# The Effect of Pronunciation Training Using Automatic Speech Recognition on Japanese EFL Learners' Listening Comprehension and Speech

**Fumio OZAWA**

## 1. Introduction

Japanese learners of English (JLE) are generally considered to have low English proficiency. In fact, it has been reported that the level of English proficiency of JLE is lower than that of other Asian nations (EF, 2023). This paper examine the reasons behind JLE's low English proficiency and explores strategies for improvement. Past research has explored many factors that can influence SLA and attested that one of the most influential factors is the age of acquisition, i.e., at what age leaners started to learn the second language. According to Johnson and Newport (1989), language learning efficiency gradually declines with age, especially after puberty. Some argue that it is difficult for L2 learners to acquire native-like pronunciation and intonation (Tahta et al., 1981). This age-related difficulty is related to various factors, including biological factors such as reduced brain plasticity, environmental influences such as poverty of input, and psychological factors (Tu et al., 2015). Since Japanese children usually start learning English after they have developed a robust first language (L1) system, the age of acquisition is likely to be a significant factor for their difficulty in learning English.

Another important factor in SLA is L1 interference, which is the interference of L1 knowledge on the learning of a second language. In terms of L2 vocabulary learning, the process of creating new word-meaning mapping in L2 can be complicated because the L1 vocabulary is already well-established and closely linked to the L1. It is known that L1 interference also influences multiple aspects of second language acquisition, including phoneme perception, production accuracy, and lexical access (Derakhshan &

Karimi, 2015). Such L1 influences on SLA can be positive and negative. For example, the L1 knowledge is helpful for L2 vocabulary learning in learning cognates or loanwords such as 'computer' or 'table'. In contrast, negative interference can often be observed with the regard to learning of phonological rules. For example, French speakers learning English often find it difficult to pronounce the first /h/ sound in words like 'hungry.' In this case, French (L1) phonological categories do not correspond with English (L2) phonemes, so transferring L1 knowledge to L2 results in negative consequence.

One classic example of L1 interference for Japanese learners is the perception and pronunciation of /l/ and /ɹ/ sounds, which JLE tend to have great difficulty in distinguishing. For example, JEL often mispronounce and misunderstand words like 'light' and 'right' or 'collect' and 'correct'. According to the *Perception Assimilation Model* (Best, 1995), this difficulty is attributed to the fact that the two sounds, which are perceptually close for Japanese learners, are assimilated into a single phonetic category (the Japanese flap /ɾ/). Phonetically, /l/ involves contact between the tongue and the alveolar ridge, while /ɹ/ requires a retroflex or bunched tongue posture. Both sounds do not exist in Japanese and thus are known to be challenging for JLE to perceive as well as pronounce. When L2 sounds differ significantly from L1 sounds, learners are more likely to perceive the L2 sounds accurately and form new phonological categories. However, when L2 sounds closely resemble existing L1 categories, L2 learners often struggle to perceive these sounds accurately. These things make it difficult for Japanese speakers to distinguish the /l/ and /ɹ/ sounds in English.

The studies mentioned above indicate that Japanese L2 learners have less accurate phonological representations for /l/ and /ɹ/ in their head. These phonological representations are accessed not only when the learners perceive the sounds but also when they produce them. Thus, those who can perceive the differences of two phonemes are usually able to distinguish these sounds in their speech. In accordance to this, previous studies showed that pronunciation training can help develop perception abilities (Linebaugh & Roche, 2015). Regarding how L2 learners end up with less phonological representations, Gor et al. proposed Fuzzy Lexical Representation (FLR) Hypothesis, which explains that L2 learners may develop weak form-meaning mappings due to their inability to establish stable phonological representations (Gor, Cook, Bordag, Chrabaszcz,

& Opitz, 2021). The FLR Hypothesis states that L2 words that have phonological categories that are absent in the L1 are more likely to have fuzzy phonological representations. This can lead to incorrect pronunciation of L2 words, such as 'parent' and 'parrot.'

