Perceptual Training of English /r/ and /l/ for Japanese Adults, Adolescents and Children

Yasuaki Shinohara

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

University College London (UCL),

Department of Speech, Hearing and Phonetic Sciences

2014

Declaration

I, Yasuaki Shinohara, 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 thesis.

Yasuaki Shinohara

Abstract

Although phoneme perception and production in second-language (L2) learners has been the focus of much research over the past few decades, the learning mechanisms and the factors that influence them are still far from understood. The purpose of this PhD research is to cast some light on the mechanisms underlying the learning of L2 phonemes and the effects of age. Japanese speakers who have problems in perception and production of the English /r/-/l/ contrast participated in three perceptual training studies. The first study examined whether Japanese adults can improve their perception and production with identification and discrimination training. The results demonstrated that both identification and discrimination training methods improved their perception and production of the English /r/-/l/ contrast, but the combination of two different training methods did not have additive effects. The second study investigated how age affects the learning of the English /r/-/l/ contrast at phonetic and phonological levels of perception. The results demonstrated that Japanese adults are disadvantaged in improving their phonetic perception of the English /r/-/l/ contrast due to their relatively fossilised brain plasticity and their developed L1 phonetic units. On the other hand, younger learners are able to improve both phonetic and phonological perception of the contrast. This may be attributed to their greater brain plasticity and less interference from undeveloped L1 phonetic categories. Finally, the third study examined how Japanese speakers improve production of English /r/ and /l/ through perceptual training, and found that perceptual training transferred to production ability in both identifiability and acoustic realisations. The improvement in production seemed to be attributable to perceptual learning. There may be a common underlying ability for perception and production, although the acoustic dimensions which link the two may be different between individuals. The theoretical implications for understanding learning mechanisms and age effects are discussed.

Keywords: second language acquisition, speech perception and production, perceptual training.

Acknowledgement

I would like to sincerely thank to my primary supervisor, Paul Iverson. From the moment he accepted me as his PhD student, I owe my deepest gratitude to him. His advice has been incisive; he encouraged me to meet challenges and gave me confidence. This thesis would not have been possible without his help. Thank you, Paul.

I am also grateful to all the professors and technicians who helped my research. Professor Valerie Hazan, my secondary supervisor, inspired me with ideas for experiments. Steve Nevard kindly supported me both in setting up equipment and with recording advice.

I would also like to express my gratitude to all my colleagues. Yasna Pereira Reyes always talked with, listened to and encouraged me. Chierh Cheng taught me how to manage a PhD, and Jose Joaquin Atria advised me about programming. Sonia Granlund, Mark Wibrow, Dorothea Hackman and Sam Green supported my English, and Mauricio Figueroa Candia gave me moral support, while Melanie Pinet baked many delicious cakes. I am also thankful to the other colleagues who make our PhD room the most comfortable and happiest working place: Nada Al-Sari, Wafa’a Alshangiti, Emma Brint, Lucy Carroll, Cristiane Hsu, Daniel Kennedy-Higgins, Young-Shin Kim, Albert Lee, Hao Liu, Xoaluan Liu, Gisela Tome Lourido, Katharine Mair, Georgina Oliver Roth, Csaba Redey-Nagy, Tim Schoof, Dong-Jin Shin, Jieun Song, Kurt Steinmetzger, Louise Stinger, and Yue Zhang.

This thesis would not have been possible without all the helpers and subjects. I am grateful to Professor Tetsuo Harada at Waseda University, Yasuyuki Takimoto, Masami Kogure, Emiko Negishi, Takayuki Baba, Wataru Asakai, Keiko Takahashi, Marina Arashiro, Tatsuhide Mizumura and Reira Hara for their support in recruiting participants for my experiments in Japan. I would also like to thank all the subjects and their families.

This PhD research was funded by the Japan Student Services Organization with a scholarship for long-term study abroad. Without this scholarship, I would not even have started my PhD. I am deeply appreciative to this organisation for their financial support.

Finally, I thank my family, and my girlfriend, Masako Kurihara. Whenever I went home to Japan, the warmth of their welcome really made me feel relaxed. Masako was always encouraging me. Thank you very much.

Table of Contents

List of Figures

List of Tables

Chapter 1: General Discussion

Language is the most frequently used tool for human beings to communicate with each other. Nevertheless, as there is considerably more than just one language in the world, English is now used by many people as a *lingua franca* in intercultural communication or as a second language (L2). Although the reasons for learning English as a second language differ between individuals, many L2 learners often desire to achieve the same level of fluency in perception and production as native English speakers. However, it is very difficult to achieve this near-native fluency in L2, and so numerous studies have proposed theories of speech perception and production to clarify the mechanisms that underlie both L1 and L2 learning, and to attempt to account for the difficulty in acquiring L2 fluency.

Research on developmental changes in speech perception and production has provided a clearer picture of why L2 learners have difficulties in perceiving and producing L2 phonemes. Newborn infants have a universal phonetic capacity to discriminate the phonetic contrasts of all languages (Eimas, 1975, Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl et al., 2006; Kuhl et al., 2008; Lasky, Syrdal-Lasky, & Klein, 1975; Werker & Lalonde, 1988). However, this capacity is lost the first few months after birth as infants detect and adapt to the distributional patterns of the linguistic input contained in infant-directed speech. (Kuhl et al., 2006; Kuhl et al., 2008; Kuhl, 2011 Tsao, Liu, & Kuhl, 2006; Werker & Tees, 1984). For example, although Japanese infants' discrimination ability for the English /r/-/l/ contrast is comparable to American English infants in the first 6-8 months of life, it declines after the age of 10-12 months, while American English infants increase their discrimination ability during the second-half of their first year of life (Kuhl et al., 2006). Once the neural phonetic representations become stable, additional language input cannot substantially change the distributional patterns (Kuhl, 2011; Kuhl et al., 2008), although infants are still capable of learning new phonetic representations by exposure to a new language (Kuhl, Tsao, & Liu, 2003). Thus, neural commitment by exposure to an ambient language seems to cause difficulty in perceiving L2 phonemes.

Although perceiving L2 contrasts is often difficult for adult listeners, some contrasts are easier than others. The difficulty in perceiving a particular L2 contrast may depend on how that contrast is perceived in the phonological space developed for L1. The Perceptual Assimilation Model (PAM; Best, 1994a, 1994b, 1995) proposes that non-native listeners assimilate unfamiliar non-native phonemes to the L1 phonemes that are most articulatorily similar, and this perceptual assimilation prevents listeners from perceiving acoustic details of non-native phonemes. Thus, how well L2 speakers identify and discriminate unfamiliar phonological contrasts can be explained by assimilation patterns: the similarity between the members of the L2 contrast and L1 phonemes. For example, when two members of an L2 contrast are perceived as acceptable tokens of two different L1 phonemes, (i.e., “Two Category assimilation”), better discrimination can be predicted. If two phonemes are perceived as equally good or bad examples of the same L1 phoneme, (i.e., “Single Category assimilation”), poor discrimination is predicted. An L2 contrast may be fairly easy to discriminate, if two phonemes are not equally good or bad tokens of the same L1 phoneme but different in goodness of fit, (i.e., “Category Goodness difference”). There is also a case that perceptual assimilation does not occur for L2 phonemes that are perceived as non-linguistic non-speech sounds, (i.e., “Non-assimilated”). Even if the phonemes in the L2 contrast are perceived as speech sounds, perceptual assimilation may still not occur (i.e., “Uncategorised”) if listeners fail to perceive them as equivalent to L1 phonemes. In this case, the degree of difficulty discriminating L2 contrasts seems to depend on how two members of an L2 contrast are perceived. If one of the L2 phonemes is perceived as equivalent to an L1 phoneme and the other fails to match any of L1 phonemes, (i.e., “Uncategorised-Categorised type”), the contrast should be relatively well discriminated. On the other hand, if both phonemes are not assimilated to any of L1 categories, (i.e., “Uncategorised-Uncategorised type”), the difficulty depends on how close they are to the same or different L1 phoneme(s). When two members of an L2 contrast are perceived as moderately similar to the same L1 phoneme, it may be difficult to discriminate. When two uncategorised members are moderately similar to different sets of L1 phonemes and perceived as distant from one another, it may be relatively easy to discriminate. Thus, PAM suggests that the difficulty in perceiving L2 phonemes depends on how L2 phonemes are assimilated to L1 phonological categories.

Within the PAM framework under discussion, the English /r/-/l/ contrast seems to be a difficult L2 contrast for Japanese speakers to identify and discriminate. Best and Strange (1992) demonstrated that Japanese /r/, which is phonetically characterised as an apico-alveolar tap [ɾ] (Vance, 2008), is the closest phoneme to English /r/ and /l/, and both English /r/ and /l/ are perceived as poor examples of a single category of Japanese /r/ or Japanese /w/. Takagi (1993) also demonstrated the Single Category assimilation pattern of English /r/ and /l/ to the Japanese /r/ phoneme. On the other hand, a recent study by Guion, Flege, Akahane-Yamada, and Pruitt (2000) demonstrated that the contrast is perceived as an Uncategorised-Uncategorised type, because both English /r/ and /l/ fall in between Japanese /r/ and Japanese /ɯr/ (i.e., an apico-alveolar tap [ɾ] preceded by a close back unrounded vowel [ɯ]). Since both English /r/ and /l/ fall in the same region, poor discrimination is predicted in PAM, and the Guion et al's result showed that this was indeed the case.

While PAM addresses the perception of non-native phonemes, the Speech Learning Model (SLM; Flege, 1995, 1999, 2002) focuses on L2 learning. SLM posits that L1 and L2 phonetic categories exist in a common phonological space, and the learning process of phonetic and phonological representations remains intact throughout an individual's lifespan. Due to the common phonological space for L1 and L2, SLM proposes that L2 phonemes which are perceptually similar/close to L1 categories are difficult to acquire. That is, even if L2 phonemes are sub-categorically different from L1 phoneme, the L2 phonemes which are categorically similar to L1 phonemes are often equated to the closest L1 phonemes. Subsequently, the sub-categorical cross-language phonetic differences between L1 and L2 phonemes merge into one phonological category. If L2 learners continue to judge the L2 phonemes as examples of their established L1 phonetic categories, the formation of new L2 phonemes may be blocked and a merged L1-L2 category would then develop, a process known as "category assimilation" in SLM terminology (Flege, 2002). For example, French-English and English-French bilinguals may produce an intermediate voice-onset time (VOT) reflecting properties of both English and French voiceless alveolar plosives, so that their phonetic categories of /t/ are different from those of monolinguals in terms of VOT (Flege, 1987, 2002). On the other hand, when L2 phonemes are not categorically equated to L1 phonemes, it seems to be easier to establish new L2 phonetic categories in a common phonological space, by increasing phonetic differences between the representations of new L2 phonemes and the closest L1 phonemes, a process known as “category dissimilation” (Flege, 2002). For example, Spanish-English bilinguals may produce a shorter VOT for Spanish /p t k/ than Spanish monolinguals do to make phonetic categories of Spanish /p t k/ distinct from their English /p t k/ (Flege & Eefting, 1986, 1987; Flege, 2002). This suggests that if sub-categorical phonetic distances are too great to equate L2 phonemes to L1 categories, category dissimilation may occur. Thus, SLM proposes that sub-categorical phonetic distances between L1 and L2 categories could predict difficulties in creating new L2 phonetic categories in a common phonological space.

In terms of the English /r/-/l/ contrast for Japanese speakers, phonetic learning is also affected by the phonetic perceptual similarities of English /r/ and /l/ to the closest Japanese phoneme /r/. English /l/ is perceptually closer to Japanese /r/, compared to the proximity of English /r/ to Japanese /r/, as Japanese speakers tend to give higher ratings for English /l/ than English /r/ in goodness-fit to Japanese /r/ (Komaki, Akahane-Yamada, & Choi, 1999; Takagi, 1993). Following this result, Aoyama, Flege, Guion, Akahane-Yamada, and Yamada (2004) examined whether English /r/ is easier to be acquired than English /l/ for Japanese speakers. Japanese adults and children who had moved to the US were tested twice in their production intelligibility of English /r/ and /l/ at 0.5 years and 1.6 years after they had arrived in the country. The results demonstrated that children improved more in their production of English /r/ than of /l/. Hence, the similarity and the dissimilarity of L1 and L2 phonetic categories seems to cause difficulties for Japanese speakers learning the English /r/-/l/ contrast.

Although the difficulty in perceiving the English /r/-/l/ contrast seems to be predicted by assimilation patterns, neither the assimilation patterns nor the proximity to L1 phonemes can predict the identification accuracy of the English /r/-/l/ contrast for Japanese speakers. Hattori and Iverson (2009) examined how individual differences in the identification accuracy of the English /r/-/l/ contrast can be explained by perceptual assimilation and phonetic similarities between L1 and L2 categories. Hattori and Iverson firstly found that English /l/ is assimilated to Japanese /r/ more readily than English /r/, and suggested that the perceptual assimilation pattern within the terminology of PAM would be a Category-Goodness difference or an Uncategorised-Categorised type, not a Single Category assimilation or Uncategorised-Uncategorised type. However, if the English /r/-/l/ contrast is a Category Goodness difference or an Uncategorised-Categorised type, fairly easy discrimination would be predicted. Hattori and Iverson also demonstrated neither assimilation nor proximity to L1 phonetic categories can predict individuals' identification, but the internal representation of the F3 for the two phonemes can account for identification accuracy. They found that Japanese speakers have native-like representations of English /r/ and /l/ in the acoustic dimensions of F1, F2, closure duration and transition duration, but there is a significant difference between English speakers and Japanese speakers in the F3 distinction for the English /r/-/l/ contrast. English speakers identify English /r/ with lower F3 and English /l/ with higher F3; Japanese speakers use more random F3 values and focus on F2 for the English /r/-/l/ contrast (Hattori & Iverson, 2009; Iverson et al., 2003; Lotto, Sato, & Diehl, 2004; Yamada & Tohkura, 1992). This underlying phonetic processing of F3 as well as category assimilation and proximity to L1 categories could contribute to the difficulty that Japanese speakers face in perceiving the English /r/-/l/ contrast.

In summary, Japanese speakers' difficulty in perceiving and producing the English /r/-/l/ contrast could be caused by multiple levels of processing. Nevertheless, it seems to be highly unlikely that these difficulties can be overcome through daily communication in English. In particular, the processing of F3 is especially resistant to change. Ingvalson, McClelland, and Holt (2011) demonstrated that F3 reliance in both perception and production is not changed as a function of length of residence in an English-speaking country, age of arrival in an English-speaking country, the amount of Japanese use or length of student status in an English environment (English education). Moreover, although F3 reliance and length of residence in an English-speaking country predict the English /r/-/l/ identification accuracy, there is no correlation between F3 reliance and length of residence. This suggests that Japanese speakers change communication strategies and/or assimilation patterns without a correspondent formation of new phoneme categories or a change of F3 representations through daily communication (Ingvalson et al., 2011; Iverson & Evans, 2009). Consequently, it seems necessary to separately improve both levels of phonological perception reflected in identification accuracy and phonetic perception reflected in processing of F3 in an artificial learning situation (e.g., training).

CHAPTER 1: GENERAL INTRODUCTION

The goal of this PhD thesis is to explore the learning mechanism of the English /r/-/l/ contrast for Japanese speakers using perceptual training. Chapter 2 investigated the effects of phonological and phonetic perceptual training of the English /r/-/l/ contrast for adult Japanese speakers. Although many training studies have been conducted to improve Japanese speakers' perception of this contrast, the studies reporting the greatest improvement combine a force-choice task with feedback using highly variable natural stimuli (Hazan, Sennema, Iba, & Faulkner, 2005; Iverson, Hazan, & Bannister, 2005; Logan, Lively, & Pisoni, 1991; Pruitt, Jenkins, & Strange, 2006). This type of training can improve the identification accuracy and production identifiability of the English /r/-/l/ contrast (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson et al., 2005). Moreover, the improved perceptual ability is retained for at least 6 months after the last training session (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994). However, there seems to be a limitation of such training techniques to improve Japanese speakers' identification accuracy by approximately 15% on average, and this may be because the underlying phonetic processing of F3 is not changed through training (Iverson et al., 2005). To address this issue, an experiment was described which used a combination of two different training methods aimed at improving higher phonological perception and lower phonetic perception, and it was examined whether or not the combined training technique has additive effects on improvement in perception and production of the English /r/-/l/ contrast.

Chapter 3 addressed age effects on learning the English /r/-/l/ contrast in perception by taking into account the age at which L2 phonemes are learned. The critical period hypothesis (Lenneberg, 1967), which has more recently been revised as the sensitive/optimal period theory (Oyama, 1976; Werker & Tees, 2005), claims that younger children are better at learning L2 speech due to their less advanced biological neural maturation. Moreover, the SLM (Flege, 1995, 1999, 2002) proposes that the level of L1 development along with age may also affect L2 learning. According to the SLM, it is hypothesised that as L1 phonetic categories develop, they become stronger attractors of unfamiliar L2 phonemes, which inhibit the forming of new L2 categories in a common phonological space. This suggests that it is relatively easier for younger learners with undeveloped L1 phonetic categories to form new L2 phonetic categories, compared to older learners with developed L1 phonetic categories. Taking into consideration these effects of biological brain plasticity and L1 phonological development on learning new L2 phonemes, a second experiment examined whether younger Japanese speakers improve their perception of the English /r/-/l/ contrasts in both phonetic and phonological levels more than older Japanese speakers through perceptual training.

Chapter 4 examined the relation between perception and production learning in L2. Regarding the acquisition of L1 speech, production ability is correlated with perception ability. As infants acquire phonemic categories through analysing distributional patterns of infant-directed speech, they start vocalising the phonetic patterns stored in their memory (Boysson-Bardies, 1993; Kuhl et al., 2008). By monitoring their own production of phonetic representations, they may develop the connection between perception and production (Kuhl et al., 2008). This suggests that perception and production are related, but perceptual ability seems to precede production. Brown and Berko (1960) demonstrated that children produce /fɪs/ for the word *fish*, but do not recognise /fɪs/ as the word *fish* in perception, and instead correctly identify the word /fɪʃ/ (the so-called “fis” phenomenon). However, in the case of the English /r/-/l/ contrast for Japanese speakers, it may not always be the case that perceptual ability precedes production, or that abilities of perception and production are correlated. Sheldon and Strange (1982) demonstrated that the production of the English /r/-/l/ contrast by Japanese speakers was more accurate than their perception. This suggests that there may be independent frameworks for perception and production in learning L2. Supporting this argument, a previous training study demonstrated no correlation in improvement between perceptual identification and production identifiability, although perceptual training improves both perceptual identification and production identifiability (Bradlow et al., 1997). Furthermore, Hattori (2009) demonstrated that articulatory training of English /r/ and /l/ does not improve perceptual identification accuracy. This chapter focused on a more detailed acoustical examination of the improvement in production of English /r/ and /l/ to investigate the relation between L2 perception and production.

Finally, in Chapter 5, the theoretical implications of the present research were discussed with particular regard to L1 interference at the phonetic and phonological levels and the influence of L1 development and neural plasticity. The efficacy of the perceptual training programme used in this research was considered, and recommendations for the development of future perceptual training programmes for learning L2 phonetic contrasts were discussed.

Chapter 2: Effects of identification and discrimination training techniques on English /r/-/l/ perception and production for adult Japanese speakers

2.1. Introduction

Late Japanese learners of English are often unable to distinguish the English /r/-/l/ contrast (Goto, 1971; Miyawaki et al., 1975). However, it has also been proved that intensive perceptual training can improve the L2 speech perception of late learners (Bradlow et al., 1997; MacKain, Best, & Strange, 1981). Several different training methods have been introduced to improve the English /r/-/l/ identification ability of Japanese speakers, but none have succeeded in improving perceptual performance to the level of native English speakers (Bradlow et al., 1997; Iverson et al., 2005; Logan et al., 1991).

This chapter describes the effects of computer-based perceptual training programmes for Japanese adults learning the English /r/-/l/ contrast. Previous studies have employed different training techniques to improve perception of this contrast. Strange and Dittmann (1984) examined the effects of a same-different discrimination training method with a synthetic stimulus series of the *rock-lock* contrast. The *rock-lock* stimuli were synthesised by interpolating between specific frequencies for F1, F2 and F3 and varying the onset of the F1 transition. During the same-different trials, standard stimuli (i.e., clear /r/-/l/ exemplars) were used for *same* trials, and standard stimuli were paired with the interpolated stimuli for *different* trials. Using this task, listeners were found to discriminate these acoustic differences within categories as well as between categories. The results demonstrated that discrimination training improved identification and discrimination at the /r/-/l/ category boundary. However, the improvement did not transfer to novel stimuli or natural recordings, suggesting that the discrimination training method was not practically useful for improving real-world English /r/-/l/ identification.

In contrast, identification training with high stimulus variability has been more successful. Logan et al. (1991) used an identification training technique with stimuli that were natural recordings of minimal-pair words spoken by multiple talkers, and found that Japanese speakers significantly improved identification accuracy of novel natural English /r/-/l/ minimal-pair words. Logan et al. (1991) claimed that the discrimination training method used by Strange and Dittmann (1984) failed to generalise to novel stimuli for three reasons. First, low-level sensory-based information learned through the discrimination training method was not helpful for identification accuracy, although the listeners became sensitive to acoustic differences. Second, the discrimination training technique used by Strange and Dittmann (1984) caused listeners to improve phonetic sensitivity within categories as well as between categories. Training perceptual sensitivity within categories was not useful for identification accuracy, because identification ability relies on perceptual sensitivity between categories. Finally, training with a synthetic stimulus series with only one phonetic environment did not transfer to other phonetic environments, because listeners began to learn context-sensitive phonetic categories rather than phonemic categories. Listeners need to be exposed to /r/-/l/ stimuli in multiple contexts produced by multiple talkers, in order to generalise across talker and context (Logan et al., 1991). The high variability of stimuli used in Logan et al. (1991) required listeners to pay attention to more robust acoustic cues separating the two categories. Japanese speakers trained in this way were able to alter their perceptual weighting of acoustic cues, and form new L2 categories with robust phonetic representations (Logan et al., 1991).

However, identification training with high stimulus variability has some limitations. Regardless of the length of the training programme, the average improvement rate in identifying the English /r/-/l/ contrast in word-initial position has only been approximately 15%, and identification accuracy has failed to reach near-perfect level (Bradlow et al., 1997; Iverson et al., 2005; Logan et al., 1991). Although Logan et al. (1991) suggested that identification with high stimulus variability might lead to the change in the weightings of different acoustic cues, Iverson et al. (2005) argued that the identification training method merely taught Japanese listeners a systematic strategy of using the closest L1 phoneme, Japanese /r/, in perceiving English /r/ and /l/. That is, the identification training technique did not alter Japanese listeners’ phonetic processing or establish their L2 phoneme categories at a phonological level, but it taught listeners how to apply their existing knowledge to identify a variety of stimuli.

Although the Perceptual Assimilation Model (PAM: Best & Strange, 1992; Best & Tyler, 2007) and the Speech Learning Model (SLM: Flege, 1995) support the idea that improving higher level categorisation is important for L2 phoneme identification accuracy, subsequent studies suggested there may be a benefit in altering the underlying phonetic processing of F3 as well as improving higher phonological categorisations. Hattori and Iverson (2009) demonstrated that individual differences in identification accuracy are not predicted by perceptual assimilation or the proximity of L2 categories to L1 categories, but by the representations of F3 frequencies for English /r/ and /l/. Iverson et al. (2003) demonstrated that unlike English speakers who are very sensitive to F3 at the English /r/-/l/ phoneme boundary, Japanese speakers are much less sensitive to this acoustic cue and instead show better discrimination of F2. In addition, Japanese speakers are also relatively sensitive to F3 *within* categories of English /r/ and /l/ compared to native English speakers (Iverson et al., 2003). These differences in auditory sensitivity may make it more difficult for Japanese speakers to identify English /r/ and /l/ (Iverson, Ekanayake, Hamann, Sennema, & Evans, 2008). This suggests that the problem of identifying English /r/ and /l/ for Japanese listeners stems from a phonetic processing level as well as a phonological processing level. Therefore, any perceptual training programme should target phonetic processing as well as phonological processing.

Auditory discrimination training may be able to alter the phonetic processing of Japanese speakers, because it improves sensitivity to acoustic contrasts (Jamieson & Morosan, 1986). So, three requirements for discrimination training can be proposed to alter Japanese speakers’ sensitivities. First, discrimination training should focus on the primary acoustic cue difference at the phoneme boundary between English /r/ and /l/, so that Japanese listeners may improve perceptual sensitivity to F3 at the boundary. Second, highly variable stimuli should be employed, in order to generalise any improvements to new phonetic environments and new talkers (Logan et al., 1991). Finally, signal processed stimuli should be used as signal processing provides fine-grained control over the manipulation of acoustic cues, and Iverson et al. (2005) found that signal processed stimuli provided comparable results to natual stimuli in phonetic training. Thus, discrimination training using a wide variety of signal processed stimuli based on multiple speakers and in a range of phonetic environments may enable improvements resulting from discrimination training to better generalise to novel natural stimuli.

Category discrimination training may also be able to alter Japanese speakers' underlying phonetic processing of F3. The category discrimination task reported in this chapter involves a three-alternative forced choice, and it requires listeners to choose one stimulus that is categorically different from the other two stimuli. The task requires participants to ignore the similarity of within-phoneme categories and to focus on the primary acoustic cue that makes one stimulus sound categorically different (Iverson, Pinet, & Evans, 2011; Logan et al., 1991). That is, listeners cannot use acoustic similarity but need to use phonological or phonetic encoding to successfully discriminate between categories (Flege, 2003; Højen & Flege, 2006; Iverson et al., 2011). As a result, listeners may be able to improve the perceptual sensitivity to F3 at the English /r/-/l/ phoneme boundary and to reduce the sensitivity to the irrelevant acoustic cues such as F2 at the boundary and F3 within the categories of English /r/ and /l/.

CHAPTER 2: EFFECTS OF IDENTIFICATION AND DISCRIMINATION TRAINING

The aims of the experiment reported in this chapter were to: (1) test whether the discrimination training method improves the identification accuracy for the English /r/-/l/ contrast, and (2) test whether the discrimination training technique supplements the effect of identification training. To address the first aim, it was hypothesised that the discrimination training may improve the identification accuracy of novel natural stimuli, because it focuses on training auditory sensitivity to F3 between English /r/-/l/ categories, and its training stimuli were highly variable in terms of phonetic environments and talkers. For the second aim, it was hypothesised that combined training of identification and discrimination would improve identification accuracy more than a single training method could achieve, as the two training methods are believed to affect different levels of processing. That is, identification training may result in improvements at the phonological level, whereas discrimination training may result in improvements at the phonetic level. Since previous studies suggested that identification training seems to provide only a limited improvement in identification accuracy at word-initial position (e.g., by approximately 15 % on average - Iverson et al., 2005), it is necessary to test whether the combination of two different training methods provides a greater improvement than either method in isolation.

To verify how Japanese speakers improve their perception of the English /r/-/l/ contrast, it is also worth examining the acoustic change in production after perceptual training. Although previous studies have reported improvements in Japanese speakers' production identifiability and intelligibility of the English /r/-/l/ contrast using a computer-based perceptual training programme (Bradlow, Akahane-

Yamada, Pisoni, & Tohkura, 1999; Bradlow et al., 1997), the speakers' production data was not acoustically analysed. If the two different training methods result in improvements at different levels of processing, they may also have different effects on production. Acoustical analyses may help show this in a fine-grained way.

Retention of learning was also investigated in this study. Although a previous study demonstrated that Japanese speakers retained improvement 6 months after the last perceptual training session (Lively et al., 1994), the present study examined the retention of acoustic improvement in production as well as the perceptual identification accuracy 4-6 weeks after completing the training programme. It may be possible that improvement through training triggers learning the contrast in daily life in English speaking environment, which may help improve the identification accuracy of the contrast more.

The identification training procedure was same as used in Iverson et al. (2005). Japanese speakers hear English minimal-pair words contrasting word-initial /r/ and /l/, and identify the initial consonant using a two-alternative forced choice task (i.e., English /r/ and English /l/). On the other hand, the discrimination task was newly developed and differed in three respects from the same-different task used by Strange and Dittmann (1984). First, a three-alternative forced choice discrimination task was used. Second, the variability of stimuli was increased using five speakers and 100 word-initial /r/-/l/ minimal-pair words (50 pairs). Finally, the discrimination training programme comprised three tasks: auditory discrimination with natural recordings (20% of trials), auditory discrimination with signal processed stimuli (40% of trials), and category discrimination with natural recordings (40% of trials).

The auditory discrimination with natural recordings allowed listeners to use all acoustic cues to choose an auditory-different stimulus. The auditory discrimination with signal-processed stimuli required listeners to perceive the F3 difference at the English /r/-/l/ boundary, because only F3 frequency was manipulated between minimal pairs. The category discrimination using natural recordings contained three different words whose initial consonant was /r/ or /l/ (e.g., *ray*, *lock*, *lamb*), and the listeners were required to click one of them whose initial consonant was categorically different from the others (e.g., *ray*).

Each training method consisted of five sessions, and Japanese listeners were trained using both identification (ID) and discrimination (DIS) training programmes. The order of the training programmes was balanced across participants so that half of the subjects performed identification training first (ID-DIS) and the other half performed discrimination training first (DIS-ID). All subjects thus received 10 training sessions in total. Three identical perception tests were conducted at pre-training, mid-training after five sessions of the first training method, and post-training after five sessions of the second training method. Each test included three perception tasks: (1) identification of English minimal-pair words contrasting word-initial /r/ and /l/, word-medial /r/ and /l/, and /r/ and /l/ in consonant clusters; (2) auditory discrimination of F3 at the English /r/-/l/ boundary, F2 at the English /r/-/l/ boundary and F3 within the English /r/ category; and (3) category discrimination. Production tests were also used at each stage to verify the effects of perceptual training on production. After 4-6 weeks of the post-training test, a further test was conducted for half the subjects to examine the retention of any training effects.

2.2. Method

2.2.1. Subjects

Fifty-five native Japanese speakers completed the pre-training test. Of these subjects, 12 were not trained because their identification accuracy was more than 75% correct, and two were excluded because they were not able to finish all sessions. This left 41 subjects in total for the data analysis.

Table 2.1 describes the information of Japanese subjects who were included in the data analysis. All subjects were native Japanese speakers with no self-reported hearing impairments. As described in the table, the number of subjects, their age, tested location and English experience (i.e., length of living in English speaking countries) were balanced between the two trainer order groups, ID-DIS and DIS-ID. In order to collect a greater variety of English experience, 22 subjects whose length of residence in English-speaking countries ranged from 1 month to 21 years and 5 months (median = 13.5 months) were tested in London, UK, and another 19 subjects whose length of living in English speaking countries ranged from 0 to 2 months (median = 0 months) were tested in Kanto area, Japan.

Trainer order	Number of subjects (female, male)	Age range (median)	Tested location (number of subjects)	Length of living in English-speaking countries (median)
Identification-Discrimination (ID-DIS)	20 (14, 6)	20-61 years (25.5 years)	UK (11)	0 month – 21 years and 5 months (2 months)
			Japan (9)	
Discrimination-Identification (DIS-ID)	21 (14, 7)	21-57 years (25 years)	UK (11)	0 month – 20 years and 5 months (2.4 months)
			Japan (10)	
Total	41 (28, 13)	20-61 years (25 years)	UK (22)	0 month – 21 years and 5 months (2 months)
			Japan (19)	

Table 2.1. Subject information: Number of subjects, sex, age, tested location, English experience (length of living in English-speaking countries) in each trainer order, identification-discrimination (ID-DIS) and discrimination-identification (DIS-ID).

2.2.2. Apparatus

The subjects were trained using headphones or earphones and laptops or laboratory PCs. They were allowed to adjust the loudness of sounds to a comfortable level. Training logs were automatically recorded and password protected, so that the subjects could not read or change the information in the logs. For testing, subjects were tested using a headphone (Sennheiser HD 280 PRO) connecting to a laboratory PC in sound-booths in London or to a laptop (Dell inspiron 1370) in a quiet room in Japan. For production tests, the microphone used in London was Rode NT1-A, and that used in Japan was Audio Technica AT2020 USB.

2.2.3. Stimuli

2.2.3.1. Natural recordings. The natural stimuli were the same as those recorded by Iverson et al. (2005). The stimuli from 12 adult Southern Standard British English (SSBE) speakers (6 females, 6 males) were digitally recorded with

44,100 16-bit samples per second, and downsampled to 22,050 samples per second. The English speakers read tokens displayed on screen one by one which was randomly taken from a list of English words including /r/-/l/ minimal pairs and dummy tokens. One hundred word-initial /r/-/l/ minimal-pair words (e.g., *lay* and *ray*) spoken by 10 speakers were used as training stimuli (Appendix A). These were used for the tasks of identification, auditory discrimination with natural stimuli and category discrimination. Another 120 minimal-pair words by two speakers were used as testing stimuli. The 120 testing stimuli which were used for identification test were 40 word-initial /r/-/l/ minimal-pair words (e.g., *rope* and *lope*), 40 word-medial /r/-/l/ minimal-pair words (e.g., *berries* and *bellies*) and 40 consonant cluster /r/-/l/ minimal-pair words (e.g., *fresh* and *flesh*) (Appendix B). For the category discrimination test, a subset of 24 minimal-pair words from the list of 40 word-initial /r/-/l/ minimal-pair words were used. The minimal-pair words used as testing stimuli were different from those in the training corpus.

