

# **Perception and Production of English /r/-/l/ by Adult Japanese Speakers**

Kota Hattori

University College London (UCL)

A dissertation submitted to University College London for the degree of Philosophy of  
Doctor in the Division of Psychology and Language Sciences

**2009**

## Declaration

I, Kota Hattori confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the dissertation.

*Kota Hattori*

服部 恒太

## Abstract

This dissertation examines perception and production of English /r-/l/ by adult Japanese speakers. This programme of research is organized into three sections, termed Study 1, Study2, and Study 3. The first study examined whether category assimilation between English /r-/l/ and Japanese /r/ was predictive of /r-/l/ identification accuracy using an individual difference approach. Japanese speakers were assessed in terms of /r-/l/ identification and assimilation of English /r-/l/ into Japanese /r/, /r-/l/ production, and perceptual best exemplars for /r/, /l/, and /r/. The results demonstrated that, although Japanese speakers strongly assimilated /l/ to /r/, category assimilation was not predictive of English /r-/l/ identification accuracy, and that only Japanese speaker's representations for F3 in /r/ and /l/ was predictive of /r-/l/ identification ability. The second study similarly took an individual difference approach and examined whether there is a relationship between perception and production of /r-/l/ measuring perception accuracies (i.e., identification, discrimination, and perceptual best exemplars) and production accuracies (i.e., acoustic measurements, and recognition accuracy by English speakers). The results demonstrated that perception and production of /r-/l/ were moderately related. However, not all aspects of /r-/l/ perception were incorporated into /r-/l/ production. The third study examined whether one-to-one pronunciation training leads to improvement in production and perception of English /r-/l/ using a multipronged approach (i.e., explicit instructions, real-time spectrograms, and feedback with signal-processed versions of their own productions). The results demonstrated that Japanese speakers could be trained to

produce native-like English /r-/l/, improving to the point that they approached a 100% accuracy ceiling in terms of how well native speakers could identify their productions. However, the training did not improve their English /r-/l/ perception at all.

## Table of Contents

Declaration.....	2
Abstract.....	3
Table of Contents.....	5
List of Tables.....	8
List of Figures.....	10
Acknowledgement.....	26
Chapter 1: General Introduction.....	29
1.1 How do individuals learn the phonemes of their language?.....	29
1.2 Perception of English /r/-/l/ by adult Japanese speakers.....	37
1.3 Dissertation overview.....	40
Chapter 2: English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy.....	42
2.1 Introduction.....	42
2.2 Experiment 1.1: Phoneme identification.....	45
2.2.1 Method.....	45
2.2.2 Procedure.....	46
2.2.3 Results.....	47
2.3 Experiment 1.2: Ratings of spoken accent.....	53
2.3.1. Method.....	54
2.3.2 Results.....	55
2.4 Experiment 1.3: Perceptual mapping of best exemplars.....	57
2.4.1 Method.....	58
2.4.2 Procedure.....	61
2.4.3 Results.....	62
2.5 General Discussion.....	72
Chapter 3: Examination of the relationship between perception and production of English /r/-/l/ by adult Japanese speakers.....	77
3.1 Introduction.....	77
3.2 Experiment 2.1: Phoneme identification.....	83
3.2.1 Method.....	83

3.2.2 Procedure .....	84
3.2.3 Results.....	84
3.3 Experiment 2.2: Phoneme discrimination.....	85
3.3.1 Method .....	85
3.3.2 Procedure .....	88
3.3.3 Results.....	88
3.4 Experiment 2.3: Perceptual mapping of best exemplars.....	93
3.4.1 Method .....	94
3.4.2 Procedure .....	94
3.4.3 Results.....	94
3.5 Experiment 2.4: Production assessment .....	100
3.5.1 Method .....	101
3.5.2 Procedure .....	102
3.5.4 Results.....	105
3.6 General discussion .....	124
Chapter 4: Examination of L2 production training effects on L2 production and perception: Training adult Japanese speakers to produce English /r/-l/ .....	128
4.1 Introduction.....	128
4.2 Experiment 3.1: /r/-l/ Production Training .....	135
4.2.1 Method .....	135
4.2.2 Procedure .....	135
4.3 Results.....	148
4.3.1 English /r/-l/ production intelligibility and goodness by English listeners... 148	
4.3.2 Pre/post test English /r/-l/ production comparison .....	152
4.3.3 English /r/-l/ production comparison between Japanese and English speakers .....	157
4.3.4 English /r/-l/ F3 contrast in word, sentence, and passage.....	159
4.3.5 English /r/-l/ identification.....	163
4.3.6 English /r/-l/ discrimination.....	164
4.3.7 Perceptual mapping of best exemplars .....	167
4.3.8 Links between English /r/-l/ production and perception.....	172

4.3.9 Predictors for English /r/-/l/ production improvement.....	174
4.4 General discussion .....	176
Chapter 5: General Conclusion.....	183
References.....	187
Appendix A: Chapter 2 subject information (part 1).....	206
Appendix B: Chapter 2 subject information (part 2).....	207
Appendix C: Chapter 3 subject information (part 1).....	208
Appendix D: Chapter 3 subject information (part 2).....	210

## List of Tables

<p>Table 2.1 Confusion matrix for bilingual identification. Japanese listeners confused /r/ with /l/ to some extent, but they rarely did so with /r/. They confused /l/ with /r/ and /r/ to the degrees in which they misidentified /r/ as /l/. They similarly confused /r/ with /l/, not with /r/.</p>	48
<p>Table 3.1 Confusion matrix for /r/-/l/ production identification. British English listeners identified most of the /l/ productions by Japanese speakers, suggesting that Japanese speakers produced identifiable /l/. However, the listeners misidentified some of the /r/ productions by Japanese speakers as /l/, suggesting that Japanese speakers had some difficulties in producing English /r/.</p>	105
<p>Table 3.2 Correlations between /r/-/l/ perception and production. The results demonstrated that there was a significant correlation between /r/-/l/ identification and /r/-/l/ production intelligibility. However, there were no correlations between /r/-/l/ identification and the other production measurements. The results also demonstrated that there were no significant correlations between discrimination sensitivities and production measurements.</p>	119
<p>Table 3.3 Correlations between /r/-/l/ best exemplar accuracy and /r/-/l/ production. The results demonstrated that there were no significant correlations between /r/-/l/ best exemplar accuracies and production measurements.</p>	122
<p>Table 4.1 Confusion matrix for pre-test /r/-/l/ production intelligibility. British English speakers listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or /d/. English speakers correctly identified most of the /l/ productions. However, they poorly identified /r/ productions; they misidentified approximately one third of /r/ productions as /l/. These suggest that Japanese speakers were able to produce identifiable /l/, but they had hard time to produce identifiable /r/.</p>	148
<p>Table 4.2 Confusion matrix for post-test production intelligibility. British English speakers listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or</p>	



/d/. English speakers correctly identified most of the /l/ productions; they identified post-test /l/ productions slightly better than they did for pre-test /l/ productions. Similarly, English speakers identified most of the post-test /r/ productions, suggesting Japanese speakers were able to produce identifiable /r/ after the pronunciation training.....149

## List of Figures

- Figure 2.1 Boxplots of English /r/-/l/ identification (i.e., measuring baseline recognition performance) and pairwise bilingual /r/-/l/-/r/ identification (i.e., indicating the degree of category assimilation). Boxplots display the medians and quartile ranges of scores, with outliers marked by circles. Standard errors from the left side of the figure are 0.02, 0.02, 0.01, and 0.01. The results demonstrated that Japanese speakers confused English /l/ and /r/ to the same extent to which they confused English /r/ and /l/. However, Japanese speakers rarely confused English /r/ and /r/.....49
- Figure 2.2 Scatterplot of the relationship between English /r/-/l/ identification accuracy and /l/-/r/ bilingual identification accuracy (i.e., the degree of category assimilation between /l/ and /r/). Higher /l/-/r/ identification scores indicate less confusion between the categories, and a lower degree of category assimilation. The results demonstrated that the two measures were not significantly correlated, suggesting that the category assimilation between /l/ and /r/ is not predictive of English /r/-/l/ identification accuracy.....51
- Figure 2.3 Scatterplot of the relationship between English /r/-/l/ identification accuracy and /r/-/r/ bilingual identification accuracy (i.e., the degree of category assimilation between /r/ and /r/). Higher /r/-/r/ identification scores indicate less confusion between the categories, and a lower degree of category assimilation. The results demonstrated that the two measures were significantly correlated, suggesting that the category assimilation between /r/ and /r/ may be predictive of English /r/-/l/ identification accuracy.....52
- Figure 2.4 Boxplots of the rated accent for Japanese speakers saying “The red robin looked across the lovely lake.” Three ratings were made for each sentence: the degree of /r/-/l/ contrast, /r/-accent, and /l/-accent. Ratings were made on a scale from 1 (poor, indicating a strong non-native accent or poor contrast) to 7 (good,

indicating little non-native accent or high contrast). Standard errors from the left side of the figure are 0.26, 0.28, and 0.15. The results demonstrated that there was a wide range of individual differences in the degrees to which English listeners perceived /r/-/l/ contrast and /r/-accent. However, there was less variance in the degree to which English listeners perceived /l/-accent.....56

Figure 2.5 A spectrogram of *lack* produced by a male British English speaker. The blue area indicates closure duration where F1, F2, and F3 generally stay stable. The yellow area indicates transition duration where F1 starts increasing and hits its peak. Closure duration for /l/ is generally longer and transition duration for the liquid is shorter.....59

Figure 2.6 A spectrogram of *rack* produced by a male British English speaker. The blue area indicates closure duration where F1, F2, and F3 generally stay stable. The yellow area indicates transition duration where F1 starts increasing and hits its peak. Closure duration for /r/ is generally shorter and transition duration for the liquid is longer..... 60

Figure 2.7 Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected similar F1 frequencies for /r/ and /l/. Japanese speakers selected similar F1 frequencies for /r/, /l/, and /r/ which were similar to those selected by English speakers. Standard errors from the left side of the figure are 14.65, 16.79, 20.45, 11.97, and 9.34 (Hz). The results demonstrated that English and Japanese speakers were not different in F1 frequencies for /r/ and /l/. The results also demonstrated that Japanese speakers selected F1 frequencies for /r/ similar to those for /r/ and /l/.....63

Figure 2.8 Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected similar F2 frequencies for /r/ and /l/. Japanese speakers selected F2 frequencies for /r/ and /l/ similar to those selected by English speakers. However, they selected higher F2 frequencies for /r/ than did they for the English

liquids. Standard errors from the left side of the figure are 27.75, 27.37, 61.65, 40.89, and 51.28 (Hz). The results demonstrated that Japanese speakers separated /r/ and /l/ in the F2 dimension as English speakers did. The results also demonstrated that Japanese speakers separated /r/ from English liquids in the F2 dimension.....64

Figure 2.9 Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese.

English speakers clearly selected lower F3 frequencies for /r/ and higher F3 frequencies for /l/ with little variance. Likewise, Japanese speakers selected lower F3 for /r/ and higher F3 for /l/ and /r/ with a wide range of variance. Standard errors from the left side of the figure are 36.33, 57.93, 50.91, 78.06, and 63.30 (Hz). The results demonstrated that Japanese speakers were particularly inaccurate in F3 frequencies for /l/ compared to English speakers. However, the results demonstrated that Japanese speakers separated /r/, /l/, and /r/ in the F3 dimension.....65

Figure 2.10 Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese.

English speakers selected short closure durations for /r/ and long closure durations for /l/. Japanese speakers selected closure durations for /r/ and /l/ similar to those by English speakers with a wide range of variance. They selected short closure durations for /r/ with less variance. Standard errors from the left side of the figure are 8.07, 7.52, 12.86, 8.31, and 3.06 (ms). The results demonstrated that Japanese speakers did not differ from English speakers in closure durations for /r/ and /l/. The results also demonstrated that Japanese speakers separated /r/ from /l/, and /l/ from /r/, but they did not separate /r/ from /r/.....66

Figure 2.11 Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese.

Japanese. English speakers selected long durations for /r/ and short closure durations for /l/. Japanese speakers similarly selected long transition durations for /r/ with a wide range of variance, similar short transition durations for /l/ and /r/ with little variance. Standard errors from the left side of the figure are 7.15, 7.03, 13.25, 3.28, and 6.33 (ms). The results demonstrated that Japanese speakers did not differ from English speakers in transition durations for /r/ and /l/. The results also demonstrated that Japanese speakers separated /r/ from /l/, and /r/ from /r/, but they did not separate /l/ from /r/..... 67

Figure 2.12 Scatterplot of the relationship between the accuracy of the English /r/ and /l/ best exemplars for Japanese speakers along the F3 dimension (i.e., distance from English averages) and their English /r/-/l/ identification accuracy from Experiment 1.1 (i.e., measuring baseline recognition performance). Japanese speakers who demonstrated lower /r/-/l/ identification accuracies generally had more distance from English averages; they were inaccurate in the F3 dimension. Japanese speakers who demonstrated higher /r/-/l/ identification accuracies had short distance from English averages; they were more accurate in the dimension. The results demonstrated that there was a significant correlation between the accuracy of the English /r/ and /l/ best exemplars and /r/-/l/ identification accuracy..... 71

Figure 3.1 Boxplot of English /r/-/l/ identification by Japanese speakers. Japanese speakers listened to naturally-spoken stimuli by British English speakers and identified /r/ and /l/ in word-initial position. The boxplot displays a median and the quartile ranges of scores. Standard error is 0.02. Japanese speakers demonstrated a wide range of /r/-/l/ identification accuracy..... 84

Figure 3.2 F2 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers and English speakers generally demonstrated similar discrimination patterns. They both demonstrated higher discrimination sensitivity at the /r/-/l/ boundary compared to discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0, 0.65, and 0.46 (d') for English speakers and

0.16, 0.18, and 0.21(d') for Japanese speakers. The results demonstrated that Japanese speakers had higher overall sensitivity than English speakers did. The results also confirmed that discrimination sensitivity at the /r/-/l/ boundary was higher than within /r/ and /l/ categories..... 89

Figure 3.3 F3 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers had lower discrimination sensitivity at the /r/-/l/ boundary than English speakers did. However, Japanese and English speakers demonstrated similar discrimination sensitivity within /r/ and /l/ categories. Standard errors from the left side of the figures are 0.48, 0.35, and 0.87 (d') for English speakers and 0.17, 0.23, and 0.40 (d') for Japanese speakers. The results confirmed that Japanese speakers were different from English speakers in terms of F3 discrimination sensitivity at the /r/-/l/ boundary; Japanese speakers were poorer at discriminating /r/ and /l/ at the boundary. The results also confirmed that Japanese and English speakers were similar in terms of F3 discrimination sensitivity within /r/ and /l/ categories; they all were poor at discriminating between /r/ exemplars and between /l/ exemplars..... 90

Figure 3.4 F2 and F3 peak discrimination sensitivity of English (ENG) and Japanese (JPN) speakers at the English /r/-/l/ boundary. The peak discrimination sensitivity in each acoustic dimension was calculated by subtracting the mean discrimination sensitivity of /r/ and /l/ categories from the discrimination sensitivity at the English /r/-/l/ boundary. Higher peak discrimination sensitivity indicates that English and Japanese speakers had high discrimination sensitivity at the /r/-/l/ boundary and low discrimination sensitivity within /r/ and /l/ categories (i.e., they were good at discriminating /r/ and /l/ at the boundary, and they were poor at discriminating between /r/ exemplars and between /l/ exemplars). Japanese and English speakers both demonstrated similar peak discrimination sensitivity in the F2 dimension. However, Japanese speakers demonstrated lower peak discrimination sensitivity in the F3 dimension than English speakers did. Standard errors from the left side of the figures are respectively 0.57, 0.22, 0.62, and 0.28 (d'). The results revealed that Japanese speakers were similar to English speakers

in terms of F2 peak discrimination sensitivity. The results also revealed that Japanese speakers were different from English speakers in terms of F3 peak discrimination sensitivity; Japanese speakers were poor at discriminating /r/ and /l/ at the boundary..... 92

Figure 3.5 Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). The best exemplars of English speakers came from Experiment 1.3 in Chapter 2. Specifically, their best exemplars which were measured using English carrier sentences are presented here. English speakers selected similar F1 frequencies for /r/ and /l/. Japanese speakers selected similar F1 frequencies for /r/ and /l/ with some wide variance for /r/. Standard errors from the left side of the figure are 16.15, 18.07, 31.98, and 13.76 (Hz). The results demonstrated that English and Japanese speakers were similar in F1 frequencies for /r/ and /l/.....95

Figure 3.6 Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected similar F2 frequencies for /r/ and /l/. Japanese speakers chose similar F2 frequencies for /r/ and /l/ which were similar to those by English speakers. Standard errors from the left side of the figure are 49.31, 35.46, 94.97, and 44.01 (Hz). The results demonstrated that, although English speakers overall preferred slightly higher F2 frequencies than Japanese speakers, English and Japanese speakers were fundamentally similar in terms of F2 frequencies chosen for /r/ and /l/..... 96

Figure 3.7 Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected low F3 frequencies for /r/ and high F3 frequencies for /l/. Japanese speakers similarly selected low F3 frequencies for /r/ and high F3 frequencies for /l/ with some wide variance for both /r/ and /l/. Standard errors from the left side of the figure are 51.29, 72.31, 72.44, and 77.45 (Hz). The results demonstrated that English /r/ and /l/ are different from each other in the F3 dimension; /r/ has low F3 frequencies and /l/ has high F3 frequencies. The results also demonstrated that, although Japanese speakers seemed to select F3 frequencies similar to those by English

speakers, Japanese speakers were specifically inaccurate in selecting F3 frequencies for /l/..... 97

Figure 3.8 Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected shorter closure durations for /r/ with little variance and longer closure durations for /l/ with some variance. Japanese speakers selected similar closure durations for /r/ and /l/ with a wide range of variance. Standard errors from the left side of the figure are 13.91, 8.96, 17.97, and 7.02 (ms). The results demonstrated that English and Japanese speakers selected similar closure durations for both /r/ and /l/..... 98

Figure 3.9 Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers generally selected long transition durations for /r/ and short transitions duration for /l/. Japanese speakers similarly selected long closure durations for /r/ with a wide range of variance, and short transition durations for /l/. Standard errors from the left side of the figure are 9.49, 8.39, 16.71, and 2.78 (ms). The results confirmed that English and Japanese speakers selected similar transition durations for both /r/ and /l/..... 99

Figure 3.10 Boxplots of English /r/ and /l/ production intelligibility. Native speakers of British English listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or /d/. Higher proportion correct indicates that English speakers correctly identified /r/ and /l/ productions of L1 Japanese speakers. The intelligibility of /l/ productions varied little; most of the Japanese speakers produced identifiable /l/. However, the intelligibility of /r/ productions widely varied; some Japanese speakers produced identifiable /r/, but some others did not. Standard errors from the left side of the figure are 0.04 and 0.02. The results demonstrated that there was a significant difference between the /r/ and /l/ production intelligibilities; English speakers identified /l/ productions better than /r/ productions..... 106

Figure 3.11 Boxplots of the goodness rating scores for /r/ and /l/ productions of Japanese speakers. British English speakers gave ratings to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese with a scale from 1 (poor)



to 7 (good). The intelligibility scores for /l/ were more or less similar among Japanese speakers, but the intelligibility scores for /r/ substantially varied among Japanese speakers. Standard errors from the left side of the figure are 0.17 and 0.09. The results demonstrated that English speakers perceived /l/ productions better than /r/ productions..... 107

Figure 3.12 Boxplots of F1 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers produced similar F1 frequencies for both /r/ and /l/. Likewise, Japanese speakers produced similar F1 frequencies for both /r/ and /l/. Standard errors from the left side of the figure are 44.02, 26.66, 41.39, and 25.88 (Hz). The results demonstrated that Japanese speakers overall produced lower F1 frequencies than English speakers did. Japanese speakers specifically differed from English speakers in producing F1 frequencies for /r/; they produced lower F1 frequencies for /r/ than English speakers did. However, Japanese speakers did not produce abnormal F1 frequencies. Therefore, it is fair to assume that Japanese speakers are fundamentally similar to English speakers in producing F1 frequencies for /r/..... 108

Figure 3.13 Boxplots for F2 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers clearly made distinctions between /r/ and /l/; they produced low F2 frequencies for /r/ and high F2 frequencies for /l/. However, Japanese speakers made less clear distinctions between /r/ and /l/. Standard errors from the left side of the figure are 27.10, 37.63, 40.66, and 26.70 (Hz). The results demonstrated that Japanese speakers were similar to English speakers in producing F2 frequencies for /l/. However, they differed from English speakers in producing F2 frequencies for /r/ (i.e., Japanese speakers were inaccurate in producing /r/)...... 110

Figure 3.14 Boxplots for F3 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers made clear distinctions between /r/ and /l/; they produced low F3 frequencies for /r/ and high F3 frequencies for /l/. However, Japanese speakers made less clear distinctions between the consonants. They produced F3 frequencies for /l/ similar to those of

English speakers, but they produced a wide range of F3 frequencies for /r/ compared to English speakers. Standard errors from the left side of the figure are 29.32, 46.07, 74.76, and 33.61 (Hz). The results confirmed that Japanese speakers were similar to English speakers in producing F3 frequencies for /l/, but they differed from English speakers in producing F3 frequencies for /r/ (i.e., they were inaccurate in producing the frequencies for /r/). . . . . 111

Figure 3.15 Boxplots for closure durations of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers generally produced shorter closure durations for /r/ and longer closure durations for /l/. Japanese speakers produced short closure durations for /r/ and long closure durations for /l/ with a wide range of variance. Standard errors from the left side of the figure are 4.72, 3.70, 5.28, and 3.17 (ms). The results demonstrated, despite such a wide range of variance, Japanese speakers were similar to English speakers in producing closure durations for /r/ and /l/. . . . . 112

Figure 3.16 Boxplots for transition durations of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers made clear distinctions between /r/ and /l/ (i.e., they produced longer transition durations for /r/ and shorter transition durations for /l/). Japanese speakers produced shorter transition durations for /l/ similar to those of English speakers. However, they produced slightly shorter transition durations for /r/ compared to English speakers. Standard errors from the left side of the figure are 2.72, 2.24, 1.66, and 1.07 (ms). The results demonstrated that Japanese speakers were similar to English speakers in producing transition durations for /l/. However, Japanese speakers produced shorter transition durations for /r/ than English speakers did. . . . . 113

Figure 3.17 The degree of English /r/-/l/ F3 contrast in words and passages. /r/ and /l/ productions of Japanese speakers were measured in terms of F3 frequencies and normalized by subtracting median F3 frequencies which were measured using the passage recordings. The F3 contrasts were calculated by subtracting normalized F3 frequencies for /r/ from normalized F3 frequencies for /l/. Bigger positive contrast indicates clearer F3 contrast between /r/ and /l/. Standard errors from the left side of the figure are 52.24, 57.27, and 63.55 (Hz). The results demonstrated

that Japanese speakers made similar /r/-/l/ F3 contrasts in words and sentences.

However, they made less clear contrasts in passages..... 115

Figure 3.18 F3 frequencies of English /r/ and /l/ productions by Japanese speakers in three different materials (i.e., words, sentences, passages). F3 frequencies of /r/ and /l/ productions were normalized by subtracting median F3 frequencies which were measured with the passage recordings. Japanese speakers produced similar F3 frequencies for /r/ in all of the materials. However, Japanese speakers had some difficulties in producing /l/. They produced similar F3 frequencies for /l/ in words and sentences, but they had hard time to produce /l/ in passages. Standard errors from the left side of the figure are 46.07, 33.61, 50.68, 38.69, 56.71, and 45.76 (Hz). The results confirmed that /r/ productions of Japanese speakers were similar among the three materials. However, /l/ productions in passages were different from the ones in words and sentences..... 116

Figure 4.1 A spectrogram of *lick* produced by a female Japanese speaker. This is her production in the 7<sup>th</sup> training session. The blue area indicates closure duration (i.e., 51 ms). The yellow area indicates transition duration (i.e., 26 ms). The average F3 frequency in the closure duration is 3052 Hz..... 138

Figure 4.2 A spectrogram of signal-processed *lick* by the female Japanese speaker from Figure 4.1. The blue area indicates enhanced closure duration. The original closure duration (i.e., 51 ms) was enhanced to 82 ms. The yellow area indicates transition duration. The original transition duration (i.e., 26 ms) was enhanced to 20 ms. The original average F3 frequency (i.e., 3052 Hz) was enhanced to 3672 Hz..... 139

Figure 4.3 A spectrogram of *rick* produced by a female Japanese speaker. This is her production in the 6<sup>th</sup> training session. The blue area indicates closure duration (i.e., 130 ms). The yellow area indicates transition duration (i.e., 88 ms). The average F3 frequency in the closure duration is 2056 Hz..... 140

Figure 4.4 A spectrogram of signal-processed *rick* by the female Japanese speaker from Figure 4.1. The blue area indicates enhanced closure duration. The original closure duration (i.e., 130 ms) was enhanced to 30 ms. The yellow area indicates transition duration. The original transition duration (i.e., 88 ms) was enhanced to

66 ms. The original average F3 frequency (i.e., 2056 Hz) was enhanced to 1500 Hz..... 141

Figure 4.5 Boxplots of English /r/ and /l/ production intelligibility for the pre-test and post-test. Boxplots display the medians and quartile ranges of scores, and outliers marked by circles. Five native speakers of British English listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese before and after the pronunciation training. They identified consonant categories by choosing /r/, /l/, /w/, or /d/. Higher proportion correct indicates that English speakers correctly identified /r/ and /l/ productions of L1 Japanese speakers. The intelligibilities of /l/ productions were similarly high before and after the training; Japanese speakers produced identifiable /l/ throughout this study. However, the intelligibilities of /r/ productions were different before and after the training. The pre-test intelligibility of /r/ productions widely varied among Japanese speakers, but the post-test intelligibility of /r/ productions was high and hardly varied; most of the Japanese speakers were able to produce identifiable /r/ after the training. Standard errors from the left side of the figure are 0.06, 0.01, 0.01, and 0.01. The results demonstrated that the intelligibility of /r/ and /l/ productions both improved after the pronunciation training..... 150

Figure 4.6 Boxplots of the goodness rating scores for pre-/post-tests /r/ and /l/ productions of Japanese speakers. British English speakers gave ratings to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese with a scale from 1 (poor) to 7 (good). The intelligibility scores for /l/ were high and similar before and after the pronunciation training. However, the intelligibility scores for /r/ were different before and after the training. The intelligibility scores for pre-test /r/ productions substantially varied. However, the intelligibility scores for post-test /r/ productions less varied. Standard errors from the left side of the figure are 0.24, 0.09, 0.11 and 0.08. The results demonstrated that post-test /r/ and /l/ productions received higher intelligibility scores, suggesting that Japanese speakers improved their /r/ and /l/ productions..... 151

Figure 4.7 Boxplots of F1 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). The data of English speakers came from

Chapter 3. For /r/, Japanese speakers produced similar F1 frequencies before and after the pronunciation training. Standard errors are 44.02, 33.23, and 34.25 (Hz) from the left side. The results demonstrated that Japanese speakers produced higher F1 frequencies for /r/ after the training. The results also demonstrated that such post-test F1 frequencies for /r/ were similar to those of English speakers. For /l/, Japanese speakers produced similar F1 frequencies before and after the training. Standard errors are 41.39, 33.05, and 33.78 (Hz) from the left side. The results demonstrated that Japanese speakers produced higher F1 frequencies after the training. The results also demonstrated that such post-test F1 frequencies for /l/ were similar to those of English speakers..... 153

Figure 4.8 Boxplots of F2 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced high F2 frequencies before the training, but they produced lower F2 frequencies after the training. Standard errors are 27.10, 52.64, and 33.72 (Hz) from the left side. The results confirmed that Japanese speakers lowered F2 frequencies after the training, and that such post-test F2 frequencies were similar to those of English speakers; Japanese speakers improved their /r/ productions in the F2 dimension. For /l/, Japanese speakers produced similar F2 frequencies before and after the training. Standard errors are 40.66, 39.91, and 43.19 (Hz) from the left side. The results confirmed that Japanese speakers produced similar F2 frequencies throughout the experiment, and that the post-test F2 frequencies were similar to those of English speakers..... 154

Figure 4.9 Boxplots of F3 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced high F3 frequencies before the training, but they produced lower F3 frequencies after the training. Standard errors are 29.32, 48.86, and 41.36 (Hz) from the left side. The results confirmed that Japanese speakers lowered F3 frequencies in the post-test., and that such post-test F3 frequencies were similar to those of English speakers; Japanese speakers improved their /r/ productions in the F3 dimension. For /l/, Japanese speakers produced similar F3 frequencies before and after the training. Standard errors are 74.76, 42.68, and 39.01 (Hz) from the left side. The results

confirmed that Japanese speakers produced similar F3 frequencies throughout the experiment, and that the post-test F3 frequencies were similar to those of English speakers..... 155

Figure 4.10 Boxplots of closure duration of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced slightly longer closure durations in the post-test than they did in the pre-test. Standard errors are 4.72, 5.01, and 4.94 (ms) from the left side. The results demonstrated that Japanese speakers produced similar closure durations throughout the experiment, and that such closure durations were similar to those of English speakers. For /l/, Japanese speakers produced longer closure durations in the post-test than they did in the pre-test. Standard errors are 5.28, 3.62, and 4.46 (ms). The results confirmed that Japanese speakers produced longer closure durations after the training, and that such post-test closure durations were similar to those of English speakers..... 156

Figure 4.11 Boxplots of transition duration of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced short transition durations before the training, but they produced long transition durations after the training. Standard errors are 2.72, 2.83, and 2.54 (ms) from the left side. The results confirmed that Japanese speakers produced longer transition durations in the post-test, and that such post-test transition durations were similar to those of English speakers; Japanese speakers improved their /r/ productions in transition duration. For /l/, Japanese speakers produced similar transition durations before and after the training. Standard errors are respectively 1.66, 1.51, and 1.23 (ms) from the left side. The results confirmed that Japanese speakers produced similar transition durations throughout the experiment, and that the post-test transition durations were similar to those of English speakers..... 157

Figure 4.12 Boxplots of F3 distance between English /r/ and /l/ productions (i.e., F3 contrast between /r/ and /l/) in three different materials (i.e., words, sentences, and passages). Normalized F3 frequencies for /r/ were subtracted from F3 frequencies for /l/. Longer positive distance indicates that Japanese speakers made clearer F3

contrasts between /r/ and /l/. In the pre-test, Japanese speakers made similar F3 contrasts in words and sentences, but they made less clear F3 contrasts in passages. In the post-test, Japanese speakers were able to make clearer /r/-/l/ contrasts in words, sentences, and passages. However, they seemed to make less clear F3 contrasts in passages. Standard errors from the left side of the figure are 60.75, 51.91, 70.84, 41.25, 80.50, and 54.48 (Hz). The results demonstrated that, although Japanese speakers had hard time to produce the /r/-/l/ contrasts in passages, Japanese speakers improved making F3 contrasts between /r/ and /l/ in all three materials after the pronunciation training..... 160

Figure 4.13 Boxplots of English /r/-/l/ identification for the pre-test and post-test.

Japanese speakers listened to 60 naturally spoken minimal pairs (i.e., 120 words) including /r/ and /l/ in word-initial positions and identified consonant categories (i.e., /r/ or /l/). Japanese speakers demonstrated similar /r/-/l/ identification accuracies with a wide range of variance in the pre-test and post-test. Standard errors are 0.02 and 0.02 from the left side of the figure. The results confirmed that, although Japanese speakers improved /r/-/l/ production, they did not improve /r/-/l/ identification at all..... 163

Figure 4.14 F2 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. The data of English speakers from Chapter 3 are displayed, but there were no cross-language data analyses because the main purpose of the analyses was to examine whether Japanese speakers change their discrimination sensitivity. Japanese speakers demonstrated similar discrimination sensitivity patterns in the pre-test and post-test; they had high discrimination sensitivity at the /r/-/l/ boundary, and low discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0.20, 0.25, and 0.28 (d') for the pre-test, and 0.21, 0.28, and 0.33 (d') for the post-test. The results demonstrated that Japanese speakers did not change their F2 discrimination sensitivity after the pronunciation training, and that they had similar high discrimination sensitivities at the /r/-/l/ boundary and within the /l/ category..... 165

Figure 4.15 F3 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers demonstrated similar discrimination sensitivity patterns in the pre-test and post-test; they had slightly high discrimination sensitivity at the /r/-/l/ boundary, and slightly low discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0.21, 0.30, and 0.51 (d') for the pre-test, and 0.31, 0.36, and 0.62 (d') for the post-test. The results demonstrated that Japanese speakers did not change their F3 discrimination sensitivity after the pronunciation training, and that they had similar high discrimination sensitivity at the /r/-/l/ boundary and within the /l/ category..... 166

Figure 4.16 Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). The data of English speakers from Chapter 3 is presented, but there was no comparison between English and Japanese speakers in this analysis because the focus of the analysis was to examine whether Japanese speakers change their best exemplars after the pronunciation training. For /r/, Japanese speakers selected similar F1 frequencies in the pre-test and post-test. Standard errors are 24.38 (pre-test) and 20.58 (post-test) Hz. The results confirmed that Japanese speakers did not change F1 frequencies of best exemplars for /r/ after the pronunciation training. For /l/, Japanese speakers chose similar F1 frequencies in the pre-test and post-test. Standard errors are 19.95 (pre-test) and 25.49 (post-test) Hz. The results demonstrated that Japanese speakers did not change F1 frequencies of best exemplars for /l/ after the training..... 168

Figure 4.17 Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers selected similar F2 frequencies in both pre-test and post-test. Standard errors are 50.25(pre-test) and 45.63 (post-test) Hz. For /l/, likewise, they selected similar F2 frequencies for /l/ in both pre-test, post-test. Standard errors are 67.04 (pre-test) and 77.6 (post-test) Hz. The results demonstrated that Japanese speakers did not change F2 frequencies of best exemplars for /r/ and /l/ after the training..... 169



Figure 4.18 Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose similar F3 frequencies with a wide range of variance in the pre-test and less variance in the post-test. Standard errors are 111.82 (pre-test) and 52.60 (post-test) Hz. For /l/, Japanese speakers chose similar F3 frequencies with a wide range of variance in the pre-test and post-test. Standard errors are 109.12 (pre-test) and 124.15 (post-test) Hz. The results demonstrated that Japanese speakers did not change F3 frequencies of best exemplars for /r/ and /l/ after the training..... 170

Figure 4.19 Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose short closure durations with a wide range of variance in the pre-test and less variance in the post-test. Standard errors are 11.12 (pre-test) and 5.24 (post-test) ms. For /l/, Japanese speakers chose similar closure durations with similar variance in the pre-test and post-test. Standard errors are 8.55 (pre-test) and 9.41 (post-test) ms. The results demonstrated that Japanese speakers did not change closure durations of best exemplars for /r/ and /l/ after the training..... 171

Figure 4.20 Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose similar transition durations with a wide range of variance in the pre-test and post-test. Standard errors are 10.08 (pre-test) and 10.33 (post-test) ms. For /l/, they chose similar short transition durations with little variance in the pre-test and post-test. Standard errors are 3.56 (pre-test) and 1.70 (post-test) ms. The results demonstrated that Japanese speakers did not change transition durations of best exemplars for /r/ and /l/ after the training..... 172

## Acknowledgement

I would like to thank my main supervisor, Dr. Paul Iverson, and second supervisor, Dr. Valerie Hazan for their great support through my doctoral training at University College London (UCL). I came to UCL after my master degree in English as a Second Language from University of Hawai'i at Mānoa (UHM) in order to learn second language acquisition from the perspectives of speech science. Dr. Iverson taught me issues in speech perception and production along with great programming skills so that I can be trained to be able to look at second language learning with speech science approaches. If I would not have encountered his inspiring studies, I would not be able to seek further education at UCL. I would also like to thank all of anonymous participants in this dissertation. It would not be possible to have this dissertation without their generous support.

I would like to dedicate this dissertation to my parents who has had great understanding to my intellectual exploration and has made this doctoral training possible. I wish I could do something more to express my appreciation. I would like to dedicate this work to my father-side grandfather, who greatly lived for 95 years without any physical problems and thought of me all the time, mother-side grandmother, who loved plants and greatly fought against physical problems, one of my uncles, who always treated me as his little brother, Dr. Craig Chaudron, who opened the door for my study at UHM, and Ok Kyoonyo, who went through hard work together at UHM. This dissertation also goes to Mark Messer, Adam Pang Qu, Daniele Da Ponte, and Jason Shaw who have always given me great support and encouraged me along the way.

*Chi cerca trova*

*E lawe i ke a`o a malama, a e `oi mau ka na`auao*

## Chapter 1: General Introduction

### *1.1 How do individuals learn the phonemes of their language?*

As far as we know, approximately 600 consonants and 200 vowels exist in the world's languages (Ladefoged, 2005, xiii). Each language contains some of the vowels and consonants in its sound system. For example, English has 24 consonants and thirteen vowels (Tsujimura, 1996). Japanese has 23 consonants and five vowels (Tsujimura, 1996). Human beings as infants initially have to learn such phonemes existing in a target language before they identify units of phonemes (i.e., words) from strings of sounds in speech. Early studies of infant speech perception have demonstrated that infants possess the ability to discriminate phonemes. Specifically, infants have high discrimination sensitivity at phonetic category boundaries rather than within phonetic category. Eimas et al. (1971) demonstrated that 1 and 4 months old American English infants had higher sensitivity at the /p/-/b/ boundary than did they within /p/ and /b/ categories. Eimas (1975) demonstrated that American English infants had higher sensitivity at the /r/-/l/ category boundary than did they within /r/ and /l/ categories. Infants are sensitive to phonetic boundaries not only in their target language but also in other languages (e.g., Aslin et al., 1981; Lasky et al., 1975, Streeter, 1976; Trehub, 1976). Streeter (1976) demonstrated that 2-months old Kikuyu infants discriminated prevoiced and voiced /b/, and voiced /b/ and voiceless /p/, despite the fact that Kikuyu has prevoiced /b/ only. Lasky et al. (1975) demonstrated that 4-6.5 month old Spanish infants demonstrated /p/-/b/ phonetic category boundary of English speaking adults rather than that of Spanish speaking adults.

These findings of categorical perception by infants led to a hypothesis that human beings are born with innate neural specification for all possible phonemes (i.e., a universal phonetic feature detector) which allows infants to choose a subset of phonemes for their native languages (Eimas and Corbit, 1973). It was also hypothesized that phonetic feature detectors which are stimulated by language input remain, but other phonetic detectors which are not used are lost (Eimas, 1975).

Studies on perception of nonhuman animals, however, falsified the phonetic feature detector hypothesis. For example, animals are able to identify speech sounds. Kuhl and Miller (1975) demonstrated that, after discrimination training, chinchillas identified /t/ and /d/ in the same manner as English speakers did, and that chinchillas showed a human-like category-boundary between /t/ and /d/. Kuhl and Miller (1978) similarly demonstrated that chinchillas identified /p/ and /b/, and /k/ and /g/ as English speakers did, and that they had a category-boundary between the voiced and voiceless consonants. Animals can also discriminate phonetic segments human-like. Kuhl and Padden (1983) demonstrated that macaques discriminate /bæ/ and /dæ/, and /dæ/ and /gæ/ as English speakers did; they had a category boundary between /bæ/ and /dæ/, and /dæ/ and /gæ/. Kuhl (1981) demonstrated that chinchillas can be trained to discriminate /ta/ and /da/. Kluender et al. (1987) trained Japanese quails to discriminate naturally spoken /dVs/ from /bVs/ and /gVs/, and demonstrated that the quail were able to categorize these three consonants. Dooling et al. (1995) demonstrated that budgerigars and zebra finches had a phonetic boundary between English /r/ and /l/. Animals are able to distinguish not only phonetic segments but also continuous speech. Ramus et al. (2000) examined whether human newborns and cotton-top monkeys discriminate unfamiliar languages (i.e.,

Dutch and Japanese) and revealed that the monkeys differentiated Dutch and Japanese sentences.

This empirical evidence from nonhuman categorical perception led to an alternative account of infant speech perception instead of the phonetic feature detector hypothesis. That is, infants do not innately have speech-specific mechanism, but they have more general auditory mechanisms to perceive and learn speech sounds. Using such mechanisms, infants initially engage in language-general perception in the first half of their first year and shift to language-specific perception in the second half of their first year; they can initially perceive realizations of phonemes in any language up to age 8 months, but they tune into their native language around age 8-10 month and decrease sensitivity to non-native languages (e.g., Best and McRoberts, 2003; Best et al., 1995; Bosch and Sebastián-Gallés, 2003, 2005; Burns et al., 2003, Cheour et al., 1998; Pegg and Werker, 1997; Polka and Werker, 1994; Riviera-Gaxiola et al., 2005; Sundara et al., 2006; Tsao et al., 2006; Werker and Lalonde, 1988; Werker and Tees, 1984; Werker et al., 1981). For example, Werker and Lalonde (1988) demonstrated that 6-8 months old English-speaking infants discriminated both the bilabial and medial (dental or alveolar) stops which are common to English and Hindi speakers, and Hindi dental and retroflex consonants. However, 11-13 month old English-speaking infants did not discriminate the Hindi consonants. Burns et al. (2003) demonstrated that 6-8 month old English monolingual and English-French bilingual infants exhibited similar category boundaries between [ba] (French /ba/) and [pa] (French /pa/ and English /ba/). However, 10-12 month old English monolingual infants shifted the category boundary between [pa] (French /pa/ and English /ba) and [p<sup>h</sup>a] (English /pa/) while 10-12 month old bilingual

infants exhibited a similar category boundary to 6-8 month old bilingual infants. Bosch and Sebastián-Gallés (2003) demonstrated that 4 month old Catalan monolingual infants and Spanish monolingual infants discriminated the two mid-front vowels of Catalan (i.e., /e/ and /ɛ/). While 8 month old Catalan monolingual infants maintained high discrimination sensitivity, 8 month old Spanish monolingual infants decreased their discrimination sensitivity to the contrast of Catalan vowels.

Current first-language (L1) perception models (e.g., Best, 1994, 1995; Best and McRoberts, 2003; Kuhl, 2000, 2004; Kuhl et al., 2008; Werker and Curtin, 2005) have hypothesized that such a shift from language-general perception to language-specific perception (i.e., declining sensitivity to non-native languages and gaining sensitivity to native languages) stems from L1 learning experience. For example, Native Language Magnet extended (NLM-e; Kuhl et al., 2008) claims that infants initially have abilities to distinguish all the sounds of human speech which derive from general auditory processing mechanisms. NLM-e claims that distributional patterns of phonetic units and infant-directed speech promote phonetic learning leading to language-specific perception. Infants possess sensitivities to the distributional properties of linguistic input and detect distributional frequencies of phonetic units in ambient speech (e.g., Kuhl et al., 1992). Acoustically exaggerated infant-directed speech helps infants to easily detect the distributions of phonetic units because such speech expands acoustic space among the phonetic units (e.g., Liu et al., 2003). NLM-e claims that the more infants receive linguistic inputs, the more they make a neural commitment to their native language (e.g., Kuhl, 2004) leading to a distortion in their perceptual map (i.e., perceptual warping); infants decrease their perceptual sensitivity near category modes and increase perceptual



ability near category boundaries. The most activated representations (i.e., prototypes) emerge from accumulated language experience, and they eventually function as *perceptual magnets* for other members of the category. This process increases the perceived similarity among the exemplars of the category. This perceptual distortion process (i.e., perceptual magnet effect) leads to facilitation in native and reduction in foreign language ability.

Processing Rich Information from Multidimensional Interactive Representations (PRIMIR; Werker and Curtin, 2005) suggests that infants shift from language-general perception to language-specific perception by developing multidimensional interactive representations. PRIMIR hypothesizes that infants initially engage in language learning using general learning mechanisms which handle statistical analyses of the language input (i.e., finding regularities in the input) along with some filters which direct attention to a small amount of information (e.g., preference for speech, infant-directed speech, and specific language-learning task). The learning mechanisms and the filters interact together so that some aspects of the rich information in speech signal (e.g., acoustic and gestural information) perceptually become more salient and easier to pick up. This process helps infants to establish interactive representations (i.e., General Perceptual plane, Word Form plane, and Phonemic plane). Infants initially generate General Perceptual plane which includes phonetic information. Clusters of the information (i.e., clusters of exemplars) found through statistical analyses turn into language-specific categories, and such categories help the formation of Word Form plane. Infants, then, extract sequences forming phonetic units in speech, and they have to match such combinations of phonetic units without meaning and object knowledge (i.e., concepts).

PRIMIR claims that infants repeatedly hear a word spoken by different speakers in different contexts and see different examples of the object category in different contexts. Infants statistically analyze which phonetic units match with which referents and eventually map the units and referents without errors. Once infants enhance the size of vocabulary, PRIMIR claims that the Phonemic plane emerges. Phonemic categories including information of phonetic variations become robust as infants expand their size of vocabulary and learn to read.

Perceptual Assimilation Model (PAM; Best, 1994, 1995; Best and McRoberts, 2003) explains the shift from language-general perception to language-specific perception considering articulatory similarity between native and non-native segments. The articulatory organ hypothesis in PAM (Best and McRoberts, 2003) predicts that sensitivity to non-native discrimination declines when phonetic contrasts include the same articulatory organs; if the degree of articulatory similarity is high between native and non-native phonetic segments sharing the same articulatory organs, older infants (e.g., 10-12 months of age) have difficulties in discriminating non-native phonetic contrasts. PAM hypothesizes that the degree of such difficulties depends on how listeners assimilate the contrasting segments into their native segments. If each non-native segment assimilates to a different native segment, PAM predicts that listeners demonstrate excellent discrimination accuracy (i.e., Two-Category assimilation, SC type). If both non-native segments assimilate to the same native segment, PAM predicts that listeners demonstrate poor discrimination accuracy (i.e., Single-Category assimilation, SC type). If both non-native segments assimilate to a same native segment with different degree of goodness to the native segment, PAM predicts that listeners demonstrate

moderate discrimination accuracy (i.e., Category-Goodness difference, CG type). If both non-native segments stay in phonetic space, but if they do not assimilate to any native segments, PAM predicts that listeners demonstrate a wide range of discrimination accuracy (i.e., Both Uncategorizable, UU type). If one of the non-native segments assimilates to a native segment and the other segments stays in phonetic space but does not assimilate to any native segments, PAM predicts that listeners demonstrate excellent discrimination accuracy (i.e. Uncategorized and Categorized, UC type). If both non-native segments stays outside of phonetic space (i.e., being perceived as non-speech sounds), PAM predicts that listeners demonstrate good discrimination accuracy.

L1 learning experience leads to facilitation in native language ability. However, such experience also leads to decline but not loss in non-native language ability (i.e., L1 interference). On the one hand, adults can certainly discriminate some contrasts of non-native phonemes, suggesting that adults maintain the ability to perceive non-native phonemes (e.g., Best et al., 1988; Polka, 1991, 1995; Tees and Werker, 1984, Werker and Tees, 1984). For example, Best et al. (1988) demonstrated that English speakers can discriminate Zulu clicks. Polka (1995) demonstrated that American English speakers were good at discriminating German /ʊ/-/ʏ/ vowel contrast. Tees and Werker (1984) demonstrated that adult English speakers who had little exposure to Hindi discriminated unvoiced aspirated dental and breathy-voiced dental consonants (i.e., /t<sup>h</sup>a/-/d<sup>h</sup>a/) and unvoiced unaspirated retroflex and dental consonants (i.e., /ʈa/-/ta/). On the other hand, adults certainly face difficulties identifying or differentiating some contrasts of L2 segments, suggesting that adults become less sensitive to perceive non-native phonemes.

For example, adult French speakers generally have difficulties to discriminate English /d/ and /ð/ (Polka et al., 2001). Adult Sinhala speakers have difficulties to identify English /v/ and /w/ (Iverson et al., 2008). Adult Norwegian speakers have some difficulties to identify English fricative consonants (e.g., /θ/, /ð/, /z/, and /ʒ/; van Dommelen, 2008). Adult Greek speakers have some difficulties discriminating some contrasts of English vowels (e.g., /æ/-/ʌ/ and /ʌ/-/a:/; Lengeris and Hazan, 2007).

