

ULTRASOUND SPEECH TRAINING FOR JAPANESE ADULTS LEARNING
ENGLISH AS A SECOND LANGUAGE

by

Haley May-Lai Tsui

B.A., Simon Fraser University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2012

© Haley May-Lai Tsui, 2012

Abstract

Japanese adults learning English as a second language often have difficulty perceiving and producing English /l/ and /ɹ/ due to specific acoustic and articulatory characteristics of these speech sounds and their absence in Japanese phonology. The current study investigated the effectiveness of using two-dimensional tongue ultrasound to teach pronunciation of these sounds to six adult native Japanese speakers. Each participant had four 45-minute training sessions over a two-week period where visual feedback from ultrasound was used to support the teaching of lingual configurations for /l/ and /ɹ/ in a variety of vowel contexts and word positions. Speech samples from participants were taken prior to training and at a two-week follow-up session. All participants were rated by expert listeners as having more accurate productions of /l/ and /ɹ/ post-training, with the most accuracy seen in word-initial clusters and as word-initial segments. The lateral /l/ showed greater improvement than /ɹ/. Acoustic and visual analyses revealed frequencies and components of tongue positioning closer to native English speaker production in words perceived to be greatly improved between pre- and post-training productions. The effect of training on perception was exploratory and did not yield analyzable results. All participants gave very positive feedback regarding the use of ultrasound for speech training, as determined by a participant questionnaire. The results suggest that incorporating lingual ultrasound in speech training can be beneficial for Japanese adults learning English liquids.

Preface

Together with my supervisor, Dr. B. May Bernhardt, and committee member Dr. Penelope Bacsfalvi, I planned the research study in this thesis. My contributions include: (i) designing the research study in consultation with Dr. Bernhardt and Dr. Bacsfalvi, (ii) recruiting participants and collecting assessment data independently, (iii) training participants with Dr. Bacsfalvi in consultation from Dr. Bernhardt, (iv) analysing and interpreting data in collaboration with Dr. Bernhardt, and (v) preparing drafts of all chapters of this thesis, with editorial comments from Dr. Bernhardt, Dr. Bacsfalvi and Dr. Marinova-Todd. This study received approval from the University of British Columbia Behavioural Research Ethics Board (UBC BREB certificate number H11-01868).

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Acknowledgements	viii
Chapter One: Introduction	1
1.1 Acquisition of a second language phonology in adulthood	1
1.1.1 Description of relevant segments: English and Japanese approximants	2
1.1.2 Production of /l/ and /ɹ/ by Japanese speakers	4
1.1.3 Syllable structure and /l/ and /ɹ/ production	6
1.1.4 Perception of English /l/ and /ɹ/ by L1 Japanese speakers	7
1.2 A Competition Model and constraint-based influences on L2 phonology	13
1.3 Additional factors that influence L2 speech development.....	15
1.4 L2 Speech training	19
1.4.1 Visual biofeedback	22
1.4.2 Electropalatography (EPG)	24
1.4.3 Ultrasound	26
1.5 Research questions	29
Chapter Two: Method	30
2.1 Recruitment	30
2.2 Participants	30
2.3 Pre-training baseline speech assessment	31
2.4 Training sessions	36
2.5 Follow-up assessment	42
2.6 Outcome measures	43
2.6.1 Judgments of segmental accuracy	43
2.6.2 Reliability	47
2.6.3 Acoustic and visual analysis	48
2.6.4 Participant feedback survey	48
Chapter Three: Results	50

3.1 General outcomes and accuracy of /ɪ/	50
3.1.1 Accuracy of /ɪ/ by word position	51
3.2 General outcomes and accuracy of /l/	53
3.2.1 Accuracy of /l/ by word position	54
3.3 Acoustic and visual analysis	56
3.4 Perception task	61
3.5 Participant feedback questionnaire	62
 Chapter Four: Discussion	 63
4.1 Production of /ɪ/	63
4.1.1 Word-initial, medial and final /ɪ/	63
4.1.2 /ɪ/ clusters	66
4.2 Production of /l/	67
4.2.1 Word-initial, medial and final /l/	68
4.2.2 /l/ clusters	69
4.3 Acoustic and visual analysis	70
4.4 Perception	71
4.5 Participant satisfaction questionnaire	73
4.6 Limitations and future research	74
4.7 Conclusion	76
 References	 78
 Appendices	 88
Appendix A. Recruitment poster	88
Appendix B. Consent form	89
Appendix C. Language experience questionnaire	91
Appendix D. Word list for assessment	93
Appendix E. Participant handout of language differences	94
Appendix F. Participant handout describing ultrasound	95
Appendix G. Word lists used in training	96
Appendix H. Word list for perception task	100
Appendix I. Participant satisfaction questionnaire	101
Appendix J. Screen shot of listener judgment program	103

List of Tables

Table 1. Consonant features for selected segments	4
Table 2. Participant characteristics	31
Table 3. Individual participant progression throughout speech sessions	40
Table 4. Frequency of /ɪ/ word ratings	50
Table 5. Frequency of /l/ word ratings	53

List of Figures

Figure 1. Schematic of word selection for analysis	44
Figure 2. Optional questionnaire: post-session questions	49
Figure 3. Perceived accuracy of /ɪ/ in singletons by word position	51
Figure 4. Perceived accuracy of /ɪ/ in clusters by word position	52
Figure 5. Perceived accuracy of /l/ in singletons by word position	54
Figure 6. Perceived accuracy of /l/ in clusters by word position	55
Figure 7. Spectrograms of one participant's improved /ɪ/ word	57
Figure 8. Ultrasound images of one participant's improved /ɪ/ word	58
Figure 9. Spectrograms of one participant's improved /l/ word	59
Figure 10. Ultrasound images of one participant's improved /l/ word	60
Figure 11. Perception task results by participant at three elicitation points	61

Acknowledgements

I am in gratitude to my supervisor, Barbara May Bernhardt, for her support, guidance and thoughtful feedback provided throughout the process of conducting this research. I would like to acknowledge and thank Penelope Bacsfalvi for jointly conducting the speech training sessions and for sharing her knowledge on using ultrasound in speech training. Thank you to Stefka Marinova-Todd for feedback about second language learning. Thank you to Bryan Gick for the use of an ultrasound machine from the UBC Speech Research Laboratory. Thank you to all of the enthusiastic participants who were a genuine pleasure to work with.

I am very grateful to Diana Lamare and Amanda Pack for volunteering to be the expert listeners and to Heidi Lipetz for feedback on an early draft. I would like to thank Osamu Takai for conducting the interviews in Japanese, providing consultation on the Japanese language and for sharing resources. Thank you to Kathryn Pasquini for collaborating on this project by doing the English participant interviews. I would like to thank Matt Tindall for creating the computer macro for data collection, helping me recruit on weekends and for his support.

I would like to acknowledge and thank my Dad, Mom and sister Ingrid for their kindness, generosity and unwavering support in all of my endeavours.

Chapter One: Introduction

1.1 Acquisition of a second language phonology in adulthood

Second language (“L2”) acquisition in adulthood is heavily influenced by the learner’s first language (“L1”) (Masuda & Arai, 2010). Learning novel speech sounds in adulthood is a unique factor of language learning because it can be profoundly impeded by neuromuscular and perceptual constraints of L1 phonology. Whereas other domains of language, such as morphosyntax or lexical development, can be learned, memorized and practiced to attain mastery, the influence of L1 in phonology can persist for years and be resistant to change (Ioup, 2008).