2.2.3.2. Synthetic stimuli. To examine how phonetic sensitivities are changed after perceptual training, synthetic stimuli of English /ra/ and /la/ were developed using a Klatt synthesizer as in Hattori and Iverson (2009). The synthetic model was based on a female Standard Southern British English (SSBE) speaker, and the stimuli were manipulated with five dimensions of acoustic cues such as closure duration, transition duration, F1, F2 and F3. The target acoustic cues with which Japanese speakers were tested are F3 at the English /r/-/l/ boundary, F2 at the English /r/-/l/ boundary and F3 within the English /r/ category. For the stimulus pair contrasting F3 at the boundary, these acoustic cues of closure duration, transition duration, F1 and

F2 were set at 64 ms, 48 ms, 327 Hz and 1196 Hz. The target F3 was contrasted between English /r/ and /l/. F3 was set at 2639 Hz for the /r/ stimulus and 3328 Hz for the /l/ stimulus. Similarly, for the stimulus pair contrasting F2 at the boundary, the acoustic cues of closure duration, transition duration, F1 and F3 were set at 64 ms, 48 ms, 327 Hz and 2965 Hz. Because the target F2 does not distinguish English /r/ and /l/, the stimulus pair was contrasted between low and high. Low F2 was set at 1051 Hz and high F2 was set at 1358 Hz. Finally, for the stimulus pair contrasting F3 within the English /r/ category, the acoustic cues of closure duration, transition duration, F1 and F2 were set at 31 ms, 81 ms, 327 Hz and 1196 Hz. The low F3 was set at 1739 Hz and the high F3 was set at 2212 Hz. Consequently, six stimuli of three parameters (i.e., F3 at the English /r/-/l/ boundary, F3 at the boundary and F3 within boundary) with two contrasts (i.e., high vs. low) were developed and used for the auditory discrimination test.

2.2.3.3. Speech Processing. For the auditory discrimination training task, six minimal pairs (i.e., *lamb-ram*, *lay-ray*, *leer-rear*, *limb-rim*, *long-wrong*, and *low-row*) were selected from the recordings of each of the 10 speakers and were re-synthesised within Praat. All acoustic cues except F3 were averaged between /r/ and /l/ minimal-pair words, and LPC filters with formant information from F1 to F4 were created from these averages for each minimal pair. After filtering LPC residuals of the 10 speakers with the LPC formant filters, the intensity of each filtered stimulus was changed to the averaged intensity of each minimal pair. The pitch contour of each stimulus was also manipulated to match the /r/ or /l/ pitch contour of the original word. Finally, the generated minimal-pair stimuli were interpolated by manipulating F3. Six stimuli were evenly interpolated between /r/ and /l/, and two extra-stimuli with 150% enhanced F3 on both sides of /r/ and /l/ were also generated. In this way, eight stimuli were created for each minimal pair.

In order to ensure that the stimuli were identifiable, a preliminary listening test was conducted. Thirteen native British English speakers participated in a two-alternative forced choice English /r/-/l/ identification test. Eight stimuli from the six minimal pairs produced by the 10 speakers were presented over 480 trials. For each of the 10 speakers, the best four of the six minimal pairs were selected based on the place of phoneme boundaries and the variance between the phoneme boundary places identified by the 13 English listeners. These two variables were calculated using psychometric graphics in R software (version 2.15.2), and the four minimal pairs whose phoneme boundary places varied least between the English listeners were selected for each speaker. Finally, the best five speakers were selected from the 10 speakers to use as the speakers for the auditory discrimination training stimuli.

For each of the four best minimal pairs selected from the six (i.e., *lamb-ram*, *lay-ray*, *leer-rear*, *limb-rim*, *long-wrong*, and *low-row*), two pairs of English /r/ and /l/ contrasts were created. Because the aim of the auditory discrimination training with signal-processed stimuli was to make Japanese speakers pay attention to the auditory difference in F3 at the English /r/-/l/ boundary, the pairs of English /r/ and /l/ were placed close to phoneme boundaries. One of the two pairs was placed very close to the phoneme boundary (i.e., close pair), while the other was the distant pair, judged as the target stimulus in the preliminary listening test. The close pairs were the English /r/ and /l/ stimuli which were often judged as a target stimulus (i.e., identified as English /r/ by 70% of native English speakers, with the other 30% of native English speakers confusing the stimulus with English /l/). The distant pairs of English /r/ and /l/ stimuli were those judged as the target stimulus (e.g., English /r/) by 85% of native English speakers and with only 15% of native English speakers confusing them with the other stimulus (e.g., English /l/). Since the differences of F3 between English /r/ and /l/ in the close pairs are not as great as those in the distant pairs, it is more difficult to identify or discriminate. Thus, the stimuli for auditory discrimination training were generated from two pairs (i.e., a close pair and a distant pair) for each of the four best minimal pairs produced by five speakers (Appendix C).

2.2.4. Procedure

2.2.4.1. Training. All subjects took part in both identification and discrimination training programmes. Each subject was assigned to either the identification-discrimination trainer order (ID-DIS) or the discrimination-identification trainer order (DIS-ID). Age, sex and the identification accuracy of word-initial /r/-/l/ minimal-pair words at pre-training test were balanced between the two trainer order groups. Each training programme comprised five sessions, so there were 10 training sessions in total. The subjects were instructed not to do more than one training session per day, but three subjects in Japan did several training sessions within a day. The speaker of the training stimuli was different each day, but the speaker order for each training programme was consistent. Each training session took approximately 30 minutes, and all subjects completed all 10 training sessions within 10 to 28 days.

CHAPTER 2: EFFECTS OF IDENTIFICATION AND DISCRIMINATION TRAINING

The identification training programme was the same as that used in Iverson et al. (2005). Training started with the English instruction, “Hello, my name is Ian. You’re going to hear my voice in the training today. Let’s get started” spoken by a synchronised animated face (Iverson et al., 2005, p. 3271). The subjects completed 300 two-alternative forced choice trials per session which comprised 100 word-initial /r/-/l/ minimal-pair words (50 pairs) repeated 3 times each. The computer screen displayed a minimal-pair (e.g. *rock*, *lock*) with a single auditory token (e.g. *rock*), and the listeners clicked the word that they thought they heard. The subjects could not listen to a trial stimulus more than once. If they answered correctly, they saw a message of *Correct* on screen with a cash register sound, and the stimulus was repeated with highlighted answer. For incorrect answers, a message *Wrong* was displayed on screen with the wrong choice crossed out and two descending beep sounds. With a highlighted visual prompt, the correct answer was repeated twice. During training, listeners could see the overall percentage of correct answers and the percentage of trials completed. After completing 300 trials, they took a short identification test of 20 trials spoken by the same talker that was used for all training in that particular session. There was no feedback while doing the short identification test, but subjects could see the percentage of correct responses at the end of the short test.

In the discrimination training, one session consisted of three tasks: auditory discrimination with natural recording (40 trials, 20% of all trials), auditory discrimination with signal-processed stimuli (80 trials, 40% of all trials) and category discrimination (80 trials, 40% of all trials). The discrimination training programme had 200 trials in total to make sure that listeners spent approximately the same amount of time as they spent on the identification training programme. The three-alternative forced choice discrimination task was same for all three sections. Three numbers (①, ②, ③) were displayed on screen, and Japanese speakers listened to three stimuli and clicked one of them whose word-initial phoneme was categorically different from the other two stimuli. The subjects could not listen to a stimulus more than once. For correct answers, a message *Correct* was displayed on screen without replaying the trial; for wrong answers the message *Wrong* was shown on screen and all 3 stimuli were repeated once with the correct number visually highlighted. The listeners could see the percentage of correct responses and the percentage of trials finished during training. The discrimination training also contained the short identification test spoken by the same speaker used in the training session. The subjects did not receive feedback during the short test, but they saw the percentage of correct responses at the end of the test.

2.2.4.2. Pre-training, mid-training, post-training and retention tests. All Japanese subjects participated in three tests, pre-training, mid-training after five sessions of a first training programme, and post-training after five sessions of a second training programme. The subjects in London also participated in the retention test. The materials of each test were identical.

The perception test comprised three different tasks: (1) identification of /r/-/l/ minimal-pairs in three different /r/-/l/ positions such as word-initial, word-medial and consonant cluster positions, (2) auditory discrimination of three stimulus pairs such as F3 at English /r/-/l/ boundary, F2 at English /r/-/l/ boundary and F3 within English

/r/ category, and (3) category discrimination. For identification, there were 40 minimal-pair words for each /r/-/l/ position so that subjects completed 120 trials in total (Appendix B). The 40 minimal pair words (20 pairs) were produced by two native English speakers (a female, a male), and each speaker produced different 20 minimal-pair words. For auditory discrimination, there were 72 trials consisting of 3 stimulus pairs with 2 possible odd stimuli of English /r/ and /l/, 3 positions for the odd stimulus and 4 repetitions for each trial. As for category discrimination, Japanese speakers completed 48 trials with 4 kinds of trial pairs with 2 possible odd stimulus of English /r/ and /l/, 3 positions for the odd stimulus and 2 speakers (a female, a male) for these stimuli (Appendix D). During the tests, subjects did not receive any feedback, and they were not allowed to listen to the same trial again. In order to test whether the improvements generalised to novel stimuli, the listeners were tested with new talkers and new words that were not used in training corpora.

For the production tests, the tasks were (1) pronouncing isolated 40 word-initial /r/-/l/ minimal-pair words which were used in the identification test and (2) reading the first third of “The Rainbow Passage” (Fairbanks, 1960) as used in Hattori (2009) (Appendix E). For the analysis of the word production task, 10 minimal-pair words were selected based on the accuracy of the following vowel’s pronunciation. Because Japanese speakers often mispronounced the vowels following English /r/ and /l/, the F3 was often affected by the mispronounced vowels due to co-articulation. The selected 10 minimal-pair words were rarely mispronounced, so that the F3 difference between English /r/ and /l/ could be more accurately analysed. Those selected minimal pairs were *race-lace*, *road-load*, *root-loot*, *rung-lung* and *wrist-list*. For the passage recordings, all 13 word-initial /r/-/l/ words involved in the passage were analysed. Those were seven /r/-words (i.e., *raindrops*, *reach*, *round*, and *rainbow* x 4) and six /l/-words (i.e., *legend*, *light*, *long*, *look*, *looking* and *looks*). F3 frequencies of the closure parts in English /r/ and /l/ were measured for both tasks, but 145 out of 2829 tokens were excluded from the analyses due to the unreliable F3 tracking measurements with Praat, leaving 2684 tokens.

2.3. Results

2.3.1. Identification

Figure 2.1 displays the identification accuracy of Japanese speakers in the two trainer orders at the pre, mid and post tests. As shown in Figure 2.1, both ID-DIS and DIS-ID trainer orders improved identification accuracy after the 10 training sessions. Moreover, both training methods of identification and discrimination improved identification accuracy for the first five training sessions, but identification training seemed to yield a greater improvement compared to discrimination training.

To verify the effects of training, a logistic mixed effects model was built for the identification analysis based on the correct/incorrect binomial responses. The best-fitting model was chosen with top-down approaches, that is, excluding ineffective random and fixed factors from a model with all potential factors. The best-fitting model included the fixed factors of testing block (pre, mid, post), trainer order (ID-DIS, DIS-ID), /r/-/l/ position (word-initial, word-medial, consonant cluster) and English experience (less experienced, more experienced). The categories of English experience were based on the length of living in English-speaking countries. The less experienced subjects had stayed in those countries for less than 2 months; the more experienced subjects had stayed in those countries for 2 months or more. Some interactions between these fixed factors also fitted into the model. The included interaction factors in the model were between testing block and trainer order, between testing block and /r/-/l/ position, between testing block and English experience, and between trainer order and /r/-/l/ position. The random factors were crossed intercepts for subject and stimulus. Stimulus in the random factors was nested within speakers, because the stimuli (words) were produced by different speakers. However, subject was not nested within any other random intercept, as all subjects took the same stimuli from the same speakers. Random slopes for testing block were excluded.

The logistic mixed-effects model on identification accuracy demonstrated that there was a significant main effect of testing block (pre, mid, post tests), $\chi^2(2) = 179.18$, $p < .001$. Although there was no significant main effect of trainer order (ID-DIS, DIS-ID), $\chi^2(1) = 1.24$, $p > .05$, the interaction between testing block and trainer order was significant, $\chi^2(2) = 14.09$, $p < .001$. These results confirmed that there was a significant improvement in the identification accuracy, and the improvement was significantly different with trainer order. The planned orthogonal contrast for the interaction between testing block and trainer order demonstrated that identification training improved identification accuracy significantly more than discrimination training for the first five sessions (pre vs. mid test), $b = -.08$, SE = .02, z = -3.63, $p < .001$, although the contrast for the last five sessions (pre & mid tests vs. post test) was not significant, $b = -.01$, SE = .01, z = -1.07, $p > .05$. This suggests that identification training was better at improving identification accuracy than discrimination training.

For testing the effect of each training method, post-hoc analyses were conducted. The results demonstrated that both identification training, *b* = .30, SE = .03, z = 9.11, *p* < .001, and discrimination training, *b* = .12, SE = .03, z = 4.07, *p* < .001, significantly improved identification accuracy of novel stimuli during the first five sessions. These suggest that both identification and discrimination training with a high-variability technique are useful for the improvement of identification accuracy. However, in the absence of a control group which received no training, it is not yet clear if the improvements seen with discrimination training are actually due to training.

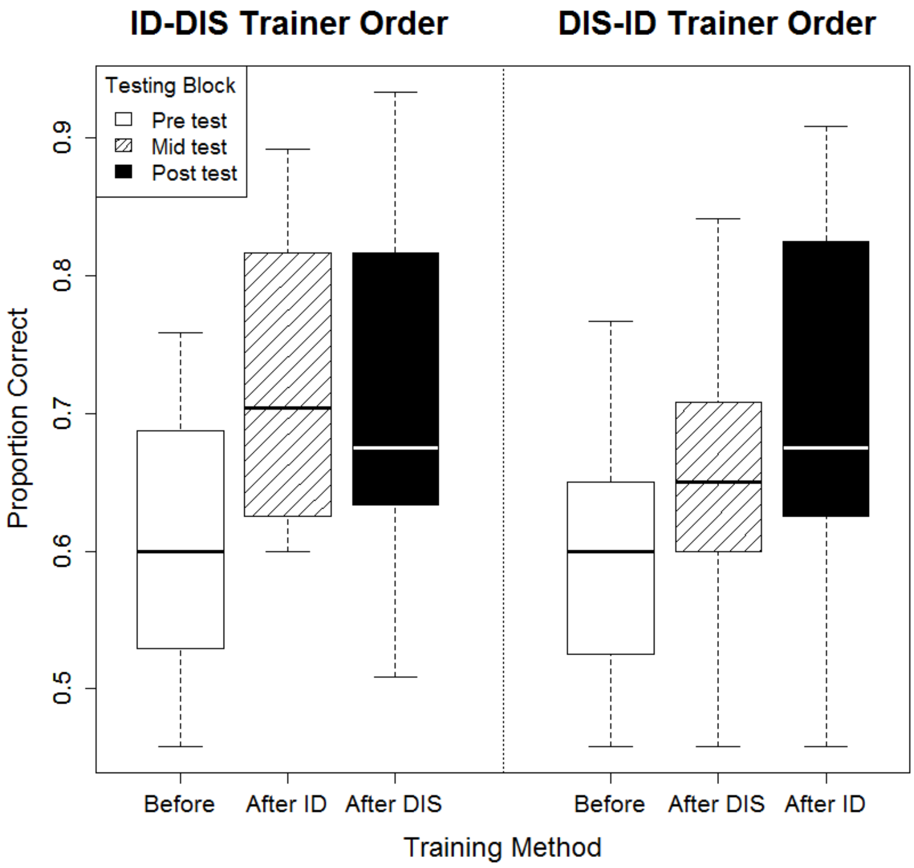


Figure 2.1. Boxplots of identification accuracy for the English /r/-/l/ contrast by adult Japanese speakers in two trainer orders, ID-DIS (left) and DIS-ID (right), at pre test (white boxes), mid test after 5 sessions of a first training method (shaded boxes) and post test after 5 sessions of a second training method (black boxes).

Figure 2.2 displays identification accuracy by /r/-/l/ position at pre, mid and post tests. Although Japanese speakers seemed to improve their identification accuracy of the trained word-initial position more than the other untrained word-medial and consonant cluster positions, their identification accuracy at all three positions was improved through 10 training sessions. This suggests that training effects transferred to the untrained /r/-/l/ positions.

This was verified with the logistic mixed effects model which demonstrated that there was a significant main effect of /r/-/l/ position, $\chi^2(2) = 14.56$, $p < .001$, and a significant effect of interaction between testing block and /r/-/l/ position, $\chi^2(4) = 18.27$, $p < .01$. The planned orthogonal contrasts of the interaction revealed that Japanese speakers improved their identification ability significantly more in word-initial /r/-/l/ positions than in other positions for both the first five sessions from pre to mid test, $b = -.05$, SE = .02, z = -3.19, $p < .01$, and the last five sessions from pre & mid tests to post test, $b = -.02$, SE = .01, z = 2.24, $p = .025$. In order to test the generalisation to untrained /r/-/l/ positions, post-hoc tests were also conducted. There was a significant improvement from pre to post test at word-medial /r/-/l/ position, $b = .28$, SE = .04, z = 7.02, $p < .001$, M_{pre} = 61.16%, M_{post} = 71.59%, and at consonant cluster position, $b = .19$, SE = .04, z = 5.10, $p < .001$, M_{pre} = 57.26%, M_{post} = 65.55%, suggesting that the effects of word-initial /r/-/l/ perceptual training transferred to the untrained /r/-/l/ positions, although the improvement in the trained word-initial position was still higher than in the other untrained positions.

Identification accuracy in the word-initial position was an average of 62.38% at pre test and 77.32% at post test, suggesting that the improvement was still limited to approximately 15%, as Iverson et al. (2005) indicated.

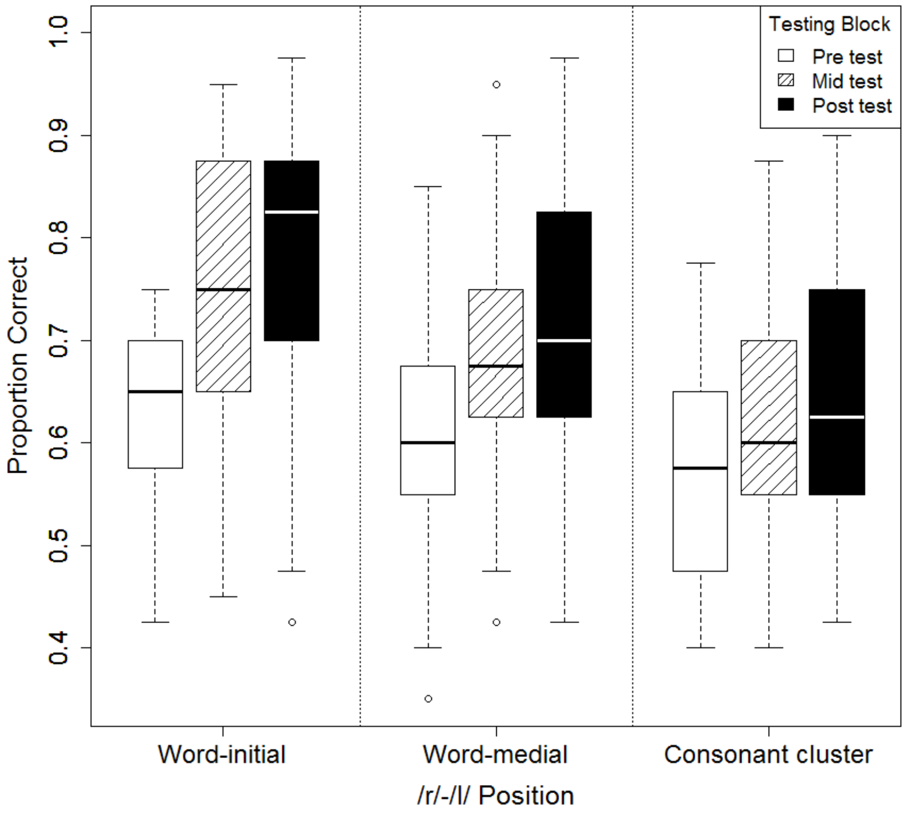


Figure 2.2. Boxplots of identification accuracy for the English /r/-/l/ contrast at word-initial, word-medial and consonant cluster positions by adult Japanese speakers at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

Figure 2.3 displays identification accuracy at pre, mid and post tests by Japanese learners who had lived in English-speaking countries for less than 2 months (less experienced) and for 2 months and more (more experienced). Both English experience groups improved identification accuracy for the first five training sessions, but the more experienced group did not seem to improve their identification accuracy for the last five session as much as the less experienced group, suggesting that the less experienced group improved identification accuracy more through the full training of 10 sessions.

The logistic mixed effect model for identification accuracy demonstrated that there was no significant main effect of English experience, $\chi^2(1) = .02$, $p > .05$, but the interaction between testing block and English experience was significant, $\chi^2(1) = 19.09$, $p < .001$. The planned orthogonal contrasts for the interaction demonstrated that Japanese speakers with less English experience improved their identification ability more than the experienced group for the last five training sessions, $b = -.06$, SE = .01, z = -4.37, $p < .001$, although there was no significant difference between the two English experience groups for the improvement for the first five training sessions, $b = -.01$, SE = .02, z = -.29, $p > .05$. This suggests that Japanese speakers with less English experience improved the identification accuracy more than Japanese speakers with more English experience over the course of 10 training sessions. However, due to the fact that the English experience was highly correlated with age, $r = .73$, n = 39, $p < .001$, it cannot be concluded that this effect can entirely be attributed to English experience only; it may be attributed to age or both age and English experience.

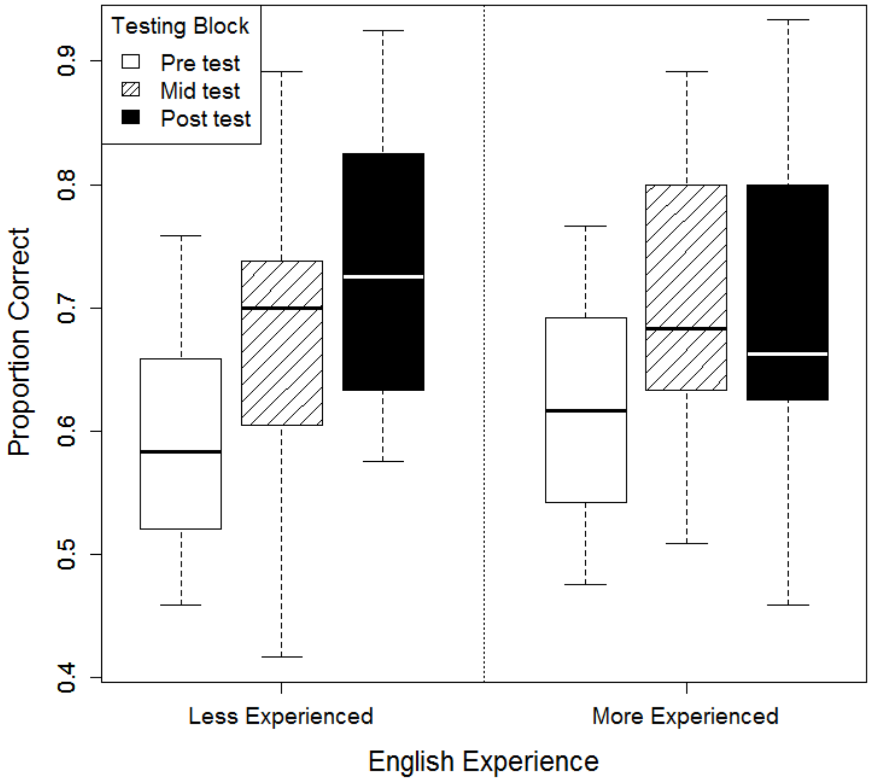


Figure 2.3. Boxplots of identification accuracy at pre, mid and post tests, by adult Japanese speakers who have stayed in English-speaking countries less than 2 months (Less experienced) and those who have stayed in English-speaking countries for 2 months or more (More experienced).

The 3-way interaction of testing block, trainer order and /r/-/l/ position was excluded from the best-fitting model on identification accuracy during the process of model comparisons. Moreover, the best logistic mixed effects model included neither the 3-way interaction of testing block, trainer order and English experience nor the 4-way interaction of testing block, trainer order, /r/-/l/ position and English experience. These results suggest that there was no significant effect of these 3-way or 4-way interactions.

2.3.2. Auditory Discrimination

Figure 2.4 displays the auditory discrimination accuracy for each stimulus pair at pre, mid and post tests. As shown in the figure, Japanese speakers seemed to improve their perceptual sensitivities to F2 at the English /r/-/l/ boundary most. They also improved their F3 sensitivity at the /r/-/l/ boundary, whereas they did not improve their F3 sensitivity within the English /r/ category through 10 training sessions. This suggests that Japanese speakers selectively improved their underlying phonetic sensitivities, but they may still have relied on their F2 sensitivity in listening to the English /r/-/l/ contrasts.

A logistic mixed effects model was built for auditory discrimination based on correct/incorrect binomial responses. The process of building a best-fitting model was the same as for the identification analyses. Instead of the fixed factor of /r/-/l/ position, stimulus pair (F3 at the /r/-/l/ boundary, F2 at the /r/-/l/ boundary, F3 within /r/ category) was included in the model. As a result of model comparisons, the best-fitting model included fixed factors of testing block, trainer order, stimulus pair (F3 difference at boundary, F2 difference at boundary, F3 difference within /r/ category), and 2-way interactions between testing block and stimulus pair, between trainer order and stimulus pair, and between testing block and English experience. The random factors in this model were crossed intercepts for subject and stimulus.

The logistic mixed effects model demonstrated that there were significant main effects of testing block, $\chi^2(2) = 31.03$, $p < .001$, and of stimulus pair, $\chi^2(2) = 28.55$, $p < .001$. The interaction between testing block and stimulus pair was also significant, $\chi^2(4) = 10.19$, $p = .037$, suggesting that the amount of improvement was different between the three stimulus pairs. Planned orthogonal contrasts for the interaction demonstrated that Japanese speakers improved F2 sensitivity at the /r/-/l/ boundary more than F3 sensitivity at the boundary and F3 sensitivity within the /r/ category from pre to mid test, b = -.05, SE= .02, z = -2.43, p = .015, although there was no significant difference in the contrast for the last five sessions (pre & mid tests vs. post test), b = .01, SE= .01, z = .83, p > .05. In addition, the improvement in F3 sensitivity at the /r/-/l/ boundary was not significantly higher than the improvement in F3 sensitivity within /r/ category for the first five sessions (pre test vs. mid test), b = .06, SE= .03, z = 1.82, p = .069, or for the last five sessions, b = .01, SE= .02, z = .46, p > .05 (pre & mid tests vs. post test). Post-hoc analyses demonstrated that Japanese speakers significantly improved sensitivity to F3 at the /r/-/l/ boundary from pre to post test, b = .13, SE = .05, z = 2.62, p < .01, as well as F2 at the /r/-/l/ boundary, b = .24, SE = .05, z = 4.70, p < .001, but not F3 within the English /r/ category, b = .09, SE = .05, z = 1.88, p = .061.

These results suggest that Japanese subjects' phonetic sensitivities did not become like those of native English speakers. Previous studies demonstrated that English speakers are sensitive to F3 differences at the /r/-/l/ boundary but not sensitive to F2 differences at the boundary or F3 differences within the English /r/ category, compared to Japanese speakers (Hattori, 2009; Iverson et al., 2003). The present study aimed to make Japanese speakers acquire such contrasts of phonetic sensitivities, but they did not achieve native-like phonetic sensitivities (i.e., sensitive to the primary acoustic cue but not to the irrelevant acoustic cues). Although

Japanese speakers’ phonetic sensitivities cannot be compared to native English speakers’ ones without an English control group, it can be assumed that their phonetic sensitivity contrasts did not become native-like. This assumption is supported by the results of Hattori (2009) which used the same stimuli for his AX discrimination task and tested native English speakers’ phonetic sensitivities.

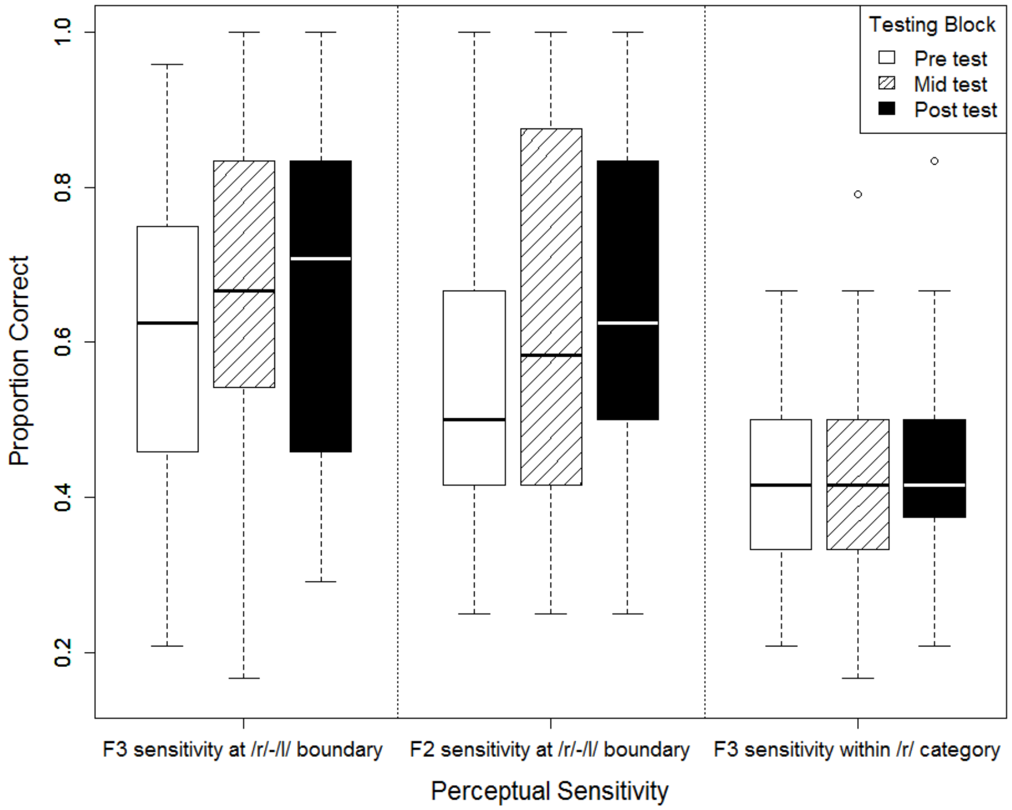


Figure 2.4. Boxplots of auditory discrimination accuracy in three stimulus pairs testing F3 sensitivity at English /r/-/l/ phoneme boundary, F2 sensitivity at English /r/-/l/ phoneme boundary, and F3 sensitivity within English /r/ category, for adult Japanese speakers at pre (white boxes), mid (shaded boxes) and post (black boxes) tests.

Figure 2.5 displays the auditory discrimination accuracy for each stimulus pair at pre, mid and post tests by Japanese speakers in the two trainer orders, ID-DIS and DIS-ID. The improvement in each stimulus pair such as F3 at the boundary, F2 at the boundary and F3 within the /r/ category appeared to be not significantly different between the two trainer orders.

The logistic mixed effect model did not include the 2-way interaction between testing block and trainer order or the 3-way interaction of testing block, trainer order and stimulus pair, because these two interactions did not fit into the best model. This suggests that both identification and discrimination training affected the underlying phonetic sensitivities in a similar way, and neither of the two trainer orders changed the perceptual sensitivities of Japanese speakers to native-like.

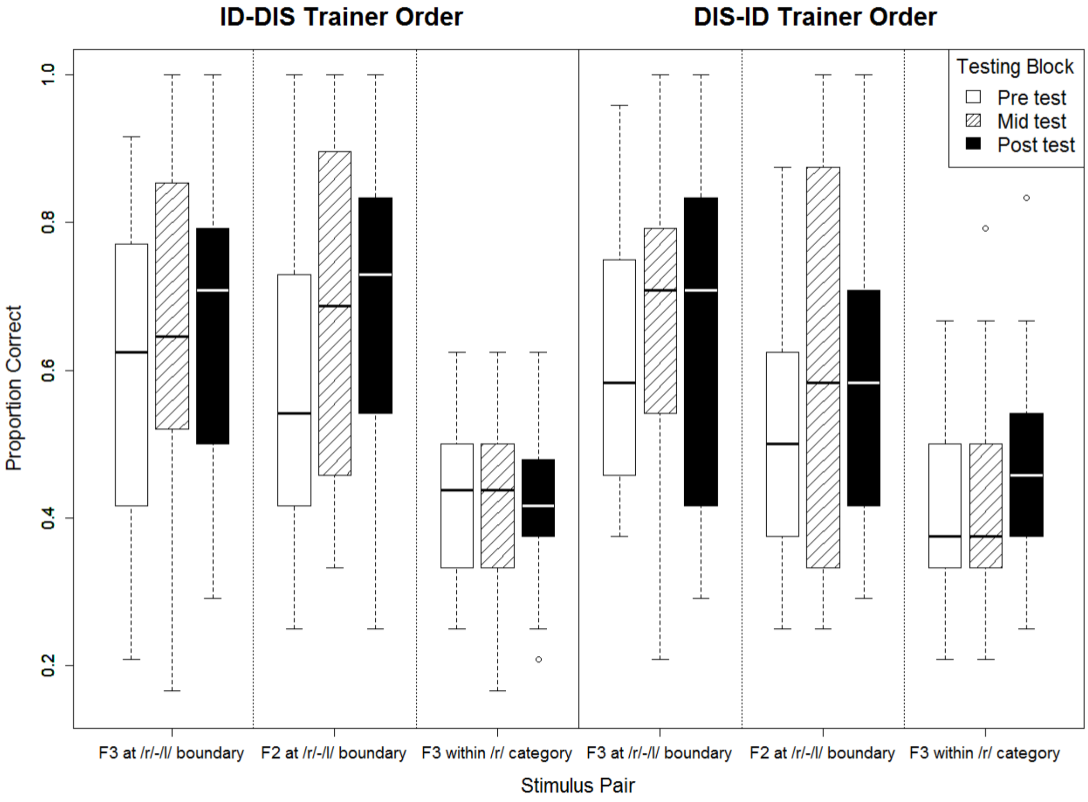


Figure 2.5. Boxplots of auditory discrimination accuracy in three stimulus pairs testing F3 sensitivity at English /r/-/l/ phoneme boundary, F2 sensitivity at English /r/-/l/ phoneme boundary, and F3 sensitivity within English /r/ category, for adult Japanese speakers in the two different trainer orders, ID-DIS (left) and DIS-ID (right), at pre (white boxes), mid (shaded boxes) and post (black boxes) tests.