Some other better-documented cases further suggest that L1 language experience interferes with L2 segments (and suprasegmental) learning. Adult Spanish speakers have difficulties in differentiating English /i/ and /ɪ/ (e.g., Escudero and Boersma, 2004; Flege, 1991; Fox et al., 1995; Morrison, 2002). They also have difficulties distinguishing Catalan vowels (e.g., Bosch et al., 2000; Pallier et al., 1997; Pallier et al., 2001, Sebastián-Gallés and Soto-Faraco, 1999). Pallier et al. (1997) examined whether Catalan-Spanish bilinguals who had exposure to Catalan from their birth and who had exposure to Spanish first and started learning Catalan at the latest at six years of age similarly identify and discriminate Catalan /e/ and /ɛ/. The results demonstrated that the late Catalan learners did not categorize the vowels and that they did not show the category boundary between the vowels. Sebastián-Gallés and Soto-Faraco (1999) similarly demonstrated that late Catalan learners have difficulties differentiating Catalan /o/ and /ɔ/. American English speakers have difficulties learning French vowels. They have difficulties to discriminate the contrast between back and front high rounded vowels (i.e., /u/-/y/; Gottfried, 1984; Levy and Strange, 2008). They inconsistently identified French /o/ and /u/ (Gottfried and Beddor, 1988). American English speakers have difficulties learning

the contrast of German /i/ and /y/ (Polka, 1995). They have difficulties to learn tonal languages as well. For example, Wang et al (1999) demonstrated that American English speakers are poor at identifying Chinese tones. Americans particularly confused Tone 2 and Tone 3. Wayland and Guion (2003) similarly demonstrated that American English speakers are poor at discriminating Thai middle and low tone contrasts.

## ***1.2 Perception of English /r-/l/ by adult Japanese speakers***

This dissertation further examines the perception and production of L2 learners. Specifically, it examines perception and production of English /r/ and /l/ by adult Japanese speakers. Previous studies have clearly demonstrated that Japanese speakers poorly identify and discriminate the consonants; Goto (1971) revealed that Japanese speakers poorly identify /r/ and /l/, and Miyawaki et al. (1975) revealed that Japanese speakers had poor discrimination sensitivity at the /r-/l/ category boundary. However, subsequent studies have demonstrated that Japanese speakers can identify and discriminate /r/ and /l/ to some extent (e.g., Mochizuki, 1981; Sheldon and Strange, 1982). When Japanese speakers gain living experience in English-speaking countries, they can overcome the problems. For example, MacKain et al. (1981) demonstrated that Japanese speakers having conversation training had native-like identification and discrimination sensitivity. Yamada (1995) demonstrated that some Japanese speakers who had lived in the United States for more than one year showed native-like /r-/l/ identification accuracy. Particularly, Japanese speakers who have earlier onset age to the environment and who have longer duration of stay demonstrated better identification accuracy. Yamada and

Tohkura (1992) similarly demonstrated that Japanese speakers whose onset age is earlier had better identification accuracy.

It is not only early or long exposure to English-speaking environments but also perceptual training that leads to better /r-/l/ identification accuracy. Specifically, Japanese speakers can improve /r-/l/ identification accuracy by going through identification tasks with high variability in stimuli (e.g., large number of minimal-pair words, multiple talkers; e.g., Iverson et al., 2005; Lively et al., 1993; Logan et al., 1991; Uther et al., 2008), and keep such improved perceptual performance for six months (Lively et al., 1994). Also audio-visual training can help Japanese speakers improve /r-/l/ identification accuracy (e.g., Hazan et al., 2005; Massaro and Light, 2003). For example, Hazan et al. (2005) had three groups of Japanese speakers take ten sessions of perceptual training using auditory stimuli, natural audiovisual stimuli, and audiovisual stimuli with a synthetic face. The results demonstrated that Japanese speakers who had training with audiovisual stimuli improved /r-/l/ identification accuracy to the same degrees as Japanese speakers who had training with auditory stimuli. Despite the fact that these previous studies in the literature demonstrate that Japanese speakers overcome their /r-/l/ problems to some extent, their improvement is not enough for them to have native-like perceptual performance.

Current L2 speech perception models attribute such learning difficulties to the phonetic similarity between native and non-native phonemes. For example, Best's Perceptual Assimilation Model (PAM) hypothesized that non-native phonemes are perceived based upon their articulatory similarities and dissimilarities to L2 listener's closest native phoneme categories (Best, 1995). That is, non-native phonemes are

perceived as either good or bad exemplars of native phonemes. Likewise, Flege's Speech Learning Model (SLM, Flege, 1995, 2003) hypothesized that the relationship between L1 and L2 phonemes affects L2 perceptual learning. L1 and L2 phonetic categories are thought to exist in the same phonological space. The L1 phonetic categories become robust as human beings grow up, and the categories become strong attractors of L2 categories. If a new L2 category has larger phonetic dissimilarity to the closest L1 category (or categories), L2 learners are likely to establish the L2 category. If a new L2 category is more similar to an L1 category (or categories), L2 learners are likely to have difficulties establishing the category. SLM hypothesizes that, if L2 learners fail to establish a L2 category, they rather create a phonetic category which includes properties of the closest L1 category and the L2 category (i.e., category assimilation). However, SLM hypothesizes that speech learning mechanisms remain intact in adulthood, and that L2 learners eventually keep all L1 and L2 categories in the same phonological space (i.e., category dissimilation).

Previous work has suggested that L1-L2 similarity relationships cause the Japanese speaker's problems with English /r/-/l/ identification. Japanese has a flap which is generally described as an apico-alveolar tap (Vance, 1987). Some scientists claimed that this phoneme is the cause of the identification problems. For example, Best and Strange (1992) claimed that both English /r/ and /l/ are perceived as poor exemplars of either Japanese /r/ or /w/. Supporting this view, Guion et al. (2000) demonstrated that Japanese speakers identified the English consonants as exemplars of Japanese /r/ with

moderate goodness rating scores. Aoyama et al. (2004) suggested, instead, that /l/ is more similar to Japanese /r/.

Although this category assimilation process has been thought to interfere with learning English /r/ and /l/, such a process may not be a negative learning factor; it can be rather a positive learning factor. Iverson et al. (2005) demonstrated that Japanese speakers tended to identify more stimuli with shorter closure and transition duration as /l/ after cue-manipulated phonetic training, suggesting that the short closure and transition duration of Japanese /r/ might have caused the assimilation between /l/ and Japanese /r/. If both English /r/ and /l/ equally assimilate to Japanese /r/, the category assimilation process is detrimental to learning the English consonants. However, if only /l/ assimilates to Japanese /r/, category assimilation could be used as a positive strategy in order to identify English /r/ and /l/ better; assimilating /l/ to Japanese /r/ and focusing on recognizing English /r/ could be easier than learning two non-native consonant categories at the same time.

### **1.3 Dissertation overview**

This dissertation investigated the perception and production of English /r-/l/ by adult Japanese speakers. Most of the previous studies generally used group analyses (i.e., eliminating individual differences), and demonstrated /r-/l/ learning difficulties. One disadvantage of such an approach is that it is difficult to examine what kind of causes



underlies difficulties in the learning. Some Japanese speakers who are good at /r-/l/ identification, for example, can be different in other perceptual behavior and/or production from some other Japanese speakers who are poor at identification; it is possible that some perceptual and/or production behavior can be predictive of /r-/l/ identification accuracy. Study 1 in Chapter 2 took an individual differences approach, rather than the traditional group approach, and examined whether category assimilation process between L1 and L2 categories is predictive of English /r-/l/ identification accuracy. Study 2 in Chapter 3 similarly took an individual differences approach and examined whether perception and production of English /r-/l/ are related. One of the motivations was that, despite the fact that English /r-/l/ learning difficulties of Japanese listeners has been in the literature nearly for 40 years, it has not been clear whether perceptual and production difficulties of Japanese speakers are related. Study 3 in Chapter 4 examined whether one-to-one pronunciation training promotes accurate /r-/l/ production. Previous studies (i.e., Bradlow et al., 1997; Bradlow et al., 1999) demonstrated that perceptual training can lead to improvement in English /r-/l/ perception and production. But it has been unknown whether pronunciation training brings similar training effects for Japanese speakers learning English. The other motivation was that, since I had training in English as a Second Language (ESL), I wanted to examine how explicit teaching combining ESL and speech science approaches changes perception and production of second language speakers.

## **Chapter 2: English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy**

### **2.1 Introduction**

Study 1 in this chapter investigated the causal relationship between the English /r/-/l/ recognition and the category assimilation process between Japanese /r/ and the English consonants. One problem with examining the causal relationship is to quantify the degree of assimilation. One possible measure is to use a goodness rating scale, but such a measure does not necessarily provide fine-grained quantitative data. For example, if a Japanese listener claims that /l/ is more similar to Japanese /r/ by giving a higher goodness score than does another Japanese listener, it is not clear whether the gap between their rating scores reflects the real difference in the degree of assimilation or a superficial difference in rating scores caused by individual rating bias. Another way to measure assimilation is acoustic analysis of speech production. For example, Lotto et al. (2004) have measured English /r/, /l/, and Japanese /r/ and found that the range of F2 and F3 formants for English /r/ and /l/ was partially overlapping with that for Japanese /r/. However, such a measure has three disadvantages. One is that English /r/, /l/, and Japanese /r/ cannot be measured and compared in all acoustic dimensions; /l/ and Japanese /r/ can use an additional dimension (e.g., a burst) which English /r/ does not use. Another disadvantage is that between-talker differences (e.g., vocal tract length and

speaking rate) introduce more variability that is hard to control. The other is that, since speech production poorly correlates with speech perception (e.g., Bradlow et al., 1997), it is not clear whether acoustic analysis of speech production provides fine-grained quantitative data to examine perceptual assimilation.

The present study instead took two new approaches in order to measure category assimilation. The first approach was to use a bilingual identification task including English /r/, /l/, and Japanese /r/. The typical identification task has been a 2-way forced-choice identification task; choosing either English /r/ or /l/. Since this task excludes Japanese /r/, it cannot examine how often Japanese speakers actually assimilate the English consonants to Japanese /r/. Therefore, the present study used the bilingual identification task for examining category assimilation process and the traditional 2-way identification task for accessing the English /r/-/l/ recognition. The second approach, instead of the acoustic measure, was to use a perceptual mapping approach (e.g., Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson et al., 2006), finding best exemplars of English /r/, /l/, and Japanese /r/. This approach can be useful, in the sense that it includes only one English-Japanese bilingual speaker (i.e., talker variability is excluded), and that it puts the consonants in the identical 5 acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration).

The aim of the present study was to examine whether category assimilation is the cause of English /r/-/l/ identification by native speakers of Japanese. If category assimilation is detrimental to learning English /r/ and /l/, Japanese speakers showing more category assimilation will have poorer English /r/-/l/ identification. In order to

examine the causal relationship, we tested Japanese speakers with a wide range of English experience and focused on individual differences. Experiment 1.1 was similar to previous studies, in that Japanese speakers identified English /r/ and /l/ with naturally spoken words by English speakers (i.e., the traditional 2-way identification task). In addition, the L2 speakers identified the English consonants and Japanese /r/ with naturally spoken monosyllables (e.g., /ra/, /la/, and /ra/) by English-Japanese bilingual speakers (i.e., the bilingual identification task). Experiment 1.2 assessed the degree of accent in English /r/ and /l/ with a goodness rating task. Japanese speakers made recordings with one sentence in English and native speakers of British English rated the goodness of English /r/, /l/ accent, and the degree of /r/-/l/ contrast. Experiment 1.3 used the perceptual mapping approach, finding best exemplars of English /r/, /l/, and Japanese /r/ in a 5-dimensional acoustic space including F1, F2, F3, closure duration, and transition duration. Japanese speakers listened to synthesized syllables (i.e., /ra/, /la/, and /ra/) embedded in natural carrier sentences in English and Japanese and found their best exemplars of English /r/, /l/, and Japanese /r/. Native speakers of British English also found their best exemplars of English /r/ and /l/ in order to provide normative data.

## **2.2 Experiment 1.1: Phoneme identification**

### **2.2.1 Method**

#### **Subjects**

Thirty-nine adult native speakers of Japanese were tested in London; three were removed from the data due to computer problems, leaving a total of 36 participants (see Appendixes A and B). Their age ranged between 19 and 48 years (median = 24.5 years). The age at which they started learning English ranged between 6 and 13 (median = 13 years). They had received English instruction for 7-25 years (median = 9 years). All participants were brought up in monolingual environments in Japan. They have lived in English-speaking countries between the range of 2 weeks and 13 years (median = 3 months). All of the participants reported no hearing problems.

#### **Apparatus and stimuli**

##### ***Apparatus***

Stimulus recordings were made in an anechoic chamber (16-bit depth; 44,100 samples/sec). Perceptual experiments were conducted in a quiet room over headphones.

##### ***Monolingual stimuli***

Monolingual stimuli were initial-position /r/-/l/ minimal-pair words (e.g., *rock* and *lock*) used in the study of Iverson et al. (2005). Four British English speakers (2 male and 2 female) recorded a total of 120 minimal-pair words. In order to avoid list intonation, stimuli were presented singly to the talkers on a computer screen.

### ***Bilingual stimuli***

Bilingual stimuli were recorded in a similar manner. Five English-Japanese bilingual speakers participated in recordings; two were native speakers of British English (one male and one female) who have majored in Japanese in a university in the UK and studied abroad in Japan. The rest of the bilingual speakers (two male and one female) were native speakers of Japanese who had lived in the UK for more than 15 years. Two of them moved to the UK before age 5, and the other speaker immigrated to the country in his 30s. The bilingual speakers recorded 30 CV monosyllabic stimuli with three consonants (i.e., /r/, /l/, and /r/) and five vowels (i.e., /i/, /e/, /a/, /o/, and /u/); they recorded each of 15 syllables twice. Note that one of the speakers could not pronounce the flap followed by /i/. Therefore, he recorded 28 stimuli only. The total number of the bilingual stimuli was 148.

### **2.2.2 Procedure**

#### **Monolingual-stimuli /r/-/l/ identification**

On each trial, the participants saw minimal-pair words on a computer screen and listened to one of the words. They gave their response by clicking on the spelled words on the screen. The participants could not replay the stimuli nor did they receive feedback. Each participant completed a short practice session and six experimental sessions. Each experimental session consisted of 20 stimuli in a random order.

### **Bilingual-stimuli /r/, /l/, and /r/ identification**

The bilingual-stimuli task was similar to the monolingual task. On each trial, the participants saw three spelled consonant categories (i.e., English /r/, /l/, and Japanese /r/) on a computer screen and listened to one of the monosyllabic stimuli. They gave responses by clicking one of the consonant categories on the screen. No replay or feedback was provided. Each participant completed a practice session and engaged in five experimental sessions. In each session, each 15 types of /r/-/l/ initial-position monosyllables were presented twice. Each session consisted of 30 trials except one session which had 28 trials.

### **2.2.3 Results**

Figure 2.1 displays the percent correct score for the monolingual-stimuli /r/-/l/ identification. Japanese speakers' monolingual-stimuli /r/-/l/ identification accuracy widely varied between 50.83 % and 98.33 % (median = 67.08 %).

Table 2.1 displays the percent correct score for the bilingual-stimuli identification among English /r/, /l/, and Japanese /r/. Japanese speakers demonstrated confusion between English /r/ and /l/. They misidentified /r/ as /l/ in 15.56 % of the trials and /l/ as /r/ in 22.39 % of the trials. They also demonstrated strong category assimilation between /l/ and Japanese /r/; they identified /l/ as Japanese /r/ in 19.44 % of the trials and Japanese /r/ as /l/ in 17.13 % of the trials. However, the L2 listeners rarely made category

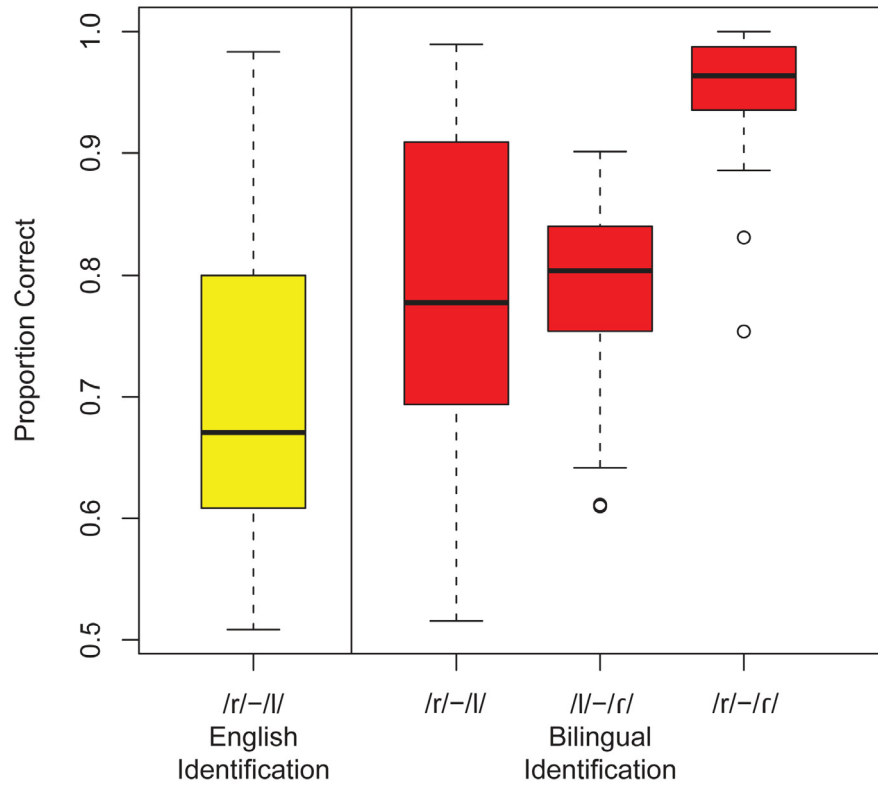
assimilation between /r/ and Japanese /r/; they identified /r/ as Japanese /r/ in only 2.39 % of the trials and Japanese /r/ as /r/ in only 5.50% of the trials.

**Table 2.1 Confusion matrix for bilingual identification. Japanese listeners confused /r/ with /l/ to some extent, but they rarely did so with /r/. They confused /l/ with /r/ and /r/ to the degrees in which they misidentified /r/ as /l/. They similarly confused /r/ with /l/, not with /r/.**

Stimulus	Response		
	/r/	/l/	/r/
/r/	82.05%	15.56%	2.39%
/l/	22.39%	58.17%	19.44%
/r/	5.50%	17.13%	77.37%

In order to examine the causal relationship between the category assimilation process and English /r/-/l/ identification, the results of the monolingual-stimuli and bilingual-stimuli identification were analyzed. First, the 3-way confusion patterns in the bilingual-stimuli identification were reduced to three pair-wise confusion patterns (i.e., percentage correct for identification between /l/ and /r/, /r/ and /r/, and /r/ and /l/). For calculating pair-wise identification accuracy, the number of the times that a Japanese speaker correctly identified two consonants was divided by the total number of stimuli for the consonants. For example, the correct responses for /r/-/l/ pair (i.e., /r/ identified as /r/, and /l/ as identified as /l/) were summed up and divided by the total number of correct and incorrect responses for the pair (i.e., /r/ identified as /l/, and /l/ as identified as /r/), omitting all trials in which the stimulus or response was from the other category (e.g., /r/).

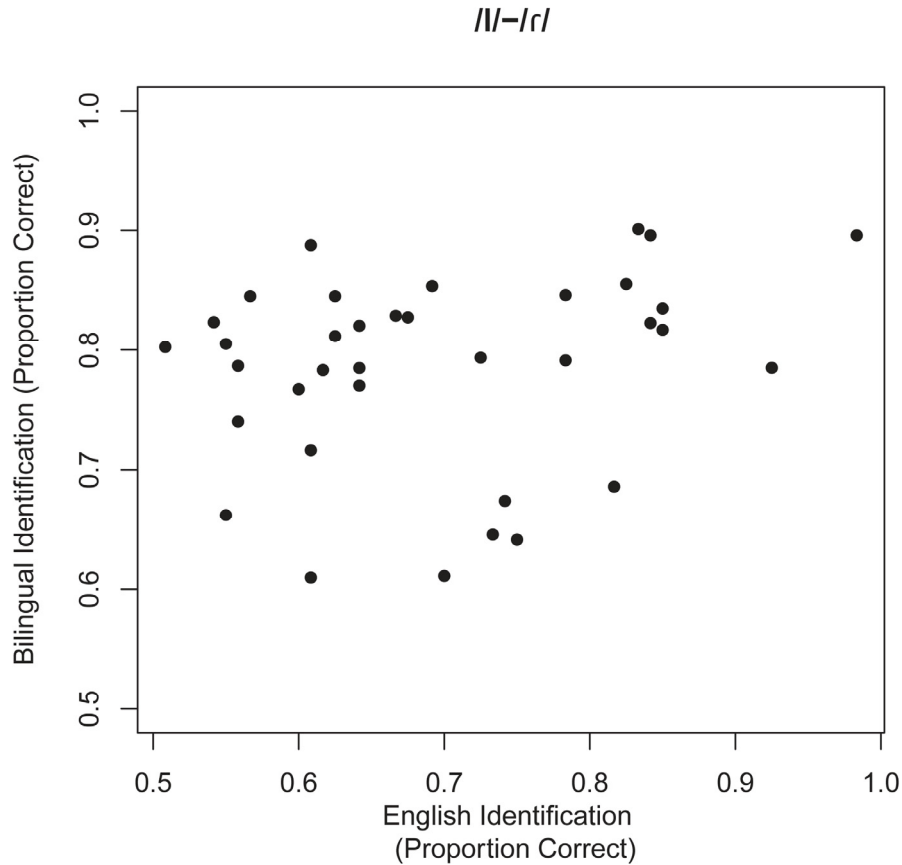




**Figure 2.1** Boxplots of English /r/-/l/ identification (i.e., measuring baseline recognition performance) and pairwise bilingual /r/-/l/-/r/ identification (i.e., indicating the degree of category assimilation). Boxplots display the medians and quartile ranges of scores, with outliers marked by circles. Standard errors from the left side of the figure are 0.02, 0.02, 0.01, and 0.01. The results demonstrated that Japanese speakers confused English /l/ and /r/ to the same extent to which they confused English /r/ and /l/. However, Japanese speakers rarely confused English /r/ and /r/.

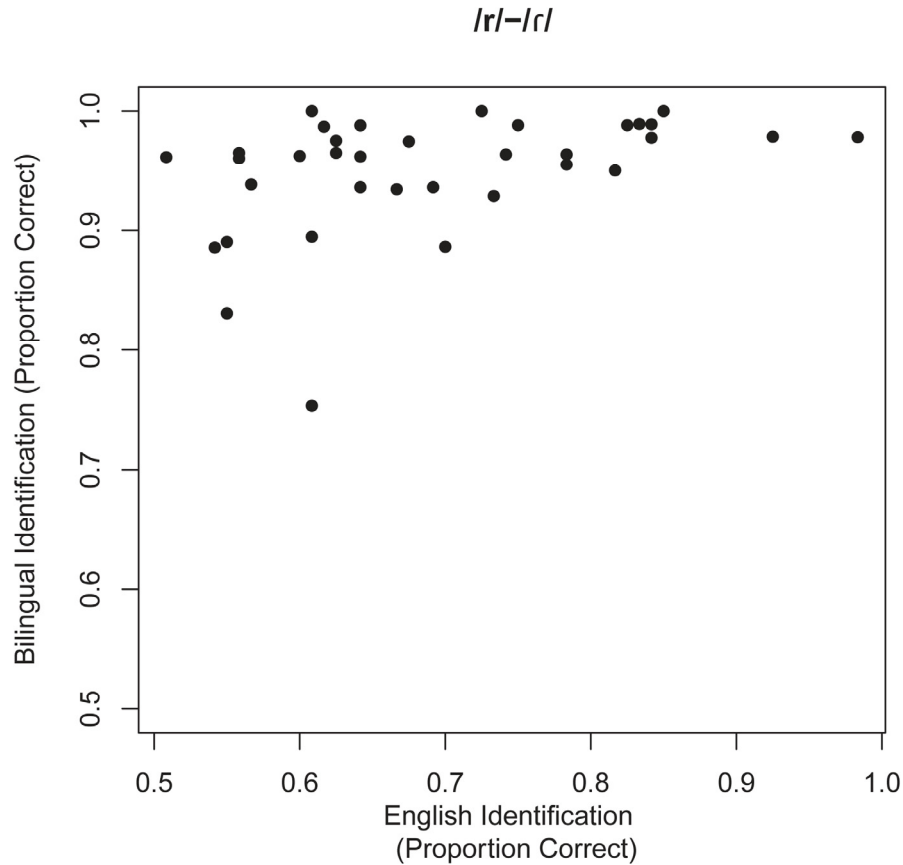
Then, each pair-wise identification accuracy and the monolingual-stimuli /r/-/l/ identification accuracy was analyzed with Pearson correlation analyses. As displayed in Figure 2.1, Japanese speakers demonstrated substantial variability in the bilingual-stimuli /r/-/l/ identification with high median (i.e., 77.70%). A paired-samples t-test revealed that there was significant difference,  $t(35) = -7.37, p < 0.001$ , between the monolingual-stimuli and bilingual-stimuli /r/-/l/ identification, but Pearson correlation analysis revealed that there was a significant correlation,  $r = 0.84, p < 0.001$ , suggesting that the Japanese listeners' English /r/-/l/ identification was consistent in both identification tasks.

Previous studies have suggested that the similarity relationship between English /r/-/l/ and Japanese /r/ causes the difficulties in identifying English /r/ and /l/. PAM hypothesized that native speakers of Japanese assimilate both English consonants to Japanese /r/ (Best and Strange, 1992). However, Aoyama et al. (2004) provided evidence that Japanese children showed greater improvement for English /r/ in perception, suggesting that /l/ is more similar to Japanese /r/. Supporting Aoyama et al (2004), the results demonstrated Japanese speakers' strong category assimilation between Japanese /r/ and /l/ (median = 80.38 %). Statistical comparison of the bilingual-stimuli /r/-/l/ and /r/-/l/ identification revealed that there was no significant difference,  $t(35) = -0.03$ ,  $p > 0.05$ , indicating that the L2 listeners assimilated /l/ to Japanese /r/ to a similar extent to which the listeners confused English /r/ and /l/. In spite of the existence of such a strong category assimilation pattern, Pearson correlation analysis between the monolingual-stimuli /r/-/l/ identification and the bilingual-stimuli /r/-/l/ identification revealed that there was no significant correlation,  $r = 0.22$ ,  $p > 0.05$ , suggesting that the /r/-/l/ category assimilation is not predictive of English /r/-/l/ identification (see Figure 2.2). In other words, there is no significant causal relationship between the English /r/-/l/ identification and /r/-/l/ category assimilation.



**Figure 2.2 Scatterplot of the relationship between English /r-/l/ identification accuracy and /l/-/r/ bilingual identification accuracy (i.e., the degree of category assimilation between /l/ and /r/). Higher /l/-/r/ identification scores indicate less confusion between the categories, and a lower degree of category assimilation. The results demonstrated that the two measures were not significantly correlated, suggesting that the category assimilation between /l/ and /r/ is not predictive of English /r-/l/ identification accuracy.**

While Japanese speakers clearly assimilated /l/ to Japanese /r/, they did not assimilate English /r/ to Japanese /r/ (median = 96.38 %). Statistical comparisons of the bilingual-stimuli /r-/l/ and /r-/r/ identification revealed that there was significant difference,  $t(35) = -9.25, p < 0.001$ , suggesting that their accuracy for /r-/r/ identification was generally better than that for English /r-/l/ identification. However, Pearson



**Figure 2.3 Scatterplot of the relationship between English /r/-/l/ identification accuracy and /r/-/r/ bilingual identification accuracy (i.e., the degree of category assimilation between /r/ and /r/). Higher /r/-/r/ identification scores indicate less confusion between the categories, and a lower degree of category assimilation. The results demonstrated that the two measures were significantly correlated, suggesting that the category assimilation between /r/ and /r/ may be predictive of English /r/-/l/ identification accuracy.**

correlation analysis revealed that there was significant relationship,  $r = 0.44$ ,  $p < 0.001$ ,

between the monolingual-stimuli /r/-/l/ identification and the bilingual-stimuli /r/-/r/

identification, suggesting that Japanese listeners whose English /r/ category was more

similar to the /r/ category had more problems with English /r/-/l/ identification. This

correlation may indicate that the /r/-/r/ category assimilation is the cause of the /r/-/l/

identification problems. However, the scatterplot in Figure 2.3 weakens this view. That is,

individuals who were poor at identifying English /r/ and /l/ had a range of /r/-/r/ scores, with many of these individuals rarely confusing /r/ and /r/; it is hard to argue that /r/-/r/ assimilations cause their English /r/-/l/ problems given that these assimilations were not made very frequently, and the more strongly assimilated pair, /l/-/r/, was not correlated with identification accuracy. The frequency of /r/-/r/ assimilation may indicate instead that there is a problem with the underlying /r/ representation, without this problem actually being caused by the proximity of /r/.

### **2.3 Experiment 1.2: Ratings of spoken accent**

Experiment 1.1 demonstrated that there was only an indirect causal relationship between category assimilation processes and the English /r/-/l/ identification, although previous studies suggested that category assimilation directly underlies the difficulties in recognizing English /r/ and /l/. Experiment 1.2 examined English /r/-/l/ production by Japanese speakers. Despite the fact that Japanese speakers have poor English /r/-/l/ identification, they can often produce tokens successfully identified as the English phonemes /r/ or /l/ (e.g., Goto, 1971; Sheldon and Strange, 1982) and improve their English /r/-/l/ production through phonetic training and audiovisual training (e.g., Bradlow et al., 1997; Hazan et al., 2005; Massaro and Light, 2003). They can also improve the /r/-/l/ production by living in English-speaking environments (e.g., Aoyama et al., 2004). The aim of Experiment 1.2 was to examine whether Japanese speakers' /r/-

/l/ production is predictive of their /r-/l/ identification, and whether category assimilation is predictive of their /r-/l/ production.

### **2.3.1. Method**

#### **Subjects**

Thirty-six Japanese from the previous experiment took part in stimuli recordings. Three British English speakers having phonetic training participated and rated the accent of L2 speaker's /r-/l/ production and the degree of contrast between the consonants.

#### **Apparatus**

The targeted sentence (i.e., *The red robin looked across from the lovely lake*) was pronounced by the 36 Japanese speakers from Experiment 1.1 and digitized (16-bit depth; 44,100 samples/sec). After the recordings were made, three native speakers of British English having phonetic training rated the recordings.

#### **Procedure**

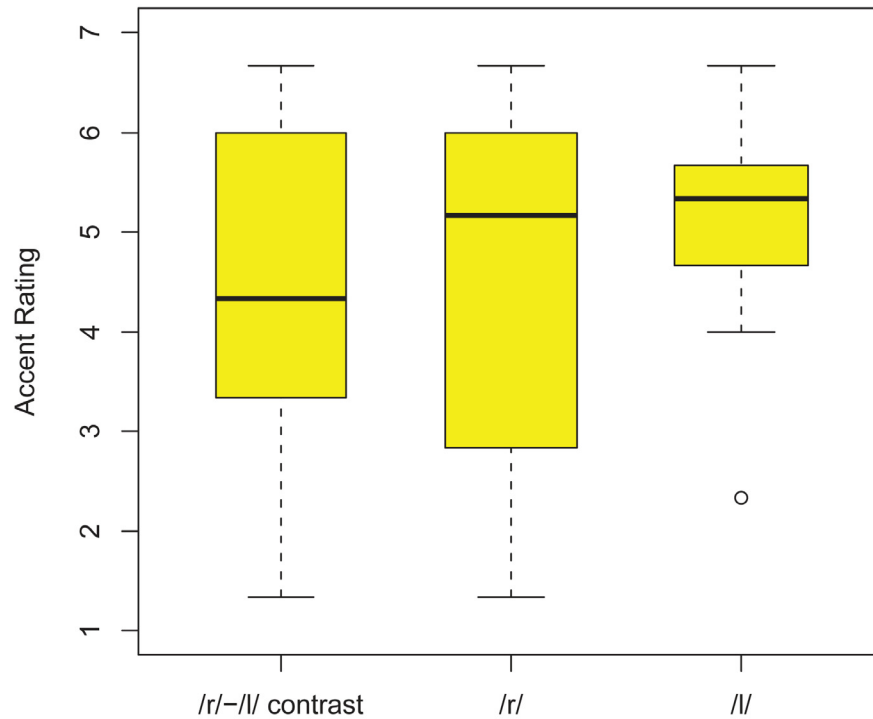
The British English listeners rated /r/-accent (i.e., the degrees of non-native accent for /r/), /l/-accent (i.e., the degrees of non-native accent for /l/), and the degree of /r-/l/ contrast by using a scale from 1 (bad) to 7 (good). For example, some Japanese speakers may produce their English /r/ with high F3 frequencies (e.g., 2600 Hz) and their English /l/ with higher F3 frequencies (e.g., 3300 Hz); such speakers will be rated as producing /r/ with a strong accent, but they will be rated as having a high degree of /r-/l/ contrast (i.e., the consonants sound different, even though /r/ may not be native-like). The English

listeners were allowed to listen to each Japanese speaker's production as many times as they wanted. Rating scores for each Japanese speaker were averaged across the three raters after confirming interrater reliability.

### 2.3.2 Results

Figure 2.4 displays goodness ratings for /r/-accent, /l/-accent, and the degree of /r-/l/ contrast. The Japanese speakers received a wide range of rating scores for the degree of the contrast (median = 4.33) and the goodness of /r/-accent (median = 5.17). In contrast, they received higher goodness /l/-accent scores (median = 5.33) with less variance.

In order to examine whether such accent-rating scores are predictive of English /r-/l/ identification, each of the rating scores and the results of monolingual-stimuli /r-/l/ identification from Experiment 1.1 were analyzed by Pearson correlation analysis. There was a significant correlation between /r/-accent and English /r-/l/ identification,  $r = 0.53$ ,  $p < 0.001$ , indicating that Japanese speakers who pronounced better /r/ were more accurate in English /r-/l/ identification. There was also a significant correlation between the degree of /r-/l/ contrast and /r-/l/ identification,  $r = 0.42$ ,  $p < 0.05$ , indicating that the L2 speakers who made better English /r-/l/ contrast in production demonstrated better accuracy in English /r-/l/ identification. However, there was no significant correlation between /l/-accent and /r-/l/ identification,  $r = 0.20$ ,  $p > 0.05$ , indicating that the degree of /l/-accent is not predictive of English /r-/l/ identification.



**Figure 2.4** Boxplots of the rated accent for Japanese speakers saying “The red robin looked across the lovely lake.” Three ratings were made for each sentence: the degree of /r/-/l/ contrast, /r/-accent, and /l/-accent. Ratings were made on a scale from 1 (poor, indicating a strong non-native accent or poor contrast) to 7 (good, indicating little non-native accent or high contrast). Standard errors from the left side of the figure are 0.26, 0.28, and 0.15. The results demonstrated that there was a wide range of individual differences in the degrees to which English listeners perceived /r/-/l/ contrast and /r/-accent. However, there was less variance in the degree to which English listeners perceived /l/-accent.

In order to examine whether the category assimilation process is predictive of English /r/-/l/ production, correlational analyses were run between the results of bilingual-stimuli /r/-/l/ identification from Experiment 1.1 and the three accent-rating scores. For the /r/-/l/ identification, there was no significant correlation with /r/-accent,  $r = 0.13$ ,  $p > 0.05$ , /l/-accent,  $r = 0.19$ ,  $p > 0.05$ , and the degree of /r/-/l/ contrast,  $r = 0.08$ ,  $p > 0.05$ , suggesting that /r/-/l/ category assimilation process is not predictive of either English /r/-/l/ perception or English /r/-/l/ production. For the /r/-/r/ identification, there



was no significant correlation with /r/-accent,  $r = 0.29$ ,  $p > 0.05$ , /l/-accent,  $r = 0.17$ ,  $p > 0.05$ , and the degree of /r/-/l/ contrast,  $r = 0.23$ ,  $p > 0.05$ , suggesting that /r/-/r/ category assimilation process is not predictive of either English /r/-/l/ perception or English /r/-/l/ production.

In summary, Japanese speakers demonstrated reasonable correlations between their /r/-/l/ production and /r/-/l/ identification. There was a significant correlation between /r/-accent and English /r/-/l/ identification, but there was not a significant correlation between /l/-accent and English /r/-/l/ identification, suggesting that Japanese speakers learned English /r/ better than /l/ and improved their perception and production. In other words, English /l/ is difficult to learn for Japanese speakers. The L2 speakers also demonstrated that category assimilation process did not predict the accuracy of English /r/-/l/ production.

### **2.4 Experiment 1.3: Perceptual mapping of best exemplars**

Experiment 1.2 demonstrated that English /r/-/l/ production by Japanese speakers is moderately predictive of their English /r/-/l/ identification, and that category assimilation was not predictive of English /r/-/l/ production. Experiment 1.3 examined whether the mental representations of Japanese speakers are predictive of English /r/-/l/ identification. In order to examine the relationship between Japanese listeners' mental representations and their English /r/-/l/ recognition, the phonetic similarity among English /r/, /l/, and Japanese /r/ was assessed by using the goodness optimization

procedure (i.e., a perceptual mapping approach) and finding best exemplar of English /r/, /l/, and Japanese /r/.

### **2.4.1 Method**

#### **Subjects**

The same 36 Japanese speakers were tested, together with 13 native speakers of British English participated in this experiment as well. Their age ranged between 18 and 62 years (median = 24). They were all born and raised in southern England. They reported no hearing problems.

#### **Stimuli and apparatus**

The stimuli were synthetic C-/a/ syllables embedded in naturally spoken English and Japanese carrier sentences (i.e., *Say [ ] again*, and *mata [ ] to itte kudasai*). The speaker was an English/Japanese bilingual who spent her childhood in both the UK and Japan, and was highly fluent in both languages.

The synthetic syllables were modelled on those of the natural talker, using a Klatt synthesizer (Klatt and Klatt, 1990). A 5-dimensional set of stimuli was created by orthogonally varying F1, F2, F3, closure duration, and transition duration (from the consonantal articulation to the following vowel). Closure duration was defined as from the point where voicing began to the point where F1 transition began to happen (see Figure 2.5 and Figure 2.6). Transition duration was from the point where F1 transition began to either 1) the point where F1 appeared at its zenith (i.e., where F1 transition

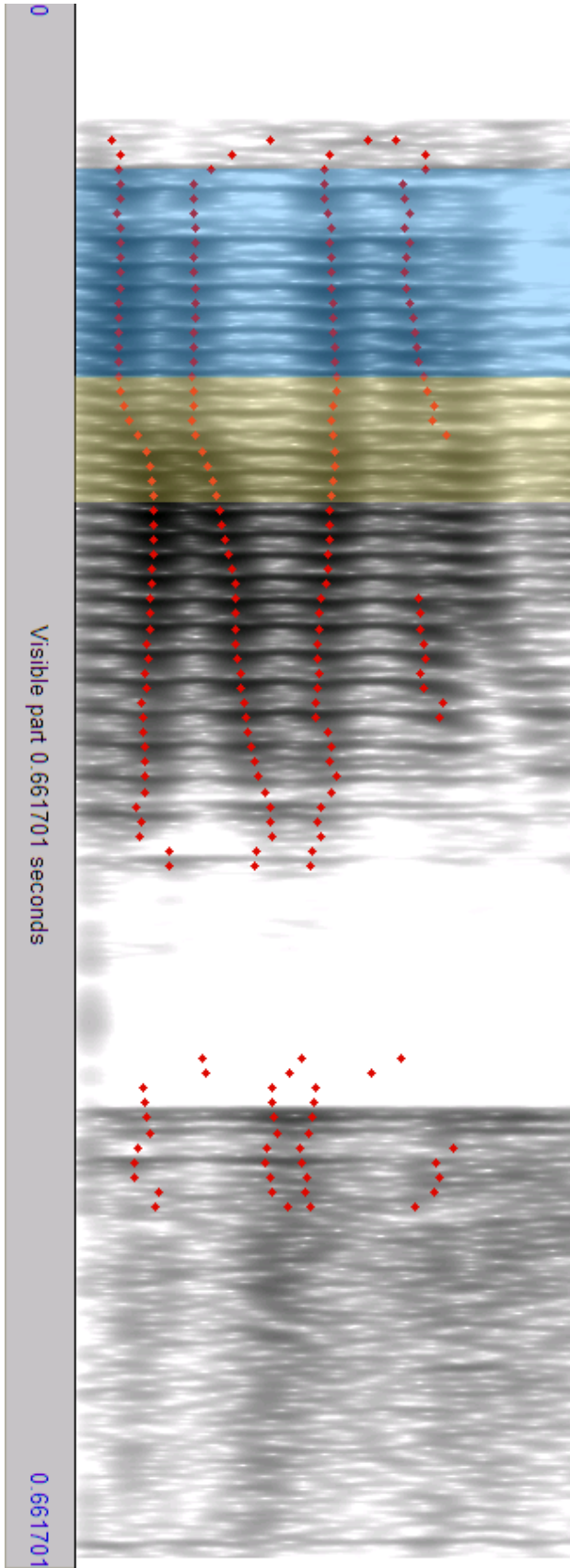
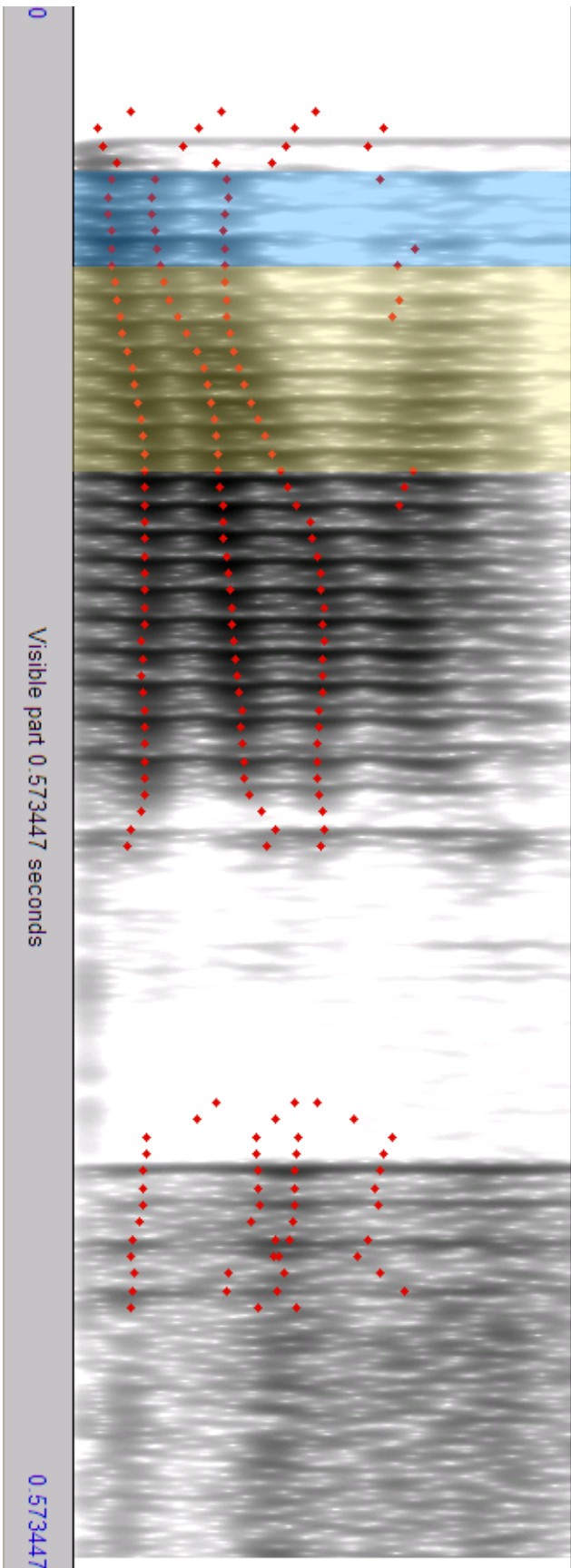


Figure 2. 5 A spectrogram of *lack* produced by a male British English speaker. The blue area indicates closure duration where F1, F2, and F3 generally stay stable. The yellow area indicates transition duration where F1 starts increasing and hits its peak. Closure duration for /l/ is generally longer and transition duration for the liquid is shorter.

Figure 2. 6 A spectrogram of *rack* produced by a male British English speaker. The blue area indicates closure duration where F1, F2, and F3 generally stay stable. The yellow area indicates transition duration where F1 starts increasing and hits its peak. Closure duration for /r/ is generally shorter and transition duration for the liquid is longer.



finished; see Figure 2.5 and Figure 2.6) or 2) the point where F3 appeared at its zenith (i.e., where F3 transition finished). The values were chosen so that they would span an acoustic space that included /r/, /l/, and /r/. F1 varied from 123-603 Hz. F2 was always at least 1 ERB (Glasburg and Moore, 1990) greater than F1, and less than 2489 Hz. F3 was always at least 1 ERB greater than F2, and less than 3951 Hz. The closure duration varied from 66 to 209 ms, and the transition duration varied from 5 to 209 ms. The frequency values were quantized in 1-ERB steps and the duration values were quantized with a log spacing (5 steps for closure duration, and 12 for transition duration). There were a total of 60,660 stimuli for each language.

### **2.4.2 Procedure**

We adapted the goodness optimization procedure that has been used previously for vowels (Evans and Iverson, 2004) to find the best exemplar of each consonant. On each trial, subjects saw a target consonant presented on the screen (i.e., "R", "L", or "Japanese R"), heard a sentence, and rated on a continuous scale whether the consonant that they heard was close or far away from the target consonant.

On each successive trial, a computer algorithm adjusted the acoustics of the stimuli to converge on a good exemplar of that consonant. The algorithm adjusted the stimuli along 7 vectors (i.e., straight-line paths) through the multidimensional space. Subjects first adjusted the stimuli on a vector that co-varied all 5 dimensions, and passed through a location in the space that corresponded to acoustic measurements of that consonant, allowing subjects to converge on a best exemplar quickly. They then searched

on vectors that varied individual dimensions (F3, F2, transition duration, closure duration, and F1). Finally, they searched along a vector that varied all 5 dimensions and allowed them to fine-tune the best exemplar that they had found thus far.

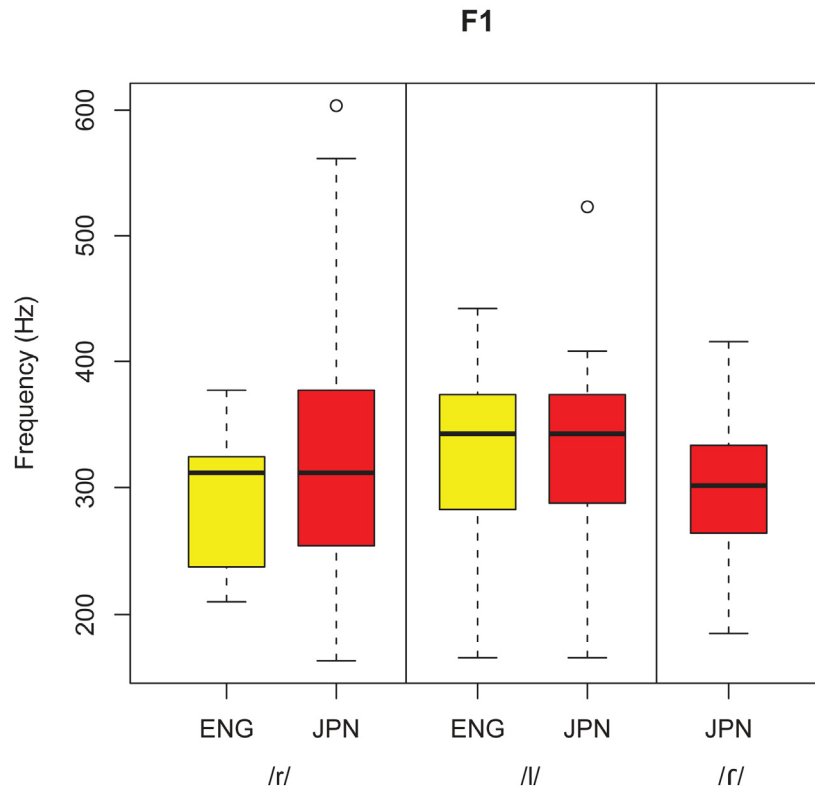
Along each of these search vectors, subjects found the best exemplar along that vector in a series of 5 trials. Subjects first heard the first two most extreme stimuli on the search vector (i.e., at the limits of the synthesis set), in a random order. Then the next 3 trials were chosen based on fitting parabolas to the previous goodness judgments (i.e., to predict which stimulus would be the smallest distance away from being a good exemplar).

