Effects of Articulatory Feedback on Japanese Learners' Perception of the English /l/–/r/

Fumio OZAWA

1. Introduction

For many Japanese learners of English (JLEs), it is difficult to develop English pronunciation close to that of native speakers. One explanation is that as previous research has shown, second language (L2) sounds are difficult to acquire when they are perceived as being close to existing first language (L1) categories. Quite a few previous studies examined whether such difficulties can be overcome by focusing on pronunciation training, i.e., whether pronunciation training leads to improvements in learners' perceptual abilities. Other studies demonstrated an important role of feedback in pronunciation training. It has, however, been pointed out that learners have relatively few opportunities to receive feedback on their pronunciation in classroom settings. Furthermore, it is also known that teachers sometimes lack confidence in their own pronunciation or in their knowledge of effective instructional techniques. As a result, these factors are likely to restrict learners' opportunities to access systematic instruction and practice on pronunciation. To address these issues, the current study aims to examine an effect of computer-based feedback on L2 learners' ability of phoneme categorization.

L2 learners' difficulty in learning correct pronunciation arises from multiple factors, including L1 influence and age of acquisition (Hartshorne, Tenenbaum, & Pinker, 2018). With respect to L1 influence, contrasts such as /s/–/θ/ and /l/–/r/ are frequently cited as especially challenging for Japanese learners of English. The English /l/–/r/ contrast is one of the phonemic contrasts that Japanese learners find most difficult to acquire, and it has therefore been the focus of extensive empirical work. This difficulty can be accounted for within theoretical frameworks such as the Perceptual Assimilation Model (PAM; Best, 1995) and the Speech Learning Model (SLM; Flege, 1995). According to PAM, Japanese /ɾ/ is perceptually assimilated to both English /l/ and /ɹ/, which in turn makes discrimination between these two English phonemes difficult. SLM, in turn, explains the difficulty in terms of learners' age and the acoustic similarity between L1 and L2 sounds, proposing that the formation of new phonetic/phonological categories becomes increasingly challenging under such conditions. Taken together, these models suggest that L1 phonological categories shape the perception and production of L2 sounds. As a result, Japanese learners often have difficulty not only perceiving the /l/–/ɹ/ contrast but also producing it in a native-like manner.

PAM-L2 (Tyler, 2021) extends the original PAM to second language learning and explains how L2 phonemes are perceived in relation to the learner's L1 phonological system. For JLEs, Japanese /ɾ/ and English /l/ and /ɹ/ are assumed to be represented within a common phonological category space for liquids. As L2 learning progresses, this shared space is gradually reorganized, and more distinct category boundaries are established between /l/, /ɹ/, and /ɾ/.

From a perspective of segmental and category-learning, PAM-L2 offers one way of characterizing how /l/, /ɹ/, and /ɾ/ are organized in learners' phonological space. A related line of research on fuzzy lexical representations has highlighted the consequences of such underspecified phonological categories for spoken-word recognition (e.g., Gor et al., 2021). From this perspective, learners often encode difficult L2 contrasts with insufficiently specified phonological features, which leads to fuzzy phonolexical representations. As a result, multiple sublexical or lexical candidates can be activated in parallel and compete for recognition, especially when L2 categories overlap with L1 categories. For Japanese /l/–/ɹ/, this framework predicts that ambiguity among /l/, /ɹ/, and /ɾ/ may create persistent uncertainty in both perception and lexical access.

To address the problem that is specific to JLEs, many previous studies examined an effect of perception training by focusing on particular segmental contrasts. For example, Shinohara and Iverson (2018) implemented high-variability identification and discrimination training in which Japanese learners heard /ɹ/–/l/ minimal pairs produced by multiple talkers and completed computer-based identification (ID) and AX discrimination tasks with trial-by-trial feedback. Complementing such perception-oriented work, production-based activities such as choral repetition (shadowing) and reading-aloud practice have also been examined. Shao, Saito, and Tierney (2023), for instance, provided several weeks of guided choral reading and repetition in classroom settings and evaluated changes in learners' comprehensibility and the perceived strength of their foreign accent.