Figure 2.6 displays the auditory discrimination accuracy of Japanese learners of English with less experience and more experience at pre, mid and post tests for each stimulus pair. Although the discrimination accuracy for each test was different between English experience groups, the improvement for each stimulus pair seemed to be unaffected by English experience. The 3-way interaction of testing block, stimulus pair and English experience was excluded from the best-fitting model, suggesting that the English experience did not significantly affect the degree of improvement difference between stimulus pairs. That is, Japanese speakers changed

F3 and F2 sensitivities at the /r/-/l/ boundary regardless the amount of English experience.

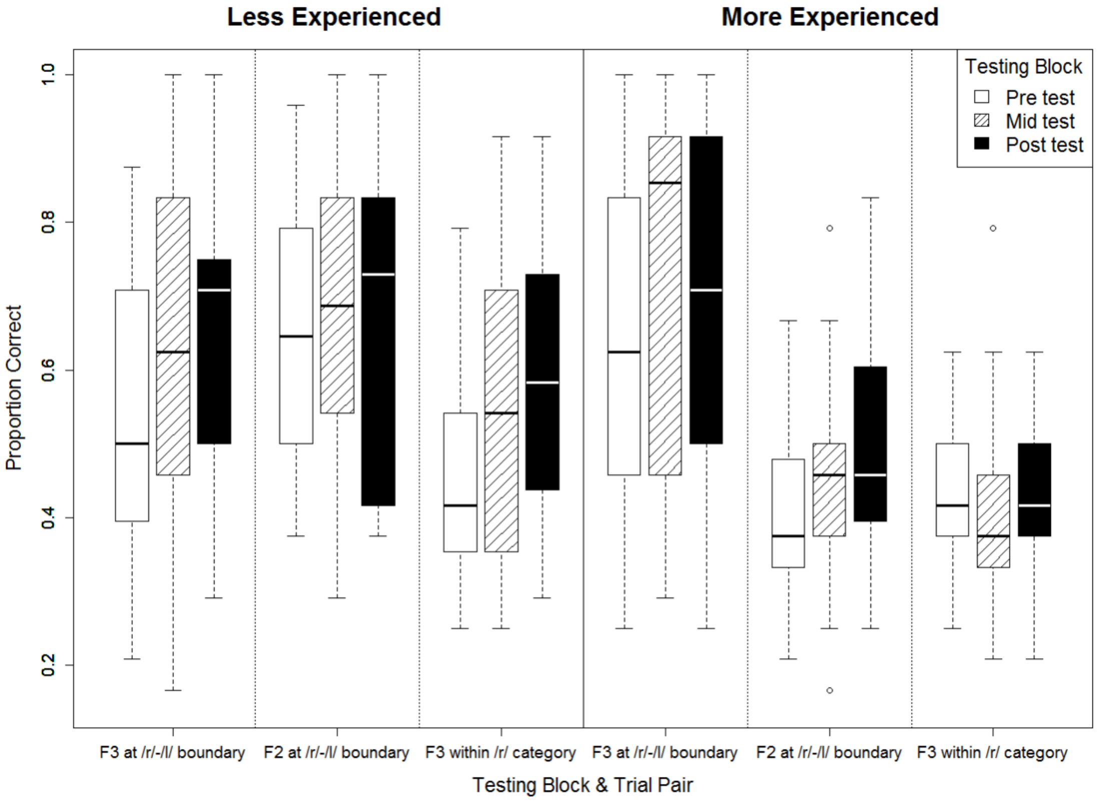


Figure 2.6. Boxplots of auditory discrimination accuracy in three stimulus pairs testing F3 sensitivity at English /r/-/l/ phoneme boundary, F2 sensitivity at English /r/-/l/ phoneme boundary, and F3 sensitivity within English /r/ category, for Japanese speaker who have stayed in English-speaking countries less than 2 months (left) and those who have stayed in English speaking countries for 2 months or more (right), at pre (white boxes), mid (shaded boxes) and post (black boxes) tests.

2.3.3. Category Discrimination

Figure 2.7 displays the category discrimination accuracy by Japanese speakers at pre, mid, and post tests. As shown in the figure, Japanese speakers in both the ID-DIS and DIS-ID trainer orders improved their category discrimination accuracy, and there was no difference in the improvement between the two trainer orders for the first five training sessions or the last five training sessions. This suggests that both identification and discrimination training improved the category discrimination accuracy to similar extents.

The best-fitting logistic mixed effects model for category discrimination included fixed factors of testing block, trainer order, English experience, and all possible 2-way and 3-way interactions. The random factors were intercepts for subject and stimulus. The logistic mixed effect model demonstrated that there was a significant main effect of testing block, $\chi^2(2) = 137.81$, $p < .001$, but the main effect of trainer order was not significant, $\chi^2(1) = 1.91$, $p > .05$. The interaction between testing block and trainer order was also not significant, $\chi^2(2) = 4.66$, $p = .097$. These results suggest that Japanese speakers significantly improved category discrimination, and there was no significant difference in the improvement between the two trainer orders. These results also suggest that the two training methods had similar effects on the improvement of the category discrimination accuracy.

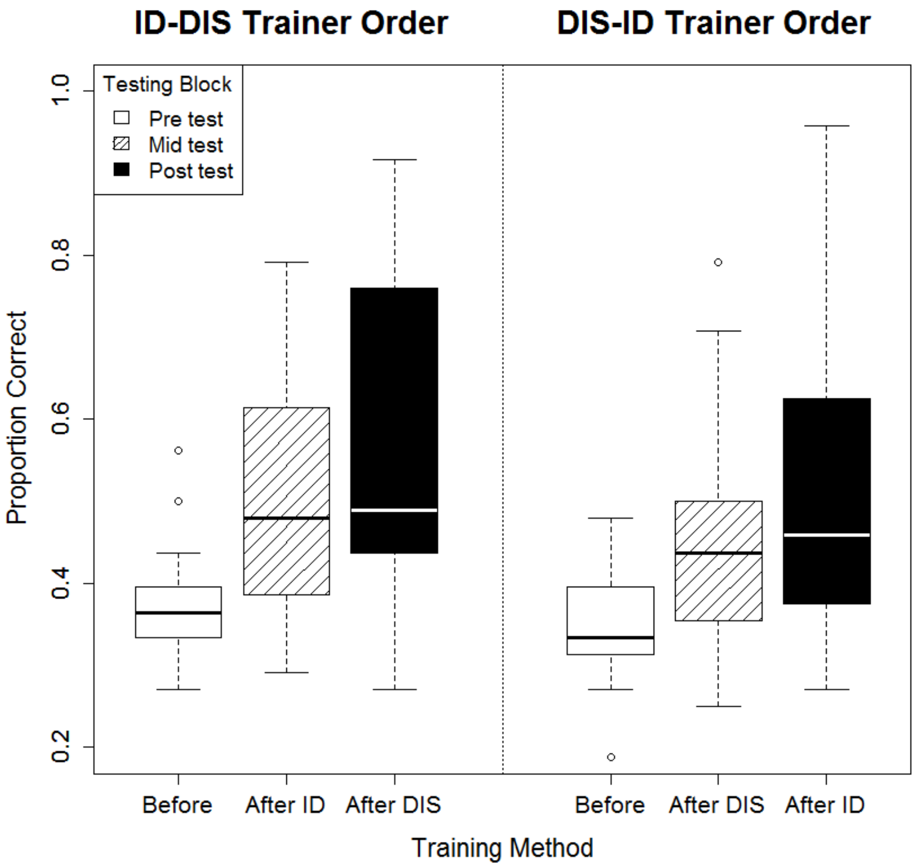


Figure 2.7. Boxplots of category discrimination accuracy at pre (white boxes), mid (shaded boxes) and post (black boxes) tests by adult Japanese speakers in two trainer orders, ID-DIS (left) and DIS-ID (right).

Figure 2.8 displays the category discrimination accuracy at pre, mid and post tests by the two English experience groups. Although both the less and the more experienced groups improved category discrimination accuracy, there appeared to be a difference in the improvement between the two English experience groups. Japanese speakers with less English experience improved category discrimination accuracy more than the experienced group over the 10 trianing sessions.

The logistic mixed effects model demonstrated that although there was no significant main effect of English experience, $\chi^2(1) = .02$, $p > .05$, the effect of the interaction between testing block and English experience was significant, $\chi^2(2) = 7.22$, $p = .027$. The planned orthogonal contrasts for the interaction demonstrated that the improvement by Japanese speakers with less English experience was significantly higher than the Japanese speakers with more English experience for the first five training sessions, $b = .08$, SE = .03, z = 2.36, $p = .018$, whereas there was no significant difference in improvement for the last five training sessions, $b = -.03$, SE = .02, z = -1.34, $p > .05$. This may be because the category discrimination ability of the less experienced group was lower than the more experienced group at pre test, although the difference was marginally significant, $\chi^2(1) = 3.41$, $p = .065$. There was no significant difference in the ability at the post test, $\chi^2(1) = .29$, $p > .05$.

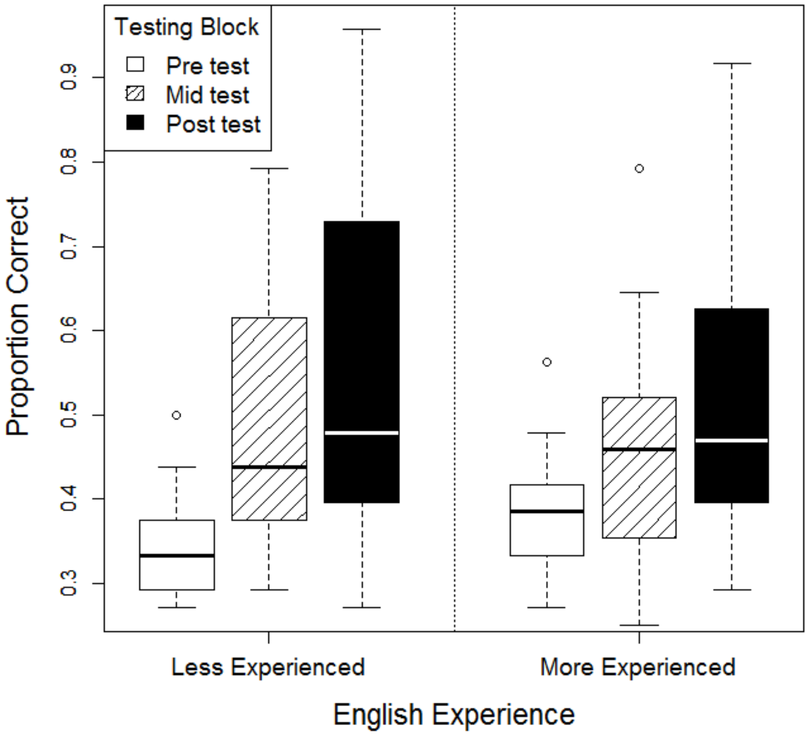


Figure 2.8. Boxplots of category discrimination accuracy at pre (white boxes), mid (shaded boxes) and post tests (black boxes) by adult Japanese speakers who have stayed in English-speaking countries less than 2 months (left) and those who have stayed in English-speaking countries for 2 months or more (right).

Figure 2.9 displays category discrimination accuracy of Japanese speakers in the two different trainer orders at pre, mid and post tests by English experience group. Japanese speakers in both ID-DIS and DIS-ID trainer orders improved the category discrimination accuracy, and the improvement difference between the two trainer orders seemed not to be affected by English experience.

The logistic mixed effects model for the category discrimination analysis demonstrated that there was no significant main effect of the 3-way interaction of testing block, trainer order and English experience, although it was marginally significant, $\chi^2(2) = 5.46$, $p = .065$. This suggests that the difference in the improvement between the two trainer orders was not significantly affected by the English experience.

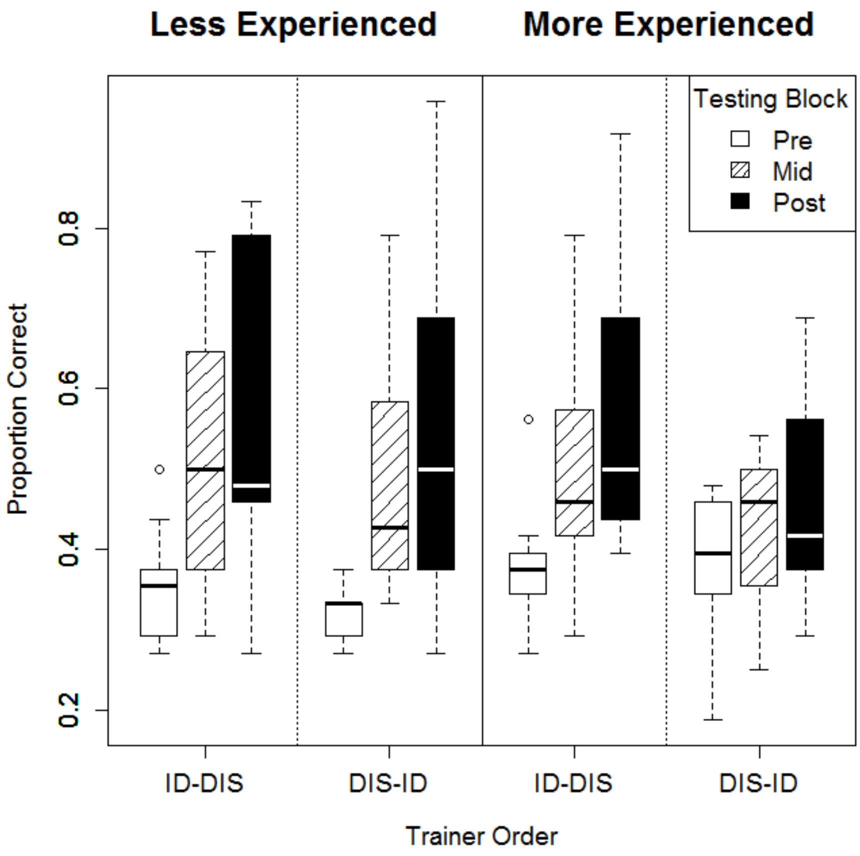


Figure 2.9. Boxplots of category discrimination accuracy at pre, mid and post tests by adult Japanese speakers in two different trainer orders (ID-DIS, DIS-ID) and in two English experience groups (Less experienced, More experienced).

2.3.4. Production of English /r/ and /l/

Figure 2.10 displays the F3 frequencies of English /r/ and /l/ produced by Japanese speakers at pre, mid and post tests. The F3 values of all /r/-/l/ tokens were normalised to the median F3 in the passage. Native English speakers distinguish

English /r/ and /l/ with F3 frequencies by producing a lower F3 for English /r/ and a higher F3 for English /l/ (Hattori & Iverson, 2009; Iverson et al., 2003; Kent & Read, 2002). As shown in the figure, Japanese speakers lowered F3 for English /r/ and raised F3 for English /l/ after perceptual training. This suggests that they learned the appropriate F3 distinctions of English /r/ and /l/ in production.

For production analyses, linear mixed effects models were used due to the continuous data of the normalised F3 measurements. As listed in Table 2.2, the best-fitting model for English /r/ and /l/ production included fixed factors of consonant (English /r/, English /l/), testing block, testing material (word, passage), trainer order, English experience, and their interactions. The interactions which are not listed on the table were excluded during the process of model comparison, suggesting that there was no significant effect of the excluded interaction factors. The random factors were subject and word. The random factor of word was nested into subject, accounting for the subjects' articulatory differences. For statistical significance of the effects of each fixed factor, *p* values are reported in the same way as the analyses for the perception data. However, due to the fact that *t* or *F*-distributions do not apply for unbalanced data, *p*-values may be anti-conservative (Baayen, Davidson, & Bates, 2008). Because of this problem, to report results of planned contrasts, Markov chain Monte Carlo (MCMC) simulations were used with the pvals.fnc function of the package *languageR* (version 1.4) within R software (Baayen et al., 2008; Baayen, 2008; Cunnings, 2012). Therefore, *p*MCMC values instead of *p* values are reported with *b*-values (estimates) and standard errors (SE) for planned contrasts and post-hoc analyses as in Whitford and Titone (2012).

Table 2.2 shows the results of the linear mixed effects model's type-III Wald chi square test which examined the effects of each fixed factor. As shown in the table, there were significant main effects of consonant and testing block. The interaction between consonant and testing block was also significant, suggesting that the change of F3 was significantly different between English /r/ and /l/. The planned orthogonal contrasts demonstrated that the difference in the F3 change between English /r/ and /l/ was significant for the first five sessions, b = 40.08, SE = 6.45, pMCMC < .001, and for the last five sessions, b = -12.64, SE = 3.71, pMCMC < .01. English /r/ was lowered by 125 Hz on average from pre to post test, while English /l/ was raised by 33 Hz.

The 3-way interaction of consonant, testing block and trainer order was excluded from the model during the model comparison. This suggests that there was no significant difference in the F3 change for the English /r/-/l/ contrast between the two trainer orders. In other words, both identification and discrimination training affected Japanese speakers by improving the English /r/-/l/ distinction by changing F3 frequency.

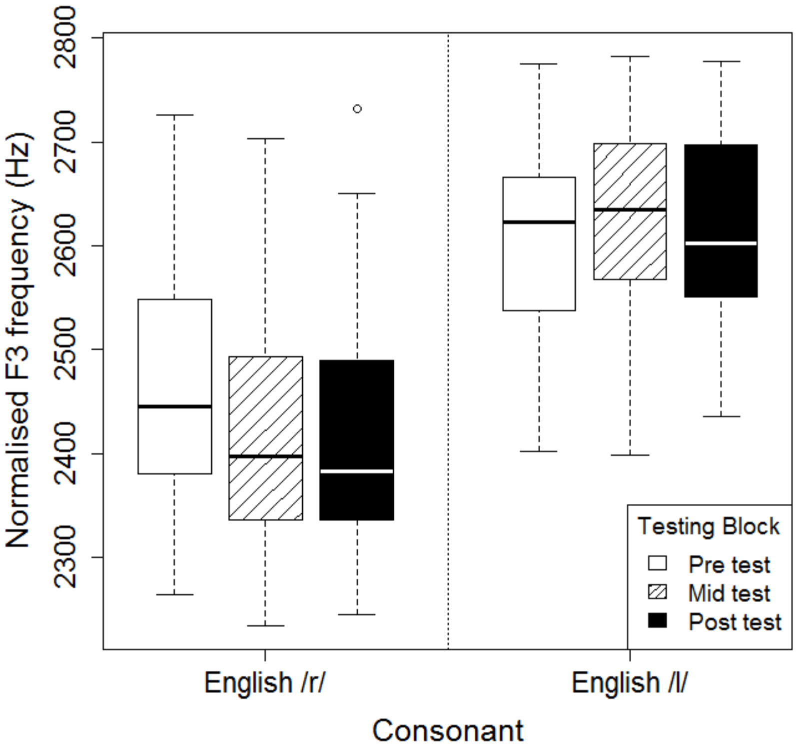


Figure 2.10. Boxplots of normalised F3 frequencies of English /r/ and /l/ produced by adult Japanese speakers at pre, mid and post tests. The F3 values of all /r/-/l/ tokens were normalised to the median F3 in the passage, and the normalised F3 values were rescaled with the formula, $F_3 = 2000 + 1200\,(F^N{}_3 - F^N{}_{3MIN}) / (F^N{}_{3MAX} - F^N{}_{3MIN})$. $F^N{}_{3MIN}$ and $F^N{}_{3MAX}$ are the minimum and maximum normalised F3 values (Thomas & Kendall, 2007).

Fixed Factors	χ^2	**Df**	***p***
Consonant	510.60	1	<. 001***
Testing block	10.46	2	<. 01**
Testing material	2.18	1	> .05
Trainer order	0.66	1	> .05
English experience	3.23	1	= .072
Consonant & Testing block	50.10	2	<. 001***
Consonant & Testing material	88.27	1	<. 001***
Testing block & Testing material	6.38	2	= .041*
Consonant & Trainer order	9.21	1	<. 01**
Testing block & Trainer order	0.80	2	> .05
Testing material & Trainer order	0.04	1	> .05
Consonant & English experience	40.17	1	<. 001***
Testing block & English experience	2.27	2	> .05
Testing material & English experience	1.33	1	> .05
Trainer order & English experience	0.21	1	> .05
Testing block & Testing material & Trainer order	3.07	2	> .05
Testing block & Testing material & English experience	0.13	2	> .05
Testing block & Trainer order & English experience	1.38	2	> .05
Testing material & Trainer order & English experience	5.60	1	= .018*
Testing block & Testing material & Trainer order & English experience	6.66	2	= .036*

Table 2.2. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on the normalised F3 frequencies of English /r/ and /l/ produced by Japanese speakers.

2.3.4.1. Production of English /r/. To examine how F3 of English /r/ and /l/ changed after training, post-hoc analyses were conducted for each consonant. Figure 2.11 displays the F3 frequencies of English /r/ production for two different testing materials (words and a passage) at pre, mid and post tests. As shown in the figure, Japanese speakers lowered F3 in the two testing materials through 10 training sessions. This suggests that they learned the rhoticity of English /r/. Although the F3 for word-reading appeared to be lower than that of passage-reading, there seemed to be no difference in the amount of the F3 change between the two testing materials.

The linear mixed effects models for English /r/ production demonstrated that there was a significant main effect of testing block, $\chi^2(2) = 64.72$, $p < .001$. The orthogonal planned contrasts for testing block demonstrated that F3 values were significantly lowered after training, from pre to mid tests, $b = -53.47$, SE = 8.37, *pMCMC* <. 001, and from pre & mid tests to post test, $b = 23.71$, SE = 4.82, *pMCMC* < .001. That is, Japanese speakers improved in producing the rhoticity of English /r/ over the course of 10 training sessions. There was also a significant main effect of testing material, $\chi^2(1) = 97.72$, $p < .001$, suggesting that F3 frequencies for word-reading ($M = 2400$ Hz) was significantly lower than those for passage-reading ($M = 2502$ Hz). Japanese speakers produced more rhotic /r/s in reading words in isolation than in reading a passage. The interaction between testing block and testing material did not fit into the best model, suggesting that there was no significant difference in improvement between the word- and passage-reading tasks.

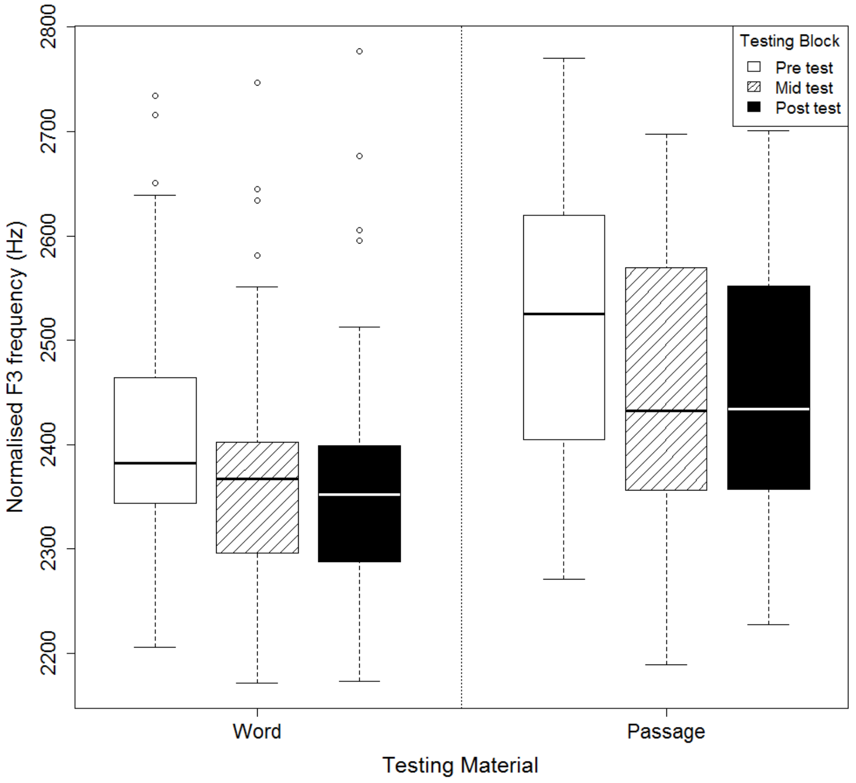


Figure 2.11. Boxplots of normalised F3 frequencies of English /r/ production by adult Japanese speakers in reading words (left) and reading a passage (right) at pre, mid and post tests.

Figure 2.12 displays the normalised F3 frequencies of English /r/ produced by Japanese speakers in the two different trainer orders, ID-DIS and DIS-ID. There seemed to be no difference in the F3 values for English /r/ between the trainer orders, and the degree of F3 change was also similar between the two trainer orders for the first five sessions and the last five sessions. This suggests that both identification and discrimination training lowered F3 to similar extents. The linear mixed effects model for English /r/ demonstrated that there was no significant main effect of trainer order, $\chi^2(1) = 1.81$, $p > .05$, and the interaction between testing block and trainer order was excluded from the best-fitting model during the process of model comparisons. That is, there was no significant difference in learning the rhoticity of English /r/ between the two training methods or between the two trainer orders.

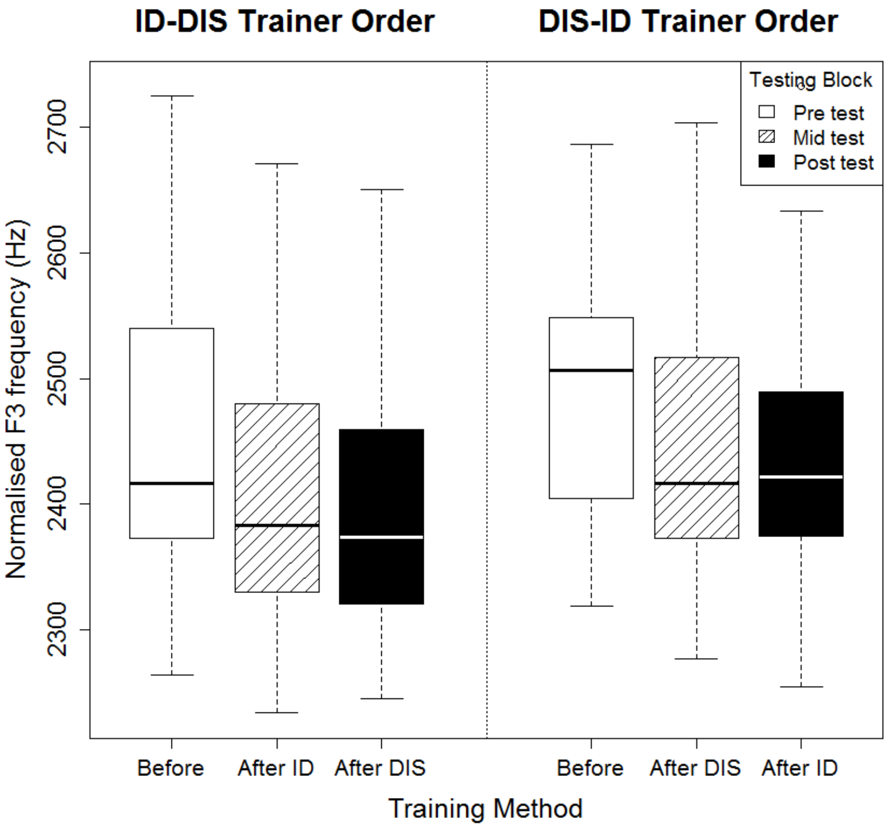


Figure 2.12. Boxplots of normalised F3 frequencies of English /r/ produced by Japanese speakers in two different trainer orders, ID-DIS (left) and DIS-ID (right), at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

Figure 2.13 displays the normalised F3 frequencies for English /r/ production at pre, mid and post tests by Japanese speakers with less and more English experience. Although the more experienced group seemed to produce lower F3 values than did the less experienced group, both English experience groups lowered F3 for English /r/ production after perceptual training. There seemed to be no difference in improvement between the two English experience groups.

The linear mixed effect model demonstrated that there was a significant main effect of English experience, $\chi^2(1) = 8.12$, $p < .01$, suggesting that Japanese speakers with more English experience pronounced English /r/ with more rhoticity. The interaction between testing block and English experience was excluded from the best-fitting model during model comparisons, suggesting that Japanese speakers in both English experience groups significantly lowered their F3 after perceptual training and the degree of improvement was not affected by English experience.

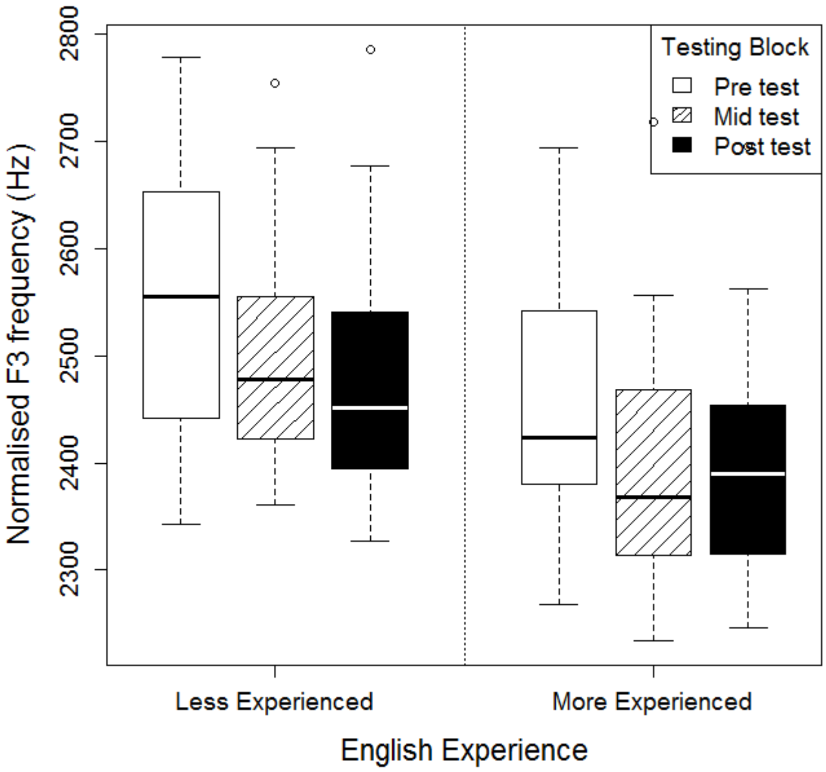


Figure 2.13. Boxplots of normalised F3 frequencies of English /r/ produced by adult Japanese speakers who had stayed in English-speaking countries for less than 2 months (Less Experienced) and those who had stayed in English-speaking countries for 2 months or more (More Experienced), at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

2.3.4.2. Production of English /l/. Figure 2.14 displays the normalised F3 frequencies of English /l/ production for each testing material at pre, mid and post tests. Unlike the English /r/ production, the F3 frequencies seemed not to change much, and the F3 values were slightly higher for the word-reading task compared to those for the passage-reading task. Regarding the interaction between testing block and testing material, there appeared to be a difference in the F3 change between the two testing materials.

However, a mixed effects model on F3 frequencies for English /l/ production showed a significant change in F3. The best-fitting model included fixed factors of testing blocks, trainer order, testing material, English experience and all possible interactions. The random factors were intercepts for subject and produced word nested into subject. The linear mixed effects model demonstrated that there was a significant main effect of testing block, $\chi^2(2) = 8.33$, $p = .016$. The planned orthogonal contrasts demonstrated that the F3 values were significantly raised from pre to mid tests, $b = 28.42$, SE = 10.03, *p*MCMC = .016, but there was no significant change for the last five training sessions from pre & mid tests to post test, $b = -3.25$, SE = 5.75, *p*MCMC > .05. This suggests that the improvement of F3 values was achieved in the first five training sessions. There was also a significant main effect of testing material, $\chi^2(1) = 28.64$, $p < .001$, suggesting that F3 values of English /l/ were significantly higher for the word-reading task than for the passage-reading task. This may be because Japanese speakers pronounced the English /l/ more carefully in isolated words than in connected speech. The interaction between testing block and testing material was also significant, $\chi^2(2) = 7.34$, $p = .025$, suggesting that Japanese speakers raised F3 for the word-reading task more than for the passage-reading task. Post-hoc analyses demonstrated that the F3 frequencies for word-reading were significantly raised from pre to post test, *b* = 46.10, SE = 12.19, *p*MCMC < .01, but not for passage-reading, *b* = -7.71, SE = 14.91, *p*MCMC > .05.