Subjects thus were able to find best exemplars for each consonant after 35 trials. Japanese subjects searched for all three consonants in both carrier sentences (Japanese and English). English speakers searched for /r/ and /l/ only, in both carrier sentences (i.e., they were told just to attend to the target syllable in Japanese, and ignore the fact that they did not understand the entire sentence).

### 2.4.3 Results

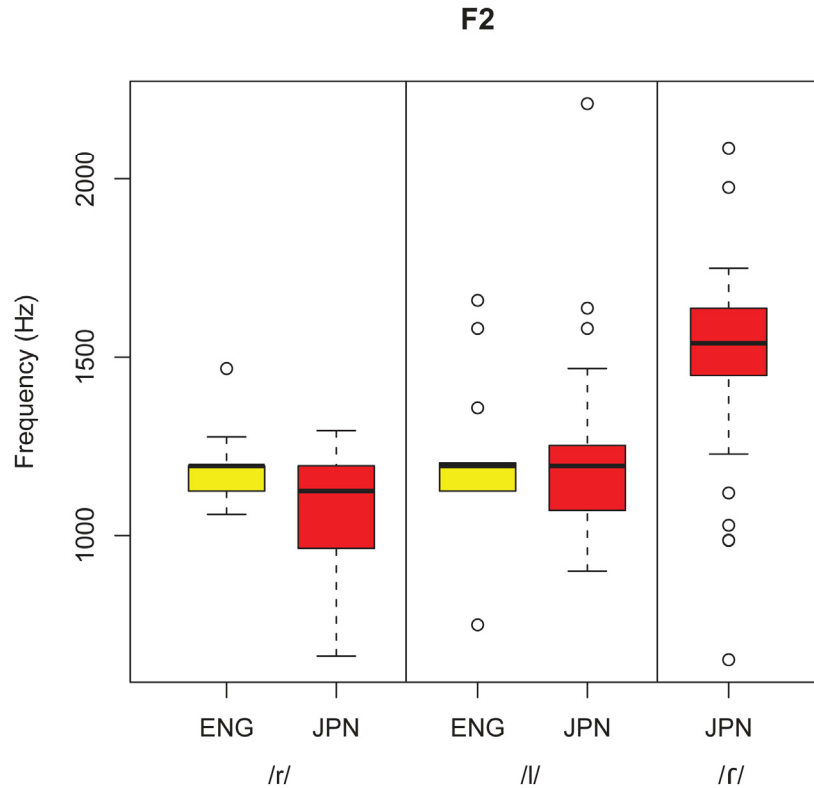
Figure 2.7 to 2.11 display the best exemplars of English /r/, /l/, and Japanese /r/. It is evident that Japanese listeners demonstrated a clear distinction between English /r/ and Japanese /r/. For example, they clearly separated the consonants in F3 (see Figure 2.9) and transition duration (see Figure 2.11). However, it is not clear whether the L2 listeners separated English /l/ and Japanese /r/. For example, they chose similar F3 and transition durations.

In order to examine whether Japanese listeners established three separate consonant categories, a one-way repeated-measures ANOVA was run for each of the five acoustic dimensions. The dependent variables were the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration). The within-subjects variable was consonant (i.e., English /r/, /l/, or Japanese /r/).



**Figure 2.7** Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected similar F1 frequencies for /r/ and /l/. Japanese speakers selected similar F1 frequencies for /r/, /l/, and /r/ which were similar to those selected by English speakers. Standard errors from the left side of the figure are 14.65, 16.79, 20.45, 11.97, and 9.34 (Hz). The results demonstrated that English and Japanese speakers were not different in F1 frequencies for /r/ and /l/. The results also demonstrated that Japanese speakers selected F1 frequencies for /r/ similar to those for /r/ and /l/.

For F1 (see Figure 2.7), there was no main effect of consonant,  $F(2, 70) = 1.15$ ,  $p > 0.05$ . Tukey HSD comparisons confirmed that the consonants are similar in the F1 dimension, /r/-/l/,  $z = -0.20$ , /l/-/r/,  $z = 1.40$ , /r/-/r/,  $z = 1.20$ ,  $p > 0.05$ .

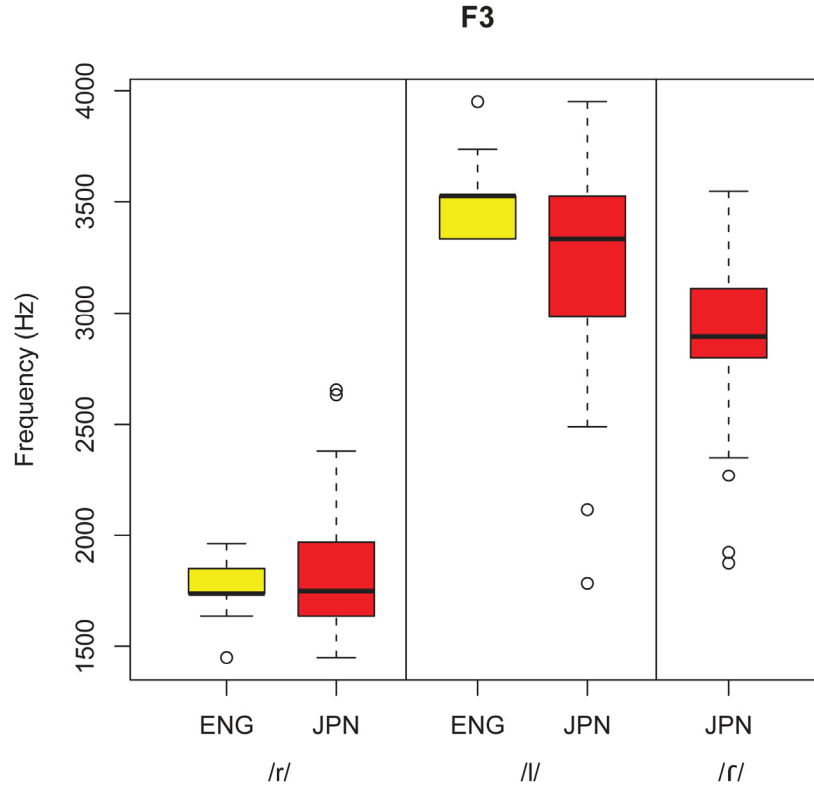


**Figure 2.8** Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected similar F2 frequencies for /r/ and /l/. Japanese speakers selected F2 frequencies for /r/ and /l/ similar to those selected by English speakers. However, they selected higher F2 frequencies for /r/ than did they for the English liquids. Standard errors from the left side of the figure are 27.75, 27.37, 61.65, 40.89, and 51.28 (Hz). The results demonstrated that Japanese speakers separated /r/ and /l/ in the F2 dimension as English speakers did. The results also demonstrated that Japanese speakers separated /r/ from English liquids in the F2 dimension.

For F2 (see Figure 2.8), there was a main effect of consonant,  $F(2,70) = 33.97$ ,  $p < 0.001$ , suggesting that the consonants are separated in Japanese speakers' mental representations. Tukey HSD comparisons revealed that there was a significant difference between /r/ and



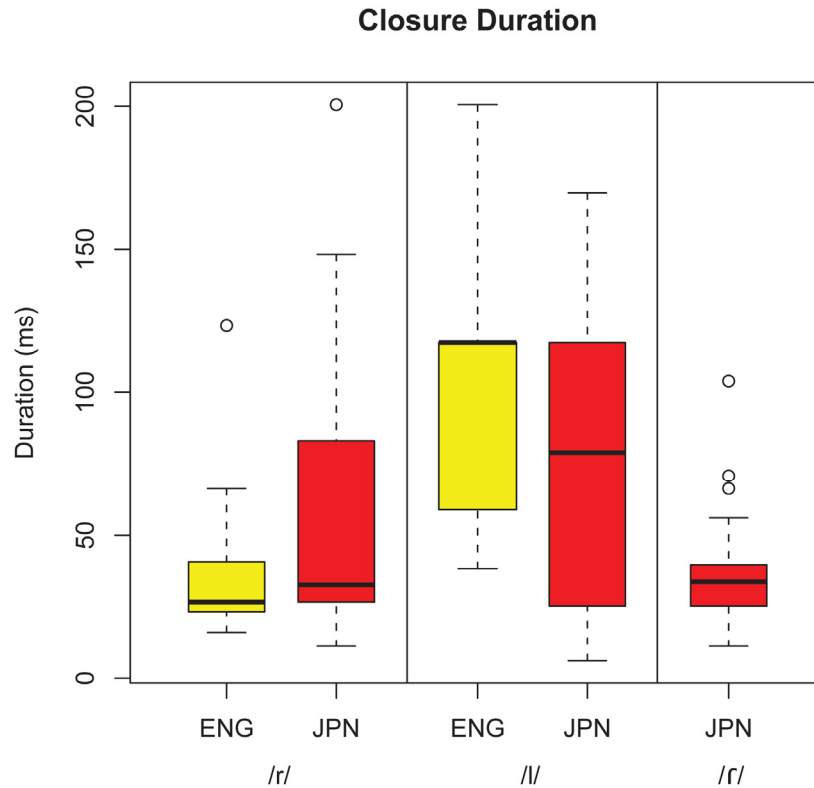
/l/,  $z = -2.76, p < 0.05$ , between /l/ and /r/,  $z = -5.35, p < 0.001$ , and between /r/ and /r/,  $z = -8.10, p < 0.001$ . This confirms that Japanese /r/ differed from both English /r/ and /l/, and that the English consonants were different in the F2 dimension.



**Figure 2.9** Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers clearly selected lower F3 frequencies for /r/ and higher F3 frequencies for /l/ with little variance. Likewise, Japanese speakers selected lower F3 for /r/ and higher F3 for /l/ and /r/ with a wide range of variance. Standard errors from the left side of the figure are 36.33, 57.93, 50.91, 78.06, and 63.30 (Hz). The results demonstrated that Japanese speakers were particularly inaccurate in F3 frequencies for /l/ compared to English speakers. However, the results demonstrated that Japanese speakers separated /r/, /l/, and /r/ in the F3 dimension.

For F3 (see Figure 2.9), there was a main effect of consonant,  $F(2,70) = 110.15, p < 0.001$ . Tukey HSD comparisons revealed that there was a significant difference between /r/ and /l/,  $z = -14.16, p < 0.001$ , between /l/ and /r/,  $z = 3.21, p < 0.01$ , and between /r/

and /r/,  $z = -10.94$ ,  $p < 0.001$ . This confirms that Japanese listeners kept three phonetic categories separated in the most important acoustic dimension to distinguish /r/ and /l/ despite the fact that they generally demonstrated poor English /r-/l/ identification accuracy in Experiment 1.1.

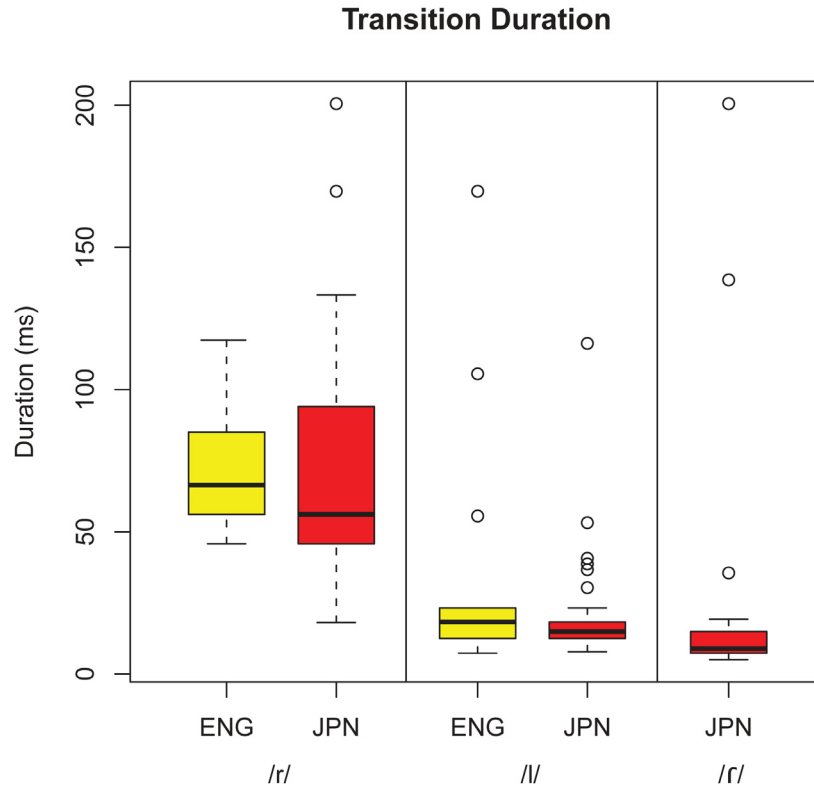


**Figure 2.10** Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected short closure durations for /r/ and long closure durations for /l/. Japanese speakers selected closure durations for /r/ and /l/ similar to those by English speakers with a wide range of variance. They selected short closure durations for /r/ with less variance. Standard errors from the left side of the figure are 8.07, 7.52, 12.86, 8.31, and 3.06 (ms). The results demonstrated that Japanese speakers did not differ from English speakers in closure durations for /r/ and /l/. The results also demonstrated that Japanese speakers separated /r/ from /l/, and /l/ from /r/, but they did not separate /r/ from /r/.

For closure duration (see Figure 2.10), there was a main effect of consonant,  $F(2,70) = 9.85$ ,  $p < 0.001$ . Tukey HSD comparisons revealed that there was a significant difference

between /r/ and /l/,  $z = -2.44$ ,  $p < 0.05$ , and between /l/ and /r/,  $z = 4.43$ ,  $p < 0.001$ . There

was no significant difference between /r/ and /r/,  $z = 1.99$ ,  $p > 0.05$ .



**Figure 2.11** Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN), and for /r/ by L1 speakers of Japanese. English speakers selected long durations for /r/ and short closure durations for /l/. Japanese speakers similarly selected long transition durations for /r/ with a wide range of variance, similar short transition durations for /l/ and /r/ with little variance. Standard errors from the left side of the figure are 7.15, 7.03, 13.25, 3.28, and 6.33 (ms). The results demonstrated that Japanese speakers did not differ from English speakers in transition durations for /r/ and /l/. The results also demonstrated that Japanese speakers separated /r/ from /l/, and /r/ from /r/, but they did not separate /l/ from /r/.

For transition duration (see Figure 2.11), there was a main effect of consonant,  $F(2,70) = 26.81$ ,  $p < 0.001$ . Tukey HSD comparisons revealed that there was a significant difference between /r/ and /l/,  $z = 6.20$ ,  $p < 0.001$ , and between /r/ and /r/,  $z = 6.47$ ,  $p < 0.001$ . There was no significant difference between /l/ and /r/,  $z = 0.27$ ,  $p > 0.05$ . This

confirms that English /r/ differed from both English /l/ and Japanese /r/ in transition duration.

Although Japanese listeners generally separated the three phonetic categories in their mental representations (i.e., F2, F3, closure duration, and transition duration), there were substantial individual differences in their best exemplars. For example, closure duration of English /r/ and /l/ widely varied and included the range of Japanese /r/. It is thus possible that some Japanese listeners may have less clear category separation while other Japanese listeners have clearer category separation. In order to examine whether such individual differences in mental representations are predictive of English /r/-/l/ identification accuracy, one-dimensional Euclidean distances between /r/ and /r/, and /l/ and /r/ were computed for the five acoustic dimensions (e.g.,  $|\text{/r/ F1 frequency} - \text{/r/ F1 frequency}|$ ). Then, correlational analyses were run with the Euclidean distances and English /r/-/l/ identification accuracy. The statistical analyses revealed that there were no significant correlations between the phonetic distance between /l/ and /r/ and English /r/-/l/ identification accuracy, F1,  $r = -0.12$ ; F2,  $r = -0.17$ ; F3,  $r = 0.11$ ; closure duration,  $r = 0.28$ ; transition duration,  $r = -0.04$ ,  $p > 0.05$ , suggesting that, although /l/ and /r/ are phonetically similar, the phonetic distance between /l/ and /r/ in mental representation is not predictive of their English /r/-/l/ identification. The statistical analyses also revealed that there were no significant correlations between the /r/-/r/ distance and English /r/-/l/ identification, F1,  $r = -0.14$ ; F2,  $r = -0.07$ ; F3,  $r = 0.26$ ; closure duration,  $r = -0.20$ ; transition duration,  $r = 0.14$ ,  $p > 0.05$ , suggesting that individual differences in the

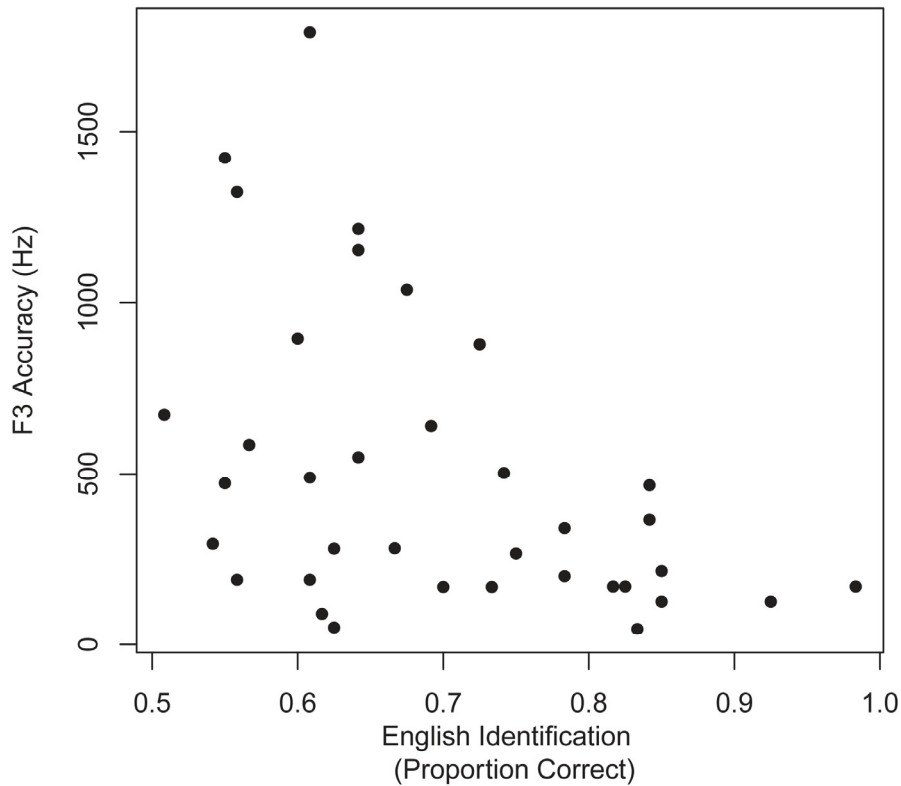
phonetic distance between /r/ and /l/ in mental representations is not predictive of English /r/-/l/ identification accuracy either.

Japanese listeners demonstrated separated three consonant categories in their mental representations. However, it is not clear yet whether they have native-like mental representations for English /r/ and /l/. In order to examine whether the best exemplars of English /r/ and /l/ by Japanese speakers are similar to those by English speakers, each of the five acoustic dimensions was analyzed using 2-way ANOVA. Dependent variables were the five acoustic parameters. A between-subject factor was language group (i.e., English or Japanese), and a within-subject factor was consonant (i.e., /r/ or /l/). For F1, there was no main effect on language group,  $F(1,47) = 0.47, p > 0.05$ . There was no main effect of consonant,  $F(1,47) = 0.55, p > 0.05$ . There was no significant interaction between language group and consonant,  $F(1,47) = 0.78, p > 0.05$ . For F2, there was no main effect of language group,  $F(1,47) = 1.88, p > 0.05$ . There was a main effect of consonant,  $F(1,47) = 9.14, p < 0.01$ , suggesting that /r/ and /l/ differed in the F2 dimension. There was no significant interaction,  $F(1,47) = 1.70, p > 0.05$ . For F3, there was no main effect of language group,  $F(1,47) = 2.18, p > 0.05$ . There was a main effect of consonant,  $F(1,47) = 296.04, p < 0.001$ , suggesting that /r/ and /l/ clearly differed in the F3 dimension. There was also a significant interaction,  $F(1,47) = 4.79, p < 0.05$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was not significant,  $t(46) = 0.86, p > 0.05$ , suggesting that Japanese speakers chose F3 frequencies similar to those by English speakers, but the effect of language group for /l/ was significant,  $t(46) = -2.64, p < 0.05$ , suggesting that English speakers systematically chose higher F3 for /l/ than Japanese speakers did. These results thus

indicate that Japanese speakers did not have completely native-like representations particularly for /l/ in the F3 dimension. For closure duration, there was no main effect of language group,  $F(1,47) = 0.19, p > 0.05$ . There was a main effect of consonant,  $F(1,47) = 14.77, p < 0.001$ , suggesting that English /r/ differed from /l/; /r/ has a shorter closure duration than does English /l/. There was a significant interaction,  $F(1,47) = 4.50, p < 0.05$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was not significant,  $t(46) = 1.08, p > 0.05$ , and the effect of language group for /l/ was not significant,  $t(46) = -1.74, p > 0.05$ , suggesting that Japanese speakers were similar to English speakers in producing closure durations for /r/ and /l/. These results thus demonstrated that, although there was a significant 2-way interaction between language group and consonant, Japanese speakers did not differ from English speakers in producing closure durations for /r/ and /l/. For transition duration, there was no main effect of language group,  $F(1,47) = 0.91, p > 0.05$ . There was a main effect of consonant,  $F(1,47) = 47.06, p < 0.001$ , suggesting that English /r/ and /l/ differ in transition duration; /l/ has shorter transition duration. There was no significant interaction,  $F(1,47) = 0.97, p > 0.05$ .

Japanese listeners thus did not have completely native-like /r/ and /l/ representations; F3 is particularly less accurate with individual differences. Pearson correlations were used to determine whether individual differences in the accuracy with which Japanese listeners represent /r/ and /l/ (i.e., their similarity to the average best exemplars chosen by native speakers of English) was predictive of identification accuracy. The accuracy of the representations was measured for each of the five acoustic dimensions, combining /r/ and /l/ using a Euclidean metric. For example, the author

calculated how far each individual's /r/ and /l/ best exemplars were from the English averages on the F3 dimension, and then combined these two values by calculating the square root of the sum of squares. The statistical analyses revealed that there was significant correlation between F3 accuracy and /r-/l/ identification accuracy,  $r = -0.45$ ,  $p < 0.01$ , suggesting that Japanese individuals who had F3 representations that were similar to those of native speakers of English were better at English /r-/l/ identification accuracy (see Figure 2.12). However, there were no significant correlations



**Figure 2.12** Scatterplot of the relationship between the accuracy of the English /r/ and /l/ best exemplars for Japanese speakers along the F3 dimension (i.e., distance from English averages) and their English /r-/l/ identification accuracy from Experiment 1.1 (i.e., measuring base line recognition performance). Japanese speakers who demonstrated lower /r-/l/ identification accuracies generally had more distance from English averages; they were inaccurate in the F3 dimension. Japanese speakers who demonstrated higher /r-/l/ identification accuracies had short distance from English averages; they were more accurate in the dimension. The results demonstrated that there was a significant correlation between the accuracy of the English /r/ and /l/ best exemplars and /r-/l/ identification accuracy.

between the other four acoustic dimensions and /r/-/l/ identification accuracy, F1,  $r = -0.32$ , F2,  $r = -0.03$ , closure duration,  $r = -0.31$ , and transition duration,  $r = -0.02$ ,  $p > 0.05$ .

## **2.5 General Discussion**

The present study used an individual differences approach to examine whether category assimilation processes between English /r/-/l/ and Japanese /r/ predict /r/-/l/ identification accuracy. The present results confirmed that English /l/ is assimilated into the Japanese /r/ category, at least to some extent. That is, in forced-choice identification tasks, Japanese speakers confused /l/ and /r/ as commonly as they confused /r/ and /l/, despite the fact that their best exemplars for /l/ and /r/ were similar to some extent (e.g., F3 and transition duration), but kept separated in mental representations. The assimilation of /r/ and /l/ was comparatively weak; individuals confused /r/ and /l/ infrequently in a forced-choice task and chose best exemplars for /r/ that were different from /l/ in terms of F3 and transition duration. The results thus suggest that /r/ and /l/ differentially assimilate into the Japanese /r/ category, with /l/ being closer, supporting the claims of Aoyama *et al.*, 2004). It is possible that previous work using subjective ratings of category goodness (e.g., Guion *et al.*, 2000; Iverson *et al.*, 2003) did not find a clearer asymmetry because it is hard for individuals to rate differences in category goodness when both tokens are poor exemplars overall.



Of interest of L2 perception models, this seems to be problematic to PAM (e.g., Best, 1995, Best et al., 2001). Best and Strange (1992) hypothesized that Japanese speakers assimilate both English /r/ and /l/ into the Japanese /r/ category (i.e., single category assimilation). However, Japanese speakers assimilated English /r/ and /l/ into Japanese /r/ category at least with different degrees, suggesting that this could be a Category Goodness (CG) difference (Best, 1995). The problem for PAM is that CG contrasts of L2 phonemes are relatively easy to discriminate, but the perception of English /r/-/l/ by Japanese speakers has apparently been one of the most difficult cases in the literature (e.g., Goto, 1971; Miyawaki et al., 1975). Given that Iverson et al. (2003) demonstrated that Japanese speakers are sensitive to acoustic variations in the F2 dimension (i.e., Japanese speakers have sensitivity to a certain relevant acoustic parameter distinguishing /r/ and /l/), Japanese speakers may be able to identify /r/ and /l/. However, they poorly identify /r/ and /l/, which calls the CG difference into question. It thus seems that PAM does not explain why Japanese speakers have difficulty with English /r/-/l/ identification.

The results of this study are also problematic for SLM. Previous work supporting SLM has suggested that language learners merge (Flege, 2003; MacKay *et al.*, 2001) or dissimilate (Flege, 1995; Flege *et al.*, 2003) L1 and L2 phonetic categories that are close enough to interact, so it was expected that Japanese learners who were better at identifying /r/ and /l/ would have assimilation patterns that would be distinct from those who identified /r/ and /l/ less accurately. The present study found little evidence that such individual differences in assimilation were related to /r/-/l/ identification performance.

There was a significant correlation between /r/-/r/ assimilation and identification

performance in Experiment 1.1, so the possibility that this assimilation makes it hard to learn to recognize these phonemes cannot be completely discarded. However, it seems unlikely that /r/-r/ assimilation is the actual cause of these learning problems because very few listeners made this confusion frequently. Moreover, the strength of /l/-r/ assimilation was stronger than /r/-r/, so SLM predicts that /l/-r/ assimilation should have an even bigger effect on category learning (Aoyama *et al.*, 2004). Instead, we found no reliable correlation of /l/-r/ assimilation with the individual differences in identification performance.

In addition, the results of Experiment 1.3 suggest that Japanese speakers were able to maintain three separate categories for /r/, /l/, and /r/, despite the fact that the subject group as a whole was not particularly good at identifying /r/ and /l/. For example, their best exemplars demonstrated that they knew that /l/ has a longer initial closure and a lower F2 than /r/; the closeness of /l/ and /r/ did not cause them to assimilate these phonemes into a single category. The only evidence for problems caused by similarity between /l/-r/ was the fact that the best exemplars for /l/ by Japanese speakers were less native-like along the F3 dimension, compared to their best exemplars for /r/. This fits SLM's predictions (Aoyama *et al.*, 2004) that assimilation causes more problems for learning /l/ than /r/. However, it is unexplained why this should affect only the F3 dimension; the best exemplars of /l/ for Japanese listeners were not significantly different from those of native speakers in any other respect.

The results of English /r/-/l/ production are also problematic to SLM although the results suggested that there is a relationship between perception and production of English /r/-/l/ by Japanese speakers, which is in accord with SLM (Flege, 1999). SLM (Flege, 1995) predicts that an L2 phoneme which has a short phonetic distance to an L1 category is difficult to acquire; learning English /l/ production is more difficult than learning /r/ production. The present results demonstrated that Japanese speakers are better at producing identifiable /l/ than /r/. This suggests that learning the production of an L2 phoneme which is more similar to an L1 phoneme may be easier than learning the production of L2 phoneme which is more dissimilar. In order to verify this point, some acoustic measurements of /l/ production by Japanese speakers are necessary. Chapters three and four will provide the measurements in detail.

Despite the fact the L2 perception models hypothesize that category assimilation processes cause difficulties for identifying L2 phonemes, the present study rather provided evidence that perceptual accuracies in mental representations may explain such L2 learning problems. The positive finding was that the representation of F3 is important for /r/-/l/ identification; individuals who had more native-like best exemplars along the F3 dimension were more accurate at identifying /r/ and /l/. F3 has been known to be problematic for Japanese speakers for a long time (e.g., Miyawaki et al., 1975), but this is perhaps the first study to demonstrate that such results on synthetic speech relate to how individuals identify real speech, produced by multiple talkers in multiple word contexts. It is worth noting that this correlation was not particularly high,  $r = -0.46$ , but this may have occurred because individuals with poor identification ability were fairly random in the F3 values that they preferred.

In summary, the present study demonstrated that Japanese speakers asymmetrically assimilated English /r/ and /l/ into Japanese /r/; /l/ assimilates to /r/ more strongly than does /r/. However, individual difference in the degree of category assimilation process was not predictive of /r/-/l/ identification accuracy. The present study, instead, demonstrated that perceptual accuracy in F3 dimension was related to identification accuracy.

## **Chapter 3: Examination of the relationship between perception and production of English /r/-/l/ by adult Japanese speakers**

### **3.1 Introduction**

Over the last 50 years, speech scientists and experimental phoneticians have theorized and examined the relationship between speech perception and speech production. For example, motor theory (e.g., Galantucci et al., 2006; Liberman et al., 1967; Liberman et al., 1952; Liberman & Mattingly, 1985, 1989) has hypothesized that human listeners perceive speech as the speaker's intended articulatory gestures (e.g., intended movement of tongue, lips, and vocal folds) but not the actual acoustic patterns generated by the articulatory gestures. Motor theory has hypothesized that human beings perceive speech with a speech-specific module (i.e., phonetic module) but not with a general perception mechanism. For speech perception, this phonetic module detects the intended articulatory gesture (i.e., neuromotor commands) from the acoustic signal and relates such information to abstract phonological knowledge (e.g., phonemes or distinctive features). For speech production, the module translates the abstract knowledge to neuromotor commands in order to produce the intended realization of phonemes. Thus, motor theory predicts that there is a very close relationship between speech perception and production.

Direct realist theory (e.g., Best, 1995; Fowler, 1981, 1984, 1986, 1989, 1994, 1996) argues that the information in the acoustic patterns (i.e., waveform) is sufficient enough to specify the actual articulatory gestures (e.g., vocal tract, lip, and tongue

movements), and hypothesizes that human beings perceive such actual articulatory gestures produced by a speaker, rather than the speaker's intended articulatory gesture (i.e., neuromotor commands), through the acoustic patterns (i.e., waveform). That is, a listener reconstructs a speaker's actual articulatory movements through the acoustic waveform which is structured by the speaker's articulators. Direct realist theory does not hypothesize any special mechanisms corresponding to the phonetic module posited motor theory. Direct realist theory rather hypothesizes that humans use general perceptual systems, which have a universal function and include the sole means by which animals can know the environmental conditions in which they successfully live. For speech perception, the perceptual systems use acoustic structure (i.e., waveform) that has been lawfully caused by articulatory movements (e.g., vocal tract, lip, and tongue movements) as information for the movements. Even though it is the waveform that sense organs convert, it is not the waveform that humans perceive. Rather, humans perceive actual articulatory gestures.

The general auditory approach (GAA; e.g., Diehl, et al., 2004) hypothesizes that human listeners perceive speech through the actual acoustic patterns but not articulatory gestures. GAA, unlike motor theory, does not posit a special module for speech perception. Rather, GAA hypothesizes that human beings perceive speech using general auditory mechanisms (i.e., audition and perceptual learning mechanisms). GAA, like motor theory and direct realist theory, posits a close relationship between speech perception and production; "production follows perception, and perception follows production" (Diehl et al., 2004, p.167). For the first half of the hypothesis, if a speaker is required to speak clearly (i.e., if there is a demand for clear auditory characteristics),

he/she may maximize interphoneme distance in the phonetic space (e.g., lowering F3 for English /r/ and raising F3 for English /l/). This perceptual demand makes the speaker's production "sharpened up". In such a case, GAA hypothesizes that speech production follows speech perception. For the second half of the hypothesis, human listeners perceive multiple acoustic cues from the acoustic waveform using general auditory mechanisms and make judgments of the abstract content of speech signals (e.g., phonemes).

Some researchers have investigated such a relationship between speech perception and production in terms of second language acquisition, namely the acquisition of L2 phonemes. The Speech Learning Model (SLM) by Flege (1999) hypothesized that the degree of L2 perception accuracy limits how accurately L2 segments are produced although some aspects of perceptual learning may not be incorporated in production. Therefore, SLM predicts that there exists moderate correlations between L2 perception and production accuracies. Some previous studies, indeed, support such a prediction (e.g., Cheng and Zhang, 2009; Flege and Schmidt, 1995; Flege et al., 1997; Flege et al., 1999; Kluge et al, 2007; Wang et al., 2003). For example, Flege and Schmidt (1995) examined English VOT productions by Spanish speakers. They revealed that Spanish speakers whose English proficiency was high showed a significant correlation between English VOT perception and production accuracies, but Spanish speakers whose English proficiency was low did not, suggesting that L2 perception and production begin to align when L2 learners gain proficiency in the L2. Flege et al. (1999) examined whether English vowel production and discrimination by Italian speakers are related. The results demonstrated that the production accuracy of

English vowels (i.e., identification accuracy by English speakers and goodness rating scores) and discrimination accuracy of English vowels were related. The previous chapter also examined whether English /r-/l/ identification accuracy and production accuracy (i.e., accent ratings) are correlated. The results demonstrated that English /r-/l/ identification accuracy was significantly correlated with the accent ratings for /r/-accent and /r-/l/ contrast. However, there was no significant correlation between the identification accuracy and /l/-accent.

Despite such empirical evidence above suggesting the existence of the relationship between L2 perception and production accuracies, some other previous studies have suggested that L2 production accuracy may precede L2 perception. That is, L2 production and perception may not be related. For example, Gass (1984) investigated perception and production of English /p-/b/ VOT by Farsi, French, Italian, Japanese, Korean, Portuguese, and Thai speakers. The results revealed that VOT of non-native speakers of English differed from that of English speakers, and that the L2 learners seemed to be able to produce English /p-/b/ with native-like VOT. Tsukada et al. (2005) examined discrimination and production of English vowels by Korean speakers. They demonstrated that Korean children are better at English vowel production than English vowel discrimination. Goto (1971) demonstrated that, although some Japanese speakers poorly identified /r-/l/, they produced identifiable /r-/l/. Sheldon and Strange (1982) similarly demonstrated that even some Japanese speakers who produce adequate /r-/l/ had less accurate /r-/l/ perception. Despite such evidence, this view is undermined by the fact that these studies did not directly examine correlations between L2 perception (e.g., identification, discrimination accuracies) and L2 production (e.g., identification and



goodness ratings by native speakers). Therefore, it is still questionable whether L2 production accuracy precedes L2 perception.

The present study further examined whether English /r/-/l/ perception and production of adult Japanese speakers are related using an individual differences approach like the previous chapter. Despite the fact that previous studies (i.e., Goto, 1971; Sheldon and Strange) suggested /r/-/l/ production may precede /r/-/l/ perception (i.e., Japanese speakers may always have better production accuracy than perceptual accuracy), it seems that Japanese speakers are widely varied in terms of /r/-/l/ production accuracy; some Japanese speakers can produce /r/-/l/, but some others poorly produce the consonants. For example, Lotto et al. (2004) acoustically measured /r/-/l/ production of Japanese speakers and demonstrated that some Japanese speakers made clear distinctions between /r/ and /l/ in the F3 dimension, but some others did not. Masaki et al. (1996) used an MRI-based analysis of the /r/-/l/ articulations and demonstrated that some Japanese speakers had native-like articulations (i.e., tongue configurations), but some others had inaccurate articulations. Flege et al. (1995) demonstrated that Japanese speakers who lived in the US over 20 years produced identifiable /r/ and /l/, but Japanese speakers who lived in the US for only two years produced poor /r/ and /l/.

The other purpose of this study was to examine how /r/-/l/ discrimination sensitivity is related to /r/-/l/ production ability and other perceptual abilities (e.g., identification and best exemplars). Previous studies (e.g., Iverson et al., 2003; Miyawaki et al., 1975; Underbakke et al., 1988) demonstrated that Japanese speakers have poor discrimination sensitivity in the F3 dimension at the /r/-/l/ category boundary; Japanese speakers seem insensitive to the most important acoustic cue distinguishing /r/ and /l/.

However, Japanese speakers are not entirely insensitive to acoustic variations in the F3 dimension. Iverson et al. (2003) demonstrated that Japanese speakers are sensitive to F3 variations within the /l/ category. The study also demonstrated that Japanese speakers have higher sensitivity to acoustic variations in the F2 dimension than do L1 English speakers. These results suggest that Japanese speakers' perceptual space is distorted in a way that interferes with the identification of /r/ and /l/.

Despite the facts that early language exposure to Japanese altered adult Japanese speakers' perceptual space and made them insensitive to F3 frequency variations at the /r/-/l/ boundary, it may be possible that they reorganize their perceptual space and become sensitive to the F3 acoustic variations at the boundary after learning English. If /r/-/l/ discrimination sensitivity is related to /r/-/l/ production, Japanese speakers who have high F3 sensitivity at the /r/-/l/ category boundary will be good at /r/-/l/ production. Similarly, Japanese speakers who become less sensitive to F2 acoustic variations may be good at production. If /r/-/l/ discrimination sensitivity is related to other perceptual abilities, it seems likely that Japanese speakers who have high F3 sensitivity at the /r/-/l/ category boundary will be good at /r/-/l/ identification and/or have native-like mental representations.

The present study measured perceptual abilities implementing a discrimination task as well as a /r/-/l/ identification task and best exemplar search. Experiment 2.1 measured /r/-/l/ identification accuracy using the identification task from Chapter 2. Experiment 2.2 assessed Japanese speakers' /r/-/l/ discrimination sensitivity in terms of F2 and F3 (i.e., within-category and category-boundary sensitivity). Experiment 2.3 measured /r/-/l/ best exemplars of Japanese speakers using the best exemplar search from

the previous chapter. The present study also measured /r-/l/ production accuracy in terms of five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), as well as recognition accuracy and goodness ratings by English speakers. Experiment 2.4 assessed Japanese speakers' /r-/l/ production in words, sentences, and passages. The /r-/l/ word productions of Japanese speakers were acoustically analyzed in the five acoustic dimensions and compared to those of English speakers. The /r-/l/ production in sentences and passages were acoustically measured in terms of F3 and compared to the production of /r-/l/ words. The word production of Japanese speakers were identified and rated by English speakers using a forced-choice identification task and a goodness rating task. The aim was to determine whether there is a relationship between /r-/l/ perception and production using all of the perceptual and production measurements.

### ***3.2 Experiment 2.1: Phoneme identification***

#### **3.2.1 Method**

##### **Subjects**

Forty-eight native speakers of Japanese were tested in London. One participant was omitted from the data because she reported that she was a dyslexic, leaving 47 participants in total (see Appendixes C and D). Their ages ranged from 18 to 67 years (median = 31 years). They started learning English between 9 and 13 years (median = 13 years), and had received instruction for 5 - 12 years (median = 8 years). All participants were born and grown up in Japanese-speaking environments in Japan. They had lived in

English-speaking countries between 1 month to 17 years and 2 months (median = 12 months). None of the participants reported having hearing problems.

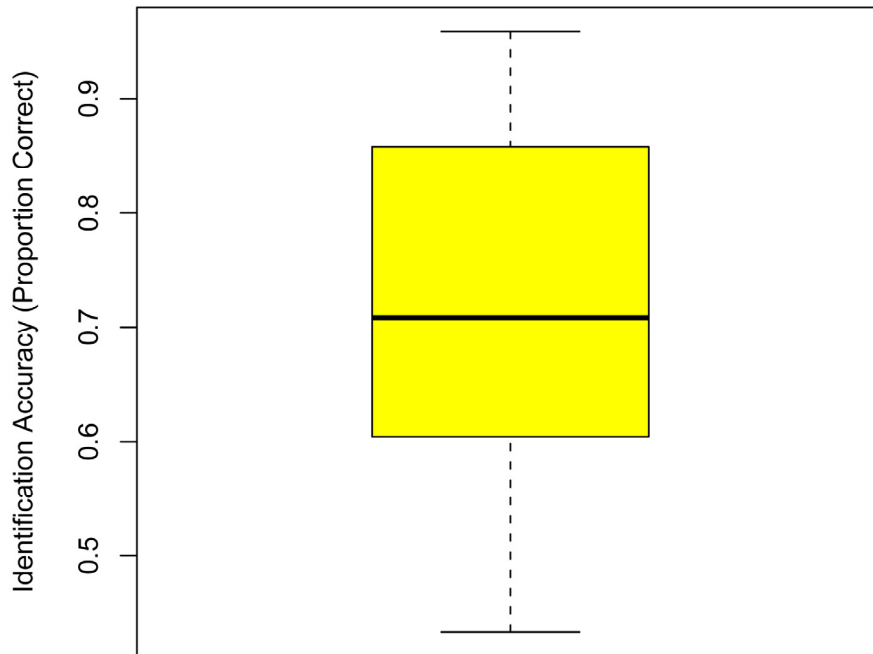
### Apparatus and stimuli

Same as Experiment 1.1 in Chapter 2.

### 3.2.2 Procedure

Same as Experiment 1.1 in Chapter 2.

### 3.2.3 Results



**Figure 3.1** Boxplot of English /r/-/l/ identification by Japanese speakers. Japanese speakers listened to naturally-spoken stimuli by British English speakers and identified /r/ and /l/ in word-initial position. The boxplot displays a median and the quartile ranges of scores. Standard error is 0.02. Japanese speakers demonstrated a wide range of /r/-/l/ identification accuracy.

Figure 3.1 displays Japanese speakers' /r/-/l/ identification accuracy. As found in the previous chapter, Japanese speakers demonstrated a wide range of identification

accuracy, ranging from 43% to 95% (mean = 71.46%). On average, they correctly identified English /r/ on 70.78 % of trials, and /l/ on 72.13 % of trials.

### **3.3 Experiment 2.2: Phoneme discrimination**

This experiment assessed the degrees of /r/-/l/ discrimination sensitivity of Japanese speakers in the F2 and F3 dimensions (i.e., to what extent Japanese speakers are sensitive to acoustic variations in the dimensions). The discrimination sensitivity in each dimension was assessed at three points; within the /r/ category, within the /l/ category, and at the /r/-/l/ category boundary. Pairs of synthetic stimuli (e.g., /ra/-/ra/, /la/-/la/, and /ra/-/la/) were generated based on the best exemplars of English speakers from Chapter 2. Japanese speakers listened to the pairs of synthetic stimuli in a same-different discrimination task. Their discrimination sensitivities were compared to those of English speakers.

#### **3.3.1 Method**

##### **Subjects**

The same Japanese speakers from Experiment 2.1 were tested in this experiment. Eight British English speakers were also tested. Their age ranged between 21 and 40 years (median = 26.5). They were all born and raised in southern England. They reported no hearing problems.

## Apparatus

All participants were tested in a sound-treated room. Stimuli were played to participants using Senheiser HD 280 headphones and Dell Optiplex GX 260. Responses were collected with Praat (Boersma and Weenik, 2009), which also controlled the presentation of stimuli.

## Stimuli

The stimuli were six pairs of /ra/ and /la/ tokens (i.e., F2 and F3 variation x locations within /r/ and /l/ and at the /r/-/l/ boundary). Four pairs of the stimuli (i.e., two pairs of /ra/-/ra/ and two pairs for /la/-/la/) were generated based on English speakers' /r/ and /l/ best exemplars in Chapter 2. For /ra/-/ra/ pairs, one pair varied only in the F2 dimension; one token had an F2 of 1051 Hz, and the other token had an F2 of 1358 Hz. The other pair varied only in the F3 dimension; one token had an F3 of 1739 Hz, and the other token had an F3 of 2212 Hz. When F2 frequencies varied, F3 frequency was set to 1739 Hz. Likewise, when F3 frequencies were varied, F2 frequency was set at 1196 Hz. All other acoustic parameters were identical; F1 = 327 Hz, closure duration = 31.36 ms, transition duration = 80.64 ms. For /la/-/la/ pairs, one pair varied only in the F2 dimension; one token had an F2 of 1051 Hz, and the other token had an F2 of 1358 Hz. The other pair varied only in F3 dimension; one token had an F3 of 3142 Hz, and the other token had an F3 of 3524 Hz. When F2 frequencies varied, F3 frequency was set to 3524 Hz. Likewise, when F3 frequencies varied, F2 frequency was set to 1196 Hz. All

other acoustic parameters were identical; F1 = 327 Hz, closure duration = 96 ms, transition duration = 16 ms.

The other two pairs of stimuli (i.e., /ra/-/la/) were generated based on an English speakers' /r/-/l/ boundary token. In order to find such a token, an identification task was run with seven native speakers of British English. Twenty-eight stimuli were generated by manipulating by F3 in seven steps (i.e., 1739, 1963, 2212, 2489, 2798, 3142, 3524 Hz) and, closure and transition duration in four levels (i.e., 31-66, 45-45, 66-31, 95-21 ms). All other acoustic parameters were identical; F1 = 312 Hz, and F2 = 1196 Hz. The results demonstrated that a boundary occurred when F3 was 2965 Hz, closure duration was 64 ms, and transition duration was 48 ms.

Based on the /r/-/l/ boundary token, two pairs of /ra/-/la/ stimuli were generated. One pair varied only in the F2 dimension; one token had an F2 of 1051 Hz, and the other token had an F2 of 1358 Hz. The other pair varied only in the F3 dimension; one token had an F3 of 2639 Hz, and the other token had an F3 of 3328 Hz. When F2 frequencies varied, F3 frequency was set to 2965 Hz. Likewise, when F3 frequencies varied, F2 frequency was set to 1196 Hz. All other acoustic parameters were identical; F1 = 327 Hz, closure duration = 64 ms, transition duration = 48 ms.

All pairs but one had a 2 ERB distance between the stimuli; each stimulus was 1 ERB distance away from the best exemplars of English speakers (i.e., either 1 ERB higher or lower from the best exemplars). For /la/-/la/ pair which varied in the F3 dimension, there was only a 1 ERB distance between the stimuli, because the /l/ best exemplar was close to the edge of the /l/ category; it was impossible to synthesize /l/

which is 1 ERB higher than the /l/ best exemplar. Therefore, we generated /l/ with the best exemplar F3 frequency and 1 ERB-lower F3 frequency (i.e., 3142 Hz).

