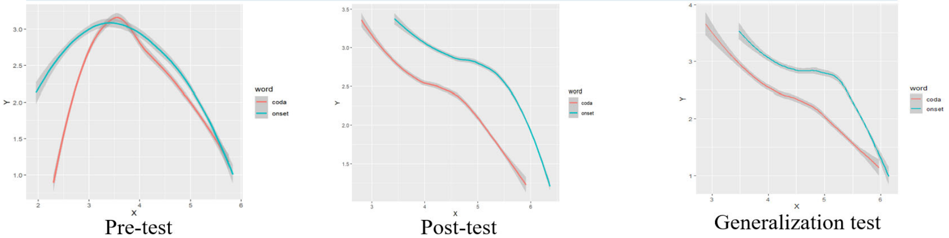
(a) SSANOVA (b) LOESS

Figure 1. SSANOVA estimate in (a) and LOESS estimate in (b) for comparison of the mean curve for onset [ɹ^{w}] and coda [ɹ] at their midpoint (American English native speaker)

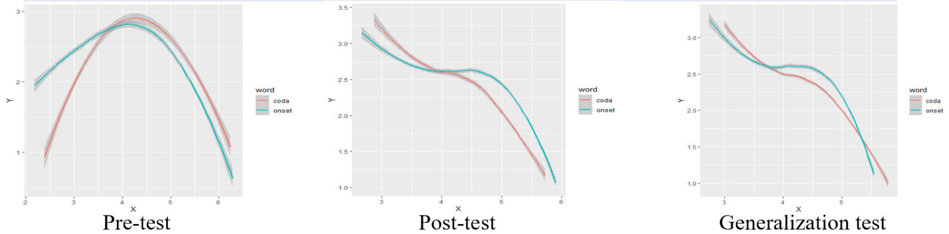
Figures 2 to 4 illustrate the production of the /ɹ/ sound by the Korean participants

(KS01 to KS08). The ultrasound images for one participant were too blurred to analyze, resulting in their exclusion from the study. The pre-test revealed three primary patterns of /ɹ/ production among the Korean learners, all of which were successfully corrected and/or improved in the post-test and generalization tests. First, KS01, KS02, KS04, and KS08 showed a similar pattern of pretest production as seen in Figure 2. That is, they did not exhibit tongue tip elevation. The onset of the sound resembled either an alveolar lateral [l] or a tap [ɾ] while the coda appeared similar to a back vowel /a/ or /ʌ/. The Korean learners in Figure 2 produced a Korean back vowel substituting the word-final coda rhotic after a front vowel as in bill or bee, but they deleted it after a back vowel as in war or heart. There were no coda smoothing splines available after the preceding back vowels. The tongue images of pre-test [ɹ] might have been interpreted as a bunched-[ɹ] because it is articulated with tongue tip lowering and tongue dorsum retraction and rising. However, it was not audibly identified as [ɹ] but as [ɾ]. The participants, on the other hand, successfully articulated the retroflex gestures in the posttest as evidenced by clear tongue tip elevation and dorsum lowering. Furthermore, the learning of /ɹ/ production was effectively generalized, as similar tongue tip and dorsum gestures were evident in the generalization test.