An important question is how L2 learners can improve these fuzzy phonological representations. Several past studies have shown a positive effect of teaching or training on improving L2 learners' phonological representations. For example, according to Shao, Saito, and Tierney (2023), reading aloud has been shown to improve listening skills. Shinohara and Iverson (2018) also found that learners' ability to differentiate between /l/ and /r/ phonemes is enhanced when they receive discrimination training that involves minimal pairs of these phonemes. Shimamune and Smith (1995) show that precise pronunciation of English words is essential to precise understanding. It is difficult for even proficient Japanese English speakers to distinguish between the /l/ and /r/ sounds (Goto, 1971). The similarity between the English /l/ and Japanese /r/ sounds is the source of this difficulty (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). According to Best (1995), second language learners find it difficult to learn pronunciation when L2 and L1 sounds are similar because they tend to hear the L2 sound as a variation of the L1 sound. On the other hand, according to the Speech Learning Model (Flege, 1995), the greater the phonological and linguistic distance between the L1 and the target language, the easier to acquire. In addition, prior research suggests that /r/ is easier to learn than /l/. According to Guion et al. (2000), training has demonstrated greater improvement in the perception of /r/ than /l/. Similarly, learners who have spent more time in the United States demonstrate higher proficiency in /r/ discrimination (Aoyama et al., 2004).

Another source of the JLE's difficulty with English is the shortage of opportunities to produce and be exposed to English in Japan. Japanese do not have enough opportunities to use English outside the classroom because Japanese rarely meet native English speakers. There is a problem in the English learning classroom. In the classroom, learners may feel anxious about speaking English in a classroom because of factors such as peer pressure, as predicted by the Affective Filter Hypothesis (Krashen, 1995). This would make it difficult for them to actively engage in pronunciation practice. As a

result, acquiring accurate phonological representations becomes even more challenging. Furthermore, many English teachers in Japan are native speakers of Japanese, which means they do not necessarily possess accurate phonological representations themselves, making it difficult for them to teach pronunciation effectively. These challenges highlight the importance and complexity of pronunciation instruction in English education in Japan.

The current research aims to examine the effect of pronunciation training using Automatic Speech Recognition (ASR) on JLEs' listening comprehension as well as speech. ASR allows learners to practice pronunciation without the fear of being observed by others, unlike classroom setting, and enables them to engage in repeated practice at their own pace. Moreover, ASR provides immediate feedback on pronunciation, which is a key feature of the technology. The effectiveness of feedback on pronunciation has been demonstrated in previous studies (Long, 1983). These features make ASR a promising tool for supporting learners in improving their pronunciation skill. This experiment was conducted two tests to measure the effect of training on listening comprehension of L2 vocabulary. The word recognition test to evaluate participants' comprehension of the words they were trained on during the pronunciation training task. The identification test was administered to examine participants' perception of /l/ and /r/ phonemes, which are known to be particularly challenging for JLE. Furthermore, to investigate the relationship between pronunciation and perception, the acoustic analysis was conducted to focus specifically on the phonemes /l/ and /r/, as these sounds pose significant challenges for JEL due to the lack of corresponding phonological categories in their L1.

In addition, the JLEs' speech to examine the effect of training on their pronunciation was analyzed. The categorization of /l/ and /r/ is primarily determined by the frequency of the third formant (F3). Additionally, transition duration and closure duration serve as secondary factors. Furthermore, changes in the frequency of the second formant (F2), which are associated with light and dark articulations and are nearly orthogonal to the /l/–/r/ contrast, also play a role. (e.g., Hattori & Iverson, 2009; Underbakke, Pola, Gottfried, & Strange, 1988; Yuan & Liberman, 2011). We focused specifically on the third formant (F3) because the ability of Japanese learners of English to accurately distinguish between /l/ and /r/ is most effectively predicted by cues

related to the third formant (F3), suggesting that the greater their reliance on this cue, the more likely they are to exhibit performance similar to that of native English speakers (Ingvalson et al., 2012).

## 2. Method

### 2.1 Participants

Participants were recruited from the student community of Seijo University. Twenty participants in total took part in the experiment. Their English proficiency levels, according to the Common European Framework of Reference for Languages (CEFR), ranged from A2 to B2. The distribution of proficiency levels was as follows: A2 (n = 3), B1 (n = 10), and B2 (n = 6). One participant had not taken any English tests, whose level of English proficiency was therefore unknown.

### 2.2 Material

The effect of training was tested using three categories of words. The first category is *loan words* (LW). These are borrowed words that are commonly written as *Katakana* in Japanese (e.g., liquid). The next category is *the words* that are difficult to pronounce (WDP). These words are difficult to pronounce for L2 learners because of irregular phoneme-letter correspondence (e.g., subtle). The last category is the words that contain the /l/ and /r/ phonemes (LRW). The /l/ and /r/ phonemes appeared either at the beginning or in the second syllable of the word (e.g., crab). We selected 10 lexical items to test in the training task for each category.