### 3.3.2 Procedure

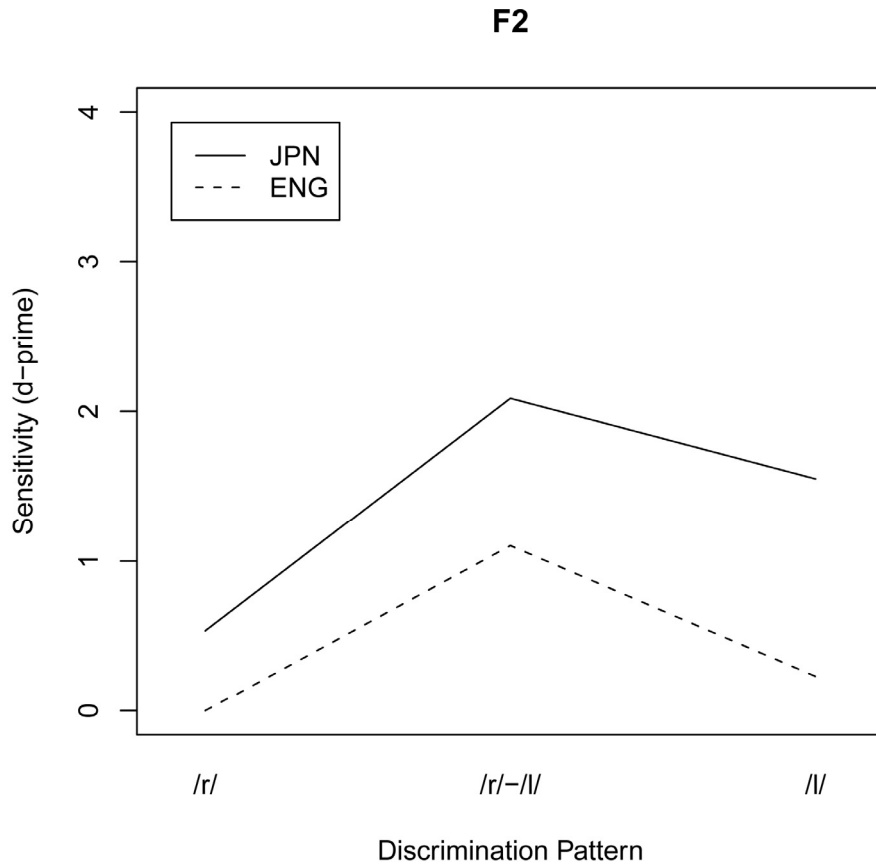
All participants heard the pairs of stimuli (i.e., /ra-/ra/, /ra-/la/, and /la-/la/) with a 300ms ISI and judged whether they were the same or different. Half were same pairs, containing two repetitions of the same stimulus. Half were different pairs, containing stimuli which varied in either the F2 or F3 dimension. All pairs were presented in all possible orders (i.e., AA, AB, BA, BB). Participants underwent a practice block of 24 trials (two same and two different for each pair) and an experimental block of 192 trials (i.e., 6 pairs x 4 orders x 8 repetitions = 192 trials). The results were analyzed using a differencing model of signal detection theory (Macmillan and Creelman, 1991) in order to calculate F2 and F3 sensitivity for each stimulus pair. Note that the /la-/la/ pair varied in the F3 dimension had only a 1 ERB distance between the stimuli, so the  $d'$  sensitivity was doubled in data analysis to make it comparable to the other pairs.

### 3.3.3 Results

Figure 3.2 displays Japanese and English speakers' discrimination sensitivity in the F2 dimension. Both Japanese and English speakers demonstrated a similar sensitivity pattern; they demonstrated higher F2 sensitivity at the English /r-/l/ category boundary than they did within the /r/ and /l/ categories. However, Japanese speakers demonstrated

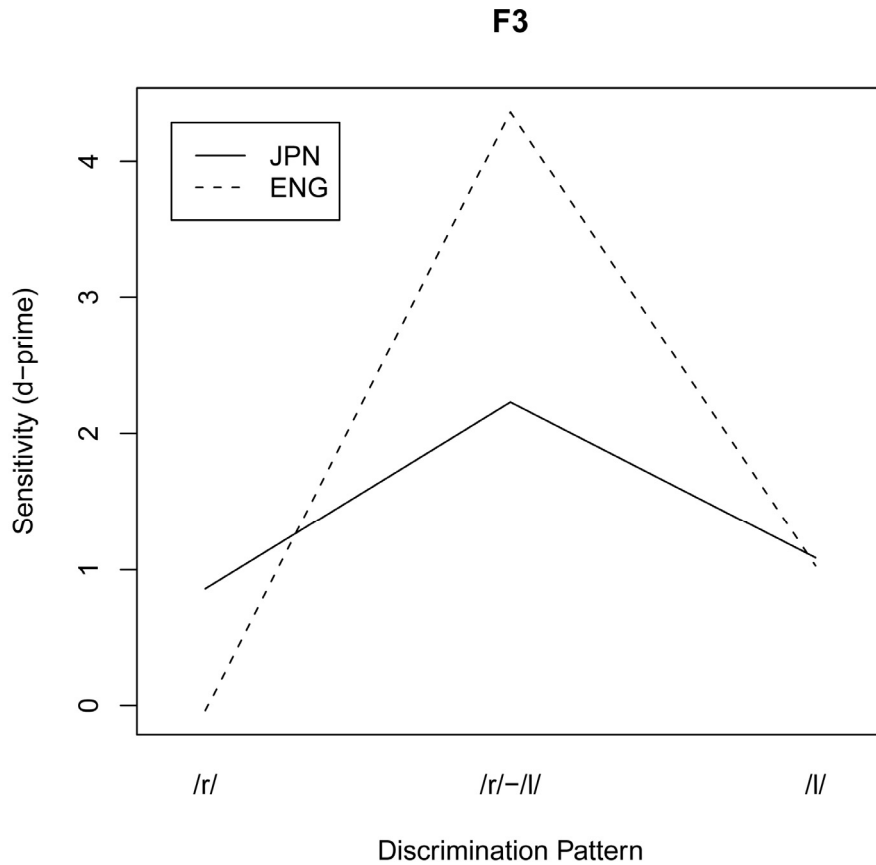


higher F2 sensitivity at the boundary than did English speakers. Similarly, Japanese speakers demonstrated higher within-category sensitivity than did English speakers. A 2-way ANOVA revealed that there was a main effect of language group,  $F(1,53) = 10.11, p < 0.01$ , confirming that Japanese speakers overall had higher F2 sensitivity. There was a main effect of discrimination pattern (i.e., within-/r/, within-/l/, and /r/-/l/ discrimination),  $F(2,106) = 20.34, p < 0.0001$ , indicating that F2 sensitivity differed among the pair



**Figure 3.2 F2 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers and English speakers generally demonstrated similar discrimination patterns. They both demonstrated higher discrimination sensitivity at the /r/-/l/ boundary compared to discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0, 0.65, and 0.46 (d') for English speakers and 0.16, 0.18, and 0.21(d') for Japanese speakers. The results demonstrated that Japanese speakers had higher overall sensitivity than English speakers did. The results also confirmed that discrimination sensitivity at the /r/-/l/ boundary was higher than within /r/ and /l/ categories.**

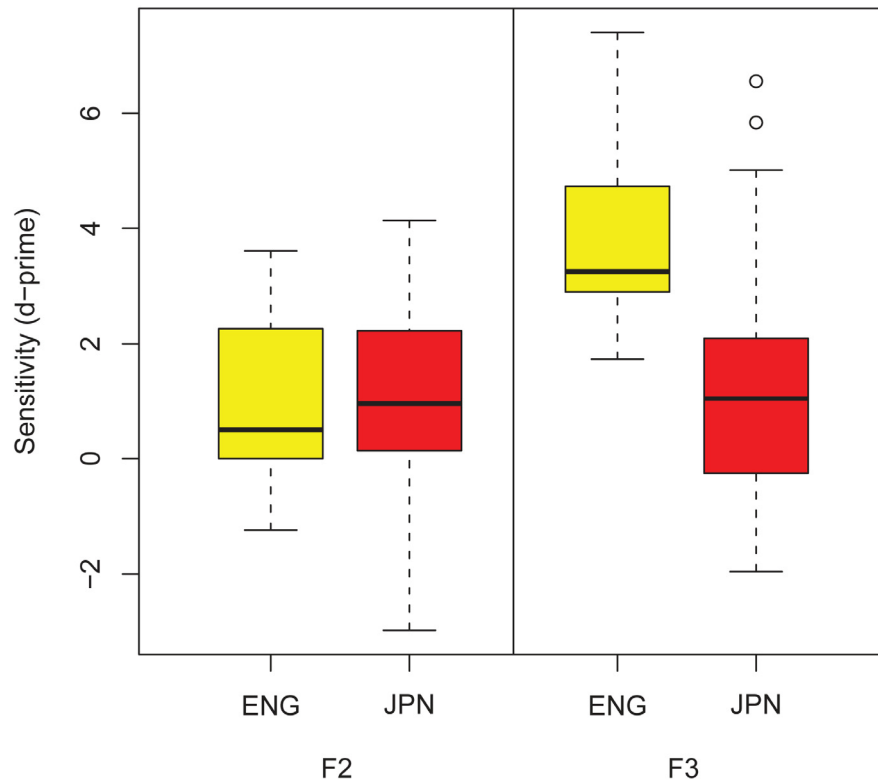
locations. There was no significant interaction,  $F(2,106) = 0.70, p > 0.05$ . Tukey HSD comparisons revealed that F2 sensitivity at the /r/-/l/ boundary was significantly different from F2 sensitivity within the /r/ and /l/ categories, /r/ and /r/-/l/ boundary,  $z = -6.35, p < 0.001$ , /l/ and /r/-/l/ boundary,  $z = -2.50, p < 0.05$ . The statistical analysis also revealed that F2 sensitivity within /r/ and /l/ categories was significantly different,  $z = -3.85, p < 0.001$ , indicating that within-/l/ category sensitivity was higher.



**Figure 3.3 F3 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers had lower discrimination sensitivity at the /r/-/l/ boundary than English speakers did. However, Japanese and English speakers demonstrated similar discrimination sensitivity within /r/ and /l/ categories. Standard errors from the left side of the figures are 0.48, 0.35, and 0.87 (d') for English speakers and 0.17, 0.23, and 0.40 (d') for Japanese speakers. The results confirmed that Japanese speakers were different from English speakers in terms of F3 discrimination sensitivity at the /r/-/l/ boundary; Japanese speakers were poorer at discriminating /r/ and /l/ at the boundary. The results also confirmed that Japanese and English speakers were similar in terms of F3 discrimination sensitivity within /r/ and /l/ categories; they all were poor at discriminating between /r/ exemplars and between /l/ exemplars.**

Figure 3.3 displays Japanese and English speakers' discrimination sensitivity in the F3 dimension. Both Japanese and English speakers demonstrated a similar sensitivity pattern; they demonstrated higher F3 sensitivity at the English /r/-/l/ category boundary than they did within /r/ and /l/ categories. However, English speakers clearly demonstrated higher sensitivity at the boundary. A 2-way ANOVA revealed that there was no main effect of language group (i.e., English or Japanese),  $F(1,53) = 0.74, p > 0.05$ . There was a main effect of discrimination pattern (i.e., within-/r/, within-/l/, or /r/-/l/ boundary discrimination),  $F(2,106) = 14.86, p < 0.0001$ , indicating that F3 sensitivity differed among the locations. Tukey HSD comparisons revealed that F3 sensitivity at the /r/-/l/ boundary was significantly different from F3 sensitivity within /r/ and /l/ categories, /r/ and /r/-/l/ boundary,  $z = -4.96, p < 0.001$ , /l/ and /r/-/l/ boundary,  $z = -4.01, p < 0.001$ . Within /r/ and /l/ category sensitivities were not different,  $z = -0.96, p > 0.05$ . There was a significant interaction between language group and discrimination patterns,  $F(2,106) = 4.90, p < 0.01$ . Simple effects analyses of the interaction revealed that the effect of language group for discrimination sensitivity at the /r/-/l/ boundary was significant,  $t(105) = -2.90, p < 0.01$ , suggesting that English speakers clearly had higher discrimination sensitivity at the boundary than Japanese speakers did, but the effects of language group for /r/,  $t(105) = 1.22, p > 0.05$ , and /l/,  $t(105) = 0.08, p > 0.05$ , were not significant. These results thus demonstrated that Japanese speakers specifically differed from English speakers in terms of F3 discrimination sensitivity at the /r/-/l/ boundary; they had lower F3 discrimination sensitivity at the boundary. However, Japanese speakers were similar to English speakers in terms of F3 discrimination within /r/ and /l/ categories.

Figure 3.4 displays the F2 peak discrimination sensitivity (i.e., F2 sensitivity which mean F2 sensitivity of /r/ and /l/ categories is subtracted from F2 sensitivity at the /r-/l/ boundary) of Japanese and English speakers. In order to examine whether Japanese speakers have higher discrimination sensitivity at the /r-/l/ boundary than do English speakers, a two-sample t-test was run. The statistical analysis revealed that there was no



**Figure 3.4 F2 and F3 peak discrimination sensitivity of English (ENG) and Japanese (JPN) speakers at the English /r-/l/ boundary. The peak discrimination sensitivity in each acoustic dimension was calculated by subtracting the mean discrimination sensitivity of /r/ and /l/ categories from the discrimination sensitivity at the English /r-/l/ boundary. Higher peak discrimination sensitivity indicates that English and Japanese speakers had high discrimination sensitivity at the /r-/l/ boundary and low discrimination sensitivity within /r/ and /l/ categories (i.e., they were good at discriminating /r/ and /l/ at the boundary, and they were poor at discriminating between /r/ exemplars and between /l/ exemplars). Japanese and English speakers both demonstrated similar peak discrimination sensitivity in the F2 dimension. However, Japanese speakers demonstrated lower peak discrimination sensitivity in the F3 dimension than English speakers did. Standard errors from the left side of the figures are respectively 0.57, 0.22, 0.62, and 0.28 (d'). The results revealed that Japanese speakers were similar to English speakers in terms of F2 peak discrimination sensitivity. The results also revealed that Japanese speakers were different from English speakers in terms of F3 peak discrimination sensitivity; Japanese speakers were poor at discriminating /r/ and /l/ at the boundary.**

significant difference in the F2 sensitivity,  $t(9.14) = 0.09, p > 0.05$ , suggesting that Japanese and English speakers had similar patterns of F2 discrimination sensitivity. Figure 3.4 also displays F3 peak discrimination sensitivity (i.e., F3 sensitivity which mean F3 sensitivity of /r/ and /l/ categories is subtracted from F3 sensitivity at the /r/-/l/ boundary) of Japanese and English speakers. It seems clear that English speakers had higher discrimination sensitivity at the /r/-/l/ boundary than did Japanese speakers. A two-sample t-test revealed that there was a significant difference in the F3 sensitivity,  $t(10.12) = -3.84, p < 0.01$ , confirming that English speakers had sharper F3 sensitivity at the English /r/-/l/ boundary than did Japanese speakers.

### **3.4 Experiment 2.3: Perceptual mapping of best exemplars**

Experiment 1.3 in Chapter 2 measured best exemplars of English /r/, /l/, and Japanese /r/ in English and Japanese carrier sentences. However, this experiment measured best exemplars of /r/ and /l/ only in English carrier sentence. Japanese speakers searched their best exemplars in the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration). Their best exemplars were compared to those of English speakers.

### **3.4.1 Method**

#### **Subject**

The Japanese speakers were the same as Experiment 2.1 and 2.2. The data of the 13 native speakers of British English from Chapter 2 was used in order to provide normative data.

#### **Apparatus and stimuli**

Same as Experiment 1.3 in Chapter 2.

#### **3.4.2 Procedure**

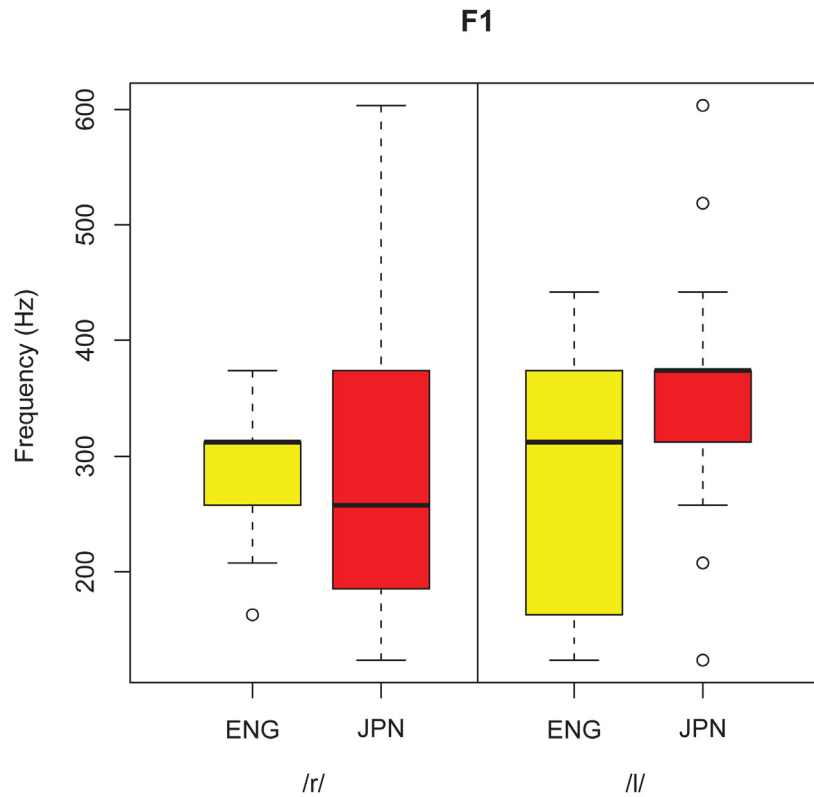
Same as Experiment 1.3 in Chapter 2.

#### **3.4.3 Results**

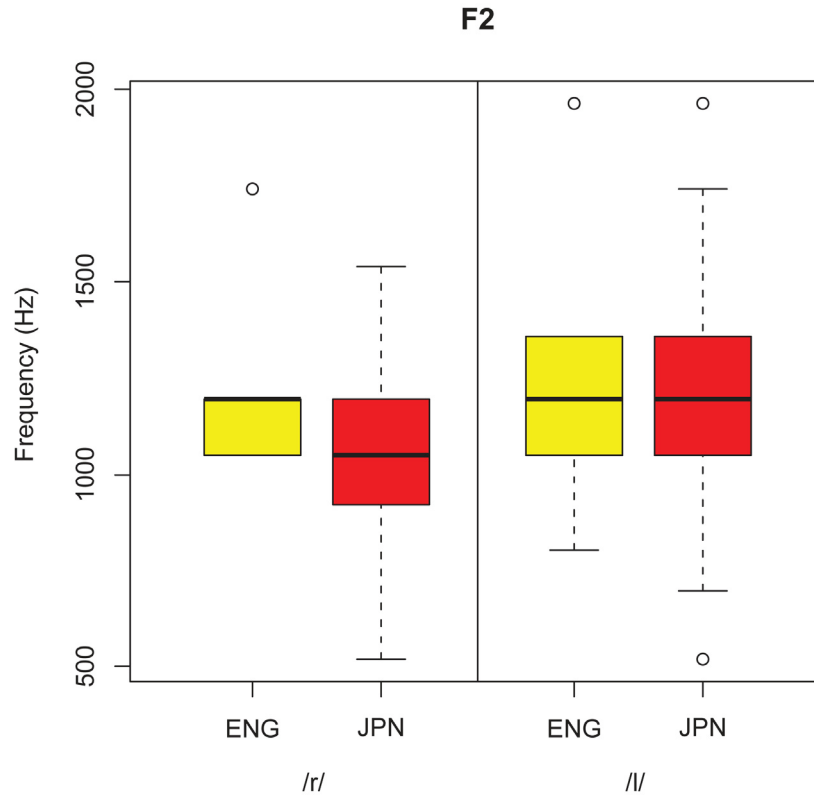
Figure 3.5 to 3.9 display perceptual best exemplars of English /r/ and /l/ by Japanese and English speakers. The best exemplars of Japanese speakers seem mostly similar to those of English speakers. For example, both language groups chose similar F1 (see Figure 3.5), F2 (see Figure 3.6), and closure duration (see Figure 3.8) for English /r/. Similarly, they chose similar F1, F2, and closure durations for English /l/. However, it is clear that Japanese speakers demonstrated some substantial variability. For example, they selected wide ranges of F3 frequencies for English /r/ and /l/ (see Figure 3.7). Similarly, they selected a wide range of transition durations for the consonants (see Figure 3.9).

Therefore, it is possible that Japanese speakers' best exemplars are different from English speaker's best exemplars. Two-way ANOVAs were separately run for each acoustic dimension to determine whether Japanese speaker's best exemplars were significantly

different from English speaker's best exemplars. The dependent variables were the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), and the between-subject factor was language group (Japanese or English), and the within-subject factor was consonant (/r/ or /l/). For F1, there was no main effect of language,  $F(1, 58) = 1.90, p > 0.05$ , but there was a main effect of consonant,  $F(1,58) = 9.22, p < 0.01$ , suggesting that /l/ F1 frequency was slightly higher than /r/ F1 frequency. There was no significant interaction,  $F(1,58) = 1.43, p > 0.05$ .



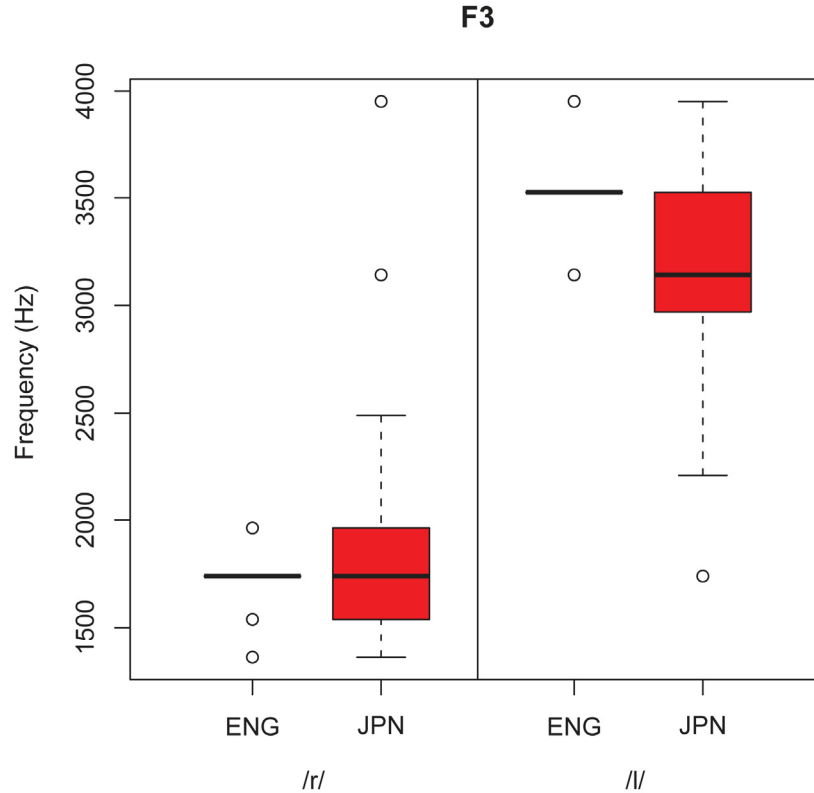
**Figure 3.5** Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). The best exemplars of English speakers came from Experiment 1.3 in Chapter 2. Specifically, their best exemplars which were measured using English carrier sentences are presented here. English speakers selected similar F1 frequencies for /r/ and /l/. Japanese speakers selected similar F1 frequencies for /r/ and /l/ with some wide variance for /r/. Standard errors from the left side of the figure are 16.15, 18.07, 31.98, and 13.76 (Hz). The results demonstrated that English and Japanese speakers were similar in F1 frequencies for /r/ and /l/.



**Figure 3.6** Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected similar F2 frequencies for /r/ and /l/. Japanese speakers chose similar F2 frequencies for /r/ and /l/ which were similar to those by English speakers. Standard errors from the left side of the figure are 49.31, 35.46, 94.97, and 44.01 (Hz). The results demonstrated that, although English speakers overall preferred slightly higher F2 frequencies than Japanese speakers, English and Japanese speakers were fundamentally similar in terms of F2 frequencies chosen for /r/ and /l/.

For F2, there was a main effect of language,  $F(1,58) = 4.62, p < 0.05$ , suggesting that English speakers preferred slightly higher F2 frequency. There was also a main effect of consonant,  $F(1,58) = 12.26, p < 0.001$ , suggesting that /l/ F2 frequency was slightly higher than /r/ F2 frequency. However, there was no significant interaction,  $F(1,58) = 1.27, p > 0.05$ .

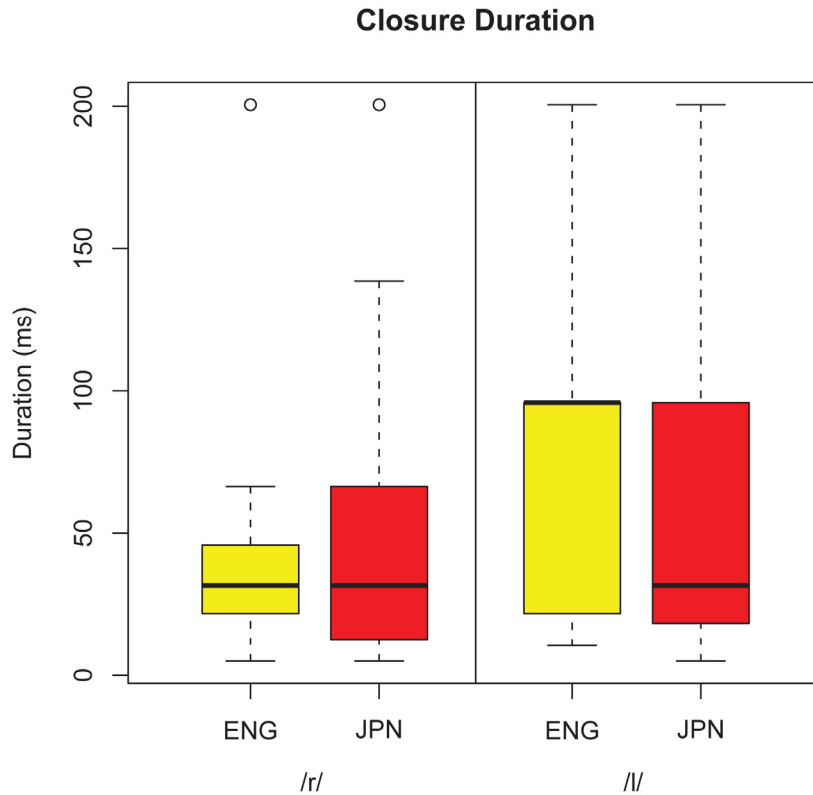




**Figure 3.7** Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected low F3 frequencies for /r/ and high F3 frequencies for /l/. Japanese speakers similarly selected low F3 frequencies for /r/ and high F3 frequencies for /l/ with some wide variance for both /r/ and /l/. Standard errors from the left side of the figure are 51.29, 72.31, 72.44, and 77.45 (Hz). The results demonstrated that English /r/ and /l/ are different from each other in the F3 dimension; /r/ has low F3 frequencies and /l/ has high F3 frequencies. The results also demonstrated that, although Japanese speakers seemed to select F3 frequencies similar to those by English speakers, Japanese speakers were specifically inaccurate in selecting F3 frequencies for /l/.

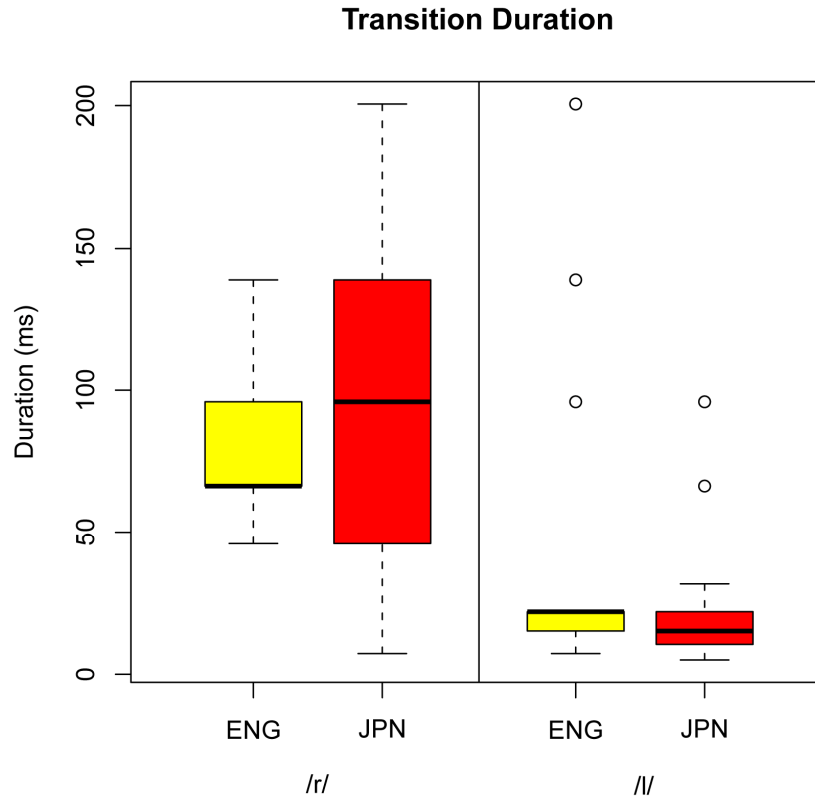
For F3, there was no main effect of language,  $F(1,58) = 0.65, p > 0.05$ . There was a main effect of consonant,  $F(1,58) = 289.18, p < 0.001$ , confirming that English /r/ and /l/ differ in the F3 dimension. There was a significant interaction between consonant and language group,  $F(1,58) = 6.78, p < 0.05$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was not significant,  $t(59) = 1.24, p > 0.05$ , suggesting that Japanese speakers chose F3 frequencies for /r/ similar to those by English speakers, but the effect of language group for /l/ was significant,  $t(59) = -2.40, p < 0.05$ ,

suggesting that English speakers systematically chose higher F3 for /l/ than did Japanese speakers. These results thus indicate that Japanese speakers did not have completely native-like representations particularly for /l/ in the F3 dimension.



**Figure 3.8** Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers selected shorter closure durations for /r/ with little variance and longer closure durations for /l/ with some variance. Japanese speakers selected similar closure durations for /r/ and /l/ with a wide range of variance. Standard errors from the left side of the figure are 13.91, 8.96, 17.97, and 7.02 (ms). The results demonstrated that English and Japanese speakers selected similar closure durations for both /r/ and /l/.

For closure duration, there were no main effects of language,  $F(1,58) = 0.60, p > 0.05$ , and consonant,  $F(1,58) = 1.50, p > 0.05$ . There was no significant interaction,  $F(1,58) = 2.91, p > 0.05$ .



**Figure 3.9** Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English (ENG) and Japanese (JPN). English speakers generally selected long transition durations for /r/ and short transition durations for /l/. Japanese speakers similarly selected long closure durations for /r/ with a wide range of variance, and short transition durations for /l/. Standard errors from the left side of the figure are 9.49, 8.39, 16.71, and 2.78 (ms). The results confirmed that English and Japanese speakers selected similar transition durations for both /r/ and /l/.

For transition duration, there was no main effect of language,  $F(1,58) = 0.78, p > 0.05$ .

There was a significant effect of consonant,  $F(1,58) = 79.78, p < 0.001$ . There was a significant interaction between consonant and language group,  $F(1,58) = 4.09, p < 0.05$ .

Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant,  $t(59) = 2.13, p < 0.05$ , suggesting that Japanese speakers chose transition durations for /r/ similar to those by English speakers, but the effect of language group for /l/ was not significant,  $t(59) = -1.12, p > 0.05$ , suggesting that English speakers systematically chose shorter transition durations for /r/ than did

Japanese speakers. These results thus indicate that Japanese speakers did not have completely native-like representations particularly for /r/ in transition duration. In short, the results demonstrated that Japanese speakers generally have native-like English /r/ and /l/ mental representations. But, they are not particularly native-like in the F3 dimension and transition duration.

### **3.5 Experiment 2.4: Production assessment**

This experiment assessed /r/-/l/ productions of Japanese and English speakers. Japanese speakers made recordings of initial-position /r/-/l/ minimal-pair words, accent-revealing sentences and passages. English speakers made recordings of such minimal pairs only. The recordings of /r/-/l/ words were measured in terms of five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration). The recordings of sentences and passages were measured in terms of F3. Japanese speaker's /r/-/l/ productions were analyzed in terms of (1) identification accuracy and goodness rating by English speakers, (2) the degree of production accuracy compared to English speakers in the five acoustic dimensions, and (3) the degree of /r/-/l/ F3 contrast in words, sentences, and passages. A relationship between /r/-/l/ production and perception was analyzed using production measurements (i.e., /r/-/l/ production identification by English speakers, goodness ratings, and production accuracy) and perceptual measurements from Experiment 2.1, 2.2, and 2.3 (i.e., /r/-/l/ identification accuracy, discrimination sensitivity, and best exemplar accuracy).

### 3.5.1 Method

#### Subjects

The Japanese speakers were the same as in Experiments 2.1, 2.2, and 2.3. Fifteen native speakers of British English were additionally tested in order to provide normative data.

#### Apparatus and stimuli

Stimulus recordings were made in a quiet sound-proof room (16-bit depth; 44,100 samples/sec). The stimuli were recorded with Radio Spares (RS) 249-946 microphone, Edirol USB Audio Capture UA-25, and a Dell Optiplex GX 260.

All Japanese speakers were recorded reading 19 initial-position /r/-/l/ minimal pairs (i.e., 38 words x 47 participants = 1786 stimuli), the accent-revealing sentence “*The red robin looked across the lovely lake*”, and a passage (i.e., the rainbow passage). Due to computer problems, four passage-recordings were missing; leaving a total of 43 passage-recordings. All participants read the materials with the experimenter before the recordings in order to make sure that they knew how to pronounce all words.

The British English recordings were from Iverson et al. (2005). Eleven speakers made recordings of 100 initial-position /r/-/l/ words (i.e., 1,100 stimuli). Four speakers made recordings of 193 words including the 38 words that Japanese speakers read (i.e., 772 stimuli). There was one sound file missing; leaving a total of 1871 stimuli (i.e., 1,100 + 772 – 1 = 1871 stimuli).

### 3.5.2 Procedure

#### Word /r/-/l/ acoustic measurements

Japanese and English speakers' /r/-/l/ productions were measured in five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration) using Praat (Boersma and Weenik, 2009). The author checked each /r/ and /l/ production and found it was hard to measure F3 frequencies in some of the productions from Japanese speakers due to a flap articulation. Such /r/ and /l/ productions were omitted; 22 out of 1786 stimuli were omitted from the analysis. Closure duration was defined as from the point where voicing began to the point where F1 transition began to happen. Transition duration was from the point where F1 transition began to either 1) the point where F1 appeared at its zenith (i.e., where F1 transition finished) or 2) the point where F3 appeared at its zenith (i.e., where F3 transition finished). Once the durations were fixed, Praat automatically calculated average F1, F2, and F3 in closure duration, and saved the spectral and temporal information in a text file. For each Japanese and English individual, F1, F2, F3, closure and transition durations were separately averaged across words for English /r/ and /l/.

In order to talker-normalize the spectral information (i.e., F1, F2, and F3), each individual's median F3 frequency across a large speech sample was measured using Praat (Boersma and Weenik, 2009). The author acoustically analyzed some points of the large speech sample by adjusting number of formants and maximum frequency. However, even such careful acoustic analyses could not completely track down formants through the entire speech sample; formant tracking was less consistent and reliable. This is because

there are some outliers due to noise. Therefore, it seemed that median F3 frequency would be more reliable than mean F3 frequency for the normalization process. Then, the median F3 frequency of each individual was subtracted from his/her mean F1, F2, and F3 frequencies of English /r-/l/ word recordings. The author acoustically analyzed each word production by adjusting number of formants and maximum frequency. Such small sound files rarely had outliers; formant tracking was consistent and reliable. Therefore, it seemed reasonable to take mean values rather than median values. This normalization process was done in order to remove between-talker differences (e.g., vocal tract length). For most Japanese speakers, median F3 frequency was measured using the rainbow passage recordings. Four Japanese speakers' median F3 frequencies were measured using the accent-revealing sentence recordings because their passage recordings were missing. For English speakers, median F3 frequencies were measured in a similar manner using the word recordings concatenated together.

### **Sentence and passage /r-/l/ F3 measurement**

Japanese speakers' English /r/ and /l/ F3 frequencies were measured in the accent-revealing sentence and the rainbow passage recordings using Praat (Boersma and Weenik, 2009). For the accent revealing sentence recordings, three /r/s (i.e., *red*, *robin*, *across*) and four /l/s (i.e., *looked*, *lovely*, *lake*) were measured (i.e., 7 stimuli x 47 participants = 329 stimuli). When stimuli did not have a stable F3 due to subjects producing some Japanese flaps, these were omitted from the acoustic measurement; 18 out of 329 stimuli (i.e., 11 /r/s and 7 /l/s) were omitted from the analysis. For each Japanese speaker, F3 frequencies were averaged across words for English /r/ and /l/, respectively. Each

speaker's median F3 was subtracted from his/her averaged English /r/ and /l/ F3 frequencies for normalization.

For rainbow passage recordings, seven /r/s (i.e., *raindrops*, *reach*, *round*, and *rainbow* x 4) and six /l/s (i.e., *legend*, *light*, *long*, *look*, *looking*, and *looks*) were measured (i.e., 13 stimuli x 43 participants = 559 stimuli). Twenty-three out of 559 stimuli (i.e., 12 /r/s and 11 /l/s) were omitted from the analysis because F3 could not be measured. For each Japanese speaker, F3 frequencies were averaged across words for English /r/ and /l/, respectively. Each speaker's median F3 was subtracted from his/her averaged English /r/ and /l/ F3 frequencies for normalization.

### **Identification and goodness rating**

Five native speakers of British English listened to the minimal-pair words spoken by Japanese speakers, chose consonant categories (i.e., /r/, /l/, /w/, /d/) and gave goodness ratings. The ratings were made on an integer scale from 1 (poor) to 7 (good). The raters underwent a practice block of 20 trials with randomly selected stimuli, and then completed 19 experimental blocks (i.e., 98 words x 19 blocks = 1786 trials). Each block had one minimal pair. The raters were allowed to repeatedly listen to each stimulus. The rating scores were collected only if raters correctly identified English /r/ and /l/. Rating scores for each stimulus were averaged across the raters after confirming high inter-rater reliability ( $r = 0.86$ ).



### 3.5.4 Results

#### English /r/-/l/ production intelligibility and goodness by English speakers

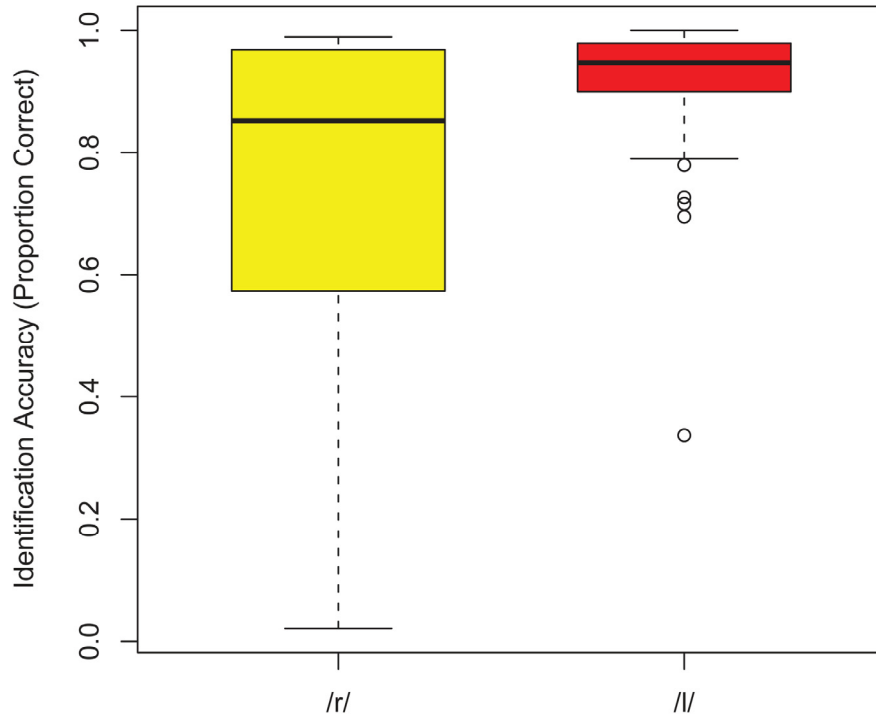
Table 3.1 displays the mean confusion matrix for the English /r/-/l/ production intelligibility task. Japanese speakers seemed to produce identifiable /l/; English listeners misidentified /l/ as /r/ in 6.43 % of the trials, /l/ as /d/ in 2.19 % of the trials, and /l/ as /w/ in 0.20 % of the trials. However, Japanese speakers seemed to have difficulties in producing /r/; English listeners misidentified /r/ as /l/ in 22.98% of the trials, /r/ as /d/ in 1.05 % of the trials, /r/ as /w/ in 2.80 % of the trials.

**Table 3.1 Confusion matrix for /r/-/l/ production identification. British English listeners identified most of the /l/ productions by Japanese speakers, suggesting that Japanese speakers produced identifiable /l/. However, the listeners misidentified some of the /r/ productions by Japanese speakers as /l/, suggesting that Japanese speakers had some difficulties in producing English /r/.**

Stimulus	Response			
	/l/	/r/	/d/	/w/
/l/	91.18%	6.43%	2.19%	0.20%
/r/	22.98%	73.16%	1.05%	2.80%

Figure 3.10 displays the English /r/-/l/ production intelligibility scores. Japanese speakers were generally homogeneous in the ability to produce English /l/; English speakers recognized most of the Japanese speakers' /l/ productions (mean = 91.18 %). But, Japanese speakers substantially varied in the ability to produce English /r/; English speakers recognized Japanese speakers' English /r/ productions with a wide range of scores (mean = 73.16 %), ranging from near 0% to near 100% correct. A two-sample t-test revealed that there was a significant difference between English /r/ and /l/ recognition

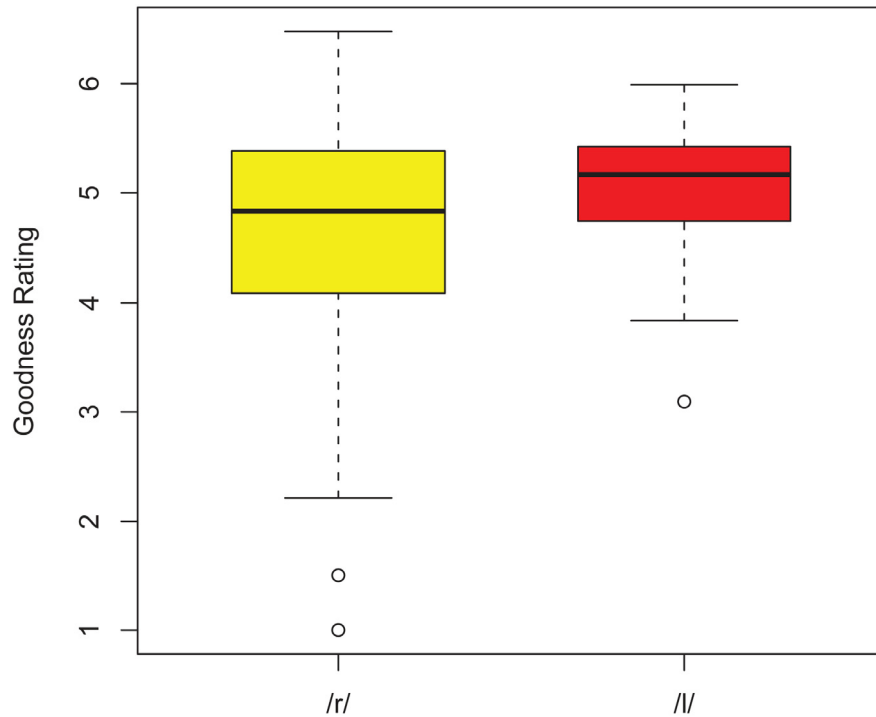
accuracy by English speakers,  $t(60.32) = 3.92, p < 0.001$ , indicating that Japanese speakers were more consistently able to produce English /l/ rather than English /r/.



**Figure 3.10** Boxplots of English /r/ and /l/ production intelligibility. Native speakers of British English listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or /d/. Higher proportion correct indicates that English speakers correctly identified /r/ and /l/ productions of L1 Japanese speakers. The intelligibility of /l/ productions varied little; most of the Japanese speakers produced identifiable /l/. However, the intelligibility of /r/ productions widely varied; some Japanese speakers produced identifiable /r/, but some others did not. Standard errors from the left side of the figure are 0.04 and 0.02. The results demonstrated that there was a significant difference between the /r/ and /l/ production intelligibilities; English speakers identified /l/ productions better than /r/ productions.

English /r/-/l/ production goodness rating scores in Figure 3.11 similarly demonstrated that Japanese speakers' /l/ productions received higher rating scores (mean = 5.02), with less variance. But, Japanese speakers' /r/ productions received substantially varied rating scores (mean = 4.61), ranging from one to near seven. A two-sample t-test revealed that there was a significant difference between English /r/ and /l/ goodness

rating scores,  $t(70.59) = -2.07, p < 0.05$ , indicating that English speakers perceived Japanese speakers' /l/ productions better than did they Japanese speakers' /r/ productions.

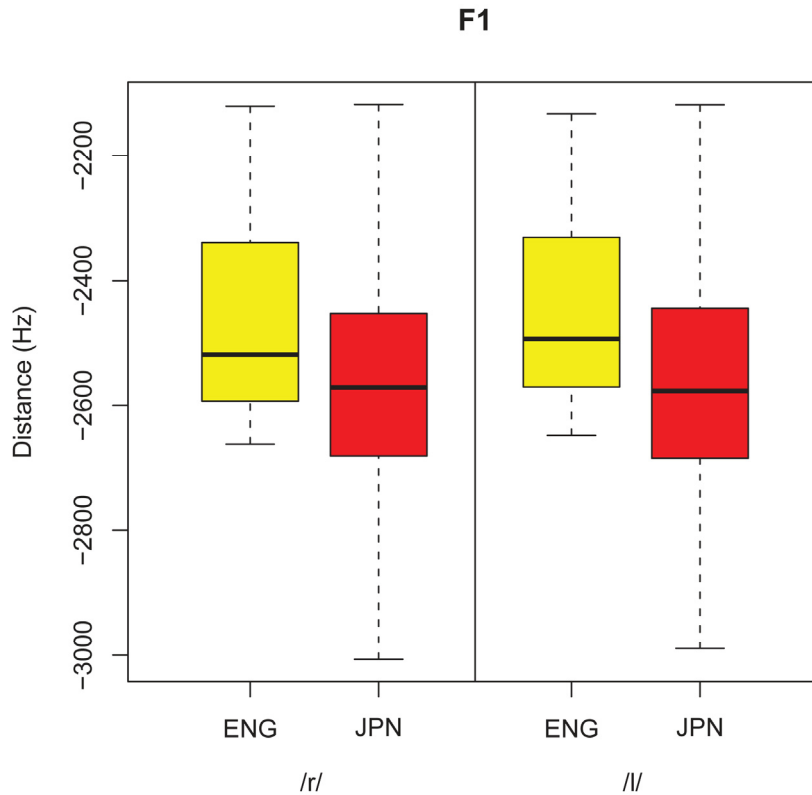


**Figure 3.11** Boxplots of the goodness rating scores for /r/ and /l/ productions of Japanese speakers. British English speakers gave ratings to 19 initial-position /r-/l/ minimal pairs produced by native speakers of Japanese with a scale from 1 (poor) to 7 (good). The intelligibility scores for /l/ were more or less similar among Japanese speakers, but the intelligibility scores for /r/ substantially varied among Japanese speakers. Standard errors from the left side of the figure are 0.17 and 0.09. The results demonstrated that English speakers perceived /l/ productions better than /r/ productions.