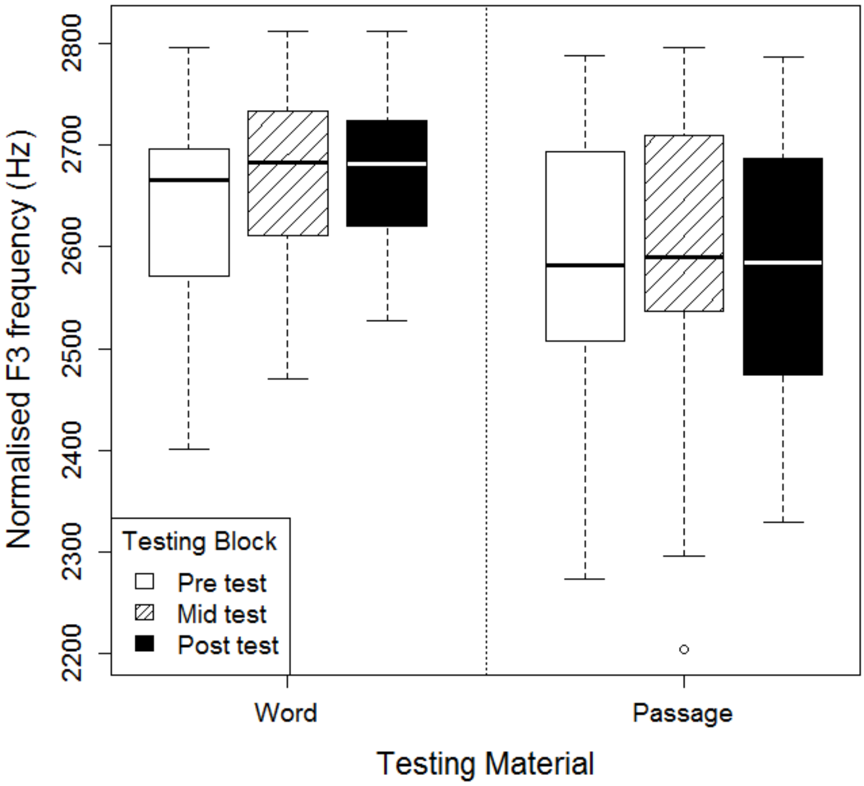


Figure 2.14. Boxplots of normalised F3 frequencies of English /l/ produced in reading words (left) and reading a passage (right) by adult Japanese speakers at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

Figure 2.15 displays the normalised F3 frequencies of English /l/ production at pre, mid and post tests for each trainer order. There seemed to be no difference in the F3 change between the ID-DIS and DIS-ID trainer orders. The linear mixed effects model also demonstrated that there was no significant main effect of trainer order, $\chi^2(1) = .11$, $p > .05$, and no significant interaction between testing block and trainer order, $\chi^2(2) = 2.05$, $p > .05$. These results suggest that the two different training methods affected the change in F3 to similar extents, and the trainer order did not affect the degree of the improvement.

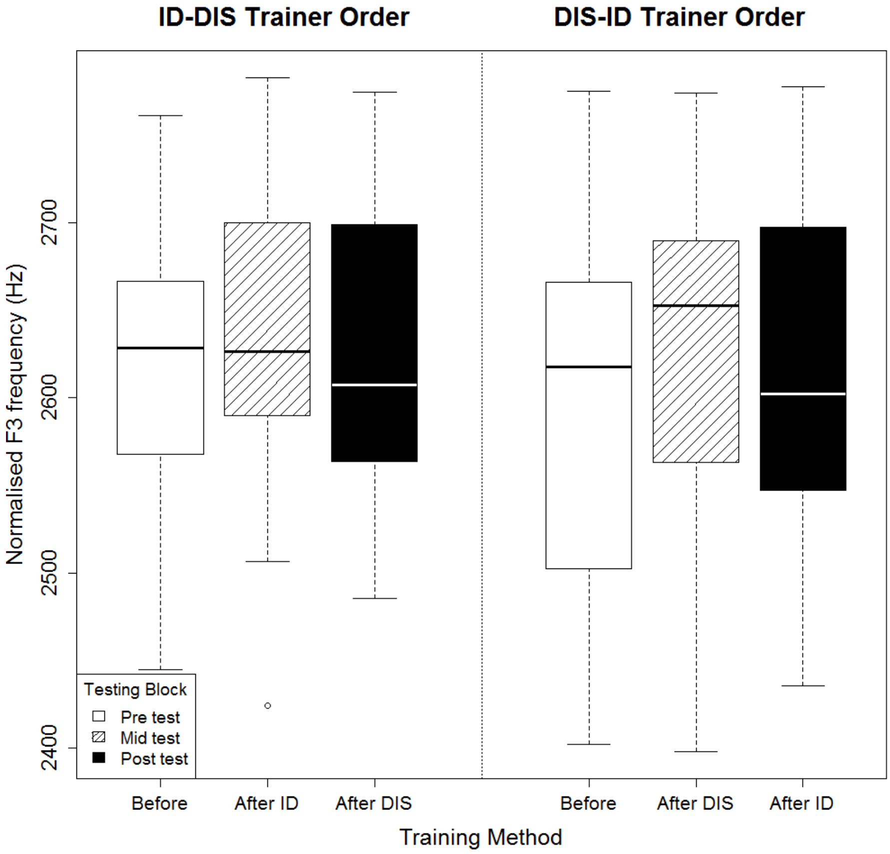


Figure 2.15. Boxplots of normalised F3 frequencies of English /l/ produced by Japanese speakers in two different trainer orders, ID-DIS (left) and DIS-ID (right), at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

Figure 2.16 displays the normalised F3 frequencies of English /l/ produced by Japanese speakers who had stayed in English-speaking countries for less than 2 months (less experienced) and those who had stayed for 2 months or more (more experienced) at pre, mid and post tests. As shown in the figure, there appeared to be a difference in the F3 change from pre to mid test between the two English experience groups. However, the logistic mixed effect model demonstrated that there was no significant main effect of English experience, $\chi^2(1) = .25$, $p > .05$, and no significant interaction between testing block and English experience, $\chi^2(2) = 2.07$, $p > .05$. This suggests that English experience did not affect the F3 change of English /l/ production.

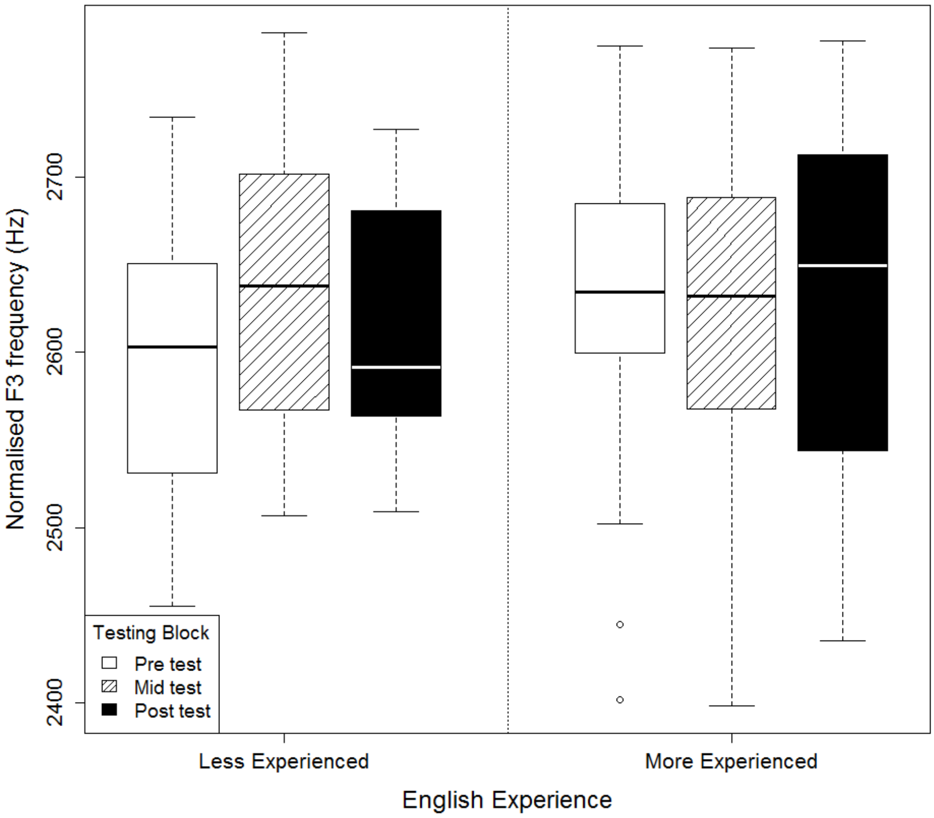


Figure 2.16. Boxplots of normalised F3 frequencies of English /r/ produced by Japanese speakers who had stayed in English-speaking countries for less than 2 months (Less Experienced) and those who had stayed in English-speaking countries for 2 months or more (More Experienced), at pre (white boxes), mid (shaded boxes) and post tests (black boxes).

In addition, although the 4-way interaction of testing block, testing material, trainer order and English experience was significant, $\chi^2(2) = 6.64$, $p = .036$, the planned orthogonal contrasts with MCMC sampling did not show any significant effect of English experience on the improvement difference in the interaction between testing material and trainer order for the first five training sessions (pre vs. mid), $b = 14.94$, SE = 10.03, *p*MCMC > .05, or for the last five training sessions (pre & mid vs. post), $b = 12.07$, SE = 5.75, *p*MCMC = .086.

2.3.5. Retention of Training Effects

2.3.5.1. Identification. Figure 2.17 displays the identification accuracy of Japanese speakers who participated in both the post-training test and the retention test that was conducted 4-6 weeks after the post-training test in London. As in the figure, Japanese learners did not demonstrate a decline in their identification accuracy, even though they did not receive any training during the 4-6 weeks. Although they did not additionally improve their identification accuracy through their daily conversations in London, they at least retained their improved identification ability.

Due to only half of the Japanese subjects participating in the retention test, a new logistic mixed effects model was used. The best-fitting model included the fixed factors of testing block (post test vs. retention test), trainer order, /r/-/l/ position, and English experience. The model also included six 2-way interactions between testing block and trainer order, between testing block and /r/-/l/ position, between trainer order and /r/-/l/ position, testing block and English experience, between trainer order and English experience, between /r/-/l/ position and English experience, and two 3-way interactions of testing block, trainer order and English experience, and of testing block, /r/-/l/ position and English experience. The English experience was categorised as less experienced and more experienced, and the breakpoint between the two experience groups was the median of the length of staying in English-speaking countries (i.e., 13.5 months). The random factors were crossed intercepts for subject and stimulus word, and the intercept of stimulus word was nested into by-speaker intercept.

Table 2.3 displays the results of the logistic mixed effects model's type-III Wald chi square test which examined each fixed factor's effect for identification accuracy at post and retention tests. As shown in the table, there was no significant main effect of testing block or trainer order. The interaction between testing block and trainer order was also not significant. These results suggest that that there was no significant decline or improvement of the identification ability 4-6 weeks after the post test, and the trainer order did not affect the decline or the improvement. Furthermore, none of the interaction factors with testing block was significant, suggesting that there was no difference in the improvement of the identification accuracy from post to retention tests according to /r/-/l/ position or English experience.

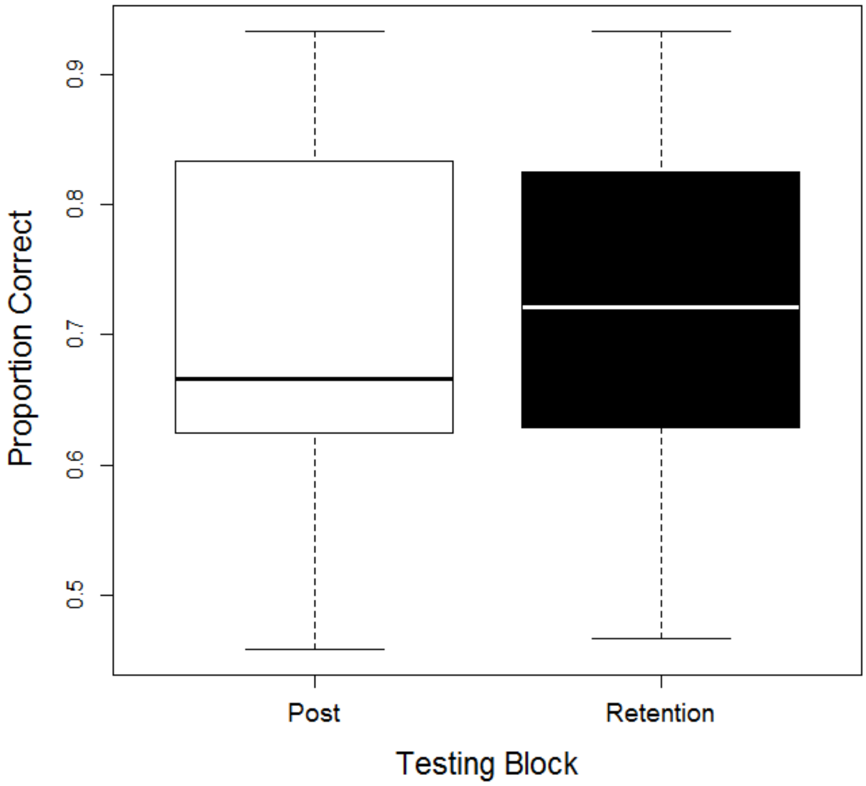


Figure 2.17. Boxplots of identification accuracy of the English /r/-/l/ contrast at post (white boxes) and retention tests (black boxes) by Japanese speakers in London.

Fixed Factors	χ^2	**Df**	***p***
Testing block	2.10	1	> .05
Trainer order	.22	1	> .05
/r/-/l/ position	22.29	2	< .001***
English experience	3.21	1	= .073
Testing block & Trainer order	.72	1	> .05
Testing block & /r/-/l/ position	1.60	2	> .05
Trainer order & /r/-/l/ position	7.47	2	= .024*
Testing block & English experience	2.05	1	> .05
Trainer order & English experience	.34	1	> .05
/r/-/l/ position & English experience	2.45	2	> .05
Testing block & Trainer order & /r/-/l/ position	3.49	1	= .062
Testing block & /r/-/l/ position & English experience	4.91	2	= .086

Table 2.3. Type-III analysis-of-variance table based on Wald chi-square tests for a logistic mixed effects model on the identification accuracy by Japanese speakers at post and retention tests in London.

2.3.5.2. Production of English /r/. Figure 2.18 displays the normalised F3 of English /r/ produced by Japanese speakers at post and retention tests. Japanese speakers seemed not to change the F3 frequency of English /r/ from post to retention test, suggesting that they may have kept the learned rhoticity of English /r/ for 4-6 weeks after the post test.

The best-fitting linear mixed effects model on the F3 of the production of English /r/ for retention effects included eight fixed factors of testing block, testing material, trainer order, English experience, 2-way interactions between testing material and trainer order, between testing material and English experience, and between trainer order and English experience, and a 3-way interaction of testing material, trainer order and English experience. The random factors were subject and word, and the random factor of word was nested into subject.

Table 2.4 displays the results of Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on F3 frequency of English /r/ production. As shown in the table, the linear mixed effects model demonstrated that there was no significant main effect of testing block. This result suggests that the lowered F3 for English /r/ had been retained for 4-6 weeks after the post test, but there was no additive learning via their daily conversations in London. Moreover, no fixed factor of the interaction with testing block was included in the best-fitting model. This suggests that the testing materials, trainer order or English experience did not affect the F3 change from post to retention test.

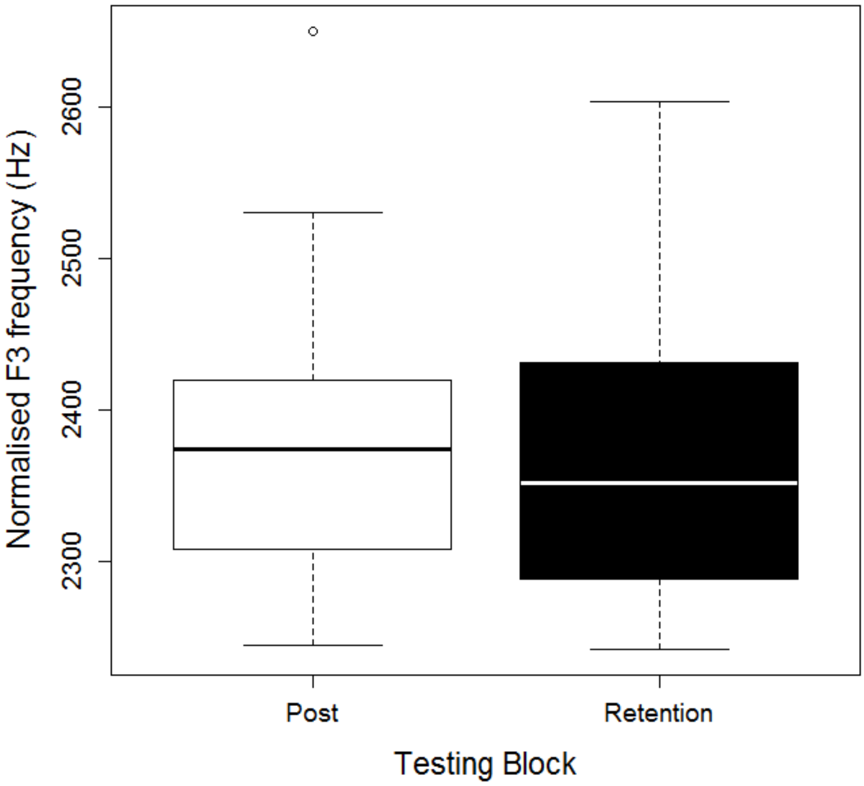


Figure 2.18. Boxplots of normalised F3 of English /r/ produced by adult Japanese speakers at post (white boxes) and retention tests (black boxes) in London.

Fixed Factors	χ^2	**Df**	***p***
Testing block	.90	1	> .05
Testing material	36.92	1	< .001***
Trainer order	.47	1	> .05
English experience	.10	1	> .05
Testing material & Trainer order	.63	1	> .05
Testing material & English experience	.62	1	> .05
Trainer order & English experience	3.61	1	= .058
Testing material &Trainer order & English experience	5.60	1	= .018*

Table 2.4. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on F3 frequency of English /r/ produced by adult Japanese speakers at post and retention tests in London.

2.3.5.3. Production of English /l/. Figure 2.19 displays the normalised F3 produced by Japanese speakers at post test and the retention test which was performed 4-6 weeks after the post test. As shown in Figure 2.19, the F3 frequency seemed not to be changed through daily conversation in London. The best-fitting linear mixed effects model on F3 of English /l/ for the retention effects included the fixed factors of testing block, testing material, trainer order, English experience and all possible interactions between these factors. The random factors were subject and word, and the random factor of word was nested into subject. The linear mixed effects model demonstrated that there was no significant main effect of testing block, χ^2 (1) = .55, $p > .05$, suggesting that the F3 values had not changed after 4-6 weeks without training.

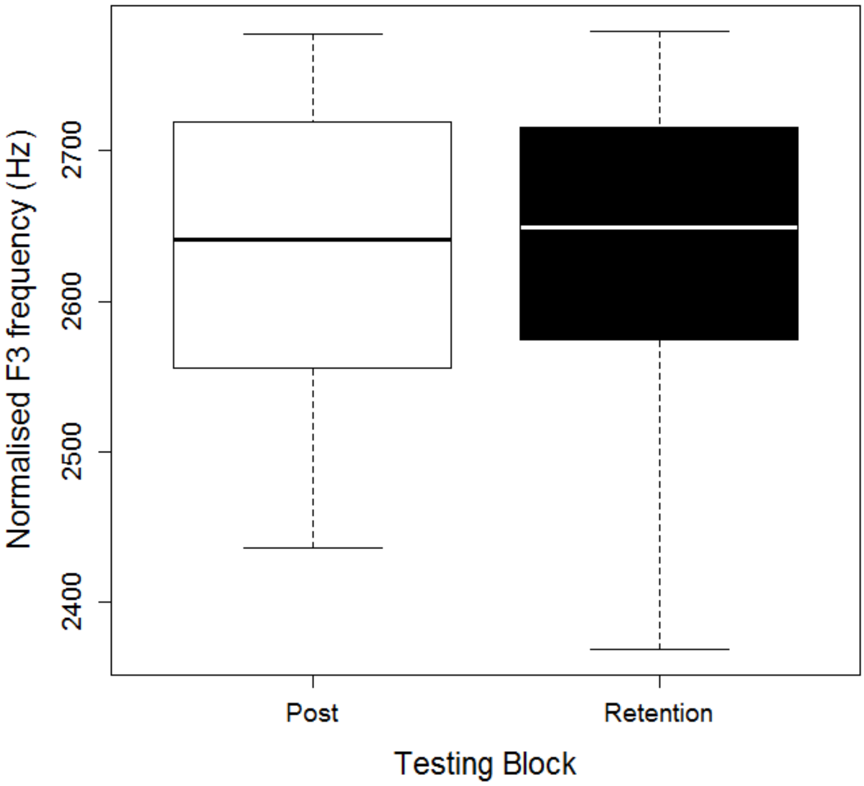


Figure 2.19. Boxplots of normalised F3 of English /l/ produced by adult Japanese speakers at post (white boxes) and retention tests (black boxes) in London.

Figure 2.20 displays the normalised F3 frequencies of English /l/ produced by the two English experience groups of Japanese speakers at post test and the retention test which was performed 4-6 weeks after the post test in London. It seems that the less experienced group slightly lowered F3 and the more experienced group slightly raised F3 frequencies. The linear mixed effect model demonstrated that the F3 change from post to retention test was significantly different by English experience, χ^2 (1) = 5.05, *p* = .025. However, the post-hoc analyses with MCMC sampling demonstrated that neither the less experienced group, *b* = -45.42, SE = 20.91, *p*MCMC = .067, nor the more experienced group, *b* = 23.39, SE = 21.53, *p*MCMC

> .05, significantly changed F3 between post and retention tests, although the F3 change by the less experienced group was marginally significant. These results suggest that both the less and the more experienced group retained the improved F3 of English /l/ production for 4-6 weeks after the post test without training.

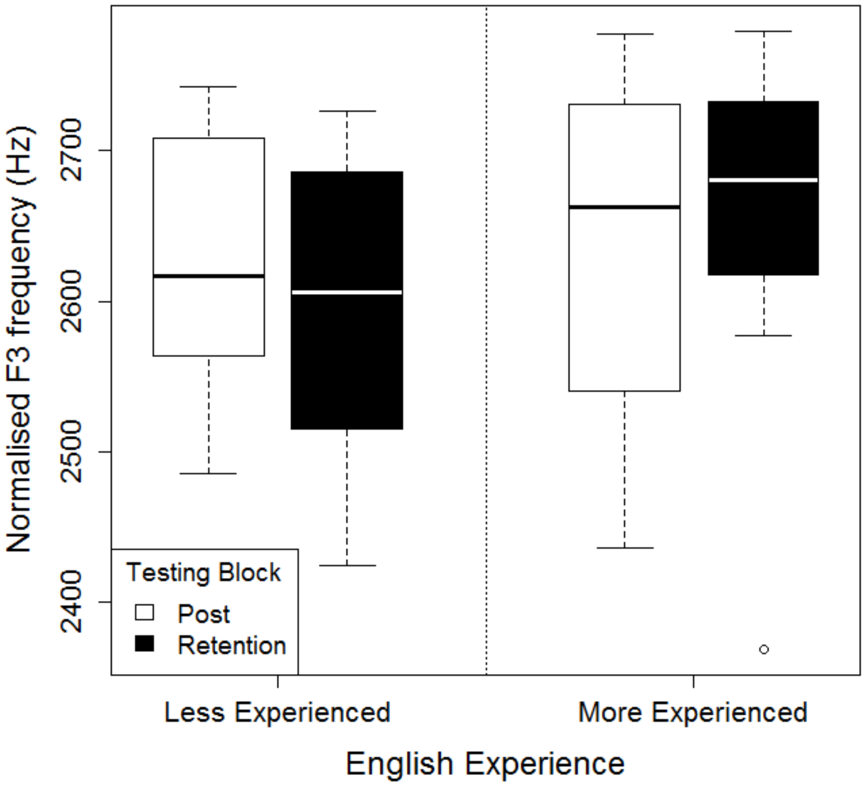


Figure 2.20. Boxplots of normalised F3 of English /l/ production at post (white boxes) and retention tests (black boxes) by adult Japanese speakers who had stayed in English-speaking countries less than 13.5 months (Less Experienced) and those who had stayed in English-speaking countries for 13.5 months or more (More Experienced).

Table 2.5 displays the results of Type-III analysis-of-variance based on Wald chi-square tests for a linear mixed effects model on F3 frequency of English /l/. As shown in the table, all interactions including testing block apart from testing block &

English experience were not significant, $p > .05$. These results suggest that there was no change in the F3 values of English /l/ between post and retention tests, and no factor affected the F3 change.

Fixed Factors	χ^2	**Df**	***p***
Testing block	0.55	1	> .05
Testing material	9.00	1	< .01**
Trainer order	.46	1	> .05
English experience	.20	1	> .05
Testing block & Testing material	2.06	1	> .05
Testing block & Trainer order	.71	1	> .05
Testing material & Trainer order	.14	1	> .05
Testing block & English experience	5.05	1	= .025*
Testing material & English experience	.17	1	> .05
Trainer order & English experience	1.29	1	> .05
Testing block & Testing material & Trainer order	1.96	1	> .05
Testing block & Testing material & English experience	.81	1	> .05
Testing block & Trainer order & English experience	.83	1	> .05
Testing material &Trainer order & English experience	.07	1	> .05
Testing block & Testing material & Trainer order & English experience	3.11	1	= .078

Table 2.5. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on F3 frequency of English /l/ produced by adult Japanese speakers at post and retention tests in London.

2.4. Discussion

One conclusion of this study is that discrimination training improved identification accuracy of novel stimuli. This finding is different from the results of Strange and Dittmann (1984). Logan et al. (1991) argued that the discrimination training used in Strange and Dittmann (1984) did not improve the identification of novel natural speech, because listeners improved their perceptual sensitivity within categories of English /r/ and /l/ as well as between the categories, and the training stimuli did not have naturalistic variability. The current discrimination training method addressed these concerns by focusing on F3 frequency differences between English /r/-/l/ categories and including variability due to different talkers and phonetic environments. The results demonstrated that a discrimination training method can improve identification accuracy of novel English words spoken by untrained talkers, provided it follows the procedures described in this chapter. It is unclear if the improvement is caused by discrimination training without examining test-retest reliability. However, it may be plausible to consider that discrimination training with high stimulus variability improves identification accuracy, comparing to the results of a previous study (Strange & Dittmann, 1984).

Another new finding is that both identification and discrimination training methods seemed to affect the perceptual abilities for the English /r/-/l/ contrast in similar ways. The hypothesis was that the discrimination training method would improve lower phonetic processing, and the identification training would improve higher phonological processing. Consequently, the combination of identification and discrimination training methods would lead to a higher improvement rate in identification accuracy. However, the results demonstrated that these two training methods had similar effects on all tasks, although the identification training method was better at improving identification accuracy than was the discrimination training method. In other words, the combination of two different training techniques might improve the perception of English /r/ and /l/ to a similar degree as 10 sessions of one training method would have achieved.

If identification and discrimination training improved the same underlying processing, what processing levels were improved? One hypothesis is that both identification and discrimination training methods improved lower level sensitivity. Firstly, Japanese speakers improved their phonetic sensitivities, although they did not improve their sensitivity to the primary acoustic cue, F3 at the /r/-/l/ boundary, more than F2 sensitivity. Secondly, both identification training and discrimination training improved category discrimination accuracy. The category discrimination task is believed to require subjects to focus on appropriate acoustic cues, and the improvement in this task may suggest that Japanese speakers become able to pay attention to the primary acoustic cue in listening to the English /r/-/l/ contrasts. These results show that identification and discrimination training techniques affected the lower phonetic levels rather than higher phonological levels.

Another plausible interpretation of the results is that both training methods taught listeners how to cope with the variability of stimuli without necessarily making fundamental changes in underlying abilities. Iverson et al. (2005) suggested that identification training with high stimulus variability does not form a new phoneme category of English /r/ or /l/. Rather, it teaches Japanese listeners to use existing category representations, for example Japanese /r/, to identify English /r/ and /l/ systematically. This may be the case for the present study. Japanese speakers may have learned the strategy to use Japanese /r/ as a reference to identify English /r/ and /l/. For instance, if the stimulus sounds similar to Japanese /r/, it would be identified as English /l/. If the stimulus sounds sufficiently different from Japanese /r/, it would be identified as English /r/. It may be plausible to consider that neither identification nor discrimination training improved certain processing levels, but that a two or three-alternative forced choice task with the variability of stimuli required Japanese speakers to learn a systematic identification strategy.

However, if Japanese speakers just learned the systematic identification technique through training, it cannot be explained how English /r/-/l/ production was improved. F3 in English /r/ was significantly lowered after training for both word- and passage-reading tasks, whereas F3 in English /l/ was significantly raised only for the word-reading task. This may be because of the acquisition of perceptual-motor mappings. Callan et al. (2003) examined localised changes in Japanese speakers' brain activity through one month of extensive perceptual training of the English /r/-/l/ contrast. The results demonstrated that Japanese speakers showed that their acquired perceptual-motor mappings reflected the enhancement of activity in the left supramarginal gyrus and Broca's area after perceptual training. Callan et al. (2003) also argued that the improvement in the identification of the English /r/-/l/ contrast may be partially attributed to the acquisition of auditory-articulatory (perceptual-motor) mappings. Thus, perceptual training may affect the speech motor areas which can be reflected in speech production acquisition, although Japanese speakers may not completely change their phonological or phonetic processing for the English /r/-/l/ contrast.

The improvement in production is in line with the prediction of the Speech Learning Model (Flege, 1995). That is, it is easier to improve the production of English /r/ than English /l/, because English /r/ is not as close to Japanese /r/ as is English /l/. The results of the present study demonstrated that the F3 of English /r/ was lowered for both the word- and passage-reading tasks, but the F3 of English /l/

was raised only for the word-reading task. This may be because English /r/ is easier to incorporate into the unoccupied region of Japanese speakers' phonological space, compared to English /l/ (see SLM: Flege, 1995). Moreover, the reason that Japanese speakers did not improve the F3 of their production of /l/ in passage-reading might be also explained by the best exemplars of English /r/ and /l/ which Japanese speakers already had in their mind. According to Hattori and Iverson (2009), Japanese speakers have more native-like cognitive representation of English /r/ in terms of F3 than that of English /l/. The better exemplar of English /r/ might have helped Japanese speakers improve their production of English /r/ more than English /l/. Another possible explanation for the less change in F3 for /l/ production is that there may not be enough space for the improvement. As Hattori and Iverson (2009) demonstrated, F3 values for English /l/ produced by Japanese speakers may have been similar to those produced by native English speakers, whereas F3 values for English /r/ may have been considerably higher than native speakers' productions. Thus, the realisations of English /l/ which Japanese speakers originally produced would not necessarily have been changed much in terms of F3.

There are two more findings in this study. One is that improved perception and production abilities were retained for at least 4-6 weeks without training. It was hypothesised that training would trigger the learning of the English /r/-/l/ contrast in daily life of L1 Japanese speakers in English-speaking countries. However, both perception and production were not significantly improved through daily conversation in English after training, but they were kept at the same levels, which is consistent with previous studies (Bradlow et al., 1999; Lively et al., 1994). The other finding is that English experience is the main factor affecting learning L2 phonemes. In the identification and category discrimination tasks, English experience negatively affected improvement. In other words, Japanese speakers with less English experience improved more than Japanese speakers with more English experience, although the factor did not affect the improvement in production. This result is not consistent with the study by Iverson et al. (2012). Iverson et al. (2012) suggested that auditory training improved perception of English vowels for both the experienced and inexperienced French speakers to similar degrees. Given that the length of living in English-speaking countries was highly correlated to the subjects' age in the present study, it cannot be concluded that the difference in improvement was caused by the English experience only. The effect of English experience in the present study may be attributable to an age effect. More pre-learned knowledge and/or more aging may constrain Japanese speakers from learning L2 phonemes in perception.

In conclusion, the discrimination training technique improved the identification accuracy of novel natural stimuli. However, the effects of the discrimination training method were similar to those of identification training at both lower phonetic levels and higher phonological levels. As a result, the current combined training methods did not provide higher improvement rates than previous studies using identification training only. Although Japanese speakers may have acquired auditory-articulatory mappings (Callan et al., 2003), both identification and discrimination training methods seemed to teach Japanese speakers how to cope with high variability (Iverson et al., 2005). Because Japanese speakers did not alter their phonetic sensitivities to match native English speakers having higher sensitivity to

F3 at the English /r/-/l/ boundary and lower sensitivity to irrelevant acoustic cues, the improvement rate of identification accuracy may be still limited to approximately 15%. A new training method that lets Japanese speakers correct their L1-interfered phonetic processing would improve the perception and production of the English /r/-/l/ contrast more than did the present training techniques. It would be also interesting to train younger learners' perception of English /r/ and /l/, because they may have greater brain plasticity and less L1 interference. The combination of identification and discrimination training would have better outcomes for younger learners, and it may be easier for them to change their phonetic processing in an L2, when compared to adults. This potential age effect was examined in the second study in Chapter 3.

Chapter 3: Age effects on learning English /r/-/l/ perception for Japanese speakers

3.1. Introduction

It is well known that language learning is affected by age. The idea of *the younger the better* in learning a language is originally derived from the critical period theory (Lenneberg, 1967). The critical period theory posits that language learning strongly relates to biology. Since the gradual specialisation of the left hemisphere for language seems to be led by biological maturation and experiential factors, language learning is unlikely to occur via specialised neural systems but only through general learning mechanism after puberty (Werker & Tees, 2005). One previous study supporting this argument demonstrated that aphasia patients who sustain damage to the language area in the left hemisphere before puberty can recover their language ability later (Lenneberg, 1967). However, patients who damage the same area after puberty often cannot overcome some of their problems, and commonly the damage leaves some trace behind (Lenneberg, 1967). As well as L1 acquisition, the critical period theory also seems applicable to L2 acquisition. Johnson and Newport (1989) demonstrated that L2 learners who arrived in L2-speaking countries earlier were better at learning L2 morphosyntax than those of later arrivals. However, after the age of 15 years, the age of arrival in L2-speaking countries did not affect the learning of L2 morphosyntax. Following this result, Johnson and Newport (1989) concluded that there is a critical period for L2 learning, and that the windows of learning L2 may close at around 15 years old. Nevertheless, this theory cannot explain why some L2 learners are able to perceive and produce L2 speech as native speakers do, even if they start learning their L2 after puberty, and also why some L2 learners who start learning an L2 before puberty sometimes retain a strong L1 accent in their L2. It seems that not only biological maturation affects language learning, and that onset, length and offset of the critical period(s) for language learning can be changed by experience (Werker & Tees, 2005).

