2013-10-02

Effects of the Use of Ultrasound in Production Training on the Perception of English /r/ and /l/ by Native Japanese Speakers

Tateishi, Miwako

UNIVERSITY OF CALGARY

Effects of the Use of Ultrasound in Production Training on the Perception of English /r/ and /l/ by Native Japanese Speakers

by

Miwako Tateishi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

DEPARTMENT OF LINGUISTICS

CALGARY, ALBERTA

SEPTEMBER, 2013

Abstract

Speech sound contrasts in second languages can be difficult for adult language learners to perceive. Numerous studies have been conducted to examine which training methods can improve perception of non-native sound contrasts. This study investigates whether production training using ultrasound as visual feedback leads to improved production and perception of non-native sound contrasts. To this end, Japanese learners of English who were beginning ESL students were trained to accurately produce English /r/ and /l/. During the training, learners were shown ultrasound images. Before and after the training, they underwent perception tests to identify whether or not the production training was successful. Results showed that the production of /l/ potentially improved. However, this did not lead to improved perception of the phoneme. Moreover, the learners' improvements varied in degree and modality. Thus, the results indicate that perception and production undergo a different developmental course with considerable individual variation.

Acknowledgements

It has been a long journey since I decided to pursue research in graduate school. During my graduate education at the University of Calgary, I have gained knowledge and experience that will help me move forward as a researcher. None of the accomplishments I have made in the MA program would have been possible without support from my professors, colleagues, friends, and family.

I am deeply grateful to my supervisors, Dr. Suzanne Curtin and Dr. Mary G. O'Brien, for their guidance and supports. I have learned a lot from them about how to conduct research and analyze results from a broad point of view. Moreover, their advice and comments on the literature review section helped me to extend my knowledge. I would not have been able to complete my research without them.

I am also grateful to Dr. Stephen Winters for his advice on the original design of my research, support for experimentation and volunteer RA recruitment, and questions and comments on my research. Moreover, I would like to thank Dr. Xiao-Jie Yang, for his thoughtful questions and comments on my research.

My training experiment is obviously the most important part of my research. I would like to thank Dr. Elizabeth Ritter and Dr. Susanne Carroll for kindly providing financial support for the experiment.

I am thankful to my professors at the University of Victoria - to Dr. Sonya Bird for her kindly letting me use her lab and ultrasound machine, as well as for her advice and support, and to Dr. John Esling for kindly letting me be a part of the Linguistics Department as a visiting graduate researcher. I am also thankful to my friends at UVic – to Thomas Magnuson for being a model talker, as well as advice and support throughout

the experiment, to Chris Coey for his assistance with laboratory equipment, and to Eoin Whitney for his help with scheduling my research participants.

Also, I would like to thank Dr. Kota Hattori for his advice and comments on the design of the training experiment. His doctoral research on spectrogram production training inspired my thesis research, and I am glad to know that we share very similar research interests.

My research would not have been completed without help from my colleagues and friends at the University of Calgary, and I would like to extend my thanks to them. Joey Windsor, Dr. Stephanie Archer, Jacqueline Jones, Rein Sastok, Sarah Greer, and Christine Merlihan - thank you for recording your speech and/or being pilot subjects for me. Thea Jochelson, Randall Thacker, and Rosie Williams – thank you for assisting me with the production data. Danica MacDonald and Vincent Carveth – thank you for your moral support. Without you, I would not have been able to overcome the challenges I encountered throughout the whole process.

Additionally, I am grateful to my friends in the United States – Dr. Sherri Novis Livengood and Jenna Luque. Sherri – I cannot even imagine how I could have done the data analysis without you. Jenna – thank you for your advice on production training from a viewpoint of a SLP.

Of course, my research would not have been possible without the research participants who were hard-working ESL students in Victoria. I talked with each of them about why they were studying English in Canada. I hope their dreams come true.

Finally, I am thankful to my parents for their understanding of my pursuit of a career as a researcher and being always there for me.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
ANOVA	Analysis Of Variance
c	Criterion location as a measure of response bias
CV	Consonant and vowel
df	Degree of freedom
d'	D-prime as a measure of perceptual sensitivity
EFL	English as a foreign language
ESL	English as a second language
F	Fisher's ratio
F1	First formant
F2	Second formant
F3	Third formant
HVPT	High variability phonetic training
L1	First language
L2	Second language
M	Sample mean
MRT	Mental rotation test
n	Number of cases in a subsample
NE	Native English
NJ	Native Japanese
NLM-e	Native Language Magnet Theory Expanded
p	Probability
PAM	Perceptual Assimilation Model
r	Estimate of the Pearson product-moment correlation coefficient
SD	Standard deviation
SLM	Speech Learning Model
U	The Mann-Whitney test statistic
VOT	Voice onset time
z	The value of a statistic divided by its standard error

Chapter One: Introduction

The ability to perceive and produce speech accurately is crucial to the acquisition of a spoken language. Learning a language's speech sound categories would be challenging if the learner could not accurately identify the spoken forms of the sounds. Moreover, deviations in the production of speech sounds could render speech unintelligible to native listeners of the language. Indeed, without experience with atypical productions, such as those exhibited in child productions, a listener often has a difficult time identifying the target word being attempted. Over the course of first language acquisition, children master the production of speech sounds and become able to discriminate their native-language speech sound categories. On the other hand, both the perception and production of speech sounds in a non-native language can be challenging to adult second language learners due to long-time experience with their native language. Numerous studies have been conducted to investigate how and when the abilities to perceive and produce speech sounds become specific to the learner's native language, and how these abilities can be modified during adulthood.

1.1 Perception of non-native sounds

Infants under six to eight months of age can perceive speech sounds not only in the language that they are learning as their native language but also many sounds that occur in other languages; however, their ability to perceive speech sounds across languages gradually declines, and perception of non-native speech sounds eventually becomes more difficult beyond 12 months (Werker & Tees, 1984a). Studies have revealed that this perceptual decline during early infancy does not reflect sensory-neural loss due to

maturation and lack of experience with non-native languages. Rather, it reflects attenuated perceptual sensitivity due to language experience during infancy. For example, Werker and Tees (1984b) showed in their subsequent study that adults do retain some of their sensitivity to acoustic cues that distinguish non-native sounds. That is, adult listeners were able to attune their perceptual sensitivity to the acoustic cues when these cues were made more salient in modified speech; however, they did not utilize this sensitivity when listening to natural speech.

This declined sensitivity to non-native speech sounds in adults is likely due to interference from previous experience with their native language (Flege, 1995, 2003; Kuhl, 2000; Kuhl et al., 2008; Kuhl et al., 2006; Iverson et al., 2003). In explaining how native speech sound systems interfere with accurate perception of non-native sounds in adults, Iverson, Ekanayake, Hamann, Sennema, and Evans (2008) and Iverson et al. (2003) suggest that adults are more likely to attend to acoustic cues that may be crucial in identifying their native sounds but are unreliable, or insufficient when used alone, in identifying non-native sounds. To illustrate, according to Iverson et al. (2008), native Sinhala speakers tend to have difficulties in discriminating between English /w/ and /v/ possibly because Sinhala has one phoneme (the labiodental approximant /ʋ/) that is similar to both of these English phonemes. Native English speakers attend to the first formant frequency (F1), the second formant frequency (F2), and frication noise amplitude in distinguishing between /w/ and /v/; however, native Sinhala speakers are likely to attend to only F1 and frication noise amplitude, which appear to be crucial in identifying the Sinhala /ʋ/ but are not sufficient in distinguishing between English /w/ and /v/ (Iverson et al., 2008).

Despite the native language interference effects, it is not impossible for adults to learn non-native sound categories. Flege (1995, 2003) proposes that the underlying capabilities of first language (L1) speech acquisition may remain available for second language (L2) speech acquisition. He further argues that in order to acquire L2 speech sounds, learners may require a large amount of speech input from native speakers of the language over a long period of time. In line with this claim, Kuhl et al. (2008) suggests that neural networks for speech sound category learning remain malleable until the learner receives an adequate amount and variety of speech input for a particular phonetic category, which can be achieved through extended residency in environments in which the language is primarily spoken (e.g., Flege, Bohn, & Jang, 1997) or short-term intensive perceptual training (e.g., Logan, Lively, & Pisoni, 1991).

Not all non-native sound contrasts are difficult for adults to perceive, however. For example, adult English speakers can discriminate isiZulu click consonants as accurately as English-learning infants (Best, McRoberts, & Sithole, 1988). Moreover, ease or difficulty in perceiving non-native contrasts appears to depend on the native language of the adult listeners. To illustrate, perception of the contrast between the unaspirated retroflex and the dental stops in Hindi is likely to pose a challenge to English speakers (Werker, Gilbert, Humphrey, & Tees, 1981), whereas perceiving this contrast is easier for native Japanese speakers (Pruitt, Jenkins, & Strange, 2005). Native Sinhala speakers tend to have difficulty in perceiving the contrast between English /w/ and /v/, whereas native speakers of Dutch perceive this contrast with near-native or native level accuracy (Iverson et al., 2008). Perception of English vowel contrasts is more likely to be challenging to native speakers of Spanish or French than native speakers of German or

Norwegian (Iverson & Evans, 2007). It appears that perception of a non-native sound contrast is more likely to be easy if the sound system of the listener's native language involves a similar contrast.

In summary, the perception of speech sounds begins as language-universal and becomes language-specific as an infant's experience with the ambient language increases, thereby interfering with perception of non-native speech sounds. Attenuated perceptual sensitivity to non-native speech sounds may be recovered through long-time experience with the language or short-term intensive training. Ease or difficulty in perceiving non-native contrasts may depend on whether the listener's native language has similar contrasts.

1.2 Production of non-native sounds

Flege (1995, 2003) claims that individuals learn L2 sound systems through the mechanisms and processes that were employed to learn the sound system of their L1. Similarly to the perception of non-native sounds, the production of non-native sounds is influenced by the sound system of the learner's native language, especially for late learners (Flege, 2003; Ioup, 2008, Major, 2008). Therefore, it may be inevitable that adult learners' production of non-native sounds bears coloring of their native sounds. For successful L2 production learning, learners are required to assess the distinct properties of the L2 and L1 speech sounds accurately, store and organize this information in long-term memory, and learn accurate articulatory gestures for the L2 sounds (Flege, 1995).

The ability to learn production of non-native speech sounds appears to decline with increasing age (Hakuta, Bialystok, & Wiley, 2003). Baker, Trofimovich, Flege,

Mack, and Halter (2008) examined production of English vowels by Korean-speaking children and adults who had lived in the United States for approximately one year. The results revealed that production of certain English vowels by the children was more intelligible to native English listeners than that of the adults. Moreover, Flege, Munro, and MacKay (1995a, 1995b) showed that production of English speech by adult Italian speakers who arrived in Canada before puberty was judged more intelligible and less foreign-accented by native English listeners than that by adult Italian speakers who arrived in Canada after puberty.

Although early onset of language learning appears to be advantageous in production learning of non-native sound categories, it may also be possible for adults to learn production of non-native sounds with extended experience with the language. Flege, Bohn, and Jang (1997) investigated whether adult learners of English who had arrived in the United States after puberty would vary in the accuracy of their production of English vowels as a function of length of residence. They found that, depending on the learners' native language and the types of the vowels used in the experiment, learners who had resided in the United States for an average of 7.3 years exhibited more accurate production of the vowels than learners who had resided in the United States for an average of 0.7 years. Therefore, adults may be capable of acquiring production of non-native sounds with higher accuracy if they have been immersed in the language sufficiently long. However, the effect of length of residence may become limited by frequent use of L1 in adults (Flege, Frieda, & Nozawa, 1997; Piske, MacKay, & Flege, 2001). In some cases, adult learners may be able to achieve native-like production accuracy, however, by compromising their native speech production to some degree.

Major (1996) showed that the voice onset time (VOT) for Portuguese voiceless stops produced by an American immigrant to Brazil (a native English speaker) did not differ from the VOT for the same consonants produced by native Portuguese speakers. However, the VOT for English voiceless stops in her spontaneous speech significantly differed from the VOT for the same consonants in spontaneous speech from native English speakers.

Adult L2 learners may be able to acquire production abilities with extended exposure to the language and reduced L1 use. However, production training with a direct teaching approach, which provides explicit instructions on articulatory gestures, has been shown to be more effective for learning production of difficult L2 sounds than simple imitations of speech from native speakers of the language (Odisho, 2003: Schmidt & Beamer, 1998). Thus, adult L2 learners may benefit more from production training than being simply immersed in an English-speaking environment.

In short, early onset of non-native speech production may have advantages. Nevertheless, adults may be able to learn to produce non-native speech sounds more accurately with extended experience with the language, provided that they minimize the use of their native language. For adult L2 learners, receiving production training may be a better and more efficient way to learn the production of difficult L2 sounds.

1.3 Intelligibility of non-native speech

According to Munro and Derwing (1995a), intelligibility of speech refers to the extent to which speech is actually understood as intended by native speakers of the language. Intelligibility can be measured in numerous ways, such as transcriptions (e.g., Munro &

Derwing, 1995a), listener judgments of production errors (Anderson-Hsieh, Johnson, & Koehler, 1992), and sentence verification (Munro & Derwing, 1995b). In one study, Munro and Derwing (1995a) examined how intelligibility is related to perceived comprehensibility (the native speakers' perception of the degree of ease or difficulty that they experienced when attempting to understand the speech) and accentedness (the degree of the deviance of the speaker's accent from that of native speakers). In this study, native English listeners perceptually evaluated the intelligibility, comprehensibility, and accentedness of English speech samples produced by native Mandarin speakers. The study revealed that more intelligible speech tended to be more easily understood, indicating a closer relationship between intelligibility and comprehensibility. On the other hand, intelligibility was shown to be more independent of accentedness. For example, some strongly foreign-accented utterances were highly intelligible to the listeners. These findings were confirmed in a subsequent study by Derwing and Munro (1997), which evaluated English speech produced by non-native English speakers from different L1 backgrounds. It has been hypothesized that comprehensibility judgments are made based more on processing difficulty experienced by listeners, whereas accentedness corresponds more to saliency of production errors, which does not necessarily impact the processing of the speech (Munro & Derwing, 1995b; Munro & Derwing, 2006). Thus, highly intelligible speech is likely to be easily processed by listeners, although it could be strongly foreign-accented. Native English speakers are likely to prefer more comprehensible speech to more foreign-accented speech (Derwing & Munro, 2009).

Both prosodic errors and segmental errors have been shown to impact intelligibility and perceived comprehensibility of non-native speech, although prosodic

errors may have a greater influence on comprehensibility than segmental errors do (Munro & Derwing, 1995a; Derwing & Munro, 1997). The importance of prosody in intelligible speech, such as correct placement of primary stress in English speech, has been demonstrated (Hahn, 2004). Further, Bent, Bradlow, and Smith (2007) showed that an English speech sample from a Mandarin speaker with the lowest segmental errors was judged by native English listeners as least intelligible, possibly due to factors such as prosodic errors, which were not measured in the experiment.

Nevertheless, segmental errors can affect the intelligibility of non-native speech, especially when they occur in word-initial positions (Bent et al., 2007). Also, the influence of segmental errors may interact with the functional load that the segments carry (Munro & Derwing, 2006). Functional load is "a measure of the work which two phonemes (or distinctive features) do in keeping utterances apart – in other words, a gauge of the frequency with which two phonemes contrast in all possible environments" (King, 1967, p. 831). Brown (1988) assessed the functional load of English segmental contrasts using various measurements and ranked them on a 10-point scale from 1 (*functional load with minimal importance*) to 10 (*functional load with maximal importance*). He further suggested that conflating segmental contrasts that carry high functional loads might be problematic in effective communications between non-native and native speakers of the language. Munro and Derwing (2006) tested this hypothesis and showed that production errors on segmental contrasts with higher functional loads tended to have greater impact on comprehensibility, which may subsequently affect intelligibility.

The results of a range of experiments show that highly intelligible non-native speech tends to be easier to process for native listeners, although it might not be free from strong foreign accents. Fewer production errors on prosody as well as on segmental contrasts that carry a high functional load are likely to result in more comprehensible speech, which may subsequently be judged as more intelligible.

1.4 Perception and production of English /r/ and /l/ by native Japanese speakers

1.4.1 Perception of /r/ and /l/

It has been well documented that adult native speakers of Japanese are likely to have difficulties in discriminating between English /r/[1] and /l/ (e.g., Goto, 1971; Miyawaki et al., 1975; Takagi & Mann, 1995). Further, Goto (1971) and Sheldon and Strange (1982) showed that some Japanese speakers had difficulties in correctly identifying the sounds not only in others' speech but also in their own speech.

Unlike English, which has two liquid phonemes, /r/ and /l/, which are an alveolar approximant and a lateral approximant respectively (Ledefoged & Maddieson, 1996), Japanese has only one liquid phoneme that is somewhat similar to both the English phonemes (Ohata, 2004). Best and Strange (1992) speculated that Japanese speakers may hear both /r/ and /l/ as poor examples of Japanese /r/, which is quite different from /r/ and /l/. The Japanese liquid /r/ is phonetically an apico-alveolar tap [ɾ] and produced by lightly touching the alveolar ridge with the tongue tip then rapidly releasing this contact (Vance, 2008). However, subsequent research has shown that Japanese speakers do not

[1] The IPA symbol for the English r sound is /ɹ/. In this thesis, /r/ is used to represent the English r, as it is conventionally used in this type of research.

hear instances of /r/ and /l/ as equally poor examples of the Japanese tap. Rather, they are more likely to hear English /l/ as the Japanese tap (Hattori & Iverson, 2009; Iverson et al, 2003; Takagi, 1993).

The primary acoustic cue that differentiates between English /r/ and /l/ is the third formant frequency (F3), which is lower for /r/ and higher for /l/ (see Figure 1.1), and distributions of instances of these phonemes can be clearly separated by F3 onset frequency (Dalston, 1974; Iverson, Hazen, & Bannister, 2005; O'Conner, Gerstman, Liberman, Delattre, & Cooper, 1957; Lotto, Sato, & Diehl, 2004).