Accent can be defined as way of speaking in an L2 that “retains features of the native language in that acoustic values in the L2 may be based on parameters of the L1” (Ioup, 2008, p. 43). The influence of L1 phonological features on L2 acquisition can be significant, both enhancing and inhibiting the ability to produce novel speech sounds in the L2. Perception of non-native phonemic contrasts in adulthood can also be challenging, sometimes resulting in perceptual assimilation where a novel segment is categorized as being within the same phonological category as a native L1 phoneme (Miyawaki et al., 1975). This can lead to difficulties perceptually distinguishing contrasts in non-native speech sounds. Since adults learning a second language often demonstrate accented speech that may limit intelligibility in the second language, the current study was undertaken to determine the effectiveness of visual feedback in speech production training of non-native phonemic contrasts. Specifically, this was explored by teaching English /l/ and /ɹ/ to native Japanese speakers learning English as a second language in adulthood. The inclusion of visual feedback in training may aid the language learner in

attaining optimal placement of the articulators for speech by making tongue positioning visible. Visual feedback may also aid the clinician in providing individualized training and feedback.

The following review discusses the unique acoustic and articulatory challenges that may be present for L1 Japanese speakers learning to perceive and produce English liquids. Additionally, linguistic and environmental factors that can influence pronunciation of L2 in general and the use of training methods in adult second language learning, including visual biofeedback (particularly with ultrasound), will be discussed.

1.1.1 Description of relevant segments: English and Japanese approximants

The difficulty for Japanese L1 speakers learning English L2 /l/ and /ɹ/ is due to a variety of articulatory (pronunciation) and acoustic (perceptual) factors. The phonological inventory of the Japanese language does not contain the equivalent of English /l/ or /ɹ/. It does contain a voiced dorso-palatal glide /j/, a voiced labio-velar glide /w/ and a voiced tap /ɾ/ (Labrune, 2012; Tsujimura, 2007). The tap is the most prototypical realization in Japanese, but allophonic variations exist. A retroflexed apico-alveolar lateral [ɭ] is realized preceding palatalized vowels and in young Japanese women's speech; a voiced alveolar lateral fricative [ɮ] occurs before high vowels /i/ and /u/. A voiced alveolar stop [d] often occurs word initially or medially in Japanese children's speech. Retroflex [ɽ] can occur before /u/ or word medially in sequences with repeated vowels, such as [aɽa] and [oɽo]. Finally, apical trills, both short [r] and long [r:], are socially indicating variants that may be present in colloquial or Tokyo male gangster speech (Labrune, 2012).

Comparatively, the English language contains four approximants: a voiced dorso-palatal glide /j/, a voiced labio-velar glide /w/, a voiced apico-alveolar lateral /l/ with a labiovelar variant and a voiced apico-prepalatal central approximant /ɹ/ (Ladefoged & Johnson, 2010). There are two primary variants of /l/ in English. Light /l/ is produced by touching the tongue tip to the alveolar ridge and relaxing one or both sides of the tongue to allow for the lateral passage of airflow. Dark /l/ is produced in the same manner, but is velarized with the tongue positioned further back. This occurs when /l/ is in coda position following a back vowel, as in the words *full*, and in consonant clusters, as in the word *told* (Bernhardt & Stemberger, 1998). The labiovelar variant also occurs before all back vowels.

Two main production patterns of English /ɹ/ have been reported (Secord, 2007). Retroflexed /ɹ/ is produced by approximating the tip of the tongue to the alveolar ridge, or just behind at the pre-palatal area, while raising the sides of the tongue to brace against the upper inner teeth, called lateral bracing or midline grooving. This causes the passage of airflow to escape sagittally through the groove created in the medial portion of the tongue. In this configuration, the tongue tip may also be curled up or bent backwards slightly. The tongue back is raised towards the palate while the tongue dorsum lowers. Lip rounding also occurs prevocalically (Bernhardt & Stemberger, 1998). For bunched /ɹ/, the tongue body is raised towards the palate and the tongue tip is pointed downwards. Lateral bracing and lip rounding may also be evident (Adler-Bock, Bernhardt, Gick & Bacsfalvi, 2007). For acoustic characteristics of these sounds, the reader is referred to section 1.1.4.

1.1.2 Production of /l/ and /ɹ/ by Japanese speakers

Due to the absence of the afore-mentioned liquid segments in Japanese phonology, Japanese speakers are not likely to have experience with the specific combination of gestural components required for English /l/ or /ɹ/ (Bradlow, 2008). English liquids /l/ and /ɹ/ are often the last speech sounds acquired in typical speech development of children learning English; they may not develop until around age 7 (Smit, 1993a), which is suggestive of their difficulty even for L1 speakers of English. Whereas glides /j/ and /w/ are more easily produced by native Japanese speakers due to their presence in the phonological inventory of Japanese, English liquids /l/ and /ɹ/ are often mispronounced by Japanese speakers learning English. Substitutions include the labio-velar approximant [w], alveolar tap [ɾ] or high back vowel [u] (Bradlow, 2008) or they may substitute each other (Flege, Takagi & Mann, 1995). English /l/ and /ɹ/ may also be erroneously assimilated into the single /ɾ/ category in perception of English consonants, contributing to their likelihood of both being replaced by /ɾ/ in speech production (Hazan, Sennema, Iba, & Faulkner, 2005; Miyawaki et al., 1975) (See Table 1).

<i>Segment:</i>	/l/	/ɹ/	/ɾ/	/w/	/u/
[sonorant]	+	+	+	+	+
[consonantal]	+	-	+	-	-
[continuant]	+	+	+	+	+
[lateral]	+				
[voiced]	+	+	+	+	+
[round]		+ (prevocalic)		+	
Labial	(Labiovelar)	Yes, Syl-initial		Yes	
Coronal	✓	✓	✓		
anterior	+	-	+		
distributed	-	+	-		
Dorsal	(Labiovelar)	(Yes)		Yes	Yes

Table 1. Consonant features for selected segments (Bernhardt & Stemberger, 1998; Hayes, 2009).

The goal of the language learner is to discriminate these phonemes into separate and distinct phonological categories and this assumes a sufficiently “plastic” phonological representation system in adults. The perceptual characteristics of these phonemes will be discussed in a following section.

Despite the absence of /l/ and /ɹ/ in the Japanese phonological inventory, research has demonstrated that native Japanese speaking adults can have success in learning these English sounds from both long-term immersion in an English-speaking environment and with specific speech training. Flege, Takagi and Mann (1995) found that, although less accurate than native English speakers, Japanese speakers with more experience speaking English (having lived in an English-speaking country for 21 years or more), received higher accuracy ratings in /l/ and /ɹ/ production than L1 Japanese speakers who had lived in an English-speaking country for approximately 2 years. Both groups of speakers had studied English at school in Japan, but were not immersed in an English-speaking environment until adulthood and neither group had received specific training, suggesting that the noted improvement was a result of simply living in an English-speaking environment. Thus, adequate exposure to English for L1 Japanese speakers may result in an improvement in production accuracy of English liquids /ɹ/ and /l/, but that this process may be very gradual and may never result in native speaker proficiency. Note that this study only assessed /l/ and /ɹ/ in word initial position in a monosyllable CVC (C=consonant, V=vowel) structure, whereas a more thorough investigation of a variety of phonetic environments and word structures would present a more representative sample of their participant’s abilities.