Previous research using training procedures suggests a link between pronunciation and listening and indicates that perception and production are closely linked. Peperkamp and Bouchon (2011) showed that learners' performance on ABX vowel discrimination tasks systematically correlated with native listeners' identification and accent ratings of the same learners' productions, suggesting that more accurate perception is associated with more target-like production. Conversely, Lambacher et al. (2005) demonstrated that high-variability vowel identification training with feedback not only improved learners' vowel perception but also led to more intelligible vowel production as judged by native listeners. In addition, a meta-analysis by Sakai and Moorman (2018) argues that sustained pronunciation training contributes to the formation and strengthening of L2 phonological representations, with perceptual gains often accompanied by measurable improvements in production. Collectively, these studies highlight the central role of pronunciation training as a means of fostering learners' perceptual and productive abilities.

It is known that feedback plays an important role in pronunciation training and in learners' phonological development (Saito & Plonsky, 2019). The effectiveness of such training varies depending on the nature of the feedback provided. Explicit feedback directly indicates learners' errors and supplies the correct form (Lyster & Ranta, 1997), thereby drawing learners' attention to the error. In contrast, implicit feedback signals errors more indirectly and encourages learners to notice and self-repair them (e.g., through recasts). A meta-analysis by Li (2010) reported a medium overall effect of corrective feedback on L2 learning and found that this effect was generally maintained over time. Moreover, explicit feedback tended to yield larger effects on immediate or short-term posttests, whereas implicit feedback was equally or even slightly more effective on long-delayed posttests. However, most of the studies included in this meta-analysis targeted morphosyntactic or lexical features rather than segmental pronunciation.

Recent research also raises a question about how effectively learners can monitor their own pronunciation. Studies comparing learners' self-ratings with expert judgments suggest that even advanced learners have difficulty evaluating the quality of their segmental production in a way that aligns with trained phoneticians, which may reflect constraints on self-monitoring imposed by L1-based perceptual biases (e.g., Dlaska & Krekeler, 2008). This limitation has motivated growing interest in technology-mediated approaches to pronunciation instruction, in which external feedback is used to make learners' errors more salient. Computer-assisted pronunciation training (CAPT) and systems based on automatic speech recognition (ASR) can provide trial-by-trial information about learners' output, including correctness judgments, scores, or visual displays comparing learner and native productions. Recent reviews and meta-analytic work suggest that such systems tend to yield medium-sized benefits over non-technological conditions, especially when they supply explicit, segment-focused feedback rather than simple transcription or correctness information (e.g., McCrocklin, 2016; Ngo, 2024). The effectiveness of feedback has been documented in training the pronunciation of /r/–vowel contrasts, and different types of feedback have been shown to influence the course of learning differently (Saito & Lyster, 2012). It, however, is not known how different types of feedback (explicit and implicit) affects the Japanese learners' learning of difficult L2 phoneme contrasts such as English /l/–/r/. Testing an influence of explicit and implicit feedback on the perception of English /l/–/r/ by Japanese learners can therefore provide an important test case for the generalizability of corrective feedback effects to L2 phonological learning. Few studies have directly compared the effects of explicit and implicit feedback on the perception of the English /l/–/r/ contrast. The present study therefore aims to clarify how these two types of feedback differentially affect changes in the perception of /l/ and /r/ for Japanese learners of English.

2. Method

2.1 Participants

A total of seventy-eight people who live in Japan were recruited through the Japanese crowdsourcing platform CrowdWorks (CrowdWorks, Inc.). Twelve participants were excluded from the analysis based on pre-specified criteria (e.g., technical issues resulting in ID–record mismatches). In total, sixty-six adult native speakers of Japanese (age range = 26–71 years) were included in the analysis.