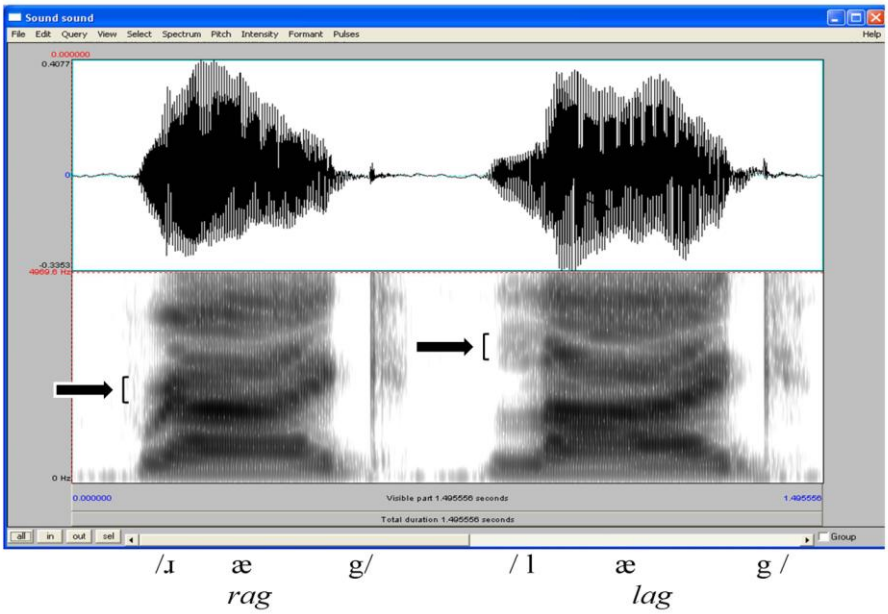

Figure 1.1. Spectrogram of *rag* and *lag*. Arrows indicate F3 frequencies

These English phonemes also differ by F2 in that F2 is slightly higher for /l/ than for /r/; however, it does not appear to be a reliable cue in discriminating the phonemes for native English speakers (O'Conner et al., 1957). This is consistent with the observations in which the F2 frequency range for /r/ overlaps with the F2 frequency range for /l/ when multiple instances of /r/ and /l/ from native English speakers are mapped (Iverson et al., 2005; Lotto et al., 2004). Additionally, English speakers appear to use differences in transition durations between /r/ or /l/ and the succeeding vowel in consonant-vowel syllables (longer transitions for /r/, and shorter transitions for /l/) as the secondary cue to the distinction between the sounds (O'Conner et al., 1957)

Japanese speakers are less sensitive to the F3 difference than native English speakers are (Best & Strange, 1992; Iverson et al., 2003; Miyawaki et al., 1975) although they exhibit sensitivity to this acoustic cue comparable to that of native English speakers when F3 contours of /r/ and /l/ in natural speech are isolated (Miyawaki et al., 1975). Lotto et al. (2004) reported that a part of the F3 frequency range for the Japanese tap overlaps more with the F3 frequency range for English /l/ than that for /r/, and the overlapped portion covers the optimal boundary between /r/ and /l/ in terms of F3. Thus, this acoustic characteristic of the tap may contribute to the reduced sensitivity to the F3 difference between the English phonemes in Japanese speakers.

Iverson et al. (2003) found that Japanese speakers show higher sensitivity to F2 differences than to F3 differences when discriminating instances of /r/ and /l/. Based on this finding, they hypothesize that F2 may be a crucial acoustic cue to the identity of the Japanese tap. Lotto et al. (2004) pointed out that instances of the Japanese tap lie in a higher F2 range while instances of Japanese /w/ lie in a lower F2 range, indicating that

Japanese speakers may weigh F2 differences heavily because of the F2 difference between these native sounds and apply this strategy when perceiving the English /r/ and /l/ contrast. Consequently, by weighing the irrelevant acoustic cue when identifying /r/ and /l/, Japanese speakers may form incorrect representations for these sound categories (Yamada, 1995).

Japanese adults are less likely to improve their perception of /r/ and /l/ than Japanese children after one-year residency in an English-speaking country (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). In addition to age of acquisition, factors such as the amount of Japanese use (Flege & McKay, 2004), length of residence (Aoyama et al. 2004; Flege & Liu, 2001), and the amount of speech input from native English speakers (Flege & Liu, 2001) may influence how likely Japanese adults acquire the perceptual ability to discriminate the phonemes. Ingvalson, McClelland, and Holt (2011) showed that among these factors, longer length of residency best predicts more accurate perception for Japanese adults. However, this may not guarantee native-like perceptual accuracy (Ingvalson et al., 2011; Takagi & Mann, 1995).

To summarize, accurate perception of English /r/ and /l/ is likely to be challenging to native Japanese speakers due to the perceived similarity between those phonemes and the Japanese liquid /r/, despite that the liquid is phonetically the tap [ɾ] that is distinct from these English phonemes. L1 speech sound systems tend to influence L2 speech learning in adult learners (Flege, 1995, 2003). Therefore, the perceived similarity between the Japanese tap and the English phonemes may lead Japanese speakers to form erroneous representations for the L2 sounds. Extended residence in an English-speaking

country may facilitate the perception of /r/ and /l/ for Japanese adults, although this does not necessarily guarantee native-like perceptual accuracy.

1.4.2 Production of /r/ and /l/ by native Japanese speakers

English /r/ and /l/ can also be difficult for native Japanese speakers to produce (e.g., Goto, 1971; Sheldon & Strange, 1982). This could be due in part to their unfamiliarity with the accurate configurations of the articulatory gestures required for these sounds (Bradlow, 2008; Wilson & Gick, 2006).

Articulations for North American English /r/ are traditionally classified as the retroflex /r/, which is produced with raising of the tongue tip toward the hard palate, or the bunched /r/, which is produced with raising of the mid-tongue dorsum or the tongue blade near the hard palate and lowering of the tongue tip (Delattre & Freeman, 1968). Studies have shown that these two articulations are at the ends of a continuum, and many variations have been observed across speakers and phonetic contexts (Alwan, Narayanan, & Haker, 1997; Delattre & Freeman, 1968; Gunther et al., 1999). Although these articulations have distinct lingual components, they lower the F3 frequency and give the same auditory impression to the listeners (Delattre & Freeman, 1968; Ladefoged & Maddieson, 1996). The low F3 corresponds to a tongue constriction made near the palate, a sublingual space, and a lip constriction that is either rounded or spread (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000).

The articulation of /l/ involves a complete closure made with the tongue and the alveolar ridge as well as a lowered tongue body, which allows for lateral airflow (Ladefoged & Maddieson, 1996). A variant of /l/, which occurs in syllable-final positions

in American English, also includes a constriction made with the tongue dorsum (Johnson, 2003).

Lotto et al. (2004) acoustically analysed Japanese speakers' productions of English /r/, /l/, and the Japanese tap by plotting multiple instances of these sound categories in terms of F2 and F3. The category boundary between /r/ and /l/ was lost due to largely overlapping distributions of the sounds on the F3 continuum. Nonetheless, the distribution of /r/ was somewhat lower than that of /l/ on the F2 continuum. Therefore, Japanese speakers' productions of /r/ and /l/ are likely to confuse native English listeners due to the ambiguous category boundary in terms of F3. Moreover, the distribution of the Japanese tap tokens substantially overlapped with the distribution of the /r/ and /l/ tokens, suggesting an influence of Japanese sound system on production of the non-native sounds.

Japanese speakers' difficulty in producing /r/ and /l/ is likely to be detrimental to the intelligibility of their English speech because the contrast between these phonemes is classified as carrying the maximal functional load (Brown, 1988). It has been shown that when adult Japanese learners of English are at an early stage of L2 learning, their productions of /l/ are judged as more intelligible than their productions of /r/ by native English speakers (Aoyama et al., 2004; Flege, Takagi, & Mann, 1995, Hattori, 2009). This unbalanced production intelligibility between /r/ and /l/ appears to become reduced as learners gain more experience with English (Flege et al., 1995, Hattori, 2009). Aoyama et al. (2004) and Hattori (2009) found that when English listeners misidentify Japanese adults' productions of /r/ and /l/, they most likely identify /l/ as /r/ and /r/ as /l/. The

researchers also found that /r/ is also misidentified as /w/ or less often as /d/ or /b/, whereas /l/ is also identified as /d/ or less often as /w/.

Similarly to perception, Japanese adults are less likely than Japanese children to acquire accurate production of /r/ and /l/ within one or two years of residence in an English-speaking country (Aoyama et al., 2004; Flege et al., 1995). For Japanese adults, extended residency has been shown to be most predictive of more intelligible and less foreign-accented production of the English phonemes when compared with other factors such as the amount of Japanese use and the amount of speech input from native English speakers (Ingvalson et al., 2011). Thus, as with perception, production learning of /r/ and /l/ may require long-time experience with the language in an English-speaking country for Japanese adults, although this may not necessarily lead to native-like production accuracy (Ingvalson et al., 2011). Interestingly, Japanese speakers' ability to produce /r/ and /l/ surpasses their ability to perceive the phonemes in some cases (Goto, 1971; Sheldon & Strange, 1982).

In brief, learning the articulations for /r/ and /l/, especially the lingual gestures which are not easily visible to the listener, can be challenging for adult Japanese speakers. Further, Japanese adults are likely to produce these two English phonemes as a single category due to the influence of Japanese sound category system, thereby confusing native English listeners. Similarly to perception, long-time experience with English may be the primary factor for more accurate production of the sounds by Japanese adults.

1.5 Training studies on /r/ and /l/

1.5.1 Laboratory perceptual training

Japanese speakers have been shown to improve their perceptual ability to identify /r/ and /l/ after receiving a short-term, intensive laboratory perceptual training called High Variability Phonetic Training (HVPT) (e.g., Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991). In this training, Japanese learners listen to multiple instances of /r/ and /l/ in a variety of phonetic contexts in natural speech from multiple native English speakers. After the training, the Japanese learners correctly identified the sounds in words significantly more often than before the training. The learners were also able to utilize their improved perceptual skill when listening to novel words produced by native English speakers who were not used for the training. A subsequent study further revealed that Japanese learners who received HVPT retained their improved perceptual ability after six months without additional training (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994).

The success of HVPT may be attributed to the variability in the stimuli. That is, by listening to a large number of instances of the new sound categories, Japanese speakers may learn which cue is most relevant in identifying the categories, thereby changing their cue-weighing strategy and forming robust representations of the categories, which can be generalized to speech from unfamiliar speakers (Iverson et al., 2003; Logan et al., 1991). HVPT has been shown to be effective in facilitating perception of other speech sound contrasts, such as Mandarin lexical tone contrasts by native speakers of English (Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999), English vowel contrasts by native speakers of German and Spanish (Iverson & Evans, 2009), as well as by native speakers of French (Iverson, Pinet, &

Evans, 2012), and a Hindi stop-retroflex contrast by native speakers of English (Pruitt, Jenkins, & Strange, 2006).

Subsequently, Bradlow, Pisoni, Akahane-Yamada and Tohkura (1997) demonstrated that HVPT promoted learning of accurate perception and production of the same sounds, although production was not targeted in the training. Bradlow, Akahane-Yamada, Pisoni and Tohkura (1999) further revealed that this improved ability to produce /r/ and /l/ was retained after three months without additional perceptual training. If perceptual training can improve perception and production of the same sounds, one may wonder if the opposite is true. That is, does production training improve production and perception of the same sounds even if perception is not a target in the training?

A pilot study by Catford and Pisoni (1970) suggests this could be possible. In this study, two groups of native English speakers, who had no knowledge of linguistics or phonetics, were trained to produce exotic speech sound contrasts that they had never heard or produced before. One group received instructions only referring to articulatory gestures of the sounds and were not allowed to listen to the sounds. The other group repeatedly listened to the sounds produced by the teacher whose mouth was covered by a screen and received no information on articulations of the sounds. After listening to the sounds, the learners were asked to mimic the sounds that they heard. The results of post-training perceptual and production tests revealed that the articulatory-only training group significantly outperformed the auditory-only training group in both production and perception of the trained sounds.

To the best of my knowledge, this is the only study that showed that purely articulatory production training may facilitate production and perception of the same

trained sounds in the absence of perceptual stimuli. However, a major limitation of this study is that it did not examine the learners' baseline performance before the training; therefore, it is difficult to determine whether the observed difference in performance between the two groups can be solely attributed to the difference in training method. Nevertheless, the study does indicate that improved production may lead to improved perception for the same sounds.

1.5.2 Production training to improve production and perception of /r/ and /l/

For teaching complex motor skills, providing learners with visual information of model actions and visual feedback on the learners' own actions has been shown to be successful (Carroll & Bandura, 1982). Visualization technology for motor skill teaching has been incorporated in speech production training. For example, Massaro and Light (2003) conducted a three-day production training experiment in which a computer-animated talking head, Baldi, taught Japanese learners of English how to produce English /r/ and /l/. Baldi's skin can be made transparent so that learners can observe the articulators in various views. During the training, Baldi also provided the learners with instructions on how to produce /r/ and /l/ verbally, and the learners' productions of these phonemes were evaluated through an automatic speech recognition system. Feedback on their productions, which was either a happy face or a sad face, was provided to the learners. Intelligibility of the learners' productions of /r/ and /l/ (judged by the speech recognition system) showed a small (4%) but significant improvement after the training. Moreover, the learners' perception of /r/ and /l/ was also improved. However, only three sets of minimal-pair words were used in the perceptual tests before and after the training, and

these words were also used as targets in the production training (no additional training targets were included). Therefore, it seems possible that the learners improved their perception of /r/ and /l/ because they repeatedly heard these test words during the training. Additionally, visual feedback on the learners' productions and explicit feedback on their articulations when they mispronounced the targets (i.e., explanations for why their articulations were not good and advice on how they can be improved) were not provided in the training. Thus, the small improvement in production could have resulted from the lack of visual and specific verbal feedback on the learners' own articulations.

Hattori (2009) explored the question of whether production training with acoustic spectrograph technology can lead to improved production and perception of /r/ and /l/ by Japanese learners of English. For learners, this training study employed: 1) explicit instructions on articulatory gestures (i.e., lip and tongue movements) by a phonetically-trained instructor; 2) acoustic spectrograms of the learners' own speech as visual feedback on their articulation; and 3) recorded speech of the learners in which the acoustic components of /r/ and /l/ were modified to be their ideal speech. Additionally, the study utilized real-time spectrograms of the learners' speech for the instructor to monitor the production of the learners. In the training, the learners initially received instructions on correct articulatory gestures for the English phonemes and practiced the production of the phonemes in isolation, consonant-vowel syllables, and monosyllabic words. While the learners were practicing, the instructor monitored lip shapes of the learners as well as real-time spectrograms of the learner's speech and provided feedback on their production. After the learners made recordings of the training targets, they saw spectrograms of the recorded production and received feedback in terms of three acoustic

components (F3 onset value, duration of the consonant, and duration of the transition from the consonant to the following vowel). Furthermore, the learners were asked to compare these spectrograms with spectrograms of their modified speech. In the modified speech, these acoustic components were enhanced by using signal processing techniques employed in Iverson et al. (2005) in order to create the learners' best pronunciations of the phonemes. For example, F3 frequencies for /r/ and /l/ were set to the values that enhanced the distinction between these phonemes.

Analyses of perception and production of /r/ and /l/ by the learners before and after the training revealed that the learners' production was significantly improved after the training; however, the training did not improve perception of the same phonemes. Thus, Hattori (2009) suggests that perception and production may not share common mental representations of speech sounds, and learners may need to build associations between perception and production. He further speculates that it may take longer to build the associations with production training. That is, production learning may transfer to perceptual learning, but the process may take place slowly.

1.5.3 Ultrasound training to improve production of /r/ and /l/

In recent years, ultrasound technology has been used in production training. Ultrasound allows learners to see the appropriate articulation and the appropriate timing of articulatory gestures for a speech sound by providing direct, dynamic images of tongue movements in both mid-saggital (front-to-back) and coronal (side-to-side) views.

The Japanese learners in the study by Hattori (2009) received feedback on their own productions in terms of acoustic components displayed in spectrograms. However,

the only direct visual information on model articulatory gestures for the learners was lip and tongue movements of the instructor as well as of a native English speaker who was videotaped. Although the learners also received advice on their own articulatory gestures, it may take time for individuals without phonetic training to interpret how certain articulatory gestures relate to the acoustic components in the spectrograms. Also, the tongue shapes and movements to articulate /r/ and /l/ are complex and are not easily seen by just looking at the mouth. Therefore, ultrasound may overcome these issues by providing direct visual information on the articulatory gestures of the sounds used in model speech and in the learners' own speech.

The utility of ultrasound in production training has been demonstrated in L1 speech remediation studies. For example, in a study by Adler-Bock, Bernhardt, Gick, and Bacsfalvi (2007), two English-speaking adolescents with persistent production difficulties learned how to produce English /r/ with ultrasound images as visual feedback on their own production. After the training, their production of /r/ became significantly more intelligible, and the F3 frequencies of /r/ in their speech became lowered. Moreover, the tongue shapes of the learners became more similar to these of typical adult English speakers. Ultrasound has also been shown to be effective when training hard-of-hearing adolescents to produce English consonants including /r/ and /l/ by providing rich visual information on the articulations for the sounds (Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003).

Ultrasound technology has also been incorporated in L2 speech training research that facilitated production learning of English /r/ and /l/ by native Japanese speakers. Gick, Bernhardt, Bacsfalvi and Wilson (2008) demonstrated in their pilot study that

Japanese learners of English can improve their production of /r/ and /l/ in a variety of phonetic contexts with ultrasound. The Japanese learners in this study participated in a 30-minute production training session, in which they were provided with information on the lingual components for production of the phonemes and real-time ultrasound images of their own productions as visual feedback. The training enabled all the learners to produce the phonemes in the contexts that were challenging to them prior to the training. However, because the learners had been trained in linguistics at a university, their knowledge in phonetics might have contributed to the success of this short-term training. Following this pilot study, Tsui (2012) revealed that Japanese learners of English without any knowledge of linguistics, and with varying degrees of experience with English can also benefit from ultrasound production training. In her study, the Japanese learners whose length of residence in English-speaking countries ranged from two months to three years improved their productions of English /r/ and /l/ after receiving a longer ultrasound training (four sessions, approximately 45 minutes per session).

These studies demonstrate the promising utility of ultrasound imaging technology for Japanese learners to overcome their persistent difficulties in producing the English phonemes. We know less about whether this training has an effect on the perception of these same phonemes. Gick et al. (2008) did not examine whether the improved production led to improved perception of the same phonemes in the Japanese learners. Additionally, Tsui (2012) conducted an exploratory investigation of the learners' perceptual ability with inconsistent numbers of perceptual tasks throughout the experiment (20 trials before the training, and 60 trials at the midpoint and after the training). Interestingly, one of the six learners participated in her study improved her

perception after the training whereas five showed decline and one showed no change. Nevertheless, it is difficult to determine specifically whether the observed changes in the learners' perception accurately resulted from the training due to the methodological issue.

1.6 Research questions

Three questions explored in the present study are: 1) Does production training using ultrasound imaging as visual feedback lead to improved production of English /r/ and /l/ by Japanese learners of English in terms of F2 and F3?; 2) Does the training improve perception of the same phonemes by the Japanese learners in the absence of perceptual training?; and 3) Does the training improve the intelligibility of the Japanese learners' productions of English /r/ and /l/?

It is predicted that ultrasound training will lead to improved production quality. That is, it will lead to changes in F2 and F3 in Japanese learners' productions that will make them more closely approximate the F2 and F3 in native English speakers' productions of the phonemes (Tsui, 2012). Second, it is predicted that utilizing visualization of tongue shape and movements as feedback in production training may facilitate the perception of /r/ and /l/ (Adler-Bock et al., 2007; Massaro & Light, 2003; Tsui, 2012). That is, the training may lead to improved perception even in the absence of perception training. Lastly, it is predicted that the training will improve the intelligibility of the Japanese learners' productions of /r/ and /l/ for native English listeners, replicating the results from the previous ultrasound training studies (Gick et al., 2008; Tsui, 2012).