Aoyama, Flege, Guion, Akahane-Yamada and Yamada (2004) also studied the effect of living in an English-speaking environment on native Japanese adults and children in producing English /l/ and /ɹ/ word initially preceding a variety of vowels (Experiment 2). The authors found that Japanese children who were 9 years old at time of initial testing did show improved accuracy of English /ɹ/ production one year later without training. However, neither children nor adults in their study improved on accuracy of /l/ production. The authors also did not find an improvement in /ɹ/ production for the adults (mean age 39.9 years old), although this may be due to the relatively short time frame (one year) between first and second elicitations, especially compared with the longer time frame (21 years) studied in Flege, Takagi, & Mann (1995).

Neither Flege et al. (1995) nor Aoyama et al. (2004) included pronunciation training in their studies. The very gradual or absent change observed in adults in such immersion studies suggests that research is needed to investigate the effectiveness of direct pronunciation training of English /l/ and /ɹ/, in order to determine whether progress can be accelerated (See Section 1.4).

1.1.3 Syllable structure and /l/ and /ɹ/ production

In addition to language differences in the segmental inventory as discussed above, Japanese and English syllable structures vary greatly. This can also affect /l/ and /ɹ/ production. The syllable structure of Japanese is (C)(j)V(V)(C); however, Japanese syllables are very rarely closed and consonant clusters are only permissible with the glide /j/ in second position and are also very rare (Ota & Ueda, 2007). The influence of these

marked structures will be discussed in the following section. In contrast, closed syllables and clusters in both onset and coda position are permissible in English syllabic structure.

Due to the different syllable shapes between these two languages, it has been observed that English consonant cluster production by adult native Japanese speakers often results in epenthetic vowel insertion between the target cluster consonants, creating two distinct syllables with context-dependent vowel insertion (Masuda & Arai, 2010; Shibuya & Erickson, 2010). The insertion of epenthetic vowels as a repair strategy to break up consonant clusters in English results in /o/ commonly inserted after alveolar sounds /t/ and /d/ and /u/ often inserted after /s/ and /b/, following Japanese phonotactics. The authors found it rarer, but not completely absent, for their participants to produce an excrescent insertion of /ə/ during a cluster simplification repair strategy, suggesting that accented speech patterns are influenced by both phonological (determined by the languages) and phonetic (general speech sound production) constraints.

Speech production and perception are interconnected, and the following section outlines perceptual issues for Japanese L1 learners of English concerning liquids and relevance for speech production.

1.1.4 Perception of English /l/ and /ɹ/ by L1 Japanese speakers

This section outlines acoustic characteristics of English /l/ and /ɹ/, describes potential difficulty for Japanese listeners, and relates perception of these liquids to production (both in general and with training).

As noted, the difficulty for native Japanese speakers in perceiving the English /l/-/ɹ/ contrast is also influenced by acoustic characteristics of these speech sounds. The /ɹ/

and /l/ typically have first formants (F1) around 250 Hz and second formants (F2) around 1,100 Hz (Ladefoged, 2006). The specific contrast between English /l/ and /ɹ/ requires the listener to attend to a differentiation between the third formant (F3), an acoustic quality that does not differentiate any segments in Japanese (Lotto, Sato, & Diehl, 2004). The F3 frequency for /ɹ/ by native English speakers is often around 2,078 Hz for females and 1,548 Hz for males (Dalston, 1975) due to the velar and pharyngeal constrictions with simultaneous lip rounding, with the frequency rising to the onset of the vowel. It is the lowest F3 onset of any phoneme in English. The F3 for /l/ by native English speakers is usually around 2,935 Hz for females and 2,523 Hz for males (Dalston, 1975). It can stay flat or raise to the vowel, as when it precedes /i/ (Ingvalson, Holt, & McClelland, 2012; Bradlow, 2008). Higher formants for /l/ are often initially reduced in intensity compared with /ɹ/ (Bradlow, 2008), but increase in intensity quickly near the onset of the vowel (Ladefoged, 2006). In their analysis of the acoustic variables of L1 English productions of [lɑ] and [ɹɑ] in comparison with the tap /ɾ/, Miyawaki et al. (1975) state that “the starting point and the transition of the third formant [of /ɾ/] seem to vary unsystematically over a range of values sufficient to distinguish the American /ɹ/ and /l/” (p. 332). The tap /ɾ/ also lacks a consistent steady state whereas “liquids...typically have short steady-state portions with an appreciable amount of sound energy preceding the formant transitions” (Miyawaki et al., 1975, p. 332), exceptions being liquids and glides in prevocalic positions.

Lack of linguistic experience attending to the third formant (F3) for Japanese native speakers contributes to the difficulties in perceiving this contrast when learning English liquids in adulthood. Learning to perceive a contrast in L2 requires learners to

establish new phonetic categories and this ability is retained, but somewhat diminished, as age of exposure to the L2 increases (Flege et al., 1995). Iverson et al. (2003) demonstrate how the phonological inventory of one's native language influences the perception of non-native languages by contrasting native Japanese speakers with native German speakers in terms of their respective abilities to discriminate between English /l/-/ɹ/. The learners had all received English language instruction in their home countries for approximately 7 years, but did not live abroad. Listeners judged the similarity of minimal pair stimuli on a 1-7 scale and their responses were mapped using a scaling technique to chart the average similarity ratings by each participant. The authors conclude that the "perceptual spaces" for the stimuli were strongly affected by the listener's native language. Whereas native English speakers were very sensitive to changes in F3 relating to /l/ and /ɹ/ differences, Japanese speakers possibly focused more on F2 when perceiving English /l/-/ɹ/, a less relevant and reliable audio cue for perceptually distinguishing this English phonemic contrast. German participants showed slight variations, but overall similar perceptual patterns to English speakers, which was expected due to the presence of light /l/ and uvular fricative /R/ in German phonology.

The relevance of perception for production is also seen in two studies without training. Lotto et al. (2004) notes the importance of attending to the F3 onset frequency to distinguish perceptually between English /l/-/ɹ/ and its effect on speech production. When comparing English /l/ and /ɹ/ audio recordings by native English versus native Japanese speakers, the English speakers had a clear boundary of F3 frequency values between /l/ and /ɹ/, with /l/ showing a higher frequency (2,000-3,500 Hz approximately) than /ɹ/ (1,000-2,000 Hz approximately). The native Japanese speakers had overlapping F3

values between these segments and did not display the same acoustic F3 contrast in their productions of /l/ (1,500-3,500 Hz approximately) and /ɹ/ (1,300-2,500 Hz approximately). Additionally, the F3 frequency for /ɹ/ was produced somewhere between /l/ and /ɹ/ (values ranged between 1,500-2,500 Hz), but was slightly closer in F3 values to /l/, demonstrating the similar acoustic properties of these segments in Japanese speakers.

For /l/, Flege et al. (1995) found no difference between F1 and F3 values between native Japanese speakers and native English speakers' productions, although the F2 onset was higher for L1 Japanese speakers. The authors also found that for /ɹ/, Japanese speakers demonstrated a higher F3 onset frequency than native English speakers. This is congruent with results from Lotto et al. (2004) who also found a wider range of F3 onset frequency values English /ɹ/ by Japanese speakers that spanned a higher frequency range than English speakers.