2.2 Procedure

The experiment was administered online using PCIbex (PennController; Zehr & Schwarz, 2018). Informed consent was obtained from all participants prior to the experiment. They received a payment upon completion of the experiment. Before starting the experiment, participants read detailed instructions about the experiment. Before beginning, participants provided completed headphone and microphone function checks. To assess the effects of training on perception of the English /l/–/r/ contrast, we conducted an identification test. In sum, the experiment comprised three phases: (a) a pre-training word identification test (ID test), (b) the pronunciation training, and (c) a post-training word identification test. The identification test was conducted before and after the training. Each experimental session typically took approximately ten minutes to complete.

2.3 Identification Test (ID test)

On each trial, participants listened to a recorded word and selected the word they think they had heard from a visually presented /l/–/r/ minimal pair (Figure 1). Each test session comprised 20 trials (two sessions—pre and post—40 trials total). In the post-test, half of the items were identical to those used in the pre-test, and the other half consisted of the complementary counterparts of those items from the same minimal pairs.

To avoid confusion, the placement of /l/ and /r/ words was fixed throughout the experiment: /l/ items are always displayed on the left side and /r/ items are

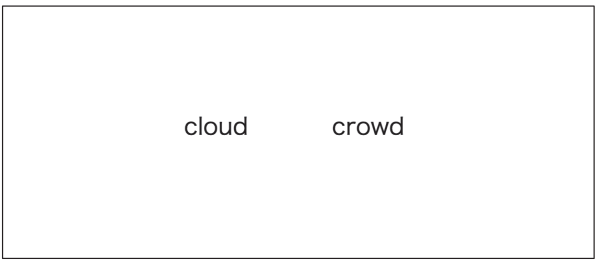

Figure 1. Screenshot of the Identification test

always displayed on the right side. Responses were registered via keyboard: the ‘F’ key was for the word on the left and the ‘J’ key was for the word on the right. No feedback was provided during the identification tests. Within each session, the order of the trials presented was randomized.

2.4 Training Task

Participants were randomly assigned to one of two groups: a feedback group and a no-feedback group. In both conditions, participants listened to a word and selected the word they believed they had heard from a visually presented /l/–/r/ minimal pair. On correct responses, a “Correct!” was displayed and then automatically advanced to the next item.

When participants responded incorrectly, the feedback differed for the two groups. For the feedback condition group (Figure 2), “Incorrect” appeared at the top of the screen, and the center of the screen displayed diagrams of articulatory phonetics of /l/ and /r/ (the position of tongue with the airflow shown in animation), along with the explanations of how to produce each phoneme above the diagram. Below, participants were required to listen to the model pronunciation of “lump” and “rump” played sequentially. They were then prompted to produce and record their pronunciation by pressing the spacebar advanced to the next trial.

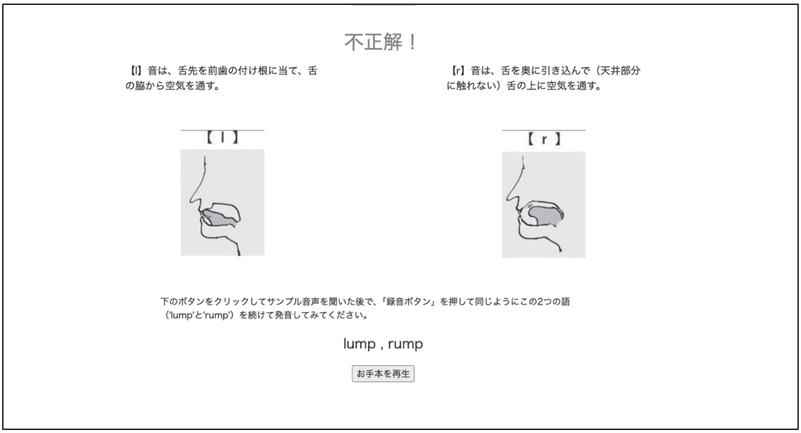

Figure 2. Screenshot of the feedback condition

In the no feedback group (Figure 3), only the “Incorrect” message was shown; no the shape of the tongue and airflow and explicit articulatory explanation were provided.