### 2.3 Procedure

Before starting the experiment, we provided detailed instructions about the experiment and asked them to sign a consent form. To evaluate an effect of training on their listening comprehension, we conducted two tests; the word recognition test and the identification test. The word recognition test was conducted after the training and one week later. The identification test was conducted before and after the training as well as one week later. Each experimental session typically took approximately 40 minutes to complete.

### 2.3.1 Training Task

Participants saw a word that they were asked to pronounce words displayed on a screen. (See Appendix A) Participants pressed a 'recording' button before pronouncing the word to ASR. We used ASR on Google Translate (https://translate.google.co.jp/). A 'play' button is also presented on the monitor is, which plays the recording of native speaker's model pronunciation of the word. Participants were instructed to press this play button and listen to the correct pronunciation after they failed to make their pronunciation correctly recognized by the ASR. They were required to repeat pronunciation training until their speech was correctly recognized. This process was repeated up to five times until the word was correctly recognized. Once it was done so, they proceeded to the next trail by pressing a 'next' button. They were instructed to proceed to a next trial if they failed to have their pronunciation recognized five times. Thirty words in total were presented to each participant, ten from each category. Figure 1 illustrates the experimental procedure.

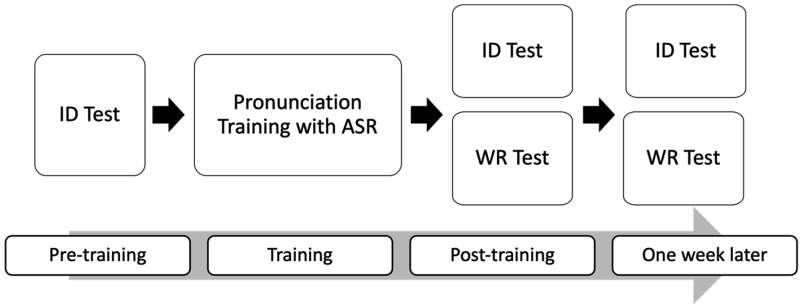


Figure 1. Experimental procedure illustrating the flow of pre-training, training, post-training, and testing after one week.

### 2.3.2 Word Recognition (WR) Test

The WR test examined participants' listening comprehension of the words that they were trained on during the training task. We are also interested in whether a training effect, if observed, would persist for a week or not. The words were presented auditorily, and participants were asked to transcribe those words on a sheet. The same thirty words that were presented during the training task in a random order were tested. This test was

administered immediately after the training task and one week later.

### 2.3.3 Identification (ID) Test

The ID test examined participants' perception of /l/ and /r/ phonemes. Participants listened to audio recording and were asked to select the correct word from the minimal pairs of /l/ and /r/ phonemes shown on the monitor (see Appendix B for the full list of items). Since none of the words used in the identification test was included in the training task, this test is intended to examine whether the training effect, if found, can generalize to the words other than those trained. This test was conducted before the training session, immediately after the session, and one week later. Forty words were presented in the test in a random order.

## 3. Data Analysis

We conducted analyses for the two tests on the correct responses (coded as 0 or 1) using Generalized Linear Mixed Models (Baayen, 2008). For the analysis of the results of the WR test, we included Category and Session (Post / After one week) as fixed factors. For the analysis of the results of the ID test, we included Type (/l/ or /r/) and Session (Pre-training / Post-training / After-one-week) as fixed factors. The interaction between the two fixed factors was initially included but was removed from the model as it was not meaningful ($p > 0.10$). In the initial model, we also included random effects for participants and items and included random slopes of Type and Session for the random effects. We selected the model with the optimal random slope structure by backward selection approach. We used the lme4 function of the lme4 package in R (version 4.3.1; CRAN project; The R Core Team, 2022). We report coefficients ($\beta$), $t$-values ($t$), standard errors (SE), and $p$-values ($p$).