Since the onset and the offset of the critical period for L2 learning is altered by experience, use of the term “sensitive period” (Oyama, 1976) or “optimal period” (Werker & Tees, 2005) has been encouraged as an alternative. Sensitive/optimal periods never cut off definitively at puberty, especially for speech learning, and the ability to learn L2 perception and production is valid at least up to adulthood (Werker & Tees, 2005). Much research has been conducted into identifying factors that affect L2 learning. Studies for L2 learning by immigrants in L2-speaking countries have demonstrated that age of arrival in those countries is one of the influential factors for learning L2 perception and production (Flege, Yeni-Komshian, & Liu, 1999). However, the age of arrival is inter-correlated with other factors such as cognitive development, quantity and quality of L2 input, amount of L1 use, and the development of L1 categories (Flege & MacKay, 2010; Flege et al., 1999). In fact, all of these factors have effects on L2 learning (Flege & MacKay, 2010; Flege et al., 1999; Flege, 2002; Ingvalson et al., 2011). Since these factors are correlated and confounded with age, studies in immigrants’ L2 learning have a limitation in isolating the effects of the confounding factors such as L1 development (SLM: Flege, 1995, 1999, 2002) and biological neural plasticity for learning language (Zhang & Wang, 2007).

Biological neural plasticity, denoted by an exclusive enhancement in brain activity, affects L2 phoneme learning. As the native language neural commitment (NLNC: Kuhl, 2000; Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005) yields a more automatic and linguistic level of processing in speech perception, the neural plasticity for non-native language is reduced. Native adult speakers who acquired their L1 automatically focus on a more linguistic level of processing with faster and shorter brain activation in particular regions while listening to speech, which prevents adult speakers from perceiving acoustic and phonetic details of speech (Zhang & Wang, 2007). This is similar to the proposition of the SLM (Flege, 1995, 1999, 2002), since the SLM suggests that the development of L1 interferes with L2 learning. However, training studies have demonstrated that phonetic learning can occur with increases, decreases and shifts in adults’ brain activation, suggesting the existence of substantial brain plasticity in adulthood (Callan et al., 2003; Golestani, Paus, & Zatorre, 2002; Golestani & Zatorre, 2004; Wang, Sereno, Jongman, & Hirsch, 2003; Zhang & Wang, 2007). The change of neural plasticity may be caused by not only prior L1 learning but also continuously updated cognitive and attentional strategies (Zhang & Wang, 2007). Thus, unlike SLM, neural plasticity may not cut off by the time of L1 acquisition, but gradually decline over the course of the life span.

In terms of learning perception and production of the English /r/-/l/ contrast by Japanese speakers, Aoyama et al. (2004) examined the effect of age on the improvement in Japanese speakers' perception and production. The results demonstrated that Japanese children who moved to the US improved their perception and production of English /r/ and /l/ more than did Japanese adults after one year of their arrival, supporting the idea that younger learners are better at learning L2 than are older learners. However, Aoyama et al. (2004) also mention that a limitation of their study is that they did not control for quality and quantity of L2 input between children and adults. Japanese adults learned English before they arrived in the US, whereas Japanese children learned after. Moreover, Japanese children might have had less L1 communication and more L2 input from native speakers of English than adults, because children communicated with native English friends and teachers at local schools. Some adults might not often have spoken English with native speakers. Therefore, the reason for the improvement in Japanese children's perception and production of the English /r/-/l/ contrast being greater than the improvement seen in adults may be due to social and educational factors.

In the present study, the age effects on learning English /r/-/l/ perception by Japanese speakers were examined, using a computer-based perceptual training programme to control the quality and quantity of L2 input. It can be hypothesized that younger learners will improve from pre-training to post-training tests more than older learners. Children may not have completed establishing their L1 phonetic categories by the age 12 years old (Hazan & Barrett, 2000), and early Japanese perception of English /r/ and /l/ may likewise be less affected by their L1 phonetic processing (Shimizu & Dantsuji, 1983). This suggests that younger learners might have some advantage in learning English /r/-/l/ due to their lower L1 interference, as well as their greater neural plasticity. Adolescents aged between 15 and 18 years old, on the other hand, may have completed constructing a system of their L1 phonetic units and may have the same phonetic processing as do adults. However, they may still have a little more advantage from their relatively greater neural plasticity compared to adults. In contrast, adults may have difficulty improving their perception of the English /r/-/l/ contrast due to their strong L1 interference and their relatively fossilised neural plasticity. Adults might be able to improve their ability to identify the contrast to some extent, but they may not be able to change their phonetic processing.

Despite this hypothesis, previous training studies have demonstrated that the amount of the improvement was comparable between adults and children. Heeren and Schouten (2008, 2010) examined child Dutch speakers learning the Finnish /t/-/tː/ length contrast via a phonetic perceptual training programme. The results demonstrated that adults and children shifted the location of a category boundary in a similar way and to a similar extent, although their perceptual sensitivities at a category boundary did not reach that of native Finnish speakers. Wang and Kuhl (2003) and Sereno and Jongman (2005) tested the improvement difference in the identification of four Chinese Mandarin tones between four age groups of American English speakers, using a 2-week programme of computerised Mandarin tone training. They found that all four age groups significantly improved their identification ability and that the degree of improvement was approximately the same between age groups. The authors explained the results of no noticeable difference in the amount of the improvement between the age groups as being due to children having a disadvantage in listening to L2 phonemes. The disadvantage was ascribed to the immaturity of cognition, such as attention, and the lack of auditory and language development (Heeren & Schouten, 2010; Wang & Kuhl, 2003). Thus, children do not have advantages in learning L2 phoneme perception through a computer-based perceptual training programme.

As well as the immaturity of cognition and the auditory sensory system, other factors may also affect child L2 speech learning. One is the difference of acoustic cue weighting between children and adults. According to the Developmental Weighting Shift (DWS) model (Nittrouer, Miller, Crowther, & Manhart, 2000; Nittrouer & Miller, 1997; Nittrouer, 1992), children focus on formant transitions when hearing speech. This may be an advantage for Japanese children learning English /r/ and /l/, because F3 transitions are the primary acoustic cues for the /r/-/l/ contrast. Children may avoid the use of temporal acoustic cues such as intensity and duration on which adult learners often focus (Hattori, 2009; Iverson et al., 2005). However, it is unclear if children always pay attention to formant transitions when listening to L2 phonemes. Going against the DWS model, Sussman (2001) demonstrated that children in general rely on the most acoustically salient cue due to their inability to use more informative cues. For Japanese adults, the most learnable salient acoustic cue for English /r/ and /l/ is not F3, but seems to be closure and transition duration (Hattori, 2009; Iverson et al., 2005). However, it remains unknown which acoustic cue is the most salient for Japanese children perceiving English /r/ and /l/. Mayo and Turk (2004) also demonstrated that children do not always rely on formant transitions, and that cue weighting depends on the segmental context. These findings, taken together, demonstrate that Japanese children may possibly have an advantage in learning the English /r/-/l/ contrast (based on the DWS model), but also that they may have a disadvantage in perceptual learning due to the immaturity of their cognition and sensory systems.

The present study trained Japanese adults, adolescents, older children and younger children with a computer-based perceptual training programme, with the aim of evaluating the effects of age on learning English /r/-/l/ perception. Japanese speakers were given 10 perceptual training sessions consisting of identification and discrimination tasks. In order to examine phonetic and phonological learning, the improvement in identification accuracy, auditory discrimination and category discrimination were compared between these age groups. In addition, the learning process was also investigated; the improvement in identification ability over the course of the 10 training sessions was compared between age groups. Although the baseline of English knowledge before pre test was different between age groups, it can be predicted that younger learners will improve quickly and boost their identification ability more than older learners will due to their greater neural plasticity and less L1 interference.

3.2. Method

3.2.1. Subjects

A total of 115 Japanese speakers were tested in Japan. Of these 115 subjects, 30 subjects were not used for the analysis; of these rejected subjects, 9 had an identification score at pre test of more than 75% correct, 5 dropped out during the training stages, 9 completed two or more training sessions in a day, 4 had experience of living in English speaking countries prior to the age of 6 years old, and 3 had a language/hearing impairment. Thus, 85 Japanese speakers completed all tasks, and their perceptual abilities were analysed for each test.

There were four age groups: younger children aged from 6;11 to 8;1 (years; months, N = 15, median = 7;9), older children aged from 8;1 to 12;4 (N = 13, median = 10;4), adolescents aged from 15;7 to 18;6 (N = 28, median = 16;1) and adults aged from 25;3 to 59;8 (N = 29, median = 29;2). The younger and older children groups were divided by the median of all children's age, although this arbitrary division of age groups may ultimately have affected the results of analyses. Apart from four adults, no subjects had lived in English speaking countries. Those four had stayed in English-speaking countries for 15, 48, 78, and 44 weeks. Adults and adolescents started learning English at schools at 12 years of age, whereas younger and older children had learned only some basic English at schools (e.g., greetings and colour names).

When subjects participated in this study, they were asked how they wanted to take part in this study and to choose from the following 3 options: (1) taking part in only pre and post tests without training (i.e., as a control group), (2) taking part in the pre test, 10 training sessions and post test (i.e., as a training group), (3) taking part in both (1) and (2) as a mixed group. Subjects who chose option (3) took the first test, then had a 10-14 day interval without training, the second test, 10 training sessions, and then the post test. Twelve subjects chose option (3), and they were counted as both control and training groups. That is, the first and the second tests were analysed as pre and post tests for the control part, and the second and the post test were analysed as pre and post test for the training part. Subsequently, 38 subjects (5 younger children, 7 older children, 12 adolescents, 14 adults) were analysed as control groups and 59 subjects (12 younger children, 13 older children, 16 adolescents, 18 adults) were analysed as training groups.

3.2.2. Apparatus

The pre and post tests were conducted in quiet rooms in several different cities in Kanto area in Japan, accompanied by the author. Stimuli were played over headphones (Sennheiser HD 205 and Sennheiser HD 201) with a user-controlled comfort level, and all responses were saved in files on computers protected by password. All training sessions were also supervised by the author, except those for adult subjects who used their own headphones/earphones and laptops on which the training software was installed. All trials and responses in training sessions were also saved in password-protected log files on the computers. Subjects were asked not to do more than one training session in a day. If they completed more than one training session in a day, those subjects' results were excluded from the analysis as detailed above.

3.2.3. Stimuli

3.2.3.1. Natural recordings. Natural recordings were used for both testing and training stimuli. Two hundred and twenty English /r/-/l/ minimal-pair words from nine adult Standard Southern British English (SSBE) speakers were digitally recorded with 44,100 16-bit samples and downsampled to 22,050 samples per second as in Iverson et al., (2005). One hundred word-initial /r/-/l/ minimal-pair words (e.g., *rock* and *lock*) spoken by five speakers were used as training stimuli (Appendix A), and the other 120 minimal-pair words by four speakers were used as testing stimuli. The 120 words were 40 /r/-/l/ word-initial (e.g., *road* and *load*), 40 /r/-/l/ word-medial (e.g., *correct* and *collect*), and 40 /r/-/l/ consonant cluster (containing /r/-/l/) minimal-pair words (e.g., *brand* and *bland*), as used in the first study in Chapter 2 (Appendix B).

3.2.3.2. Synthetic stimuli and signal processing. Synthetic stimuli were used for testing perceptual sensitivities in the auditory discrimination test, and signal-processed stimuli were used for the task of auditory discrimination training. The methods of generating those stimuli were the same as described in Chapter 2.

3.2.4. Procedure

3.2.4.1. Training. A total of 59 subjects received 10 training sessions, using a computer-based perceptual training programme which consisted of the recordings of five SSBE speakers. The programme was called “Samurai English Training”, and subjects were apprentices to become international Samurai. Each training session included three tasks: identification, auditory discrimination and category discrimination. Subjects were instructed how to play the game in Japanese before each task. During training, they were shown their percentage correct and the percentage completed on each task.

In the identification task (90 trials), subjects heard a stimulus and clicked on the minimal-pair word that they thought they heard (e.g., *rope* vs. *lope*). There were always the same two characters displayed on screen next to the /r/ and /l/ words, which were an R-man wearing a blue hat and an L-man wearing a yellow hat (i.e.,

). Because of these two characters, subjects did not have to read or understand words but needed to focus on the first phoneme to judge R or L. For only the first 10 trials in each session, they heard both model /r/-/l/ minimal-pair words before a trial to remind them of the phonemes. When the model /r/-/l/ stimuli were produced, the R-man and the L-man jumped slightly to let the subjects remember each of the /r/ and /l/ phonemes. For the remainder of the 80 trials, subjects did not hear the models but only the trial stimuli. When they clicked a correct answer, they saw a message せいかい (*Correct*) on the screen and heard a cash register sound. The correct stimulus was replayed once with the correct character (the R-man or the L-man) jumping. In the case that subjects clicked a wrong answer, they saw a message ざんねん (*Bad Luck*) and the wrong choice was crossed out on the screen. They also heard two beep sounds with different pitches, and the correct stimulus was replayed twice with the correct character (the R-man or the L-man) jumping. After every 20 trials, a message *Level-up* was shown on the screen to let the subjects have a break and to encourage them to carry on. After completing all 90 trials, they were rewarded with a sword and a magic wand for the next task of auditory discrimination.

In the auditory discrimination task (48 trials), subjects heard three signal-processed stimuli and chose the one that sounded different from the other two. The three signal-processed stimuli were English /r/-/l/ minimal-pair words which were acoustically all the same apart from their F3 values. To mark correct answers, Japanese subjects needed to focus on the F3 difference at the /r/-/l/ boundary. The distant pairs used in the first study were used in this present study. Three identical objects numbered 1, 2 and 3 on their fronts were displayed on the screen, and a moving black cursor highlighted the object and its number when the stimulus was being played. When subjects clicked the correct numbers, they could cut the object and number with their swords and magic wands. They also heard cash register sounds and saw a message せいかい (*Correct*) on screen. There was no repetition for the correct trial. In the case that they clicked a wrong answer, they were not able to cut the object or the number and heard two beep sounds with a message ざんねん (*Bad Luck*) displayed on the screen. All the three stimuli were replayed once with a highlighted correct answer. There were four kinds of objects such as husumas (Japanese traditional sliding screens), shoujis (Japanese traditional sliding paper doors), rocks and iron shields, and the objects displayed on screen were changed from softer (e.g., husuma) to stiffer ones (e.g., iron shields) after every 10 trials. Subjects were also rewarded with a new sword or a new magic wand after every 10 trials.

In the category discrimination task (60 trials), subjects heard three stimuli alongside three red devils being displayed on the screen, and clicked on the one devil that produced a categorically different stimulus from the other two. The stimuli in a trial were natural recordings of a word-initial /r/-/l/ minimal pair produced by different speakers, so subjects needed to ignore the speaker differences and the irrelevant acoustic cues for the /r/-/l/ categories. They needed to pay attention only to the F3 that made a stimulus categorically different. If they marked a correct answer, they received a treasure from the devil alongside cash register sounds and the message せいかい (*Correct*) displayed on the screen. In the case that they clicked a wrong answer, they saw the message ざんねん (*Bad Luck*) on screen, the devil that produced the categorically different stimulus became bigger, and all three stimuli were replayed once.

After completing all three training tasks in each session, subjects took a short identification test. There were 20 identification trials spoken by the trained speaker of each session, and subjects did not receive any feedback during the short test. At the end, they saw their percentage correct of the short test.

3.2.4.2. Pre/Post tests. Japanese subjects took pre and post tests to assess their perceptual improvement through the 10 training sessions. The tasks were identification, auditory discrimination and category discrimination. In the identification test, subjects saw an English /r/-/l/ minimal pair with the characters of the R-man and the L-man (as described earlier) on screen, and clicked the one that they thought they heard. Because younger Japanese children did not know the English phonemes of /r/ and /l/, all subjects heard an example English /r/-/l/ minimal pair, *rack* and *lack*, to remember how the /r/-/l/ phonemes sound. To ensure everyone understood the task, they also had two practice trials of identification without feedback, consisting of *rack-lack* and *rock-lock*. After the practice trials, subjects were examined on their identification of 40 /r/-/l/ word-initial minimal-pair words (20 pairs), 40 /r/-/l/ word-medial minimal-pair words (20 pairs) and 40 consonant cluster (containing /r/-/l/) minimal-pair words (20 pairs). After every 10 trials, a message such as *Level-up* was shown on the screen to let the subjects have a break and to encourage them to do more trials. All stimuli were novel English words spoken by two untrained speakers.

In the auditory discrimination task, subjects heard three synthesised /ra/-/la/ stimuli from three identical white cats on screen, and clicked the cat that produced a different sound from the other two. As explained in Chapter 2, there were three stimulus pairs contrasting high and low F3 at the /r/-/l/ boundary, high and low F2 at the /r/-/l/ boundary, and high and low F3 within the /r/ category. Subjects took 72 trials in total (3 stimulus pairs with 2 possible odd stimuli of /ra/ and /la/, and 3 odd stimulus positions with 4 repetitions). After every 18 trials, they saw a message of *Level up* or *Break Time* on the screen, to allow them to have a break.

The final task was category discrimination. Three dogs were displayed on the screen and produced a word-initial /r/-/l/ minimal pair with three different speakers' voices (e.g., *lace*, *lace*, *race*). Subjects did not have to identify the phoneme but needed to ignore speaker differences and to click the dog producing a categorically different word-initial consonant (e.g., *race*). To ensure that all subjects understood the task, all subjects practiced the task with 12 trials of the English /s/-/g/ contrasts of two minimal pairs before being tested with the /r/-/l/ contrast. The stimuli of the /r/-/l/ contrast were eight minimal pairs (i.e., *lord-roared, race-lace, lied-ride, loon-rune*, *lug-rug, load-road*, *lapse-wraps*) produced by the four SSBE speakers (two females, two males) who were not used as training corpora. The three stimuli in a trial were produced by three randomly chosen SSBE speakers, and the odd stimulus of /r/ or /l/ was equally balanced in frequencies and positions. Japanese speakers took 48 trials in total because of the 8 minimal pairs with the 2 odd stimuli at 3 different positions. After each 12 trials, subjects saw a message *Level up* with the stars they had been rewarded with on the screen, and had a break if they wanted. During each test, subjects received no feedback, and they were not allowed to listen to a trial again.

3.3. Results

3.3.1. Identification

Figure 3.1 displays identification accuracy at pre and post tests by the four age groups of Japanese speakers who had 10 training sessions (training) and those who had no training (control). As can be seen from the figure, while no age group in the control group improved their identification accuracy from pre to post test, all age groups in the training group improved their identification accuracy. However, there appeared to be a difference in the amount of improvement affected by training between age groups. The training effects on the improvement were higher for older children and adolescents than for the youngest and oldest age groups, although there was a large amount of variance in the improvement for older children in the training group.

For statistical analysis, a logistic mixed effects model was used. To make a decision of how the factor of age should be included in the model, the age factor was tested in the model fit. Amongst the possibilities of a linear variable, a log linear variable, a segmented continuous variable and a categorical variable, age as a categorical variable was the best fit in the model. The categorical variable of age had four levels: younger children aged 6;11-8;1 (years; months), older children aged 8;1-12;4, adolescents aged 15;7-18;6, and adults aged 25;3-59;8.

The best-fitting logistic mixed effects model on identification accuracy included the fixed factors of training (i.e., yes: training, no: control), testing block (pre, post), /r/-/l/ position (i.e., word-initial, word-medial, consonant cluster), age group (i.e., younger children, older children, adolescents, adults) and all possible 2-way interactions. Although the 3-way interaction of testing block, /r/-/l/ position and age group was excluded from the model during the process of the model comparisons, all the other 3-way interactions fitted into the model. The 4-way interaction of training, testing block, /r/-/l/ position and age group did not fit in the model and was excluded. The random factors were intercepts of subjects and stimulus word. The stimulus word was nested into speaker, as the two native English speakers produced different words.

Table 3.1 shows the results of a logistic mixed effects model's type-III Wald chi square test which examined the effects of each fixed factor for identification accuracy. As shown in the table, the interaction between training and testing block was significant, suggesting that the amount of improvement in identification accuracy from pre to post test by trained Japanese speakers was significantly higher than by untrained Japanese speakers in the control group. In other words, training affected the improvement of identification accuracy. In addition, as expected from Figure 3.1, there was also a significant 3-way interaction of training, testing block and age group. This suggests that the amount of improvement affected by training was different between the four age groups. Planned contrasts for the 3-way interaction of training, testing block and age group demonstrated that there was a significant quadratic relation, $b = -.08$, SE = .03, z = -2.59, $p < .01$. This suggests that training affected the improvement of identification accuracy for older children and adolescents more than for the youngest and oldest groups. These results revealed that older children had some advantages in learning to identify the English /r/-/l/ contrast, and adolescents after puberty were still able to improve their identification ability more than adults.

Although older children and adolescents improved their identification accuracy by virtue of training more than the youngest and oldest age groups did, post-hoc analyses for each age group demonstrated that there was a significant interaction between training and testing block for adults, $\chi^2(1) = 28.70$, $p < .001$, as well as for adolescents, $\chi^2(1) = 37.34$, $p < .001$, and older children, $\chi^2(1) = 31.08$, $p < .001$, but not for younger children, $\chi^2(1) = 2.77$, $p = .096$. These results suggest that training affected the improvement of identification accuracy for all age groups except younger children.

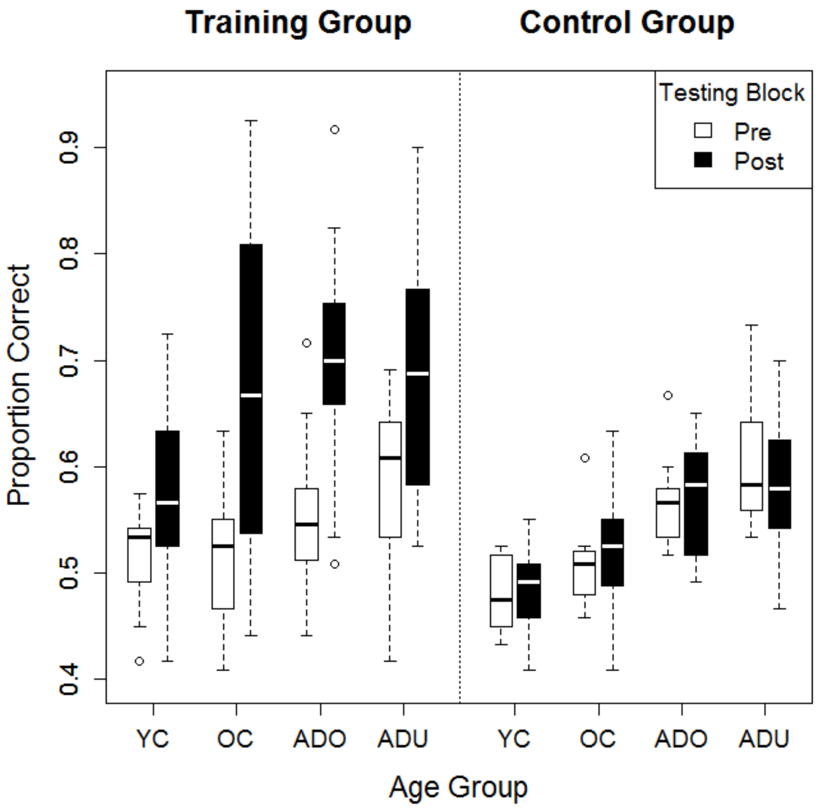


Figure 3.1. Boxplots of identification accuracy for the English /r/-/l/ contrast by trained (left) and untrained (right) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	28.01	1	< .001***
Testing block	83.48	1	< .001***
/r/-/l/ position	18.65	2	< .001***
Age group	21.10	3	< .001***
Training & Testing block	80.32	1	< .001***
Training & /r/-/l/ position	27.00	2	< .001***
Training & Age group	2.08	3	> .05
Testing block & /r/-/l/ position	45.04	2	< .001***
Testing block & Age group	12.96	3	< .001**
/r/-/l/ position & Age group	8.98	6	> .05
Training &Testing block & /r/-/l/ position	28.49	2	< .001***
Training &Testing block & Age group	7.96	3	=.047*
Training & /r/-/l/ position & Age group	17.06	6	< .01**

Table 3.1. Type-III analysis-of-variance table based on Wald chi-square tests for a logistic mixed effects model on /r/-/l/ identification accuracy by trained and untrained Japanese speakers at pre and post tests.

Figure 3.2 displays identification accuracy of the three /r/-/l/ positions (i.e., word-intial, word-medial, consonant cluster) by Japanese speakers who had 10 training sessions and those who had no training between pre and post tests. As can be seen from the figure, identification accuracies of all /r/-/l/ positions were improved by Japanese speakers in the training group more than by those in the control group. However, training seemed to affect the improvement in identification accuracy of the trained word-intial position more than in the word-medial and consonant cluster positions.

As shown in Table 3.1, the 3-way interaction of training, testing block and /r/-/l/ position was significant. This suggests that training affected the improvement of identification accuracy, but the training effect was different between the three /r/-/l/ positions. The planned orthogonal contrasts for the 3-way interaction demonstrated that training affected the improvement of word-initial identification accuracy significantly more than that in word-medial and consonant cluster positions, $b = -.05$, SE = .02, z = -5.28, $p < .001$. However, there was no significant difference in the training effect between word-medial and consonant cluster positions, $b = .01$, SE = .02, z = .77, $p > .05$.

Even though there were still significantly fewer training effects in word-medial and consonant cluster positions than in the trained word-initial position, it appears that word-initial phonetic perceptual training affected the improvement of identification accuracy for the word-medial and consonant cluster positions as well as the word-initial position. Post-hoc analyses for each position demonstrated that there was a significant interaction of training and testing block for all of the word-initial, $\chi^2(1) = 90.78$, $p < .001$, word-medial, $\chi^2(1) = 17.06$, $p < .001$, and consonant cluster positions, $\chi^2(1) = 9.06$, $p < .01$. These results suggest that training affected the improvement of identification accuracy for all the three /r/-/l/ positions.

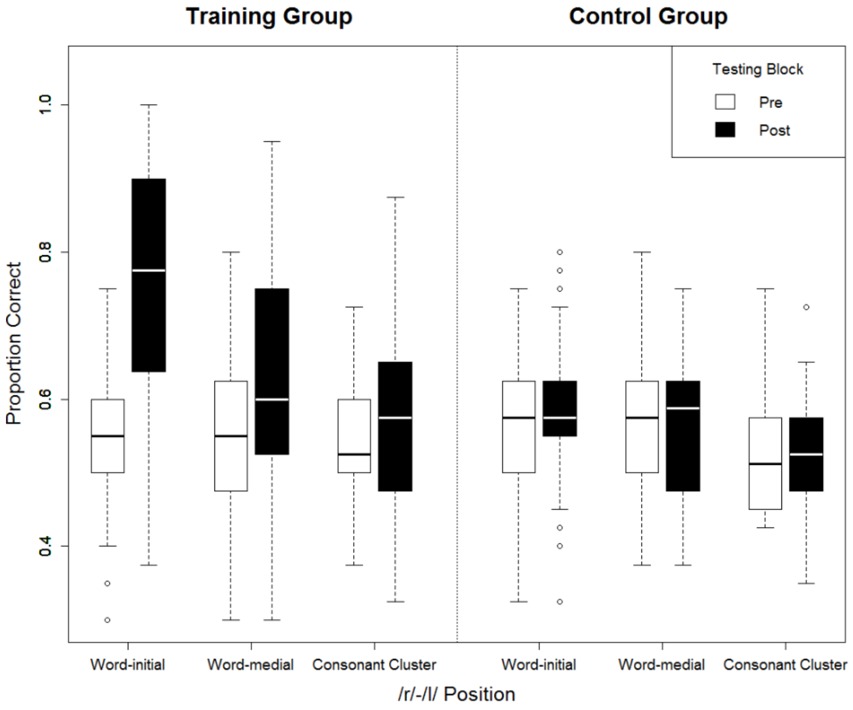


Figure 3.2. Boxplots of identification accuracy for the English /r/-/l/ contrast at word-initial, word-medial and consonant cluster positions by trained (left) and untrained (right) Japanese speakers at pre (white boxes) and post tests (black boxes).

The 4-way interaction of training, testing block, /r/-/l/ position and age group did not fit in the model. This suggests that the effects of the significant 3-way interaction of training, testing block and /r/-/l/ position did not significantly differ by age group.

Although there was no significant effect of the 4-way interaction, in order to examine whether the improvement in word-initial /r/-/l/ identification accuracy by younger learners was limited to approximately 15% on average, descriptive analyses for word-initial identification accuracy were conducted for trained Japanese speakers. In terms of improvement in percentage correct on average, older children improved their word-initial English /r/-/l/ identification accuracy most, $M_{\text{post-pre}}$ =

26.67%, followed by adolescents, $M_{\text{post-pre}}$ = 24.22%, younger children, $M_{\text{post-pre}}$ = 18.85%, and adults, $M_{\text{post-pre}}$ = 15.60%. This result suggests that there is no 15% average improvement limitation for younger learners, although previous studies demonstrated that adult learners often showed this limitation in their improvement (Iverson et al., 2005).

3.3.2. Auditory Discrimination

Because different acoustic cues were tested in a higher order interaction, auditory discrimination was analysed for each stimulus pair (i.e., F3 difference at the English /r/-/l/ boundary, F2 difference at the English /r/-/l/ boundary, F3 difference within the English /r/ category).

Figure 3.3 displays auditory discrimination accuracy of F3 difference at the English /r/-/l/ boundary by the four age groups of Japanese speakers who had 10 training sessions and those who had no training between pre and post tests. As shown in the figure, all age groups of trained Japanese speakers, except adults, improved their F3 sensitivity at the English /r/-/l/ boundary more than the equivalent age groups who had no training.

Table 3.2 shows the results of a logistic mixed effects model's type-III Wald chi square test which examined each fixed factor's effect on F3 sensitivity at the English /r/-/l/ boundary. As listed in the table, the best-fitting logistic mixed effects model for F3 sensitivity at the boundary included the fixed factors of training, testing block, age group, and three 2-way interactions and a 3-way interaction. The random factors were crossed intercepts of subject and stimulus. As shown in the table, the logistic mixed effects model demonstrated that there was a significant interaction between training and testing block. This suggests that trained Japanese speakers improved their F3 sensitivity at the English /r/-/l/ boundary from pre to post test significantly more than untrained Japanese speakers did. That is, training affected the improvement of the F3 sensitivity at the boundary from pre to post test. In addition, the 3-way interaction of training, testing block and age group was also significant, suggesting that training effects were significantly different between the four age groups. Post-hoc analyses of the 3-way interaction demonstrated that training significantly affected the improvement in F3 sensitivity at the English /r/-/l/ boundary for adolescents, $\chi^2(1) = 13.47$, $p < .001$, but not for adults, $\chi^2(1) = .20$, $p > .05$, older children, $\chi^2(1) = .55$, $p > .05$, or younger children, $\chi^2(1) = .91$, $p > .05$. Although Figure 3.3 suggests that the older children and younger children in the training group seemed to improve their sensitivity to the primary acoustic cue for the /r/-/l/ contrast more than subjects in the corresponding age groups in the control group, there was no significant training effects for them.

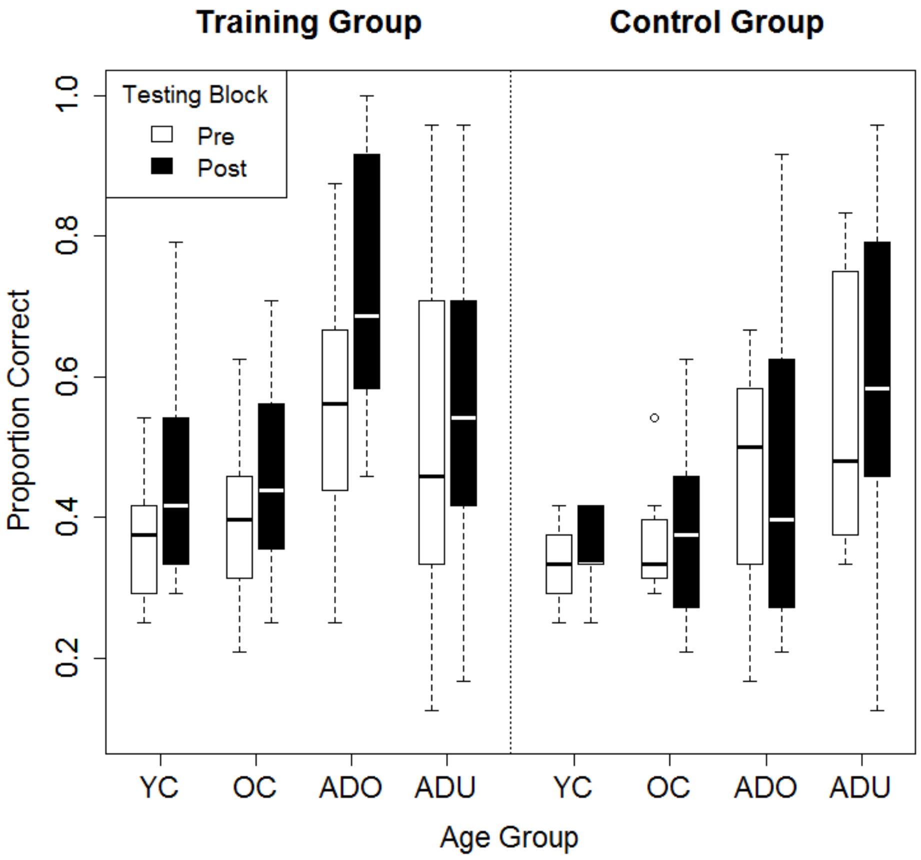


Figure 3.3. Boxplots of auditory discrimination accuracy of F3 sensitivity at the English /r/-/l/ boundary, by both trained (left) and untrained (right) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	8.02	1	< .01**
Testing block	13.40	1	< .001***
Age group	20.56	3	< .001***
Training & Testing block	5.81	1	= .016*
Training & Age group	8.34	3	= .039*
Testing block & Age group	1.86	3	> .05
Training & Testing block & Age group	8.95	3	= .030*

Table 3.2. Type-III analysis-of-variance table based on Wald chi-square tests of a logistic mixed effects model on auditory discrimination for F3 sensitivity at the English /r/-/l/ boundary by trained and untrained Japanese speakers at pre and post tests.