Figure 3. Screenshot of the no-feedback condition

However, in both conditions, participants listened to model pronunciation of /l/ and /r/ sounds(‘lump’ and ‘rump’). Participants then produced the pair once, and their repetition was recorded.

3. Analysis

Total of 78 learners initially took part in the experiment. For the ID analyses, data from 12 participants were excluded because of technical problems with ID matching, leaving 66 participants in the final dataset.

For the ID test, mean identification accuracy was first computed for each combination of Session and feedback condition. Figure 4 shows mean identification accuracy in the identification test by Session (pre-test, training, post-test) and Group (feedback vs. no-feedback), with standard error bars.

Trial-level identification accuracy (1 = correct, 0 = incorrect) was then analyzed using a generalized linear mixed-effects model (GLMM; Baayen, 2008) with a binomial distribution and logit link, fitted with the glmer function in the lme4 package in R (version

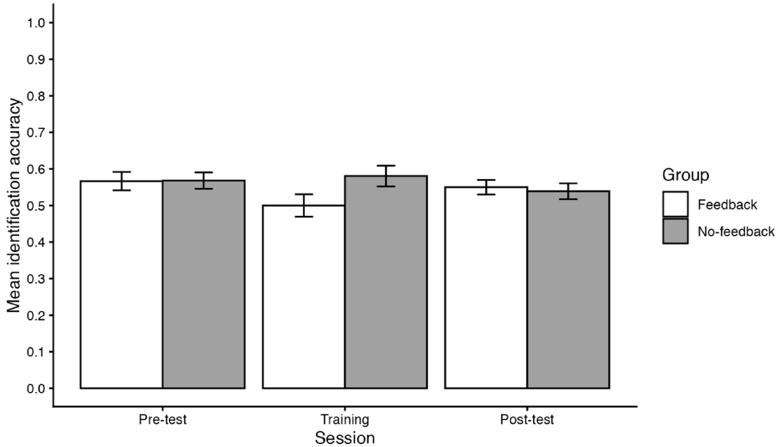

Figure 4. Mean identification accuracy for the pre-test, training, and post-test for each feedback group

4.3.2; R Core Team, 2024).

Model selection began with a full model that included the three-way interaction among Session, Phoneme, and Group, as well as all associated lower-order interactions. We then compared simpler, nested models using likelihood-ratio tests and removed interaction terms when dropping them did not significantly worsen model fit. As a result, all interactions involving Group were excluded, whereas interaction between Session and Phoneme remained significant and was therefore retained in the final model.

The fixed effects were Position (word-initial vs. word-middle), Session (pre-training vs. training vs. post training) , Phoneme (/l/ vs. /r/), Group (feedback vs. no-feedback), age, gender (Male vs. Female), training-phase accuracy, and the interaction between Session and Phoneme. Random intercepts were specified for participants and items. The final model included 3,300 observations from 66 participants and 40 words (AIC = 4419.7, BIC = 4486.8, log-likelihood = −2198.8).

The results of the analysis are summarized in Table 1. As shown in the table, the main effect of Group was not significant. The main effect of Position was significant. It suggests that the accuracy of identification is lower with the words that contain the /l/ or /r/ phonemes at the word-middle position than those that contain them at the

Table 1. Fixed effects from the generalized linear mixed-effects model predicting trial-level identification accuracy.