### English /r-/l/ production comparison between Japanese and English speakers

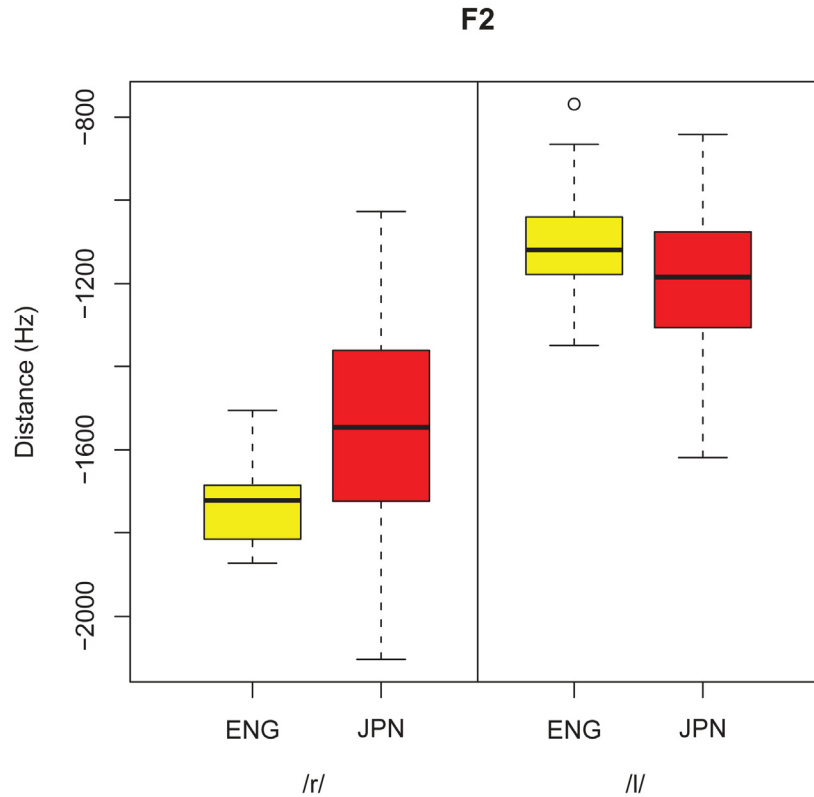
Since English speakers demonstrated such substantial variability in /r-/l/ intelligibility scores and goodness ratings, it is possible that Japanese speakers' English /r-/l/ productions, particularly English /r/ productions, were acoustically varied. Figure 3.12-

3.16 display /r-/l/ productions of English and Japanese speakers in five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration). In order to examine whether Japanese speakers' English /r/ and /l/ productions were different from those of English speakers, 2-way ANOVAs were separately run for each acoustic dimension. The dependent variables were the five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration), the between-subject factor was language group (Japanese or English), and the within-subject factor was consonant (/r/ or /l/).



**Figure 3.12** Boxplots of F1 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers produced similar F1 frequencies for both /r/ and /l/. Likewise, Japanese speakers produced similar F1 frequencies for both /r/ and /l/. Standard errors from the left side of the figure are 44.02, 26.66, 41.39, and 25.88 (Hz). The results demonstrated that Japanese speakers overall produced lower F1 frequencies than English speakers did. Japanese speakers specifically differed from English speakers in producing F1 frequencies for /r/; they produced lower F1 frequencies for /r/ than English speakers did. However, Japanese speakers did not produce abnormal F1 frequencies. Therefore, it is fair to assume that Japanese speakers are fundamentally similar to English speakers in producing F1 frequencies for /r/.

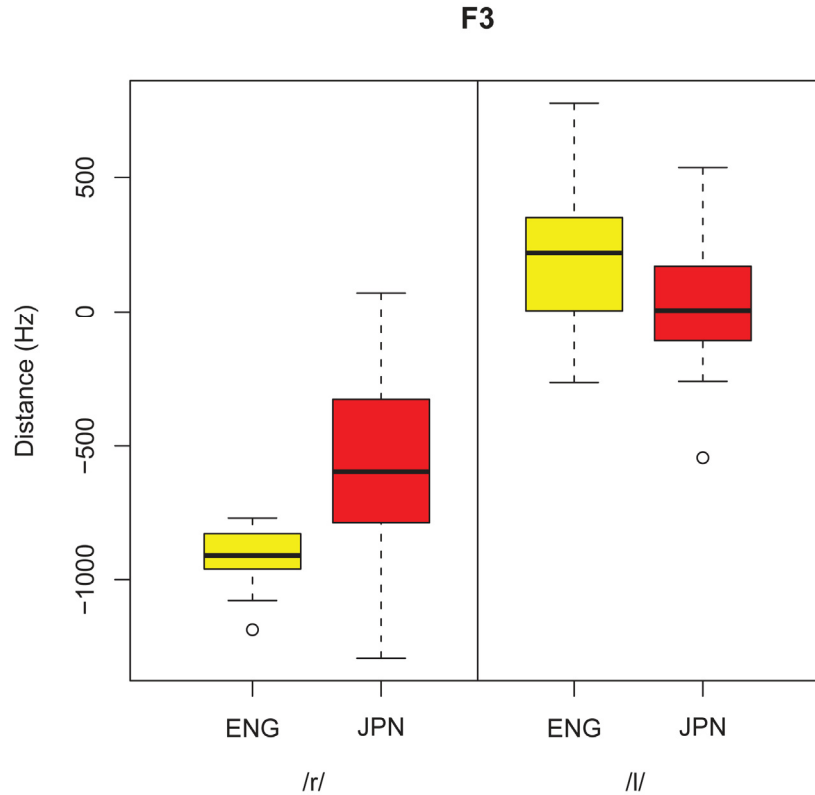
For F1, there was a main effect of language,  $F(1,60) = 4.44, p < 0.05$ , suggesting that Japanese speakers produced lower F1 than did English speakers. There was no main effect of consonant,  $F(1,60) = 0.19, p > 0.05$ . There was a significant interaction between language group and consonant,  $F(1,60) = 5.41, p < 0.05$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant,  $t(59) = 2.39, p < 0.05$ , suggesting that Japanese speakers produced lower F1 for /r/ than did English speakers, but the effect of language group for /l/ was not significant,  $t(59) = 0.38, p > 0.05$ , suggesting that Japanese speakers were similar to English speakers in producing F1 frequencies for /l/. These analyses indicate that Japanese speakers differed from English speakers in producing F1 frequencies for /r/. However, F1 frequencies of Japanese speakers were not substantially different from those of English speakers, and it is fair to consider that Japanese and English speakers are fundamentally similar in producing F1 frequencies for /r/. For F2, there was no main effect of language,  $F(1,60) = 0.69, p > 0.05$ . There was a main effect of consonant,  $F(1,60) = 200.08, p < 0.001$ , confirming that English /r/ and /l/ differed in the F2 dimension. There was a significant interaction,  $F(1,60) = 18.28, p < 0.001$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant,  $t(59) = 4.83, p < 0.001$ , suggesting that Japanese speakers were inaccurate in producing F2 frequencies; they produced higher F2 for /r/ than did English speakers. However, the effect of language group for /l/ was not significant,  $t(59) = 0.39, p > 0.05$ , suggesting that Japanese speakers were similar to English speakers in producing F2 frequencies for /l/. These results thus indicate that Japanese speakers were inaccurate in producing F2 frequencies specifically for English /r/.



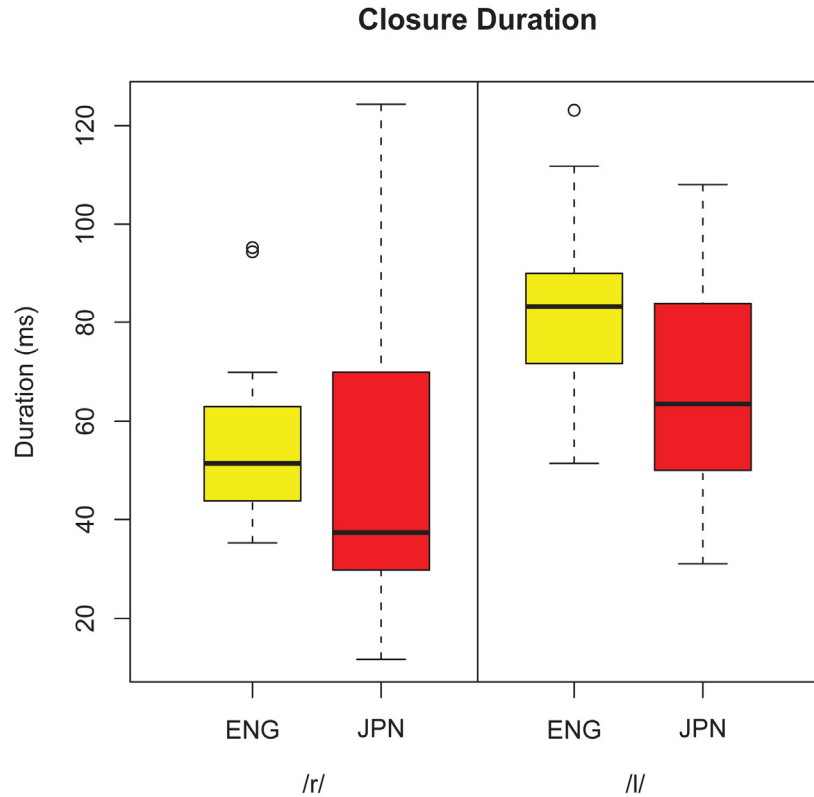
**Figure 3.13** Boxplots for F2 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers clearly made distinctions between /r/ and /l/; they produced low F2 frequencies for /r/ and high F2 frequencies for /l/. However, Japanese speakers made less clear distinctions between /r/ and /l/. Standard errors from the left side of the figure are 27.10, 37.63, 40.66, and 26.70 (Hz). The results demonstrated that Japanese speakers were similar to English speakers in producing F2 frequencies for /l/. However, they differed from English speakers in producing F2 frequencies for /r/ (i.e., Japanese speakers were inaccurate in producing /r/).

For F3, there was no main effect of language,  $F(1,60) = 2.54$ ,  $p > 0.05$ . There was a main effect of consonant,  $F(1,60) = 285.80$ ,  $p < 0.001$ , confirming that English /r/ and /l/ differed in the F3 dimension. There was a significant interaction,  $F(1,60) = 26.75$ ,  $p < 0.001$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant,  $t(59) = 5.02$ ,  $p < 0.001$ , suggesting that Japanese speakers were inaccurate in producing F3 frequencies; they produced higher F3 for /r/ than did English speakers. However, the effect of language group for /l/ was not significant,  $t(59) = -0.37$ ,  $p > 0.05$ , suggesting that Japanese speakers were similar to

English speakers in producing F3 frequencies for /l/. These results thus indicate that Japanese speakers were inaccurate in producing F3 frequencies specifically for English /r/.



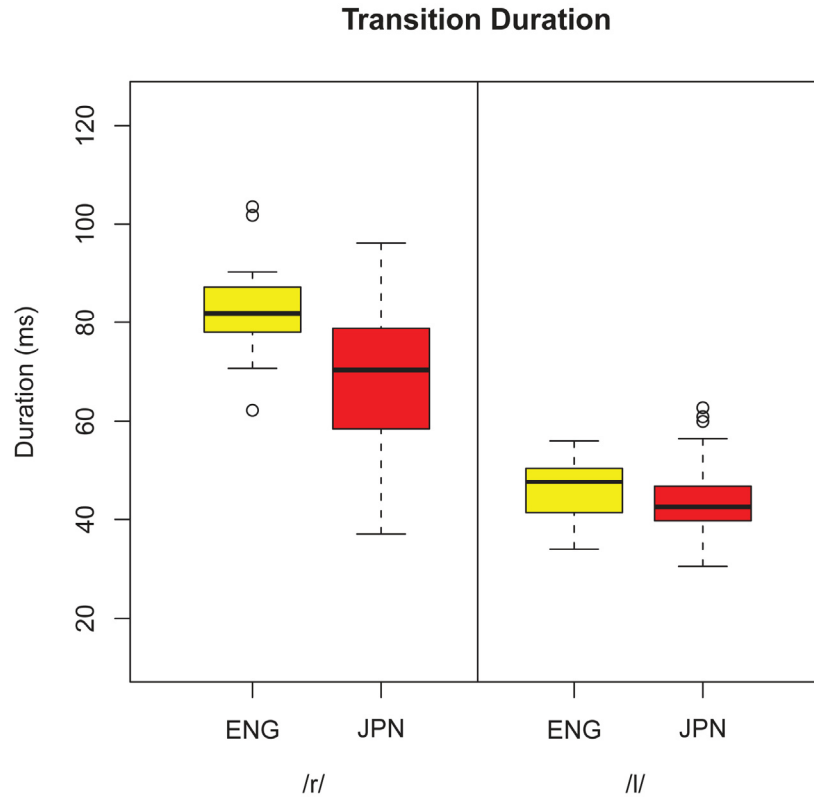
**Figure 3.14** Boxplots for F3 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers made clear distinctions between /r/ and /l/; they produced low F3 frequencies for /r/ and high F3 frequencies for /l/. However, Japanese speakers made less clear distinctions between the consonants. They produced F3 frequencies for /l/ similar to those of English speakers, but they produced a wide range of F3 frequencies for /r/ compared to English speakers. Standard errors from the left side of the figure are 29.32, 46.07, 74.76, and 33.61 (Hz). The results confirmed that Japanese speakers were similar to English speakers in producing F3 frequencies for /l/, but they differed from English speakers in producing F3 frequencies for /r/ (i.e., they were inaccurate in producing the frequencies for /r/).



**Figure 3.15** Boxplots for closure durations of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers generally produced shorter closure durations for /r/ and longer closure durations for /l/. Japanese speakers produced short closure durations for /r/ and long closure durations for /l/ with a wide range of variance. Standard errors from the left side of the figure are 4.72, 3.70, 5.28, and 3.17 (ms). The results demonstrated, despite such a wide range of variance, Japanese speakers were similar to English speakers in producing closure durations for /r/ and /l/.

For closure duration, there was nearly a main effect of language,  $F(1,60) = 3.9968$ ,  $p = 0.05012$ . There was a main effect of consonant,  $F(1,60) = 70.50$ ,  $p < 0.001$ , confirming that English /r/ closure duration was shorter than English /l/ closure duration. There was no significant interaction,  $F(1,60) = 2.43$ ,  $p > 0.05$ .





**Figure 3.16** Boxplots for transition durations of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). English speakers made clear distinctions between /r/ and /l/ (i.e., they produced longer transition durations for /r/ and shorter transition durations for /l/). Japanese speakers produced shorter transition durations for /l/ similar to those of English speakers. However, they produced slightly shorter transition durations for /r/ compared to English speakers. Standard errors from the left side of the figure are 2.72, 2.24, 1.66, and 1.07 (ms). The results demonstrated that Japanese speakers were similar to English speakers in producing transition durations for /l/. However, Japanese speakers produced shorter transition durations for /r/ than English speakers did.

For transition duration, there was a main effect of language,  $F(1,60) = 4.44$ ,  $p < 0.05$ .

There was no main effect of consonant,  $F(1,60) = 0.19$ ,  $p > 0.05$ . There was a significant interaction,  $F(1,60) = 5.41$ ,  $p < 0.05$ . Simple effects analyses of the interaction revealed that the effect of language group for English /r/ was significant,  $t(59) = -4.34$ ,  $p < 0.001$ , suggesting that that Japanese speakers were inaccurate in producing transition durations; they produced shorter transition durations for /r/ than did English speakers. However, the effect of language group for /l/ was not significant,  $t(59) = -0.55$ ,  $p > 0.05$ , suggesting

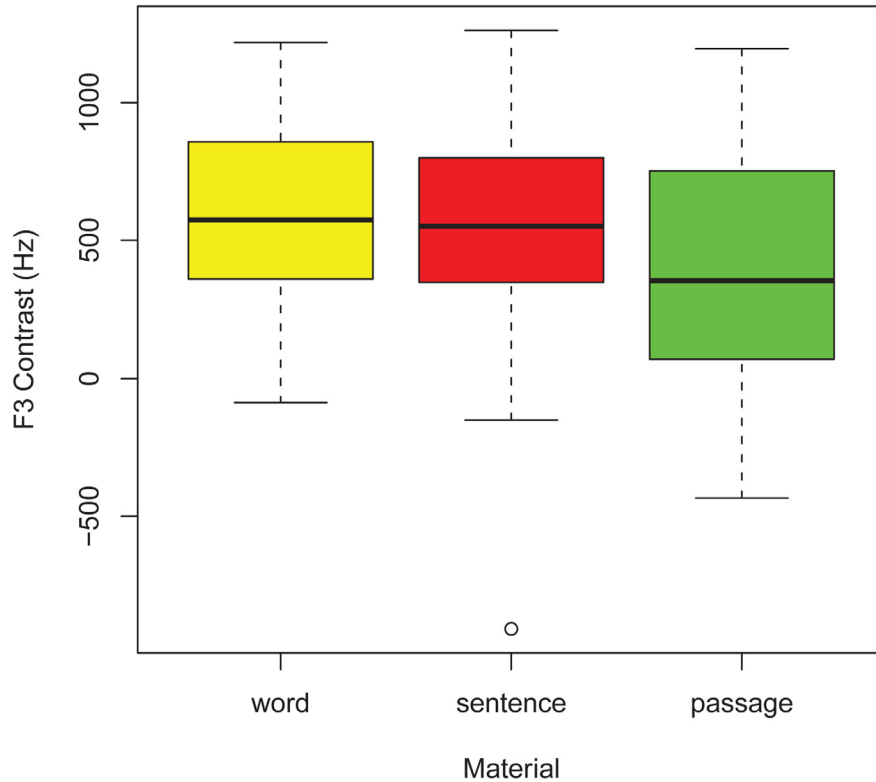
that Japanese speakers were similar to English speakers in producing transition durations for /l/. These results thus indicate that Japanese speakers were inaccurate in producing transition durations specifically for English /r/.

### **English /r/-/l/ F3 contrast in word, sentence, and passage**

The results thus demonstrated that Japanese speakers did not produce clear English /r/-/l/ distinctions in F2, F3, and transition duration. In order to further examine Japanese speakers' English /r/-/l/ productions, their productions in different materials (i.e., words, sentences, and passages) were analyzed in terms of F3 frequencies. Specifically, the degree of English /r/-/l/ contrast in F3 was examined across the materials.

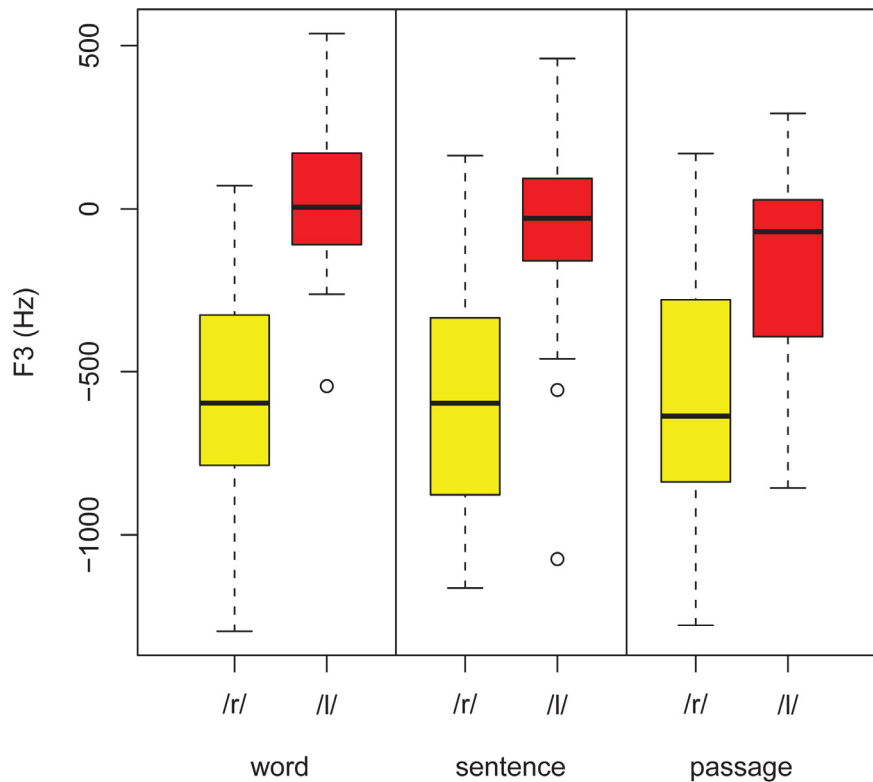
Figure 3.17 displays the degree of English /r/-/l/ F3 contrast in three different materials. It seems that Japanese speakers made similar degree of English /r/-/l/ contrast in words and sentences. But, they made a less clear contrast in passages. An ANOVA was run to determine whether Japanese speakers made a different degree of English /r/-/l/ contrasts across the materials. The dependent variables were distances between /l/ and /r/ in the F3 dimension, and material (words, sentences, or passages) was within-subject factor. The statistical analysis revealed that there was a main effect of material,  $F(2,88) = 9.07, p < 0.001$ , suggesting that Japanese speakers had a hard time to produce English /r/-/l/ contrast in passages. Tukey HSD comparisons revealed that the English /r/-/l/ contrast in the passages was significantly different from the contrast in words,  $z = 4.03, p < 0.001$ , and sentences,  $z = 3.30, p < 0.01$ , demonstrating that Japanese speakers had the hardest

time to produce English /r-/l/ contrast in passages. There was no significant difference in English /r-/l/ contrast between words and sentences,  $p > 0.05$ .



**Figure 3.17** The degree of English /r-/l/ F3 contrast in words and passages. /r/ and /l/ productions of Japanese speakers were measured in terms of F3 frequencies and normalized by subtracting median F3 frequencies which were measured using the passage recordings. The F3 contrasts were calculated by subtracting normalized F3 frequencies for /r/ from normalized F3 frequencies for /l/. Bigger positive contrast indicates clearer F3 contrast between /r/ and /l/. Standard errors from the left side of the figure are 52.24, 57.27, and 63.55 (Hz). The results demonstrated that Japanese speakers made similar /r-/l/ F3 contrasts in words and sentences. However, they made less clear contrasts in passages.

Figure 3.18 displays the F3 frequencies separately for /r/ and /l/ in words, sentences, and passages. Japanese speakers demonstrated similar F3 frequencies for English /r/ across the three kinds of material. However, they demonstrated different F3 frequencies for English /l/ across the materials; Japanese speakers seemed to produce poor /l/ in passages, compared to /l/ in words and sentences. One-way ANOVAs were



**Figure 3.18** F3 frequencies of English /r/ and /l/ productions by Japanese speakers in three different materials (i.e., words, sentences, passages). F3 frequencies of /r/ and /l/ productions were normalized by subtracting median F3 frequencies which were measured with the passage recordings. Japanese speakers produced similar F3 frequencies for /r/ in all of the materials. However, Japanese speakers had some difficulties in producing /l/. They produced similar F3 frequencies for /l/ in words and sentences, but they had hard time to produce /l/ in passages. Standard errors from the left side of the figure are 46.07, 33.61, 50.68, 38.69, 56.71, and 45.76 (Hz). The results confirmed that /r/ productions of Japanese speakers were similar among the three materials. However, /l/ productions in passages were different from the ones in words and sentences.

separately run for each consonant. For English /r/, there was no main effect of material,  $F(2,88) = 1.12, p > 0.05$ , confirming that Japanese speakers produced similar F3 frequencies for English /r/ across the materials. Tukey HSD comparisons also confirmed that there were no significant differences in English /r/ productions across material, words – sentences,  $z = 1.26$ , words – passages,  $z = -0.10$ , sentences – passages,  $z = -1.32$ ,  $p > 0.05$ . For English /l/, the ANOVA analysis revealed that there was a main effect of material,  $F(2,88) = 11.50, p < 0.0001$ , suggesting that Japanese speakers had difficulties

producing /l/ in passages. Tukey HSD comparisons revealed that English /l/ productions in passages were significantly different from /l/ productions in words and sentences, passages – words,  $z = 4.79, p < 0.001$ , passages – sentences,  $z = 2.66, p < 0.05$ . There was no significant difference between words and sentences /l/ productions, words – sentences,  $z = 2.20, p > 0.05$ .

### **Links between English /r/-/l/ production and perception**

The perceptual and production data revealed that adult Japanese individuals varied in /r/-/l/ perception and production. For example, there are wide ranges of variability in /r/-/l/ identification (see Figure 3.1), F3 discrimination sensitivity at the /r/-/l/ boundary (see Figure 3.4), and mental representations of the F3 dimension (see Figure 3.7). Likewise, there are wide ranges of variability in the degree to which native speakers of British English identified and rated /r/ production of Japanese speakers (see Figure 3.10 and Figure 3.11), and the degree to which /r/-/l/ productions of Japanese speakers are similar to English speakers in the F3 dimension (see Figure 3.14). Considering such individual differences, it is possible, for example, that Japanese speakers who have good /r/-/l/ identification accuracy are good at producing the consonants; there may be a correlation between /r/-/l/ identification accuracy and the degree in which English speakers identify /r/-/l/ productions of the Japanese speakers. In order to examine such a relationship between /r/-/l/ perception and production, Pearson correlational analyses were run using the individual differences in perceptual and production data.

Table 3.2 displays the correlations between /r-/l/ identification (i.e., how Japanese speakers identified /r/ and /l/) and /r-/l/ production of Japanese speakers (i.e., how English speakers identified and rated /r-/l/ production of Japanese speakers, and to what extent /r-/l/ production of Japanese speakers were similar to English speakers). The table also displays the correlations between /r-/l/ discrimination sensitivity (i.e., how Japanese speakers are sensitive to acoustic variations in the F2 and F3 dimensions) and /r-/l/ production of Japanese speakers. In order to examine the correlations, multiple comparisons were conducted using Bonferroni correction. The critical p-value was set to 0.01. Correlational analyses were divided into five sets. The first family of test involved /r-/l/ identification and /r-/l/ production. The other four families of tests involved /r-/l/ discrimination (i.e., F2 average discrimination sensitivity, F2 peak discrimination sensitivity, F3 average discrimination sensitivity, F3 peak discrimination sensitivity) and /r-/l/ production.

For /r-/l/ identification, Pearson correlational analyses revealed that there was a significant correlation with /r-/l/ production intelligibility (i.e., how well English speakers overall identified /r-/l/ production of Japanese speakers),  $r = 0.56$ ,  $p < 0.001$ . The statistical analyses revealed that there was a marginal correlation with /r-/l/ production goodness (i.e., how English speakers overall gave ratings to /r-/l/ production of Japanese speakers),  $r = 0.41$ ,  $p = 0.02114$ . The statistical analyses also revealed that there was a marginal correlation with transition duration production accuracy,  $r = -0.45$ ,  $p = 0.011662$ , suggesting that Japanese speakers who had good control in this temporal aspect were good at identifying /r/ and /l/. There were no significant correlations between /r-/l/ identification and other production accuracies, F1,  $r = -0.08$ , F2,  $r = -0.25$ ,

**Table 3.2 Correlations between /r-/l/ perception and production.** The results demonstrated that there was a significant correlation between /r-/l/ identification and /r-/l/ production intelligibility. However, there were no correlations between /r-/l/ identification and the other production measurements. The results also demonstrated that there were no significant correlations between discrimination sensitivities and production measurements.

	<i>/r-/l/</i> perception by Japanese				
	Identification	Discrimination Sensitivity			
		F2 average	F2 peak	F3 average	F3 peak
<i>/r-/l/</i> production intelligibility	0.56 ***	0.09	-0.18	-0.14	0.21
<i>/r-/l/</i> production goodness	0.41	0.05	-0.25	-0.12	0.07
F1 production accuracy	-0.08				
F2 production accuracy	-0.25	0.03	-0.02		
F3 production accuracy	-0.24			0.13	-0.07
Closure duration production accuracy	0.36				
Transition duration production accuracy	-0.45				

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$

F3,  $r = -0.24$ , closure duration,  $r = 0.36$ ,  $p > 0.05$ . These results thus suggest that Japanese speakers who are good at identifying /r-/l/ are also good at producing /r-/l/; there is a relationship between /r-/l/ perception and production.

For F2 average discrimination sensitivity (i.e., F2 sensitivity which is averaged across sensitivities within /r/ and /l/ categories and at the /r-/l/ boundary), there were no significant correlations with /r-/l/ production intelligibility,  $r = 0.09$ , /r-/l/ production goodness,  $r = 0.05$ , and F2 production accuracy (i.e., the degrees in which /r-/l/ production of Japanese speakers was similar to English speakers in the F2 dimension),  $r = 0.03$ ,  $p > 0.01$ . For F2 peak discrimination sensitivity (i.e., F2 sensitivity which mean sensitivity within /r/ and /l/ categories is subtracted from sensitivity at the /r-/l/ boundary), there was no significant correlation with /r-/l/ production intelligibility,  $r = -0.18$ , /r-/l/ production goodness,  $r = -0.17$ , and F2 production accuracy,  $r = -0.02$ ,  $p > 0.05$ . For F3 average discrimination sensitivity, there was no significant correlation with /r-/l/ production intelligibility,  $r = -0.14$ , /r-/l/ production goodness,  $r = -0.12$ , and F3 production accuracy,  $r = 0.13$ ,  $p > 0.01$ . For F3 peak discrimination sensitivity, there was no significant correlation with /r-/l/ production intelligibility,  $r = 0.21$ , /r-/l/ production goodness,  $r = 0.07$ , and F3 production accuracy,  $r = -0.07$ ,  $p > 0.01$ . These results thus suggest that there is no relationship between /r-/l/ discrimination sensitivity and /r-/l/ production; L2 phoneme discrimination processes may not be incorporated into L2 production processes.

Table 3.3 displays correlations between /r-/l/ best exemplar accuracy (the degrees in which /r-/l/ best exemplars of Japanese speakers were similar to those of English speakers) and /r-/l/ production. In order to examine whether the best exemplar accuracy



is related to /r-/l/ production, multiple comparisons were conducted using family-wise Bonferroni correction. The critical p-value was set to 0.01. Correlational analyses were divided into five sets (i.e., F1, F2, F3, closure duration, and transition duration). Each family of test involved /r-/l/ best exemplar accuracy in one acoustic dimension, /r-/l/ production (i.e., /r-/l/ production intelligibility, and goodness), and production accuracy in the acoustic dimension. For example, multiple correlational analyses were conducted using F3 best exemplar accuracy, /r-/l/ production, and F3 production accuracy.

For F1 best exemplar accuracy, there was no significant correlation with /r-/l/ production intelligibility,  $r = 0.13$ , /r-/l/ production goodness,  $r = 0.05$ , and F1 production accuracy,  $r = -0.09$ ,  $p > 0.01$ . For F2 best exemplar accuracy, there was no significant correlation with /r-/l/ production intelligibility,  $r = -0.21$ , /r-/l/ production goodness,  $r = -0.06$ , and F2 production accuracy,  $r = 0.32$ ,  $p > 0.01$ . For F3 best exemplar accuracy, there was no significant correlation with /r-/l/ production intelligibility,  $r = -0.10$ , /r-/l/ production goodness,  $r = -0.14$ , and F3 production accuracy,  $r = -0.23$ ,  $p > 0.01$ . For closure duration best exemplar accuracy, there was no significant correlation with /r-/l/ production intelligibility,  $r = -0.07$ , /r-/l/ production goodness,  $r = 0.01$ , and closure duration production accuracy,  $r = -0.23$ ,  $p > 0.01$ . For transition duration best exemplar accuracy, there was no significant correlation with /r-/l/ production intelligibility,  $r = 0.10$ , /r-/l/ production goodness,  $r = 0.10$ , and transition duration production accuracy,  $r = -0.15$ ,  $p > 0.01$ . These results thus demonstrated that /r-/l/ discrimination sensitivity and best exemplar accuracy are not related to /r-/l/ production. That is, not all of L2 perceptual aspects are related to L2 production.

**Table 3.3 Correlations between /r-/l/ best exemplar accuracy and /r-/l/ production. The results demonstrated that there were no significant correlations between /r-/l/ best exemplar accuracies and production measurements.**

	<i>/r-/l/ best exemplar accuracy</i>			
	F1	F2	F3	Closure duration
<i>/r-/l/ production intelligibility</i>	0.13	-0.21	-0.10	-0.07
<i>/r-/l/ production goodness</i>	0.05	-0.06	-0.14	0.01
F1 production accuracy	-0.09			
F2 production accuracy		0.32		
F3 production accuracy			0.25	
Closure duration production accuracy				-0.23
Transition duration production accuracy				
				-0.15

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

### **Links among English /r-/l/ perception measurements**

One of the purposes of this study was to examine how /r-/l/ discrimination sensitivity is related to /r-/l/ production and perception. The results demonstrated that the discrimination sensitivity is not predictive of /r-/l/ production. In order to examine whether the discrimination sensitivity is linked to other perceptual measurements (i.e., /r-/l/ identification and best exemplars), multiple comparisons were conducted using family-wise Bonferroni correction. The critical p-value was set to 0.01. Correlational analyses were divided into four sets (i.e., F2 average discrimination sensitivity, F2 peak discrimination sensitivity, F3 average discrimination sensitivity, and F3 peak discrimination sensitivity). Each family of test involved a discrimination sensitivity, /r-/l/ identification accuracy, and best exemplar accuracy in either the F2 or F3 dimension. For example, multiple correlational analyses were conducted using F3 peak discrimination sensitivity, /r-/l/ identification, and F3 best exemplar accuracy.

For F2 average discrimination sensitivity, there were no significant correlations with /r-/l/ identification,  $r = 0.19$ , and F2 best exemplar accuracy,  $r = -0.06$ ,  $p > 0.01$ . For F2 peak discrimination sensitivity, there was no significant correlation with /r-/l/ identification,  $r = -0.03$ , and F2 best exemplar accuracy,  $r = -0.26$ ,  $p > 0.01$ . For F3 average discrimination sensitivity, there was no significant correlation with /r-/l/ identification,  $r = 0.08$ ,  $p > 0.01$ , suggesting that having high general sensitivity in the F3 dimension is not related to /r-/l/ identification. There was no significant correlation with F3 best exemplar accuracy,  $r = 0.04$ ,  $p > 0.01$ . For F3 peak discrimination sensitivity, there was a marginal correlation with /r-/l/ identification,  $r = 0.35$ ,  $p = 0.02934$ ,

suggesting that Japanese speakers who had high sensitivity at the /r-/l/ boundary and low sensitivity in /r/ and /l/ categories were good at identifying /r/ and /l/. However, there was no significant correlation with F3 best exemplar accuracy,  $r = -0.27, p > 0.01$ . The results thus demonstrated that /r-/l/ discrimination sensitivity is weakly related to /r-/l/ identification and best exemplar accuracy only in the F3 dimension.

### **3.6 General discussion**

The main purpose of this study was to examine whether there is a relationship between English /r-/l/ perception and production by adult Japanese speakers. The results demonstrated that there are individual differences in identifying /r/ and /l/ with a wide range of variability. Such differences were moderately related to the degree in which English speakers identified /r-/l/ production of Japanese speakers. This finding supports the view of theoretical frameworks such as motor theory (e.g., Galantucci et al., 2006; Liberman et al., 1967; Liberman et al., 1952; Liberman & Mattingly, 1985, 1989), direct realism (e.g., Best, 1995; Fowler, 1981, 1984, 1986, 1989, 1994, 1996), and general auditory approach (e.g., Diehl et al., 2004) that there is a close relationship between speech perception and production. One of the possible reasons why previous studies (i.e., Goto, 1971; Sheldon and Strange, 1982) suggested that /r-/l/ production of Japanese speakers might precede /r-/l/ perception is that they had small samples in their studies. The present study is probably the first study to demonstrate that how individual differences in /r-/l/ perception and production by adult Japanese speakers are correlated with each other.

Another purpose of this study was to examine how English /r/-l/ discrimination sensitivity is related to other perceptual and production processes. The present study demonstrated that /r/-l/ discrimination sensitivity is weakly incorporated into other perceptual processes; Japanese speakers who have high F3 discrimination sensitivity at the English /r/-l/ boundary and low sensitivity in /r/ and /l/ categories are good at identifying /r/ and /l/. However, the discrimination sensitivity was hardly related to the other perceptual processes. This suggests that perceptual processes may be more or less independent; these processes are probably more task-specific. The present study also demonstrated that /r/-l/ discrimination sensitivity is not incorporated into /r/-l/ production processes. Such a finding is in accord with the hypothesis of SLM (Flege, 1999) that the degree of L2 perception accuracy limits how accurately L2 segments are produced although some aspects of perceptual learning may not be incorporated into L2 production processes. The present study suggests that, at least in the case of /r/-l/ perception and production, discrimination sensitivity has little to do with how accurately Japanese speakers produce /r/ and /l/. If this SLM hypothesis is revised based on the present study, it may be stated as followed; the degree of L2 perception accuracy limits how accurately L2 segments are produced although perceptual learning in terms of discrimination sensitivity may not be incorporated into L2 production. In order to further examine and modify the hypothesis, more observations looking at the link between L2 perception and production are necessary in the future.

The other purpose of this study was to assess how adult Japanese speakers produce English /r/ and /l/. Previous studies (e.g., Flege et al., 1995; Lotto et al., 2004; Masaki et al., 1996) demonstrated that there are individual differences in /r/-l/ production.

The present study confirmed such a view and demonstrated that Japanese speakers are generally poorer at producing English /r/ than /l/. Adult Japanese speakers produced higher F2, F3, and shorter transition duration for English /r/ compared to British English speakers. However, they are generally able to produce English /l/ despite the fact that their F3 was lower than British English speakers.

One possible explanation for such inaccurate /r/ productions of Japanese speakers is that they may not have accurate articulatory knowledge in their mental representations and suffer from producing the consonants. A possible factor which makes the articulation learning difficult is substantial individual differences in English speakers' tongue configurations. Some previous studies (e.g., Alawan et al., 1997; Espy-Wilson et al., 2000; Hagiwara, 1995; Westbury et al., 1998; Zhang et al., 2003; Zhou et al., 2007, Zhou et al., 2008) have reported that American English speakers generally use retroflex /r/ and bunched /r/ with a wide range of tongue configurations among individuals. Retroflex /r/ uses a raised (or curled) tongue tip and lowered tongue dorsum, and bunched /r/ uses a lowered tongue tip and a raised tongue dorsum. If Japanese speakers detect and learn articulatory gestures of retroflex and bunched /r/ with the individual differences, which generally have opposite tongue configurations in tongue tip and tongue dorsum, they may get confused, suffer from generalizing the tongue configurations, and end up using inaccurate articulatory gestures. Chapter 4 will examine whether Japanese speakers can learn accurate articulatory gestures of English /r/ and /l/.

Of interest to L2 speech perception and production models, the /r/-/l/ production ability of Japanese speakers is problematic for SLM, as suggested in Chapter 2. The model hypothesizes that learning L2 phoneme is easier if the phoneme is dissimilar to L1

### Chapter 3 The relationship between /r/-/l/ perception and production

phonemes. In the case of /r/-/l/ learning, Japanese speakers are expected to be able to acquire English /r/ easier than English /l/ because Chapter 2 demonstrated that English /r/ is perceptually more dissimilar to Japanese flap. However, the present results demonstrated that Japanese speakers are generally good at producing an identifiable /l/ but not /r/. If learning /l/ is more difficult than learning /r/, it will be the case that Japanese speakers have to increase their F3 frequency in order to produce clear /l/. Since /l/ with lower F3 is already sufficient to be identified as /l/ by English speakers, such a tiny modification will require more attention and make this subtle learning very difficult.

## **Chapter 4: Examination of L2 production training effects on L2 production and perception: Training adult Japanese speakers to produce English /r/-l/**

### **4.1 Introduction**

Chapter 3 demonstrated that there is a relationship between /r/-l/ perception and production. In particular, there was a moderate correlation between /r/-l/ production accuracy (i.e., the accuracy with which English speakers identified /r/-l/ productions of Japanese speakers) and /r/-l/ identification accuracy (i.e., to what extent Japanese speakers identified English /r/-l/). However, the /r/-l/ productions of Japanese speakers were hardly related to the other perceptual measures (e.g., discrimination). Although some theoretical frameworks such as motor theory (e.g., Galantucci et al., 2006; Liberman et al., 1967; Liberman et al., 1952; Liberman & Mattingly, 1985, 1989), direct realist theory (e.g., Best, 1995; Fowler, 1981, 1984, 1986, 1989, 1994, 1996), and general auditory approach (e.g., Diehl et al., 2004) hypothesized a close relationship between speech perception and production, the relationship found in Chapter 3 was not particularly strong. Considering these results, it seems reasonable to question whether speech perception and production share common underlying mechanisms to process speech. The present chapter further examined whether /r/-l/ perception and production share common underlying mechanisms by determining whether /r/-l/ pronunciation training changes the /r/-l/ production and perception of Japanese speakers.

Some previous studies (e.g., Bradlow et al., 1997; Bradlow et al., 1999; Rochet, 1995, Wang et al., 2003) used perceptual training and demonstrated that this leads to



improved L2 perception and production, suggesting that there may be transfer mechanisms between speech perception and production. For example, Rochet (1995) examined whether Mandarin Chinese speakers improve perception and production accuracy of French voice onset time (VOT). Twelve Mandarin Chinese speakers underwent six 30-minute sessions of perceptual training identifying synthetic /bu-/pu/ stimuli. In pre/post-tests, the participants identified synthetic voiced/voiceless labial (i.e., /p/ and /b/) followed by /u/, /a/, and /i/, and dental, and velar (i.e., /t/, /d/, /k/, and /g/) followed by the vowel /u/. They also imitated and identified such monosyllables spoken by a native speaker of French. The results demonstrated that the Mandarin Chinese speakers perceptually shifted their VOT to a French native-like VOT when they listened to the synthetic stimuli, and that they improved the identification accuracy of natural voiceless stimuli (i.e., /pa/, /pi/, /pu/, /tu/, and /ku/), but not voiced stimuli (i.e., /ba/, /bi/, /bu/, /du/, and /gu/). The results also demonstrated that Mandarin Chinese speakers improved and produced more native-like VOTs with both voiceless and voiced consonants.

Wang et al. (2003) examined whether American English speakers improve tone perception and production accuracy of Mandarin Chinese. Eight out of 16 American English speakers underwent a 2-week perceptual training program (i.e., eight 40-minute sessions) identifying naturally spoken monosyllabic Mandarin words pairwise (i.e., Tones 1 and 2, Tones 1 and 3, Tones 1 and 4, Tones 2 and 3, Tones 2 and 4, Tones 3 and 4). In pre/post-tests, the participants identified 100 naturally spoken stimuli and made recordings of 80 stimuli. In the post-test, they additionally identified 120 new stimuli. The results demonstrated that American English speakers improved in their tone

identification accuracy of Mandarin Chinese. Such perceptual improvement was extended to new stimuli. American English speakers also improved in tone production accuracy after the training; they produced both pitch and contour more similar to Mandarin Chinese speakers' production. The results demonstrated that the perception and production of Mandarin Chinese tone were highly correlated in both pre-test and post-test.

Bradlow et al. (1997) similarly examined whether Japanese speakers improve perception and production of English /r/ and /l/. Eleven Japanese speakers underwent 45 sessions (over a period of 3-4 weeks) of perceptual training identifying 68 naturally spoken /r/-l/ minimal pairs with feedback. In pre-/post-tests, Japanese speakers identified 16 minimal pairs and made recordings of 55 English words including /r/ and /l/. In the post-test, they additionally identified 195 new stimuli. The results demonstrated that Japanese speakers improved their /r/-l/ identification accuracy by 16 percentage points after the training, and that Japanese speakers' /r/-l/ production accuracy significantly improved; American English speakers identified post-test production better than pre-test production. The results thus demonstrated that there is a link between perception and production in the sense that perceptual learning transferred to improved /r/-l/ production. The results also demonstrated that the amount of improvement in /r/-l/ perception and production were poorly correlated. Bradlow et al. (1999) further examined whether Japanese speakers retain such training effects on /r/-l/ perception and production for three months, and demonstrated that Japanese speakers did retain better /r/-l/ identification and production accuracy.

The present study further investigated Japanese speaker's English /r/-l/ perception and production by using one-to-one pronunciation training and by determining

whether such training promotes accurate /r/-/l/ production and perception. Given that perceptual training studies (i.e., Bradlow et al., 1997; Bradlow et al., 1999; Rochet, 1995, Wang et al., 2003) demonstrated transfer of perceptual learning to improvement in speech production, and that Chapter 3 demonstrated that there is a relationship between speech perception and production, it seems that there are common mental representations underlying perception and production. If there is such a common representation, it may be possible that phonetic category learning through production-focused training similarly promotes better /r/-/l/ perception and production accuracy. That is, Japanese speakers may be able to produce better /r/-/l/ (e.g., /r/ with lower F3 and /l/ with higher F3), and change their /r/-/l/ perception (e.g., native-like identification, discrimination sensitivity at the /r/-/l/ boundary, and best exemplars). Alternatively, it may be possible that /r/-/l/ production training is only effective for acquiring specific motor commands for improved /r/-/l/ production (e.g., native-like tongue configurations) and is not sufficient on its own to result in improvement in speech perception. That is, Japanese speakers will improve /r/-/l/ pronunciation, but they will not change their /r/-/l/ perception (e.g., poor /r/-/l/ identification, low discrimination sensitivity at the /r/-/l/ boundary, and inaccurate best exemplars).

The present study combined two approaches to teach English /r/-/l/ pronunciation. The first approach was to use explicit instructions and feedback. Pronunciation has been recently taught using computer assisted language learning (CALL) system based on automatic speech recognition (ASR; e.g., Burlison, 2007; Chou, 2005, Dalby and Kewley-Port, 1999; Eskenazi, 1999; Cucchiarini et al., 2007; Kewley-Port et al., 1991; Mich et al., 2006; Neri et al., 2006). For example, Mich et al. (2006) developed a CALL

system called PARLING in order to facilitate English pronunciation of Italian children. Italian children listened to some famous stories for children with a visual display. They could check sample pronunciations of vocabulary in the stories and make recordings of their own pronunciation. PARLING analyzes the recordings in real time and tells whether words are correctly pronounced. If the system does not recognize a children's pronunciation, children are asked to repeat the mispronounced words. Neri et al. (2006) developed a CALL system called Dutch CAPT (computer assisted pronunciation training) in order to facilitate L2 Dutch speaker's vowel and consonant productions. The L2 learners participate in role-plays, answer questions by pronouncing one of the possible answers, and pronounce minimal pairs. If the ASR algorithm does not recognize learner's recordings, the learners see the transcription of word indicating mispronounced segments with red letters. Dalby and Kewley-Port (1999) developed a CALL system called PRONTO in order to facilitate American English speaker's Spanish pronunciation and Mandarin Chinese speaker's English pronunciation. L2 learners identify naturally spoken stimuli, imitate aurally presented words, and produce visually presented words. The PRONTO evaluates the imitation and production and gives feedback along with identification accuracy using a bar chart. Yamada et al. (1998) developed software which Japanese speakers can make recordings and compare their /r/-/l/ production to English speakers' /r/-/l/ production using spectrograms.

Despite the fact that these CALL systems are successful in promoting L2 learner's pronunciation, there seems to be some disadvantages. One disadvantage is that feedback with the CALL systems may not be helpful for some L2 learners, in that the feedback does not tell the learners why their pronunciation is good or bad. In order to correct

inaccurate pronunciation, the L2 learners may rather need explicit feedback on their mispronunciation (e.g., feedback on spectral and temporal aspects of pronunciation). Another disadvantage is that, even if L2 learners know their production is inaccurate, they may need explicit instructions on articulation. Otherwise, L2 learners may continuously reinforce poor pronunciation and such pronunciation may result in fossilization. The other disadvantage is that, when non-native speakers make direct comparisons with native speaker's production, it is usually difficult for L2 learners to notice differences between their pronunciation and native speaker's pronunciation due to between-talker differences (e.g., gender, vocal track length, and speaking rate). Therefore, the present study rather used a phonetically trained Japanese-English speaker as an instructor in order to provide precise and explicit feedback on English /r/-/l/ production of Japanese speakers. The present study also used signal processing techniques of Iverson et al. (2005) modifying the English /r/-/l/ production of Japanese speakers (i.e., enhancing F3, closure duration, and transition duration) in order to remove the between-talker differences. In this way, Japanese speakers can easily compare their pronunciation and their "ideal" pronunciation.

The second approach was to use real-time spectrograms in order to track Japanese speakers' speech productions. Some clinical studies have previously used spectrograms for perceptual evaluation of patients' speech production (e.g., Chaney, 1988; Hagiwara et al., 2002; Huer, 1989; Maxwell and Weismer, 1982; Weismer et al., 1981). For example, Chaney (1988) spectrographically analyzed correct and misarticulated semivowels (i.e., /w/, /r/, /l/, and /j/) of American children and demonstrated differences in semivowel productions between normal and articulation impaired children. Huer (1989) used

acoustic tracking to examine whether a 10-year-old girl substituting /w/ for /r/ changed her F2 transition and F2 values over a 70-day remediative period. Hagiwara et al. (2002) implemented acoustic analysis to document the speech of a 6-year-old child with persistent /r/-distortion. They demonstrated how the /r/ pronunciation of the child acoustically changed before and after receiving speech therapy; they clearly showed that F3 formant changed after the therapy. Following these clinical studies, the present study implemented real-time spectrograms into the pronunciation training. This allowed an instructor to meticulously observe F3 of English /r/-l/.