Chapter Two: Production Training Experiment

2.1 Introduction

The present study investigated effects of ultrasound production training on the production and perception of English /r/ and /l/ in the absence of perceptual training. Experiment 1 explored whether native Japanese learners of English would improve their production of English /r/ and /l/ as a result of the training. Changes in the ability to produce /r/ and /l/ were gauged in terms of F2 and F3 frequencies. The learners' productions before and after the training were also compared with native English speakers' productions of /r/ and /l/ in terms of the same acoustic measures. As discussed in Chapter 1, for native English speakers' productions, /r/ and /l/ clearly differ by F3 (low for /r/, high for /l/) and slightly differ by F2 (low for /r/, high for /l/) although F2 ranges for the phonemes overlap. In contrast, for Japanese speakers' productions, /r/ and /l/ are not quite distinctive in F3 but somewhat more distinctive in F2 (Lotto et al., 2004). Thus, Experiment 1 was intended to examine: 1) whether F3 for /r/ and /l/ would become more distinct in Japanese learners' productions; and 2) whether F2 for the same phonemes would become less distinct in the learners' productions after the training. It was hypothesized that Japanese learners' production of /r/ and /l/ would become more distinct in F3 and less distinct in F2, as seen in English speakers' productions of the phonemes, if the ultrasound production training is effective.

Experiment 2 explored whether the same Japanese learners would improve their perception of /r/ and /l/ as a result of the training. Changes in the ability to identify these phonemes accurately were gauged in terms of percentages of correct identification of the phonemes. If Japanese learners of English improve their perception of the phonemes, this

will provide support for the influence of production on subsequent perception (Massaro & Light, 2003; Tsui, 2012). It was hypothesized that Japanese speakers would improve their abilities to identify /r/ and /l/ accurately if the acoustic measures in their productions become similar to those in English speakers productions for the phonemes.

2.2 General experiment design

The production training experiment comprised three stages: 1) pre-training perception test and production recordings; 2) production training; and 3) post-training perception test and production recordings. The training comprised five separate sessions. Following the experimental design by Hattori (2009), each session lasted approximately 30 minutes, and only one session took place per day. The entire experiment took place over a three-week period. The number of training sessions was originally planned to be 10 (as in Hattori, 2009). However, it was reduced to five due to technical, time, and monetary constraints. The pre-training perception test and recordings were conducted on the day before the first training session, and the post-training perception test and recordings were completed immediately after the fifth training session. The perception test took place before the production recordings. The experiment schedule is outlined in Table 2.1 below.

Table 2.1. Experiment schedule

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
1) Perception test 2) Production recordings	Production training				1) Production training 2) Perception test 3) Production recordings

Prior to the pre-training test and recordings, Japanese learners completed a questionnaire detailing their educational background and language learning experience, as well as personal information such as age and gender. They also completed the Mental Rotation Test (MRT) (Vandenberg & Kuse, 1978), which measures spatial processing abilities. In the test, the learners were asked to look at a given two-dimensional object and find the same object rotated in a set of dissimilar two-dimensional objects. The production training required the learners to process, understand, and use the representations of the tongue in ultrasound images. Therefore, the MRT was included to explore whether the ability to mentally manipulate objects required for this test would correspond to better understanding and use of ultrasound images of the tongue, which might, in turn, lead to improved production. To this end, correlations between MRT scores and intelligibility of the learners' productions on the post-test were examined (Results are presented in Chapter 3). After the completion of the study, learners completed a questionnaire asking whether the training using the ultrasound was helpful for them. All instructions were given in Japanese in order to accommodate the learners' language as well as to avoid potential influence of English use on the learners'

performance. The experiment was conducted in the Speech Research Laboratory at the University of Victoria. Recordings of auditory stimuli for the pre-/ post-training perception tests and auditory prompts for the production recordings were made in the Phonetics Laboratory at the University of Calgary.

2.3 Production training

2.3.1 Method

2.3.1.1 Participants

Participants were 10 native Japanese speakers (four male and six female) ranging from 18 to 30 years of age (mean age: 24.6 years). One additional participant completed the pre-training perceptual test and production recordings but withdrew because she could not be available for the production training as well as the post-training perceptual test and production recordings. Her data from the pre-training perceptual test and production recordings were excluded from analysis. All participants were recruited through advertisements posted in the University of Victoria and a Japanese grocery store in Victoria, British Columbia, as well as a website targeted to Japanese communities in Canada. All of the participants were attending ESL programs offered at the University of Victoria English Learning Centre, Camosun College Interurban Campus, or private ESL schools in downtown Victoria. All the participants and their parents were born and raised in Japan. Participants had been living in Canada no more than four months (except one who had been living in Victoria for nine months), and none had lived in any other English-speaking countries before coming to Canada. All participants had received formal English instruction for a minimum of six years, starting at the age of 12 or 13, in

middle school and high school; five had also received additional English instruction in junior college or university for periods ranging from one year to three years. In addition, four had attended private English language schools or learned English from a private tutor for periods ranging from four months to three years. None spoke a language other than Japanese and English fluently, and none reported speech or hearing impairments.

Although all participants had had English learning experience in Japan, it should be noted that English is taught as a foreign language (EFL) in Japan, with more emphasis on writing and reading than on speaking and listening and little emphasis on pronunciation, in formal education. Moreover, although many private EFL schools offer small-group or private lessons taught by native English-speaking teachers, students rarely have opportunities to speak English with native English speakers or listen to English speech outside classrooms.

2.3.1.2 Apparatus

For the production training, a LOGIQe portable ultrasound machine (GE Healthcare) was used. According to researchers who have utilized ultrasound technology for lingual shape and movement analyses, ultrasound images of soft tissue are obtained through the echo patterns of ultra high-frequency sound waves emitted by and reflected back to piezoelectric crystals contained under the upper surface of a transducer (Gick, 2002; Stone, 2005). In order to image tongue shapes and movements, the transducer is placed against soft tissue under the chin; by rotating it by 90 degrees, both mid-sagittal (front-to-back) and coronal (side-to-side) views of the tongue can be captured (Gick, 2002). During the training, the transducer was hand-held by the learners themselves.

2.3.1.3 Training targets

Isolated /r/ and /l/, six consonant-vowel syllables (/ri/, /li/, /ru/, /lu/, as well as /ræ/ and /læ/ as in *rack* and *lack* respectively), and six monosyllabic minimal-pair words (*reek, leak, room, loom, rack,* and *lack*) were selected as a total of 14 targets for the production training. The minimal pair words were also used for the pre- and post-training production recordings. A native English (NE) speaker (male) recorded ultrasound images of his own production of the targets by using video recording and editing software (Sony Vegas Pro) installed in a computer at the Speech Research Laboratory at the University of Victoria. These ultrasound images were provided to learners during the training as a model of tongue shapes and movements to produce the targets. The purpose of presenting the model images was to show learners how the lingual components of /r/ and /l/ were realized in the NE speaker's production of the targets. The NE speaker produced each target six times, and these six utterances were recorded as a single video clip. He recorded the tongue movements for the first three utterances in a mid-sagittal view and for the next three utterances in a coronal view. Audio signals of his production were simultaneously recorded with the video clip. During the training, the ultrasound machine and a lap-top computer displaying the recorded ultrasound images of the NE speaker's production were placed side by side.

As discussed in Chapter 1, North American /r/ generally has the two articulation types (the retroflex /r/ and the bunched /r/), although there are many variants of them across speakers and phonetic contexts. For this production training, the bunched /r/ was selected because the NE speaker who was selected as a production model used this type of articulation in prevocalic positions.

2.3.1.4 Procedure

Learners underwent the production training individually. The training progressed from production of isolated /r/ and /l/ to production of the consonant-vowel (CV) syllables, and ultimately to production of the monosyllabic words (Table 2.2). This approach was chosen to ensure that the learners progressed through the stages equally for both phonemes and because this form of graduated training was employed by Hattori (2009) and Tsui (2012).

Table 2.2. Training schedule

Stage 1	Stage 2	Stage 3
Isolated /r/ and /l/	/ri/, /ru/, /ræ/ /li/, /lu/, /læ/	*reek, room, rack* *leak, loom, lack*

The first training session began with instructions on correct articulatory movements for /r/ and /l/, and each of the subsequent training sessions began with a review of what the learners had learned in the previous session. In each training session, ultrasound images of learners' productions and corresponding audio signals were selectively recorded in order for the experimenter (the author), who is a phonetically-trained, English-Japanese bilingual, to evaluate progress and identify difficulties for each learner for individualizing successive training. Following the approach used by Gick et al. (2008), the recorded images were also shown to the learner himself or herself for discussions with the experimenter in order to promote intellectual involvement in the

training process and self-awareness of his or her own articulation for the learner. During the discussion, the learner first looked at the images of his or her own production and the NE speaker's production for a given target side by side on a computer screen. Then, the learner was asked to describe similarities and differences between his or her production and the NE speaker's production by referring to general tongue shapes, shapes of specific parts of the tongue, and movements of various tongue parts. The images were frozen when the learner was asked to describe the shape of the tongue at a particular time point. When the learner was asked to describe movements of the tongue, the images were presented continuously.

2.3.1.4.1 Production of isolated /r/ and /l/

At the beginning of the first training session, learners were asked to describe how to produce /r/ and /l/. They all knew about raising of the tongue tip for /r/ because the retroflex /r/ is typically taught in school in Japan as well as in ESL programs in Victoria; however, very few learners could describe how to produce /l/. Subsequently, the experimenter drew diagrams of the oral cavity and the tongue in both mid-sagittal and coronal views and explained articulatory gestures for /r/ and /l/. When teaching articulatory gestures of /r/, the experimenter first taught learners that /r/ has two types of articulations, and that they would be trained to produce the bunched /r/. Production of /r/ involves: 1) lowering of the tongue tip for the bunched /r/, or raising of it for the retroflex /r/; 2) raising of the tongue dorsum or the tongue blade near the palate for the bunched /r/; 3) retraction of the tongue root toward the pharyngeal wall; 4) contact of the sides of the tongue dorsum with the back upper molars and the palate (Gick et al., 2008); and 5)

rounding of the lips, which tend to occur when producing /r/ before stressed vowels (Delattre & Freeman, 1968). Because learners were new to the bunched /r/ but somewhat familiar with the retroflex /r/, the experimenter explained the articulatory gestures for the bunched /r/ in comparison with these for the retroflex /r/. Subsequently, the experimenter explained articulatory gestures for /l/, which include: 1) contact of the tongue tip with the alveolar ridge; 2) retraction of the tongue dorsum toward the uvula or into upper pharynx; and 3) lowering of the sides of the tongue, through which the air flowing from the lungs escapes (Gick et al., 2008).

After the initial instructions, learners sat in front of the ultrasound machine and were instructed on how to hold and place the transducer for mid-sagittal and coronal views of the tongue. They were encouraged to move the tongue around in the mouth freely while looking at the display in order to gain familiarity with the ultrasound images of the tongue. Learners appeared to gain a general idea of how tongue shapes and movements were captured in the images after a few minutes of practice. The initial instructions lasted approximately five minutes, and the initial practice using ultrasound lasted approximately five minutes as part of the first 30-minute training session.

Following the initial practice, learners began practicing the production of /l/ first because /l/ is generally easier for native Japanese speakers to produce than /r/ (Hattori, 2009). Learners were presented with ultrasound images of the model production of /l/ from the NE speaker and discussed how the articulatory gestures for /l/ were realized in the model production with the experimenter. Learners practiced producing the phoneme while looking at real-time images of their own production displayed on the ultrasound machine. Although many learners became able to produce /l/ with relative ease, a few

exhibited weak release of the air through the sides of the tongue, which affected the auditory impression of the sound. They tended to have a large part of the anterior tongue surface contacting part of the palate in addition to the alveolar ridge, and they did not lower the sides of the tongue adequately. When that occurred, they were asked to ensure that only the tongue tip was touching the alveolar ridge, and that the air was flowing through the sides of the tongue. The experimenter determined the learners' mastery of the production of /l/ by evaluating tongue shapes and movements in the images in terms of the three articulatory components for /l/ (Gick et al., 2008) as criteria, as well as auditory impressions of the learners' productions.

Once the experimenter determined that learners were able to produce /l/, they progressed to the production of /r/. They were presented with the ultrasound images of the model production of /r/ from the NE speaker and discussed how the articulatory gestures for the bunched /r/ were realized in the model production with the experimenter. It was evident that most learners were experiencing difficulty in lowering the tongue tip when they practiced producing the bunched /r/: They could not help raising it, as they had previously learned to produce the retroflex /r/. Because this problem persisted for some learners, less emphasis was placed on the tongue-tip lowering movement. In addition, some learners tended to produce /r/ as /w/. When this occurred, the experimenter asked them to describe how parts of the tongue were moving and how the tongue was shaped in order to raise self-awareness of their own articulation. Throughout the training, learners were allowed to look at the images of the model production again if necessary. The experimenter determined the learners' mastery of the production of /r/ by evaluating tongue shapes and movements in the images, and lip shapes in terms of the five

articulatory components for /r/ (Delattre & Freeman, 1968; Gick et al., 2008) as criteria, as well as auditory impressions of the learners' productions

2.3.1.4.2 Production of CV syllables

Once the experimenter determined that learners were able to produce isolated /r/ and /l/, they progressed to production of the CV syllables. For each target syllable, they were presented with ultrasound images of the model production from the NE speaker, and they practiced producing the syllable while looking at real-time images of their own productions displayed on the ultrasound machine. During the training, learners were allowed to look at the images of the model production again if necessary. They were instructed on correct pronunciation of the vowels as well if they consistently mispronounced them. They initially practiced producing the syllables containing /l/ and progressed to the syllables containing /r/.

While many learners became able to produce the /l/-vowel syllables with relative ease, some showed a tendency to produce /l/ as sounds resembling an alveolar stop (/t/ or /d/) or the alveolar tap [ɾ]. This is because the contact between the tongue-tip and alveolar ridge was made too short and released abruptly during the transition from /l/ to the following vowel. When this occurred, the learners were taught to hold the contact somewhat longer and release it more slowly and smoothly. In addition, a few learners did not release the air through the sides of the tongue adequately because a large part of the anterior tongue surface was touching the palate, or the sides of the tongue were not lowered enough. When that occurred, they were taught to use only the tongue tip to make the contact and ensure that the sides of the tongue were lowered enough so that the air

could flow through them. If learners had difficulty in making correct tongue movements, they were encouraged to produce the syllables more slowly and adjust their speech rate later. Moreover, a few learners practiced producing /ili/, /ulu/, and /ælæ/ because producing /l/ in intervocalic positions was easier for them. Once they became able to produce /l/ in intervocalic positions, they practiced producing the syllables without the initial vowels.

Once the experimenter determined that learners were able to produce the /l/-vowel syllables, they progressed to the /r/- vowel syllables. Since most learners experienced great difficulty in co-articulating /r/ and the following vowel in these syllables, they were encouraged to produce the syllables slowly and adjust their speech rate later. Some learners practiced producing /r/ in intervocalic positions (/iri/, /uru/, and /æræ/) or after /w/ (/wri/, /wru/, and /wræ/) because the adjustments made the production easier for them. As they became able to produce the modified syllables, they practiced to produce the syllables without /w/ or the initial vowels. In general, not all the syllables were equally difficult for learners. For example, some found co-articulation of /r/ and the high vowels (/ri/ and /ru/) more difficult than /r/ and the low vowel (/ræ/), whereas others found /ræ/ more difficult to produce than /ri/ and /ru/. Because some learners tended to produce /r/ as /w/ in these syllables, they were taught to ensure that they were making the correct tongue movements for /r/ and to make the transition between /r/ and the following vowel more slowly. A couple of the learners tended to have the tongue tip or the tongue tip and blade contact the anterior part of the oral cavity, such as the alveolar ridge, palate, or behind the upper front teeth, when producing /ri/ and /ru/. Moreover, for the same learners, the tongue tip tended to contact behind the lower front teeth when they were

producing /ræ/. Those learners were asked to look in the mirror while producing the syllables and ensure that the tongue tip and blade were not touching anywhere in the mouth. The experimenter determined the learners' mastery of the production of the syllables based only on their productions of /r/ and /l/.

2.3.1.4.3 Production of monosyllabic words

Because many learners required longer training time for the production of the consonant-vowel syllables and were not able to produce all the syllables, only two learners progressed to the production of the monosyllabic words. For each target word, they were presented with the ultrasound images of the model production from the NE speaker and practiced producing the word while looking at real-time images of their own production displayed on the ultrasound machine. They were allowed to look at the model images again if necessary. In addition, they were instructed on correct pronunciation of the vowels as well if they consistently mispronounced them. The learners initially practiced producing the words containing /l/ and progressed to the words containing /r/. The experimenter determined that both of the learners became able to produce all the monosyllabic words by the end of the last training session based only on their productions of /r/ and /l/.

As seen in Table 2.3, all learners completed the isolated /r/ and /l/ stage. Further, two learners completed the CV syllable stage and subsequently the word stage, whereas eight learners remained in the CV syllable stage.

Table 2.3. Progress of individual learners through training sessions. Numbers denote session numbers (e.g., 1 = Session1)

	Content of training		
Learner	Isolation	CV syllable	Word
NJ1	1	2 - 5	–
NJ2	1	2 - 5	–
NJ3	1	2 - 4	4,5
NJ4	1	2 - 4	4,5
NJ5	1, 2	2 - 5	–
NJ6	1	2 - 5	–
NJ7	1	2 - 5	–
NJ8	1	2 - 5	–
NJ9	1	2 - 5	–
NJ10	1	2 - 5	–

2.4 Experiment 1: Production recordings

2.4.1 Method

2.4.1.1 Participants

Participants were the same as the participants in the production training.

2.4.1.2 Prompts

Twenty minimal-pair words and additional 20 non-minimal-pair monosyllabic words containing /r/ and /l/ word-initially were selected as a total of 40 prompts to be presented visually and aurally for the production recordings. The minimal-pair words were the same as those used by Iverson et al. (2005). A new NE speaker (male) recorded the auditory prompts in a sound-attenuated booth in the Phonetics Laboratory at the University of Calgary. All of the recorded prompts were sampled at 44,100 Hz, and peak

amplitudes of each prompt were normalized at 70 dB using Praat (Boersma & Weenink, 2009). The auditory prompts were provided to participants in order to ensure that they knew how to pronounce the entire word before the recording began. Additionally, 40 slides were prepared as visual prompts, and each slide showed the orthographic representation of the word to be produced in the center. Each slide was presented to the Japanese learners using a computer, followed by the presentation of the corresponding auditory prompt, before the recording began. The list of the prompts is provided in Appendix A.

2.4.1.3 Procedure

The Japanese learners made production recordings individually in the sound attenuated booth in the Speech Research Laboratory at the University of Victoria. They were asked to articulate the prompt words after they were presented with the visual and auditory prompts. In each recording, they saw an orthographic representation of the word to be produced on the computer screen and heard the word through speakers while looking at the slide. They were allowed to listen to the word twice if necessary. All utterances were digitally recorded and sampled at 44,100 Hz using Audacity (Audacity Team, 2012). Peak amplitudes of each recorded utterance were normalized at 70 dB using Praat (Boersma & Weenink, 2009). Prompts and procedures were identical for the pre-test and the post-test.