Fixed effect	β	SE	z	p
Intercept	0.235	0.073	3.210	.001**
Position	−0.362	0.128	−2.832	.005**
Session (pre vs. post)	−0.099	0.080	−1.234	.217
Phoneme (/l/ vs. /r/)	0.056	0.083	0.677	.498
Group (feedback vs. no-feedback)	−0.089	0.111	−0.801	.423
Age	0.000	0.006	0.034	.973
Gender	−0.121	0.112	−1.086	.278
Training accuracy	0.085	0.033	2.590	.010**
Session × Phoneme	0.361	0.166	2.173	.030*

word-initial position. Training accuracy also showed a positive effect. In addition, the interaction between Session and Phoneme was significant. By contrast, the main effects of Session, Phoneme, age, and gender were not significant.

The absence of a main effect of Group indicates that overall identification accuracy did not differ reliably between the feedback and no-feedback conditions. The significant main effect of Position suggests that some word positions were systematically easier to identify than others. The positive effect of training accuracy shows that learners who performed better during the training phase also tended to achieve higher scores on the identification test. The significant interaction between Session and Phoneme indicates that the /l/–/r/ contrast changed over time, with /r/ becoming relatively easier to identify than /l/ from pre- to post-test.

The non-significant main effects of Session, Phoneme, age, and gender suggest that, once the other predictors were taken into account, these factors did not produce reliable overall differences in identification accuracy.

Figure 5 illustrates the interaction between Session and Phoneme by plotting mean identification accuracy for /l/ and /r/ at pre- and post-test. Further comparisons examined a simple effect of Session for each phoneme. The results showed that the simple effect of Session was statistically significant for /l/ ($\beta = -0.28$, SE = 0.12, $z = -2.42$, $p = .015$) but not for /r/ ($\beta = 0.08$, SE = 0.12, $z = 0.71$, $p = .48$). This suggests

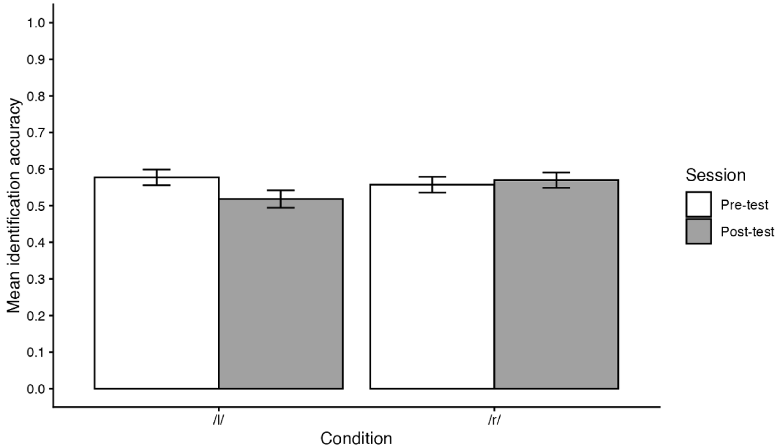

Figure 5. Mean /l/ and /r/ identification accuracy for the pre-test and post-test

that the identification accuracy for /l/ decreased from the pre-test to the post-test while that for /r/ did not change.

4. Discussion

The present study investigated whether providing explicit articulatory feedback during /l/–/r/ pronunciation training facilitates Japanese learners' perception of this contrast. To address this question, an articulatory-feedback condition was compared with an auditory-only feedback condition. The mixed-effects analyses did not reveal reliable overall differences in ID test accuracy between the feedback and no-feedback groups. Thus, within the relatively short training period, the articulatory diagrams and L1 explanations did not seem to provide any clear additional benefit over simple listen-and-repeat practice.

Instead, individual differences in training success emerged as a predictor of test performance. Learners who achieved higher accuracy during the training phase tended to show better identification performance. Crucially, a significant interaction between Session and Phoneme showed that the two sounds changed differently over time. Identification accuracy for /l/ declined from pre- to post-test, whereas accuracy for /r/

remained relatively stable across sessions.