KS01's production of /ɹ/ in pre-, post- and generalization tests



KS02's production of /ɹ/ in pre-, post- and generalization tests



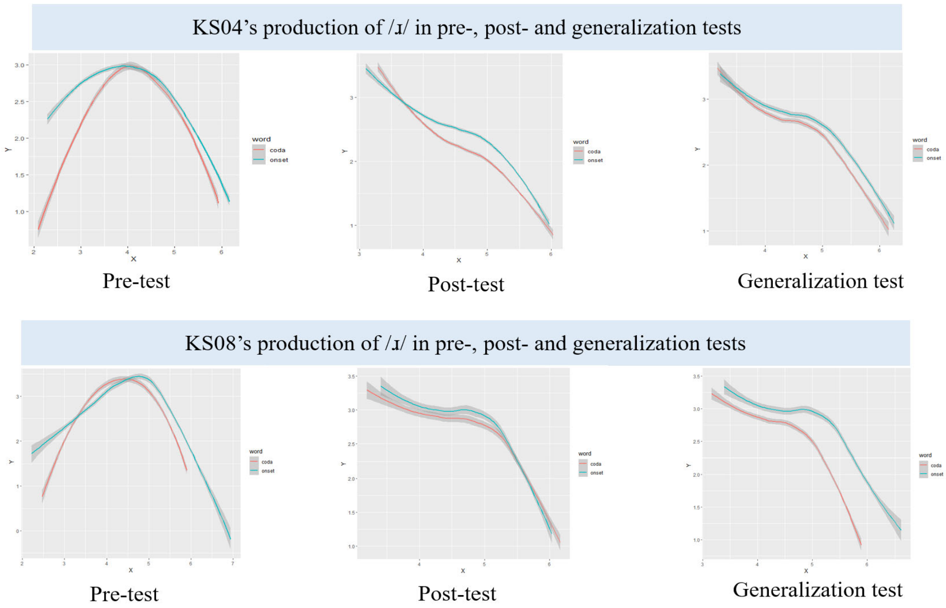


Figure 2. Korean speakers' production of /ɹ/ across pre-test, post-test, and generalization test (KS01, KS02, KS04, and KS08)

As shown in the splines of the four Korean participants in Figure 2, the tongue dorsum was lowered in both onset and coda positions, although the degree of tongue tip elevation and the manner of root retraction varied among participants. KS01 and KS08 exhibited a more extended degree of tongue tip elevation as well as tongue root retraction for the onset position. Notably, KS01 alternated between onset and coda retroflexion gestures in the post-test and generalization tests, suggesting that the degree of tongue tip curling was more extreme in the onset position. This may be attributed to KS01's hyperarticulation of /ɹ/ in word-initial position. The tongue tip curling gestures were completed with tongue tip elevation and dorsum retraction, which appeared to induce a posterior movement of the entire tongue. KS08, on the other hand, produced almost identical retroflexion gestures for both onset and coda in post-test, but the onset seemed to be strengthened or hyperacticualted in the generalization test.

Second, KS03 and KS06 shared similar patterns of /ɹ/ production in the pre-test, although they ultimately succeeded in articulating the retroflex sounds after training and generalizing their learning. As shown in Figure 3, the onset was not correctly articulated as the tongue tip lowered. It resembled the alveolar lateral or the tap, but the coda was successfully produced with the tongue tip raised and the tongue dorsum lowered. These retroflexion gestures were consistent in post- and generalization tests, suggesting that they were sufficiently accurate at articulating English retroflex /ɹ/ in word-final coda position. The onset production which was poor before training improved after training and it was successfully sustained in the generalization test.

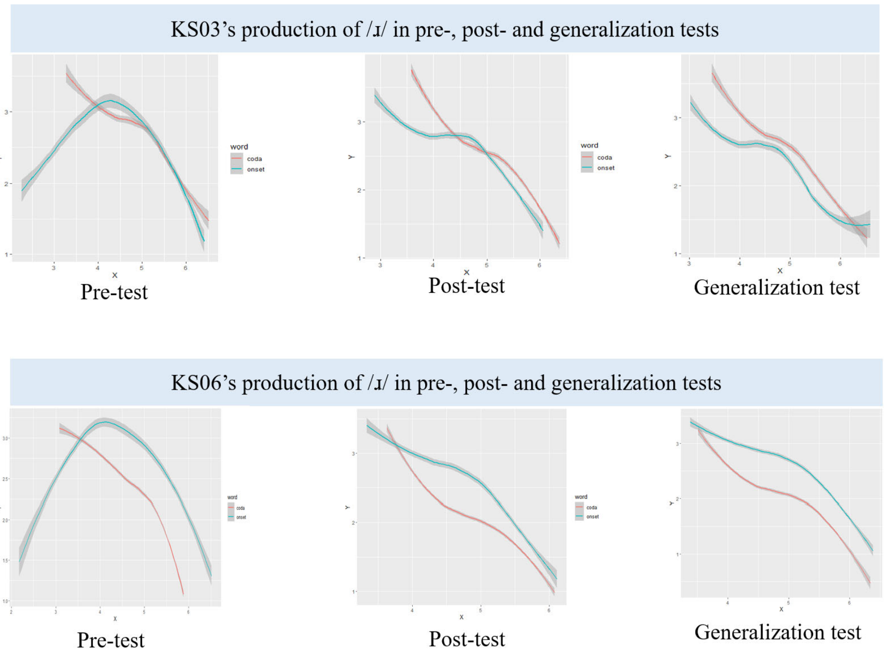


Figure 3. Korean speakers' production of /ɹ/ across pre-test, post-test, and generalization test (KS03 and KS06)