As demonstrated in Miyawaki et al. (1975), native Japanese and native English speakers perform equally well when perceiving an isolated F3 contrast in non-speech sounds, suggesting that perception of F3 within linguistic sound contrasts may be a unique type of auditory perception or that the issue of perception in speech stimuli is not isolatable in an F3 difference. Indeed, in one study, training that focused specifically on perception of F3 frequencies in synthetic speech was not effective for native Japanese speakers learning to perceive the /l/-/ɹ/ contrast (Ingvalson et al., 2012). However, this study did not use natural speech stimuli in training and only trained one syllable type with one vowel (/la/-/la/). Furthermore, the authors did state that the group who was trained and received corrective feedback did show "limited improvement", but specific values were not provided. This is in contrast to Bradlow, Akahame-Yamada, Pisoni and

Tohkura (1999) who found robust perception training effects when using natural speech contrasts. Bradlow et al. (1999) may have achieved better outcomes in perception accuracy than Ingvalson et al. (2012) due to the larger sample size, the use of four phonetic environments rather than one and the use of a variety of English speakers in training (Bradlow et al., 1999). Importantly, outcomes may have furthermore been affected by the presence of feedback. All Japanese participants received feedback about their performance accuracy in Bradlow et al. (1999), but only half (six) of the participants in Ingvalson et al. (2012) received feedback. As mentioned, some improvements occurred for the six participants in Ingvalson et al., (2012) who were provided performance feedback.

Studies have demonstrated that native Japanese speaking adults can benefit from teaching that focuses on perceptual distinction of the categorical boundaries of English /l/ - /ɹ/ with two (Aoyama et al., 2004), ten (Hazan et al., 2005); (Iverson, Hazan, & Bannister, 2005) or twelve training sessions (Zhang et al., 2009). The exception is the Ingvalson (2012) study mentioned above.

Additionally, training in perceptual contrast identification has shown to transfer some benefits to production accuracy as well, even when pronunciation is not the focus of training (Hazan et al., 2005). Bradlow, Akahane-Yamada, Pisoni and Tohkura (1999) found that nine native Japanese speaking participants increased their production accuracy after perceptual contrast training for distinguishing /l/-/ɹ/ at a 3-month follow-up assessment. These authors stress the importance of including variability in a training program with regards to teaching perception of novel segments in a variety of word positions, phonetic environments and with a variety of native speakers. These studies

demonstrate the strong association between speech perception and production and that training in one domain may positively affect the other.

Interestingly, there is evidence to suggest that accurate production of English /l/ and /ɹ/ for native Japanese speakers may precede accurate perception, which may be contrary to some pedagogical views that perception must be mastered prior to pronunciation training (Goto, 1971). Six of the 11 native Japanese participants in Goto (1971) were considered to be “excellent in English conversation” and have a higher degree of accuracy in production of /l/ and /ɹ/ compared with the Japanese speakers with less English language experience. However, these experienced speakers still did not perform at the level of native English speakers for discrimination of /l-/ɹ/, even when their own accurate productions of /l-/ɹ/ were recorded and played back to them. The author suggests that the speakers might be relying on “kinetic sensation of their own speech organs” (p. 321) to produce the contrast.

Goto (1971) also notes that higher accuracy in auditory discrimination tasks was found when the American speakers were well known to the Japanese listener, suggesting the possibility of familiarity in increasing auditory perceptual abilities in non-native contrasts. With regards to an interactive activation model, which will be discussed in further detail in section 1.2, it is speculated that the acoustic characteristics associated with a familiar speaker may decrease the processing demands required of second language learners, allowing more processing resources to be allocated to phonologic information of the signal. I turn now to a discussion of models relevant to Japanese L2 learning English liquids.

1.2 A Competition Model and constraint-based influences on L2 phonology

As noted in the previous section, the difficulty inherent in perceiving and producing the /l/-/ɹ/ contrast for Japanese speakers may be attributed to: (1) their absence in Japanese phonology, creating novel articulation patterns that non-native English speakers must learn; (2) their underlying featural similarity to particular Japanese glides, the tap and vowels, increasing the difficulty with which they are systematically differentiated into separate phonemic categories; and (3) the lack of experience non-native listeners may have with attending to acoustic featural contrasts required to differentiate /l/ and /ɹ/, specifically F3 onset frequency.

Constraints from both the speaker's L1 and linguistic universal markedness constraints affect the development of non-native pronunciation in adulthood. As a speaker's experience with the specific features of L1 phonology can influence L2 pronunciation in adulthood, accent for people with a similar language background is often alike as a result of the phonotactic constraints of being highly tuned and practiced in one's L1. For the adult L2 learner who already has a fully developed phonological system for L1, pronunciation training in L2 typically focuses on training of new phonological contrasts that are present in the L2, but absent in L1 (Schmidt & Beamer, 1998).

A psycholinguistic, interactive model that aims to account for typical, impaired and second language acquisition and processing is the Competition Model (Presson & MacWhinney, 2010). It explains L2 phonological learning with principles common to other aspects of language processing, with reference to an underlying competition of interpretation (reception) and output alternatives (production). Weighting in the

competition is based on a variety of cues and constraints in the system that may be linguistic universals or more language-specific. L2 accent is partly affected by cue costs influencing language-universal phonological processes. A cue cost refers to a demand for processing resources by the system. For example, if the system is requiring a speaker to produce a difficult, non-native segment, the cost of doing so will be high because of the additional requirements of the speaker to concentrate on moving the articulators in a novel pattern and compare their auditory signal with that of a native speaker production to confirm accuracy.

Some constraints are more language-specific, but are common among the world's languages. As previously mentioned, in the Japanese language, if two consonants are in an onset, the second position must be filled by the glide /j/ (Ota & Ueda, 2007).

Therefore, clusters containing any other segment in second position are considered marked and not permissible (*Complex) and may result in epenthesis or deletion of the first consonant in the cluster (Hancin-Bhatt, 2008). Closed syllables are also marked in Japanese (NoCoda). To test the predictions of degree of markedness on perception in L2 English learners, Eckman and Iverson (1994) found that L1 Japanese adults made more errors on a discrimination task on marked coda clusters than less marked ones. It was also mentioned that liquids (*/l, ɹ/) are absent in the Japanese phonological inventory.

Japanese speakers bring these constraints, among others, to the process of learning English in adulthood. According to the Competition Model, it is partly these underlying featural differences between L1 and L2 that can result in accent through transfer that influences competition provided by highly activated, and therefore lower cue cost, L1 units (MacWhinney, 1992).

There are interactions between typological markedness and L1-L2 overlap with regard to difficulty of acquisition of novel L2 sounds (Eckman, 2008). Generally, a more marked segment is more difficult to acquire, unless the speaker has experience with the structure from the L1. L2 segments that are different from L1 and more marked will be the most difficult segments to acquire (Eckman, 2008). An example is English /ɪ/ that both is absent in Japanese phonology and is not common in the world's languages. Therefore, there is influence from both typological markedness and a comparison of L1 and L2 that can contribute to the development of native-like pronunciation in L2.

Regarding the trajectory of L2 phonological learning, the Competition Model predicts that in the initial stages of L2 learning, the influence of L1 in terms of cue costs and phonological constraints will be strong. When L1 has a significant featural, segmental or suprasegmental overlap with L2, the transfer will be positive with regards to both perception of contrasts and production in articulation (MacWhinney, 1992). Gradually it is predicted that the language learner will focus more on matching the articulatory output to the target L2 and learning will be facilitated by successful perception of L2 contrasts. According to this model, the learner must perceive inaccuracies of their own L2 productions in order to benefit from the phonological loop and better match productions to their targets.