Overall, this pattern suggests that training may have made learners' /l/ category less stable. Some tokens that had previously been identified as /l/ appear to have become more ambiguous after training and were sometimes perceived as /ɹ/ or as /ɾ/, leading to increased confusion rather than a clear improvement in /ɹ/ perception. This interpretation is compatible with accounts that assume a shared phonological space for Japanese /ɾ/ and English /l/ and /ɹ/, in which ambiguous tokens can activate multiple liquid categories during perception.

Finally, accuracy systematically differed across segmental positions within the word. This indicates that phonological context also influenced the difficulty of the /l/–/ɹ/ contrast. The findings in this study provide a more detailed picture of how Japanese learners' perception of the English /l/–/ɹ/ contrast changes following a short session of pronunciation training.

In the overall analysis, the mixed-effects models did not reveal a reliable difference in identification accuracy between the feedback and no-feedback groups. This pattern suggests that, within the relatively short training period, providing visual information about tongue position and airflow and brief Japanese explanations of how to produce /l/ and /ɹ/ did not provide any clear additional benefit beyond simple listen-and-repeat practice.

At the same time, individual differences in training success emerged as a key factor associated with test performance. Learners who achieved higher accuracy during the training phase tended to perform better on the identification test. The significant interaction The significant interaction between Session and Phoneme indicated that the two target sounds did not change in parallel over time. Identification accuracy for /l/ decreased from pre- to post-test, whereas accuracy for /ɹ/ showed increase. As a result, an /ɹ/ advantage emerged at post-test.

From PAM-L2 perspective, this asymmetric pattern is compatible with the idea that Japanese learners initially represent Japanese /ɾ/ and English /l/ and /ɹ/ within a common phonological category space. On this view, L2 learning involves a gradual restructuring of this shared /l/–/ɹ/–/ɾ/ category space. Previous studies have suggested that, for Japanese learners, English /ɹ/ tends to be perceived as more distinct from Japanese /ɾ/ than English /l/ is. Therefore, /ɹ/ is more likely to be represented as a separate L2 category, whereas /l/ often remains in a region of phonological overlap with both the L1 tap and /ɹ/ (e.g., Aoyama et al., 2004). In light of these proposals, the present training may have reshaped the shared /l/–/ɹ/–/ɾ/ space in a way that increased ambiguity for items near the /l/ region rather than producing a straightforward improvement in /ɹ/. Some tokens that had previously been heard as /l/ may have begun to be mapped variably to /l/, /ɹ/, or /ɾ/, which would reduce /l/ identification accuracy without yet yielding a clear gain for /ɹ/.

This interpretation aligns with Gor's fuzzy lexical representations hypothesis, which attributes L2 listening difficulties to fuzzy phonological and phonolexical representations. According to this view, learners encode difficult L2 contrasts with insufficiently specified features, so that multiple sublexical and lexical candidates are activated in parallel and compete for recognition. Applied to the present data, when Japanese learners attempt to identify /l/ tokens after training, they may face competition among at least three viable candidates—/l/, /ɹ/, and Japanese /ɾ/—within a partially restructured liquid space. The resulting increase in representational ambiguity and candidate competition could temporarily destabilize /l/ identification, even as the representation of /ɹ/ becomes more robust. In this sense, the /l/ accuracy and the/ɹ/ accuracy reflect a transitional stage in the reorganization of the /l/–/ɹ/–/ɾ/ system, rather than a complete absence of learning. One possible explanation for the lack of a reliable difference between the feedback and no-feedback groups is that the articulatory diagrams and L1 explanations did not sufficiently support learners' self-monitoring of their own productions. A large body of work has shown that L2 learners often struggle to evaluate their own pronunciation accurately. Even advanced learners have difficulty judging the quality of their productions in a way that matches expert ratings, and these findings suggest that learners' self-assessments are strongly constrained by L1-based perceptual biases.