2.4.2 Analysis

In order to assess changes in the Japanese learners' productions, acoustic measurements were made for the initial segments in each word produced before and after the training. Of a total of 800 utterances from the pre- and post-training recordings (40 prompts × 10 learners × 2 recording conditions), 46 utterances (23 utterances from the pre-training recordings and 23 utterances from the post-training recordings) were excluded from the analysis because the onset consonants were either missing or pronounced as stop consonants, in which formant frequencies were absent, in these utterances. The F2 and F3 frequency values of the initial segments in each of the remaining 754 utterances were measured by taking the average F2 and F3 values for the steady state of the segment, in which the formants are relatively stable from the beginning of the segment to the beginning of transition from the segment to the following vowel, using Praat (Boersma & Weenink, 2009).

For normative data, speech samples of five NE speakers (two male and three female), who were undergraduate or graduate students at the University of Calgary, were collected. They articulated the 40 prompt words used for the pre- and post-training recordings individually. The recordings were made in the Phonetics Laboratory at the University of Calgary. All of the recorded utterances were sampled at 44,100 Hz, and peak amplitudes of each utterance were normalized at 70 dB using Praat (Boersma & Weenink, 2009). The F2 and F3 frequency values of the initial segments for each of the 200 utterances (40 words × 5 speakers) were measured using Praat (Boersma & Weenink, 2009).

In order to eliminate individual differences in formant frequency due to anatomical or physiological differences across the native Japanese (NJ) learners and the NE speakers, all formant frequency measurements were normalized for each learner and for each speaker, using the following z-score transformation formula (Lobanov, 1971).

$$F_i^N = (F_i - \mu_i) / \sigma_i$$

In this equation, *i* denotes the formant number (e.g., 1 = Formant 1), F_i denotes an individual formant frequency measurement for a talker, F_i^N denotes the measurement transformed to a z-score, μ_i denotes the average formant frequency value across all /r/ and /l/ productions for the talker, and σ_i denotes the standard deviation of the formant frequency value for that talker. This normalization method has been shown to be effective in eliminating formant differences across genders, which most likely derive from differences in vocal tract length, while retaining phonemic (and also regional) variation (Adank, Smits, & van Hout, 2004).

To illustrate how this method works, I will demonstrate formant transformations for tokens from two of the NE speakers (NE1 and NE2) as an example. For NE1 (male), the averaged F3 frequency all across his productions of /r/ and /l/ (μ_3) and the standard deviation for the μ_3 (σ_3) are 2135.95 Hz and 758.14 Hz respectively. A normalized measurement (a z-score) for the F3 frequency for his token of /r/ in *rack* (F_3 = 1518 Hz) is calculated as below:

$$F_3^N = (1518 \text{ Hz} - 2135.95 \text{ Hz}) / 758.14 \text{ Hz} = -0.82$$

Likewise, a normalized measurement (a z-score) for the F3 frequency for his token of /l/ in *lack* (F_3 = 3186 Hz) is calculated as below:

$$F_3^N = (3186 \text{ Hz} - 2135.95 \text{ Hz}) / 758.14 \text{ Hz} = 1.39$$

For NE2 (female), the averaged F3 frequency all across her productions of /r/ and /l/ (μ_3) and the standard deviation for the μ_3(σ_3) are 2369.38 Hz and 717.47 Hz respectively. A normalized measurement (a z-score) for the F3 frequency for her token of /r/ in *rack* (F_3 = 1794 Hz) is calculated as below:

$$F_3^N = (1794 \text{ Hz} - 2369.38 \text{ Hz}) / 717.47 \text{ Hz} = -0.80$$

Likewise, a normalized measurement (a z-score) for the F3 frequency for her token of /l/ in *lack* (F_3 = 2892 Hz) is calculated as below:

$$F_3^N = (2892 \text{ Hz} - 2369.38 \text{ Hz}) / 717.47 \text{ Hz} = 0.73$$

The differences in normalized measurements between the phonemes for each speaker are attributed to across-phoneme variation. The differences in normalized measurement between the speakers for each phoneme is attributed to within-phoneme (and possibly regional) variation.

After normalization, mean F2 and F3 frequencies within phoneme were calculated for each NJ learner and averaged across the learners. Likewise, mean F2 and F3 frequencies within phoneme were calculated for each NE speaker and averaged across the speakers.

2.4.3 Results

2.4.3.1 Acoustic analysis

For statistical tests for the acoustic analysis, an alpha level of .05 (two-tailed) was used. Tables of descriptive statistics for original acoustic measurements are provided in Appendix C and Appendix D for reference.

Figure 3.3 displays results of F2 measurements collapsed across words. For the NJ learners' productions, there was a large decline in the mean F2 for /r/ from -0.24 (*SD* = 0.49) at pre-test to -0.52 (*SD* = 0.32) at post-test. Similarly, the mean F2 for /l/ largely declined from 0.59 (*SD* = 0.59) at pre-test to 0.23 (*SD* = 0.42) at post-test. On the other hand, for /r/, the mean F2 frequency for the NE speakers' productions as control (*M* = -0.05, *SD* = 0.28) was higher than the mean F2 frequencies for the NJ learners' productions at pre-test and post-test. Moreover, for /l/, the mean F2 frequency for the NE speakers' productions (*M* = 0.05, *SD* = 0.28) was lower than the mean F2 frequencies for the NJ learners' productions at pre-test and post-test. The small difference in mean F2 between the phonemes for the NE speakers' productions is consistent with the previous observations (Dalston, 1974; Iverson et al., 2005; Lotto et al., 2004; O'Conner, et al., 1957).

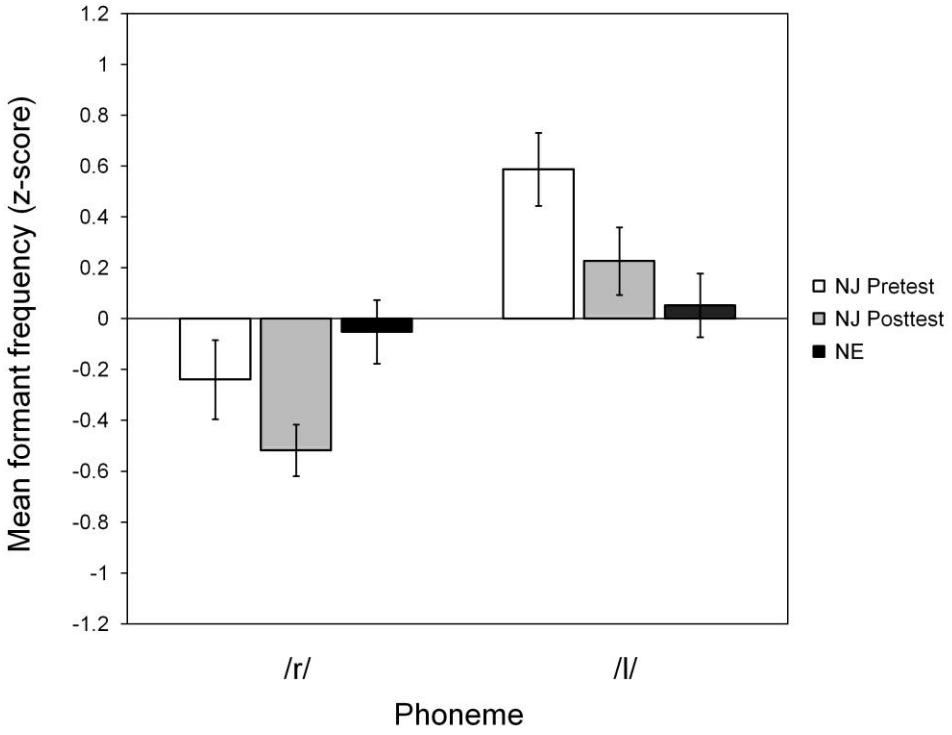

Figure 2.1. F2 frequencies for the NJ learners' productions at pre-test and at post-test and the NE speakers' productions as a function of phoneme. Frequency values were converted into z-scores for normalization. Error bars represent standard errors.

A two-way repeated ANOVA with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors was performed in order to examine the observed differences in F2 between the testing sessions for the NJ learners' productions. The main effect of phoneme was significant, $F(1, 9) = 21.79$, $p = .001$. However, there was no significant main effect of testing session, $F(1, 9) = 3.17$, $p = .109$, nor interaction of phoneme and testing session, $F(1, 9) = 0.14$, $p = .715$. This indicates that the F2 for /l/ was higher than the F2 for /r/ for the NJ learners' productions, regardless of the testing session.

In order to examine whether the F2 frequencies for the NJ learners' productions of /r/ and /l/ at pre-test and post-test significantly differed from the F2 frequencies for the NE speakers' productions of the same phonemes, Mann-Whitney tests were performed across language groups (pre-test NJ vs. NE, post-test NJ vs. NE) for each phoneme. The Mann-Whitney test, a non-parametric test for between-subjects analyses, was used because the NJ and NE groups were small and unequal in sample size (n = 10 for NJ, n = 5 for NE). For /r/, the F2 for the NJ groups' production at pre-test was not significantly lower than the F2 for the NE group's production, U = 22.00, z = -0.37, p = .768. However, the F2 for the NJ group's production for /r/ at post-test was significantly lower than the F2 for the NE group's production, although this difference was marginal, U = 9.00, z = -0.961, p = .052. On the other hand, for /l/, the F2 for the NJ group's production at pre-test was significantly higher than the F2 for the NE group's production, U = 8.00, z = -2.08, p = .04. However, the F2 for the NJ group's production at post-test was not significantly higher than the F2 for the NE group's production, U = 20.00, z = -0.61, p = .594. Therefore, the analysis suggests that F2 for the NJ groups' production became lower than F2 for the NE group's production after the training for /r/, whereas F2 for the NJ group's production became similar to F2 for the NE group's production after the training for /l/. In other words, the F2 difference between the phonemes produced by the NJ group did not become reduced as much as the F2 difference between the phonemes produced by the NE group, although for /l/, F2 was moving in the right direction.

Results of F3 measurements collapsed across words are displayed in Figure 2.2. For the NJ group's production, the mean F3 for /r/ declined from -0.65 (SD = 0.50) at pre-test to -0.74 (SD = 0.46) at post-test, whereas the mean F3 for /l/ showed negligible

decline from 0.73 (*SD* = 0.22) at pre-test to 0.72 (*SD* = 0.34) at post-test. On the other hand, the mean F3 for the NE group's production (M = -0.96, SD = 0.02) was lower than the mean F3 for the NJ group's production at pre-test and post-test for /r/, whereas the mean F3 for the NE group's production (M = 0.96, SD = 0.03) was higher than the mean F3 for the NJ group's production at pre-test and post-test for /l/. Note that the small standard deviation values for the NE group's F3 frequencies for the phonemes indicate that the NE speakers' productions of /r/ and /l/ were consistent in terms of F3. That is, when producing /r/, the NE speakers consistently make a tongue constriction in the palatal area and a lip constriction (rounded or unrounded), which lowers F3 frequencies for /r/ (Espy-Wilson et al., 2000).

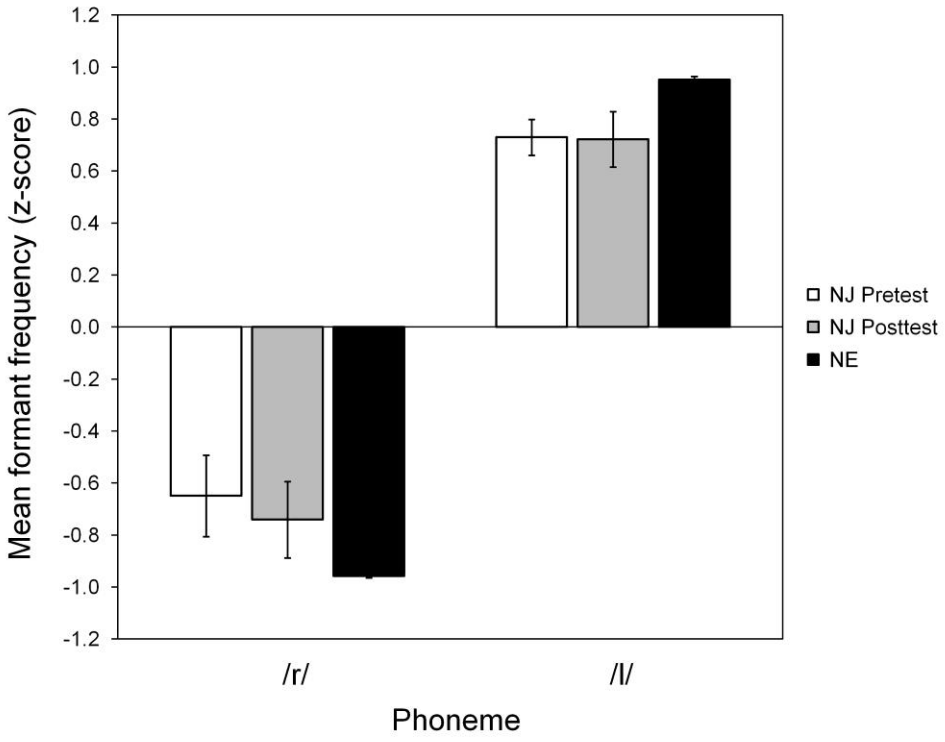

Figure 2.2. F3 frequencies for the NJ learners' productions at pre-test and at post-test and the NE speakers' productions as a function of phoneme. Frequency values were converted into z-scores for normalization. Error bars represent standard errors.

A two-way repeated ANOVA with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors was performed in order to examine the observed differences in F3 between the testing sessions for the NJ learners' productions. The analysis revealed that the main effect of phoneme was significant, $F(1, 9) = 70.32$, $p < .001$. On the other hand, the main effect of testing session and the interaction of phoneme and testing session were not significant, $F(1, 9) = 0.12$, $p = .734$ for testing session, $F(1, 9) = 0.13$, $p = .723$ for phoneme and testing session. Thus, similarly to the difference in

F2 between the phonemes, the analysis indicates that the F3 for /l/ was higher than the F3 for /r/ for the NJ learners' productions, regardless of the testing session.

Mann-Whitney tests were performed across language groups (pre-test NJ vs. NE, post-test NJ vs. NE) for each phoneme. For /r/, the difference in F3 between the NJ group's production at pre-test and the NE group's production was marginally significant, $U = 9.50$, $z = -1.90$, $p = .06$, whereas the F3 for the NJ group's production at post-test was not significantly higher than the F3 for the NE group's production, $U = 15.00$, $z = -1.23$, $p = 0.254$. For /l/, the F3 for the NJ group's production at pre-test was not significantly lower than the F3 for the NE group's production, $U = 11.00$, $z = -1.72$, $p = .099$. Likewise, the F3 for the NJ group's production at post-test was not significantly lower than the F3 for the NE group's production, $U = 13.00$, $z = -1.47$, $p = .165$. The analysis suggests that the NJ group's F3 for /r/ became similar to the NE group's F3 after the training. Moreover, the NJ group's F3 was similar to the NE group's F3 for /l/ before and after the training.

Figure 2.3 displays distributions of /r/ and /l/ productions from the NE group in a F2 × F3 space. The space clearly separates two distributions along the F3 axis, with productions of /l/ in the positive plane and productions or /r/ in the negative plane. Moreover, the /l/ productions are more scattered whereas the /r/ productions are more clustered in terms of F3. This reflects the large difference in mean F3 and the difference in error bar size between the phonemes (see Figure 2.2). On the other hand, the distributions for /r/ and /l/ greatly overlap along the F2 axis. In addition, F2 frequencies for the /r/ and /l/ productions vary substantially, suggesting that those phonemes may not be reliably discriminated by F2 differences alone. These observations are consistent with

the small difference in mean F2 between the phonemes and the overlapping error bars (see Figure 2.1). These distribution patterns generally apply to all the NE speakers.

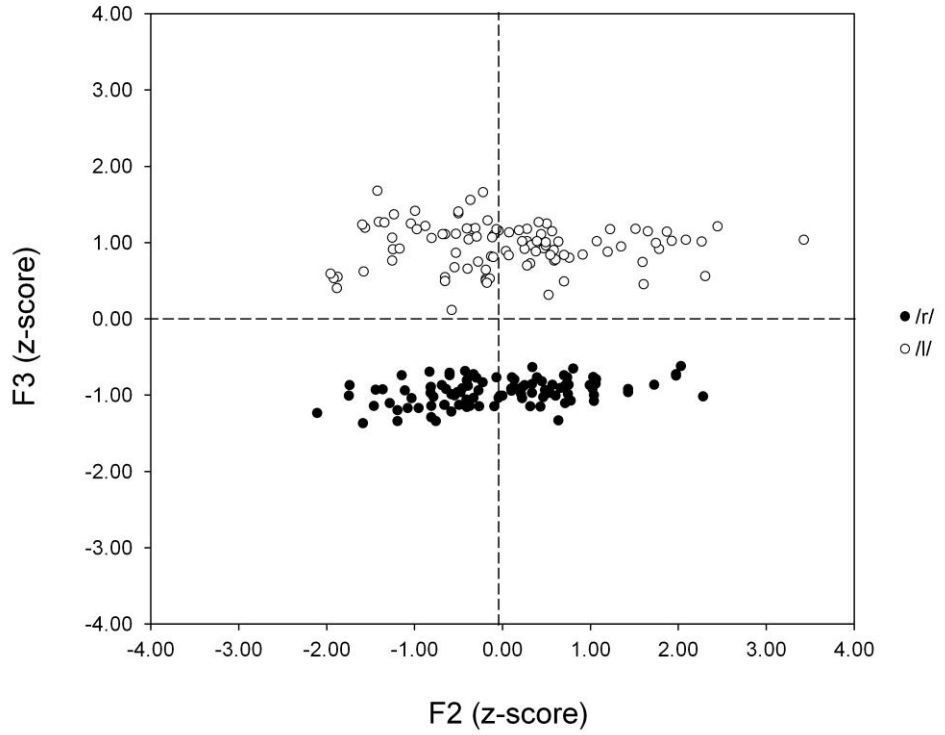

Figure 2.3. Scatter plot of F2 and F3 frequencies for the NE speakers' productions of /r/ and /l/. Frequency values were converted into z-scores for normalization.

Distributions of /r/ and /l/ productions from the NJ group at pre-test differ substantially from these from the NE group, as Figure 2.4 displays. Comparison of this figure and the previous figure show differences in production between the two language groups in terms of F2 and F3 and greater variability and overlap in the NJ group's production. The two distributions overlap and do not show clear separation along the F3 axis. Moreover, the distributions spread into both positive and negative planes although the centers of them stay in opposite planes. F3 frequencies for productions of the

phonemes vary substantially, and this tendency is greater for /r/. This may reflect the high mean F3 relative to the NE group's mean F3 (see Figure 2.2). The distributions show large overlap in both positive and negative planes along the F2 axis as well. Further, F2 frequencies greatly vary for both phonemes. The center of the /l/ distribution is placed in the positive plane, whereas the center of the /r/ distribution appears to be placed in the negative plane. Recall that the mean F2 for /l/ at pre-test was higher than the NE group's mean F2 (see Figure 2.1). This difference appears to be supported by the heavy clustering of /l/ productions in the positive plane.