### 3.1 Results of Word Recognition (WR) Task

The data from twenty participants were analyzed. The responses were scored as 'correct' if participants correctly transcribed the word that they heard. Figure 2 shows the rate of correct answers in the WR test for each level of Category for post-training and

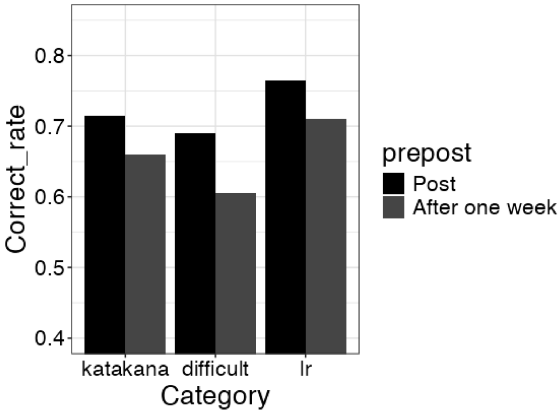


Figure 2. The rate of correct answers in the WR test for each Category for post-training and after one week

after one week.

This thesis was analyzed the number of correct answers with Category and Session as fixed factors. Among the three levels of Category, the loan words were set to be the reference level from which the differences for the other two levels were examined. The analysis showed a significant main effect of Session, showing significant decrease from the post-training Session to the one-week-later follow-up session ($\beta$ = -0.38, SE = 0.14, $z$ = -2.75, $p < 0.001$). There was no significant effect of Category (WDP: $p = 0.23$, LRW: $p = 0.71$) and no interaction between Session and Category (WDP: $p = 0.68$, LRW: $p = 0.95$).

The results showed that the accuracy rate declined after one week. Since this drop is considered as loss of a training effect. Thus, this in turn suggests that training did help the participants to perceive the trained words and indicates that pronunciation training was effective in improving participants' ability to recognize words. However, the effect did not persist for one week. Yet, it cannot be concluded that the effect completely disappeared because no baseline test was conducted prior to the training.

There are three participants whose accuracy rate was very low (below 50%) in the post-training session of the word recognition test. Since it is likely that they were unfamiliar with the trained words before the training session, which would negatively

affect learning of the pronunciation of the words, we decided to exclude these participants from further analyses.

### 3.2 Results of Identification Test

The ID test investigated accuracy in perceiving the minimal pairs of /l/ and /r/ phonemes for pre-training, post-training, and after one week. As stated above, the data from three participants were excluded, resulting in the data of seventeen participants for the ID test.

We analyzed the responses in the ID test, Pre-training, Post-training and the one-week-later session. Figure 3 shows the rate of correct answers separately for the words with the /l/ phoneme and those with /r/ phoneme in the ID test for Pre-training and Post-training. The results showed a significant effect of Session. The analysis revealed significant improvement in accuracy from Pre-training to Post-training ($\beta = 0.24$, SE = 0.12, z = 1.99, $p = 0.05$). There was no significant effect of Type ($p = 0.30$). There was no interaction between Secession and Type ($p = 0.45$).

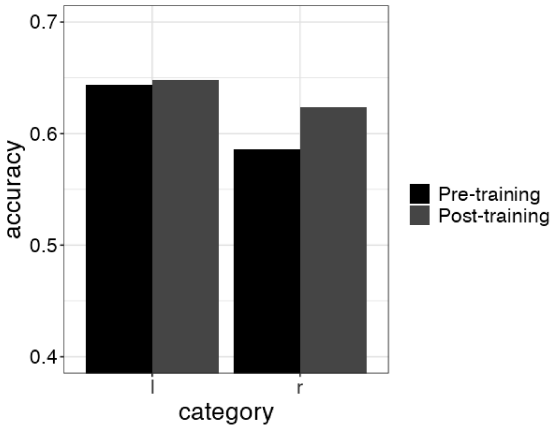


Figure 3. The rate of correct answers in the ID test for Pre-training and Post-training.

Figure 4 presents the rates of correct answers in the ID tests for Post-training and After-one-week. No effect of Session ($p = 0.09$), ($p = 0.19$), and no interaction between Session and Type were observed ($p = 0.19$).