1.3 Additional factors that influence L2 speech development

In addition to linguistic, processing and acoustic parameters, a number of additional parameters affect pronunciation in L2. Age of L2 acquisition, amount of L2 use, length of immersion in an L2 environment and the amount of instruction the L2 are

relevant, in addition to speaker gender, motivation and skill. Production training is also relevant and is discussed further in the next section.

In terms of age, it was previously believed that the end of a critical period exposure to minimize accented speech is approximately 11-12 years of age (Tahta, Wood, & Loewenthal, 1981; Scovel, 1988; Flege, Yeni-Komshian, & Liu, 1999). Long's (1990) meta-analysis of L2 phonology learning found studies showing a range of success for achievement native-like proficiency for L2 learned between the ages of 6 and 12 years of age, with all participants having non-native-like pronunciation (accented speech) if exposure to the L2 occurred after 12 years of age. However, more recent studies investigating influence of L1 Korean (Flege et al., 1999; Flege et al., 2006) and L1 Italian (Flege, Frieda, & Nozawa, 1997) on L2 English learning suggest that a critical or sensitive period for phonological acquisition with native-like proficiency is much earlier, i.e., at approximately 3-6 years of age. These latter studies reported age of acquisition to have the strongest positive correlation with degree of accented speech, compared with all other environmental variables examined. Flege et al. (1999) also note that age of acquisition did not affect other domains of language, such as morphosyntactic abilities. Proficiency in these areas steadily increased as a result of amount of education in the L2. However, degree of accent in these studies was based on native speaker judgments and the authors do not discuss which features are present that contribute to accented speech in individuals of these language backgrounds. The stimuli in these studies were elicited imitations of English sentences. It is not known whether stimuli were selected to adequately represent the English speech sound inventory.

Other studies showing the relevance of age were Piske, MacKay and Flege (2001), Flege et al. (1999, 2006) and Oyama (1976). Oyama (1976) studied Italian immigrants to the United States, finding age of acquisition to be a strong predictor of degree of foreign accent compared with length of residence in the United States, also suggesting that the development of an accent is related more to age of immersion than length of exposure. This is consistent with results from Riney and Flege (1998) that found after an initial period of rapid learning when first immersed in an L2 speaking environment, length of residence did not correlate with a decrease in the degree of accented speech for adults learning an L2.

Further to 'immersion', amount of L2 use as a predictive factor of degree of accent has yielded conflicting results in the small number of studies where this was examined. Flege and Fletcher (1992) found no significant difference for the amount of English L2 use and degree of accent for speakers of Spanish L1. However, Purcell and Suter (1980) found that a composite variable combining the number of years residing in the United States and the number of months of co-habitation with a L1 English speaker was the third most predictive factor of degree of accent in an L2, behind L1 background and aptitude for oral mimicry. Ingvalson, McClelland and Holt (2011) found that both age of acquisition and length of residence in the United States contributed to the degree of accent in Japanese L1 speakers and English /l/-/ɹ/ perception and production proficiency; longer residences and earlier age of exposure correlated positively with higher accuracy scores in both perception and production.

Other personal variables that have been studied with regards to degree of foreign accent in an L2 in adulthood are gender and personal motivation level. Studies that have

examined gender as a predictive variable have either found no influence or a female advantage (Piske et al., 2001). The gender effect reportedly decreases when age of L2 learning and amount of L2 experience are controlled for (Asher & Garcia, 1969; Flege et al., 1995). Motivation level has also been positively correlated to pronunciation outcomes in some studies (Elliott, 1995; Moyer, 1999), but not to native speaker proficiency levels, even with participants who were considered highly motivated for occupational reasons (Moyer, 1999). Oyama (1976) and Thompson (1991) did not find any effect of motivation level in their participants on L2 pronunciation accuracy. Conflicting results may be partially influenced by motivation being a difficult variable to control and measure and it is often not measured in many L2 studies (Piske et al., 2001). Motivation may also be subject to a response bias from participants who do not want to appear unmotivated to the researcher and it can be argued that most participants who are volunteering for research are motivated.

In addition to motivation to increase accuracy in L2 pronunciation, some learners of a second language may experience motivation to maintain an accent due to its link with personal and cultural identity. A group of individuals who speak with shared linguistic features may psychologically benefit from association or membership within a culture based on their common way of speaking (Underwood, 1988). Therefore, it may not always be the goal to eradicate all forms of influence of L1 on L2 pronunciation. Marx (2002) describes how increasing proficiency in an L2, in an attempt to achieve native pronunciation, affected pronunciation in the L1. This in turn affected her sense of cultural identity and membership in the L1 culture during re-integration. However, in a study of 100 L2 speakers of English, respondents stated that when a breakdown in L2

communication occurs, a pronunciation problem contributes at least 55% of the time (Derwing, 2003). The participants, who all had various L1 backgrounds, almost unanimously agreed that they would like to pronounce English as a native speaker and that doing so would not negatively impact their cultural identity in the L1. Thus, individual factors certainly influence motivation in pronunciation training and modification.

Research presented here shows that a variety of interacting variables contribute to accent. Future research is warranted to determine the relative weight of the various factors. The next section discusses more in depth the topic of speech training in accent modification, the central topic for the current study.

1.4 L2 Speech training

While adults may no longer possess phonological acquisition capabilities equivalent to those of children, adults can learn to produce non-native phonemic contrasts with exposure and training (Flege et al., 1995; Hazan et al., 2005) The amount of formal instruction in the L2 has not generally proven to be predictive of degree of accent. However, this may be because the focus in language learning classes is typically on vocabulary development, morphosyntax and grammatical structure with pronunciation receiving very little direct training in many language classes. In the 1960s in Canada, pronunciation training was a priority in L2 teaching classrooms, with the goal of learners achieving native proficiency in speech production. Teaching pronunciation gradually fell out of favour when learners were not achieving this goal, leading to the decrease or shift in pronunciation training to focus on suprasegmentals rather than segmentals (Levis,

2005). However, there has been a recent resurgence of pronunciation training focusing on the goal of increasing intelligibility for successful communication in conversational English (Saito & Lyster, in press). Of the L2 training programs that have focused on teaching L2 pronunciation directly, more training effects have been observed (Moyer, 1999).

The importance of a direct teaching approach is often emphasized in pronunciation training because adults with persistent phonological habits in their first language do not tend to make gains with simplistic “repeat-after-me” teaching paradigms (Odisho, 2003; Schmidt & Beamer, 1998). Articulation training, manner descriptions and perceptual skills are useful for adult clients, and consideration of the L1 phonological system is crucial to predict probable areas of pronunciation strengths and difficulties in the L2.

One study that used a less direct approach, however, was Saito and Lyster (in press). In their study of 25 native Japanese speakers, all improved in production of English /ɪ/ in comparison with a control group. Instruction entailed corrective feedback of production accuracy with word-initial and -medial /ɪ/ words for four 1-hour sessions over a period of three weeks. This study was unique in that participants were not informed that pronunciation would be assessed and feedback for /ɪ/ accuracy was embedded in a course on English argumentative skills. Skills in training included using the target words in preparation for a debate and practicing words in public speaking tasks. The control group received an equivalent time of comparable instruction on debate skills, but without focus on specific words with /ɪ/. Outcome measures included native speaker listener ratings and acoustic analyses, namely a drop in the third formant (F3) that corresponds with more

accurate /ɪ/ productions. However, later follow-up of /ɪ/ production was not done to determine whether maintenance was achieved and it is not exactly clear whether the participants were completely blind as to the purpose of the experiment as they were aware that they were working with words containing English /ɪ/.