Figure 3.4 displays the auditory discrimination accuracy for F2 difference at the English /r/-/l/ boundary by the four age groups of trained and untrained Japanese speakers. For F2 sensitivity at the boundary, test-retest reliability appeared to be very low, as Japanese speakers in the control group improved their sensitivity to F2 at the boundary. As can be seen from the figure, training may have affected older children and adults in the training group by not improving their F2 sensitivity. However, this result may not be reliable. Because older children and adults in the control groups seemingly randomly improved their sensitivity to F2, the improvement differences between the training and the control groups in these age groups may not be substantial.

Table 3.3 shows the results of a logistic mixed effects model's type-III Wald chi square test which examined each fixed factor's effect on F2 sensitivity at the English /r/-/l/ boundary. The best-fitting logistic mixed model included the fixed factor of training, testing block, age group, and all possible 2-way and 3-way interactions; the random factors were crossed intercepts of subject and stimulus. As shown in the table, the logistic mixed effects model demonstrated that there was a significant 2-way interaction between training and testing block. This suggests that there was a significant difference in the change of sensitivity to F2 at the English /r/-/l/ boundary from pre to post test between the training and control groups. There was also a significant 3-way interaction of training, testing block and age group, suggesting that the training effects were different between age groups. As expected from Figure 3.4, post-hoc analyses for each age group demonstrated that there was a significant 2-way interaction of training and testing block for older children, $\chi^2(1) = 7.26$, $p < .01$, and adults, $\chi^2(1) = 6.31$, $p = .012$, but not for younger children, $\chi^2(1) =$

1.32, $p > .05$, or adolescents, $\chi^2(1) = 1.42$, $p > .05$. These results suggest that training affected older children and adults by not improving their sensitivity to F2 at the boundary. However, this effect seems not to be caused by training but by control groups' random performance improvements in the task familiarisation.

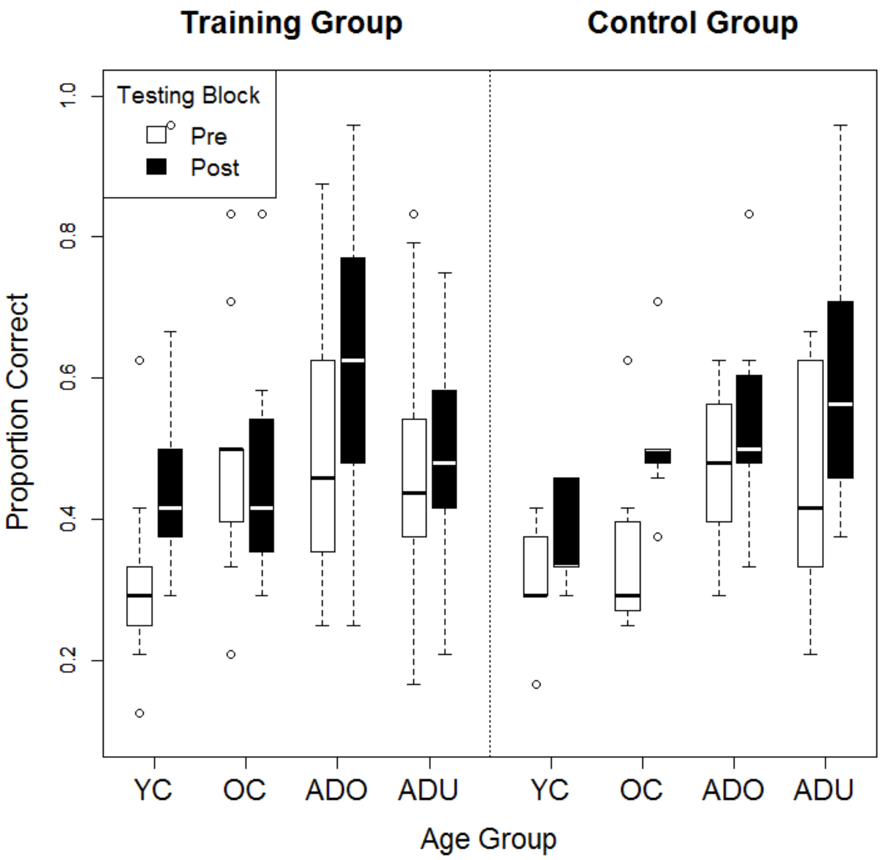


Figure 3.4. Boxplots of auditory discrimination accuracy of F2 sensitivity at the English /r/-/l/ boundary, by both trained (left) and untrained (right) Japanese speakers, broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	3.49	1	= .062
Testing block	35.37	1	< .001***
Age group	13.21	3	< .01**
Training & Testing block	1.49	1	> .05
Training & Age group	.62	3	> .05
Testing block & Age group	1.53	3	> .05
Training & Testing block & Age group	13.95	3	< .01**

Table 3.3. Type-III analysis-of-variance table based on Wald chi-square tests of a logistic mixed effects model on auditory discrimination for F2 sensitivity at the English /r/-/l/ boundary by trained and untrained Japanese speakers at pre and post tests.

Figure 3.5 displays auditory discrimination accuracy of F3 sensitivity within the English /r/ category for both trained and untrained Japanese speakers at pre and post tests. As can be seen from the figure, training did not affect the change in F3 sensitivity within the English /r/ category from pre to post test. Although the younger and older children in the training group seemed to reduce their perceptual sensitivity to this irrelevant acoustic cue, the training effect was not substantial.

Table 3.4 shows the results of a logistic mixed effects model’s type-III Wald chi square test which examined each fixed factor’s effect on F3 sensitivity within the English /r/ category. As listed in the table, the best-fitting logistic mixed effects model included fixed factors of training, testing block, age group and two 2-way interactions between training and age group, and between testing block and age group. The 2-way interaction between training and testing block and the 3-way interaction of training, testing block and age group were excluded from the model during the process of model comparisons, suggesting that there was no significant effects of these interactions. In other words, training did not affect the change in F3 sensitivity within the English /r/ category from pre to post test, and there was no difference in the training effects between the four age groups.

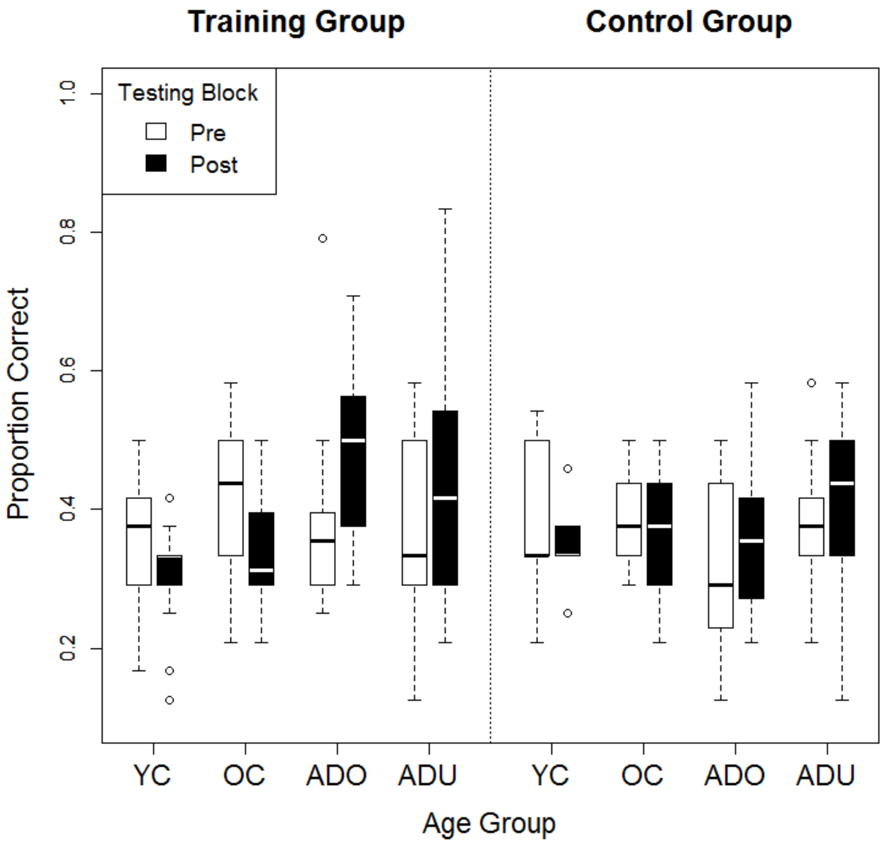


Figure 3.5. Boxplots of auditory discrimination accuracy of F3 sensitivity within the English /r/ category, by both trained (left) and untrained (right) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	3.49	1	= .062
Testing block	35.37	1	< .001***
Age group	13.21	3	< .01**
Training & Age group	.62	3	> .05
Testing block & Age group	1.53	3	> .05

Table 3.4. Type-III analysis-of-variance table based on Wald chi-square tests of a logistic mixed effects model on auditory discrimination for F3 sensitivity within the English /r/ category by trained and untrained Japanese speakers at pre and post tests.

3.3.3. Category Discrimination

Figure 3.6 displays the category discrimination accuracy by the four age groups of Japanese speakers who had 10 training sessions and those who had no training between pre and post tests. Although no age group who had no training improved their category discrimination accuracy from pre to post test, all age groups who had 10 training session improved their category discrimination accuracy. Similar to identification accuracy, training affected the improvement of category discrimination accuracy for older children and adolescents more than for the oldest and youngest age groups.

Table 3.5 shows the results of a logistic mixed effects model's type-III Wald chi square test which examined each fixed factor's effect for category discrimination. The best-fitting logistic mixed effects model included the fixed factors of training, testing block, age group and all possible 2-way and 3-way interactions; the random factors were crossed intercepts for subject and stimulus. As shown in the table, the logistic mixed effects model demonstrated that there was a significant interaction between training and testing block. This suggests that training affected the improvement in category discrimination accuracy from pre to post test. Furthermore, the significant 3-way interaction of training, testing block and age group suggests that the amount of improvement affected by training were different by age group. The planned contrasts for the 3-way interaction demonstrated that there was a significant quadratic relation with age group, $b = -.14$, SE = .05, z = -2.83, $p < .01$. This suggests that the improvement affected by training was more for older children and adolescents than for the youngest and oldest age groups.

Although there was a significant quadratic relation in the 3-way interaction, post-hoc analyses for each age group demonstrated that there was a significant interaction between training and testing block for adults, $\chi^2(1) = 13.00$, $p < .001$, as well as for adolescents, $\chi^2(1) = 40.89$, $p < .001$, and older children, $\chi^2(1) = 17.45$, $p < .001$, but not for younger children, $\chi^2(1) = 1.70$, $p > .05$. These results suggest that training affected the improvement of category discrimination for all age groups except younger children.

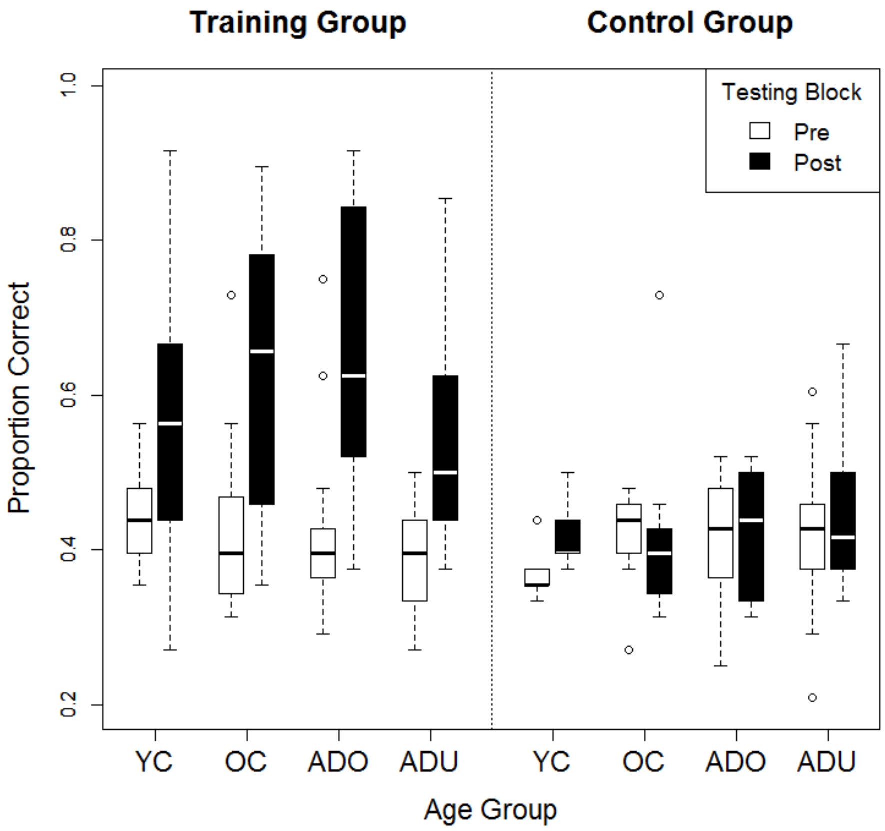


Figure 3.6. Boxplots of category discrimination accuracy by both trained (left) and untrained (right) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	37.94	1	< .001***
Testing block	91.31	1	< .001***
Age group	3.17	3	> .05
Training & Testing block	51.84	1	< .001***
Training & Age group	1.66	3	> .05
Testing block & Age group	3.38	3	> .05
Training & Testing block & Age group	9.49	3	= .023*

Table 3.5. Type-III analysis-of-variance table based on Wald chi-square tests of a logistic mixed effects model on category discrimination by trained and untrained Japanese speakers at pre and post tests.

3.3.4. Training Sessions

Figure 3.7 displays the identification accuracy of each training session. The figure shows the proportion correct of short identification tests (consisting of 20 trials) which were conducted at the end of each training session. The curved line for each age group is the predicted improvement slopes obtained by the best-fitting logistic mixed effects model. As can be seen from the figure, the improvement slopes for older children and adolescents are steeper than for adults. In other words, older children and adolescents improved more quickly and consistently than did adults.

The best-fitting logistic mixed effects model for the identification accuracy of training sessions included the random factor of word nested into speaker. The fixed factors were training session (i.e., 1 to 10), age group and the interaction between training session and age group. As the fixed factor of training session did not linearly relate to identification accuracy in the model, it was scaled by the Box-Cox transformation to improve the model fit (Box & Cox, 1964). The fixed factor of training session was transformed with lambda = .135.

The logistic mixed effects model demonstrated that there were significant main effects of training session, $\chi^2(1) = 97.16$, $p < .001$, and age group, $\chi^2(3) = 10.45$, $p = .015$. There was also a significant interaction between training session and age group, $\chi^2(3) = 15.39$, $p < .01$, suggesting that the steepness of the improvement slope was significantly different between the four age groups. Non-orthogonal contrasts demonstrated that the improvement slopes of adolescents, $b = .22$, SE = .06, z = 3.42, $p < .001$, and older children, $b = .24$, SE = .08, z = 3.10, $p < .01$, were significantly steeper than the adults' slope, although the improvement slope of younger children was not significantly different from the adults' one, $b = .10$, SE = .07, z = 1.51, $p > .05$. Furthermore, the improvement slopes of adolescents, $b = .11$, SE = .07, z = 1.73, $p = .083$, and older children, $b = .14$, SE = .08, z = 1.74, $p = .081$, were steeper than the younger children's slope, although the difference was only marginally significant. These results suggest that there seems to be a tendency for older children and adolescents to improve their identification accuracy more quickly and consistently than the youngest and oldest groups do. This implies the possibility that older children and adolescents might have continued to improve if they had had more than 10 training sessions.

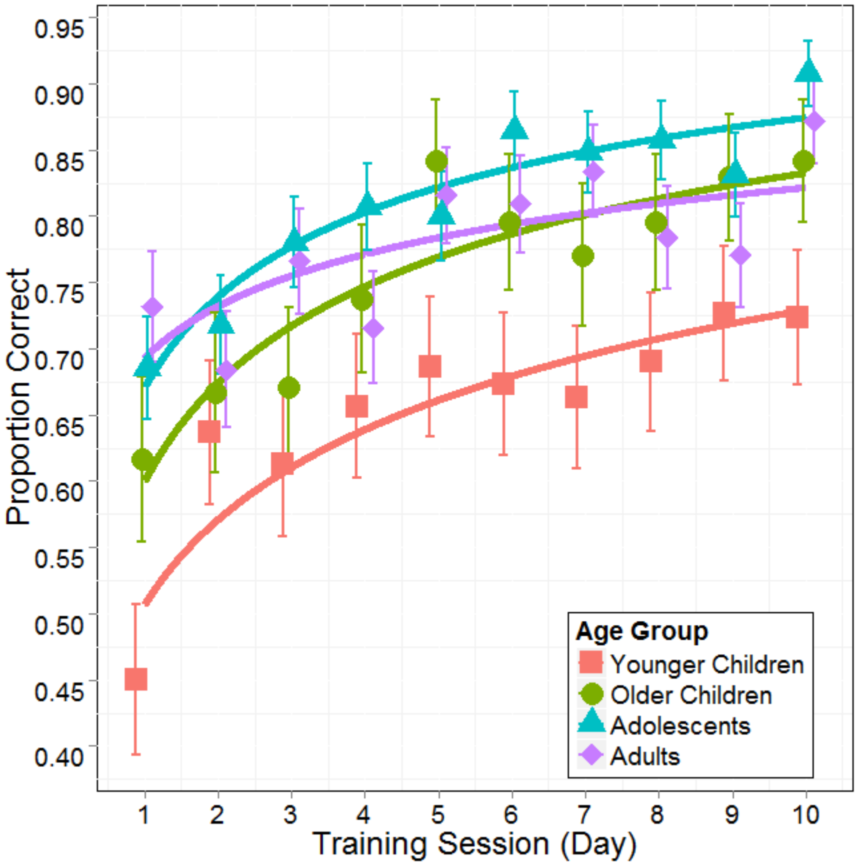


Figure 3.7. Identification accuracy of English /r/-/l/ (with confidence intervals) at each short test (20 trials) after each training session, performed by Japanese subjects broken down by the age groups of Younger Children (red), Older Children (green), Adolescents (blue) and Adults (purple). The curved line of each age group shows the predicted proportion correct obtained by a logistic mixed effects model.

3.4. Discussion

The present study examined how native Japanese speakers in four age groups improve their perception of English /r/ and /l/, using a computer-based perceptual training programme. As the sensitive/optimal period theory (Oyama, 1976; Werker & Tees, 2005) and SLM (Flege, 1995, 1999, 2002) suggest, younger learners improved their perception more than adults did when their L2 input was identical in quality and quantity. Regarding the phonological levels reflected in identification accuracy, adolescents and older children improved more than the youngest and oldest age groups. With regard to acoustic/phonetic levels reflected in auditory discrimination and category discrimination, adolescents improved their phonetic sensitivity to the primary acoustic cue (i.e., F3 at the /r/-/l/ boundary), and older children and adolescents improved their category discrimination abilities more than the youngest and oldest age groups.

These age effects may be ascribed to neural plasticity and L1 interference. Adults improved their perception of English /r/ and /l/, but their improvement in identification accuracy seemed to be limited to approximately 15% on average, and they did not improve in their sensitivity to the primary acoustic cue. This may be because their relatively fossilised neural plasticity constrains their learning ability, and their developed L1 phonetic units prevented them from learning English phonetic processing. On the other hand, adolescents improved both at the phonological and phonetic levels of perception. This may be because their neural plasticity is still flexible for language learning. Although the adolescents' L1 phonetic units may have been developed and the L1 categories may have interfered with learning English phonetic processing (SLM: Flege, 1995, 1999, 2002), they were still able to improve their phonological identification and phonetic sensitivity to the primary acoustic cue. Older children also improved the phonological and phonetic levels of perception more than adults, despite the immaturity of their sensory system and cognition. Although the improvement by older children was similar to that by adolescents, older children's improvement in phonetic and phonological perception may be caused by not only their greater neuroplasticity for language learning. Less interference from their L1 may have also a factor in their improvement. As children under the age of 12 years have not yet completed the establishment of their L1 categories (Hazan & Barrett, 2000), their L1 phonetic categories may not be the powerful attractors in perceiving L2 phonemes (SLM: Flege, 1995, 1999, 2002).

These findings suggest that younger learners' improvement in identification accuracy by training may not be limited to the same extent as those of adults. Iverson et al. (2005) demonstrated that adult Japanese speakers seem to be limited in their improvement of identification accuracy to around 15% on average. Hattori and Iverson (2009) found that Japanese speakers' phonetic processing of F3 contributes to difficulties in identifying the English /r/-/l/ contrast, and suggested that it would be necessary to improve the lower levels of phonetic perception. However, the phonetic processing of F3 is difficult to improve through daily communication in English, even in English-speaking countries (Ingvalson et al., 2011). The present study addressed younger learners' improvement in their perception of English /r/ and /l/ with the hypothesis that they are able to change their phonetic processing of F3 more than adults are. The results demonstrated that Japanese adolescents and older children learned English phonetic processing and improved their identification accuracy by much more than 15% on average. Although it was still difficult to completely improve both phonological and phonetic levels of perception within 10 training sessions, if Japanese younger learners continue training in the perception of English /r/ and /l/ then they may improve their phonetic and phonological abilities to a greater extent.

These results are not in line with previous studies of Heeren and Schouten (2010) and Wang and Kuhl (2003). These previous studies demonstrated that the improvements in discriminating consonant duration and L2 lexical tones were comparable between adults and children, but the present study clearly demonstrated the age effects on learning L2 segmental perception. This inconsistent result between previous studies and the present study can be attributed to the difference in the tested acoustic cues. The present study examined subjects' sensitivities to spectral cues such as formant frequencies, but not pitch perception or the duration of consonants. Because Japanese older children's perception is not affected by their L1 phonetic processing as much as adults' perception is, and because they might rely more on perceiving formant transitions than do adults as the DWS model suggests (Nittrouer et al., 2000; Nittrouer & Miller, 1997; Nittrouer, 1992), children may have the advantage when learning L2 phoneme contrasts whose primary acoustic cues are formant frequencies.

However, contrary to the hypotheses based on the sensitive/optimal period (Oyama, 1976; Werker & Tees, 2005) and SLM (Flege, 1995, 1999, 2002), the present study demonstrated that younger children did not improve in their perception of English /r/ and /l/ more than older children. This may be because of the immaturity of their phonemic awareness as well as the immaturity of their selective attention and auditory sensory system (Berman & Friedman, 1995; Halliday, Taylor, Edmondson-Jones, & Moore, 2008). According to Ogawa and Tanemura (2001), Japanese children tend to acquire basic phonological or moraic awareness by the age 4 or 5 years, but their phonemic awareness of segmenting moras needs more time to develop (Endo, 1991; Matsumoto, 2004, 2006). Supporting this argument, Mann (1986) demonstrated that Japanese children aged between 6 and 7 years old were not able to delete word-initial phonemes, whereas 9-10 year-old children were significantly better at doing so. These findings indicate that the younger children in the present study may have difficulty in understanding the concept of word-initial phonemes of training stimuli (i.e., word-initial English /r/-/l/ minimal pair words).

The development of phonemic awareness may also explain the reason why older children showed a wide variance in /r/-/l/ identification accuracy. Children aged from 8 to 12 years old were in the middle of developing their phonemic awareness. For some, it might hot have been fully developed, and so they might not have understood the concept of word-initial phonemes. Even if they understood the concept of word-initial phonemes, it may not have been possible for some children to generalise to other phonetic environments. As mentioned above, for Japanese children, there is a sharp development in phonemic awareness from 9 to 11 years of age (Mann, 1986), and word-initial consonant segmentations are easier than word-medial/final vowel segmentations (Endo, 1991). This suggests that some older children may still have had difficulties understanding the concept of word-medial and consonant cluster phonemes. Although they may be able to access new L2 phoneme categories of English /r/ and /l/ in word-initial position, the improvement would not transfer to the other phonetic positions for subjects who have not completely acquired phonemic awareness.

In conclusion, the goal of the present study was to develop a clearer picture of the effects of age on L2 speech perceptual learning, by controlling the quality and quantity of L2 input. Although younger children aged under 8 years old may have a problem in learning L2 phoneme perception through a computer-based perceptual training programme, it has been demonstrated that younger learners have some advantage for learning L2 speech perception due to their greater brain plasticity and less L1 interference. This suggests that both the sensitive/optimal period hypothesis (Oyama, 1976; Werker & Tees, 2005), positing the effects of biological neural maturation, and the Speech Learning Model (Flege, 1995), proposing the effects of the L1 phonetic attraction, are supported. The optimal timing for learning L2 phoneme perception with a training programme appears to be somewhere between childhood and adolescence.

Although older children had an advantage for learning the perception of English /r/ and /l/, it is not clear how long the training effects last for children. Adults retain their improved ability of identification for at least 6 months (Lively et al., 1994), but children may not retain this ability for the same period of time, or at all. Future studies are necessary to investigate the retention of learning effects for younger learners. Moreover, there is no research as yet into whether perceptual training transfers to younger learners’ production of English /r/ and /l/. The production improvement in identifiability and acoustic realisations by the four age groups are examined in Chapter 4.

Chapter 4: Perceptual training effects on production and the link between perception and production

4.1. Introduction

In L1 speech development, it is intuitively convincing that there is a connection between perception and production. According to Kuhl et al. (2008), when infants are exposed to an ambient language after birth, their sensitivity to the distributional patterns of segmental and supra-segmental features leads to phonetic learning. During their acquisition of L1 phonetic categories, they also develop a connection between speech perception and production by vocalising and imitating the phonetic realisations they perceive. Their vocalisations and imitations seem to cause language-specific speech production, and monitoring their own production may accelerate learning language-specific speech perception. Thus, acquiring perception and production in L1 seems to be correlated and cannot be separated.

On the other hand, the relation between perception and production in an L2 seems to be different from that in L1. In regard to the English /r/-/l/ contrast for Japanese speakers, perceptual ability seems to be preceded by the ability to produce. Sheldon and Strange (1982) demonstrated that some Japanese speakers' production of English /r/ and /l/ were more correctly identified by native English speakers than by the speakers themselves. Moreover, previous training studies suggest that there are separate underlying abilities for perception and production in an L2. Hattori

(2009) demonstrated that articulatory training which improved Japanese speakers' productions to native-like levels did not improve perceptual identification accuracy at all. Bradlow et al. (1997) demonstrated that although perceptual training significantly improved Japanese speakers' productions, there was no correlation in improvement between perceptual identification and production identifiability by native listeners.

Bradlow et al. (1997) explained their study's results referring to three theories in L1 speech perception: motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), direct realist theory (Fowler, 1996), and general auditory and learning approaches to speech perception. However, none of these explain the findings of recent studies in L2 learning. The motor theory proposes that listeners perceive speech sounds based on articulatory gestures including neuromotor commands to the articulators, or intended gestures. While production improvement can be explained by the acquired articulatory gestures through perceptual training, this theory cannot explain one-way training effects (e.g., articulatory training does not improve perceptual identification accuracy; Hattori, 2009). In addition, this theory cannot explain why there appears to be no correlation between perceptual identification improvement and production identifiability improvement after perceptual training (Bradlow et al., 1997; Iverson et al., 2011).

Similar to motor theory, direct realist theory proposes that the objects of speech perception are articulatory gestures, not acoustic events. This theory asserts that the actual structured vocal tract movements are the objects rather than the intended gestures, and denies a speech-specific mechanism. However, this theory still cannot explain one-way training effects (Hattori, 2009) or the independence of perception and production improvement (Bradlow et al., 1997; Iverson et al., 2011).

In contrast to these two theories, general auditory approaches do not assert that there is a medium such as an articulatory gesture that relates to production, but propose that the speech sounds are perceived as acoustic signals constructing phonemes and words. This theory can explain the independence of perception and production, but does not explain the finding that perceptual training can improve L2 learners' production without any instruction or feedback in pronunciation (Bradlow et al., 1997). Although monitoring learners' own production with their improved internal acoustic targets in perception may help learners improve their production, the general auditory approaches posit that improvement in production occurs during actual productions. Since perceptual training does not require productions, the improvement in production cannot be explained by general auditory approaches. Thus, none of the three theories can fully explain the theoretical frameworks of L2 speech perception and production learning.

It may be more plausible to consider that there is a link between perception and production in L2. Bradlow et al. (1997) considered that their findings supported motor theory and direct realist theory, because their perceptual training of English /r/ and /l/ improved Japanese speakers' productions of these phonemes without any instruction in production. They interpreted the results showing no correlation in improvement between perception and production as evidencing non-parallel processes of perception and production within individual subjects. That is, two separate processes of perception and production proceed at different rates within individuals, whereas perception training transfers to the production domain (Bradlow et al., 1997). However, there remains the question of what process of production is affected by perceptual training. There may be some acoustic realisations improving more than others, and these acoustic changes may be correlated with perceptual improvement, but no study has yet researched the production improvement in terms of acoustic details.

The present study examined the acoustic improvement in production of English /r/ and /l/ by the same Japanese speakers who participated in the perceptual training in Chapter 3. There were three main aims of this study. Firstly, this study examined the improvement in production identifiability and acoustic realisations. The previous studies by Bradlow et al. (1999, 1997) tested identifiability and intelligibility of Japanese speakers' production of the English /r/-/l/ contrast but not acoustic change. Since no previous studies have researched acoustic improvement in production through perceptual training, this study examined how closure duration, transition duration and F3 frequency changed after training. Secondly, age effects on production improvement were tested. Since perceptual training effects were different between age groups as described in the previous chapter, production improvements in identifiability and in acoustic measurements would also be expected to be different. Older children and adolescents would be expected to improve their production identifiability and acoustic realisations more than the youngest and oldest groups, because older children and adolescents improved in both perceptual identification accuracy of the English /r/-/l/ contrast and phonetic processing of F3 more than the other age groups did. Finally, the link in improvement between perception and production was examined using correlation tests. If there are two separate underlying factors for speech perception and production in L2, no correlation would be expected between the perception and production domains.

4.2. Method

4.2.1. Subjects

Same as the experiment in Chapter 3 (pp. 95-97).

4.2.2. Apparatus

Same as the experiment in Chapter 3 apart from perception tests (p. 97).

4.2.3. Stimuli

Same as the experiment in Chapter 3 apart from perception tests (pp. 97-98).

4.2.4. Procedure

4.2.4.1. Training. Same as the experiment in Chapter 3 (pp. 98-101).

4.2.4.2. Pre/Post tests. Before and after the 10 training sessions, the Japanese speakers had their production of word-initial English /r/-/l/ minimal pairs examined (e.g., *reap-leap*). Recording was carried out in quiet rooms in Japan, using an Audio Technica AT2020 USB microphone connected to a laptop. The task was to repeat 40 novel /r/-/l/ minimal-pair words (20 pairs) at least 2500 ms after an auditory prompt. Each word was displayed on screen with an orthographic visual prompt, and accompanied by an audio clip of the word read by a female SSBE speaker as an auditory prompt. After the auditory prompt, a male SSBE speaker said “read the word after the tone”, which was then followed by a beep sound. After hearing the beep sound, subjects produced the same word they saw and heard. The word order was random but the same for all subjects at pre and post tests. The inter-stimulus interval (ISI) including the instructions and the beep sound was 2500 ms.

In order to measure the identifiability of English /r/-/l/ production, 10 native British English speakers who were phonetically trained identified the stimuli using a task of two-alternative forced choice of English /r/ and /l/. Due to the high number of produced stimuli (40 English /r/-/l/ minimal-pair words from 85 Japanese speakers at pre and post tests), each English subject was tested with two minimal pairs (four words). That is, English subjects identified all Japanese speakers' productions at pre and post tests but only for two minimal pairs each, so that 20 minimal pairs (40 words) were completed by 10 English speakers. English subjects were able to replay the stimulus as many times as they wanted, and they received no feedback.

Three acoustic cues for the English /r/-/l/ contrast (closure duration, transition duration and F3) were also measured to investigate the improvement in the production domain. To avoid the effect of the following vowels on the formant frequencies, five minimal pairs whose vowels were similar to those in the Japanese vowel system were chosen (i.e., *raw-law*, *reap-leap*, *rest-lest*, *root-loot*, *rug-lug*). Care was taken to avoid minimal pairs in which the spelling differed between the two words to ensure there was no effect from orthography.