Recent reviews of technology-mediated pronunciation instruction argue that learners' self-monitoring is grounded in their own perception of their speech. As a result, external feedback that makes errors salient is considered crucial for effective learning. In our training, the feedback was limited to static visual and written information about how to articulate /l/ and /ɹ/, presented as diagrams of tongue position and airflow, accompanied by brief L1 explanations in Japanese about how to pronounce /l/ and /r/. In other words, the intervention provided "how to articulate" information but did not give them explicit feedback on whether their pronunciations were correct on each trial.

By contrast, studies using automatic speech recognition (ASR) or other forms of computer-assisted pronunciation training (CAPT) typically supply objective, trial-by-trial information about learners' output. This information can take the form of correct/incorrect judgments, simple scores for each attempt, or visual comparisons between learner and native productions. Such feedback, which clearly shows how learners' speech differs from the target, has been shown to promote noticing and autonomous adjustment of pronunciation. A recent meta-analysis on ASR-based pronunciation training reports a medium overall benefit of ASR relative to non-ASR conditions. It also finds larger effects when the ASR system provides explicit corrective feedback (e.g., highlighting mispronounced segments or displaying waveforms/spectrograms) than when it merely transcribes learner speech or indicates correct/incorrect responses. Overall, these findings suggest that articulatory information alone, as implemented in the present feedback condition, may not be sufficient by itself to bring about substantial changes in learners' phonological categories. More robust gains may require combining articulatory information with richer, outcome-oriented feedback that makes the gap between learners' productions and target forms perceptually clear.

In addition, work on visual and articulatory feedback indicates that this type of information tends to be effective when integrated with auditory models and when learners are helped to map what they see (e.g., tongue position or formant patterns) onto what they hear. In our relatively short training, learners were exposed to static articulatory diagrams and written instructions, but they did not receive dynamic visualizations of their own speech or explicit guidance on how changes in articulation would affect the acoustic signal. This design likely limited the extent to which the provided diagrams could support the refinement of phonological representations or strengthen the link between production and perception. As a result, it may have contributed to the absence of a clear advantage for the feedback group over simple listen-and-repeat practice.

A decrease in /l/ identification and no change for /r/ suggests that it may have destabilized learners' /l/ category. Rather than producing parallel gains for both sounds, it may have left /l/ more unstable or more in competition with /ɹ/ in learners' perception. One plausible interpretation is that some tokens that had previously been heard as /l/ began to be perceived as /ɹ/. It may increase competition between the two categories instead of a straightforward improvement in /l/ perception. From a pedagogical perspective, this pattern indicates that /l/ may require more sustained and more tightly focused support than /ɹ/. Learners may benefit from additional practice that reduces ambiguity around /l/. For example, by using minimal pairs that isolate /l/ in different word positions and by explicitly contrasting /l/ with the Japanese liquid. Teachers can also draw learners' attention to the subtle acoustic and articulatory cues that distinguish /l/ from both /ɹ/ and /ɾ/, with the aim of stabilizing a dedicated /l/ category rather than allowing /l/ responses to drift toward /ɹ/. At the same time, /ɹ/ identification did not change over training, in contrast to the decline observed for /l/. This stability is compatible with the view that even relatively simple high-variability listen-and-repeat activities can begin to support the reorganization of L2 categories for segments that are acoustically more distinct from their L1 counterparts, even if clear improvements were not yet observable in the present short intervention. Furthermore, learners who performed better during the training phase also tended to achieve higher scores on the identification test. This association suggests that training performance largely reflected pre-existing individual differences in /l/–/ɹ/ perception rather than a strong additional effect of the brief intervention.

Several limitations of the present study point to directions for future research. The training consisted of a single short session, which may not have been sufficient for differences between feedback types to emerge. Longer-term interventions that span multiple training sessions will be needed in future research. Such designs should also include delayed post-tests. This would make it possible to determine whether articulatory feedback yields cumulative advantages over time and whether any benefits are retained. In addition, the present implementation of feedback provided only static articulatory information and did not support systematic self-monitoring of learners' own productions.