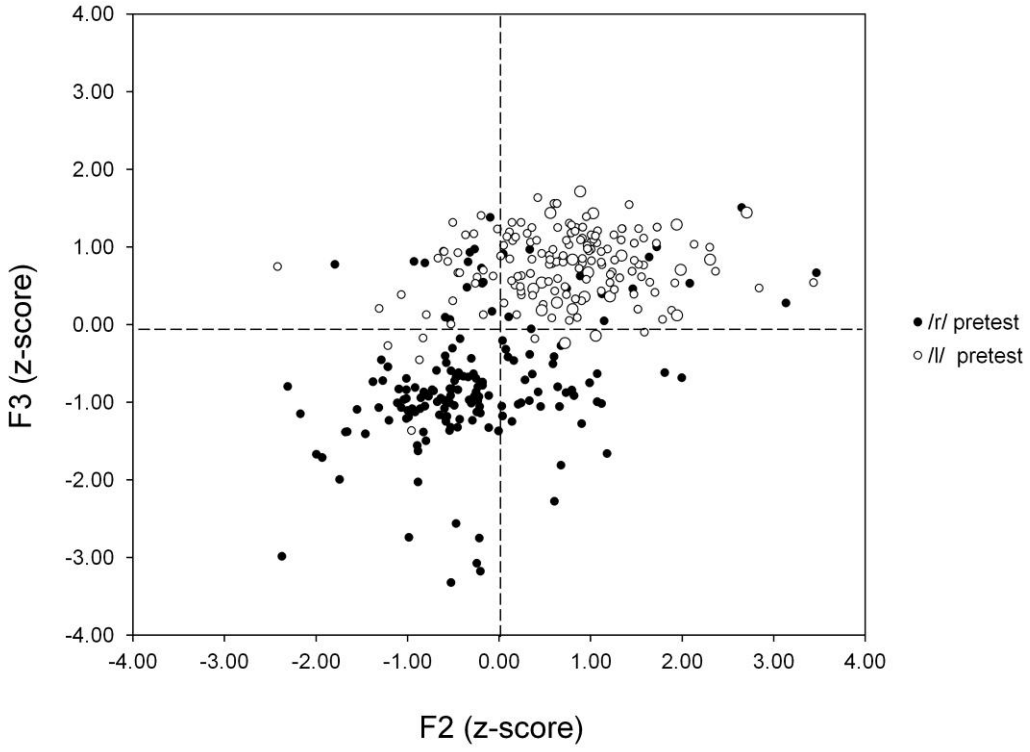

Figure 2.4. Scatter plot of F2 and F3 frequencies for the NJ learners' productions of /r/ and /l/ at pre-test. Frequency values were converted into z-scores for normalization.

Distributions of /r/ and /l/ productions from the NJ group at post-test are somewhat different, as shown in Figure 2.5. Although there is still overlap between their distributions along the F3 axis, the /r/ productions become more clustered in the negative plane whereas the /l/ productions become more clustered in the positive plane. Although the distributions overlap along the F2 axis as well, the /l/ distribution appears to be somewhat shifted in the negative direction. Further, more /r/ productions become clustered in the negative plane, reflecting the NJ group's low mean F2 for /r/ relative to the NE group's mean F2 for /r/ (see Figure 2.1).

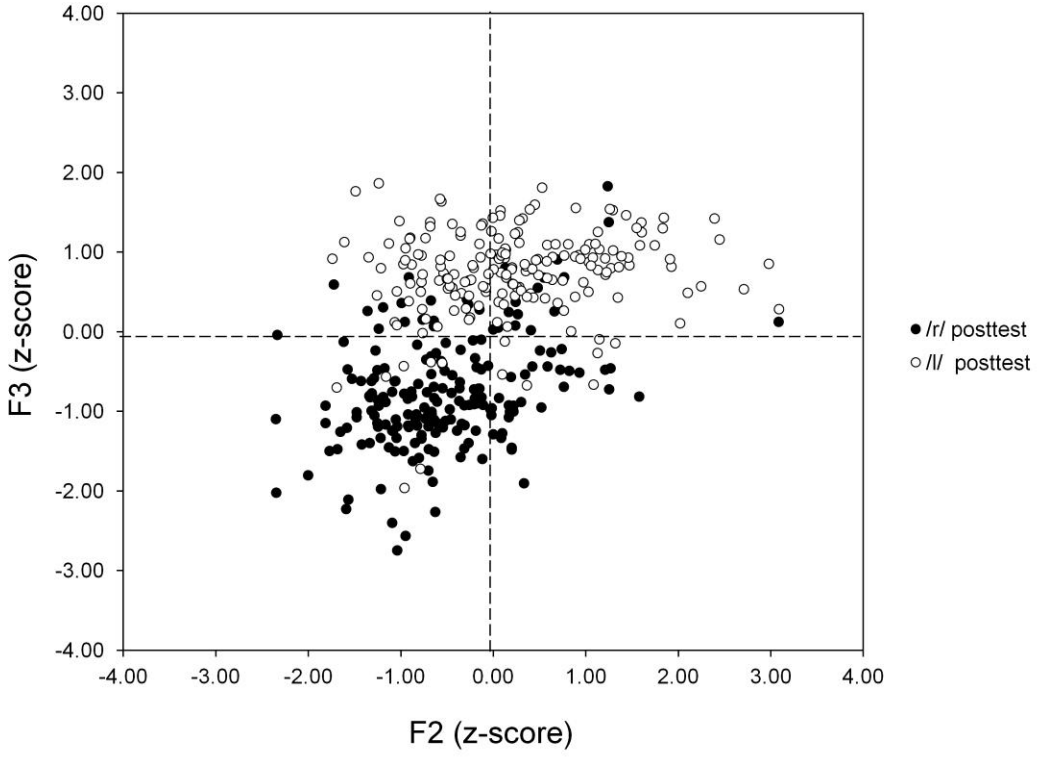

Figure 2.5. Scatter plot of F2 and F3 frequencies for the NJ learners' productions of /r/ and /l/ at post-test. Frequency values were converted into z-scores for normalization.

2.5 Experiment 2: Perception tests

2.5.1 Method

2.5.1.1 Participants

Participants were the same as the participants in the production training.

2.5.1.2 Stimuli

Sixty sets of minimal-pair monosyllabic English words which contrast /r/ and /l/ word-initially were selected as auditory stimuli for the perceptual tests (120 words in total). None of the words were used for the production recordings or the production training. The minimal-pair words are the same as those used by Iverson et al. (2005) except two, which do not constitute a minimal-pair in North American English. The male NE speaker who recorded the auditory prompts for the production recordings and a female NE speaker individually recorded the stimuli in a sound-attenuated booth in the Phonetics Laboratory at the University of Calgary. A total of 240 stimuli (120 words × 2 speakers) were sampled at 44,100 Hz, and peak amplitudes of each stimulus were normalized at 70 dB using Praat (Boersma & Weenink, 2009). The stimuli were divided into two sets, and each set contained 120 stimuli, comprising 60 words produced by the male speaker and the other 60 words produced by the female speaker. That is, Set 1 included Pairs 1 to 30 produced by the male speaker and Pairs 31 to 60 produced by the female speaker. Set 2 included Pairs 1 to 30 produced by the female speaker and Pairs 31 to 60 produced by the male speaker. Each participant was randomly assigned to either of the stimulus sets. The list of the stimulus words is provided in Appendix B.

2.5.1.3 Procedure

Immediately before the recordings outlined in Section 2.4 above, NJ learners underwent perception tests individually in the sound-attenuated booth in the Speech Research Laboratory at the University of Victoria. At the beginning of each trial, orthographic representations of two words from a minimal-pair (e.g., *right* and *light*) were displayed on the computer screen. One of the words from the pair was positioned at the bottom right, and the other was positioned at the bottom left. While seeing the pair words on the screen, learners heard one of the words over headphones and were asked to select the word that they thought they had heard by pressing a key corresponding to the word. Before the test, the learners completed a practice block of two trials in order to gain familiarity with the task. No feedback on the learners' responses was provided in the test trials and practice trials. The test comprised two blocks, and each block comprised 60 trials (2 blocks × 60 trials = 120 trials). Stimulus words starting with /r/ were presented on the right, and stimulus words starting with /l/ were presented on the left on the computer screen. Each stimulus was presented only once. Presentation order was randomized within block and across learners, and the test lasted approximately 10 minutes. Stimuli and procedures were identical for the pre-test and the post-test.

2.5.2 Analysis

In order to assess changes in the NJ learners' perception, correct identification percentages for /r/ and /l/ were first calculated for each learner for each testing session (pre-test and post-test). Next, the percentages of correct identification were averaged across learners for each testing session.

Moreover, changes in the learners' perceptual sensitivity to the contrast between the phonemes were assessed using d' (d-prime), a measure of sensitivity used in Signal Detection Theory (Green & Swets, 1966; Macmillan & Creelman, 2005). d' values were calculated using the following formula:

$$d' = z(H) - z(F)$$

In this equation, $z(H)$ denotes the proportion of hit responses (i.e., correctly identifying /r/ tokens) transformed to a z-score, and $z(F)$ denotes the proportion of false-alarm responses (i.e., incorrectly identifying /l/ tokens as /r/ tokens) transformed to a z-score. The d' value of zero indicates the lack of sensitivity to the contrast between /r/ and /l/ (i.e., the learner is unable to discriminate the phonemes). d' values were calculated for each learner for each testing session and were subsequently averaged across learners for each testing session.

Additionally, changes in the learners' response bias were assessed using c (criterion location), a measure of response bias used in Signal Detection Theory (Green & Swets, 1966; Macmillan & Creelman, 2005). c values were calculated using the following formula:

$$c = -0.5[z(H) + z(F)]$$

As in the previous formula for d' calculations, $z(H)$ denotes the proportion of hit responses (i.e., correctly identifying /r/ tokens) transformed to a z-score, and $z(F)$ denotes the proportion of false-alarm responses (i.e., incorrectly identifying /l/ tokens as /r/ tokens) transformed to a z-score. Negative c values indicate the learners' response bias toward /r/, and positive c values indicate the learners' response bias toward /l/. The c value of zero indicates unbiased responses. c values were calculated for each learner for

each testing session and were subsequently averaged across learners for each testing session.

2.5.3 Results

2.5.3.1 Perceptual accuracy

For all statistical tests for the perception data analysis, an alpha level of .05 (two-tailed) was used. As Figure 2.6 shows, the mean percent correct identification of /r/ for the NJ learners declined from 66.50 (*SD* = 11.29) at pre-test to 61.50 (*SD* = 11.34) at post-test. On the other hand, the mean percent correct identification of /l/ increased from 55.83 (*SD* = 9.85) at pre-test to 60.67 (*SD* = 12.20) at post-test.

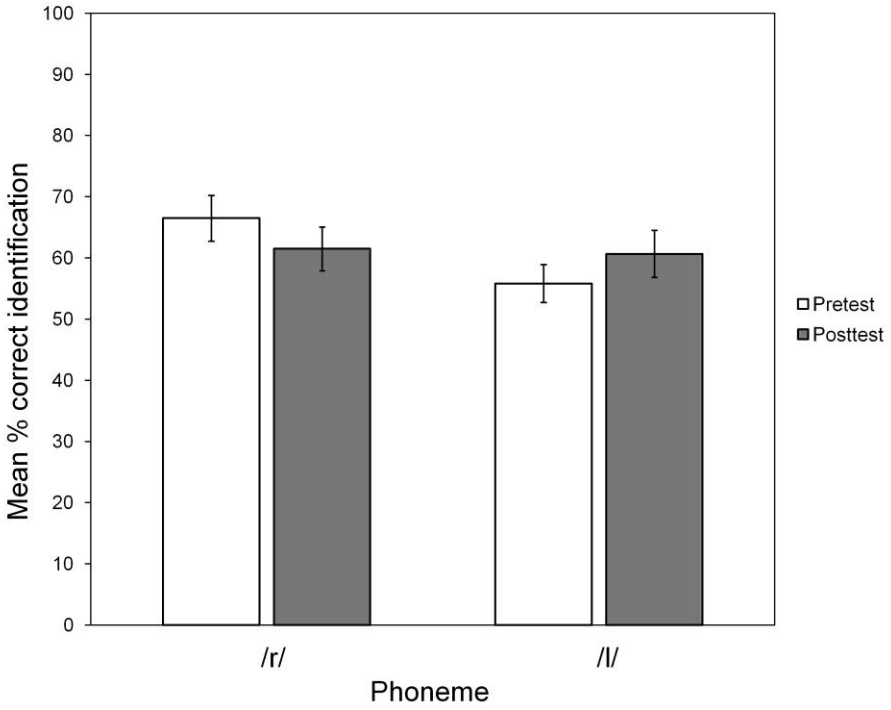

Figure 2.6. Percentages of correct identification scores for the NJ learners in the perceptual tests as a function of phoneme and testing session. Error bars represent standard errors.

In order to examine the observed changes, a 2 × 2 × 2 mixed ANOVA was conducted with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors, as well as stimulus set (Set 1, Set 2) as a between-subject factor. The stimulus set was included as a factor in the analysis in order to examine whether particular combinations of the NE talkers and stimulus words influenced the learners' perception. The analysis revealed no significant main effects of phoneme, $F(1, 8) = 1.70$, $p = .229$, testing session, $F(1, 8) = 0.001$, $p = .98$, or stimulus set, $F(1, 8) = 0.15$, $p = .71$. There were no significant interaction effects of 1) phoneme and stimulus set, $F(1, 8) = 0.49$, $p = .51$, 2) testing session and stimulus set, $F = 0.36$, $p = .564$, or 3) phoneme, testing session and stimulus set, $F(1, 8) = 0.04$, $p = .843$. However, an interaction of phoneme and testing session was significant, $F(1, 8) = 5.87$, $p = .042$. A simple effect analysis revealed that there was a marginally significant difference in identification accuracy between the phonemes at pre-test, $F(1, 8) = 5.07$, $p = .054$. However, the difference in identification accuracy between the phonemes at post-test was not significant, $F(1, 8) = 0.03$, $p = .871$. This indicates that although the NJ learners were more likely to identify /r/ than /l/ before the training, this tendency disappeared after the training. In addition, there was no significant effect of testing session on phoneme, indicating that the level of perceptual accuracy did not change after the training, regardless of the phoneme type.

2.5.3.2 Perceptual sensitivity and response bias

The mean d' value showed negligible decline from 0.60 ($SD = 0.42$) at pre-test to 0.59 ($SD = 0.51$) at post-test. In order to examine the observed change, a paired samples t-test was conducted with testing session as the within-subject factor. The difference in

perceptual sensitivity (d') between the testing sessions was not significant, $t(9) = 0.01$, $p = .996$, indicating that the perceptual sensitivity to the phoneme contrast did not improve significantly after the training.

The mean c value increased from -0.15 ($SD = 0.19$) at pre-test to -0.01 ($SD = 0.21$) at post-test. Note that negative values indicate response bias toward /r/. In order to examine the observed change, a paired samples t-test was conducted with testing session as the within-subject factor. The difference in response bias (c) was significant, $t(9) = -2.50$, $p = .034$. Therefore, the result suggests that the learners' bias to select /r/ became significantly reduced after the training.

2.6 Discussion

Although the F3 for the NJ learners' productions of /r/ and /l/ did not significantly change over time, it became close to that for the NE speakers' productions of the phonemes after the training. The lowered F3 which is a major characteristic for /r/ is related to a tongue constriction near the palate and a lip constriction (Espy-Wilson et al., 2000). Thus, the observed change in F3 suggests that the NJ learners became able to make the tongue and lip gestures for /r/ which are similar to these of the NE speakers. However, the difference in mapped tokens between the two language groups indicates that the NJ learners' productions of /r/ and /l/ were still not fully distinct on the F3 continuum despite this improvement.

Despite the learner's improvement in production in terms of F3 and the corresponding articulatory gestures, the learners' perceptual abilities to discriminate between the phonemes did not change significantly over time. At the same time, their

tendency to provide /r/ responses became significantly reduced after the training, although their sensitivity to the contrast between the phonemes did not improve. Therefore, the results indicate that the production training helped the learners to reduce their response bias, although it did not lead to improved perception of the contrast between /r/ and /l/. It is possible that the learners' subjective familiarity with the stimulus words (e.g., having heard or said the word frequently, or having never heard or said the word) affected their performance in the perceptual tests, as native Japanese listeners are more likely to misidentify English /r/ and /l/ in words that are less familiar to them (Flege, Takagi, & Mann, 1996). Nevertheless, the production training might have not helped the learners to improve their perception of the phoneme contrast.

Chapter Three: Perceptual evaluation of production by English listeners

3.1 Introduction

As presented in Chapter 2, native Japanese (NJ) learners of English articulated 40 English words starting with /r/ or /l/ before and after the training. In addition to the acoustic analyses of the NJ learners' productions of English /r/ and /l/, the intelligibility and goodness of each production were evaluated by native English (NE) listeners. To this end, NE listeners performed a phoneme identification task and a goodness rating task. In the phoneme identification task, NE listeners were asked to identify sounds in the initial segments of each word produced by a NJ learner. In the goodness rating task, NE listeners were asked to rate how good the production of the segment was as an example of the sound category (/r/ or /l/) which they selected for the segment. It has been shown that NE listeners are more likely to identify /r/ and /l/ accurately for Japanese productions of the phonemes, if the productions are judged as better articulated (Bradlow et al., 1997). Moreover, it has also been shown that an increase in identification accuracy is more likely to be accompanied by an increase in goodness rating (Hattori, 2009; Hazen, Senema, Iba, & Faulkner, 2005), which may reflect an improvement in comprehensibility more than a reduction of accentedness because of the close relationship between intelligibility and comprehensibility (e.g., Munro & Derwing, 1995a). Therefore, it was predicted that changes in identification accuracy and goodness rating by NE listeners would be indicative of changes in the NJ learners' articulations of the phonemes. It was hypothesized that NE listeners would correctly identify the NJ learners' productions of /r/ and /l/ after the training more frequently and rate them higher if the learners improved their productions of the phonemes. Additionally, correlational analyses for the production

intelligibility and the NJ learners' perceptual accuracy were performed in order to explore the relationship between production and perception (presented in Chapter 4). An alpha level of .05 (2-tailed) was used for all statistical tests.

3.2 Method

3.2.1 NE listeners

Listeners were three phonetically trained native English speakers, who were undergraduate students at the University of Calgary. They performed the phoneme identification task and the goodness rating task individually as volunteers in a testing room in the Phonetics Laboratory at the University of Calgary. In order to avoid influence of knowledge about this research on the listeners' judgments, they were not provided with detailed information about the data and the purpose of the research. They were told that they would evaluate English words spoken by non-native speakers of English.

3.2.2 Stimuli

Stimuli were a total of 800 utterances from the pre- and post-training recordings of the NJ learners (40 prompts × 10 learners × 2 testing sessions). These utterances were randomly mixed across testing conditions within learner, and the presentation order was randomized across the learners. Due to a technical issue, peak-amplitudes of each utterance were normalized again at 65 dB using Praat (Boersma & Weenink, 2009). Each listener evaluated all utterances over a two week period at their own pace.