For cross-speaker comparisons of acoustic changes, F3 frequencies of English /r/ and /l/ were normalised between subjects. In the procedure, F2 frequencies of the following vowels were measured. After calculating the grand mean of the log F2 values, a scale factor was created for each subject with the log grand mean divided by each subject's log mean of F2. After scaling F3 values with the scale factor for each subject, those values were rescaled into Hertz.

Out of 1810 tokens in total, only 1284 were included in the acoustic analyses. There were 388 tokens of apico-alveolar tap [ɾ] and 138 tokens of other consonants (e.g., /b/ for English /r/) that were often produced by children. These 526 tokens were not acoustically measurable in duration and formants, and were therefore excluded from acoustic analysis.

4.3. Results

4.3.1. Identifiability of English /r/-/l/ Production

Figure 4.1 displays the improvement in production identifiability of English /r/ and /l/ from pre to post test by the four age groups of trained and untrained Japanese speakers. Compared to the control groups, who had no training between pre and post tests, all age groups in the training group seemed to improve their production identifiability of the English /r/-/l/ contrast. However, adolescents seemed to improve their production identifiability the most by training.

Table 4.1 shows the results of a logistic mixed effects model's type-III Wald chi square test which examined each fixed factor's effect for production identifiability of English /r/ and /l/. As shown in the table, the best-fitting logistic mixed effects model on production identifiability included fixed factors of training (yes: training, no: control), testing block (pre, post), age group (younger children, older children, adolescents, adults) and perceptual identification improvement. The perceptual identification improvement was taken as a simple difference score of the proportion correct from pre to post test in perceptual identification of word-initial English /r/-/l/ minimal-pair words. This was included because it significantly improved the model fit to data. However, the effect of perceptual identification improvement in the model is not mentioned in this section, because the link between perception and production improvement is reported in more details in correlation tests in a later section (4.3.3. Links between Perception and Production Improvement). Six 2-way and three 3-way interactions were also included in the model. The random factors were crossed intercepts of English listener, word and Japanese speaker; random slopes for testing block were excluded.

As in Table 4.1, although there was no significant interaction between training and testing block, the effect of the 3-way interaction of training, testing block and age group was significant. This significant 3-way interaction suggests that the training effect for the improvement in production identifiability was different according to age group. Planned contrasts for the 3-way interaction demonstrated that the training effect on the improvement was in significant quadratic, b = -.19, SE = .07, z = - 2.69, $p < .01$, and cubic relations with age group, b = -14, SE = .07, z = - 2.05, $p = .041$. These results suggest that training affected adolescents’ improvement the most, followed by older children’s, adults’ and younger children’s improvements. Post-hoc analyses for each age group with fixed factors of training, testing block and their interaction demonstrated that there was a significant interaction between training and testing block for adults, $\chi^2(1) = 8.91$, $p < .01$, adolescents, $\chi^2(1) = 22.34$, $p < .001$, and older children, $\chi^2(1) = 6.17$, $p = .013$, but not for younger children, $\chi^2(1) = 2.17$, $p > .05$. These results suggest that training affected the improvement in production identifiability for all age groups except younger children. Although the training improved the production identifiability of adults and older children to a similar extent, adolescents and older children improved more than the youngest and oldest age groups.

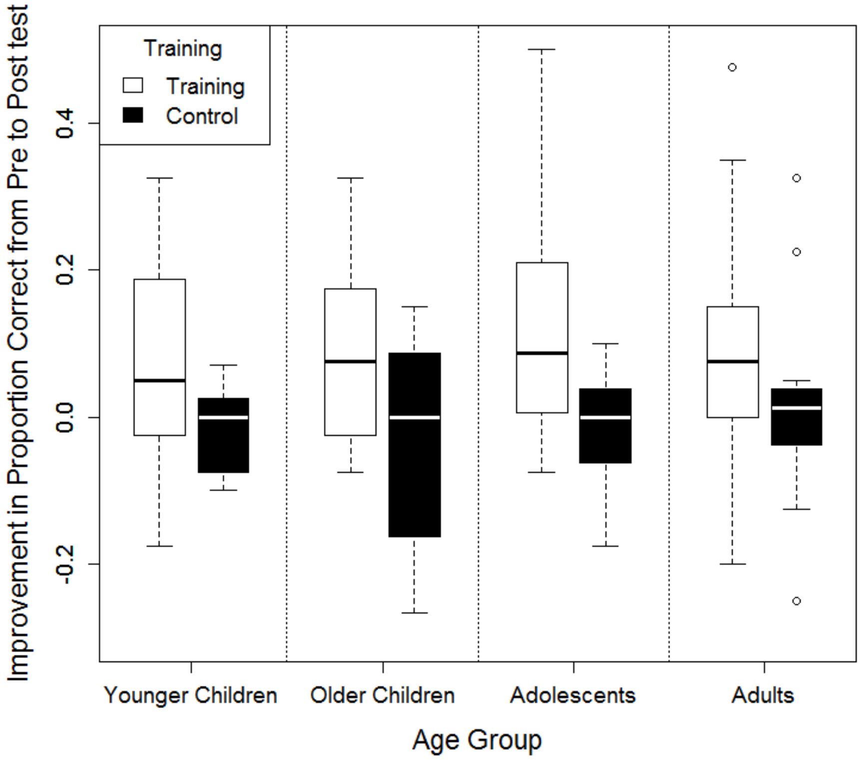


Figure 4.1. Boxplots of identifiability improvement of English /r/-/l/ production from pre to post test by trained (white boxes) and untrained (black boxes) Japanese speakers broken down by the age groups of Younger Children, Older Children, Adolescents and Adults.

Fixed Factors	χ^2	**Df**	***p***
Training	3.09	1	= .079
Testing block	1.70	1	> .05
Age group	17.39	3	< .001***
Perceptual identification improvement (PII)	.78	1	> .05
Training & Testing block	1.30	1	> .05
Training & Age group	4.77	3	> .05
Training & PII	.10	1	> .05
Testing block & Age group	6.88	3	= .075
Testing block & PII	1.00	1	> .05
Age group & PII	4.71	3	> .05
Training & Testing block & Age group	13.05	3	< .01**
Training & Testing block & PII	12.23	1	< .001***
Testing block & Age group & PII	25.53	3	< .001***

Table 4.1. Type-III analysis-of-variance table based on Wald chi-square tests for a logistic mixed effects model on English /r/-/l/ identifiability produced by trained and untrained Japanese speakers at pre and post tests.

4.3.2. Acoustic Change in Production

Acoustic changes of English /r/ and /l/ productions from pre to post tests were analysed with linear mixed effects models on closure duration, transition duration and F3 frequency. For the analyses of fixed factors' effects, ordinal *p*-values were reported. For planned contrasts, due to the fact that the ordinal *p*-values may be anti-conservative for unbalanced data, Markov chain Monte Carlo (MCMC) simulations were used with the *pvals.fnc* function in the package *languageR* of the software R (2.15.2) (Baayen et al., 2008; Cunnings, 2012). *P*MCMC values instead of the ordinal *p* values were reported with *b*-values (estimates) and standard errors (SE) as in Whitford and Titone (2012). The random factors were crossed intercepts of subject and word for all analyses. The random slopes for testing block were not included.

4.3.2.1. Closure duration. Figure 4.2 displays the closure duration of English /r/ and /l/ produced by Japanese speakers who had 10 training sessions and those who had no training between pre and post tests. English /r/ should be produced with a shorter closure duration, whereas English /l/ should be produced with a longer closure duration (Kent & Read, 2002). There seemed to be an improvement in the closure duration for both English /r/ and /l/ productions by the trained Japanese speakers, compared to untrained Japanese speakers. In other words, training affected Japanese speakers in that they slightly shortened the closure duration of English /r/, and slightly lengthened that of English /l/.

Table 3.2 shows the results of the linear mixed effects model's type-III Wald chi square test which examined the effects of each fixed factor for closure duration of the English /r/ and /l/ productions. As in the table, the linear mixed effects model included fixed factors of training, consonant (English /r/, English /l/), testing block, age group, and all possible interactions. The linear mixed effects model demonstrated that there was a significant 3-way interaction of training, consonant and testing block. This suggests that training affected Japanese speakers by changing their distinction of English /r/ and /l/ productions by using different closure durations.

Post-hoc analyses demonstrated that Japanese speakers in the training group significantly improved the distinction of English /r/ and /l/ by using different closure durations, $\chi^2(1) = 9.73$, $p < .01$, and that Japanese speakers in the control group did not improve the distinction, $\chi^2(1) = 2.08$, $p > .05$. However, the improvement in the distinction with the closure duration by trained Japanese speakers was not attributed to both shortening English /r/ and lengthening English /l/, but only to lengthening the closure duration of English /l/ by 12.69 ms on average, $\chi^2(1) = 13.47$, $p < .001$.

Japanese speakers in the training group shortened the closure duration of English /r/ after training by 2.12 ms on average, but it was not significant, $\chi^2(1) = 1.13$, $p > .05$.

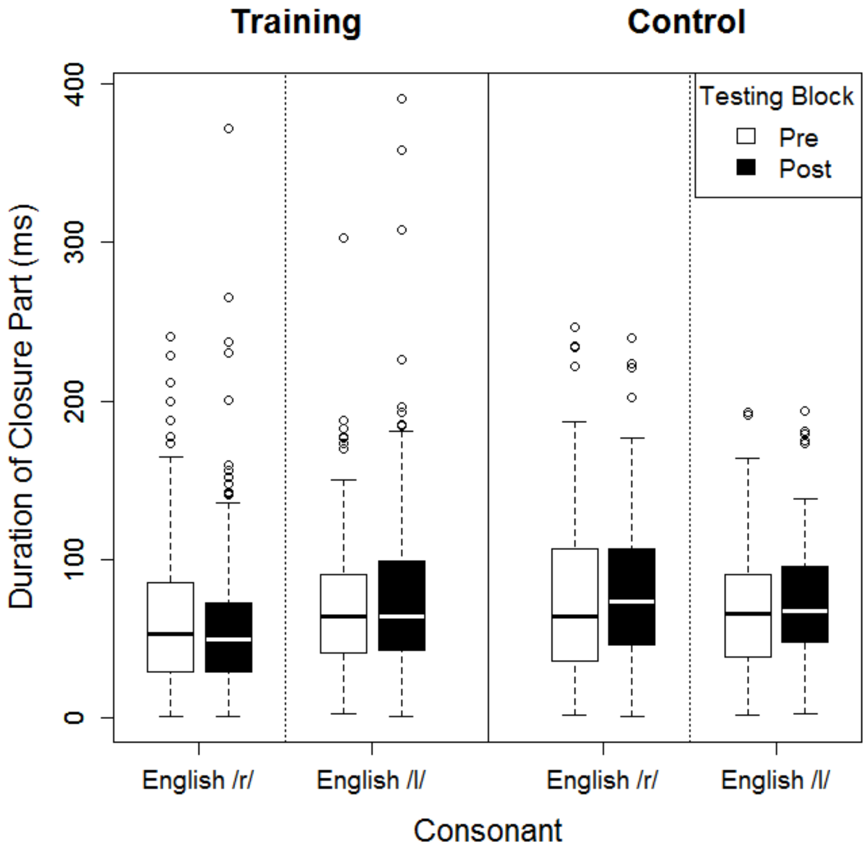


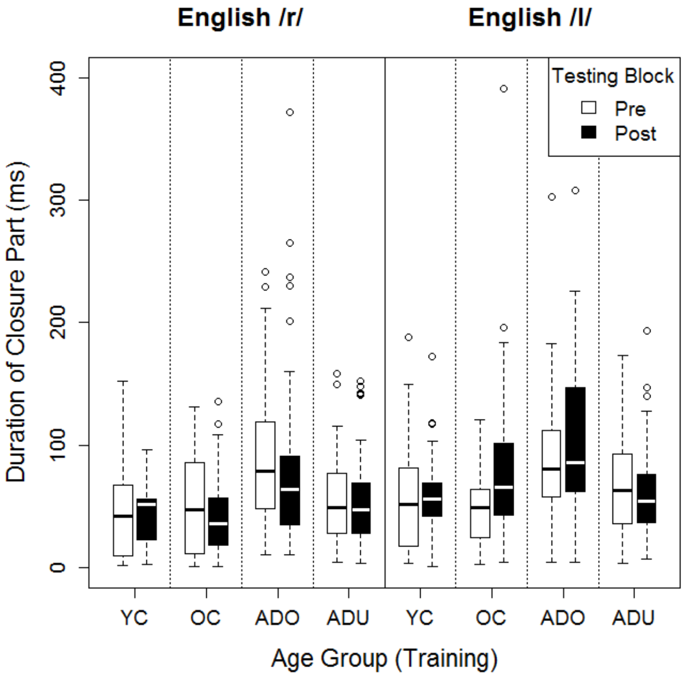
Figure 4.2. Boxplots of closure duration of English /r/ and /l/ produced by trained (left) and untrained (right) Japanese speakers at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	0.21	1	> .05
Consonant	1.23	1	> .05
Testing block	3.77	1	= .052
Age group	12.89	3	< .01**
Training & Consonant	8.87	1	< .01**
Training & Testing block	0.01	1	> .05
Consonant & Testing block	0.68	1	> .05
Training & Age group	2.75	3	> .05
Consonant & Age group	13.75	3	< .01**
Testing block & Age group	5.16	3	> .05
Training & Consonant & Testing block	8.36	1	< .01**
Training & Consonant & Age group	10.70	3	= .013*
Training & Testing block & Age group	1.60	3	> .05
Consonant & Testing block & Age group	6.41	3	= .093
Training & Consonant & Testing block & Age group	8.87	3	= .031*

Table 4.2. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on closure duration of English /r/ and /l/ produced by trained and untrained Japanese speakers at pre and post tests.

Figure 4.3 displays the closure duration of English /r/ and /l/ produced by the four age groups of trained and untrained Japanese speakers at pre and post tests. For English /r/ produced by Japanese speakers in the training group, older children and adolescents shortened the closure duration from pre to post test. On the other hand, for English /l/ produced by Japanese speakers in the training group, these older children and adolescents lengthened the closure duration from pre to post test. However, these age effects on the duration change cannot be found from either English /r/ or /l/ production in the control groups, although younger children in the control group seemed to lengthen the closure duration of English /r/. These results suggest that older children and adolescents in the training group appeared to improve the distinction of the English /r/-/l/ contrast with closure duration, compared to those age groups who had no training.

As shown in Table 4.2, the linear mixed effects model demonstrated that there was a significant 4-way interaction of training, consonant, testing block and age group. This suggests that training affected the closure duration distinction for the English /r/-/l/ contrast differently by age group. Planned orthogonal contrasts demonstrated that training effects for younger children, older children and adolescents were significantly different from those for adults, $b = -1.48$, SE = .52, *pMCMC* < .01. Although there was a significant difference in the training effects between age groups, post-hoc analyses for each age group demonstrated that training affected only adolescents improving the distinction of the English /r/-/l/ contrasts with closure duration. The 3-way interaction of training, consonant and testing block was marginally significant for adolescents, $\chi^2(1) = 3.46$, $p = .063$. Although there was also a significant 3-way interaction for younger children, $\chi^2(1) = 10.27$, $p < .01$, this may be attributed to younger children in control group randomly changing their productions between pre and post tests. As they had a difficulty producing English words in this task, younger children’s acoustic changes may not be reliable. Finally, there was no significant training effects on closure duration change for the English /r/-/l/ contrast in adults, $\chi^2(1) = 0.82$, $p > .05$, or older children, $\chi^2(1) = 1.66$, $p > .05$.



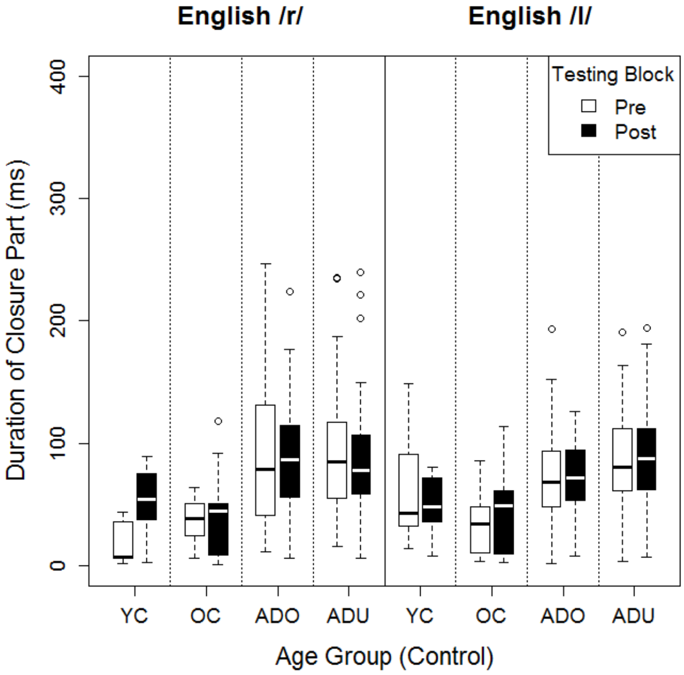


Figure 4.3. Boxplots of closure duration of English /r/ and /l/ produced by trained (top) and untrained (bottom) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

4.3.2.2. Transition duration. Figure 4.4 displays the transition duration of English /r/ and /l/ at pre and post tests by trained and untrained Japanese speakers. English /r/ should be produced with a longer transition duration, whereas English /l/ should be produced with a shorter transition duration (Kent & Read, 2002). As shown in the figure, Japanese speakers in the training group lengthened the transition duration of English /r/ and slightly shortened the transition duration of English /l/. On the other hand, Japanese speakers in the control groups did not change the transition duration for either English /r/ or English /l/ between pre and post tests. These results suggest that training had the effect of lengthening the transition duration of English /r/ and shortening the transition duration of English /l/.

Table 4.3 shows the results of the linear mixed effects model’s type-III Wald chi square test which examined the effects of each fixed factor for transition duration of the English /r/ and /l/ productions. As in the table, the linear mixed effects included fixed factors of training, consonant, testing block, age group and all possible interactions. The results demonstrated that there was a significant 3-way interaction of training, consonant and testing block, suggesting that training affected the change in transition duration for the English /r/-/l/ contrast.

Post-hoc analyses demonstrated that Japanese speakers in the training group improved the distinction of the English /r/-/l/ contrast by changing their transition durations, $\chi^2(1) = 55.25$, $p < .001$, but there was no improvement in the distinction for those in the control group, $\chi^2(1) = .40$, $p > .05$. The improvement in the distinction by trained Japanese speakers was attributed to both lengthening the transition duration of English /r/ by 13.68 ms on average, $\chi^2(1) = 56.48$ $p < .001$, and shortening the transition duration of English /l/ by 2.78 ms on average, $\chi^2(1) = 6.12$, $p = .013$.

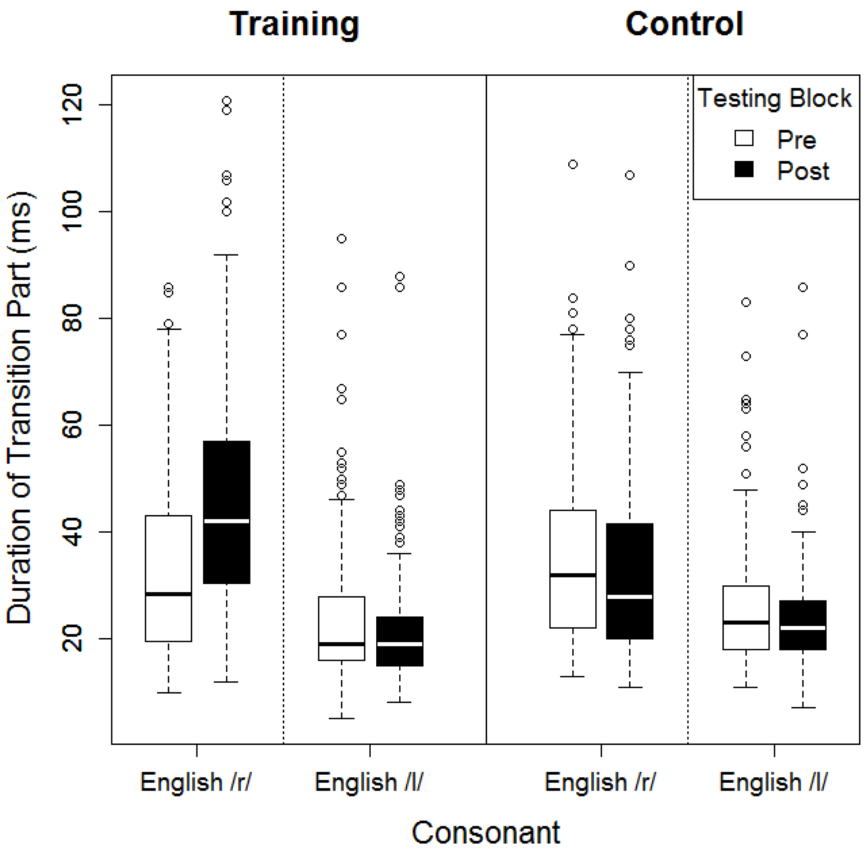


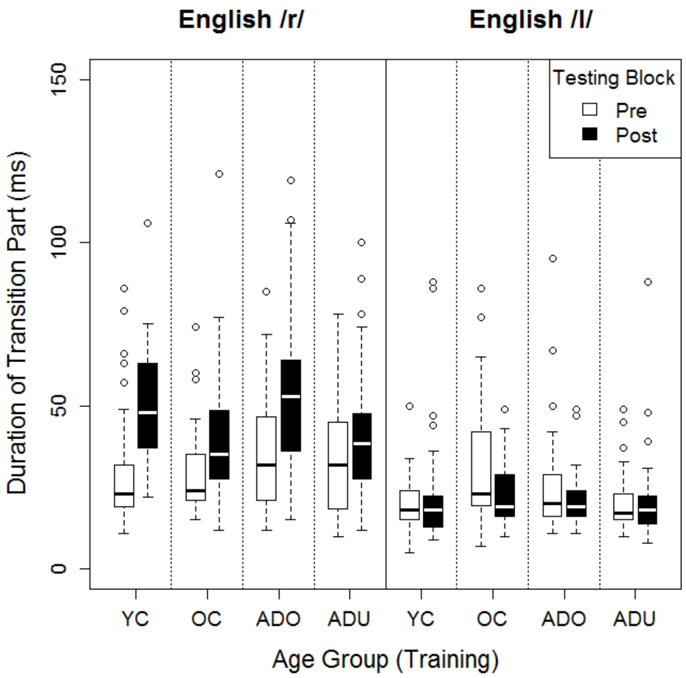
Figure 4.4. Boxplots of transition duration of English /r/ and /l/ produced by trained (left) and untrained (right) Japanese speakers at pre (white boxes) and post tests (black boxes).

Fixed Factors	χ^2	**Df**	***p***
Training	0.35	1	> .05
Consonant	50.13	1	< .001***
Testing block	0.74	1	> .05
Age group	2.45	3	> .05
Training & Consonant	20.54	1	< .001***
Training & Testing block	21.13	1	< .001***
Consonant & Testing block	18.09	1	< .001***
Training & Age group	2.09	3	> .05
Consonant & Age group	42.02	3	< .001***
Testing block & Age group	8.27	3	= .041*
Training & Consonant & Testing block	16.25	1	< .001***
Training & Consonant & Age group	7.69	3	= .013*
Training & Testing block & Age group	10.00	3	= .053
Consonant & Testing block & Age group	1.13	3	> .05
Training & Consonant & Testing block & Age group	9.08	3	= .028*

Table 4.3. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on transition duration of English /r/ and /l/ produced by trained and untrained Japanese speakers at pre and post tests.

Figure 4.5 displays the transition duration of English /r/ and /l/ by the four age groups of Japanese speakers at pre and post tests. For the training group, younger children and adolescents lengthened the transition duration of English /r/ more than older children and adults did. Older children and adolescents in the training group also shortened the transition duration of English /l/ more than the youngest and the oldest groups. These changes in the transition duration for the English /r/-/l/ contrasts cannot be found in the control group. From these results, it may be clear that training affected younger children, older children and adolescents by improving their distinction between English /r/ and /l/ by altering the transition duration more than the adults did.

As shown in Table 4.3, the linear mixed effects model for the transition duration demonstrated that there was a significant 4-way interaction of training, consonant, testing block and age group. This suggests that training affected transition durations for the English /r/-/l/ contrasts differently for the four age groups. Planned orthogonal contrasts for the 4-way interaction demonstrated that the changes in the transition duration for the English /r/-/l/ contrast by younger children, older children and adolescents were significantly different from those of adults, $b = .67$, SE = .22, $pMCMC < .01$. Post-hoc analyses for each age group demonstrated that there was a significant 3-way interaction of training, consonant and testing block for adolescents, $\chi^2(1) = 11.19$, $p < .001$, older children, $\chi^2(1) = 7.29$, $p < .01$, and younger children, $\chi^2(1) = 4.47$, $p = .034$, but not for adults, $\chi^2(1) = .02$, $p > .05$. From these results, it can be concluded that adolescents, older children and younger children improved their distinction of English /r/ and /l/ production by changing the transition duration more than adults did.



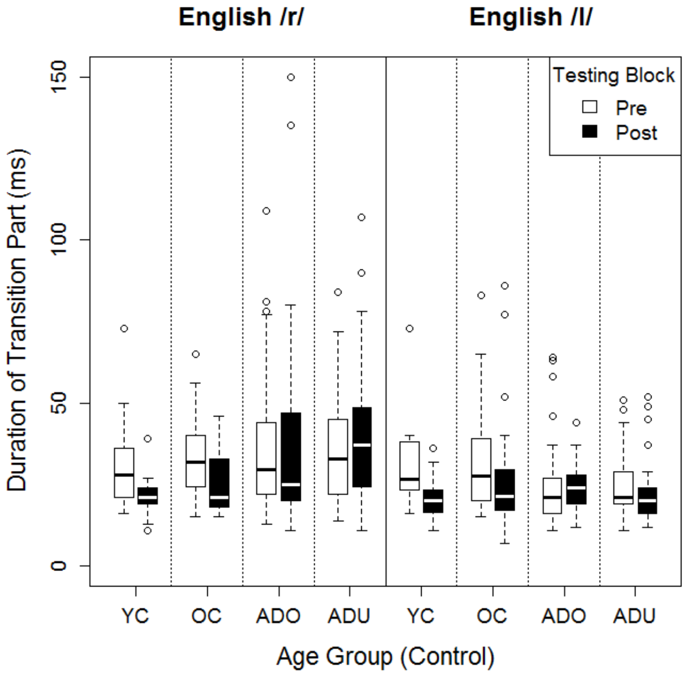


Figure 4.5. Boxplots of transition duration of English /r/ and /l/ produced by trained (top) and untrained (bottom) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

4.3.2.3. F3 frequency. Figure 4.6 displays the normalised F3 frequency for English /r/ and /l/ produced by Japanese speakers at pre and post tests. English /r/ should be produced with a low F3, while English /l/ should be produced with a high F3 (Hattori & Iverson, 2009; Kent & Read, 2002). As shown in the figure, Japanese speakers in the training group lowered the F3 for English /r/ and slightly raised the F3 for English /l/ more than Japanese speakers in the control group did.

Table 4.4 shows the results of the linear mixed effects model's type-III Wald chi square test which examined the effects of each fixed factor for F3 frequency of the English /r/ and /l/ productions. As listed in the table, the best-fitting linear mixed effects model on F3 frequency included the fixed factors of training, consonant, testing block and age group. All possible interactions were also included in the model except the 4-way interaction of training, consonant, testing block and age group, the 3-way interaction of training, consonant and age group, and the 3-way interaction of consonant, testing block and age group. These interaction factors were excluded during the process of model comparisons, suggesting that there were no significant effects of these interactions. Although there was no significant 4-way interaction, the 3-way interaction of training, consonant and testing block was significant, suggesting that the training affected the F3 change for the English /r/-/l/ contrast.

As shown in Figure 4.6, post-hoc analyses for training group demonstrated that trained Japanese speakers significantly improved their F3 distinction of the English /r/-/l/ contrast from pre to post test, $\chi^2(1) = 19.74$, $p < .001$. On the other hand, post-hoc analyses for the control group demonstrated that Japanese speakers who had no training did not significantly change their F3 distinction for the contrast from pre to post test, $\chi^2(1) = .39$, $p > .05$.

Furthermore, post-hoc analyses for each consonant of trained Japanese subjects demonstrated that the improvement in the F3 distinction for the contrast was attributed to both lowering the F3 for English /r/ by 121 Hz on average, $\chi^2(1) = 7.55$, $p < .01$, and raising the F3 for English /l/ by 119 Hz on average, $\chi^2(1) = 13.30$, $p < .001$.

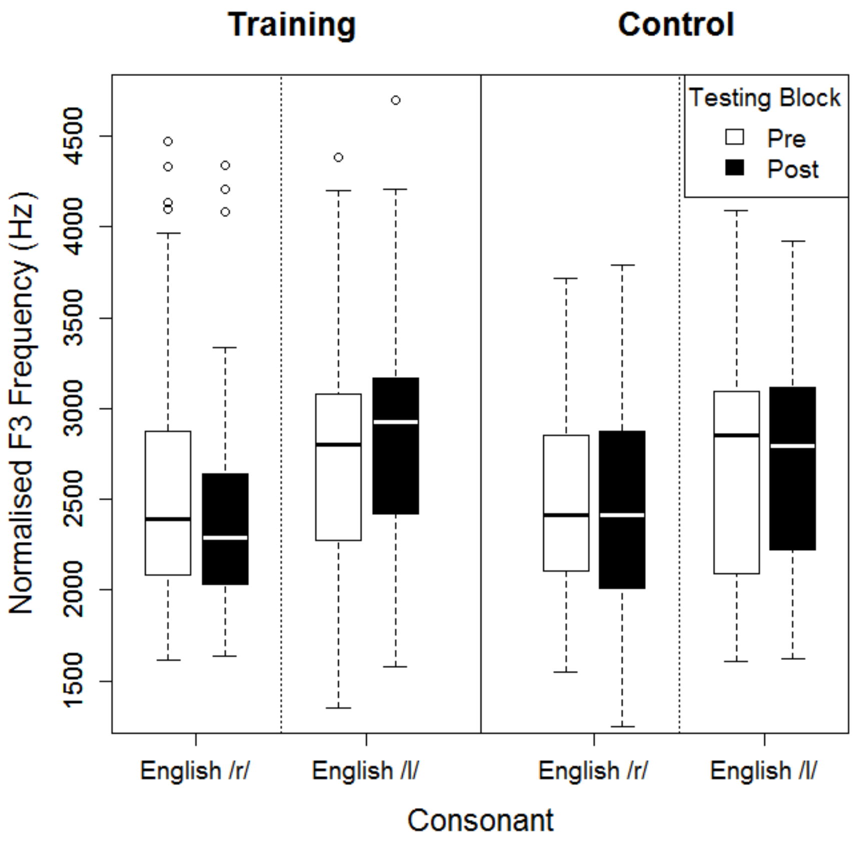
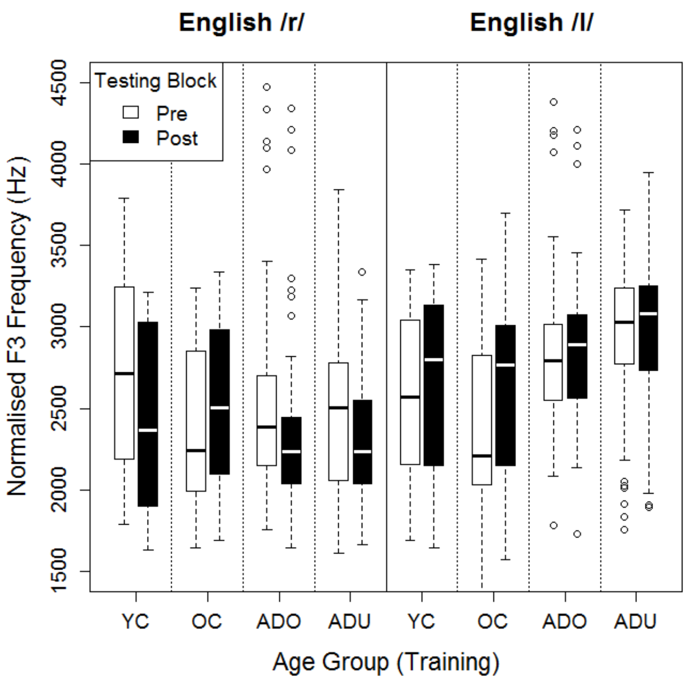


Figure 4.6. Boxplots of normalised F3 frequencies of English /r/ and /l/ production by trained (left) and untrained (right) Japanese speakers at pre (white boxes) and post tests (black).

Fixed Factors	χ^2	**Df**	***p***
Training	0.19	1	> .05
Consonant	5.92	1	= .015*
Testing block	0.04	1	> .05
Age group	4.48	3	> .05
Training & Consonant	4.69	1	= .030*
Training & Testing block	0.28	1	> .05
Consonant & Testing block	9.65	1	< .01**
Training & Age group	0.17	3	> .05
Consonant & Age group	85.35	3	< .001***
Testing block & Age group	2.18	3	> .05
Training & Consonant & Testing block	4.48	1	= .034*
Training & Testing block & Age group	8.75	3	= .033*

Table 4.4. Type-III analysis-of-variance table based on Wald chi-square tests for a linear mixed effects model on normalised F3 frequencies of English /r/ and /l/ produced by trained and untrained Japanese speakers at pre and post tests.