KS03 and KS06 showed individual differences in their production. KS03 produced distinctive allophonic variations of the retroflex similarly to the native speaker; that is, the tongue tip rose higher in coda position while the dorsum simultaneously retracted. However, such allophonic distinction was not found in KS06's production in post- or generalization tests; tongue tip elevation was similar between onset and coda positions, and tongue root retraction was observed in onset rather than coda.

Finally, KS05 and KS07 were commonly accurate at articulating retroflexion gestures across pre-, post-, and generalization tests. Both tongue tip rising and dorsum lowering were well executed. What attracted our attention was tongue root retraction for the onset variant of /ɹ/. The root retracted more extensively in the production of onset as opposed to coda unlike the native speaker. In addition, the tongue tip elevated higher in onset position after training, almost as high as in coda position shown in the smoothing splines of the post-test. This indicates that both participants strengthened and/or hyperarticuated the onset variant of the retroflex.

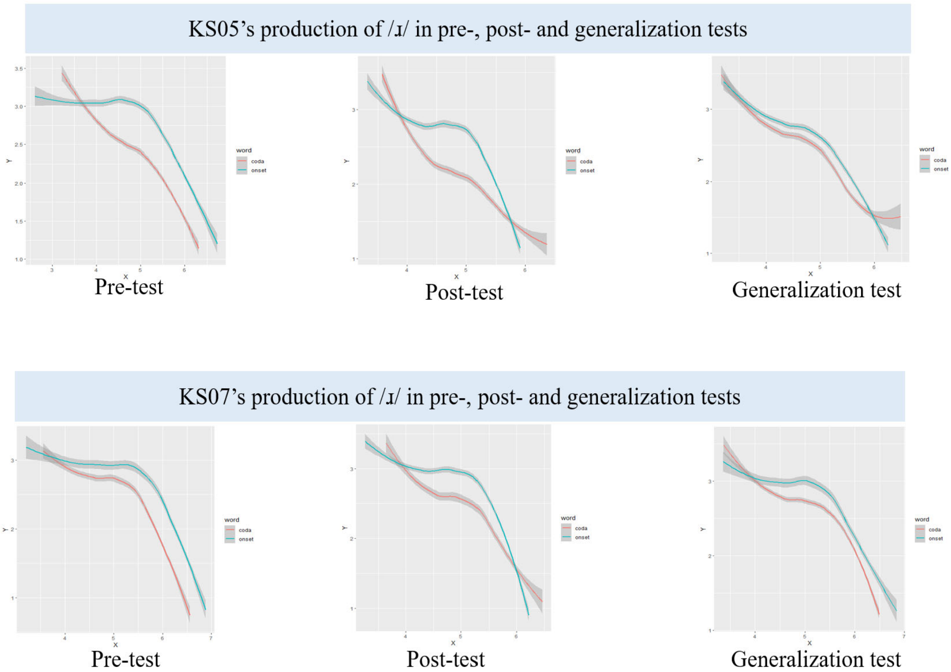


Figure 4. Korean speakers' production of /ɹ/ across pre-test, post-test, and generalization test (KS05 and KS07)

3. Discussion

This study demonstrated that ultrasound production training was sufficiently effective for Korean learners in acquiring the American English retroflex sounds. As noted by Carroll and Bandura (1982), visual feedback is highly effective in enhancing learners' performance when acquiring articulatory movements of the speech organs. Similarly, the corrective feedback provided by ultrasound imaging appeared to be significantly effective for Korean L2 participants in learning the English retroflex /ɹ/, as the majority of participants showed substantial improvement following the training. More specifically, the retroflexion gestures displayed in the ultrasound videos, particularly the rising of the tongue tip and the lowering of the tongue dorsum, were successfully integrated into participants' motor control for articulating English retroflex sounds.