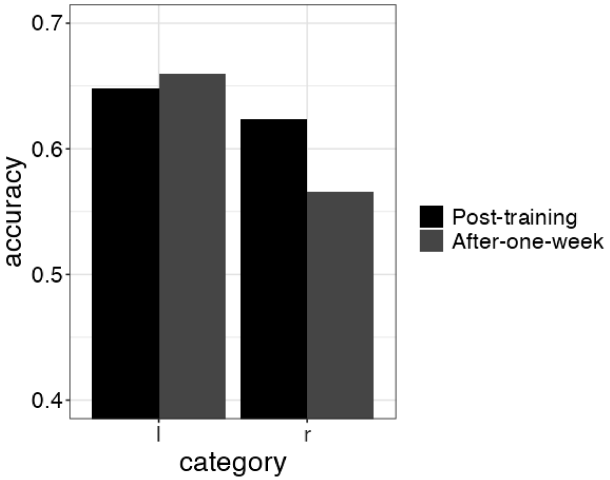


Figure 4. The rate of correct answers in the ID test for Post-training and After-one-week

The improvement in accuracy from pre-training to post-training suggests that pronunciation training positively affected participants' listening comprehension ability of the /l/ and /r/ phonemes. We next report the acoustic analysis focusing on /l/ and /r/ in order to examine the possibility that the observed changes in perception may be accompanied by the changes in speech production during the training.

### 3.3 Acoustic Analysis

For the analysis of acoustic properties of the speech produced during the training session, we examined the F3 values of the extracted consonant segments. The mean F3 value across the entire extracted range was calculated to provide consistent assessment of the acoustic properties of the target consonants. First, we selected only the LRW words and extracted the pronunciation of the /l/ and /r/ phonemes. The selected speech data were then converted to mono, to avoid the influence of differences in volume and frequency between the left and right channels on the analysis results, as part of preprocessing before being analyzed using the speech analysis software Praat (Boersma & Weenink, 2011, 2015). Next, we extracted the consonantal portions of the /l/ and /r/ phonemes. Vowel segments were excluded by defining the boundary of /l/ and /r/ as

the point immediately preceding the rise in F3 (third formant) values.

We compared the acoustic properties of the /l/ and /r/ pronunciation for the first attempt during each training trial and those for the successful attempt, regardless of how many times they tried. We excluded the trials in which participants succeeded in pronouncing a target word for the first attempt and also those in which they failed to make the ASR recognize their pronunciation at the 5th attempt.

Specifically, /l/ is characterized by relatively higher F3 values, whereas /r/ exhibits significantly lower F3 values. This acoustic distinction serves as a critical cue for native English speakers in perceiving and differentiating these two sounds (Hattori & Iverson, 2009; Underbakke, Pola, Gottfried, & Strange, 1988; Yuan & Liberman, 2011). Thus, we predict that if pronunciation training using ASR in fact affects JLEs' production, their pronunciation of /l/ should have higher F3 value, approximating to the F3 value of native speakers' production after training where as that of /r/ should have lower F3 value after training.

We analyzed the F3 values using Linear Mixed Effect models (Baayen, 2008). We included Session (First attempt/Successful attempt) and Type (/l/-/r/) as fixed factors. The interaction between the two fixed factors was initially included but was removed from the model since it was not significant ($p > 0.10$). In the initial model, we included random effects for participants and random slopes of Type and Session for the two random effects. We selected the model with the optimal random slope structure by backward selection approach. Table 1 presents the descriptive statistics of the F3 values of the /l/ and /r/ phonemes for native speaker's model pronunciation, L2 learners' pronunciation in the first attempts and the successful attempts.

Figure 5 shows the mean F3 values for /l/ and /r/ in the First attempt and

Table 1. Descriptive Statistics Formants of /l/ and /r/ for Native's model, JLE's first attempt, and JLE's successful attempt

| | Native's model | | JLE's first attempt | | JLE's successful attempt | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD |
| l | 2798.25 | 236.41 | 2403 | 265.87 | 2470.39 | 326.3 |
| r | 2169.17 | 199.26 | 2283.34 | 428.64 | 2187.94 | 317.92 |

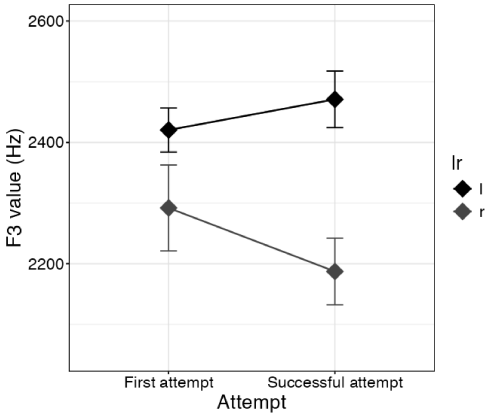


Figure 5. The Interaction between Session (First attempt/Successful attempt) and Type (/l/ and /ɾ/)

Successful attempt. The main effect of Session was not significant ($p = 1.38$), nor was the main effect of Type ($p = 0.78$). However, there was significant interaction between Session and Type ($\beta = -194.18$, SE = 79.26, $t = -2.45$, $p = 0.02$).