Future studies should therefore compare the current paradigm with versions that provide richer, performance-contingent feedback. For example, they could incorporate

ASR-based scoring, color-coded evaluations of individual segments, or guided comparison between learners' recordings and native models. Such designs would make it possible to test whether making the discrepancy between learners' output and target forms more salient leads to stronger gains in both perception and production. Finally, the present outcomes were assessed solely with an auditory identification test on a limited set of word-level stimuli. Further work is therefore needed to examine how similar training affects /l/–/r/ production, perception in more varied lexical and sentence contexts, and learners with different proficiency levels or L1 backgrounds.

5. Conclusion

The present study examined whether providing explicit articulatory feedback during /l/–/r/ pronunciation training would facilitate Japanese learners' perception of this contrast. Across both groups, no reliable overall differences in identification accuracy were observed. This pattern suggests that, within the short intervention period, adding static visual information about tongue position and airflow together with brief L1 explanations did not yield a clear advantage over simple listen-and-repeat training. At the same time, learners who achieved higher accuracy during the training phase tended to perform better on the identification test. This suggests that how successfully learners carry out the training task itself is a key factor in determining learning outcomes.

The interaction between Session and Phoneme revealed an asymmetric change in the two target sounds: /l/ identification declined from pre- to post-test, whereas /r/ identification remained at approximately the same level. This pattern suggests that some items that had previously been heard as /l/ became more ambiguous after training and were sometimes perceived as /r/ or as /ɾ/, leading to increased confusion rather than a clear improvement in /r/ perception. This pattern is consistent with accounts based on PAM-L2 and fuzzy lexical representations, which propose that Japanese /ɾ/, English /l/, and English /r/ are accommodated in a shared phonological space and that L2 learning involves a gradual restructuring of this system. From this perspective, the present findings may reflect an intermediate stage in the reorganization of the /l/–/r/–/ɾ/ system. At this stage, the category for /l/ appears particularly unstable:

/l/ occupies an ambiguous region that overlaps with both /r/ and the Japanese liquid, so that multiple liquid candidates can be activated when learners try to recognize /l/.

Pedagogically, the results suggest that even simple listen-and-repeat activities using varied word tokens can help reshape learners' L2 phonological categories. This type of practice may be especially useful for sounds that are relatively distinct from learners' L1 categories, such as English /r/ for Japanese learners. However, the decline in /l/ identification indicates that /l/ requires more sustained and targeted instructional support, including focused practice on /l/ in different word positions and explicit contrasts with the Japanese liquid.

The fact that the feedback and no-feedback conditions did not differ in overall accuracy highlights the need for pronunciation training that goes beyond static articulatory information and incorporates richer, outcome-oriented feedback—such as ASR-based or visually enhanced knowledge-of-results—that makes discrepancies between learners' productions and target forms perceptually salient.

Acknowledgement

I am sincerely grateful to Professor Manabu Arai of the Faculty of Economics for providing valuable advice on the framework and approach of this research. I am also indebted to Associate Professor Yumiko Mizusawa. Her guidance enriched my knowledge, and her invaluable advice supported me. Finally, I thank my family and colleagues for supporting me all along.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*, 233–250.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using* R. Cambridge University Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, *36*(4), 506–516.

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Frontiers in Psychology*, *12*, 1–19. phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/and /r/. *Journal of Phonetics*, *32*, 233–250.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309–365.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, *19*(1), 37–66.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25–42.

R Core Team (2024). *R: A Language and Environment for Statical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*(2), 595–633.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708.

Shao, Y., Saito, K., & Tierney, A. (2023). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly*, *57*(1), 33–63.

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of Interspeech 2011* (pp. 161–164). International Speech Communication Association.

Tyler, M. D. (2021). Perceived Phonological Overlap in Second-Language Categories: The Acquisition of English /r/ and /l/ by Japanese Native Listeners. *Languages*, *6*(1), 4.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). (https://www.pcibex.net/).