3.2.3 Procedure

3.2.3.1 Phoneme identification task

In this task, listeners heard the recorded utterances from the NJ learners, one at a time, and were asked to identify sounds in the initial segments of each utterance. In each trial, they saw the incomplete orthographic representation of a word (the spelling of the word without the initial segment) on a computer screen while listening to an utterance of the word from a NJ leaner. They were asked to select one out of a set of sound categories (/r/, /l/, /d/, /b/, /t/, /w/, the alveolar tap /ɾ/, and *others*) displayed on the screen for the missing segment. They were allowed to listen to the utterance again if necessary. If the listeners selected *others*, they were asked to describe the sound by typing in a description in a dialog box displayed on the screen. The orthographic representations were provided because some utterances were difficult to segment. In order to avoid influence of the orthography on the listeners' judgments, they were told that they might not hear real English words. Also, they were instructed to identify the sound they thought they had actually heard, not the sound which was supposed to fill in the missing segment.

3.2.3.2 Goodness rating task

If the listeners identified the initial segment as /r/ or /l/ in the phoneme identification task, regardless of whether the response was correct or not, they were subsequently prompted to rate how good the sound was as an example of the selected sound category. They saw a 7-point scale ranging from 1 (*bad*) to 7 (*good*) on the computer screen and were asked to select a point on the scale by pressing an on-screen button corresponding to the point.

They were allowed to listen to the utterance again if necessary. They were encouraged to use the entire scale when rating the sounds.

3.3 Analysis

In order to analyse the data from the phoneme identification task, correct identification percentages for /r/ and /l/ within testing session was first calculated for each NE listener for each NJ learner. Next, the percentages of identification were averaged across NE listeners for each NJ learner and subsequently averaged across NJ learners.

In order to analyse the data from the goodness rating task, rating scores for correctly identified segments were used. Mean rating scores for /r/ or /l/ within testing session were first calculated for each NE listener for each NJ learner. Next, the mean rating scores were averaged across NE listeners for each NJ learner and subsequently averaged across NJ learners.

3.4 Results

3.4.1 Phoneme identification task

Figure 3.1 displays mean percent intelligibility scores for the NJ learners' productions of English /r/ and /l/ judged by the NE listeners. The mean intelligibility score for /r/ increased slightly from 74.17 (*SD* = 21.24) at pre-test to 75.33 (*SD* = 27.53) at post-test. There was a greater increase in the mean intelligibility score for /l/ from 75.83 (*SD* = 19.01) at pre-test to 89.00 (*SD* = 18.99) at post-test.

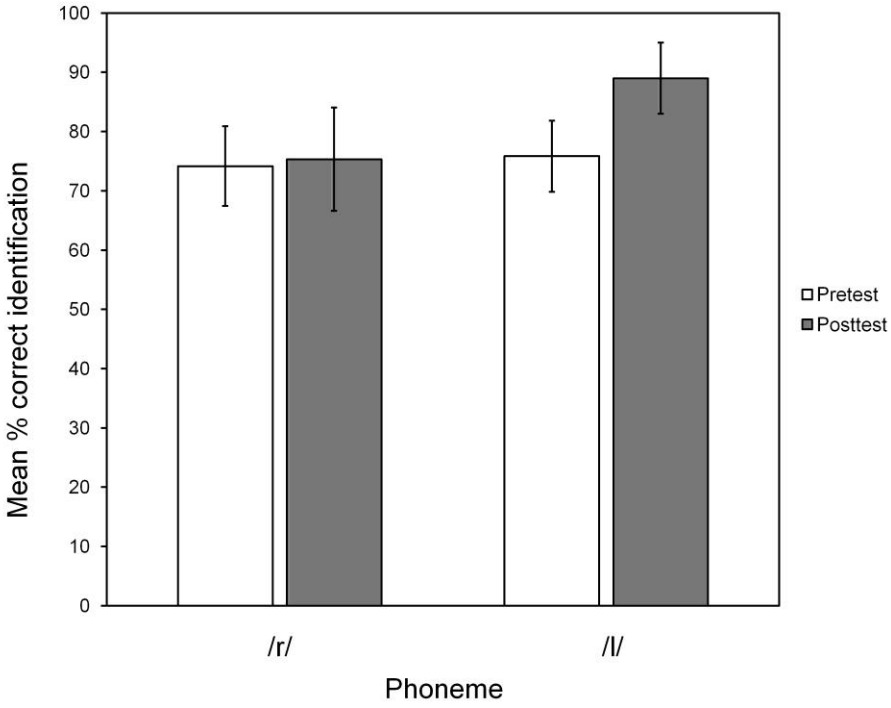

Figure 3.1. Percentages of intelligibility scores for the NJ learners' productions judged by the NE listeners in the intelligibility judgment task as a function of phoneme and testing session. Error bars represent standard errors.

In order to examine the observed increases in production intelligibility, a two-way repeated-measures ANOVA was performed with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors. There were no significant main effects of phoneme, $F(1, 9) = 1.05$, $p = .333$, or testing session, $F(1, 9) = 2.74$, $p = .133$. Moreover, the interaction effect of phoneme and testing session was not significant, $F(1, 9) = 0.88$, $p = .374$. Although the analysis suggests that the training did not improve the intelligibility of the NJ learners' /r/ and /l/ productions, the lack of significance might be attributed to the large variability in the NE listeners' responses, as the standard deviation values indicate.

As displayed in Table 3.1 and Table 3.2, the NE listeners most frequently misidentified /r/ as /l/ for the NJ learners' productions at pre-test and at post-test, and this trend was greater for the productions at post-test.

Table 3.1. Confusion matrix of the NE listeners' responses in the intelligibility judgment task for the NJ learners' productions of English /r/ (in percent)

	Response								
Condition	/r/	/l/	/d/	/b/	/t/	/w/	/ɾ/	Other	Total
Pretest	**74.17**	7.83	2.83	4.67	0.50	0.00	1.50	8.50	100.00
Posttest	**75.33**	11.17	0.33	5.50	0.00	0.00	0.33	7.33	100.00

Table 3.2. Confusion matrix of the NE listeners' responses in the intelligibility judgment task for the NJ learners' productions of English /l/ (in percent)

	Response								
Condition	/r/	/l/	/d/	/b/	/t/	/w/	/ɾ/	Other	Total
Pretest	3.00	**75.83**	9.50	5.67	0.33	0.17	1.00	4.50	100.00
Posttest	5.67	**89.00**	0.83	2.50	0.00	0.00	0.17	1.83	100.00

The NE listeners also selected /d/, /b/, /t/, /ɾ/, and *other* when they misidentified /r/ for the productions of /r/ at pre-test. The listeners selected /d/, /t/, /ɾ/, and *other* less frequently whereas they selected /b/ more frequently for the productions of /r/ at post-test. This suggests that when the NJ learners mispronounced /r/ before the training, they were more likely to make direct contact between the tongue tip and the alveolar ridge. In contrast, it appears that the learners were more likely to make the contact appropriate for /l/ or protrude the lips too much when they mispronounced /r/ after the training.

On the other hand, /l/ was most frequently misidentified as /d/ for the productions at pre-test and as /r/ for the productions at post-test. Moreover, the listeners were less likely to misidentify /l/ as /d/, /b/, /t/, /ɾ/ and *other* for the productions at post-test. This suggests that when the learners mispronounced /l/ before the training, the contact between the tongue-tip and the alveolar ridge was inappropriately made for the phoneme in terms of timing and manner. However, it appears that the learners became able to make the appropriate midline contact for /l/ after the training.

3.4.2 MRT and its relationship to production intelligibility

NJ learners completed the MRT, which measures their spatial processing abilities. A high score on this test represents a high spatial processing ability. The percentage of correct responses on the MRT was calculated for each NJ learner (M = 50.50, SD = 22.17). In order to analyze whether variation in response accuracy level on the MRT correlates with variation in production intelligibility at post-test, Pearson's correlation coefficients and their levels of statistical significance were calculated for each phoneme. For /r/, there was no significant relationship between the learners' response accuracy level on the MRT and production intelligibility, r = -.46, p = .183. Likewise, there was no significant relationship between the learners' response accuracy level on the MRT and production intelligibility for /l/, r = .18, p = .616. Thus, it appears that having a higher spatial processing ability does not necessarily indicate better production intelligibility for /r/ and /l/. This might suggest that a higher spatial processing ability does not necessarily correspond to better understanding and use of the visual information in ultrasound images, although production intelligibility is not a direct measure of these skills.

3.4.3 Goodness rating task

Word-initial segments which the NE listeners correctly identified as /r/ or /l/ in the intelligibility task were further rated on a goodness rating scale for 1 (*bad*) to 7 (*good*). As Figure 3.2 shows, the mean rating score for /r/ slightly declined from 4.32 (*SD* = 0.75) at pre-test to 4.20 (*SD* = 0.66) at post-test. Likewise, the mean rating score for /l/ showed negligible decline from 4.79 (*SD* = 0.59) at pre-test to 4.73 (*SD* = 0.64) at post-test.

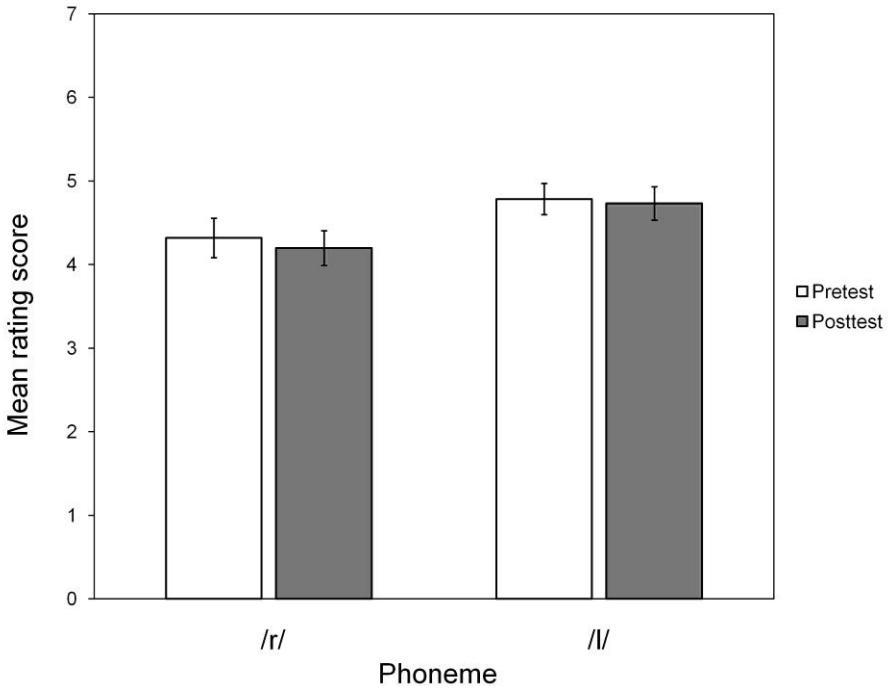

Figure 3.2. Rating scores for the NJ learners' productions judged by the NE listeners in the goodness rating task as a function of phoneme and testing session. Error bars represent standard errors.

A two-way repeated-measures ANOVA with phoneme (/r/, /l/) and testing session (pre-test, post-test) as within-subject factors revealed no significant main effects of

phoneme, $F(1, 9) = 3.18$, $p = .108$, or testing session, $F(1, 9) = 1.64$, $p = .233$. Further, the interaction of phoneme and testing session was not significant, $F(1, 9) = 0.07$, $p = .802$.

3.5 Discussion

Overall, intelligibility and goodness of the NJ learners' productions of /r/ and /l/ did not significantly improve after the training. Moreover, the MRT scores did not correlate with production intelligibility, which implies that having a higher spatial processing ability might not necessarily indicate that the learner can understand the visual information in ultrasound displays better and utilized them to improve his or her production.

However, the changes in production intelligibility might be masked by the large variability in the NE listeners' responses especially for /l/. In fact, the results of the acoustic analyses suggest that the NJ group's production of the phonemes became more distinctive on the F3 continuum after the training and more similar to the NE group's production of the same phonemes on the continuum. Because F3 is a crucial acoustic cue in discriminating between /r/ and /l/ (e.g., O'Conner et al., 1957), it might have been easier for the NE listeners to discriminate between the phonemes produced by the NJ learners after the training. Nevertheless, this potential improvement in production intelligibility for /l/ was not reflected in the production goodness for the same phoneme, whereas both intelligibility and goodness were improved for production of /r/ and /l/ in previous studies (Hattori, 2009; Hazen et al., 2005). This suggests that the NE listeners experienced the same level of difficulty when they tried to identify /l/ in the NJ learners'

utterances recorded before and after the training, although they identified the phoneme more frequently in the utterances recorded after the training.

Chapter Four: General discussion

4.1 Introduction

The goals of this study were to investigate: 1) whether the production training using ultrasound would improve productions of English /r/ and /l/ by Japanese learners of English in terms of F2 and F3; 2) whether the training would facilitate the accurate perception of /r/ and /l/ by the learners in the absence of perception training; and 3) whether the production training would improve the intelligibility of the learners' productions of English /r/ and /l/ for English listeners. The results indicate that the ultrasound production training potentially improved the Japanese learners' productions of /l/ although this did not lead to improved perception of the same phoneme in the learners. At the same time, the training helped to reduce the learners' tendency to provide /r/ responses without changing their perceptual abilities. Further, the data from individual learners demonstrate considerable variability in learning. Overall, the results of this study suggest that production learning did not have a direct impact on perceptual learning for these learners, although the evidence is not strong. Further, production learning and perception learning possibly take different courses.

In the following sections, the results are discussed in relation to previous research as well as models of speech perception and L2 speech learning. Finally, future directions and implications of the findings for L2 teaching are proposed.

4.2 Acoustic analysis

In the production training study, different results emerged depending on the speech sound. Although there was no significant change in F3 between the pre-test production

and the post-test production in the Japanese learners, the comparison between the two language groups (Japanese vs. English) after the training suggests that there were changes in the quality of the Japanese learners' productions of the phonemes. Because F3 is the primary cue for distinguishing between /r/ and /l/ for native English speakers, the phonemes produced by the Japanese learners after the training might have become as distinguishable as the phonemes produced by native English speakers. The low F3 for /r/ is related to a tongue constriction toward the palate and a lip constriction (Espy-Wilson et al., 2000). Thus, the Japanese learners might have improved in their ability to make these gestures more accurately after the training, although they might not have been able to do so as consistently as the native English speakers.

There was also no significant change in F2 between the pre-test production and the post-test production in the learners. However, the comparison between the language groups demonstrated that the F2 for the Japanese learners' productions of /r/ became too low when compared with the English speakers, whereas these groups did not differ by F2 for their productions of /l/ at post-test. Because lowered F2 corresponds to a tongue constriction made in the pharyngeal region in addition to a lip constriction (Delattre & Freeman, 1968; Johnson, 2003), the learners might have retracted the tongue excessively into the pharynx for the articulation of /r/. Interestingly, this suggests that the learners also made distinctions between the phonemes by retracting the tongue. That is, they retracted the tongue in the way which is not quite appropriate for /r/ when producing /r/, whereas they became able to retract the tongue appropriately for /l/ when producing /l/. Taken together, the acoustic analysis indicates that the learners' productions of /r/ became more distinct from their productions of /l/ but less native-like in terms of the

tongue retraction gesture, whereas their productions of /l/ became more native-like after the training.

4.3 Production learning

According to the acoustic analysis, the Japanese learners' productions of /l/ showed greater improvement than their productions of /r/. Also, /l/ appeared to be easier to produce for most of the learners, which was also observed in Hattori (2009), based on the smaller amount of time spent to train them on this phoneme as well as their self-report. A possible explanation for the greater improvement and ease of production for /l/ is that the articulation for /l/ is similar to the articulations for the alveolar tap [ɾ] and the alveolar stop /d/ in that they all involve a direct contact between the tongue tip and the alveolar ridge. Catford and Pisoni (1970) revealed that articulations of novel speech sounds can be learned in reference to the speech sounds the learners already know. In their experiment, one of the two groups of native English speakers who were trained on production of novel exotic speech sounds[2] only received articulatory information on the sounds. The information referred to English sounds that share some articulatory gestures with the sounds to be learned. As a result, the English speakers became able to produce the sounds without hearing the sounds during the training. Therefore, the English speakers just modified the articulations for the speech sounds they were already familiar with. Similarly, the Japanese learners in the present study were familiar with [ɾ] and /d/

[2] The authors did not indicate what language(s) the learners were trained on. These speech sounds are likely to be adapted from several different languages, e.g., the voiceless dorso-palatal fricative [ç], the glottalic egressive dorso-velar stop [k'], and the close back unrounded vowel [ɯ].

because both sounds exist in Japanese. Therefore, the learners might have become able to produce /l/ by modifying the articulatory gestures for these familiar sounds, which were referred to in the training. Indeed, it appeared to be easier for them to understand and learn the articulation for /l/ in reference to these Japanese sounds, rather than contrasting it with the articulation for /r/, which they were not quite familiar with because there are no similar sounds in Japanese.

On the other hand, /r/ does not resemble any phonemes in Japanese in terms of articulatory gestures. /r/ is perceptually similar to [ɾ], /w/, or the high back unrounded vowel /ɯ/ in Japanese for native Japanese speakers (Bradlow, 2008). However, articulations of these Japanese sounds do not involve lingual gestures made near the palate. Although one learner (NJ9) showed large improvement in production intelligibility for /r/ after the training, it might have been challenging for the learners to be trained to produce /r/ in absence of any familiar articulations that they could refer to in their native language. In fact, no sound categories in Japanese were referred to for the production training for /r/. It is also possible that the learners' knowledge in the production of the retroflex /r/ interfered with the training which focused on the production of the bunched /r/. Because the learners were likely to have used the tongue-tip raising gesture when producing /r/ based on their knowledge prior to the training, this lingual gesture persisted in many of the learners. Thus, the lack of reference sounds in Japanese combined with the challenge the learners faced when altering lingual gestures for the novel articulation might have attenuated effects of the training for /r/.

4.4 Perception learning

The Japanese learners' perception of /r/ and /l/ did not improve after the training despite the changes in F2 and F3 in their productions of the phonemes, although the learners' response bias became reduced after the training. Although it is possible that the training did not facilitate the learners' perceptual accuracy, a possible factor for this lack of improvement is the learners' subjective familiarity with the stimulus words. It has been shown that native Japanese speakers are more likely to identify /r/ and /l/ in words that they have heard or said more frequently, especially if the speakers have been immersed in an English-speaking environment for four years or less (Flege et al., 1996). Although none of the 120 stimulus words were misidentified by all the Japanese learners in the present study, nine words (one /r/-word and eight /l/-words) appear to have posed difficulties to more than half the learners both at pre-test and at post-test. Because the learners had gained experience with English in Canada for a short time (less than four month for nine learners, and nine months for one learner), it is possible that their less frequent encounter with (or use of) these nine words affected their perception of the words. Further, frequency of occurrence of the words in speech might have contributed to the learners' unfamiliarity with the words. Indeed, according to Brigham Young University-British National Corpus, six out of the nine words are less frequently used in speech relative to their minimal-pair counterparts (Davis, 2004). Although it is possible that the learners' lack of familiarity with these words interfered with their perception of the words, they constitute a small portion of the entire stimulus words. Thus, it seems less likely that the learners' subjective familiarity with the stimulus words substantially contributed to the lack of improvement in perception.