Korean participants, rated as having intermediate proficiency in the foreign accent task, exhibited an imbalanced learning process between onset and coda retroflexes. The majority of participants (KS01, KS02, KS03, KS04, KS06, and KS08) struggled with articulating the word-initial (onset) variant of /ɹ/, as illustrated in Figures 2 and 3. However, two participants (KS03 and KS06) successfully produced the word-final (coda) variant, while the remaining four continued to face difficulties. KS03 and KS06 appear to be at a more advanced stage of learning compared to KS01, KS02, KS04, and KS08. This suggests that the onset retroflex [ɹ^{w}] is more challenging to acquire than the coda retroflex [ɹ]. For the word-initial onset variant, participants typically produced an alveolar lateral [l] or a tap [ɾ], with the tongue tip contacting the alveolar ridge. Although their tongue positions were similar, the lateral articulation was temporally longer than the tap. As acoustic measurements of duration were not conducted in this study, it cannot be precisely determined which sound these six Korean learners produced. However, given that the Korean liquid /l/ is allophonically realized as a tap [ɾ], those who struggled to pronounce the English retroflex likely assimilated it to [ɾ].

The Speech Learning Model (SLM) posits that an L2 sound is likely to assimilate to its corresponding L1 sound when they are phonetically similar (Flege 1995). According to Cathcart (2012), both tap and retroflex sounds share articulatory and acoustic similarities, specifically involving a concave tongue shape and a lower F3 in CV structure. This provides a compelling explanation for the inaccurate production of the onset /ɹ/ before training. In Korean, /l/ is allophonically realized as [ɾ] in onset position, and the English [ɹʷ] is phonetically closer to the Korean tap [ɾ] in the same context, potentially leading to greater perceptual confusion. Consequently, acquiring the ability to perceive and produce this sound is likely to require more time, and even after it is learned, retaining this ability may prove challenging.

What might explain the easier acquisition of the coda retroflex compared to the onset variant in the cases of KS03 and KS06? The Speech Learning Model (SLM) suggests that an L2 sound is more easily learned when it is more dissimilar to its L1 counterpart. It is well known that the Korean liquid /l/ has an allophonic variant [l] in the coda position. The coda retroflex [ɹ] in L2 English corresponds to the L1 Korean [l] in the same syllabic position. These sounds are phonetically distinct, and in some languages, such as English, they are phonologically distinct as well, occupying different phonemic statuses. Korean learners may have perceived the L2 [ɹ] in the coda position as a different sound from the L1 Korean [l], facilitating easier learning compared to the onset retroflex. This may account for the observed imbalance in learning between the onset and coda retroflexes for KS03 and KS06. In terms of L2 sound acquisition, Korean learners appear to acquire the coda retroflex before the onset variant due to the phonetic similarities and dissimilarities between the L1 Korean and L2 English sounds in the same syllabic position.

Despite having the same proficiency level, participants KS01, KS02, KS04, and KS08, who were unable to accurately produce either the onset or coda retroflex, seem to be in the initial stage of learning. In contrast, KS03 and KS06 had already mastered the coda variant at the start of the production training, although they still struggled with the onset retroflex, suggesting they are at an intermediate learning stage. Finally, KS05 and KS07 successfully articulated the tongue tip raising and dorsum lowering gestures, though there was some inconsistency in tongue root retraction between the onset and coda positions. This suggests that they had completed the learning of both the onset and coda variants of American English /ɹ/ and may be in the final stage of learning.

4. Conclusion

The current study has shown that ultrasound-based production training was effective on Korean L2 learners' articulation of the English retroflex /ɹ/. Real-time ultrasound feedback on retroflexion gestures seemed to play an informative role. They compared the ultrasound images of their own with those of native model's articulation, successfully applying what they had learned to their motor control of retroflexion gestures. All the participants showed significant improvements in articulatory accuracy post-training, particularly in tongue tip elevation and dorsum lowering. These improvements were sustained a month later, indicating effective generalization, suggesting that ultrasound feedback can successfully facilitate the learning of complex L2 sounds like the English retroflex /ɹ/. What was worth noting in the experiment was a notable asymmetry in the learning process between onset and coda retroflex allophones. The coda variant ([ɹ]) was learned more readily than the onset variant ([$ɹ^w$]). This discrepancy is attributed to the greater phonetic dissimilarity between the Korean L1 [l] and the English L2 [ɹ] in coda positions, compared to the dissimilarity between [ɾ] and [$ɹ^w$] in onset positions. This finding suggests that Korean learners are likely to acquire the coda variant of the English retroflex earlier than the onset variant in their learning development of the English L2 retroflex.