A subset of Japanese speakers from Chapter 3 participated in ten sessions of one-to-one pronunciation training. Twenty-eight participants completed a series of perceptual and production experiments (i.e., English /r/-l/ identification, discrimination, perceptual mapping of /r/-l/ best exemplars, production recordings of /r/-l/ words, sentences, and passages) before and after the training. All experiments are identical to ones in Chapter 3. The aims were to determine whether /r/-l/ articulatory training facilitates /r/-l/ production of Japanese speakers, and whether such training also leads to perceptual changes (e.g., improvement in /r/-l/ identification, change in discrimination sensitivity, and change in /r/-l/ best exemplars).

## **4.2 Experiment 3.1: /r/-l/ Production Training**

### **4.2.1 Method**

#### **Subjects**

Twenty-eight participants from the previous chapter participated in the present study. Their ages ranged from 21 to 67 years old (median = 34 years old). They started learning English between 9 and 13 years old (median = 13 years old), and had received instruction for 6 - 12 years (median = 8 years). All participants grew up in Japanese-speaking environments in Japan. They had lived in English-speaking countries between 1 month to 17 years and 2 months (median = 1 year and 2 months). None of the participants reported having hearing problems.

#### **Apparatus**

Participants were tested and trained in a sound-proof room. Pre- and post-tests were run using a Radio Spares (RS) 249-946 microphone, Edirol USB Audio Capture UA-25 device, and Dell Optiplex GX 260, and Senheiser HD 280 headphones. Training sessions were run using a hand mirror, SFS real-time speech spectrogram, and Praat (Boersma and Weenik, 2009) along with the recording equipment.

### **4.2.2 Procedure**

#### **Pre/post testing**

All Japanese participants were tested using all of the perception and production measurements in Chapter 3 (i.e., English /r/-l/ identification, /r/-l/ discrimination,

perceptual mapping of /r/-/l/ best exemplar, and /r/-/l/ production recordings). The pre-tests were carried out right before the first training session, and the post-tests were carried out right after the last training session. Since these 28 participants were involved in the previous study in Chapter 3, their data from the chapter was used for their pre-test.

#### **4.2.2.2 Pronunciation training**

##### ***Training curriculum***

The pronunciation training comprised 10 sessions, each taking approximately 30-40 minutes to complete. The participants could participate in one session per day. They completed the training over a 2 to 3 week period.

##### ***Training material***

##### **Training words**

Only three minimal-pair words (i.e., *lack*, *rack*, *lick*, *rick*, *loom* and *room*) were used in the entire training. Participants practiced and made recordings of the words in each session.

##### **Real-time spectrogram**

SFS real-time spectrograms were used in order to assess participants' F3 formants. The instructor demonstrated how the real-time spectrogram works in the first session. He produced /ra/, /la/ and /ra/ in order to let participants know how these three consonants differ in terms of F3, closure duration, and transition duration. Participants were allowed to make some recordings so that they became familiar with the software.



In the training sessions, the instructor used the real-time spectrogram when participants exercised to produce English /r/ and /l/, and provided feedback. For example, if he found that participants' English /r/ F3 was too high (e.g., 2500 Hz), he told participants to check their tongue position, tongue shape, and lip shape. If he found that participants were producing good English /r/, he provided positive feedback and encouraged the participants to maintain the articulation and produce the consonant.

#### **Enhanced English /r/-/l/ stimuli**

The instructor chose participants' best pronunciations of the target words (i.e., *lack*, *rack*, *lick*, *rick*, *loom* and *room*) in each training session and signal-processed the initial portion of the recordings. The processing combined changes in closure and transition durations, and changes in F3 frequencies (see Figures 4.1, 4.2, 4.3, and 4.4). The duration changes were made using the PSOLA function in Praat (Boersma and Weenink, 2009). The duration of the closure and transition intervals were independently manipulated based on English speakers' best exemplars in Chapter 2; English /r/ closure and transition durations were set to 31 and 66 ms, and English /l/ closure and transition durations were set to 82 and 20 ms. The F3 frequency changes were made via LPC analysis and resynthesis within Praat; the LPC parameters (e.g., prediction order and frequency cutoff) were hand selected for each recording so that the analysis correctly tracked the formant in the spectrograms. The F3 frequencies were then manipulated; the English /r/ F3 value was set close to the value of F2 (e.g., 1200-1800 Hz), and English /l/ F3 value was set at the middle value of F3 and F4 (e.g., 2800-3700 Hz). The final stimulus was created by filtering the LPC residual with the new LPC parameters. In order to improve the naturalness of the recordings, the high-frequency energy that was removed

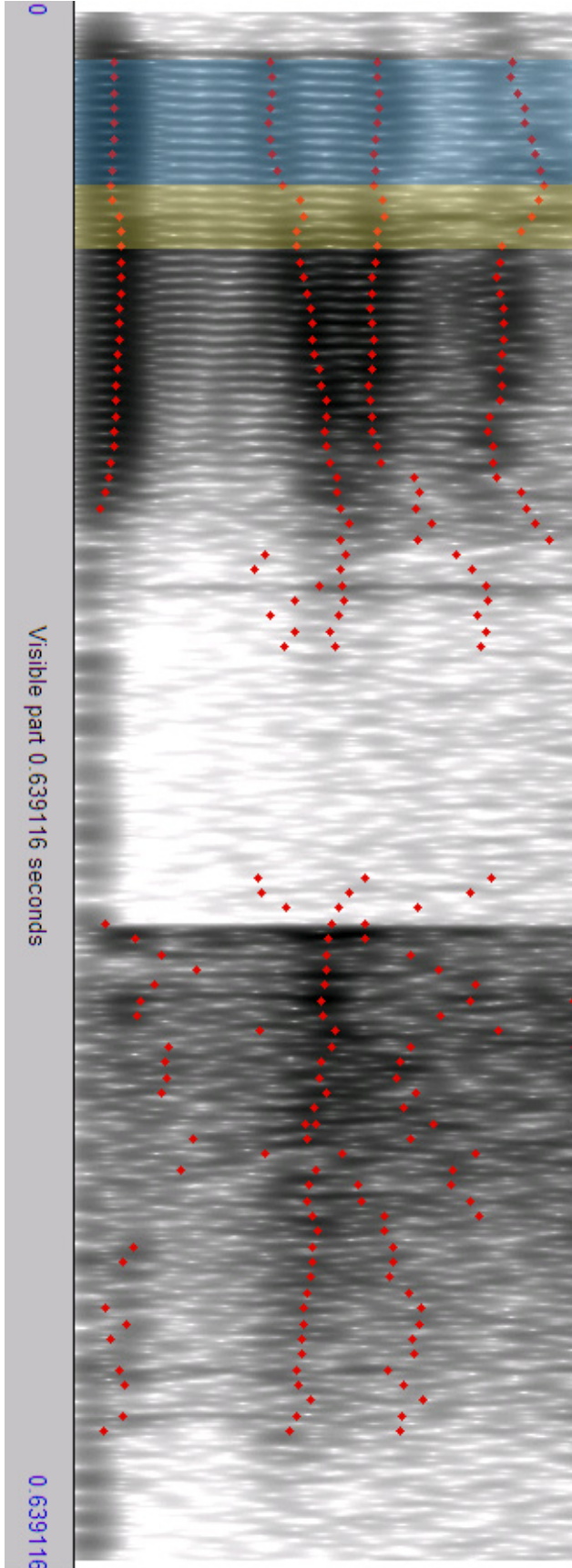


Figure 4. 1 A spectrogram of *lick* produced by a female Japanese speaker. This is her production in the 7<sup>th</sup> training session. The blue area indicates closure duration (i.e., 51 ms). The yellow area indicates transition duration (i.e., 26 ms). The average F3 frequency in the closure duration is 3052 Hz.

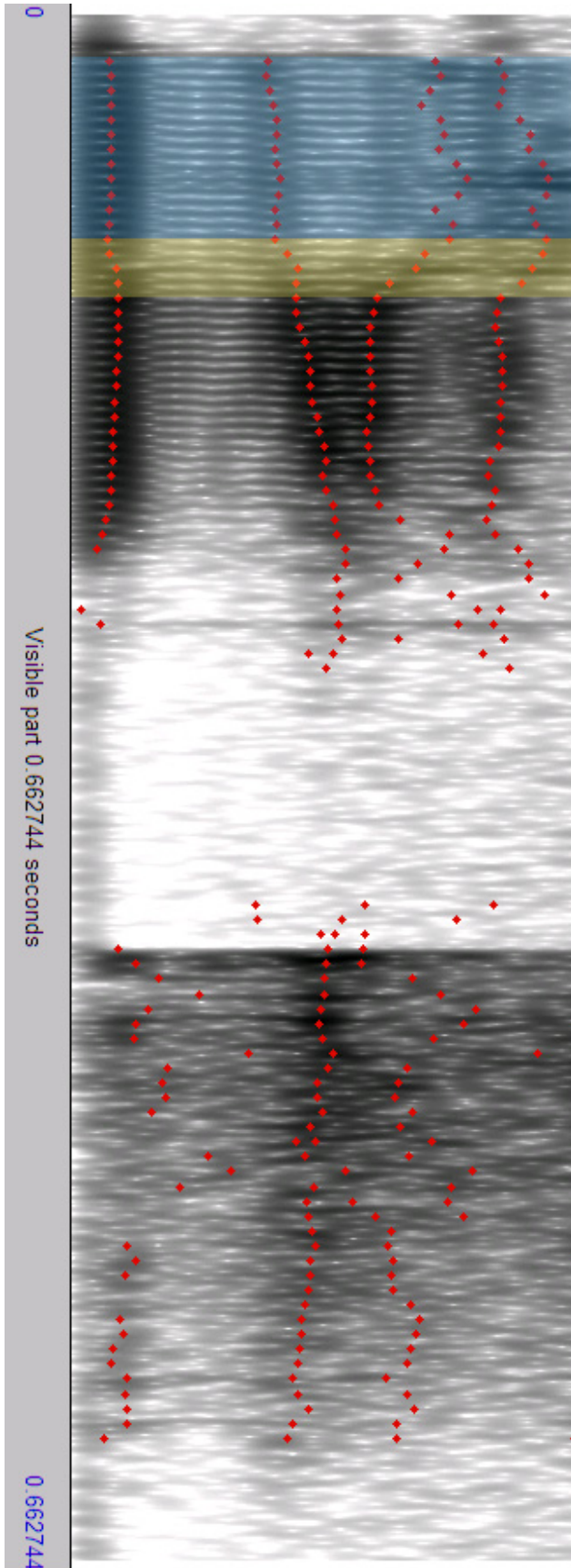


Figure 4. 2 A spectrogram of signal-processed *lick* by the female Japanese speaker from Figure 4.1. The blue area indicates enhanced closure duration. The original closure duration (i.e., 51 ms) was enhanced to 82 ms. The yellow area indicates transition duration. The original transition duration (i.e., 26 ms) was enhanced to 20 ms. The original average F3 frequency (i.e., 3052 Hz) was enhanced to 3672 Hz.

Figure 4. 3 A spectrogram of *rick* produced by a female Japanese speaker. This is her production in the 6<sup>th</sup> training session. The blue area indicates closure duration (i.e., 130 ms). The yellow area indicates transition duration (i.e., 88 ms). The average F3 frequency in the closure duration is 2056 Hz.

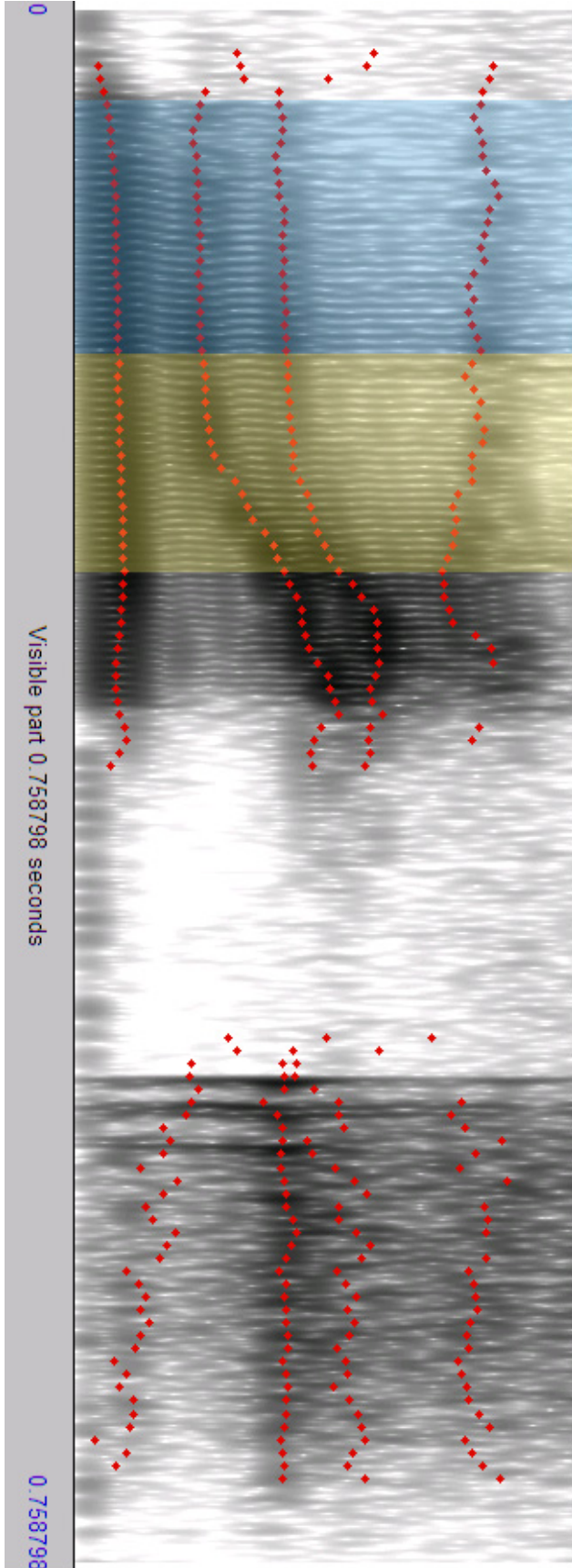
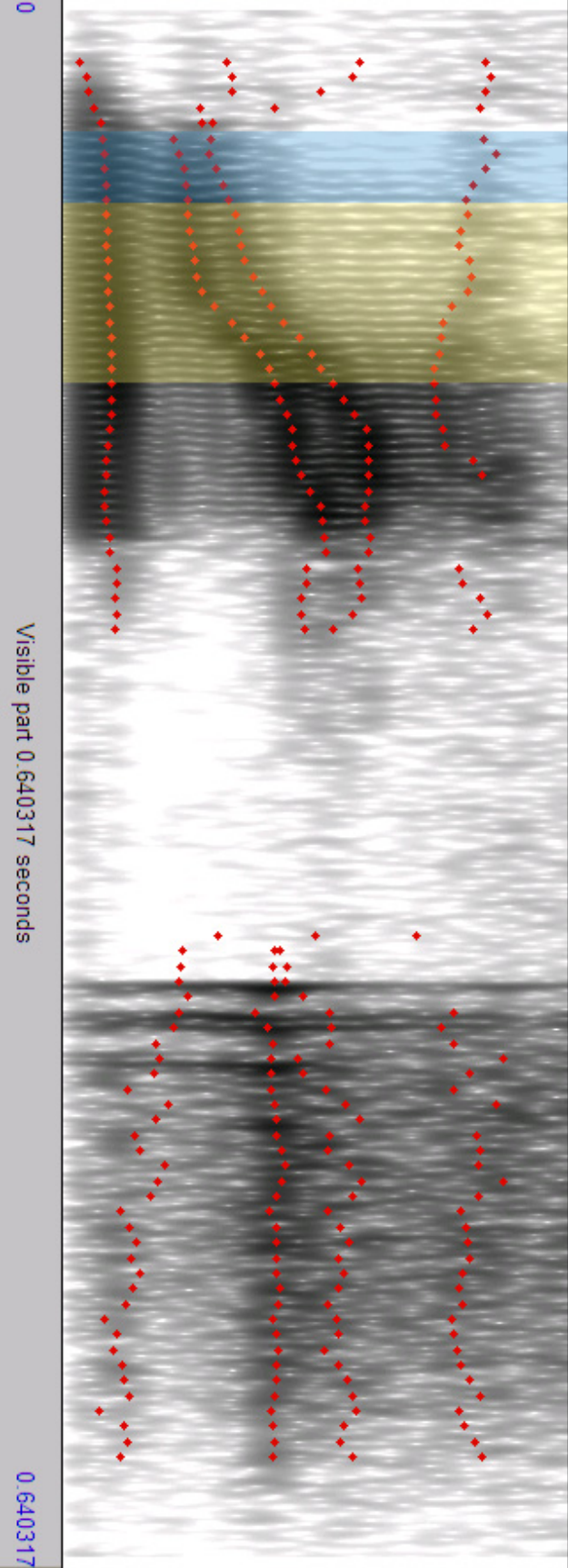


Figure 4. 4 A spectrogram of signal-processed *rick* by the female Japanese speaker from Figure 4.1. The blue area indicates enhanced closure duration. The original closure duration (i.e., 130 ms) was enhanced to 30 ms. The yellow area indicates transition duration. The original transition duration (i.e., 88 ms) was enhanced to 66 ms. The original average F3 frequency (i.e., 2056 Hz) was enhanced to 1500 Hz.



by LPC (i.e., energy that was above the cut-off frequency) was added back into the signal following the LPC manipulations.

Such enhanced stimuli were used in the beginning of each training session except the first session. The participants and instructor carefully listened to each pair of the target words (e.g., original *rack*, and enhanced *rack*) two or three times. The instructor asked the participants whether they noticed any differences between original and enhanced recordings. He then explained, using spectrograms, how the enhanced recordings were different from original recordings in terms of F3 frequency, closure duration, and transition duration.

### **4.2.2.3 Instructions for English /r/**

#### ***Stage 1***

Participants initially watched a video clip a few times in which an English speaker pronounced *wrens*. The instructor asked participants to pay attention to the native speaker's lip shape. He then replayed the video in slow motion a few times so that participants could see how the English speaker moved his upper lip toward his nose and made slight lip rounding. Participants were informed that there are individual differences in lip shapes and that lip rounding may be optional for them. The instructor also provided articulatory information in terms of tongue shape and position. He drew the side-face picture of a tongue with a raised tongue tip and a raised tongue dorsum (i.e., retroflex-type English /r/). He also opened his mouth and demonstrated the retroflex tongue shape

by holding the tongue shape steady so that participants could get the idea of how the tongue shape looked.

Once participants received such phonetic information, they were asked to produce *ar* or *bar*. This is because Japanese speakers seemed to produce good English /r/ in word final position. Participants were asked to hold a hand mirror and pay attention to their tongue position and lip shape. They repeatedly produced *ar* or *bar* until the instructor confirmed their F3 going down somewhere under 2200 Hz in the real-time spectrogram. Whenever participants were struggling to produce English /r/, the instructor asked participants where their tongue position was, and whether they had a raised tongue tip (i.e., retroflex) in order to build up their awareness to their own articulation. Based on how their English /r/ productions sounded, he experimentally asked participants to further raise/lower their tongue dorsum and/or curl their tongue tip. Based on how their lip rounding looked, he experimentally asked participants either to change their lip shapes (e.g., more lip rounding, and less lip rounding by lifting the edges of upper lip) or not to use lip rounding at all.

After confirming that participants were producing English /r/ to some extent, participants were asked to make recordings of *ar* or *bar* repeating after the instructor. They were asked to keep /r/ duration long (e.g., 1-2 seconds, if possible). The instructor immediately replayed the /r/ part in the recordings using spectrograms and let participants know how English /r/ sounded like with their own voice. Participants then made recordings of *arack*, *arick*, or *aroom* with a relatively long /r/ duration (e.g., 100 ms or longer), repeating after the instructor. The instructor replayed the recordings from the point where /r/ closure duration was roughly 50-70 ms so that the recordings sounded as

if the participants pronounced *rack*, *rick*, or *room*. Participants went through this stage until they were able to keep F3 frequency as low as in word final position.

### **Stage 2**

Once participants were consistently producing English /r/ in word final position, watching the audio-visual material became optional. Participants warmed up by pronouncing *ar* or *bar* with long /r/ duration, *arack*, *arick*, and *aroom*. In the training, participants were asked to produce *ar*, stop the pronunciation, and check their tongue shape, position and lip shape so that participants learned and confirmed their own correct articulation. Participants were asked to repeatedly produce /ar/ and /r/ until they were able to produce isolated /r/.

Confirming their ability to produce such English /r/, participants attempted to produce /r/ with a following vowel (i.e., /ra/, /ri/, and /ru/). Many participants initially had troubles producing the monosyllables because they could not move their tongue and lips to produce English /r/ with a short duration (e.g., 30- 60 ms), and they could not have a smooth coarticulation with the following vowel (e.g., some participants tended to keep lip rounding and ended up not producing a clear /a/). Therefore, participants were told to keep closure and transition durations long (e.g., longer than 100 ms); the instructor told participants to produce /ra/ as if they were slowly yawning and saying /ra/ using a hand mirror. Note that participants' vowels were corrected if they had vowel problems (e.g., Some Japanese speakers produced unrounded /u/). A similar approach was taken to facilitate participants' /ri/ and /ru/ productions. Participants produced such exaggerated monosyllables several times while the instructor was checking their lip shape and



tracking their F3 formant in a real-time spectrogram. Participants followed the instructor and made separate recordings of each monosyllable and target word a few times in each session (e.g., participants produced *ra* and *rack* repeating after the instructor three times). The participants and the instructor immediately checked each recording using spectrograms, and instructor provided feedback on their F3, closure and transition durations. Note that feedback priority was given to F3 and temporal aspects were ignored unless participants produced extreme durations. They also made comparisons between the instant recordings and enhanced recordings from the previous session. Participants went through this stage until they were able to produce consistent English /r/ productions in word initial position.

### **Stage 3**

Once participants were able to produce consistent initial-position /r/, they attempted to adjust their closure and transition durations. This is because some participants tended to produce longer durations due to the training instructions. The goal was to match participants' closure and transition durations to English speakers' /r/ best exemplars as much as possible (i.e., 31.36 ms for closure duration, and 80.64 ms for transition duration).

Participants warmed up by producing target words (i.e., *rack*, *rick*, and *room*) repeating after the instructor twice or three times. In the training, the instructor carefully controlled his closure and transition durations in order to match his English /r/ production to English speakers' best exemplar as much as he could. He asked participants to attend to the temporal aspects of his /r/ production before they made recordings of the target words. Participants made three repetitions of a target word in each recording while the

instructor observed the participants' lip shape. They immediately listened to their recordings a few times, and the instructor asked them to choose the best production among the three. Then, the instructor and participants looked into each attempt using spectrograms and checked their closure and transition durations along with the F3 formant. If their closure and/or transition durations were too short or long, participants were asked to listen to original and enhanced recordings from the previous session. When participants had difficulties in adjusting their temporal aspects or noticing the differences between the recordings, the instructor selected the closure duration with a following vowel from each sound file and interchangeably played the recordings. Then, participants repeated after the enhanced recording in order to get close to the ideal closure and transition duration.

## **Instructions for English /l/**

### ***Stage 1***

Participants initially watched a video clip a few times in which an English speaker pronounced *lens*. The instructor asked participants to pay attention to the native speaker's tongue tip. He then replayed the video in slow motion a few times so that participants could see how the English speaker moved his tongue tip toward his alveolar ridge and released his tongue tip. Participants were informed that the tongue tip touches the alveolar ridge and stays steady for roughly 100 ms. This information was given in order to make sure that English /l/ has a longer closure duration compared to Japanese flap which has an extremely short or no closure duration. Participants were also informed that English /l/ and Japanese flap are similar in terms of transition duration.

Once participants received such phonetic information, the instructor asked them to carefully watch how his tongue tip was moving while he was producing *lens*. He purposely exaggerated the closure duration so that participants noticed that there is closure duration. Participants were asked to hold a hand mirror, pay attention to their tongue tips, and produce *la* or *lack* by keeping their closure duration long. They repeated this exercise until the instructor confirmed some closure duration along with high F3 (e.g., 2800 Hz). Whenever participants were struggling in producing English /l/, the instructor asked participants to exaggerate closure and transition durations; they were asked to produce *la* or *lack* as if they were yawning.

### **Stage 2**

Once participants were able to produce consistent initial-position /l/, they attempted to adjust their closure and transition durations. This is because some participants tended to produce longer durations due to the training instructions. The goal was to match participants' closure and transition durations to English speakers' /l/ best exemplar as much as possible (i.e., 96 ms for closure duration, and 16 ms for transition duration). Training was given in the same manner as stage 3 for English /r/ production. Note that most of the participants received training in order to adjust temporal aspects of English /l/ since they were able to produce English /l/ with a high F3 before the training, and thus they spent less time to learn English /l/, compared to English /r/.

### 4.3 Results

#### 4.3.1 English /r/-/l/ production intelligibility and goodness by English listeners

Table 4.1 displays the pre-test mean confusion matrix for the English /r/-/l/ production intelligibility task. Five British English listeners identified and gave ratings to English /r/-/l/ productions by Japanese speakers. Japanese speakers seemed to have difficulties producing /r/; English listeners misidentified /r/ as /l/ in 30.60 % of the trials, /r/ as /d/ in 1.58 % of the trials, and /r/ as /w/ in 3.38% of the trials. However, Japanese speakers seemed to produce identifiable /l/; English listeners misidentified /l/ as /r/ in 4.89 % of the trials, /l/ as /d/ in 3.05 % of the trials, and /l/ as /w/ in 0.08 % of the trials.

**Table 4.1 Confusion matrix for pre-test /r/-/l/ production intelligibility. British English speakers listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or /d/. English speakers correctly identified most of the /l/ productions. However, they poorly identified /r/ productions; they misidentified approximately one third of /r/ productions as /l/. These suggest that Japanese speakers were able to produce identifiable /l/, but they had hard time to produce identifiable /r/.**

Stimulus	Response			
	/l/	/r/	/d/	/w/
/l/	91.99%	4.89%	3.05%	0.08%
/r/	30.60%	64.44%	1.58%	3.38%

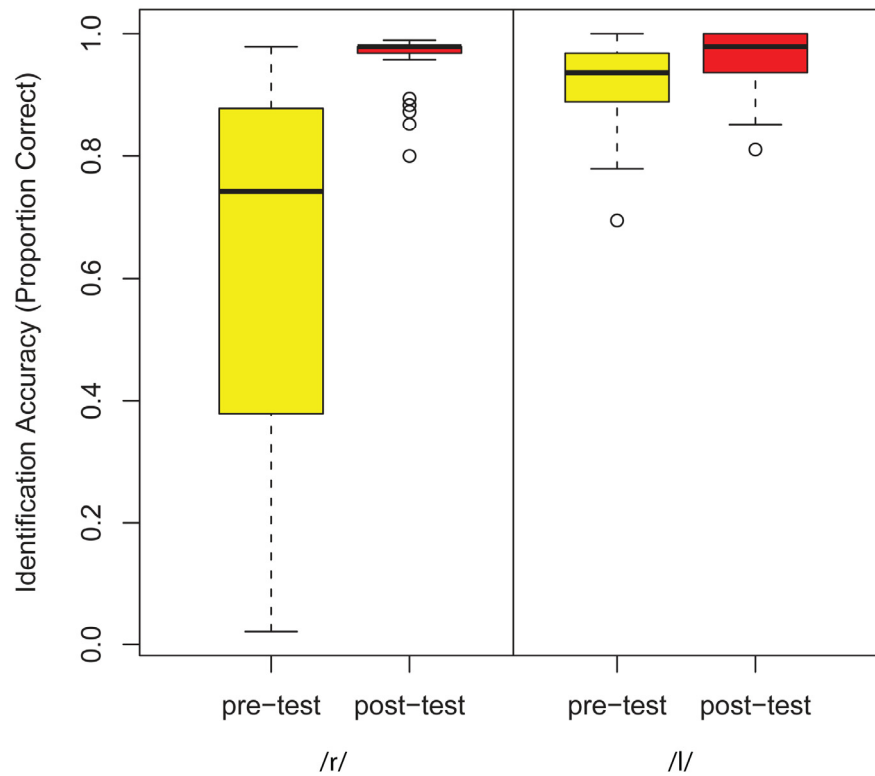
Table 4.2 displays the post-test mean confusion matrix for the English /r/-/l/ production intelligibility task. Japanese speakers were clearly able to produce English /r/ after the pronunciation training; English listeners misidentified /r/ as /l/ in only 1.62 % of the trials, /r/ as /d/ in 0 % of the trials, and /r/ as /w/ in 2.63 % of the trials. Japanese speakers seemed to be consistently able to produce an identifiable /l/, and they seemed to

stop producing a Japanese flap; English listeners misidentified /l/ as /r/ in 4.06 % of the trials, /l/ as /d/ in 0 % of the trials, and /l/ as /w/ in 0.08 % of the trials.

**Table 4.2 Confusion matrix for post-test production intelligibility. British English speakers listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese and identified consonant categories choosing /r/, /l/, /w/, or /d/. English speakers correctly identified most of the /l/ productions; they identified post-test /l/ productions slightly better than they did for pre-test /l/ productions. Similarly, English speakers identified most of the post-test /r/ productions, suggesting Japanese speakers were able to produce identifiable /r/ after the pronunciation training.**

Stimulus	Response			
	/l/	/r/	/d/	/w/
/l/	95.90%	4.06%	0%	0.08%
/r/	1.62%	95.75%	0%	2.63%

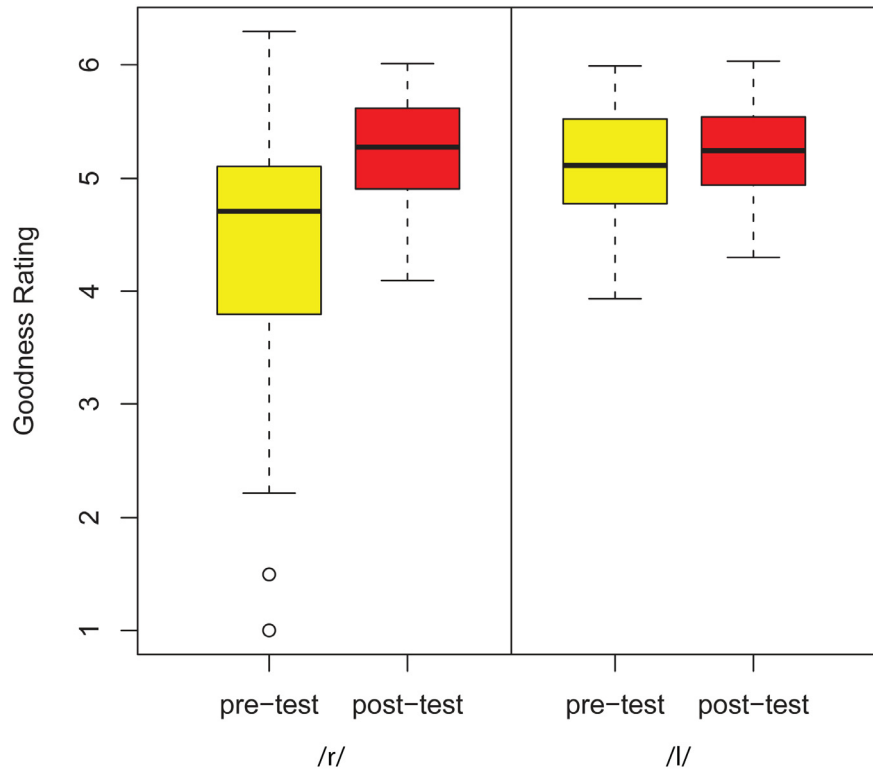
The results thus suggest that Japanese speakers' English /r/ production as a group generally improved after the training, and their /l/ production was good and consistent. Figure 4.5 displays Japanese individuals' /r/-/l/ production identification accuracy by English listeners. Pre-test English /r/ production accuracy widely varied, ranging from near 0 % to near 100 % correct (average = 64.4%). However, the identification of post-test /r/ production was near 100% (average = 95.7%), suggesting that Japanese speakers improved their /r/ pronunciation. The identification of pre-test English /l/ production varied less, compared to pre-test English /r/, ranging from 69% to 100% (average = 92%). Similarly, the identification of post-test /l/ production varied little, ranging from 81% to 100% (average = 96%), suggesting that Japanese speakers seemed to produce intelligible /l/ even before the training.



**Figure 4.5** Boxplots of English /r/ and /l/ production intelligibility for the pre-test and post-test. Boxplots display the medians and quartile ranges of scores, and outliers marked by circles. Five native speakers of British English listened to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese before and after the pronunciation training. They identified consonant categories by choosing /r/, /l/, /w/, or /d/. Higher proportion correct indicates that English speakers correctly identified /r/ and /l/ productions of L1 Japanese speakers. The intelligibilities of /l/ productions were similarly high before and after the training; Japanese speakers produced identifiable /l/ throughout this study. However, the intelligibilities of /r/ productions were different before and after the training. The pre-test intelligibility of /r/ productions widely varied among Japanese speakers, but the post-test intelligibility of /r/ productions was high and hardly varied; most of the Japanese speakers were able to produce identifiable /r/ after the training. Standard errors from the left side of the figure are 0.06, 0.01, 0.01, and 0.01. The results demonstrated that the intelligibility of /r/ and /l/ productions both improved after the pronunciation training.

In order to examine whether Japanese speakers' English /r/ and /l/ production became more identifiable after the training, paired t-tests were run for each consonant. For /r/, there was a significant difference between pre-test and post-test recognition accuracies,  $t(27) = -5.42$ ,  $p < 0.0001$ , suggesting that Japanese speakers' /r/ production became more intelligible to British English listeners. For /l/, there was also a significant

difference,  $t(27) = -2.39$ ,  $p < 0.05$ , suggesting that, although Japanese speakers' /l/ productions were highly intelligible before the training, their /l/ productions still became slightly more intelligible after the training.



**Figure 4.6** Boxplots of the goodness rating scores for pre-/post-tests /r/ and /l/ productions of Japanese speakers. British English speakers gave ratings to 19 initial-position /r/-/l/ minimal pairs produced by native speakers of Japanese with a scale from 1 (poor) to 7 (good). The intelligibility scores for /l/ were high and similar before and after the pronunciation training. However, the intelligibility scores for /r/ were different before and after the training. The intelligibility scores for pre-test /r/ productions substantially varied. However, the intelligibility scores for post-test /r/ productions less varied. Standard errors from the left side of the figure are 0.24, 0.09, 0.11 and 0.08. The results demonstrated that post-test /r/ and /l/ productions received higher intelligibility scores, suggesting that Japanese speakers improved their /r/ and /l/ productions.

Figure 4.6 displays similar patterns in goodness ratings for English /r/ and /l/ productions. The goodness ratings for English /r/ widely varied in the pre-test, ranging from 1 to 6.3 (average = 4.3). However, the goodness ratings for post-test /r/ varied less, ranging from 4.1 to 6 (average = 5.2). The goodness ratings for pre-test English /l/ were

less diverse compared to pre-test /r/, ranging from 3.9 to 6 (average = 5.25). The goodness ratings for post-test /l/ were similarly less diverse, ranging from 4.3 to 6 (average = 5.3), suggesting that Japanese speakers were constantly able to produce /l/.

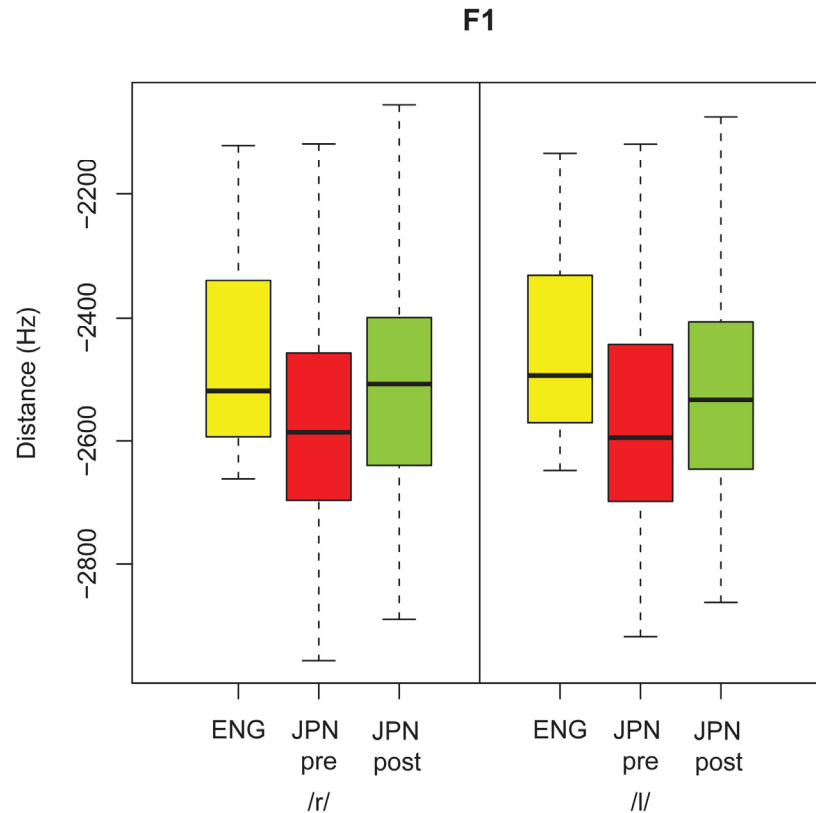
In order to examine whether Japanese speakers' production improvements were also reflected in goodness ratings, paired t-tests were run for pre-/post-test English /r/ and /l/ rating scores. There were significant differences between pre-test and post-test /r/ production,  $t(27) = -4.06, p < 0.001$ , and /l/ production,  $t(27) = -2.65, p < 0.05$ , confirming that Japanese speakers' English /r/ and /l/ pronunciation became more intelligible to British English listeners after the training.

### 4.3.2 Pre/post test English /r/-/l/ production comparison

The results thus demonstrated that Japanese speakers were able to produce more identifiable English /r/ and /l/ syllables after the training. However, it is not clear how exactly Japanese speakers acoustically improved their English /r/-/l/ pronunciation. Japanese speakers' English /r/ and /l/ productions were therefore measured in five acoustic dimensions (i.e., F1, F2, F3, closure duration, and transition duration). The Japanese speakers' pre- and post-test word recordings (38 words x 28 participants x 2 tests = 2128 stimuli) were measured using the same acoustic measurement procedure in Chapter 3. Fifty-one out of 1064 stimuli in the pre-test (i.e., 20 /r/s, and 31 /l/s) had difficult-to-measure F3 frequencies and were omitted. None of the stimuli in the post-test were omitted. All stimuli were normalized by subtracting median F3 frequencies, which were measured using passage recordings. For each Japanese speaker, F1, F2, F3, closure



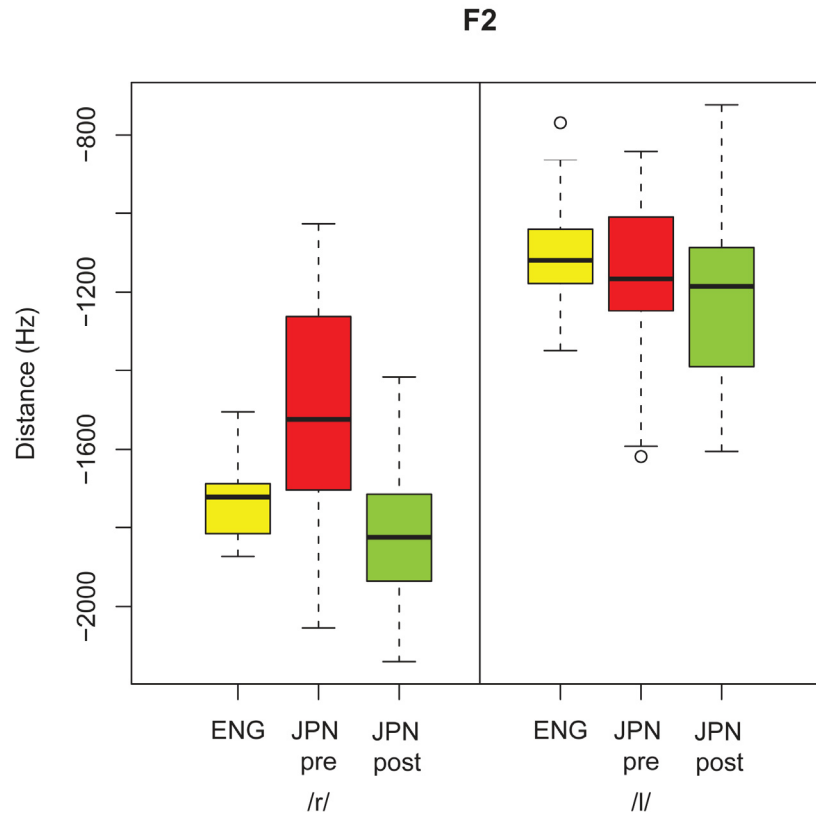
duration and transition duration were separately averaged across words for English /r/ and /l/.



**Figure 4.7** Boxplots of F1 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). The data of English speakers came from Chapter 3. For /r/, Japanese speakers produced similar F1 frequencies before and after the pronunciation training. Standard errors are 44.02, 33.23, and 34.25 (Hz) from the left side. The results demonstrated that Japanese speakers produced higher F1 frequencies for /r/ after the training. The results also demonstrated that such post-test F1 frequencies for /r/ were similar to those of English speakers. For /l/, Japanese speakers produced similar F1 frequencies before and after the training. Standard errors are 41.39, 33.05, and 33.78 (Hz) from the left side. The results demonstrated that Japanese speakers produced higher F1 frequencies after the training. The results also demonstrated that such post-test F1 frequencies for /l/ were similar to those of English speakers.

Figure 4.7 to 4.11 display /r/-/l/ production of English and Japanese speakers in five acoustic dimensions. In order to examine whether Japanese speakers acoustically changed their English /r/ and /l/ pronunciation, five sets of paired t-tests were run; one set of paired t-tests was run for each of the five acoustic dimensions (i.e., F1, F2, F3, closure

duration, and transition duration). Each set included one paired t-test for /r/ production and one paired t-test for /l/ production. Bonferroni correction was done for each set of paired t-tests in order to adjust p-values. The critical p-value was set to 0.01.

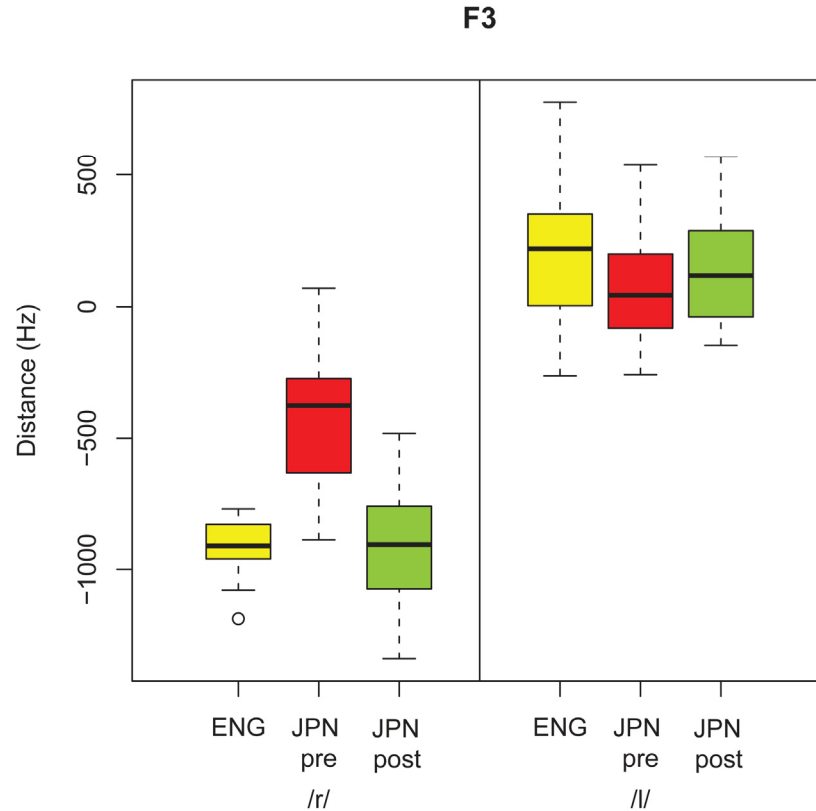


**Figure 4.8** Boxplots of F2 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced high F2 frequencies before the training, but they produced lower F2 frequencies after the training. Standard errors are 27.10, 52.64, and 33.72 (Hz) from the left side. The results confirmed that Japanese speakers lowered F2 frequencies after the training, and that such post-test F2 frequencies were similar to those of English speakers; Japanese speakers improved their /r/ productions in the F2 dimension. For /l/, Japanese speakers produced similar F2 frequencies before and after the training. Standard errors are 40.66, 39.91, and 43.19 (Hz) from the left side. The results confirmed that Japanese speakers produced similar F2 frequencies throughout the experiment, and that the post-test F2 frequencies were similar to those of English speakers.

For F1, there were significant differences for both /r/,  $t(27) = -6.15, p < 0.0001$  and /l/,  $t(27) = -3.96, p < 0.001$ , indicating that Japanese speakers increased F1 for both consonants after the training. For F2, there was a significant difference for /r/,  $t(27) =$

7.36,  $p < 0.0001$ , indicating that Japanese speakers lowered their F2 after the training.

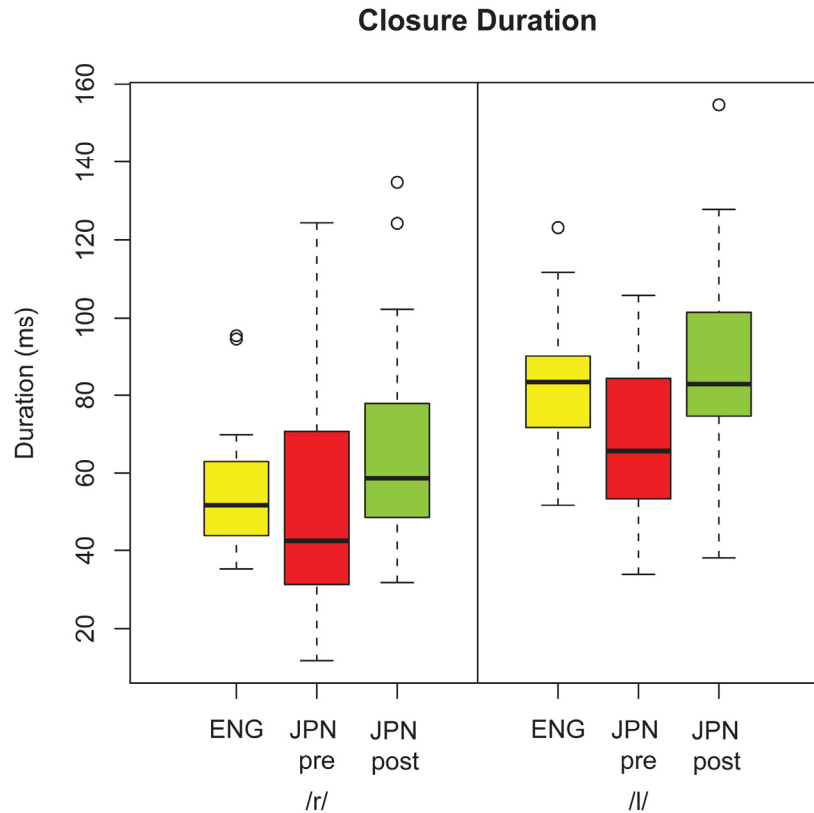
However, there was no significant difference for /l/,  $t(27) = 1.90$ ,  $p > 0.01$ .



**Figure 4.9** Boxplots of F3 frequencies of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced high F3 frequencies before the training, but they produced lower F3 frequencies after the training. Standard errors are 29.32, 48.86, and 41.36 (Hz) from the left side. The results confirmed that Japanese speakers lowered F3 frequencies in the post-test., and that such post-test F3 frequencies were similar to those of English speakers; Japanese speakers improved their /r/ productions in the F3 dimension. For /l/, Japanese speakers produced similar F3 frequencies before and after the training. Standard errors are 74.76, 42.68, and 39.01 (Hz) from the left side. The results confirmed that Japanese speakers produced similar F3 frequencies throughout the experiment, and that the post-test F3 frequencies were similar to those of English speakers.