4.5 Production intelligibility

The intelligibility of the Japanese learners' productions did not improve significantly. However, intelligibility judgments given by only three listeners are likely to vary greatly, as the large standard deviation values indicate, thereby weakening the effect of the training on production intelligibility, especially for /l/. In fact, the acoustic analysis indicates that the learners' productions of /l/ improved more than their productions of /r/. Thus, it is possible that the learners' productions of /l/ became more intelligible to native English speakers after the training. This potential improvement, however, was not accompanied with improvements in goodness of the productions. This suggests that the Japanese learners' productions of /l/ were not perceived as better examples of the category after training, possibly due to the difficulty the English listeners experienced when trying to identify the phoneme in the learners' speech recorded after training, although they identified /l/ as intended more frequently in the speech.

This potential improvement in production intelligibility for /l/ appears to confirm the findings by Aoyama et al. (2004), Flege et al. (1995), and Hattori (2009) in which intelligibility of /l/ tokens exceeds that of /r/ tokens for adult Japanese learners at an early stage of L2 learning. At the same time, Aoyama et al. (2004) speculates that this greater production intelligibility for /l/ for Japanese adults may be resulted from substitutions of the Japanese tap [ɾ] for the English phoneme. Japanese adults tend to substitute [ɾ] for English /r/ and /l/, and this tendency is greater for /l/ (Riney, Takada, & Ota, 2000). Further, [ɾ] is more likely to be identified as English /l/ by native English listeners (Sekiyama & Tohkura, 1993). Thus, the seemingly high intelligibility of /l/ tokens may be due to frequent substitutions of [ɾ], which tends to be heard as English /l/ by native

English listeners (Aoyama et al., 2004). However, the English listeners who made intelligibility judgments in the present study were phonetically trained, and [ɾ] was included in the response alternatives. Phonetically trained listeners may be more sensitive to differences between English /l/ (or /r/) and the Japanese tap [ɾ]. When they perceive [ɾ], they may select the corresponding response alternative. Thus, it seems less likely that the English listeners in the present study selected /l/ as a response when the Japanese learners substituted [ɾ] for /l/, although it is still possible that the listeners might have heard some [ɾ] tokens as /l/. On the other hand, it is uncertain whether or not the English listeners in Aoyama et al. (2004), Flege et al. (1995), and Hattori (2009) were phonetically trained because detailed educational backgrounds of the listeners were not provided. Further, [ɾ] was not included as a response alternative in these studies (Aoyama et al., 2004; Flege et al., 1995; Hattori, 2009). Thus, even if the listeners were phonetically trained and detected [ɾ] substitutions, they might have had to select response alternatives which do not correspond to [ɾ] (and the closest alternative could be /l/). This seems possible especially for Flege et al. (1995) because /r/ and /l/ were the only response alternatives for intelligibility judgments. Therefore, the potential improvement in production intelligibility for /l/ in the present study might not be solely accounted for by substitutions of the Japanese tap.

When the English listeners in the present study misidentified /l/ tokens from the pre-test, they were most likely to select /d/, followed by /b/. /r/ was the third most selected category. On the other hand, when the English listeners misidentified /l/ tokens from the post-test, they were most likely to select /r/, followed by /d/. The finding for the misidentified pre-test /l/ tokens is interesting because both Aoyama et al. (2004) and

Hattori (2009) found that /r/ was most likely to be selected, followed by /d/ when /l/ tokens were misidentified.

In the present study, the English listeners were most likely to select /l/ when they misidentified /r/ tokens from the pre-test and post-test. This aligns with the findings from Aoyama et al. (2004) and Hattori (2009) in which /l/ was most frequently selected for misidentified /r/ tokens. However, the second most selected category for misidentified /r/ tokens was /b/ in the present study, whereas /w/ was the second most selected category in the previous studies (Aoyama et al., 2004; Hattori, 2009). /w/ was never selected in the present study. This suggests that some of the Japanese learners might have made the lip rounding gesture by protruding the lips excessively, which made the lips closed.

Interestingly, the English listeners' misidentification patterns for the pre-test production and the post-test production appears to reflect the influence of Japanese on the learners' productions and how they were modified throughout the training. For the production of /l/, the learners might have been likely to substitute the Japanese sounds (/d/, /t/, [ɾ]), which are close to /l/ in terms of articulation, when they mispronounced it. In fact, tongue gestures appropriate for these Japanese sounds but not for /l/ were often observed for some learners during the training. The findings from Aoyama et al. (2004) and Hattori (2009) also indicate substitution of /d/ for /l/. However, the learners appear to have modified the tongue gestures to be more appropriate for /l/ after the training. For /r/, the learners might have had difficulties in producing /r/ without direct contact between the tongue and passive articulators such as the alveolar ridge before the training. Further, the misidentification patterns suggest some influence of Japanese on the learners' productions of /r/ (substitutions of [ɾ]). After the training, when the learners

mispronounced /r/, they appear to be more likely to substitute /l/ or make an excessive lip rounding gesture that produced a sound like /b/. Aoyama et al. (2004) also reported misidentification of /r/ as /b/ by English listeners, although it was less frequent compared with /l/ or /w/.

4.6 Relationship between production intelligibility and perceptual accuracy

4.6.1 Individual Japanese learners' performance in production and perception

Close inspection of individual Japanese learners' performance revealed considerable variation in degree and modality (i.e., perception and production) of improvement across learners (see Table 4.1 and Table 4.2). Perception accuracy indicates percentages of correct identification of /r/ and /l/ by the Japanese learners from the perception tests (Experiment 2). Production intelligibility indicates percentages of correct identification of /r/ and /l/ by the English listeners from the intelligibility judgment task.

Table 4.1. Individual NJ learners' perception accuracy and production intelligibility scores for /r/ at pre-test and at post-test (in percent)

	Perception			Production		
Learner	Pretest	Posttest	Difference	Pretest	Posttest	Difference
NJ1	66.67	50.00	-16.67	51.67	50.00	-1.67
NJ2	48.33	50.00	**1.67**	95.00	95.00	0.00
NJ3	78.33	55.00	-23.33	83.33	90.00	**6.67**
NJ4	73.33	68.33	-5.00	83.33	75.00	-8.33
NJ5	65.00	63.33	-1.67	86.67	96.67	**10.00**
NJ6	78.33	85.00	**6.67**	75.00	78.33	**3.33**
NJ7	73.33	68.33	-5.00	46.67	10.00	-36.67
NJ8	50.00	66.67	**16.67**	88.33	68.33	-20.00
NJ9	75.00	48.33	-26.67	36.67	98.33	**61.67**
NJ10	56.67	60.00	**3.33**	95.00	91.67	-3.33
Average	66.50	61.50	-5.00	74.17	75.33	**1.17**

Table 4.2. Individual NJ learners' perception accuracy and production intelligibility scores for /l/ at pre-test and at post-test (in percent)

	Perception			Production		
Learner	Pre-test	Post-test	Difference	Pre-test	Post-test	Difference
NJ1	53.33	41.67	-11.67	55.00	88.33	**33.33**
NJ2	71.67	73.33	**1.67**	98.33	98.33	0.00
NJ3	48.33	50.00	**1.67**	100.00	98.33	-1.67
NJ4	46.67	48.33	**1.67**	93.33	96.67	**3.33**
NJ5	55.00	61.67	**6.67**	46.67	98.33	**51.67**
NJ6	68.33	76.67	**8.33**	70.00	90.00	**20.00**
NJ7	58.33	55.00	-3.33	85.00	96.67	**11.67**
NJ8	40.00	56.67	**16.67**	85.00	86.67	**1.67**
NJ9	63.33	73.33	**10.00**	65.00	36.67	-28.33
NJ10	53.33	70.00	**16.67**	60.00	100.00	**40.00**
Average	55.83	60.67	**4.84**	75.83	89.00	**13.17**

Moreover, their improvements in perception and production interacted with the type of phoneme. Four learners (NJ4, NJ5, NJ8, and NJ10) improved in both modalities only for /l/ whereas one learner (NJ6) improved in both modalities for both phonemes. Further, for the learners NJ5, NJ8, and NJ10, improvement in one of the modalities occurred for /r/ as well (improvement in production for NJ5 and improvement in perception for NJ8 and NJ10). Further, two learners (NJ3 and NJ9) improved in both modalities as well. However, their improvement in production occurred only for /r/ whereas their improvement in perception occurred only for /l/. Two learners (NJ1 and NJ7) improved only in production for /l/ whereas one learner (NJ2) improved only in perception for both phonemes.

4.6.2 Relationship between pre-test performance and change in performance

In order to analyze whether the variation in performance level before the training correlate with the variation in degree of changes in performance for the same modality and the same phoneme, Pearson's correlation coefficients and their levels of statistical significance (at an alpha level of .05, 2-tailed) were calculated within phoneme for each modality. The degree of changes in production performance (i.e., production intelligibility) was calculated by subtracting the averaged percentage of correct identification of /r/ or /l/ for the pre-training productions across English listeners from the averaged percentage of correct identification of /r/ or /l/ for the post-training productions across English listeners for each Japanese learner. Likewise, the degree of changes in perception performance (i.e., perceptual accuracy) was calculated by subtracting the percentage of correct identification of /r/ or /l/ for the pre-training perceptual test from the

percentage of correct identification of /r/ or /l/ for the post-training perceptual test for each Japanese learner.

For /r/, there was no significant relationship between the Japanese learners' production intelligibility level at pre-test and the degree of changes in production intelligibility, r = -.31, p = .382. Similarly, there was no significant relationship between the Japanese learners' perceptual accuracy level at pre-test and the degree of changes in perceptual accuracy for /r/, r = -.60, p = .067. On the other hand, for /l/, there was a marginally significant, moderate negative relationship between the learners' production intelligibility level at pre-test and the degree of changes in production intelligibility, r = -.62, p = .055. However, the learners' perceptual accuracy level at pre-test was not significantly correlated with the degree of changes in perceptual accuracy for /l/, r = -.15, p = .69. Therefore, the analysis indicates that the learners whose productions of /l/ were less intelligible before the training showed greater amounts of improvements in production intelligibility for /l/.

4.6.3 Relationship between change in production and change in perception

Six learners improved in either of the modalities for /r/ (see Figure 4.1). For the only learner (NJ6) who improved in both modalities for /r/, the amount of improvement in perception slightly exceeded the amount of improvement in production.

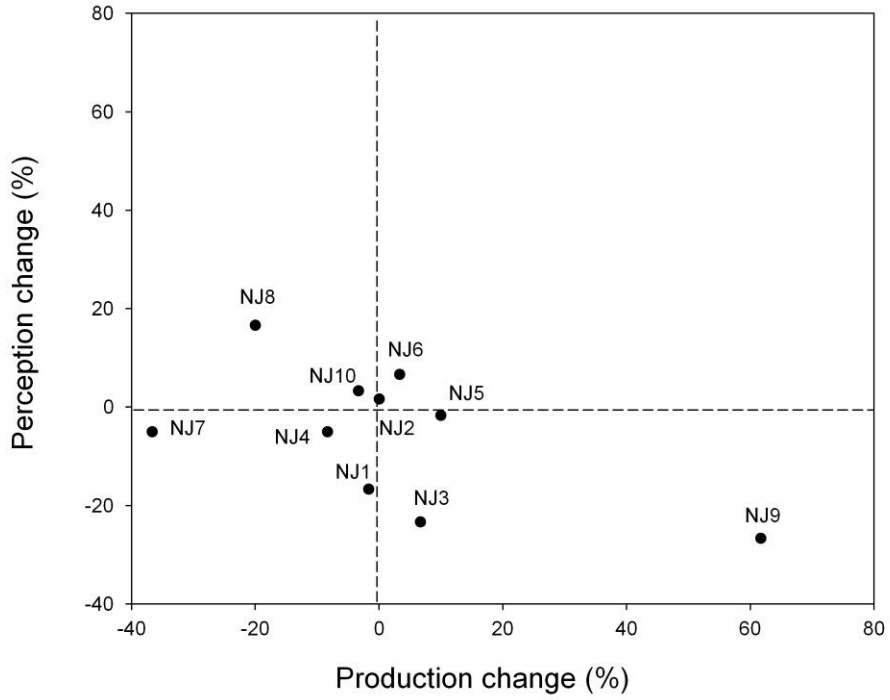

Figure 4.1. Scatter plot of production changes and perception changes for individual NJ learners for /r/. Perception changes were calculated by subtracting mean percentages of correct identification scores at pre-test from mean percentages of correct identification scores at post-test for each learner. Production changes were calculated by subtracting mean percentages of intelligibility scores at pre-test from mean percentages of intelligibility scores at post-test for each learner.

In contrast, all learners improved in either or both of the modalities for /l/ (see Figure 4.2). For most of the learners who improved in both modalities for /l/, the amount of improvement in production exceeded the amount of improvement in perception. This is the case for all learners, except NJ8, whose improvement in perception was greater than her improvement in production.

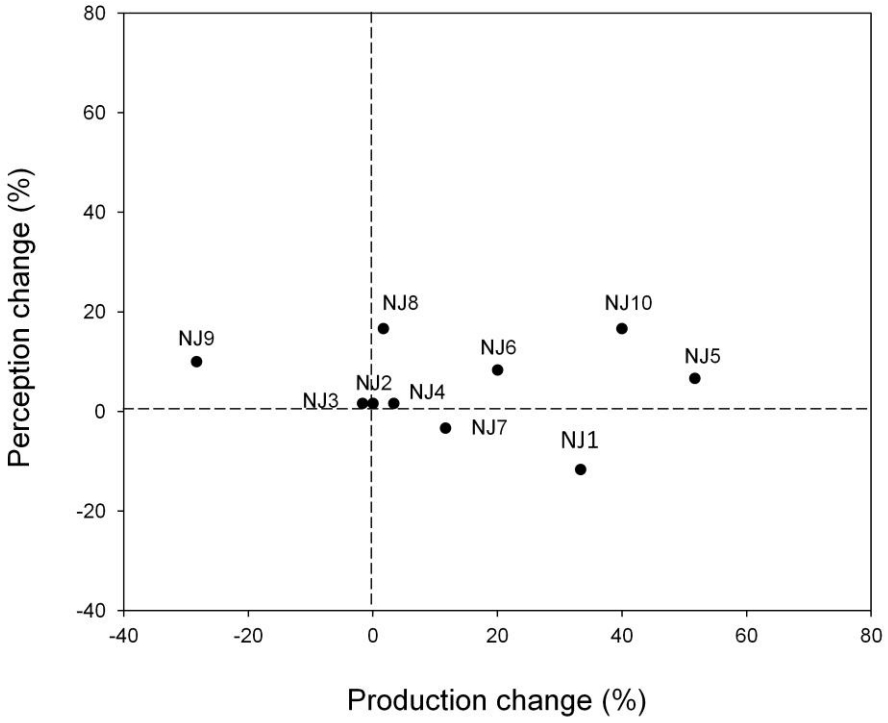

Figure 4.2. Scatter plot of production changes and perception changes for individual NJ learners for /l/. Perception changes were calculated by subtracting mean percentages of correct identification scores at pre-test from mean percentages of correct identification scores at post-test for each learner. Production changes were calculated by subtracting mean percentages of intelligibility scores at pre-test from mean percentages of intelligibility scores at post-test for each learner.

In order to analyze whether the variation in degree of changes in production intelligibility correlates with the variation in degree of changes in perceptual accuracy for the same phoneme, Pearson's correlation coefficients and their levels of statistical significance (at an alpha level of .05, 2-tailed) were calculated for each phoneme. The correlation between the degree of changes in production intelligibility and the degree of changes in perceptual accuracy for /r/ was not significant, r = -.59, p = .074. Likewise, the correlation between the degree of changes in production intelligibility and the degree

of changes in perceptual accuracy for /l/ was not significant, r = -.08, p = .837, suggesting that the amount of improvement in production intelligibility was not related to the amount of improvement in perceptual accuracy for either of the phonemes.

Inspection of individual learners' performance in production and perception revealed that improvement patterns varied considerably across learners in terms of degree, modality and phoneme type. Furthermore, for learners whose productions of /l/ were less intelligible to the English listeners before the training, their productions of the phoneme became more intelligible to the listeners after the training. Finally, the amount of improvement in production intelligibility did not correlate with the amount of improvement in perceptual accuracy. Although this may be because most of the learners did not complete all the training stages, there was no alignment between the degree of changes in production intelligibility and the degree of changes in perception accuracy even for the learners who completed all the stages (NJ03 and NJ4). Thus, the lack of correlations suggests that production learning and perception learning may be independent.

The analysis found that the learners' productions of /l/ potentially became more native-like after the training. However, this did not align with their perceptual accuracy for the same phoneme. This finding appears to correspond to the previous study (Hattori, 2009), in which Japanese learners improved their productions of /r/ and /l/ but not their perception of the phonemes. Further, the data from individual learners in the present study support the finding that some improvements in production can emerge in absence of perceptual changes, at least for L2 learners (Brière, 1966; Goto, 1970; Sheldon & Strange, 1982). Goto (1970) and Sheldon and Strange (1982) found that the ability to

produce English /r/ and /l/ exceeded the ability to perceive the phonemes for some Japanese speakers. Similarly, Brière (1966) found that the ability to produce some phonemes in Arabic, French, and Vietnamese exceeded the ability to perceive these phonemes for some monolingual English speakers who had not learned any of these languages. Additional evidence comes from a study on production-perception relationships for non-native segmental contrasts, which found that improvement in production emerged when production training was coupled with perception training whereas improvements in production and perception emerged with perception training (Baese-Berk, 2010).

However, it should be noted that the greater production intelligibility than the perceptual accuracy for /l/ in the present study could be accounted for by the visual prompts that were provided to the learners during the recordings. Seeing orthographic representations can lead to fewer errors in production tasks (Piske, Flege, McKay, & Meador, 2002), and read speech tend to be more intelligible than spontaneous speech, especially when learners are at an early stage of L2 learning (e.g., Flege et al. 1995). Thus, production might not precede perception if tokens of /r/ and /l/ in spontaneous speech are examined.

4.7 Theoretical implications

4.7.1 Production learning and perception learning in L2

Two L2 speech learning models (Speech Learning Model: Flege, 1995, 2003; Perceptual Assimilation Model: Best, 1994) propose that an L2 speech sound category is more likely to be formed if it is perceptually more distant from the closest L1 sound category. PAM

makes predictions about perceptual learning and does not explicitly address effects of interactions between the L1 and L2 on production learning. On the other hand, SLM proposes that accurate perception guides production learning for accurate production of L2 sounds. Thus, SLM would predict that production of a L2 phoneme is more learnable if the phoneme is perceptually less similar to the closest L1 phoneme. However, this might not always be the case for L2 speech production learning. The Japanese alveolar tap [ɾ] is perceptually more similar to /l/ than /r/ for Japanese speakers, but the articulation for /l/ appears to be learned more easily than that for /r/. Therefore, the production of L2 phonemes might be more learnable if the phoneme is similar to the closest L1 sound phoneme in terms of articulation.

The present study found that the degree of changes in production did not correlate with the degree of changes in perception, which is also in line with previous production training studies (Baese-Berk, 2010; Hattori, 2009). These findings would appear to contradict claims of SLM (Flege, 1995) that production learning requires accurate auditory representations as targets, and the accuracy level of the auditory representation confines the accuracy level of production in L2 speech learning. Perhaps some L2 learners may be skilled at making correct gestures by receiving explicit instructions without having awareness of the resulting sounds.