References

Aliaga-García, Cristina. and Joan C. Mora. 2009. Assessing the effects of phonetic training on L2 sound perception and production. *Recent Research in Second Language Phonetics/phonology: Perception and Production.* 231: 10-27.

Antolík, Tanja Kocjančič. 2020. Ultrasound tongue imaging in second language learning. *Studie z Aplikované Lingvistiky* 11(1): 109-119.

Antolík, Tanja Kocjančič, Claire Pillot-Loiseau, and Takeki Kamiyama. 2013. Comparing the effect of pronunciation and ultrasound trainings to pronunciation training only for the improvement of the production of the French/y/-/u/ contrast by four Japanese learners of French. In *Proceeding sof Colloque Ultrafest VI.* Edinburgh.

Antolík, Tanja Kocjančič, and Jan Volín. 2019. Ultrasound tongue imaging for vowel remediation in Czech English. In Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 3651-3655. Melbourne, Australia.

Antolík, Tanja Kocjačič, C. Pillot-Loiseau, and T. Kamiyama. 2019. The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training. *Journal* *of Second Language Pronunciation* 5: 72-79

Articulate Instruments Ltd. 2008. *Ultrasound stabilisation headset users manual: Revision 1.4.* Edinburgh: Articulate Instruments Ltd.

Articulate Instruments Ltd. 2012. *Articulate assistant advanced user guide: Version 2.14.* Edinburgh: Articulate Instruments Ltd.

Bacsfalvi, Penelope. 2010. Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech Language, Pathology and Audiology* 34(3): 206-217.

Best, Catherine T. 1994. The emergence of native-language phonological influence in infants: a perceptual assimilation hypothesis. In Judith. C. Goodman and Howard. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words*, 167-224. Cambridge, MA: MIT Press.

Bird, Sonya and Brian Gick. 2018. Ultrasound biofeedback in pronunciation teaching and learning. In B. Yegnanarayana, C. Chandra Sekhar, Shrikanth Narayanan, S. Umesh, S. R. M. Prasanna,

Hema A. Murthy, Preeti Rao, Paavo Alku and Prasanta Kumar Ghosh (eds.), *Proceedings of ISAPh 2018 International Symposium on Applied Phonetics*, 5-11.

Bradlow, Ann, Reiko Akahane-Yamada, David B. Pisoni, and Yohichi Tohkura. 1999. Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61(5): 977-985.

Carroll, Wayne R. and Albert Bandura. 1982. The role of visual monitoring in observational learning of action patterns: Making the unobservable observable. *Journal of Motor Behavior* 14(2): 153-167.

Catford, John and David B. Pisoni. 1970. Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54(7): 477-481.

Cathcart, Chundra. 2012. Articulatory variation of the alveolar tap and implications for sound change. *UC Berkeley Phonology Lab Annual Report* 8(8): 76-110.

Cenoz, Jasone and Ma Luisa Garcia Lecumberri. 1999. The acquisition of English pronunciation: Learners' views. *International Journal of Applied Linguistics* 9(1): 3-15.

Cleland, Joanne, James M. Scobbie, and Alan A. Wrench. 2015. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics and Phonetics* 29(8-10): 575-597.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1): 407-415.

Espy-Wilson, Carol Y., Suzzane Boyce, Michel Jackson, Shrikanth Narayanan, and Abeer Alwan. 2000. Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America* 108(1): 343-356.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (eds.), *Speech perception and linguistic experience: Issues in crosslanguage* *research*, 233-277. Timonium, MD: York Press.

Flege, James E. 2003. Assessing constraints on second-language segmental production and perception. In Niels Schiller Olaf, and Antje Meyer (eds.), *Phonetics and phonology in language comprehension and production: Differences and Similarities*, 319-355. Berlin: Mouton de Gruyter.