In order to examine the pattern of the interaction, we analyzed the data separately for each level of Type. The result showed that the difference in the F3 value between the first attempt and the successful attempt was not significant for the /l/ pronunciation ($\beta = 68.96$, SE = 60.42, $t = 1.14$, $p = 0.26$) but it was significant for the /ɾ/ pronunciation ($\beta = -102.11$, SE = 48.15, $t = -2.12$, $p = 0.04$). This demonstrates that the F3 value of /l/ did not change from first attempt to successful attempt, whereas the F3 value of /ɾ/ decreased significantly from first attempt to successful attempt during the pronunciation training.

## 4. Discussion

The current study investigated the effect of pronunciation training with ASR on the perception of English words by JLE. The results are summarized as follows. The results of the WR test show significant decrease in comprehension accuracy from the post-training test to the test after one week. This means that the participants answered

more correctly immediately following the training than after one week, suggesting that the training improved their listening comprehension ability. It also suggests that the training effect did not persist for one week although it is not clear that the training effect disappeared completely. It is important to consider how long the effect of training will persist.

The result of the ID test showed that participants made more correct answers in the post-training test than in the pre-training training. This demonstrates that ASR-assisted pronunciation training was effective and improved the participants' ability to distinguish between the /l/ and /ɾ/ phonemes. The analysis also indicated that the accuracy was not significantly different between the post-training session and the one-week-later session. This suggests that the effects of pronunciation training did not disappear completely after one week. It also show that participants were able to discriminate the words which were not trained, so the training effect on the perception of /l/ and /ɾ/ was generalized.

An acoustic analysis of the third formant (F3) of /l/ and /ɾ/ was conducted to investigate potential change in pronunciation during the training session. Our hypothesis is that L2 learners may be able to learn correct phonological representations by practicing their pronunciation using ASR. This suggests that the improvement in L2 learners' listening comprehension should have resulted from the improvement in the relevant phonological representations through pronunciation training. Therefore, it is crucial to examine whether their pronunciation has in fact improved during the training. The analysis revealed an asymmetrical influence of training session on the /l/ and /ɾ/ pronunciation. It showed that while there was no significant change in the F3 value of the /l/ pronunciation, the F3 value of /ɾ/ decreased significantly from the first attempt to the success for attempt, approaching the F3 value of the native English speaker's model pronunciation.

No change in the F3 value of /l/ following the training can be attributed to the fact that English /l/ is assimilated to the Japanese flap /ɾ/. This finding is consistent with Best's (1995) Perception Assimilation Model, which posits that non-native phonemes are perceived in relation to the closest existing categories in the native phonological system. In the case of the /l/ and /ɾ/ pronunciation, /l/ appears to be assimilated

into the Japanese flap /ɾ / category. In contrast, the change in the F3 value of /ɹ/ be interpreted by Flege's (1995) Speech Learning Model (SLM). This model posits that the formation of a new phonological category is possible if the distance between the non-native and native phonemes is adequately large. The distance between Japanese flap /ɾ/ and English /ɹ/ is larger, so it is easier for JLE to acquire /ɹ/ than /l/. The sound /ɹ/ is phonetically and linguistically distant from the Japanese flap /ɾ/, and this distance likely contributed to the observed shift in the F3 value of /ɹ/ toward that of native English speakers in this study. In contrast, /l/ is phonetically similar to the Japanese flap /ɾ/, so it may lack a clearly established phonological model in JLEs. As a result, /l/ is assimilated into the Japanese flap /ɾ/. In this study, the result from acoustic analysis is consistent with the SLM. There was no significant decrease from the first attempt to the successful attempt in /l/ whereas there was significant decrease from first attempt to successful attempt in /ɹ/. Consequently, our participants' /ɹ/ were more likely to adapt the /ɹ/ pronunciation, approximating to the native English F3 value during pronunciation training, whereas /l/, was assimilated with the Japanese flap /ɾ/.