Further to what adults bring to or know about the learning process or training method, they may have insights into personal learning strategies, including meta-cognitive and meta-linguistic abilities that are highly beneficial for learning. A combined “bottom-up” phonetic/phonological and “top-down”, metacognitive/metalinguistic approach to articulation training can be useful for teaching L2 phonology (Odisho, 2003). A method that supports these aspects of learning in adults is the use of visual biofeedback technology in speech training. The auditory speech input is supplemented by visual images (“bottom-up”) that can be used cognitively (“top-down”) to support motor learning. The addition of visual feedback in otherwise unobservable movements has been shown to be beneficial for adults learning complex, novel motor patterns and the combination of providing an adequate model with visual feedback of one’s performance can positively facilitate learning (Carroll & Bandura, 1982).

According to the Competition Model, there is a high cue cost associated with violating the principles of L1 (such as *Complex, *NoCoda, */ɪ, l/), which are highly engrained in adult speakers. Yet reranking of these constraints (demotion) is necessary to achieve intelligible pronunciation in L2. Imitation of new speech sounds is a high-demand cognitive-motor activity. Supplementing speech input with visual information (discussed below) may lower the cognitive demands of this learning process to benefit second language learners by providing visual motor targets to imitate in speech

production. Additionally, training production of speech sounds may also lower the demands associated with perception of non-native speech contrasts by increasing the familiarity of relevant auditory input through exposure.

1.4.1 Visual biofeedback

Four decades ago, Catford and Pisoni (1970) observed that visual biofeedback may be helpful in teaching L2 phonology. However, several types of Computer Aided Pronunciation Training (CAPT) tools provide the speech learner with feedback about accuracy, but lack the ability to provide specific instruction for articulation (Badin, Youssef, Bailly, Elisei, & Hueber, 2010); or they provide articulation instruction, but are unable to provide specific feedback for the individual learner because the training is on a computer (Massaro & Light, 2003). For example, one feedback method uses real-time audio-visual formants in spectrograms to teach sound class distinctions of non-native phonetic contrasts (Öster, 1998), but spectrographic information is abstract and more useful for certain contrasts over others.

Other systems use animated views of articulatory positioning, but often do not take into account the influence of the individual's first language phonology on accented speech patterns and provide general statements about accuracy without providing specific feedback. One exception is the Automat for Accent Reduction (AZAR) that was developed in Germany (Jokisch, Koloska, Hirschfeld, & Hoffmann, 2005). In using this visual articulation feedback tool with Russian L1 speakers learning German as L2, the authors developed a set of allophonic variants of German phones based on the Russian phonological inventory in conjunction with a recording function that allows for human

feedback. Although this tool combines multi-modal elements for L2 learning, the learner cannot visualize their own unique patterns of muscle use, which can be an important factor in shaping production of novel speech sounds (see Gibbon et al., 1991, discussed in Section 1.4.2).

Training studies have also been conducted with the digital talking head, Baldi. Massaro and Light (2003) examined the utility of Baldi to teach identification and production of English /l/ and /ɪ/ to 11 native Japanese speakers. Half of the participants used Baldi initially with the digital “skin” removed, allowing the participants to see the articulators (tongue, palate, teeth) within the oral cavity. The other half of the participants started training using Baldi in a natural condition with the skin present. Their within-subjects design showed improvements in all participants, regardless of condition (whether they received training with articulators visible first or last). Perhaps improvement was possible regardless of timing of introduction of the articulators visible condition. It would be beneficial in the future to use a between-subjects design to compare whether seeing within the oral cavity is beneficial over a group that does not receive any training with articulators visible. Their participants had very positive feedback about the use of Baldi, specifically, the condition with articulators visible. In another study, Hazan, Sennema, Ida and Faulkner (2005) compared the use of audio-only to audio-visual feedback with Baldi in training for another group of native Japanese speakers learning English /ɪ/ and /l/ speech contrasts. For perception, participants improved identification accuracy of non-native speech contrasts after training, regardless of whether they received audio-only or audio-visual information. For production, the group that received audio-visual (hearing speech with the speaker’s face visible)

information improved more than the audio-only group in pronunciation of /ɪ/ and /l/. However, while the perceptual training experiment had beginner to intermediate level English speakers, the participants tested in the later production task were mainly phonetics PhD students and trained EFL teachers, who would be at a distinct advantage to both perceiving and producing sound contrasts of a non-native language. Due to the lack of representative sample for the production study, the effect of perception training with visual feedback on the production by an average English L2 learner needs to be explored. The outcomes of the two studies support the use of visual feedback technology in teaching non-native speech contrasts to adult L2 learners. More direct articulatory feedback has also been useful in L2 training, as discussed in the next two sections.

1.4.2 Electropalatography (EPG)

Electropalatography (EPG) has been used in speech training to provide information about lingual-palatal contact patterns during speech production; for example, it has shown positive learning effects, for individuals with persistent /s/ difficulty (Gibbon & Hardcastle, 1987), children with hearing impairments (Bernhardt, Gick, Bacsfalvi & Ashdown, 2003), adults with acquired apraxia of speech (Howard and Varley, 1995) and in an exploratory treatment study of speech sound disorders in five children with cerebral palsy (Nordberg, Carlsson, & Lohmander, 2011).

EPG has also been used to teach novel speech sounds to L2 learners. Two adult native Japanese speakers were able to achieve more optimal tongue-to-palate contact patterns for English /l/ and /ɪ/ when training with EPG in four 45-minute sessions over a period of 2 weeks, demonstrating the usefulness of visual feedback technology for use

with adult L2 learners (Gibbon, Hardcastle & Suzuki, 1991). The authors also observed substantive idiosyncratic pre-training differences in tongue-palate contact during pronunciation of English /l/ and /ɹ/ for their two participants, recognizing the need for individualized speech assessment and training. The participants learned to use the visual feedback tools quickly and made significant gains in speech production of these segments during the training period.

However, EPG has a relatively high cost of fitting and creating individual palates plates for each speaker. Additionally, the EPG plate itself requires a period of familiarization for each user in order to prevent its interference with natural speech production while wearing the acrylic plate. Participants in some L2 learning EPG studies have been required to wear a palate 2 weeks prior to training to (1) reduce excess saliva production and (2) become accustomed to wearing the palate during natural speech (Schmidt & Beamer, 1998). The authors did not specify for how long each day participants wore their palates, but it has been suggested that participants wear a plate for at least 24 hours prior to engaging in speech research (Gibbon et al., 1991).

In their study teaching a variety of English L2 phonemes to three Thai L1 speakers twice a week for approximately 6 months using EPG, Schmidt and Beamer (1998) found that /ɹ/ showed the most variability prior to training, emphasizing the need for individualized assessment and training. Variability existed in the width of medial tongue grooving and the location of initial tongue-palate contact during /ɹ/ production, and was realized as closer to a trilled [r] or tap [ɾ]. The authors were successful in teaching production of several English sounds and suggest that the use of EPG for L2 speakers may be more efficient and less frustrating than traditional pronunciation training

that does not include a feedback component. Furthermore, the authors also note the individualized patterns of tongue-palate contact, made available by EPG, and individual progression through training in each participant which indicates a need for an individualized training approach.