4.7.2 Relationship between production and perception

If speech production and perception share a single common mental representation or are tightly related, changes in production could be reflected in perception through the tight link between the modalities. For example, according to Motor Theory (e.g., Liberman &

Mattingly, 1985), a single phonetic representation, which consists of information about the listener's own articulatory gestures, underlies production and perception. In other words, listeners perceive speech by referring to their own articulatory gestures. Thus, Motor Theory would predict that gaining articulatory skills for a phoneme modifies the mental representation (which consists of information about the articulatory gestures) of the phoneme, thereby facilitating perception of the same phoneme. The Direct Realist Account (e.g., Best, 1995; Fowler, 1986), on the other hand, posits that listeners perceive the actual articulatory gestures of the speaker through the integrated perceptual system that senses articulatory events. That is, perception and production are tightly linked in the system. Thus, under the direct realist account, gaining articulatory skills would be transferred to perception via the tight link between the two modalities. However, similarly to the findings from earlier studies (Baese-Bark, 2010; Hattori, 2009), evidence for transfer of learning from production to perception was not observed in the present study.

During the post-training perceptual tests, a few learners read the two alternative words displayed on the computer screen out loud after they heard the auditory stimulus in every trial. According to one of those learners, she was trying to hear the words she produced in order to see which one of the words would match with the auditory stimulus that she just heard. Considering that these learners' productions were judged as highly intelligible at post-test, this anecdote would seem to imply that separate representations underlie speech production and perception. In other words, learners must have an auditory representation for perception. Nevertheless, these learners might not have been able to gain an auditory representation through the production training. Alternatively, the

production training might have allowed them to gain an auditory representation. However the representation might not be sufficiently accurate for improved perception to occur. In fact, perceptual accuracy of one of these learners at post-test was at the chance level (50% accuracy) for /l/ and slightly above the chance level for /r/. Further, his perceptual accuracy had slightly improved for /l/ but declined for /r/. The other learner's perceptual accuracy at post-test was relatively high for /l/ and below the chance level for /r/. Moreover, her perceptual accuracy showed negligible improvements for both phonemes.

The present study does not provide strong supporting evidence for separate representations underlying speech production and perception. Moreover, it is possible that those two modalities share common representations but access them through different processes. Alternatively, production learning needs to reach a certain threshold level in order to alter the underlying single representation for transfer to occur. However, if separate representations are assumed, as evidence from previous studies would appear to suggest (Baese-Bark, 2010; Hattori, 2009), these representations might not be completely independent. In fact, transfer of perception learning to production through perception training has been observed for English /r/ and /l/ in Japanese speakers (Bradlow et al., 1997) as well as for other non-native segmental and suprasegmental contrasts in native English speakers (Baese-Berk, 2010; Wang, Jongman, & Sereno, 2003). Such transfer of learning across modalities may not occur if production and perception are not associated in any way. It is not clear, however, why there is such asymmetry in transfer between the two modalities.

The Native Language Magnet Model, expanded (NLM-e) (Kuhl et al., 2008) suggests existence of a cross-modal link between production and perception. The model

proposes that connections emerge during L1 learning. That is, production learning is initially guided by perception, and connections between perception and production become formed as they influence one another. Similarly to NLM-e, SLM posits that connections between production and perception could be formed in L2 as well, as it also claims that perception guides production learning. Therefore, the transfer could be explained, assuming that a linkage is developed through perception learning.

Although these models do not predict production learning without auditory representations, it appears possible at least in L2 learning. Likewise, none of the models explicitly state whether the link can be developed from production to perception. The link might emerge with production learning, but the process in which the link is formed might be different from the link that emerges with perception learning. Hattori (2009) speculates that development of an association from production to perception could occur, however, at slow rates. Moreover, Baese-Berk (2010) revealed that production learning may suppress or slow down perceptual learning when training is provided in both modalities, whereas perceptual learning with perceptual training can predict changes in production. Therefore, building a linkage from production to perception might be effortful and time-consuming, whereas it might be easier and faster in the reverse direction. Alternatively, the linkage emerged from production might be fragile and become strengthened over time whereas the linkage emerged from perception might be more robust.

The present study showed no correlation in the degree of changes between production and perception. It is possible that the five sessions of production training were not sufficiently long for correlations to occur. However, such lack of correlations has

been observed with more production training sessions (10 sessions in Hattori, 2009) or with perception training (45 sessions in Bradlow et al., 1997, and eight sessions in Iverson et al., 2012). An emerging account for such lack of correlation in developments between production and perception is that production and perception may take distinct developmental processes underlain by different representations (Iverson et al., 2012). This proposal appears to agree with the considerable variability observed in the individual learners' data from the present study, although it does not provide convincing evidence for a close relationship between production and perception.

Finally, what are possible factors that inhibited perceptual learning in the production training? If development of accurate auditory representations requires a large amount of speech input comprising various words from multiple speakers (e.g., Logan et al., 1991), the production training might have failed to provide a sufficient environment for the development to emerge. Hattori (2009) hypothesizes that the amount of the speech input in his experiment might have been insufficient. He further speculates that the limited talker and word variability of the speech input in his experiment might have contributed to the absence of perception learning if the amount of the input was sufficient. Therefore, the learners in the present study might have not received a sufficient amount of speech input. Further, if the amount of the input was sufficient, the talker and word variability of the input might have been limited. In fact, the learners listened to a small number of targets (14 targets for the two learners who completed all training stages and eight targets for the eight learners who did not progress to the word training stage) produced by themselves, the model speaker, and the experimenter during the training. However, it has been shown that robust perception learning can emerge with

perception training using stimuli from a single talker whereas adding production of the stimuli to the perception training disrupts perceptual learning (Baese-Bark, 2010). This suggests that the low talker variability is unlikely to be the factor for the lack of perceptual learning in the present study. Alternatively, the quality of the speech input in the training might not have been optimal for perceptual learning. Flege (2003) points out that successful L2 speech learning requires input from native speakers of the language. Thus, learners' listening to their own speech might not help unless the proficiency of their production is near native speaker levels. Another potential factor might be reduced attention to the speech during the training due to the demand of the production training. The learners had to learn to control their articulators for the complex articulatory gestures and modify their existing knowledge of articulations which conflicted with the new articulatory information. Moreover, while ultrasound provides detailed visual information of lingual gestures of the learners and the model speaker, more focused attention may be required to process this rich information. In other words, because the training forced the learners to focus on production itself substantially, extra resources might not have been available for the learners to attend to and process the speech input (Baese-Berk, 2010; Ferreira & Pashler, 2002).

4.8 Future directions and implications in L2 teaching

While ultrasound technology has been shown to be a powerful tool in L2 production training (Gick et al., 2008; Tsui, 2012), it is not immune to some drawbacks. Because fixed speech organs and bones are not imaged by ultrasound, it does not show the location of the tongue or parts of the tongue relative to passive articulators, such as teeth,

the alveolar ridge, palate, and velum (Gick et al., 2008). Moreover, although the temporal resolution is higher than other visualization technologies, 30 frames per second, which is typical for ultrasound machines, can be somewhat slow for capturing rapid tongue movements (Gick et al., 2008; Stone, 2005). That is, ultrasound does not capture tongue tip movements for stop and tap sounds, which were common misarticulations for /r/ and /l/ in the training in the present study. On the other hand, the frame rate is appropriate for slower articulations such as approximants. For these reasons, in addition to inspecting the images, the experimenter asked the learners about the location and movements of the tongue based on the sounds they were making when necessary. Despite these issues, incorporating ultrasound technology into production training can be beneficial for L2 learners (Gick et al., 2008; Tsui, 2012). In fact, all learners in the present study stated in the post-training questionnaire that they found the ultrasound training helpful and enjoyable.

It is possible that significant production improvement did not occur for the learners as a group because only two learners were able to progress to the word training stage. However, greater improvements were observed among learners who ended the training at the CV syllable stage. Thus, it appears that they were able to generalize the articulation skills from the syllables to the production of words. Nevertheless, providing more training sessions would most likely lead to more robust production learning. Due to a few constraints, only five sessions were provided in the present study, which was not sufficient for most of the learners to master all the training stages. With more training sessions, all the learners might have been able to complete the word training stage, and they may have shown a greater improvement in their production as a group. Additionally,

the training targets include only three vowels (/i/, /u/, /æ/), which could have limited the learners' improvements in their productions of the prompt words for the recordings. Because the prompt words varied in vowel context, training the learners on /r/ and /l/ in a greater variety of vowel contexts might have lead to greater improvement in production.

Perceptual learning might occur if learners with more various language backgrounds are included in a study such as this one. For example, the length of residence in Canada was limited to less than one year in the present study. Learners with longer lengths of residence might be able to learn the articulations more quickly because of their long-time experience with English, thereby allowing more resources for perception learning to emerge. On the other hand, if the learner has interacted with non-native English speakers with foreign accents frequently, inaccurate articulatory representations might have been formed and could be resistant to change. Consequently, it might take more time and resources for the learners to alter the representations. It would also be interesting to explore how motivational factors would influence production learning and perception learning. Higher motivations would promote production learning and subsequently perceptual learning.

A potential application of the findings to L2 teaching would be the use of ultrasound technology in L2 production training, which has also been suggested by other researchers (Gick et al., 2008; Tsui, 2012), although the present study found that the lack of reference points (passive articulators) in ultrasound images and the frame rate could be a limitation. In addition, referring to articulations of the learner's L1 phonemes would facilitate L2 production learning. At the same time, it might be important to check the learner's knowledge in articulation of the L2 phoneme to be learned in advance in order

to determine challenges that the learner is likely to face. Finally, it should be noted that improvement in production does not necessarily indicate that the learner has become able to perceive the phoneme to the same degree, which is also suggested by Sheldon and Strange (1982). That is, perception training is necessary in order to improve perceptual abilities. It has been shown that perception training can facilitate improvements in production (Baese-Bark, 2010; Bradlow et al., 1997; Wang et al., 2003), but producing the training words could interfere with the effects of perception training on perception learning (Baese-Bark, 2010) . Therefore, providing perceptual training before production training may effectively improve both production and perception for L2 learners.

4.9 Summary

The present study explored the possibility that ultrasound production training improves production and perception of English /r/ and /l/ by Japanese learners. Overall, it replicated previous findings (Baese-Bark, 2010; Hattori, 2009) and suggests that perceptual learning might not occur with production training. It also provides implications applicable to L2 teaching. Future research with more production training sessions and learners with various experiences with the trained language would reveal whether production training leads to improved perception and further our understanding of the relationship between production learning and perception learning.

References

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099-3107.

Adler-Bock, M., Bernhardt, B.M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*, 16, 128-139.

Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data: Part II. The rhotics. *Journal of the Acoustical Society of America*, 101(2), 1078-1089.

Anderson-Hsieh, J., Johnson, R.,& Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmental, prosody, and syllable structure. Language Learning, 42(4), 529-555.

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

Audacity Team, (2012). Audacity (Version 2.0.0) [Computer program]. Retrieved from http;//audacity.sourceforge.net/

Baese-Berk, M. M. (2010). *An examination of the relationship between speech perception and production* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, 51(4), 317-342.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. In O. S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emile Flege* (pp. 331-347). Amsterdam, Netherland: John Benjamins Publishing Company.

Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, 17(3), 199-216.

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

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

Best, C. T., McRoberts, F. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305-330.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.17) [Computer program]. Retrieved from http://www. praat.org/

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In J. G. H. Edwards & M.L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 287-308). Amsterdam, Netherland: John Benjamins Publishing Company.

Bradlow, A.R., Akahane-Yamada, R.A., Pisoni, D.B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/ IV: Lont-term retention of learning in perception and production. *Perception and Psychoacoustics*, 61, 977-985.

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

Brière, E. J. (1966). Investigation of phonological interference. *Language*, 42(4), 768-796.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593-606.

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

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal*, 54(7), 477-481.

Dalston, R. M. (1974). Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*, 57(2), 462-489.

Davis, M. (2004). Brigham Young University-British National Corpus [Online corpus]. Retrieved from http://corpus.byu.edu/bnc/

Delattre, P. & Freeman, D.C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 44, 29-68.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1-16.

Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *The Canadian Modern Language Review*, 66(2), 181-202.

Espy-Wilson, C. Y., Boyce, S., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America*, 108(1), 343-356.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology*, 28(6), 1187-1199.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller and A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and Similarities* (pp. 319-355). Berlin, Germany: Mouton de Gruyter.

Flege, J. E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Flege, J. E., Frieda, E. M. & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169-186.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second languages. *Studies in Second Language Acquisition*, 23, 527-552.

Flege, J. E., & McKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 24, 1-34.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995a). Effects of age of second-language learning on the perception of English consonants. *Speech Communication*, 16, 1-26.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995b). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125-3134.

Flege, E. F., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, 38(1), 25-55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults perception of /ɹ/ and /l/. *Journal of the Acoustical Society of America*, 99(2), 1161-1173.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, 14, 3-28.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, *32*(2), 113-121.

Gick, B., Bernhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. H. Edwards & M.L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 315-328). Amsterdam, Netherland: John Benjamins Publishing Company.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Gunther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105(5), 2854-2865.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motive the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201-223.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). A test of the critical-period hypothesis for second language acquisition. *Psychological Science*, 14(1), 31-38.

Hattori, K. (2009). *Perception and production of English /r/-/l/ by adult Japanese speakers* (Doctoral dissertation). Retrieved from http://discovery.ucl.ac.uk/19204

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

Hazen, V., Senema, V., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication,* 47, 360-378.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, 39, 571-584.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 41-62). Amsterdam, Netherland: John Benjamins Publishing Company.

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.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, 122(5), 2842-2854.

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

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. *Journal of the Acoustical Society of America*, 118(5), 3267-3278.

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, B47-B57.

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

Johnson, K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Malden, MA: Blackwell Publishing.

King, R. D. (1967). Functional load and sound change. *Language*, 43(4), 831-852.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceeding of the National Academy of Sciences of the United States of America*, 97(22), 11850-11857.

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.

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.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's language*. Oxford, United Kingdom: Blackwell Publishers Limited.

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

Lively, S.E., Logan, J.S., & Pisoni, D.B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of Acoustical Society of America, 94*, 1242-1225.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96(4), 2076-2087.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of Acoustical Society of America*, 49(2), 606-608.

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

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In J. Slifka, S. Manueal, & 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.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R. C. (1996). L2 acquisition, L1 loss, and the critical period hypothesis. In R. Allan & J. James (Eds.), *Second-language speech: Structure and process* (pp. 149-159). Berlin, Germany: Mouton de Gruyter.

Major, R. C. (2008). Transfer in second language phonology. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 63-94). Amsterdam, Netherland: John Benjamins Publishing Company.

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

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975). An effect off linguistic experience: The discrimination of [r] and [l] by native Japanese speakers and English. *Perception and Psycholinguistics*, 18, 331-340.

Munro, M. J. & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech*, 38, 289-306.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520-531.

O'Conner, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word*, 13, 25-43.

Odisho, E. Y. (2003). *Techniques of teaching pronunciation in ESL, bilingual and foreign language classes*. Munich, Germany: Lincom Europa.

Ohata, K. (2004). Phonological difference between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners. Asian EFL Journal, 6(4).

Piske, T., Flege, J. E., McKay, I. R. A., & Meador, D. (2002). The production of Engish vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59, 49-71.

Piske, T., MacKay, I. R. A. & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, 191-215.

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. *Journal of the Acoustical Society of America*, 119(3), 1684-1696.

Riney, T. J., Takada, M., & Ota, M. (2000). Segments and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, 34(4), 711-737.

Schmidt, A. M., & Beamer, J. (1998). Electropalatography treatment for training Thai speakers of English. *Clinical Linguistics & Phonetics*, 12(5), 389-403.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, 21, 427-444.

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.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 455-501.

Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9402438)

Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, 16, 379-405.

Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language* (Master's thesis). Retrieved from https://circle.ubc.ca/handle/2429/43348

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

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of 3-dimentional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599-604.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106(6), 3649-3658.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech-perception. *Child Development*, 52(1), 349-355.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75(6), 1866-1878.

Wilson, I., & Gick, B. (2006). Ultrasound technology and second language acquisition research. In M.G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp.148-152). Somerville, MA: Cascadilla Proceedings Project.

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

APPENDIX A: LIST OF PROMPT WORDS

1	reek	leak
2	rate	late
3	rack	lack
4	room	loom
5	row	low
6	rip	lip
7	red	led
8	rice	lice
9	rug	lug
10	rock	lock
11	reed	leash
12	rib	limp
13	rail	lame
14	wrench	ledge
15	rat	land
16	ripe	like
17	roof	loop
18	roach	loaf
19	rub	love
20	rod	log

APPENDIX B: LIST OF STIMULUS WORDS

1	reach	leach
2	reef	leaf
3	reap	leap
4	Rick	lick
5	rid	lid
6	rift	lift
7	rim	limb
8	rink	link
9	writ	lit
10	wrist	list
11	rear	leer
12	raid	laid
13	rake	lake
14	rain	lane
15	ray	lay
16	raise	laze
17	race	lace
18	rent	lent
19	wrens	lens
20	rest	lest

21	rare	lair
22	rad	lad
23	rag	lag
24	ram	lamb
25	rank	lank
26	raft	laughed
27	ramp	lamp
28	wrap	lap
29	raps	lapse
30	rise	lies
31	rife	life
32	right	light
33	rhyme	lime
34	Rhine	line
35	rind	lined
36	ride	lied
37	rowed	loud
38	rout	lout
39	rude	lewd
40	rune	loon
41	root	loot
42	ruse	lose

43	rook	look
44	roan	loan
45	roves	loaves
46	robe	lobe
47	rose	lows
48	rob	lob
49	wrong	long
50	Ross	loss
51	raw	law
52	roared	lord
53	rump	lump
54	rush	lush
55	rust	lust
56	rung	lung
57	road	load
58	roam	loam
59	rope	lope
60	rot	lot

APPENDIX C: TABLE OF DESCRIPTIVE STATISTICS FOR ORIGINAL F2 MEASUREMENTS (IN HZ)

Group	Variable		*n*	*M*	*SD*	Range	
						Min	Max
NJ	/r/						
		Pretest	10	1364.81	225.22	1038.65	1728.47
		Posttest	10	1307.52	194.78	1017.55	1580.50
	/l/						
		Pretest	10	1552.26	153.18	1312.65	1797.56
		Posttest	10	1474.49	201.10	1116.35	1694.05
NE	/r/		5	1153.64	99.61	1065.50	1302.00
	/l/		5	1162.90	124.55	993.80	1290.90

APPENDIX D: TABLE OF DESCRIPTIVE STATISTICS FOR ORIGINAL F3 MEASUREMENTS (IN HZ)

Group	Variable		*n*	*M*	*SD*	Range	
						Min	Max
NJ	/r/						
		Pretest	10	2182.20	337.52	1555.85	2609.00
		Posttest	10	2130.74	388.60	1622.30	2849.15
	/l/						
		Pretest	10	2874.29	265.82	2602.15	3335.05
		Posttest	10	2885.54	255.09	2532.69	3300.60
NE	/r/		5	1665.73	221.84	1429.70	2026.80
	/l/		5	2935.87	102.28	2842.20	3062.75