For F3, there was a significant difference for /r/,  $t(27) = 8.43$ ,  $p < 0.0001$ , indicating that Japanese speakers lowered their F3 after the training. However, there was no significant difference for /l/,  $t(27) = -1.31$ ,  $p > 0.01$ . For closure duration, there was a marginal difference for /r/,  $t(27) = -2.58$ ,  $p = 0.03112$ , suggesting that Japanese speakers may

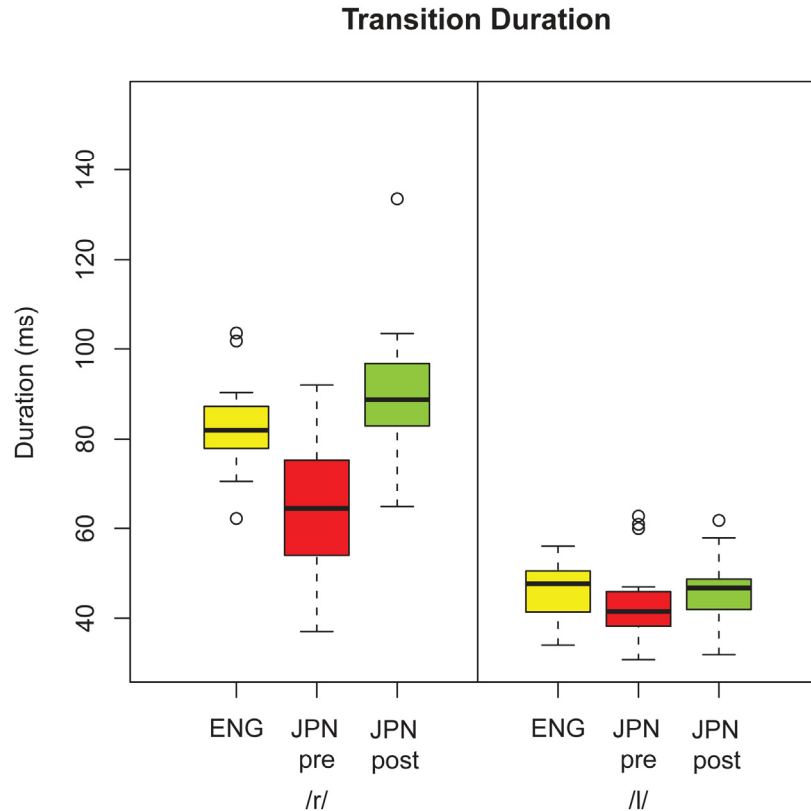
produced longer closure durations after the training. There was a significant difference for /l/,  $t(27) = -5.65, p < 0.0001$ , indicating that Japanese speakers produced longer closure durations after the training.



**Figure 4.10** Boxplots of closure duration of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced slightly longer closure durations in the post-test than they did in the pre-test. Standard errors are 4.72, 5.01, and 4.94 (ms) from the left side. The results demonstrated that Japanese speakers produced similar closure durations throughout the experiment, and that such closure durations were similar to those of English speakers. For /l/, Japanese speakers produced longer closure durations in the post-test than they did in the pre-test. Standard errors are 5.28, 3.62, and 4.46 (ms). The results confirmed that Japanese speakers produced longer closure durations after the training, and that such post-test closure durations were similar to those of English speakers.

For transition duration, there was a significant difference for /r/,  $t(27) = -8.86, p < 0.0001$ , indicating that Japanese speakers produced longer transition durations after the training.

There was no significant difference for /l/,  $t(27) = -2.04, p > 0.01$ .



**Figure 4.11** Boxplots of transition duration of English /r/ and /l/ productions by L1 speakers of English (ENG) and Japanese (JPN). For /r/, Japanese speakers produced short transition durations before the training, but they produced long transition durations after the training. Standard errors are 2.72, 2.83, and 2.54 (ms) from the left side. The results confirmed that Japanese speakers produced longer transition durations in the post-test, and that such post-test transition durations were similar to those of English speakers; Japanese speakers improved their /r/ productions in transition duration. For /l/, Japanese speakers produced similar transition durations before and after the training. Standard errors are respectively 1.66, 1.51, and 1.23 (ms) from the left side. The results confirmed that Japanese speakers produced similar transition durations throughout the experiment, and that the post-test transition durations were similar to those of English speakers.

### 4.3.3 English /r/-/l/ production comparison between Japanese and English speakers

The results demonstrate that Japanese speakers changed their English /r/ production in all five acoustic dimensions, and their /l/ production only in F1 and closure duration. However, it is not clear to what extent their post-test English /r/-/l/ productions became native-like. Therefore, their post-test productions were compared with British

English speakers' /r/-/l/ productions in five acoustic dimensions. Two-way ANOVA analyses tested whether English /r/ and /l/ productions varied with language group for each acoustic dimension. The acoustic values were entered in separate analyses as dependent variables, consonant (i.e., /r/ or /l/) was the within-subject factor, and language group (i.e., Japanese or English) was the between-subject factor. There were no main effects of language group for F2,  $F(1,41) = 0.29, p > 0.05$ , F3,  $F(1,41) = 1.58, p > 0.05$ , closure duration,  $F(1,41) = 1.47, p > 0.05$ , transition duration,  $F(1,41) = 2.13, p > 0.05$ . However, there was a significant main effect of language group for F1,  $F(1,41) = 18.69, p < 0.0001$ , suggesting that Japanese speakers' F1 was slightly higher than English speakers' F1. There was a significant interaction between consonant and language group for F1,  $F(1,41) = 13.17, p < 0.001$ . Simple effects analyses of the interaction revealed that the effects of language group for English /r/,  $t(40) = -0.80, p > 0.05$ , and /l/,  $t(40) = -1.59, p > 0.05$ , were not significant, suggesting that Japanese speakers were similar to English speakers in producing F1 frequencies for /r/ and /l/. These results thus demonstrated that, although there was a significant 2-way interaction between language group and consonant, Japanese speakers did not differ from English speakers in producing F1 frequencies for /r/ and /l/. There was no significant interaction between language group and consonant for the other acoustic dimensions,  $p > 0.05$ , suggesting that their productions were overall very similar to L1 English speakers after training. Unsurprisingly, there were main effects of consonant for F2,  $F(1,41) = 388.88, p < 0.0001$ , F3,  $F(1,41) = 685.20, p < 0.0001$ , closure duration,  $F(1,41) = 61.59, p < 0.0001$ , and transition duration,  $F(1,41) = 393.13, p < 0.0001$ , demonstrating that English /r/ and /l/ productions differ in these dimensions. There was no significant main effect of

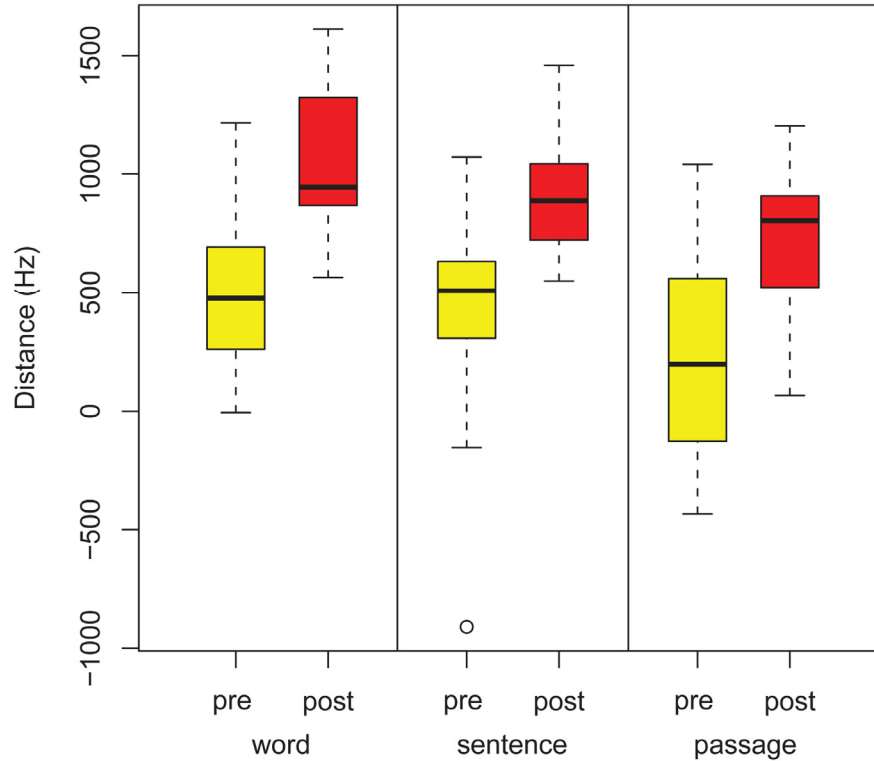
consonant for F1,  $F(1,41) = 1.79$ ,  $p > 0.05$ . The results thus demonstrate that training made their production extremely similar to those of native speakers of English.

#### 4.3.4 English /r/-/l/ F3 contrast in word, sentence, and passage

The results demonstrated that Japanese speakers improved their English /r/-/l/ pronunciations after the training and produced the consonants in a native-like way. That is, pronunciation training at the word level was effective to correct Japanese speakers' pronunciation. However, it is possible that such training could also have effects on English /r/-/l/ productions for more complex materials (e.g., sentences and passages); Japanese speakers could make a clearer contrast of the consonants. Therefore, F3 frequencies of English /r/ and /l/ in sentence and passage recordings were measured.

Japanese speakers' pre- and post-test sentence and passage recordings were measured using the same acoustic measurement procedure in Chapter 3. For accent-revealing sentence recordings, there were 392 tokens (i.e., 7 tokens x 28 participants x 2 tests = 392 tokens). Thirteen out of 196 tokens in the pre-test (i.e., 3 /r/s, and 10 /l/s) did not have clear F3 frequencies and were omitted; five of out 196 tokens in the post-test (i.e., 5 /l/s) were omitted. For passage recordings, there were 624 tokens (i.e., 13 tokens x 24 participants x 2 tests = 624 tokens). The recordings of four participants were missing due to computer problems. Fifteen out of 312 tokens in the pre-test (i.e., 7 /r/s and 8 /l/s) did not have clear F3 frequencies and were omitted; four out of 312 tokens in the post-test (i.e., 2 /r/s and 2 /l/s) were omitted. Each Japanese speaker's English /r/ and /l/ F3 frequencies were separately averaged across tokens, and the F3 frequencies were normalized by subtracting his/her median F3 values. In order to measure the degree of

English /r/-/l/ contrast, the normalized English /r/ F3 was subtracted from the normalized English /l/ F3.



**Figure 4.12** Boxplots of F3 distance between English /r/ and /l/ productions (i.e., F3 contrast between /r/ and /l/) in three different materials (i.e., words, sentences, and passages). Normalized F3 frequencies for /r/ were subtracted from F3 frequencies for /l/. Longer positive distance indicates that Japanese speakers made clearer F3 contrasts between /r/ and /l/. In the pre-test, Japanese speakers made similar F3 contrasts in words and sentences, but they made less clear F3 contrasts in passages. In the post-test, Japanese speakers were able to make clearer /r/-/l/ contrasts in words, sentences, and passages. However, they seemed to make less clear F3 contrasts in passages. Standard errors from the left side of the figure are 60.75, 51.91, 70.84, 41.25, 80.50, and 54.48 (Hz). The results demonstrated that, although Japanese speakers had hard time to produce the /r/-/l/ contrasts in passages, Japanese speakers improved making F3 contrasts between /r/ and /l/ in all three materials after the pronunciation training.

Figure 4.12 displays the distance between English /r/ and /l/ in different materials (i.e., word, sentence, and passage). It seemed that Japanese speakers overall made better English /r/-/l/ contrasts after the pronunciation training. However, the degree of improvement varied among the materials. For example, Japanese speakers seemed to



make similar contrasts in words and sentences, but they seemed to make a less clear contrast in passages. In order to examine whether Japanese speakers improved their English /r/-/l/ F3 in the sentences and passages, a 2-way repeated measure ANOVA was run. The acoustic values (i.e., the degree of English /r/-/l/ F3 contrast) were entered as dependent variables, testing condition (i.e., pre-test or post-test) was the within-subject factor, and material (i.e., word, sentence, or passage) was the between-subject factor. The statistical analysis revealed that there was a significant main effect of testing condition,  $F(1, 127) = 146.44, p < 0.0001$ , confirming that Japanese speakers overall improved their English /r/-/l/ pronunciation after the training. There was a significant main effect of material,  $F(2, 127) = 15.32, p < 0.0001$ , suggesting that Japanese speakers had a harder time making English /r/-/l/ contrasts at the passage level. There was no significant interaction between material and testing condition,  $F(2, 127) = 0.65, p > 0.05$ , suggesting that, although Japanese speakers had a harder time making the contrast in the passage, the amount of learning was not significantly different for the three different types of materials. Tukey HSD comparisons revealed that the English /r/-/l/ contrast in passages was significantly different from the contrast in words,  $z = 3.58, p < 0.01$ , and sentences,  $z = 2.89, p < 0.05$ , demonstrating that Japanese speakers had the hardest time making the English /r/-/l/ contrast in passages. There was no significant difference in English /r/-/l/ contrast between words and sentences,  $z = 0.72, p > 0.05$ .

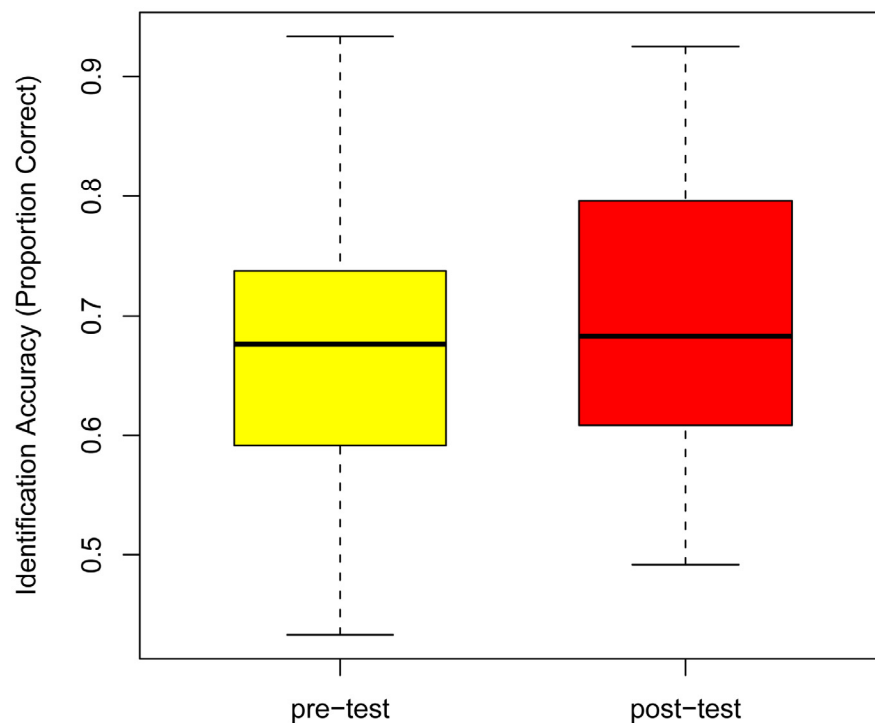
In order to further verify that Japanese speakers had a harder time producing English /r/-/l/ contrasts in passages, and that the amount of learning was not different across the types of materials, 2-way repeated measure ANOVAs were separately run for English /r/ and /l/. The acoustic values (i.e., normalized English /r/ and /l/ F3 frequencies)

were entered as dependent variables, testing condition (i.e., pre-test or post-test) was the within-subject factor, and material (i.e., word, sentence, and passage) was the between-subject factor. For English /r/, there was a main effect of material,  $F(2, 127) = 3.25, p < 0.05$ , confirming that Japanese speakers had a harder time producing /r/ in passages. There was a significant main effect of testing condition,  $F(1, 127) = 129.09, p < 0.0001$ , confirming that Japanese speakers learned better English /r/ pronunciation through the training. There was no significant interaction,  $F(2, 127) = 2.68, p > 0.05$ , confirming that the amount of learning did not differ among the types of materials. Tukey HSD comparisons revealed that there was no significant difference across materials (sentences – passages,  $z = -1.06, p > 0.05$ , words – passages,  $z = -0.06, p > 0.05$ , words – sentences,  $z = 1.04, p > 0.05$ ), demonstrating that Japanese speakers were consistently able to produce similar English /r/ in all types of materials.

For /l/, there was a main effect of material,  $F(2, 127) = 10.57, p < 0.0001$ , confirming that Japanese speakers had a harder time producing /l/ in passages. There was a significant main effect of testing condition,  $F(1, 127) = 9.55, p < 0.01$ , confirming that Japanese speakers learned better English /l/ pronunciation through the training. There was no significant interaction,  $F(2, 127) = 1.92, p > 0.05$ , confirming that the amount of learning did not differ among the types of materials. Tukey HSD comparisons revealed that English /l/ productions in passages were significantly different from /l/ production in words,  $z = 4.54, p < 0.001$ , and sentences,  $z = 2.55, p < 0.05$ , confirming that Japanese speakers had a harder time producing /l/ in passages. There was no significant difference in /l/ production between words and sentences,  $z = 2.08, p > 0.05$ .

### 4.3.5 English /r/-/l/ identification

The production results demonstrated that Japanese speakers improved in English /r/ pronunciation and produced consistent /l/. A previous study, Bradlow et al. (1997) suggested that perceptual training was effective in improving English /r/-/l/ pronunciation. Although participants in the present study received intensive production training, it is possible that such training was effective to improve English /r/-/l/ identification.



**Figure 4.13** Boxplots of English /r/-/l/ identification for the pre-test and post-test. Japanese speakers listened to 60 naturally spoken minimal pairs (i.e., 120 words) including /r/ and /l/ in word-initial positions and identified consonant categories (i.e., /r/ or /l/). Japanese speakers demonstrated similar /r/-/l/ identification accuracies with a wide range of variance in the pre-test and post-test. Standard errors are 0.02 and 0.02 from the left side of the figure. The results confirmed that, although Japanese speakers improved /r/-/l/ production, they did not improve /r/-/l/ identification at all.

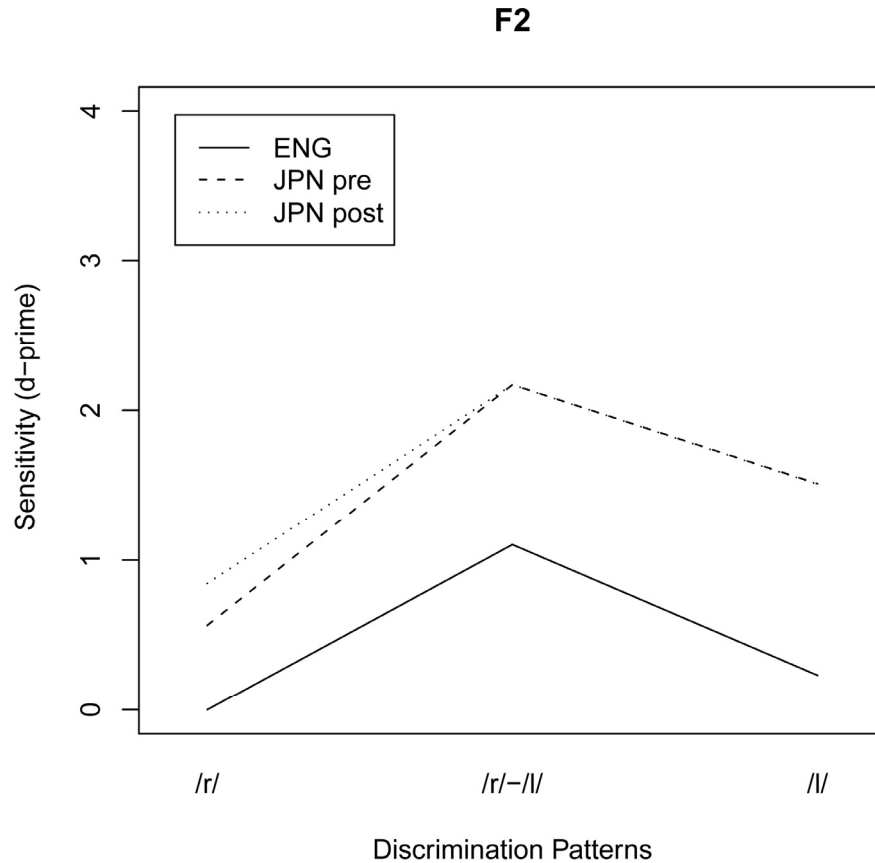
Figure 4.13 displays the results of English /r/-/l/ identification. Japanese speakers demonstrated a wide range of English /r/-/l/ identification (mean = 67.67%) in the pre-test. They similarly demonstrated a wide range of /r/-/l/ identification accuracy (median =

68.33%). In order to examine whether Japanese speakers improved their English /r/-l/ identification, a paired t-test was run on the pre- and post-test English /r/-l/ identification accuracy. The statistical analysis revealed that there was no significant difference between their pre- and post-test English /r/-l/ identification accuracy,  $t(27) = -0.72, p > 0.05$ , suggesting that the pronunciation training did not improve Japanese speakers /r/-l/ identification.

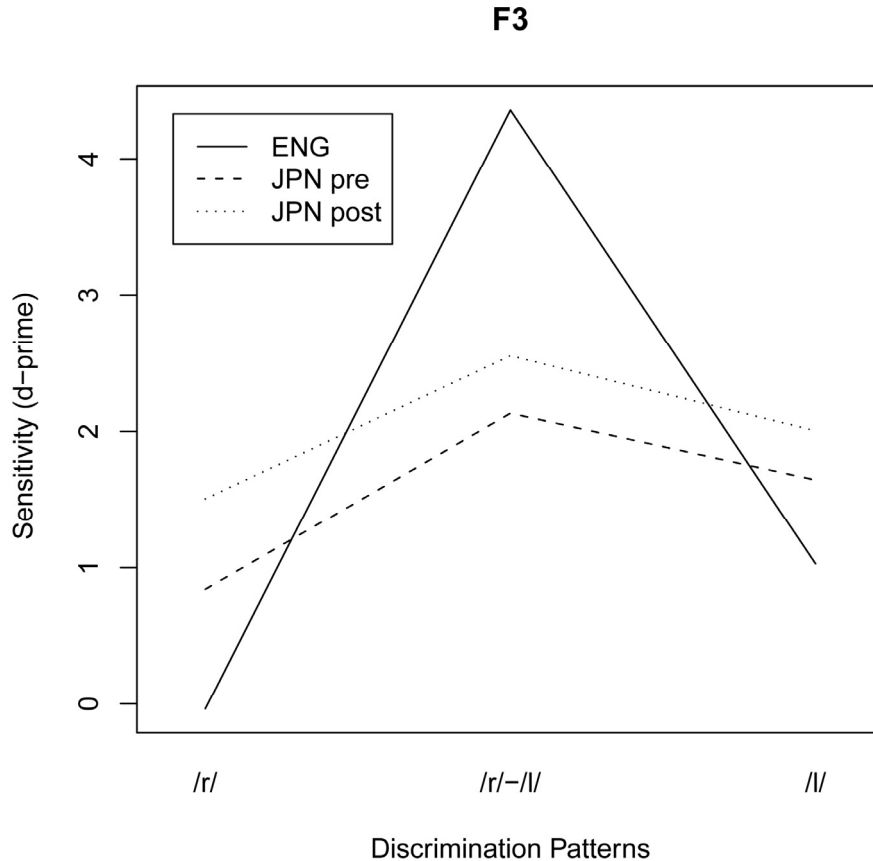
### 4.3.6 English /r/-l/ discrimination

Figure 4.14 displays the results of AX discrimination in the F2 dimension. Japanese speakers' F2 sensitivity at the English /r/-l/ boundary did not change, and it remained higher than their F2 sensitivities within /r/ and /l/ categories. In order to examine whether Japanese speakers changed their F2 sensitivities after the training, a 2-way repeated measure ANOVA was run. F2 sensitivity values were entered as dependent measures, discrimination type (i.e., within-/r/ and within-/l/, or /r/-l/ boundary) was the within-subject factor, and testing condition (pre- or post-test) was the between-subject factor. The statistical analysis demonstrated that there was no main effect on testing condition,  $F(1, 135) = 0.22, p > 0.05$ , confirming that Japanese speakers did not change their F2 sensitivity after the training. There was a main effect of discrimination type,  $F(2, 135) = 18.33, p < 0.0001$ , suggesting that Japanese speakers may have more sensitivity at the /r/-l/ boundary than within English /r/ or /l/ categories. There was no interaction between testing condition and discrimination types,  $F(2, 135) = 0.22, p > 0.05$ . Tukey HSD comparisons revealed that F2 sensitivity at the /r/-l/ boundary was significantly different from F2 sensitivity within the /r/ category,  $z = 4.68, p < 0.001$ . F2 sensitivity

within /r/ category was significantly different from F2 sensitivity within /l/ category,  $z = 2.76$ ,  $p < 0.05$ , suggesting that Japanese speakers had more within category sensitivity for /l/ than did they for /r/. There was no significant difference between F2 sensitivity at the /r/-/l/ boundary and F2 sensitivity within the /l/ category,  $z = -1.92$ ,  $p > 0.05$ .



**Figure 4.14 F2 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. The data of English speakers from Chapter 3 are displayed, but there were no cross-language data analyses because the main purpose of the analyses was to examine whether Japanese speakers change their discrimination sensitivity. Japanese speakers demonstrated similar discrimination sensitivity patterns in the pre-test and post-test; they had high discrimination sensitivity at the /r/-/l/ boundary, and low discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0.20, 0.25, and 0.28 ( $d'$ ) for the pre-test, and 0.21, 0.28, and 0.33 ( $d'$ ) for the post-test. The results demonstrated that Japanese speakers did not change their F2 discrimination sensitivity after the pronunciation training, and that they had similar high discrimination sensitivities at the /r/-/l/ boundary and within the /l/ category.**



**Figure 4.15 F3 discrimination sensitivity within /r/ and /l/ categories and at the English /r/-/l/ boundary for Japanese and English speakers. Japanese speakers demonstrated similar discrimination sensitivity patterns in the pre-test and post-test; they had slightly high discrimination sensitivity at the /r/-/l/ boundary, and slightly low discrimination sensitivity within /r/ category and /l/ category. Standard errors from the left side of the figure are 0.21, 0.30, and 0.51 ( $d'$ ) for the pre-test, and 0.31, 0.36, and 0.62 ( $d'$ ) for the post-test. The results demonstrated that Japanese speakers did not change their F3 discrimination sensitivity after the pronunciation training, and that they had similar high discrimination sensitivity at the /r/-/l/ boundary and within the /l/ category.**

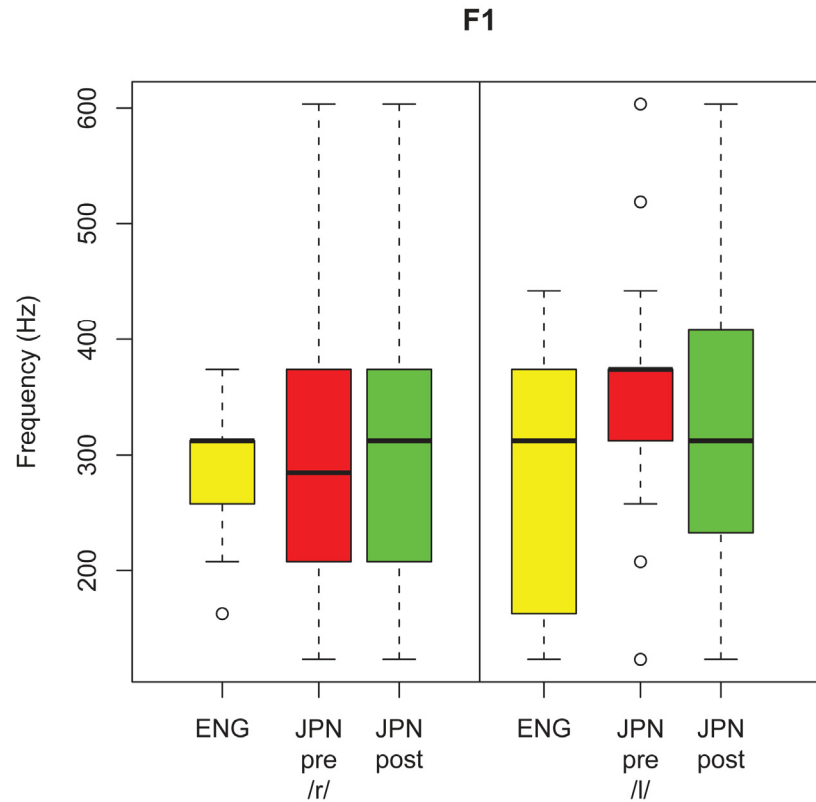
Figure 4.15 displays the results of AX discrimination in the F3 dimension. In the pre-test, Japanese speakers seemed to have slightly higher sensitivity at the /r/-/l/ category boundary compared to sensitivities within the /r/ and /l/ categories. In the post-test, they demonstrated the same pattern of sensitivity. In order to examine whether Japanese speakers changed their F3 sensitivities after the training, a 2-way repeated measure ANOVA was run. F3 sensitivity values were entered as dependent measures,

type of discrimination (i.e., within /r/ and /l/, and /r/-/l/ boundary) was the within-subject factor, and testing condition (pre- or post-test) was the between-subject factor. The statistical analysis demonstrated that there was no main effect of testing condition,  $F(1, 135) = 2.39, p > 0.05$ , confirming that Japanese speakers did not change their F3 sensitivity after the training. There was a main effect of discrimination type,  $F(2, 135) = 4.71, p < 0.05$ , suggesting that Japanese speakers had higher F3 sensitivity at the English /r/-/l/ boundary. There was no interaction between testing condition and types of discrimination,  $F(2, 135) = 0.09, p > 0.05$ . Tukey HSD comparisons revealed that F3 sensitivity at the /r/-/l/ boundary was significantly different from F3 sensitivity within the /r/ category,  $z = 2.39, p < 0.05$ . However, F3 sensitivity at the /r/-/l/ boundary was not significantly different from F3 sensitivity within the /l/ category,  $z = -0.90, p > 0.05$ , and F3 sensitivity within the /l/ category was not significantly different from one within the /r/ category,  $z = 1.49, p > 0.05$ .

### 4.3.7 Perceptual mapping of best exemplars

Figures 4.16 to 4.20 displays the perceptual best exemplar of English /r/ and /l/ by Japanese and English speakers. The results suggested that Japanese speakers did not change their English /r/ and /l/ best exemplars in most of the acoustic dimensions after the training; their F1, F2, closure duration, and transition duration looked similar. However, the results suggested that Japanese speakers may have changed their /r/ F3 and closure duration after the training; they chose lower F3 and shorter closure duration with less variability. In order to examine whether Japanese speakers changed their perceptual best exemplars, 2-way repeated-measures ANOVAs were run separately for each

acoustic dimension. The acoustic values were entered in separate analyses as dependent variables, consonant (/r/ or /l/) was treated as a between-subject factor, and testing condition (pre-test or post-test) was treated as a within-subject factor.

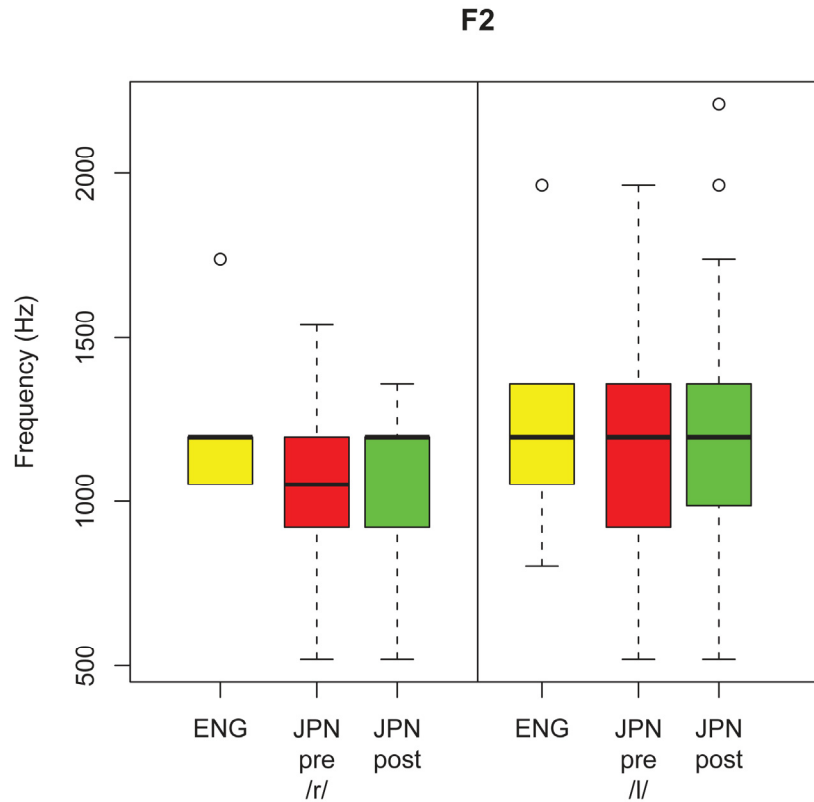


**Figure 4.16** Boxplots of F1 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). The data of English speakers from Chapter 3 is presented, but there was no comparison between English and Japanese speakers in this analysis because the focus of the analysis was to examine whether Japanese speakers change their best exemplars after the pronunciation training. For /r/, Japanese speakers selected similar F1 frequencies in the pre-test and post-test. Standard errors are 24.38 (pre-test) and 20.58 (post-test) Hz. The results confirmed that Japanese speakers did not change F1 frequencies of best exemplars for /r/ after the pronunciation training. For /l/, Japanese speakers chose similar F1 frequencies in the pre-test and post-test. Standard errors are 19.95 (pre-test) and 25.49 (post-test) Hz. The results demonstrated that Japanese speakers did not change F1 frequencies of best exemplars for /l/ after the training.

For F1, there were no main effects of testing condition,  $F(1,81) = 0.27, p > 0.05$ , and consonant,  $F(1,81) = 3.77, p > 0.05$ . There was no significant interaction,  $F(1,81) = 0.96, p > 0.05$ . For F2, there was no main effect of testing condition,  $F(1,81) = 0.15, p >$

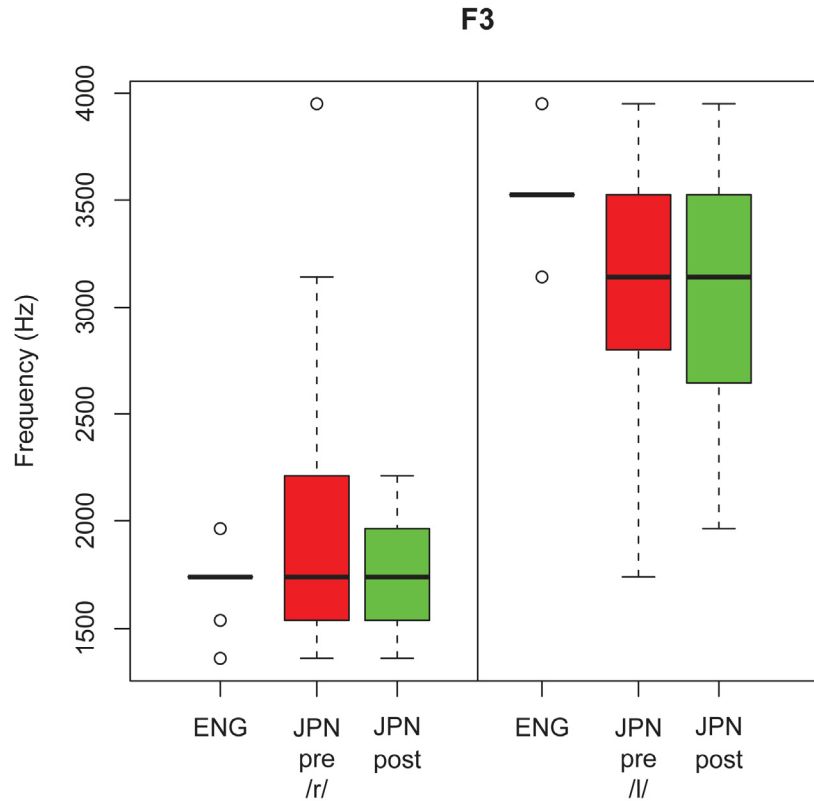


0.05. There was a main effect of consonant,  $F(1,81) = 7.63, p < 0.01$ , suggesting that Japanese speakers separated English /r/ and /l/. There was no significant interaction,  $F(1,81) = 0.39, p > 0.05$ .



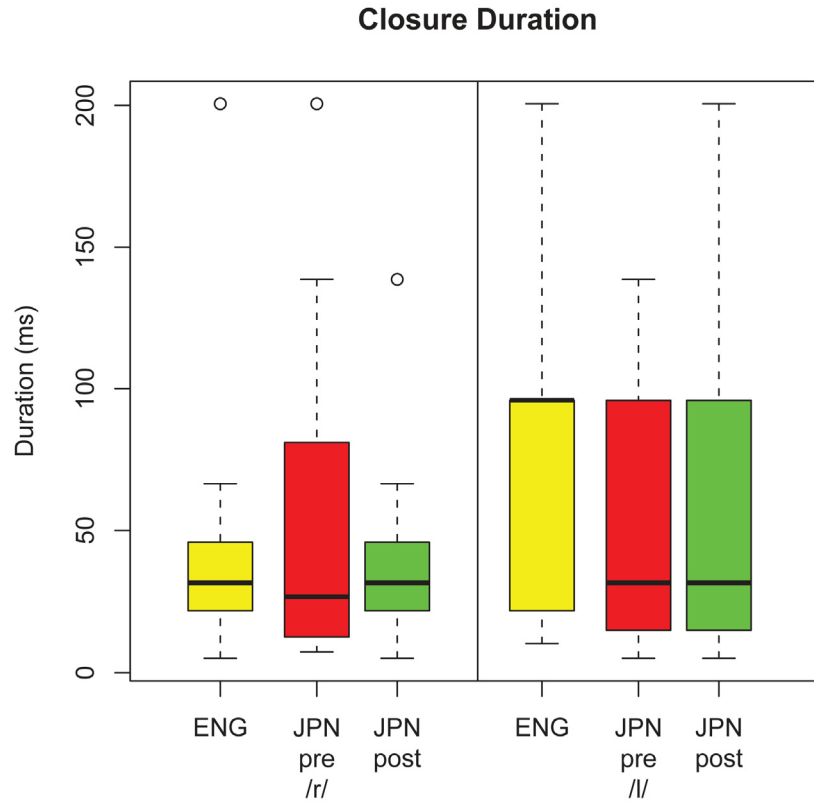
**Figure 4.17** Boxplots of F2 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers selected similar F2 frequencies in both pre-test and post-test. Standard errors are 50.25(pre-test) and 45.63 (post-test) Hz. For /l/, likewise, they selected similar F2 frequencies for /l/ in both pre-test, post-test. Standard errors are 67.04 (pre-test) and 77.6 (post-test) Hz. The results demonstrated that Japanese speakers did not change F2 frequencies of best exemplars for /r/ and /l/ after the training.

For F3, there was no main effect of testing condition,  $F(1,81) = 0.82, p > 0.05$ . There was a main effect of consonant,  $F(1,81) = 158.07, p < 0.001$ , confirming that Japanese speakers separated /r/ and /l/ in the F3 dimension. There was no significant interaction,  $F(1,81) = 1.93, p > 0.05$ , suggesting that, despite the fact that there was less variability in /r/ F3, Japanese speakers did not change their best exemplars.

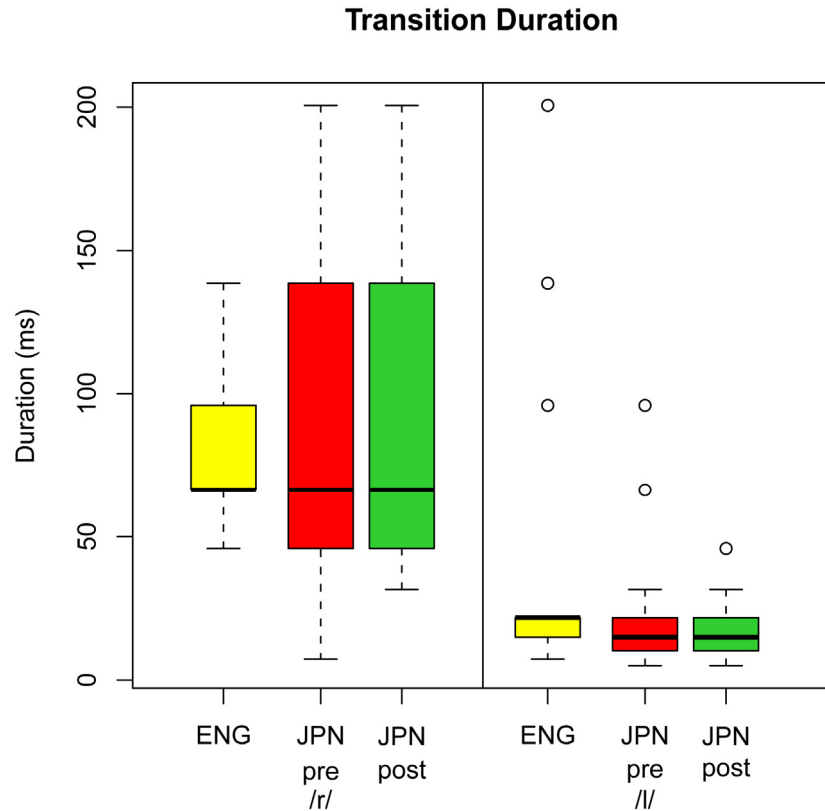


**Figure 4.18** Boxplots of F3 frequencies of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose similar F3 frequencies with a wide range of variance in the pre-test and less variance in the post-test. Standard errors are 111.82 (pre-test) and 52.60 (post-test) Hz. For /l/, Japanese speakers chose similar F3 frequencies with a wide range of variance in the pre-test and post-test. Standard errors are 109.12 (pre-test) and 124.15 (post-test) Hz. The results demonstrated that Japanese speakers did not change F3 frequencies of best exemplars for /r/ and /l/ after the training.

For closure duration, there was no main effect of testing condition,  $F(1,81) = 0.96, p > 0.05$ . There was no main effect of consonant,  $F(1,81) = 0.57, p > 0.05$ . There was no significant interaction,  $F(1,81) = 2.01, p > 0.05$ . For transition duration, there was no main effect of testing condition,  $F(1,81) = 0.03, p > 0.05$ . There was a main effect of consonant,  $F(1,81) = 85.83, p < 0.001$ , suggesting that Japanese speakers separated English /r/ and /l/; longer duration for /r/ and shorter duration for /l/. There was no significant interaction,  $F(1,81) = 0.15, p > 0.05$ .



**Figure 4.19** Boxplots of closure duration of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose short closure durations with a wide range of variance in the pre-test and less variance in the post-test. Standard errors are 11.12 (pre-test) and 5.24 (post-test) ms. For /l/, Japanese speakers chose similar closure durations with similar variance in the pre-test and post-test. Standard errors are 8.55 (pre-test) and 9.41 (post-test) ms. The results demonstrated that Japanese speakers did not change closure durations of best exemplars for /r/ and /l/ after the training.



**Figure 4.20** Boxplots of transition duration of the best exemplars for /r/ and /l/ by L1 speakers of English and Japanese (JPN). For /r/, Japanese speakers chose similar transition durations with a wide range of variance in the pre-test and post-test. Standard errors are 10.08 (pre-test) and 10.33 (post-test) ms. For /l/, they chose similar short transition durations with little variance in the pre-test and post-test. Standard errors are 3.56 (pre-test) and 1.70 (post-test) ms. The results demonstrated that Japanese speakers did not change transition durations of best exemplars for /r/ and /l/ after the training.

### 4.3.8 Links between English /r/-/l/ production and perception

The results thus demonstrated that Japanese speakers did not change their English /r/-/l/ identification accuracy, F2 and F3 sensitivities, and that they generally did not change perceptual best exemplars, despite the fact that they improved English /r/ pronunciation and produced consistent /l/. However, it is possible that individual differences on these measures are still correlated. In order to examine whether the amount of change in production and the amount of change in perception are related, Pearson

correlation analyses were used to test whether the individual differences were correlated. The amount of change in production and perception was calculated by subtracting the pre-test values from post-test values. For example, if English speakers recognized a Japanese speaker's /r/-/l/ productions 90 % of the time in the post-test and 70 % of the time in the pre-test, the amount of change was 20 %.

There was no significant correlation between the amount of change in /r/-/l/ production accuracy (i.e., /r/-/l/ production intelligibility by English speakers) and the amount of change in Japanese speakers' /r/-/l/ identification accuracy,  $r = 0.12$ ,  $p > 0.05$ , suggesting that Japanese speakers who improved their English /r/-/l/ production did not necessarily improve their English /r/-/l/ identification. There were no significant correlations between the amount of change in /r/-/l/ production accuracy and change in any of the F2 and F3 discrimination sensitivity measurements, F2 peak sensitivity,  $r = -0.09$ , F2 average sensitivity,  $r = 0.11$ , F3 peak sensitivity,  $r = 0.21$ , F3 average sensitivity,  $r = -0.26$ ,  $p > 0.05$ . There were no significant correlations between the amount of change in /r/-/l/ production accuracy and changes in best exemplars in the five acoustic dimensions, F1,  $r = 0.16$ , F2,  $r = 0.01$ , F3,  $r = 0.16$ , closure duration,  $r = -0.03$ , transition duration,  $r = 0.14$ ,  $p > 0.05$ .

The same analyses were conducted separately for /r/ and /l/. For the amount of change in /r/ production accuracy, there was no significant correlation with English /r/ identification accuracy,  $r = 0.32$ ,  $p > 0.05$ . There were no significant correlations with F2 and F3 within-/r/ category sensitivity, F2 within-/r/ sensitivity,  $r = 0.06$ , F3 within-/r/ sensitivity,  $r = -0.04$ ,  $p > 0.05$ . There were no significant correlations with changes in English /r/ best exemplar, F1,  $r = 0.36$ , F2,  $r = -0.01$ , F3,  $r = -0.01$ , closure duration,  $r =$

-0.07, transition duration,  $r = 0.37$ ,  $p > 0.05$ . For the amount of change in /l/ production accuracy, there was no significant correlation with English /l/ identification accuracy,  $r = 0.08$ ,  $p > 0.05$ . There were no significant correlations with F2 and F3 within-/l/ category sensitivity, F2 within-/l/ sensitivity,  $r = -0.17$ , F3 within-/l/ sensitivity,  $r = -0.04$ ,  $p > 0.05$ . There were no significant correlations with changes in English /l/ best exemplar, F1,  $r = 0.29$ , F2,  $r = 0.03$  F3,  $r = -0.06$ , closure duration,  $r = 0.26$ , transition duration,  $r = -0.02$ ,  $p > 0.05$ .

#### 4.3.9 Predictors for English /r/-/l/ production improvement

The results demonstrated that English /r/-/l/ production learning did not lead to English /r/-/l/ perceptual learning at all; pronunciation training was exclusively effective for production learning. However, it may be possible that Japanese speaker's pre-training perceptual knowledge and production ability can be predictive of their English /r/-/l/ production learning. Pearson correlation analyses were run between the amount of English /r/-/l/ production learning (i.e., improvement in /r/-/l/ production intelligibility by English speakers) and pre-training perception and production measurements (i.e., English /r/-/l/ identification accuracy, F2 and F3 sensitivities, and /r/-/l/ production intelligibility by English speakers).