Gick, Bryan, Barbara Bernhardt, Penelope Bacsfalvi, and Ian Wilson. 2008. Ultrasound imaging applications in second language acquisition. *Phonology and Second Language Acquisition* 36(6): 309-322.

Gu, Chong. 2002. *Smoothing Spline ANOVA Models*. New York, NY: Springer.

Hattori, Kota. 2009. *Perception and production of English /r/-/l/ by adult Japanese speakers*. PhD Dissertation. University College London.

Hayes, Bruce. 2009. *Introductory phonology*. West Sussex: Blackwell Publishing.

Huensche, Amanda and Annie Tremblay. 2015. Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics* 52: 105-120.

Iverson, Paul, Melanie Pinet, and Bronwen G. Evans. 2012. Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics* 33(1): 145-160.

Iverson, Paul, Valerie Hazan, and Kerry Bannister. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America* 118(5): 3267-3278.

Kondaurova, Maria and Alexander Francis. 2010. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of Phonetics* 38(4): 569-587.

Linebaugh, Gary and Thomas Bernard Roche. 2015. Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning* 9(1): A1-A17.

Lively, Scott, David Pisoni, Reiko Yamada, Yohichi Tohkiura, and Tsuneo Yamada. 1994. Training Japanese listeners to identify /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America* 96(4): 2076-2087.

Massaro, Dominic W. and Joanna Light. 2003. Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. *Proceedings of Eurospeech (Interspeech), 8th European Conference on Speech Communication and Technology,* 2249-2252, Geneva, Switzerland.

McCandliss, Bruce, Julie Fiez, Athanassions Protopapas, Mary Conway, and James McClelland. 2002. Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience* 2(2): 89-108.

Munro, Murray J. 1998. The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition* 20(2): 139-154.

Rvachew, Susan. 1994. Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research* 37(2): 347-357.
Sisinni, Bianca, Sonia d'Apolito, Barbara G. Fivela, and Mirko Grimaldi. 2016. Ultrasound articulatory training for teaching pronunciation of L2 vowels. In Pixel (ed.) *Proceedings of ICT for language learning*, 265-270.
Tateishi, Miwako. 2013. *Effects of the use of ultrasound in production training on the perception of English /r/ and /l/ by native Japanese speakers*. Master's Thesis. University of Calgary
Tsui, Haley May-Lai. 2012. Ultrasound speech training for Japanese adults learning English as a second language. Unpublished Maters's Thesis. University of British Columbia.
Preston, Jonathan L. and Marion Leaman. 2014. Ultrasound visual feedback for acired apraxia of speech: A case report. *Aphasiology* 28(3): 278-295.
Wang, Yuedong. 2011. *Smoothing splines: Methods and applications.* Boca Raton, FL: CRC Press.
Wilson, Ian. 2014. Using ultrasound for teaching and researching articulation. *Acoustic Science and Technology* 35(6): 285-289.

Appendix I
Recording passage for Foreign Accent Rating

When Frank was young, his job was to repair bicycles. He saved money for years. And he bought a small workshop of his own. In a few years the small workshop had become a large factory. Frank smiled when he remembered his hard early years and the long road to success.

Appendix II
Stimuli for training, pretest, posttest, and generalization test

Pretest & Training		Posttest		Generalization test	
onset	coda	onset	coda	onset	coda
right	beer	right	beer	race	poor
read	war	read	war	rent	care
row	fire	row	dear	wrong	fire
wrist	mire	road	fair	wrist	mire
rock	tire	rise	fear	rock	tire
rat	gear	rat	gear	ridge	hire
rest	soar	rest	soar	ruin	pure
rank	mere	rank	chair	ring	mere
rich	share	range	near	rich	share
roof	heart	wrist	tour	roof	heart

Joo-Kyeong Lee
Professor
Department of English Language and Literature
163 Seoul Siripdae-ro, Dongdaemun-gu
Seoul 02504, Korea
E-mail: jookyeong@uos.ac.kr

Received: 2024. 08. 27.
Revised: 2024. 10. 02.
Accepted: 2024. 10. 04.