1.4.3 Ultrasound

Another technology, ultrasound, also has been used in recent years for speech training. Ultrasound uses high frequency sound waves transmitted through a transducer containing piezoelectric crystals (Gick, Bernhardt, Bacsfalvi, & Wilson, 2008). It is non-invasive, non-ionizing and does not place the user in any real discomfort. For speech training, when the probe (transducer) is held against the speaker's neck under the chin and above the larynx, ultrasound provides visual information about the speaker's tongue shape and movement in either a midsagittal or coronal plane (Wilson & Gick, 2006). An image results when ultrasonic waves are transmitted upwards by the probe and reflected back downwards when they hit the space immediately above the tongue's surface (Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005). (See ultrasound images in Chapter 3).

Similarly to how speakers gain information about speech sounds from facial cues, visualization of the tongue can be beneficial during speech training, especially for sounds that are complex articulatorily or not visually salient due to articulation being primarily further back within the oral cavity, such as English /l/ and /ɹ/ (Badin et al., 2010).

Ultrasound images of the tongue provide the client and clinician with specific information about tongue placement and movement beyond what can be obtained from

other feedback sources such as a mirror, spectrogram information or electropalatography (EPG) (Bernhardt et al., 2005).

Recently, ultrasound technology has become portable, more affordable and thus more clinically accessible, making it a potentially valuable tool in the fields of speech sciences and training/therapy to display tongue positioning and movement. An ultrasound machine can be used for multiple clients, greatly minimizing the cost of fitting and creating personalized palates as required for EPG. The clinician can model correct articulation and tongue movements, providing the learner with a visual articulatory target to mimic. The learner can compare their own tongue movements with the models, so that both the clinician and learner can observe the learner's tongue movements and create/practice modifications as needed. In this way, both the learner and clinician can visualize the learner's tongue positioning and provide feedback about tongue placement. Real-time visual feedback has been shown to be more beneficial for motor learning in adults than delayed visual feedback through the use of videotapes, for example (Carroll & Bandura, 1982).

Ultrasound has shown promise for use with teaching speech production to deaf and hard of hearing individuals (Bernhardt, Gick, Bacsfalvi, Ashdown, 2003), adolescents with persistent /r/ difficulties that have been resistant to other forms of treatment (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007) and as treatment for adult apraxia of speech secondary to stroke (McNeil et al., 2000).

Research in the use of ultrasound with English L2 learners is sparse. Ultrasound was used in a pilot study on accent modification for adult Japanese L1 learning English L2 (Gick, Bernhardt, Bacsfalvi & Wilson, 2008). In this study, three native Japanese

speakers, who were also linguistics students, were trained on production of English /l/ and /ɹ/ using ultrasound for 1 hour each. All participants made production accuracy gains in this short time period for both /l/ and /ɹ/ and had highly positive feedback with regards to their experience and their perception of the usefulness of using ultrasound for speech training. This pilot study suggests the applicability of ultrasound technology for L2 learners. Adult L2 learners are often articulate in their L1 and do not have speech, language or learning disabilities, suggesting that they may progress at a fast pace, have internal motivation for practice and be cognitively able to benefit from the visual images of both the clinician's and their own tongues that ultrasound technology provides. Gick et al.'s (2008) pilot study demonstrated that L2 learners without speech, language or learning disabilities are typically able to interpret visual tongue images shown on the ultrasound screen very easily with a brief orientation (Gick et al., 2008).

However, this pilot study did not control for variables such as age of acquisition, length of residence in an English-speaking country or use word lists controlling for vowel contexts and word shape. It also did not follow-up with participants to determine whether long-term maintenance of gains made in the single session was achieved. The present study was designed to extend and expand Gick et al.'s (2008) pilot study and to contribute further to the literature of training for Japanese speakers on production of English liquids with visual feedback support. Questions and predictions for the study follow in Section 1.5.

1.5 Research questions

1. Can L1 Japanese speakers learn to produce English /l/ and /ɹ/ accurately in a variety of phonetic contexts and word positions after speech training sessions using ultrasound?
2. If changes in speech are observed, are the Japanese productions closer in approximation to native English speaker productions, both acoustically and articulatorily?
3. Does ultrasound training in speech production affect perception of the same contrasts, even if perception is not targeted in training?

Specific predictions and hypotheses are that L1 Japanese adults will be successful in learning production of English /l/ and /ɹ/ after four training sessions focusing on the lingual components for these sounds using ultrasound as a visual feedback tool. It is predicted that these changes will be reflected in the acoustic and visual analysis of the spectrographic and ultrasound images to closer approximate native English speaker productions. Finally, it is predicted that gains made in production will positively affect perceptual accuracy of the same contrast.

Chapter Two: Method

2.1 Recruitment

Recruitment posters (Appendix A) containing a description of the research study and contact information of the principal investigator and co-investigators were placed in a variety of locations on the University of British Columbia Vancouver campus. Locations included the Student Union Building, Japanese Students' Association club room, Freidman Building, Buchanan Building, Education Building, Asian Studies Building and the English Language Institute. Approval for posting was obtained when required by each building. Information regarding participation was also disseminated by teachers in the English Language Institute on campus and included in an electronic newsletter by the UBC Institute of Asian Research (IAR) distributed to their mailing list on September 24, 2011. The content of the message was as follows:

“The School of Audiology and Speech Sciences at UBC is currently recruiting native Japanese speakers who have learned or are learning English as a second language for a research study looking at the effectiveness of using tongue ultrasound as a pronunciation teaching device.

Each participant will receive 4 one-on-one sessions with a registered speech-language pathologist or a final year Masters student in speech-language pathology to work on perception and pronunciation of English "r" and "l" in November 2011. Participants will also be asked to attend one pre-session assessment and one post-session assessment where audio recordings will be taken.

Interested participants can contact Haley Tsui at hmtsui@interchange.ubc.ca to sign up for the sessions or to ask any questions.”

2.2 Participants

Six native Japanese speakers between the ages of 19-28 (mean age = 22.8 years) participated in this study (4 female; 2 male). As determined by a hearing screening at the time of baseline assessment, all participants had normal hearing ability. Additionally,

none of the participants had ever had speech therapy or been diagnosed with a speech or language impairment either in their first or second language. All participants were deemed to have typical pronunciation in the Japanese language, as determined by short conversations done by a native Japanese speaker who is also a PhD student in speech-language pathology at the University of British Columbia.

Participant #	Sex	Age	Age of beginning formal English instruction	# months in ESC*	% speaking English in daily life
1	F	19	13 years	2	50%
2	M	22	6 years	2	75%
3	F	22	17 years	24	75%
4	F	24	12 years	6	90%
5	F	22	13 years	2	100%
6	M	28	13 years	36	75%

Table 2. Participant Characteristics (*ESC = English-speaking country: Western Canada or North-western United States)

2.3 Pre-training baseline speech assessment

A 45-minute baseline assessment was conducted 2 weeks prior to commencement of the training sessions. The assessment and all training sessions were held at the School of Audiology and Speech Sciences at the University of British Columbia. The baseline assessments were conducted individually with each participant by the thesis author and included the following tasks:

- a) Read and sign informed consent form (See Appendix B)
- b) Language experience questionnaire (See Appendix C)
- c) Audio recording of word list #1 (See Appendix D)
- d) Hearing screening

- e) Audio recording of word list #2
- f) Perception task
- g) Audio and visual recording of word list #3

a) Informed consent

All participants provided informed consent for participation in this study by signing a consent form (Appendix B). Each participant was given a copy of their signed consent form for their own records.

b) Language experience questionnaire

Each participant filled out a Language experience questionnaire (Appendix C) with the thesis author. This included questions about age of first exposure to the English language, information about formal English instruction, self-rating of speech accuracy for /l/ and /ɹ/ and degree of motivation to participant and practice. See Table 2 above for participant characteristics obtained from this questionnaire.