There was a very high correlation between the amount of English /r/-/l/ production learning and their pre-test /r/-/l/ production intelligibility,  $r = -0.98$ ,  $p < 0.001$ , demonstrating that Japanese speakers who had poorer English /r/-/l/ production ended up learning English /r/-/l/ pronunciation more. There was also a significant correlation

between the amount of English /r/-l/ production learning and pre-test English /r/-l/ identification accuracy,  $r = -0.44$ ,  $p < 0.05$ , suggesting that Japanese speakers who had poorer English /r/-l/ identification accuracy ended up learning English /r/-l/ pronunciation more, despite the fact that they did not improve in their English /r/-l/ identification accuracy. There were no significant correlations between the amount of English /r/-l/ production learning and F2 and F3 discrimination sensitivities, F2 peak sensitivity,  $r = 0.23$ , F2 average sensitivity,  $r = -0.20$ , F3 peak sensitivity,  $r = -0.20$ , F3 average sensitivity,  $r = 0.28$ ,  $p > 0.05$ . There were no significant correlations between the amount of English /r/-l/ production learning and best exemplar accuracies, F1,  $r = -0.17$ , F2,  $r = 0.14$ , F3,  $r = -0.03$ , closure duration,  $r = 0.23$ , transition duration,  $r = 0.07$ ,  $p > 0.05$ .

The same analyses were also conducted separately for /r/ and /l/. There was a strong correlation between the amount of English /r/ production learning and pre-test /r/ production intelligibility,  $r = -0.99$ ,  $p < 0.001$ , demonstrating that Japanese speakers who had poorer English /r/ production ended up learning English /r/ pronunciation more. There was also a significant correlation between the amount of English /r/ production learning and English /r/ identification accuracy,  $r = -0.44$ ,  $p < 0.05$ , suggesting that Japanese speakers who had poorer English /r/ identification accuracy ended up learning English /r/ pronunciation more, despite the fact that they did not improve their English /r/ identification accuracy. There were no significant correlations between the amount of English /r/ production learning and F2 and F3 within-/r/ category sensitivity, F2 sensitivity,  $r = -0.19$ , F3 sensitivity,  $r = 0.17$ ,  $p > 0.05$ . There were no significant correlations between the amount of English /r/ production learning and /r/ best exemplar

in four acoustic dimensions, F1,  $r = -0.19$ , F2,  $r = -0.19$ , F3,  $r = -0.18$ , closure duration,  $r = -0.05$ ,  $p > 0.05$ . There was, however, a significant correlation between the amount of English /r/ production learning and the best exemplar transition duration,  $r = -0.38$ ,  $p < 0.05$ , suggesting that Japanese speakers who preferred shorter transition duration before the training ended up learning English /r/ pronunciation more.

There was a strong correlation between the amount of English /l/ production learning and pre-test /l/ production intelligibility,  $r = -0.82$ ,  $p < 0.001$ , demonstrating that Japanese speakers who had poorer English /l/ production ended up learning English /l/ pronunciation more. There was also a significant correlation between the amount of English /l/ production learning and English /l/ identification accuracy,  $r = -0.38$ ,  $p < 0.05$ , suggesting that Japanese speakers who had poorer English /l/ identification accuracy ended up learning English /l/ pronunciation more, despite the fact that they did not improve their English /l/ identification accuracy. There were no significant correlations between the amount of English /l/ production learning and F2 and F3 within-/l/ category sensitivity, F2 sensitivity,  $r = 0.07$ , F3 sensitivity,  $r = 0.05$ ,  $p > 0.05$ . There were no significant correlations between the amount of English /l/ production learning and /l/ best exemplar in all five acoustic dimensions, F1,  $r = -0.32$ , F2,  $r = -0.07$ , F3,  $r = -0.11$ , closure duration,  $r = -0.16$ , transition duration,  $r = 0.03$ ,  $p > 0.05$ .

#### **4.4 General discussion**

The present study examined whether Japanese speakers improve English /r/-/l/ production and perception after receiving one-to-one pronunciation training. The results demonstrated that there was substantial improvement of English /r/ productions and slight



improvement of /l/ productions at the word level by adult Japanese speakers; the production intelligibility of /r/ by English speakers improved by an average of 31.3 percentage points (i.e., pre-test 64.4%, post-test 95.7%), and the production intelligibility of /l/ by English speakers improved by an average of four percentage points (i.e., pre-test, 92%, post-test, 96%). This confirms that pronunciation training is particularly effective for promoting accurate English /r/ productions, and that Japanese speakers are already able to produce accurate /l/ even before the training. Such improved /r/ and /l/ productions of Japanese speakers turned out to be similar to the /r/-/l/ productions of British English speakers. The results also demonstrated that there was significant improvement of /r/ and /l/ productions at the sentence and passage levels; Japanese speakers produced similar degrees of /r/-/l/ F3 contrasts at these two levels as they did at word level. This confirms that the effects of pronunciation training generalized to /r/-/l/ productions in continuous speech.

Such successful /r/-/l/ production learning provided strong evidence that the present training procedures are effective for Japanese speakers. Explicit instructions and feedback related to tongue position and shape, and lip shape enhanced Japanese speaker's attention to their articulatory movements. That particularly led to the spectral improvement in /r/ productions; Japanese speakers lowered F2 and F3 for English /r/. Explicit instruction and feedback related to temporal aspects of English /r/ and /l/ enhanced Japanese speaker's attention to temporal aspects of their pronunciation; Japanese speakers increased closure duration for both /r/ and /l/, and transition duration for /r/ after the training. L2 learners seem capable of learning details of non-native segments (e.g., articulatory movements and temporal information) as long as specialists

(e.g., phoneticians, teachers) orient the L2 learner's attention to specific aspects of L2 production. All Japanese speakers demonstrated such capability in learning L2 segments in the present study, but Japanese speakers who have poor knowledge of /r/-/l/ perception and production particularly learned better through the training. This may suggest that, if L2 learners receive explicit instruction at an early stage of learning, this will prevent L2 learners from establishing erroneous phonetic categories and articulatory movements, and lead to quick improvement in L2 phoneme learning.

The Japanese speaker's ability to acquire native-like English /r/ and /l/ production through pronunciation training confirms that L2 learners maintain their ability to learn L2 speech production, which is in accord with the claim of SLM (Flege, 1995) that language learning mechanisms and processes remain intact even in adulthood. However, the present results seem problematic for the claim of SLM (Flege, 1999) that the degree of accuracy in L2 segment perception limits the degree of accuracy in L2 segment production; despite the fact that Japanese speakers become capable of producing native-like English /r/ and /l/, they did not improve /r/-/l/ perceptual accuracies. It may be in natural settings that the degree of accuracy in L2 perception limits the degree of accuracy in L2 production or vice versa. Chapter 3 tested Japanese speakers without any training and certainly provided evidence supporting the view of Flege (1999). However, training settings may differ from natural settings, in that the degree of L2 perceptual accuracy may not put any limits on learning L2 production or vice versa, and as a result it looks like the degree of perceptual accuracy does not limit the degree of production accuracy.

Why did Japanese speakers improve their /r/-/l/ production, but not /r/-/l/ perception? One possible explanation for such unchanged perceptual behavior is that

Japanese speakers did not have sufficient listening to modify their perceptual knowledge and develop /r/-/l/ phonetic categories. The present study strictly limited the amount of listening through the entire training. Japanese speakers listened to their original and enhanced recordings, the pronunciation of the instructor when they repeated after him, and possibly their bone-conducted pronunciation. Even if such amount of listening was sufficient, Japanese speakers received impoverished stimuli, in that talker and stimulus variability was low (i.e., two talkers, and three minimal-pair words). This may be another reason why Japanese speakers did not change their perceptual knowledge after training.

The other possible explanation is that the type of training might interfere with /r/-/l/ perceptual learning of Japanese speakers. Japanese speakers did not have any across-category pronunciation training. They were instructed to focus on either /r/ or /l/, and pronounce and listen to three words (e.g., *rack*, *rick*, and *room*). That is, Japanese speakers perceived phonetic variations within either the /r/ or /l/ category all the time. Previous studies (e.g., Maye and Gerken, 2000; Maye et al., 2002) demonstrated that the types of distributional patterns of stimuli can change phonetic category learning. For example, Maye and Gerken (2000) examined whether differences in distributional patterns of stimuli affect the discrimination accuracy of English voiceless unaspirated /t/ and voiced /d/, which English speakers perceive as members of the /d/ category. They manipulated an 8-step continuum of /ta/-/da/. Two groups of English speakers received the same amount of exposure to the stimuli, but one group of English speakers (i.e., bimodal group) had more exposure to the stimuli from near the endpoints of the continuum (i.e., stimuli 2 and 7) and the other group (i.e., unimodal group) had more exposure to the stimulus in the center of the continuum (i.e., stimuli 4). The results

demonstrated that the bimodal group had better discrimination of the stimuli after training, suggesting that the types of distributional patterns of stimuli in the laboratory setting can promote or interfere with L2 phoneme learning. It is plausible that, if Japanese speakers were instructed to pronounce and listen to /r/-/l/ minimal pairs in the present study (i.e., bimodal learning), they would be able to improve at least /r/-/l/ identification accuracy. Future studies need to examine how the types of pronunciation training (i.e., unimodal or bimodal training) change /r/-/l/ production and perceptual accuracies.

Previous perceptual training studies (i.e., Bradlow et al, 1997, Rochet, 1995; Wang et al., 2003) demonstrated a transfer of perceptual learning to improvement to speech production, but the present study did not find the reverse direction of transfer. It is possible that such transfer mechanisms are available only from perceptual learning to production learning even if speech perception and production are closely related. L1 speech perception models (e.g., NLM-e; Kuhl et al., 2008, PAM; Best, 1995, PRIMIR; Werker and Curtin, 2005) hypothesize that infants gain perceptual knowledge and establish phonetic categories before they have accurate L1 speech production, implying that infants may use and transfer perceptual learning to tune their L1 speech production. It may be natural for even adult L2 learners who have already developed motor skills for speech production to gain perceptual knowledge first. This may lead to improvements in L2 speech production.

The other explanation may be that the effects of pronunciation training may transfer to improvement in L2 perceptual learning, but such transfer effects may take some time to be seen in the L2 learner's perceptual behavior. In other words, there may not be common mental representations underlying for both speech perception and

production. Rather, there may be somewhat independent mental representations for speech perception and production respectively, and L2 learners may have to build up networks linking perception and production. This may be the reason why the amount of perceptual learning and the amount of production learning are poorly correlated (e.g., Bradlow et al., 1997), and the reason why the pronunciation training in the present study did not lead to improvement in /r/-/l/ perception at all. Regarding the applied goal of L2 phoneme learning in perception and production, it thus seems that using both perceptual and production training may be the best option for L2 learners to achieve native-like speech perception and production.

The present study clearly demonstrates for language teachers that technical /r/-/l/ pronunciation training helps L2 learners. It would be difficult for language teachers to acquire some of the technical skills used here (e.g., acoustically enhanced speech) to orient the L2 learner's attention to specific aspects of the consonants. Language teachers are usually required to take an introductory phonetics course, but this is not sufficient to gain knowledge of English /r/-/l/ (e.g., teachers will not typically learn the F3 difference between the consonants). In addition, language teachers are required to teach second or foreign languages for the sake of communication and exams; L2 segment teaching is trivial to most teachers. What the present study offers for teachers are the acoustic details of word-initial /r/ and /l/ production by English and Japanese speakers (i.e., F1, F2, F3, closure duration, and transition duration), and experimental teaching methods using spectrograms. If teachers are required to teach /r/-/l/ pronunciation, they can show their students how /r/ and /l/ are different using spectrograms (e.g., lower F3 for /r/, higher F3 for /l/, shorter transition duration for /l/, and longer transition duration for /r/). Adult

Japanese speakers were convinced with the differences in visual information (e.g., shape of F3 line); adult learners like some concrete evidence to understand acoustic differences between /r/ and /l/. Teachers can, then, use the experimental teaching instructions in this study using real-time spectrograms, make recordings of students, and let students compare their pronunciations to ones by native speakers of English, instead of using enhanced recordings. Participants were fond of visually checking their productions, and that they gained confidence when they perceived similarities between original and enhanced recordings. For teaching purposes, even comparisons between recordings of students and English speakers may suffice.

## Chapter 5: General Conclusion

Some speech scientists and experimental phoneticians have investigated perception and production of English /r/-/l/ by Japanese speakers and examined why L2 speakers have difficulties learning L2 phonemes and whether the learners can be trained to overcome such difficulties. This dissertation initially investigated whether category assimilation processes between L1 and L2 phonetic categories are the cause for the problems of Japanese speakers. Japanese speakers demonstrated an asymmetric pattern of category assimilation for the /r/-/l/ contrast; they assimilated /l/ into Japanese /r/ category more strongly than they assimilated /r/. However, such a strong category assimilation process was not predictive of /r/-/l/ identification accuracy.

This result may suggest that category assimilation processes are not the direct cause of L2 speech perception difficulties even though current L2 speech perception models (e.g., PAM; Best, 1995; SLM; Flege, 1995) hypothesized that the phonetic similarity between L1 and L2 phonemes causes L2 phoneme learning problems. If category assimilation does not explain why L2 speakers have difficulties learning L2 phonemes, what else can explain such difficulties? The present results suggest that the perceptual difficulty is specifically localized to the F3 dimension. Chapter 2 demonstrated that individuals who have difficulty identifying /r/ and /l/ have less accurate best exemplars in the dimension. Chapter 3 further demonstrated that individual differences in F3 discrimination sensitivity at the /r/-/l/ boundary is correlated with F3 performance. It is plausible that some kind of auditory processes related to F3 frequencies may make /r/-/l/ learning difficult for adult Japanese speakers. However, it is

still unclear why Japanese speakers have difficulty with F3. This result is likely a factor specific to English /r-/l/, in that it probably cannot explain all types of L2 learning difficulties; category assimilation may explain the difficulties with other phonemes even though it does not explain this particular example. Causes of L2 phoneme learning problems may thus need to be examined case by case, because a single explanation may not apply to all situations.

Regardless of what exactly causes L2 perceptual and production problems, it is a fact that there are some L2 speakers who can have native-like perception accuracy. Chapter 2 and 4 demonstrated that some Japanese speakers can identify English /r-/l/ in the same way as English speakers. It is also a fact that there are some L2 speakers who can produce native-like L2 sounds. Chapter 3 demonstrated that some Japanese speakers can produce /r/ and /l/ in the same way as English speakers. Previous studies (e.g., Flege et al., 1995; Flege et al., 2003; MacKay et al., 2001, Yamada, 1995) suggested that such individual differences may stem from age of arrival and length of residence in L2 speaking environments. It will be interesting to examine whether such factors are related to the perceptual and production measures in this dissertation. If these factors do not explain the individual differences, researchers need to consider other possible factors to explain the diverse perceptual and production performance among Japanese speakers.

Although some Japanese speakers have difficulties identifying English /r-/l/, previous studies demonstrated that Japanese speakers can be trained to be able to identify the L2 segments better (e.g., Hazan et al., 2005; Iverson et al., 2005; Lively et al., 1993, Logan et al., 1991; Uther et al., 2008; Wang et al., 1999). Chapter 4 demonstrated that it is also possible that L2 speakers can be trained to be able to produce L2 segments in a



native-like way; adult Japanese speakers became capable of producing native-like English /r/ and /l/ after one-to-one pronunciation training. However, Japanese speakers did not improve their /r/-/l/ perception accuracy at all. Given that /r/-/l/ production was not related to /r/-/l/ discrimination sensitivity in Chapter 3 and that production training did not change Japanese speakers' /r/-/l/ perception in Chapter 4, it seems fair to assume that there are somewhat separated mental representations for speech perception and production. If so, what researchers have to do is to identify how these mental representations are linked and what kinds of training can facilitate transfer between the two domains. What sounds reasonable at this moment is to use both perceptual and production training. This is perhaps the only way to let L2 speakers improve both L2 perception and production.

Future research could also try to use different kinds of approaches to production training. Chapter 4 examined production training using minimal-pair words. This approach was effective for teaching Japanese speakers to produce native-like /r/ and /l/ in words, sentences, and passages. But, it is unknown whether such an approach is also effective for Japanese speakers producing the consonants in spontaneous speech. In order to further enhance Japanese speakers' /r/-/l/ production ability, it may be necessary to consider some additional training. For example, it may be helpful to use some tongue twisters in training. This may give good opportunities for Japanese speakers to master accurate tongue configurations in difficult continuous speech. It may be also helpful to use conversation tasks (e.g., telling directions), which include some /r/-/l/ words. Recording such tasks and letting Japanese speakers listen to their own speech along with feedback from instructors may lead to improved /r/-/l/ production in spontaneous speech.

The present one-to-one pronunciation training involved many interactions between the instructor and Japanese speakers. This interaction in L2 settings may be a new area of research. Native Language Magnet extended (NLM-e; Kuhl et al., 2008) suggests that social interaction (e.g., infant-directed speech) is very important for infants for developing their perceptual ability. It may likewise be the case that L2 learners need sufficient social interaction in order to develop L2 phonetic categories. It would be interesting to acoustically analyze foreigner talk or teacher talk and examine the distributional patterns of phonetic units. For example, it may be possible that English speakers and teachers who know that Japanese speakers have /r/-/l/ problems may exaggerate the F3 formant or the transition for /r/ in conversational or classroom contexts. If they do so, it would be possible to examine whether such exaggerated speech facilitates Japanese speakers' /r/-/l/ identification accuracy.

In conclusion, this dissertation provided a detailed examination of the perception and production of English /r/ and /l/ by adult Japanese speakers. The results challenged previous findings, but generated more questions than answers. Further work is necessary to understand the origin of the learning difficulties of Japanese speakers, how they can best be trained, and the link between perception and production.

## References

- Alwan, A., Narayanan, S., & Haker, K. (1997). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II : The rhotics," *Journal of the Acoustical Society of America*, **101**, 1078-1089.
- 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 /l/," *Journal of Phonetics*, **32**, 233-250.
- Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Percy, A. J. (1981). "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," *Child Development*, **52**, 1135-1145.
- Best, C. T. (1994). "The emergence of native-language phonological influence in infants: a perceptual assimilation hypothesis," in *The development of speech perception: the transition from speech sounds to spoken words*, edited by J. C. Goodman & H. C. Nusbaum, (Cambridge, MA: MIT press), pp. 1676-224.
- Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171-204.
- Best, C. T., & McRoberts, G. W. (2003). "Infant perception of non-native consonant contrasts that adults assimilate in different ways," *Language and Speech*, **46**, 183-216.
- Best, C. T., & Strange, W. (1992). "Effects of phonological and phonetic factors on cross-language perception of approximants," *Journal of Phonetics*, **20**, 305-330.

- Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," *Journal of the Acoustical Society of America*, **109**, 775–794.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 345-360.
- Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). "Divergent developmental patterns for infants' perception of two nonnative consonant contrasts," *Infant Behavior and Development*, **18**, 339-350.
- Boersma, P., and Weenik, D. (2009). "Praat: Doing phonetic by computer," retrieved from <http://www.praat.org> (Last viewed July, 2009).
- Bosch, L. & Sebastián-Gallés, N. (2005). "Developmental changes in the discrimination of vowel contrasts in bilingual infants," In *Proceedings of the 4th International Symposium on Bilingualism*, edited by J. Cohen, K. T. McAlister, K. Rolstad, and J. MacSwan (Somerville, MA: Cascadia Press), pp. 354-363.
- Bosch, L. & Sebastián-Gallés, N. (2003). "Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life," *Language and Speech*, **46**, 217-243.
- Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). "First and second language vowel perception in early bilinguals," *European Journal of Cognitive Psychology*, **12**, 189-221.

- Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). "Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production," *Journal of the Acoustical Society of America*, **101**, 2299-2310.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," *Perception and Psychophysics*, **61**, 977-985.
- Burns, T. C., Werker, J. F., & McVie, K. (2003). "Development of phonetic categories in infants raised in bilingual and monolingual environments," In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27<sup>th</sup> Annual Boston University Conference on Language Development* (Vol. 1, pp. 173-184). Somerville, MA: Cascadilla Press.
- Burleson, D. F. (2007). "Improving intelligibility of non-native speech with computer-assisted phonological training," *IULC Working Papers Online*, 7, Article 07-05. Retrieved May 20, 2009, from <https://www.indiana.edu/~iulcwp/pdfs/07-Burleson5.pdf>
- Chaney, C. (1988). "Acoustic analysis of correct and misarticulated semivowels," *Journal of Speech and Hearing Research*, **31**, 275-287.
- Cheng, B. & Zhang, Y. (2009). "The relationship between speech perception and production in second language learners," *Journal of the Acoustical Society of America*, **125**, 2754.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). "Development of language-specific phoneme representations in the infant brain," *Nature Neuroscience*, **1**, 351-353.

- Chou, F. C. (2005). "On-line English pronunciation training system with automatic speech recognition technologies," In P. Kommers & G. Richards (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005* (pp. 3819-3824). Chesapeake, VA: AACE.
- Cucchiaroni, C., Neri, A., De Wet, F., & Strik, H. (2007). "ASR-based pronunciation training: Scoring accuracy and pedagogical effectiveness of a system for Dutch L2 learners," In *Proceedings of Interspeech 2007*, 2181-2184.
- Dalby, J., & Kewley-Port, D. (1999). "Explicit pronunciation training using automatic speech recognition technology," *CALICO Journal*, **16**, 425-445.
- Diehl, R. L., Lotto, A. J. & Holt, L. L. (2004). "Speech perception. Annual Review of Psychology," **55**, 149-179.
- Dooling, R. J., Best, C. T., & Brown, S. D. (1995). "Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulates*) and zebra finches (*Taeniopygia guttata*)," *Journal of the Acoustical Society of America*, **97**, 1839-1846.
- Eimas, P. D. (1975). "Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants," *Perception and Psychophysics*, **18**, 341-347.
- Eimas, P. D., & Corbit, J. D. (1973). "Selective adaptation of linguistic feature detectors," *Cognitive Psychology*, **4**, 99-109.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). "Speech perception in infants," *Science*, **171**, 303-306.

- Escudero, P., & Boersma, P. (2004). "Bridging the gap between L2 speech perception research and phonological theory," *Studies in Second Language Acquisition*, **26**, 551–585.
- Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). "Acoustic modeling of American English /r/," *Journal of the Acoustical Society of America*, **108**, 343-356.
- Eskenazi, M. (1999). "Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype," *Language Learning & Technology*, **2**, 62-76.
- Evans, B. G., and Iverson, P. (2004). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *Journal of the Acoustical Society of America*, **115**, 352–361.
- Evans, B. G., and Iverson, P. (2007). "Plasticity in vowel perception and production: A study of accent change in young adults," *Journal of the Acoustical Society of America*, **121**, 3814–3826.
- Flege, J. E. (1991). "The interlingual identification of Spanish and English vowels: Orthographic evidence," *Quarterly Journal of Experimental Psychology*, **43**, 701–731.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233-277.
- Flege, J. E. (1999). "The relation between L2 production and perception," In *Proceedings of the XIVth International Congress of Phonetics Sciences*, edited by J. Ohala, Y.

- Hasegawa, M. Ohala, D. Granville & A. Bailey (Berkeley, CA: Department of Linguistics, Univ. of California at Berkeley), pp. 1273-1276.
- Flege, J. E. (2003). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.
- Flege, J. E., and Schmidt, A. M. (1995). "Native speakers of Spanish show rate-dependent processing of English stop consonants," *Phonetica*, **52**, 90-111.
- Flege, J. E., Bohn, O-S. and Jang, S. (1997). "The production and perception of English vowels by native speakers of German, Korean, Mandarin, and Spanish," *Journal of Phonetics*, **25**, 437-470.
- Flege, J. E., MacKay, I. R. A., & Meader, D. (1999). "Native Italian speakers' perception and production of English vowels," *Journal of the Acoustical Society of America*, **106**, 2973-2987.
- Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). "Interaction between the native and second phonetic subsystems," *Speech Communication*, **40**, 467-491.
- Flege, J. E., Takagi, N., & Mann, V. (1995). "Japanese adults can learn to produce English /r/ and /l/ accurately," *Language and Speech*, **38**, 25-55.
- Fowler C. A. (1981). "Production and perception of coarticulation among stressed and unstressed vowels," *Journal of Speech, Language, and Hearing Research*, **24**, 127–139.
- Fowler C. A. (1984). "Segmentation of coarticulated speech in perception," *Perception and Psychophysics*, **36**, 359–368.



- Fowler C. A. (1986). "An event approach to the study of speech perception from a directrealist perspective," *Journal of Phonetics*, **14**, 3–28
- Fowler C. A. (1989). "Real objects of speech perception: a commentary on Diehl and Kluender." *Ecological Psychology*, **1**, 145–160.
- Fowler C. A. (1994). Speech perception: direct realist theory. In *The Encyclopedia of Language and Linguistics*, ed. RE Asher, pp. 4199–4203. Oxford: Pergamon.
- Fowler C. A. (1996). "Listeners do hear sounds, not tongues," *Journal of the Acoustical Society of America*, **99**, 1730–1741.
- Fox, R. A., Flege, J. E., & Munro, M. J. (1995). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," *Journal of the Acoustical Society of America*, **97**, 2540-2551.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). "Motor theory of speech perception reviewed," *Psychonomic Bulletin & Review*, **13**, 361-377.
- Gass, S. (1984). "Development of speech perception and speech perception abilities in adult second language learners," *Applied Psycholinguistics*, **5**, 51-74.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, **47**, 103–138.
- Goto, H. (1971). "Auditory perceptoin by normal Japanese adults of the sounds "L" and "R"," *Neuropsychologia*, **9**, 317-323.
- Gottfried, T. L. (1984). "Effects of consonant context on the perception of French vowels," *Journal of Phonetics*, **12**, 91-114.
- Gottfried, T. L., & Beddor, P. S. (1988). "Perception of temporal and spectral information n French vowels," *Language and Speech*, **31**, 57-75.

- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *Journal of the Acoustical Society of America*, **107**, 2711–2724.
- Hagiwara, R. (1995). "Acoustic realizations of American /r/ as produced by women and men," *UCLA Phonetics Laboratory Working Papers*, **90**, 1-187.
- Hagiwara, R., Fosnot, S. M., & Alessi, D. M. (2002). "Acoustic phonetics in a clinical setting: a case study of /r/-distortion therapy with surgical intervention," *Clinical Linguistics and Phonetics*, **16**, 425-441.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," *Speech Communication*, **47**, 360-378.
- Huer, M. B. (1989). "Acoustic tracking of articulation errors: [r]," *Journal of Speech and Hearing Disorders*, **54**, 530-534.
- Iverson, P., and Evans, B. G. (2003). "A goodness optimization method for investigating phonetic categorization," Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona.
- Iverson, P., and Evans, B. G. (2007). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," *Journal of the Acoustical Society of America*, **122**, 2842–2854.
- Iverson, P., Hazan, V., and Bannister, K. (2005). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r-/l/ to Japanese adults," *Journal of the Acoustical Society of America*, **118**, 3267–3278.

- Iverson, P., Smith, C. A., and Evans, B. G. (2006). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration," *Journal of the Acoustical Society of America*, **120**, 3998–4006.
- Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. G. (2008). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch soeakers," *Journal of Experimental Psychology Human Perception and Performance*, **34**, 1305-1316.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition*, **87**, B47–B57.
- Kewley-Port, D. Watson, C. S., Elbert, M., Maki, D., & Reed, D. (1991) "The Indiana speech training aid (ISTRA) II: Training curriculum and selected case studies," *Clinical Linguistics and Phonetics*, **5**, 13-38.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *Journal of the Acoustical Society of America*, **87**, 820–857.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). "Japanese quail can learn phonetic categories," *Science*, **237**, 1195-1197.
- Kluge, C. D., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). "The relationship between perception and production of English nasal codas by Brazilian learners of English," *Proceedings of Interspeech 2007 conference*, (Antwerp, Belgium, August 27-31, 2007), pp. 2297-2300.

- Kuhl, P. K. (1981). "Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories," *Journal of the Acoustical Society of America*, **70**, 340-349.
- Kuhl, P. K. (2000). "Phonetic learning: a pathway to language," *National Academy of Sciences USA*, **97**, 11850-11857.
- Kuhl, P. K. (2004). "Early language acquisition: cracking the speech code," *Nature Reviews Neuroscience*, **5**, 831-843.
- Kuhl, P. K., Miller, J. D. (1975). "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," *Science*, **190**, 69-72.
- Kuhl, P. K., Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *Journal of the Acoustical Society of America*, **63**, 905-917.
- Kuhl, P. K., & Padden, D. M. (1983). "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," *Journal of the Acoustical Society of America*, **73**, 1003-1010.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science*, **255**, 606-608.
- Kuhl, P. K., Conboy, B. T., Coffery-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)," *Philosophical Transaction of the Royal Society B*, **363**, 979-1000.

- Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of language* (Blackwell, Oxford, UK).
- Lasky, R. E., Syrdal-Lasky, A., & Klein R. E. (1975). "VOT discrimination by four to six and a half month old infants from Spanish environments," *Journal of Experimental Child Psychology*, **20**, 215-225.
- Lengeris, A., & Hazan, V. (2007). "Cross-language perceptual assimilation and discrimination of southern British English vowels by Greek and Japanese learners of English," *16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1641-1644.
- Levy, E. S., & Strange, W. (2008). "Perception of French vowels by American English adults with and without French language experience," *Journal of Phonetics*, **36**, 141-157.
- Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). "An association between mother's speech clarity and infants' speech discrimination skills," *Developmental Science*, **6**, F1-F10.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, **243**, 489-494.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review*, **74**, 431-461.

- Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, **65**, 497-516.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). "Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories," *Journal of the Acoustical Society of America*, **94**, 1242-1255.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). "Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories," *Journal of the Acoustical Society of America*, **96**, 2076-2087.
- Logan, J. D., Lively, S. E., & Pisoni, D. B. (1991). "Training Japanese listeners to identify English /r/ and /l/: A first report," *Journal of the Acoustical Society of America*, **89**, 874-886.
- Lotto, A. J., Sato, M., & Diehl, R. L. (2004). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," in *From Sounds to Sense: 50 + Years of Discoveries in Speech Communication*, edited by J. Slifka, S. Manuel, and M. Matthies (MIT Research Laboratory in Electronics, Cambridge, MA), pp. C-181-C186.
- MacKain, K. S., Best, C. T., Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," *Applied Psycholinguistics*, **2**, 369-390.

- MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (2001). "Category restructuring during second-language (L2) speech acquisition," *Journal of the Acoustical Society of America*, **110**, 516–528.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. New York: Cambridge University Press.
- Masaki, S., Akahane-Yamada, R., Tiede, M. K., Shimada, Y., & Fujimoto, I. (1996). "An MRI-based analysis of the English /r/ and /l/ articulations," *The Ninth International Conference on Spoken Language Processing*, Pittsburgh, USA, 1581-1584.
- Massaro, D. W., and Light, J. (2003). "Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/," In *Proceedings of Eurospeech (Interspeech)*, 8<sup>th</sup> European conference on speech communication and technology, Geneva, Switzerland. 2249-2252.
- Maxwell, E. M., & Weismer, G. (1982). "The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops," *Applied Psycholinguistics*, **3**, 29-43.
- Maye, J., & Gerken, L. (2000). "Learning phonemes without minimal pairs," In S. C. Howell, S. A. Fish & T. Keith-Lucas (Eds), *Proceedings of the 24th Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.
- Maye, J., Werker, J. F., & Gerken, L. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition*, **82**, B101-B111.

- Mich, O., Neri, A., & Giuliani, D. (2006). "The effectiveness of a computer assisted pronunciation training system for young foreign language learners," *In Proceedings of CALL Conference 2006*, (University of Antwerp, Language Institute Antwerp, Belgium), 135-143.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujumura, O. (1975). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Perception and Psycholinguistics*, **18**, 331-340.
- Mochizuki, M. (1981). "The identification of /r/ and /l/ in natural and synthesized speech," *Journal of phonetics*, **9**, 283-303.
- Morrison, G.S. (2002). "Perception of English /i/ and /ɪ/ by Japanese and Spanish Listeners: Longitudinal results," In G.S. Morrison & L. Zsoldos (eds.), *Proceedings of the North West Linguistics Conference 2002*. Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association. 29-48.
- Neri, A., Cucchiarini, C., & Strik, H. (2006). "ASR-based corrective feedback on pronunciation: does it really work?," *Proceedings of Interspeech 2006*, Pittsburgh, PA, 1982-1985.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). "A limit on behavioral plasticity in speech perception," *Cognition*, **64**, B9-B17.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries," *Psychological Science*, **12**, 445-447.



- Pegg, J. E., & Werker, J. F. (1997). "Adult and infant perception of two English phones," *Journal of the Acoustical Society of America*, **102**, 3742-3753.
- Polka, L. (1991). "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions," *Journal of the Acoustical Society of America*, **89**, 2961-2977.
- Polka, L. (1995). "Linguistic influences in adult perception of non-native vowel contrasts," *Journal of the Acoustical Society of America*, **97**, 1286-1296.
- Polka, L., & Werker, J. F. (1994). "Developmental changes in perception of nonnative vowel contrasts," *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421-435.
- Polka, L., Colantonio, C., & SUNDARA, M. (2001). "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern," *Journal of the Acoustical Society of America*, **109**, 2190-2201.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). "Language discrimination by human newborns and by Cotton-top tamarin monkeys," *Science*, **288**, 349-351.
- Riviera-Gaxiola, M., Silvia-Pereyra, J., & Kuhl, P. K. (2005). "Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants," *developmental Science*, **8**, 162-172.
- Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 379-410.

- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). "Online processing of native and non-native phonemic contrasts in early bilinguals," *Cognition*, **72**, 111-123.
- Sheldon, A., & Strange, W. (1982). "The acquisition of /r/ and /l/ by Japanese listeners of English: Evidence that speech production can precede speech perception," *Applied Psycholinguistics*, **3**, 243-261.
- Streeter, L. A. (1976). "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," *Science*, **259**, 39-41.
- Sundara, M., Polka, L., & Genesee, F. (2006). "Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English," *Cognition*, **100**, 369-388.
- Tees, R. C., & Werker, J. F. (1984). "Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech," *Canadian Journal of Psychology*, **38**, 579-590.
- Trehub, S. E. (1976). "The discrimination of foreign speech contrasts by infants and adults," *Child Development*, **47**, 466-472.
- Tsao, F. -M., Liu, H. -M., & Kuhl, P. K. (2006). "Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants," *Journal of the Acoustical Society of America*, **120**, 2285-2294.
- Tsujimura, N. (1996). *An introduction to Japanese linguistics* (Blackwell, Oxford, UK).
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). "A developmental study of English vowel production and perception by native English adults and children," *Journal of Phonetics*, **33**, 263-290.

- Underbakke, M., Polka, L., & Gottfried T. L. (1988). "Trading relations in the perception of /r/-/l/ by Japanese learners of English," *Journal of the Acoustical Society of America*, **84**, 90-100.
- Uther, M., Uther, J., Athanasopoulos, P., Singh, P. & Akahane-Yamada, R. (2007). "Mobile Adaptive CALL (MAC). A lightweight speech-based intervention for mobile language learners," *Proceedings of Interspeech 2007 conference*, (Antwerp, Belgium, August 27-31, 2007), pp. 2329-2332
- Vance, T. (1987). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).
- van Dommelen, W. A. (2008). "The perception of English consonants by Norwegian listeners: A preliminary report," In *Proceedings of the XXIst Swedish Phonetic Conference*, edited by A. Eriksson and J. Lindh (Gothenburg, Sweden: Department of Linguistics, University of Gothenburg), pp. 45-48.
- Wang, Y., Jongman, A., & Sereno, J. A. (2003). "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *Journal of the Acoustical Society of America*, **113**, 1033-1043.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). "Training American listeners to perceive mandarin tones," *Journal of the Acoustical Society of America*, **106**, 3649-3658.
- Wayland, R., & Guion, S. (2003). "Perceptual discrimination of Thai tones by native and experienced learners of Thai," *Applied Psycholinguistics*, **24**, 113-129.

- Weismer, G., Dinnsen, D., and Elbert, M. (1981). "A study of the voicing distinction associated with omitted, word-final stops," *Journal of Speech and Hearing Disorders*, **46**, 320–327.
- Werker, J. F., & Curtin, S. (2005). "PRIMIR: A developmental framework of infant speech processing," *Language Learning and Development*, **1**, 197-234.
- Werker, J. F., & Lalonde, C. E. (1988). "Cross-language speech perception: Initial capabilities and developmental change," *Developmental Psychology*, **24**, 672-683.
- Werker, J. F., & Tees, R. C. (1984). "Phonemic and phonetic factors in adult cross-language speech perception," *Journal of the Acoustical Society of America*, **75**, 1866-1878.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). "Developmental aspects of cross-language speech perception," *Child Development*, **52**, 349-355.
- Westbury, J. R., Hashi, M., Lindstorm, M. J. (1998). "Differences among speakers in lingual articulation for American /ɹ/," *Speech Communication*, **26**, 203-226.
- Yamada, R. A. (1995). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp.305–320.
- Yamada, R., and Tohkura, Y. (1992). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," *Perception and Psychophysics*. **52**, 376–392.

- Yamada, T., Adachi, T., & Advanced Telecommunications Research Institute International. (1998). “*Eigo listening kagakuteki joutatsuhou* [Scientific training for English listening],” Tokyo: Koudansha.
- Zhang, Z., Boyce, S., Espy-Wilson, C., & Tiede, M. (2003). “Acoustic strategies for production of American English “retroflex” /r/,” *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, Barcelona, Spain, 1125-1128.
- Zhou, X., Espy-Wilson, C. Y., Tiede, M., & Boyce, S. (2007). “An articulatory and acoustic study of “retroflex” and “bunched” American English rhotic sound based on MRI,” *Proceedings of Interspeech 2007 conference*, (Antwerp, Belgium, August 27-31, 2007), pp. 54-57.
- Zhou, X., Espy-Wilson, C. Y., Bpyce, S., Tiede, M., Holland, C., & Choe, A. (2008). “A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/,” *Journal of the Acoustical Society of America*, **123**, 4466-4481.

## Appendix A: Chapter 2 subject information (part 1)

Subject	Gender	Age	University / College major	English learning onset age	Total time of learning English (years)	Length in language school in Japan
J 01	M	20	English	12	8	16 months
J 02	F	25	International relationship	13	10	
J 03	F	19	English	12	8	
J 04	F	21	English	6	15	11 years
J 05	F	24	Tourism	13	10	12 months
J 07	F	20	English	12	8	
J 08	F	38		13	25	5 years
J 09	F	40	International Child Health Nursing	13	11	1 year
J 10	M	21	English	10	10	
J 11	M	20		13	8	2 years
J 12	M	21	Economics	13	9	1 year
J 14	F	28	Japanese	12	12	
J 15	F	33	Culture Study	13	10	3 years
J 16	F	25	Medicine	12	8	6 months
J 17	F	36	Class Management	10	24	
J 18	F	18	Linguistics	13	7	3 years
J 19	F	19	English	12	8	6 years
J 20	F	31	Sociology	13	9	
J 21	F	22	Hair Dresser	13	6	
J 22	F	46		13	6	2 years
J 23	F	21	English	13	9	1 year
J 24	F	19	English	12	7	5 years
J 25	F	41	Graphics	13	6	
J 26	F	33	English Japanese Applied Linguistics	13	9	
J 27	F	27		13	6	
J 28	F	30	Drama	12	10	
J 29	F	21	Communication	13	8	
J 30	F	48	Cognitive Psychology	13	10	2 years 6 months
J 31	F	22	English	13	10	
J 32	F	20	English	13	7	
J 33	M	32	Accounting	13	8	2 years
J 34	F	33	English	6	12	2 years
J 35	F	30	International Cultural Studies Applied Linguistics	12	10	
J 37	F	26		13	9	3 years
J 38	M	22		13	9	
J 39	F	22	Communication	10	12	3 months

## Appendix B: Chapter 2 subject information (part 2)

Subject	Length of residence	Usage of spoken English	Usage of written English	Self-evaluated spoken English	Self-evaluated written English
J 01	1 week (UK)	0%	0%	primary	primary
J 02	3 months (UK)	40%	20%	intermediate	primary
J 03	1 week (UK)	25%	35%	primary	primary
J 04	12 months (AUS)	5%	5%		
J 05	5 years 6 months (UK)	40%	20%	advanced	intermediate
J 06	1 week (UK)	0%	0%	primary	primary
J 07	9 months (UK)	30%	20%	primary	primary
J 08	7 months (Philippines)	80%	80%	advanced	advanced
	7 months (Djibouti)				
J 09	3 months (UK)	10%	20%	primary	primary
J 10	1 week (UK)	3%	3%	primary	primary
J 11	2 months (Spain)	5%	1%	primary	primary
	1 week (UK)				
J 12	1 year 9 months (UK)	10%	0%	primary	primary
J 13	2 weeks (UK)	90%	50%	intermediate	intermediate
J 14	4 months (UK)	60%	50%	intermediate	intermediate
J 15	4 years 9 months (UK)	40%	50%	intermediate	advanced
J 16	1 week (UK)	10%	30%	primary	primary
J 17	1 week (UK)	20%	34%	primary	primary
J 18	7 months (Canada)	90%	50%	intermediate	intermediate
	1 year 3 months (US)				
J 19	3 months (UK)	40%	50%	primary	primary
J 20	5 months (UK)	20%	10%	primary	primary
J 21	1 week (UK)	3%	3%	primary	primary
J 22	1 week (UK)	0%	0%	primary	intermediate
	1 year 3 months				
J 23	(Canada)	50%	0%	intermediate	primary
	4 years (UK)				
J 24	1 year 8 months (Ireland)	10%	5%	intermediate	intermediate
	2 years (UK)				
J 25	3 months (UK)	10%	10%	primary	primary
J 26	1 year 6 months (Italy)	50%	0%	primary	primary
	1 month (UK)				
J 27	1 week (UK)	20%	20%	intermediate	intermediate
J 28	13 years (UK)	60%	50%	advanced	advanced
J 29	3 months (UK)	20%	10%	primary	intermediate
J 30	1 week (UK)	0%	0%	primary	intermediate
J 31	1 month 2 weeks (UK)	0%	80%	primary	intermediate
J 32	7 years 5 months (UK)	30%	30%	intermediate	primary
J 33	3 years 5 months (UK)	80%	80%	advanced	advanced
J 34	2 months (UK)	40%	60%	primary	intermediate
J 35	1 week (UK)	30%	60%	intermediate	intermediate
J 36	1 week (UK)	15%	30%	intermediate	intermediate

## Appendix C: Chapter 3 subject information (part 1)

Subject	Gender	Age	University / College major	English learning onset age	Total time of learning English (years)	Length in language school in Japan
J 01	F	34	French	9	10	3 years
J 02	M	26	Social Welfare	13	10	
J 03	F	40	Nursing	13	7	2 months
J 04	F	34	Law	10	10	2 years
			International Politics			
J 05	F	23	International Relationship	12	6	9 months
J 06	F	33	Spanish	9	6	1 year 6 months
J 07	F	27	Brain Neurology	12	6	
J 08	F	30	Educational Sociology	12	8	1 year 6 months
J 09	F	31	Law	12	7	
J 10	F	36	Japanese	13	8	
J 11	F	33	Sociology	12	10	3 months
J 12	F	24	English	11	12	2 years
			Politics and Economics			
J 13	F	24	Cultural Studies	13	11	3 years
J 14	F	39		12	8	
			Health & Physical Education			
J 15	F	67		12	8	
J 16	M	21	Law	13	8	
J 17	F	34		13	6	
J 18	F	37	Dental Hygiene	2	7	3 months
J 19	F	35	Business Management	13	8	
J 20	F	28	Communication	12	11	
J 21	F	33		12	8	
J 22	M	47	Economics	13	10	1 year 6 months
J 23	M	20	Communication	12	8	
J 24	F	30	Medicine	12	6	
J 25	F	28	Medicine	10	8	6 months
J 26	F	18		13	6	
J 27	F	30	Accounting	13	10	
J 28	F	40	Cloth Design	13	6	4 years
J 29	F	27		13	8	
J 30	F	25	Cloth Design	13	10	
J 31	F	31	Nursing	13	8	6 months
J 32	M	29	Medicine	13	12	1 year 6 months
J 33	F	42	Graphics Design	12	6	
J 34	M	34	English	13	10	
J 35	F	39	English	12	10	2 years
J 36	F	31	Drama	13	10	



Subject	Gender	Age	University / College major	English learning onset age	Total time of learning English (years)	Length in language school in Japan
J 37	F	31	Chemistry	12	6	
J 38	F	24		10	8	5 months
J 39	F	31	English	12	8	
J 40	M	25		12	6	
J 41	F	45	Music	12	8	1 month
J 42	F	40	Graphics Design	13	7	4 months
J 43	F	30	Music	12	8	3 years
J 44	F	38	Medicine	13	8	3 years
J 45	M	31	Commerce	13	10	1 year 6 months
J 46	F	35	Music	13	8	
J 47	M	21	Economics	13	6	

## Appendix D: Chapter 3 subject information (part 2)

Subject	Length of residence	Usage of spoken English	Usage of written English	Self-evaluated spoken English	Self-evaluated written English
J 01	1 year 10 months (US) 7 years (UK)	30%	5%	intermediate	intermediate
J 02	3 months (UK)	60%	30%	intermediate	intermediate
J 03	2 weeks (UK)	20%	30%	primary	primary
J 04	4 months (UK) 4 months (US)	65%	20%	intermediate	intermediate
J 05	1 week (UK) 9 months (Canada) 3 months (NZ)	60%	20%	advanced	advanced
J 06	1 year 7 months (UK)	30%	10%	intermediate	intermediate
J 07	9 months (UK)	80%	0%	intermediate	primary
J 08	1 year (Indonesia) 7 months (UK)	10%	5%	primary	primary
J 09	3 months (UK)	80%	10%	primary	intermediate
J 10	1 year 11 months (UK)	0%	0%	primary	primary
J 11	3 years 1 month (UK)	50%	50%	primary	intermediate
J 12	1 month (UK)	20%	5%	primary	primary
J 13	3 years (Belgium) 1 year (UK)	5%	5%	primary	intermediate
J 14	1 month (UK)	50%	50%	primary	intermediate
J 15	6 years (UK)	0%	0%	primary	primary
J 16	1 month (UK) 2 months (US)	30%	30%	intermediate	intermediate
J 17	1 year (UK)	30%	50%	primary	primary
J 18	11 months (Canada) 2 years (UK)	60%	70%	intermediate	intermediate
J 19	5 months (Italy) 8 months (UK)	0%	20%	intermediate	intermediate
J 20	4 years (UK)	90%	30%	intermediate	intermediate
J 21	2 years (UK)	0%	0%	primary	primary
J 22	16 years (UK) 7 years (Italy) 1 year 4 months (US)	10%	10%	advanced	advanced
J 23	4 months (UK)	90%	30%	intermediate	advanced
J 24	7 months (UK)	0%	0%	intermediate	intermediate
J 25	1 week (UK)	40%	40%	intermediate	intermediate
J 26	1 week (UK)	0%	0%	primary	primary
J 27	4 months (UK)	30%	0%	intermediate	intermediate
J 28	4 years 6 months (UK)	65%	50%	intermediate	primary
J 29	10 months (NZ) 1 year 4 months (UK)	40%	30%	intermediate	intermediate
J 30	1 year (UK)	80%	80%	intermediate	intermediate
J 31	8 months (UK)	10%	20%	intermediate	intermediate

Subject	Length of residence	Usage of spoken English	Usage of written English	Self-evaluated spoken English	Self-evaluated written English
J 32	1 week (UK)	10%	10%	primary	intermediate
J 33	6 years 4 months (UK) 11 months (Canada)	20%	10%	intermediate	primary
J 34	1 year 8 months (UK)	10%	30%	intermediate	advanced
J 35	1 year (Canada) 1 year 2 months (UK)	30%	0%	intermediate	intermediate
J 36	1 year 6 months (Italy) 1 year 5 months (UK)	80%	30%	intermediate	intermediate
J 37	4 months (UK)	5%	5%	primary	primary
J 38	2 months (UK)	70%	30%	intermediate	intermediate
J 39	1 year 9 months (UK)	80%	80%	advanced	advanced
J 40	2 months (UK)	30%	30%	primary	intermediate
J 41	5 years (US) 1 month (UK)	20%	20%	intermediate	advanced
J 42	6 years 2 months	30%	20%	primary	intermediate
J 43	1 year (UK) 2 years 10 months (UK)	70%	50%	intermediate	intermediate
J 44	(UK)	30%	30%	intermediate	intermediate
J 45	3 years 4 months (UK)	30%	60%	intermediate	intermediate
J 46	7 years (UK)	40%	10%	primary	primary
J 47	10 months (UK)	50%	70%	primary	intermediate