Since the one-week-later follow-up session in this study showed decline in accuracy in the WR test. The future research should consider creating experimental groups with varying time intervals to evaluate the persistence of the training effect.

The difficulties faced by JLEs in acquiring accurate English pronunciation concern not only consonants but also vowels. Research has shown that vowels also pose significant challenges for JLE. For instance, Ingram and Park (1997) highlight that JLE often struggle with perceiving and producing certain vowel contrasts in English, such as /i/ versus /ɪ/ or /æ/ versus /ɛ/. These difficulties also arise due to the phonetic and phonological differences between Japanese and English vowel systems. Japanese has a relatively small inventory of vowels compared to English, which can result in assimilation of multiple English vowels sounds into a single Japanese category. This phenomenon leads to reduced perceptual sensitivity and difficulty in achieving native-like vowel production. Given these findings, we consider it to be valuable to extend the scope of research beyond consonants, by examining whether the ASR-based pronunciation training used in this study can also be effective for improving vowel acquisition among JLEs. Future experiments could include minimal pair training of vowels, similar to the

approach used for /l/ and /r/.

The future research should consider creating experimental groups with varying time intervals to evaluate the persistence of training effects since the one-week follow-up period in this study showed a decline in accuracy in the WR test.

To conclude, this thesis demonstrated that pronunciation training using ASR proved effect in improving the Japanese learners' listening comprehension of English words. It improved the ability of JLE to distinguish between /l/ and /r/. The results indicated that pronunciation training using ASR is useful for JLEs to develop more accurate phonological representations in English, which are accessed on their perception as well as production of the L2.

**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, colleagues and friends 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. Cambridge: Cambridge University Press.

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

Boersma, P., &Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.2) Retrieved from http://www.praat.org/

Derakhshan, A., & Karimi, E. (2015). The Interference of First Language and Second Language Acquisition. *Theory and Practice in Language Studies*, *5*(10), 2112.

EF Education First. (2024). *EF EPI 2024: EF English proficiency index*. https://www.efjapan.co.jp/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi site/reports/2024/ef-epi-2024-japanese.pdf

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). York Press.

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.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r.' *Neuropsychologia*, *9,* 317–323.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*, 2711–2724.

Hattori, K., & Iverson, P. (2009). English /r/–/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. The Journal of the Acoustical Society of America, 125(1), 469–479.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly, 17*, 359–382.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics*, *18*(3), 375–394.

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*, 33–63.

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics, 66*, 242–251. Shimamune and Smith (1995)

Tahta, S., Wood, M., & Loewenthal, K. (1981). Age Changes in the Ability To Replicate Foreign Pronunciation and Intonation. *Language and Speech*, *24*(4), 363–372.

Tu, L., Wang, J., Abutalebi, J., Jiang, B., Pan, X., Li, M., Gao, W., Yang, Y., Liang, B., Lu, Z., & Huang, R. (2015). Language exposure induced neuroplasticity in the bilingual brain: A follow-up fMRI study. *Cortex*, *64*, 8–19.

Underbakke, M., Pola, L., Gottfried, J., & Strange, W. (1988). Trading relations in the perception of /r/–/l/ by Japanese learners of English. The Journal of the Acoustical Society of

America, 84(1), 90–100.
Yuan, J., & Liberman, M. (2011). /l/ variation in American English: Corpus approach. The Journal of Speech Sciences, 1(2), 35–46.

## Appendix

Appendix A

The words used in the pronunciation training and the Word Recognition test are shown below.

| Loanwords: **LW** | Words that are Difficult to Pronounce: **WDP** | L or R words: **LRW** |
|---|---|---|
| champagne | suitable | crab |
| vaccine | yawn | rub |
| sofa | tomb | elect |
| liquid | diagnosis | crowd |
| vitamin | aisle | grass |
| alcohol | voucher | play |
| genre | subtle | bleed |
| theme | dawn | lice |
| ideology | signature | crime |
| chaos | diameter | ray |

Appendix B

The words used in the Identification test are shown below.

| **Minimal pair words** | | | |
|---|---|---|---|
| alive | arrive | belly | berry |
| clash | crash | clue | crew |
| collect | correct | flame | frame |
| flight | fright | fly | fry |
| glass | grass | glow | grow |
| lace | race | lack | rack |
| late | rate | lead | read |
| light | right | long | wrong |
| lot | rot | love | rub |
| low | row | lust | rust |