Although there was no signficant 4-way interaction in the linear mixed effects model for F3 frequency, there seemed to be a difference in the direction for the F3 change by age group. Figure 4.7 displays the normalised F3 frequency for English /r/ and /l/ produced by the four age groups of Japanese speakers at pre and post tests. Although no subjects in any age group changed F3 for the English /r/-/l/ contrast in the control group, younger children, adolescents and adults in the training groups lowered their F3 for English /r/ by 191 Hz, 144 Hz and 172 Hz on average. Older children in the training group, on the other hand, raised F3 for English /r/ by 66 Hz on average. In addition, while the F3 for English /l/ was raised by adults, *M* = 98 Hz, adolescents, *M* = 77 Hz, and younger children, *M* = 71 Hz, to small extents, older children raised F3 for English /l/ more than the other three age groups, *M* = 275 Hz. It may be possible that these results demonstrated that some developmental factor during childhood affects the direction of the F3 change for the English /r/ and /l/ productions.



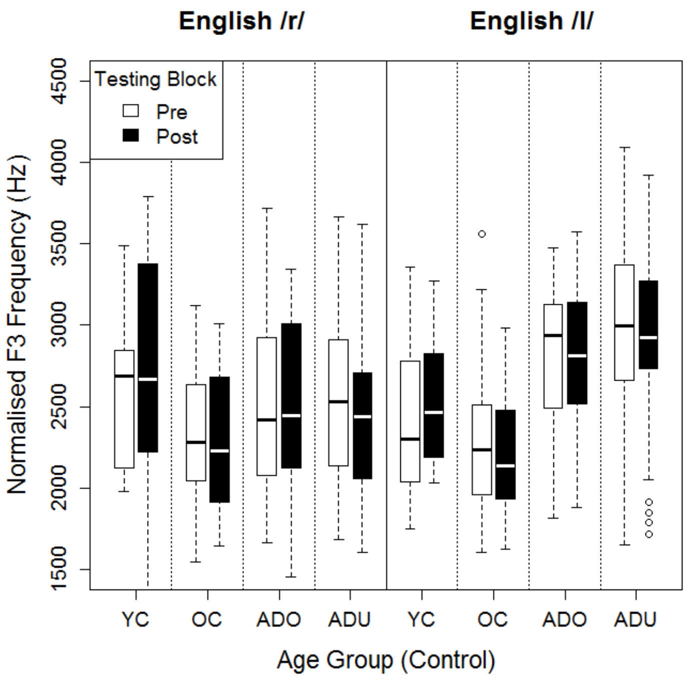


Figure 4.7. Boxplots of normalised F3 frequencies of English /r/ and /l/ produced by trained (top) and untrained (bottom) Japanese speakers broken down by the age groups of younger children (YC), older children (OC), adolescents (ADO) and adults (ADU), at pre (white boxes) and post tests (black boxes).

4.3.3. Links between Perception and Production Improvement

Table 4.5 shows results of 15 correlation tests between perception and production ability changes for English /r/ and /l/ for all trained Japanese speakers. The change in each testing variable was calculated by subtracting the mean values of each variable at pre test for each subject from those values at post test. Pearson's *r* was used for the correlation tests. As shown in the table, changes in perceptual identification accuracy were tested in correlations with changes in all production variables. In addition, changes in perceptual F3 sensitivity at the English /r/-/l/ boundary and F3 sensitivity within the English /r/ category were tested in correlations with changes in production identifiability, F3 for English /r/ production and F3 for English /l/ production. Finally, changes in perceptual F2 sensitivity at the English /r/-/l/ boundary and perceptual category discrimination accuracy were tested in correlations with production identifiability. Since 15 correlations were tested, *p* values were adjusted with Holm's method (Aickin & Gensler, 1996; Holm, 1979).

Table 4.5 shows the results of the 15 correlation tests. As shown in the table, change in perceptual identification accuracy by Japanese speakers was significantly correlated to change in production identifiability by native English speakers, $r = .37$, df = 57, $p = .048$. This result suggests that Japanese speakers who improved perceptual identification accuracy also improved production identifiability.

There was also a significant positive correlation between perceptual F3 sensitivity at the English /r/-/l/ boundary and production identifiability, $r = .39$, df = 57, $p = .035$. This suggests that Japanese speakers who became sensitive to the primary acoustic cue for the English /r/-/l/ contrast in perception improved their production identifiability. These significant correlations in improvements of perception and production suggest that there is a link between perception and production in L2.

***p* < .05 * (Holm-adjusted *p*)**

	Prod ID	**Prod R cldur**	**Prod R trdur**	**Prod R F3**	**Prod L cldur**	**Prod L trdur**	**Prod L F3**
Perc ID	**0.37***	-0.16	0.16	0.19	-0.09	0.17	0.29
F3 sen bnd	**0.39***			-0.12			0.16
F2 sen bnd	0.02						
F3 sen r	0.08			0.08			-0.06
Perc CD	0.31						

Table 4.5. Correlation matrix of changes from pre-training to post-training test in perceptual identification (Perc ID), perceptual F3 sensitivity at the English /r/-/l/ boundary (F3 sen bnd), perceptual F2 sensitivity at the English /r/-/l/ boundary (F2 sen bnd), perceptual F3 sensitivity within the English /r/ category (F3 sen r), perceptual category discrimination (Perc CD), production identifiability (Prod ID), closure duration of English /r/ production (Prod R cldur), transition duration of English /r/ production (Prod R trdur), F3 of English /r/ production (Prod R F3), closure duration of English /l/ production (Prod L cldur), transition duration of English /l/ production (Prod L trdur), F3 of English /l/ production (Prod L F3).

4.4. Discussion

One of the main findings of the present study is that Japanese speakers improved their English /r/-/l/ production through perceptual training. Not only did they significantly improve their production identifiability, they also made a distinction between English /r/ and /l/ in all acoustic measurements of closure duration, transition duration and F3 frequency. For temporal cues, Japanese speakers lengthened the transition duration of English /r/ and the closure duration of English /l/, and shortened the transition duration of English /l/. For spectral cues, Japanese speakers lowered their F3 for English /r/ and raised their F3 for English /l/.

Another main finding is that age affects production improvement. All age groups except younger children improved their production identifiability, and adolescents improved their identifiability the most among the four age groups. In terms of acoustic change, older children and adolescents improved their distinction of English /r/-/l/ production more than adults did in closure and transition duration, although younger children improved their distinction with transition duration to a similar amount as did older children and adolescents. These results of age effects on production identifiability and acoustic changes may be due to perceptual learning. Older children and adolescents improved both phonetic and phonological perception more than adults did. This improvement may have helped these older children and adolescents improve their production identifiability and acoustic realisations of the English /r/-/l/ contrast.

Although there was no significant difference in the improvement of the distinction for the English /r/-/l/ contrast with F3 frequency by age group, there remains one question: why did older children not lower their F3 frequency for the English /r/ production? Developmental studies would suggest that older children may have had a difficulty in their articulatory command. Since English-speaking children under 8 years old tend to have difficulties in pronouncing English /r/ (Smit, Hand, Freilinger, Bernthal, & Bird, 1990), it is likely that all children including Japanese children would have difficulty pronouncing English /r/. Although Aoyama et al. (2004) demonstrated that Japanese children aged around 10 years old were able to improve their production of English /r/ more than English /l/ one year after their arrival in the US, the Japanese-speaking environment may have contributed to a delay in the improvement of articulatory commands for English /r/. However, since younger children lowered their F3 for English /r/, older children should have been able to learn the rhoticity of English /r/ at the developmental stage. Somehow, instead of learning the rhoticity of English /r/, older children here improved their F3 of English /l/ more than the other three age groups. English /l/ should have been more difficult to improve than English /r/ according to SLM (Flege, 1995), but older children improved F3 for English /l/ more than for English /r/. It is not possible to identify what caused this improvement difference between older children and younger children from the current results, but this implies that another developmental factor might have affected learning L2 phonemes.

Following these two findings of production improvement and age effects, there seems to be a connection in improvement between perception and production. Correlation tests also demonstrated that the production identifiability improvement was positively correlated with perceptual identification improvement and with improvement of F3 sensitivity at the English /r/-/l/ boundary. Although two significant correlations out of 15 correlation tests between perception and production improvements may not strongly suggest that there is a common underlying ability governing all perception and production abilities, it may be unlikely that there are two independent abilities for perception and production.

These results are not consistent with the findings of previous studies. Bradlow et al. (1997) demonstrated that there was no significant correlation between perceptual identification improvement and production identifiability improvement. This may be because of analysis differences. Bradlow et al. (1997) tested the relation in improvements between perception and production with Spearman’s rank-order correlation tests with only 11 subjects. If the rank-order was different between perception and production improvement for a few subjects, the results would not show a significant correlation. On the other hand, the present study tested the correlation by 59 Japanese speakers with Pearson's *r*, so the results here were more robust. The present study also included many younger Japanese speakers than did the previous study, and the subjects' differences may also have contributed to the inconsistencies.

If there is a common underlying ability linking speech perception and production, then there is an unsolved question of why the articulation training by Hattori (2009) did not improve perceptual ability at all. Although his articulatory training improved adult Japanese speakers' production to native levels, the articulatory training did not transfer increased skill to the perception domain. This may be because his articulatory training did not fundamentally change the underlying mental representations of English /r/ and /l/. Kuhl et al. (2006, 2008) suggest that infants develop the connection between perception and production in L1, while comparing their own production with phonetic representations stored in their memories. This monitoring process may affect the internal representations of L1 phonetic units. If this is also the case for L2, Japanese speakers may need to compare their own production of English /r/ and /l/ with their internal representations of those phonemes when they practice pronouncing English /r/ and /l/. However, Japanese subjects in Hattori (2009) practiced pronunciation of English /r/ and /l/ by looking at spectrograms on a screen and at their articulatory movements with mirrors. Japanese subjects might have focused on these visual aids, but they might not have compared their own production of English /r/ and /l/ with their mental representations. This may provide one explanation for why Hattori's (2009) articulatory training did not lead to a perceptual improvement.

On the other hand, perceptual training seems to affect underlying mental representations of English /r/ and /l/. Although the perceptual levels affected by training may be different between age groups due to the effects of brain plasticity and L1 interference, many Japanese subjects may have improved their internal representations of English /r/ and /l/ and acquired auditory-articulatory (perceptual-motor) mappings (Callan et al., 2003). This fundamental improvement may have contributed to a production improvement in identifiability and acoustic realisations.

As described in the introduction to this chapter, L2 speech learning does not correspond to any of the three theories in L1 speech perception, namely motor theory (Liberman et al., 1967; Liberman & Mattingly, 1985), direct realist theory (Fowler, 1996) or general auditory and learning approaches to speech perception. Motor theory and direct realist theory propose that the articulatory gestures are the objects people perceive as speech sounds, which provides an explanation for why perceptual training improves production identifiability without any production instruction (Bradlow et al., 1997; Iverson et al., 2011). However, these theories cannot explain why articulatory training does not improve perceptual identification accuracy (Hattori, 2009). General auditory and learning approaches propose that speech sounds are perceived as acoustic signals, so that the independence of perception and production as reported in Hattori (2009) can be explained. However, this theory does not explain why perceptual training improves production identifiability, unless subjects actually produce target L2 phonemes and perceive their own productions.

To explain the link between speech perception and production in L2, it may be more important to consider how leaners acquire L2 speech rather than considering the objects they perceive. If Japanese speakers improve their underlying mental representations of English /r/ and /l/ by training, then it might be that they transfer the perceptual learning to the production domain and simultaneously transfer the production learning to the perception domain. Future studies should be conducted to investigate what factors are essential to make training affect both perception and production domains. Examining changes in brain activities of Japanese speakers after perceptual and production training for the English /r/-/l/ contrast may supplement the interpretation of behavioural results.

Finally, the results of the present study suggest an important characteristic of production identifiability. Since the production identifiability was based on native English speakers' phonological judgement of Japanese speakers' English /r/-/l/ productions, any changes which distinguish English /r/ and /l/ in production would affect this identifiability. In other words, production identifiability can be improved by any acoustic changes moving production in the direction of native speakers' production, even if the improved acoustic cues for production are different between subjects. Since there was no significant correlation between perceptual improvement and any acoustic changes in production, it may be plausible to consider that the perceptual improvement is not related to any common acoustic realisations in production for all subjects. It may instead be related to different acoustic realisations for each individual. This might explain why there were significant correlations between production identifiability improvement and perceptual improvements (i.e., identification, F3 sensitivity at the English /r/-/l/ boundary).

In conclusion, Japanese speakers improved their production of English /r/ and /l/ through perceptual training. They improved both their production identifiability and their acoustic distinction for the English /r/-/l/ contrast in production, and training effects were different between age groups. Older children and adolescents improved the distinction for the English /r/-/l/ contrast in closure duration more than younger children and adults did, and they also improved their distinction of transition duration for the contrast more than adults did. Following these production improvements through perceptual training, it is possible that there is an underlying common ability in improvement between perception and production in L2, although the acoustic realisations which result from the common ability seems to be different between individuals. Since more questions were raised concerning age effects on production improvement, future studies should be conducted to research the effects on production learning from developmental factors throughout childhood.

Chapter 5: General discussion

The aim of this PhD research was to contribute to the clarification of the learning mechanisms and age effects for Japanese speakers learning the English /r/-/l/ contrast through perceptual training. The first study investigated the effects of two different perceptual training methods contrasting identification and discrimination for Japanese adults. It was hypothesised that identification training would improve phonological levels of perception, reflected in identification accuracy, and that discrimination training would improve phonetic processing for the English /r/-/l/ contrast. The improvement of the phonetic processing is essentially the process of acquiring higher sensitivity to the primary acoustic cue (i.e., F3 at the English /r/-/l/ boundary) and lower sensitivity to irrelevant acoustic cues (i.e., F2 at the English /r/-/l/ boundary and F3 within the English /r/ category). However, the results demonstrated that both identification and discrimination training improved identification of the English /r/-/l/ contrast, although identification training improved identification ability more than discrimination training did. Japanese speakers also improved their sensitivities to F2 and F3 at the English /r/-/l/ boundary through both types of training. Since both training methods had similar effects, there was no additive effect from the combination of the two different training techniques.

One possible conclusion was that Japanese adults may have simply learned how to cope with a high variability of stimuli over the course of both training programmes, as Iverson et al. (2005) suggested. That is, Japanese adults may learn a strategy for identifying English /r/ and /l/ using their L1 alveolar tap [ɾ]. If the stimulus sounds similar to Japanese /r/, they would identify it as English /l/; if the stimulus sounds different from Japanese /r/, they would identify it as English /r/ (Iverson et al., 2005). Supporting this argument that Japanese speakers did not create new phonemic categories of English /r/ and /l/ in their phonological space, a previous perceptual training study with fMRI measurements demonstrated that adult Japanese speakers' brain activity for the /r/-/l/ contrast does not resemble the brain activity of an easy L2 contrast of /b/-/g/ after one month perceptual training (Callan et al., 2003). Although increases and shifts of brain activity were found in this previous study, this seemed to be attributed to general processing strategies, task-related experience, and/or session-related effects (Callan et al., 2003). Callan et al. (2003) also argued that the identification of the English /r/-/l/ contrast is facilitated in part by the acquisition of auditory-articulatory (perceptual-motor) mappings, which can be reflected in speech production acquisition (Hickok & Poeppel, 2000). This perceptual-motor mapping may have also occurred in the first study of this PhD research, and it may be why adult Japanese speakers lowered F3 for English /r/.

The second study examined age effects on learning the perception of English /r/ and /l/. The sensitive/optimal period hypothesis (Oyama, 1976; Werker & Tees, 2005) and the Speech Learning Model (Flege, 1995, 1999, 2002) led me to infer that younger learners would have some advantage in learning L2 phonemes. Supporting this hypothesis, this study demonstrated that older children and adolescents can improve phonetic and phonological perception of the contrast being investigated better than is the case for adults. Not only did older children aged 8-12 years old improve their word-initial English /r/-/l/ identification ability by more than 26%, but they also improved their category discrimination accuracy more than adults did. Adolescents aged 15-18 years old also improved their identification ability more than adults did, and they improved their sensitivity to the primary acoustic cue for the English /r/-/l/ contrast as well as their category discrimination accuracy.

These findings in the second study support both the sensitive/optimal period hypothesis (Oyama, 1976; Werker & Tees, 2005) and the Speech Learning Model (Flege, 1995, 1999, 2002). For instance, the reasons why older children improved their perception of English /r/ and /l/ at both phonetic and phonological levels may be that their brain plasticity is still flexible and that they have less L1 interference. Since establishing L1 phonetic units is not completed by adolescence (Flege & Eefting, 1986; Hazan & Barrett, 2000; Lee, Potamianos, & Narayanan, 1999), L2 phonetic learning may not be inhibited by the strong L1 phonetic attractions (Flege, 1995), although they may have disadvantages from their immaturity of sensory system and cognition. Adolescents also improved both phonetic and phonological perception more than adults did. This may be attributed to their relatively greater brain plasticity and their developed cognition. On the other hand, adults did not improve either phonetic or phonological perception to the same level as either older children or adolescents. This may be because they have the disadvantage of their L1 interference and their relatively fossilised brain plasticity. Thus, there may be both effects of brain plasticity and L1 interference on learning L2 speech perception.

The final study investigated how the production of English /r/ and /l/ were improved through perceptual training. Both production identifiability and acoustic distinction of English /r/ and /l/ with closure duration, transition duration and F3 frequency were improved by Japanese speakers through perceptual training. In terms of age effects, production identifiability was improved by all age groups except younger children, and older children and adolescents were better at learning the distinction of English /r/ and /l/ with closure and transition duration than adults. Following these production improvements through perceptual training, it may be plausible to consider that there is a relationship between perception and production in L2. Older children and adolescents improved closure duration and transition duration realisations for the English /r/-/l/ contrast more than adults did. This may be attributed to their higher perceptual improvement. Moreover, correlation tests demonstrated that there were significant relations in improvement between perception and production domains. This also suggests that there is a common underlying ability linking perception and production improvements, although the acoustic realisations that result from this common ability may be different between individuals.

However, this conclusion does not correspond with the results of Hattori (2009), and so further explanation is needed. Hattori (2009) found that articulation training improved Japanese speakers' production of English /r/ and /l/ to native-levels, but his articulation training did not transfer to the perception domain. This may be because his training did not affect the underlying mental representations of English /r/ and /l/. Hattori's articulatory training method focused on teaching Japanese adults places and manners of articulation and formant movements on spectrograms. In other words, Japanese speakers may not have had to listen carefully to their own pronunciation but instead needed to monitor the relation between their articulatory movement and visual display of spectrogram formants. As L1 developmental studies suggest, vocalising and imitating L1 phonetic units should accompany listening to the speakers’ own production. Infants compare their own productions with phonetic representations stored in their memory, and this process may accelerate warping universal phonetic capacity to language-specific perception and production (Kuhl et al., 2006, 2008). If this is also the case for L2, Japanese subjects in Hattori (2009) would have needed to monitor their own pronunciations, comparing them to their mental representations of English /r/ and /l/. However, the training being investigated did not require them to do this. Consequently, they may have learned a pronunciation technique without affecting their internal representations of English /r/ and /l/. From this argument, it can be concluded that only when Japanese speakers improve their underlying mental representations of English /r/ and /l/ would training transfer learning in a domain to the other domain.

In regard to the present perceptual training programme, younger learners seemed to improve their underlying mental representations of English /r/ and /l/. For example, older children and adolescents improved their perception for the English /r/-/l/ contrast more than adults. This perceptual improvement may be reflected by the change of internal representations of English /r/ and /l/, and this fundamental learning in perception may cause their higher production improvement in using closure and transition duration cues for the English /r/-/l/ contrast. However, there are still unresolved questions regarding why the older children did not lower F3 for English /r/ production, and why adults improved F3 distinction for the contrasts as much as other age groups did. Future studies are necessary to clarify the learning mechanism and the age effects in more detail.

Finally, there remains the most important question of whether younger Japanese speakers formed new phonemic categories in their phonological space. Adolescents showed “acquired distinctiveness” in perception after training, as they increased sensitivity to F3 at the /r/-/l/ boundary (Heeren & Schouten, 2010; Liberman et al., 1961; Pisoni, 1991). Despite this improvement in L2 phonetic perception, it is still unclear if new L2 phonemic categories were created in the learners’ phonological space. Future studies should be conducted to examine how younger learners acquire the perception of the English /r/-/l/ contrast: whether they merge both English /r/ and /l/ categories in their Japanese phonemic categories (“category assimilation” in the SLM terminology: Flege, 2002), or they dissimilate the English /r/ and /l/ categories from Japanese phonemes (“category dissimilation” in the SLM terminology: Flege, 2002). In addition, testing brain responses may demonstrate how Japanese speakers acquire the perception of the contrast. Although the interpretation of brain activities is still difficult, brain responses supplement the interpretation of behavioural data (Callan et al., 2003; Zhang et al., 2009). Such future research may contribute to the development of the L2 education system.

In conclusion, this PhD research investigated the learning mechanism and age effects for Japanese speakers learning the English /r/-/l/ contrast through perceptual training. The learning mechanism has been gradually clarified through the history of studies in L2 speech perception and production, but in gaining greater understanding of the mechanism, we end up raising more questions. This research will never end, because it may not be possible to conclusively prove all phenomena in L2 learning. Nevertheless, further research will edge science towards a clearer understanding and take us closer to at least a working explanation. Language learning is a highly sophisticated function and needs to be researched with a variety of factors both from nature and nurture. Since language is the single most utilised tool for communication between people on earth, we should keep trying to solve the mystery of learning language for the unity of the world.

References

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *American journal of public health*, *86*(5), 726–728.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. doi:10.1016/S0095-4470(03)00036-6

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Berman, S., & Friedman, D. (1995). The development of selective attention as reflected by event-related brain potentials. *Journal of Experimental Child Psychology*, *59*, 1–31.

Best, C. T. (1994a). Learning to perceive the sound pattern of English. In C. Rovee-Collie & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 8, pp. 217–304). Hillsdale, NJ: Ablex Publishers.

Best, C. T. (1994b). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sound to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

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

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., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, *26*(2), 211–252.

Boysson-Bardies, B. de. (1993). Ontogeny of language-specific syllabic productions. In Benedicte de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & M. John (Eds.), *Developmental neurocognition: speech and face processing in the first year of life* (pp. 353–364). doi:10.1007/978-94-015-8234-6_29

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 & psychophysics, 61*(5), 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development*, *31*(1), 1–14.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*(1), 113–124. doi:10.1016/S1053-8119(03)00020-X

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. doi:10.1177/0267658312443651

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341–347. doi:10.3758/BF03211210

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306. doi:10.1126/science.171.3968.303

Endo, M. (1991). Phonological awareness of Japanese young children and learning to read and write yoo-on (small-sized kana character). *Japanese Journal of Educational Psychology*, *39*, 448–454.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed., p. 127). New York: Harper & Row.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

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

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101–131). Mahwah, NJ: Lawrence Erlbaum Associates.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of henning wode* (pp. 217–244). Trier, Germany: Wissenschaftlicher Verlag Trier.

Flege, J. E. (2003). A method for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in Clinical Linguistics* (pp. 19–44). Padova: UniPress.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*, 155–171.

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers of English. *Journal of Phonetics*, *15*, 67–83.

Flege, J. E., & MacKay, I. R. A. (2010). What accounts for “age” effects on overall degree foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 65–82). Bern, Switzerland: Peter Lang.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*(1), 78–104. doi:10.1006/jmla.1999.2638

Fowler, C. a. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730–1741.

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, *35*(5), 997–1010.

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage*, *21*(2), 494–506. doi:10.1016/j.neuroimage.2003.09.071

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, *9*, 317–323.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults’ perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711–2724.

Halliday, L. F., Taylor, J. L., Edmondson-Jones, a M., & Moore, D. R. (2008). Frequency discrimination learning in children. *The Journal of the Acoustical Society of America*, *123*(6), 4393–4402. doi:10.1121/1.2890749

Hattori, K. (2009). *Perception and production of English /r/-/l/ by adult Japanese speakers* (Doctoral thesis, University College London, United Kingdom). University College London. Retrieved from http://discovery.ucl.ac.uk/19204/1/19204.pdf

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*(1), 469–79. doi:10.1121/1.3021295

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396. doi:10.1006/jpho.2000.0121

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*(3), 360–378. doi:10.1016/j.specom.2005.04.007

Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: how sensitivity changes along acoustic dimensions that contrast phoneme categories. *The Journal of the Acoustical Society of America*, *124*(4), 2291–2302. doi:10.1121/1.2967472

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish /t-tː/ distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics*, *38*(4), 594–603. doi:10.1016/j.wocn.2010.08.005

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, *4*(4), 131–138.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, *119*(5), 3072–3084. doi:10.1121/1.2184289

Holm, S. (1979). Board of the Foundation of the Scandinavian Journal of Statistics, *6*(2), 65–70.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*(4), 571–584. doi:10.1016/j.wocn.2011.03.003

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 speakers. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1305–1316. doi:10.1037/0096-1523.34.5.1305

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. doi:10.1121/1.3148196

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267. doi:10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. doi:10.1016/S0

Iverson, P., Pinet, M., & Evans, B. G. (2011). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 1–16. doi:10.1017/S0142716411000300

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215.

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

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech* (Second edi.). New York: Thomson Learning.

Komaki, R., Akahane-Yamada, R., & Choi, Y. (1999). Effects of native language on the perception of American English/r/and/l: Cross-language comparison between Korean and Japanese. *Technical report of IEICE*, *SP 99-45*,, 39–46.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11850–7. doi:10.1073/pnas.97.22.11850

Kuhl, P. K. (2011). Brain mechanisms underlying the critical period for language: Linking theory and practice. In A. M. Battro, S. Dehaene, & W. J. Singer (Eds.), *Human neuroplasticity and education* (Vol. 117, pp. 33–59). Vatican City: The Pontifical Academy of Sciences.

Kuhl, P. K., Conboy, B. T., Coffey-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 Transactions of the Royal Society B: Biological Sciences*, *363*, 979–1000. doi:10.1098/rstb.2007.2154

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. doi:10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–9101. doi:10.1073/pnas.1532872100

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.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468.

Lenneberg, E. H. (1967). *The biological foundations of language*. *Exploring the mind. Fall*. New York: Wiley.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379–388.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

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. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Lotto, A., Sato, M., & Diehl, R. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. C–181–C–186). Cambridge, MA: MIT Research Laboratory in Electronics.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*(04), 369–390. doi:10.1017/S0142716400009796

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition*, *24*, 65–92. doi:10.1016/0010-0277(86)90005-3

Matsumoto, H. (2004). Development of phonemic awareness and the beginning process of learning to read in Japanese speaking children. *Reseach and Clinical Center for Child Development Annual Report*, *26*, 67–77.

Matsumoto, H. (2006). Relationship between phonemic awareness and syllabic awareness during the initial stage of learning to read in Japanese children. *Reseach and Clinical Center for Child Development Annual Report*, *28,* 77–87.

Mayo, C., & Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America*, *115*(6), 3184–3194. doi:10.1121/1.1738838

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of and by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, *20*(3), 351–382.

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America*, *101*(4), 2253–2266.

Nittrouer, S., Miller, M. E., Crowther, C. S., & Manhart, M. J. (2000). The effect of segmental order on fricative labeling by children and adults. *Perception & Psychophysics*, *62*(2), 266–284.

Ogawa, A., & Tanemura, J. (2001). Contributions of phoneme operating ability in acquisition of kana reading. *The Japan Journal of Logopedics and Phoniatrics, 42*(3), 220–226.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283. doi:10.1007/BF01067377

Pisoni, D. B. (1991). Comment: Modes of processing speech and nonspeech signals. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 225–238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696. doi:10.1121/1.2161427

Sereno, J., & Jongman, A. (2005). Language learning in adults and children. In *Proceedings of Workshop on Plasticity in Speech Perception* (p. 142). London, UK.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*, 243–261.

Shimizu, K., & Dantsuji, M. (1983). A Study on the Perception of/r/and/l/in Natural and Synthetic Speech Sounds. *Studia phonologica*, *15*, 1–14.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The IOWA articilation norms project and its nebraska replication. *Journal of Speech and Hearing Disorders*, *55*, 779–798.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145.

Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *The Journal of the Acoustical Society of America*, *109*(3), 1173–1180. doi:10.1121/1.1349428

Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view*. (Doctoral dissertation, University of California, Irvine). Retrieved from http://search.proquest.com/dissertations/docview/304042958/fulltextPDF/14286C787BA778B2C85/1?accountid=14511

Thomas, E., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. Retrieved December 05, 2013, from http://ncslaap.lib.ncsu.edu/tools/norm/

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. *The Journal of the Acoustical Society of America*, *120*(4), 2285–2294. doi:10.1121/1.2338290

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, UK: Cambridge University Press.

Wang, Y., & Kuhl, P. K. (2003). Evaluating the “Critical Period” hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress on Phonetic Sciences* (pp. 1537–1540). Barcelona: Casual Productions.

Wang, Y., Sereno, J. a, Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience, 15*(7), 1019–1027. doi:10.1162/089892903770007407

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*(5), 672–683. doi:10.1037/0012-1649.24.5.672

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. doi:10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychology*, *46*(3), 233–251. doi:10.1002/dev.20060

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. doi:10.3758/s13423-011-0179-5

Yamada, R. A. (1995). Age and acquisition of second language speech sounds perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 305–320). York Press.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & psychophysics*, *52*(4), 376–392.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *NeuroImage, 46*(1), 226–240. doi:10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–20. doi:10.1016/j.neuroimage.2005.02.040

Zhang, Y., & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, *10*(02), 147–160. doi:10.1017/S1366728907002908

Appendix A

Natural recordings of English /r/-/l/ minimal-pair words used for training in Chapter 2 and 3

/r/ words	/l/ words	/r/ words	/l/ words	/r/ words	/l/ words
rack	lack	rear	leer	roves	loaves
rad	lad	rent	lent	rob	lob
rag	lag	rice	lice	robe	lobe
raid	laid	rick	lick	rock	lock
rake	lake	rid	lid	wrong	long
ram	lamb	rise	lies	rook	look
rain	lane	rife	life	room	loom
rank	lank	rift	lift	ross	loss
rate	late	right	light	rot	lot
raft	laughed	rim	limb	rowed	loud
roars	laws	rhyme	lime	rout	lout
ray	lay	rhine	line	row	low
raise	laze	rind	lined	rose	lows
reach	leach	rink	link	rump	lump
reef	leaf	rip	lip	rush	lush
reek	leak	writ	lit	rust	lust
red	led	roan	loan		

These natural recordings of 100 minimal-pair words were used for the training tasks of identification, auditory discrimination with natural stimuli and category discrimination in Chapter 2. These were also used for identification training in Chapter 3. These natural stimuli used for training were produced by 10 SSBE speakers.

Appendix B

English /r/-/l/ minimal-pair words for identification tests in Chapter 2 and 3

Word-initial position		**Word-medial position**		**Consonant Cluster position**	
/r/ words	**/l/ words**	**/r/ words**	**/l/ words**	**/r/ words**	**/l/ words**
race	lace	arouse	allows	brand	bland
ramp	lamp	arrive	alive	broom	bloom
raps	lapse	bereave	believe	brunt	blunt
raw	law	berries	bellies	brush	blush
reap	leap	boring	bawling	cramp	clamp
rest	lest	coring	calling	crime	climb
ride	lied	correct	collect	crowd	cloud
road	load	erect	elect	frame	flame
roam	loam	fairy	fairly	fresh	flesh
roared	lord	farrow	fallow	froze	flows
root	loot	horror	holler	fruit	flute
rope	lope	marrow	mallow	grass	glass
rude	lewd	mirror	miller	graze	glaze
rug	lug	parrot	palate	grew	glue
rune	loon	pirate	pilot	grow	glow
rung	lung	poring	palling	praise	plays
ruse	lose	starring	starling	prank	plank
wrap	lap	tarry	tally	prod	plod
wrens	lens	terror	teller	spray	splay
wrist	list	whirring	whirling	sprint	splint

These minimal-pair words were produced by two English speakers (one female, one male) who were not used for training stimuli.

Appendix C

Signal processed minimal pairs used for auditory discrimination training in Chapter 2

Session number (speaker)	**Minimal-pairs**	
	/r/ words	**/l/ words**
Session 1 (EG: female)	ray	lay
	rear	leer
	wrong	long
	row	low
Session 2 (AJ: male)	ram	lamb
	ray	lay
	rear	leer
	rim	limb
Session 3 (MK: female)	ray	lay
	rear	leer
	rim	limb
	wrong	long
Session 4 (TG: male)	ram	lamb
	ray	lay
	rim	limb
	row	low
Session 5 (RW: female)	ram	lamb
	ray	lay
	rear	leer
	row	low

There were close and distant pairs for each minimal pair (see 2.2.3.3. Speech Processing, pp. 37-39).

Appendix D

Trial lists including 24 word-initial English /r/-/l/ minimal-pair words for category discrimination test in Chapter 2

Trial number	Stimulus 1	Stimulus 2	Stimulus 3
1	lest	root	wrap
2	rest	loot	wrap
3	rest	root	lap
4	rest	loot	lap
5	lest	root	lap
6	lest	loot	wrap
7	lope	raw	rung
8	rope	law	rung
9	rope	raw	lung
10	rope	law	lung
11	lope	raw	lung
12	lope	law	rung
13	lied	wrist	race
14	ride	list	race
15	ride	wrist	lace
16	ride	list	lace
17	lied	wrist	lace
18	lied	list	race
19	lens	roam	roared
20	wrens	loam	roared
21	wrens	roam	lord
22	wrens	loam	lord
23	lens	roam	lord
24	lens	loam	roared

These 24 trial pairs were produced by two SSBE speakers (one female, one male) comprising 48 trials. The order of the 24 trial pairs was randomised for each speaker.

Appendix E

The first third of “The Rainbow Passage” used as a reading material for production tests in Chapter 2

The Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colours. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation.

“The Rainbow Passage” was cited from the page 127 of Grant Fairbanks’ (1960) book, “Voice and Articulation Drillbook 2nd Edition” (New York: Harper & Row).