The length of time living in an English-speaking country prior to participating in the research study ranged from 2 months to 3 years (mean length of time = 12 months). For all participants except one, the place of residence was Vancouver, Canada. The exception was Participant 3 who spent one year in Vancouver, Canada and one year in the north-western United States. Due to the fact that there is a Japanese community in Vancouver, and participants might not be speaking English 100% of the time, participants were asked to estimate the amount of English use in daily life. All participants stated that they spoke English at least 50% of the time in daily life in Canada, with the majority of

participants rating their English use at 75% or higher. Contexts in which English was used included conversations at school or home with strangers, friends, roommates and significant others.

The age of first exposure to the English language ranged from 6 years of age to 17 years of age (mean age of first exposure = 12.3 years). All participants' first exposure to the English language was through formal English language instruction in Japan consisting of instruction on English vocabulary and grammar. All participants mentioned that their English instruction did not include pronunciation teaching.

Participants were asked to self-rate their productions of /l/ and /ɹ/ on a 4-point rating scale as either (1) Exactly on target, (2) Almost on target, (3) Somewhat on target or (4) Not at all on target. For /l/, all participants rated themselves as (3) Somewhat on target and listed a variety of reasons as to why /l/ was difficult to produce. When asked what they thought made /l/ a difficult segment to learn, the participants responded with several reasons. These included not having /l/ in Japanese, /l/ being difficult to hear in typical English conversation, difficulty with understanding what to do with one's tongue to produce /l/ and not being able to consciously think about /l/ production in conversation where speech is typically fast.

All participants rated their /ɹ/ productions on the same 4-point scale and all rated themselves as (3) Somewhat on target for pronouncing English /ɹ/. Participants were also asked what about the /ɹ/ segment they thought was difficult. Variables identified were difficulty with pronouncing it based on position in the word and its difficulty to produce in connected speech, compared with in single words.

Participants were asked to rate their level of motivation to participate and practice on a 4-point scale. The scale included (1) Extremely motivated, (2) Very motivated, (3) Slightly motivated and (4) Not very motivated. Five participants rated themselves as (2) Very motivated and participant one stated that she was (1) Extremely motivated. Participants were also asked to provide their reasons for participation in this study. The responses included to improve pronunciation of English /l/ and /ɹ/, to “not think so much during conversation” and to take advantage of opportunities to practice English. Participant two stated that he was motivated to improve English pronunciation for employment opportunities.

c) Data collection

A 44-word list comprising 24 words with /ɹ/ and 20 with /l/ were elicited in a variety of word positions (Appendix D). Three words for word initial, medial, final singletons and final clusters were elicited. For word initial clusters, eight word initial /l/ cluster and twelve word initial /ɹ/ clusters were elicited. This is due to the greater variety of permissible consonant clusters containing /ɹ/ in English. Vowel contexts included high front /i/ and /ɪ/, mid-front /e/, low front /æ/, mid-central /ə/, high back /u/, mid-back /ʌ/ and a variety of English diphthongs. Three audio recordings of a word list were taken for each participant during baseline assessment to ensure reliability of the elicited sample. The collection of three tokens per word to comprise a representative sample has also been done in previous research with native Japanese speakers (Aoyama et al., 2004; Lotto et al., 2004) and native Thai speakers using EPG to learn English (Schmidt & Beamer, 1998). For this report, the analysis used data from the third elicitation because

participants were more familiar with the word set and needed less cueing at the time of the third recording.

The words were elicited in phrases in response to pictures using Microsoft Office PowerPoint 2003 (Version 11.8335.8341). Phrases were: “I want to see _____” for word-initial and word-medial or word-initial cluster targets. Words with /ɹ/ or /l/ in final position or in final clusters were elicited in the phrase “I want to see _____ be”. The word “be” was chosen because it is a labial segment with neutral tongue positioning, decreasing any potential co-articulation effects on the preceding target segment while still eliciting the word in a carrier phrase.

Participants were initially shown the word list and asked to identify whether any words were unfamiliar to them, but were usually not shown the word list during elicitation to avoid any potential effects of orthography on production unless they did not know the word. Flege et al. (1995) found that the Japanese speakers with less than 2 years of experience with English were more accurate in English liquid production in reading than in spontaneous speech. The authors suggest that reading written words may overestimate the performance ability of some speakers in an L2.

d) Hearing screening

All participants passed a pure-tone hearing screening that was completed using a portable audiometer with attached circumaural headphones at 1000, 2000 and 4000 Hz at 20dB.

e) Perception task

Due to the interaction between perception and production of non-native phonemic contrasts, participants in the current study completed an exploratory perception task where participants were asked to identify /ɪ/ and /l/ words in 10 minimal pairs. The minimal pair word list for this task can be found in Appendix H. For this activity, each participant was shown two pictures displayed side by side depicting the minimal pair words (ex. *lock* and *rock*) in Microsoft Office PowerPoint 2003 (Version 11.8335.8341). An audio recording of one of the words spoken by the author, a female native English speaker who is about the same age as the participants, was played to the participants. They were then instructed to point to the picture that matched the spoken word. No written words were provided to prevent potential orthographic influence.

2.4 Training sessions

Each participant had four individual training sessions lasting 45 minutes each. The four sessions were held over two consecutive weekends where each participant attended speech sessions on both Saturday and Sunday. The design for the current study was based on Bernhardt et al. (2008) where assessments were taken prior to and after ultrasound training which included modelling, introducing sounds first as silent gestures and then adding voice. The number and timing of sessions was similar to previous work with adult second language pronunciation training using biofeedback (Gibbon, Hardcastle, & Suzuki, 1991).

The training sessions were held in the same room as the assessments at the University of British Columbia School of Audiology and Speech Sciences. A registered

speech-language pathologist (SLP), Dr. Penelope Bacsfalvi, Clinical Assistant Professor at the School of Audiology and Speech Sciences, UBC, and a final year Masters student in speech-language pathology, the author, jointly conducted all training sessions.

Consultation was also provided by the thesis supervisor, Dr. Barbara May Bernhardt, also an SLP, who met each participant and observed at least one training session for each.

The first training session consisted of the following:

- a) Discussion of selected differences between Japanese and English phonologies (see Appendix E for participant handout)
- b) Description of how ultrasound works (see Appendix F for participant handout)
- c) Orientation to the ultrasound machine; clinician modeling of lingual components of /l/ and /ɹ/
- d) Teaching gestural components of /l/ and /ɹ/
- e) Participant practice with ultrasound

The following is a description of each of these activities:

a) Discussion of language differences

Some pertinent differences between English and Japanese phonologies were discussed and an informational sheet was provided to each participant (Appendix E). This information was given in order to heighten awareness of some of the influencing factors of L1 on English pronunciation. Participants were given information about syllable

structure, which is most often (C)V (C=consonant; V=vowel) in Japanese and (C)V(C) in English. Also mentioned was that English phonology permits a variety of clusters containing two to four segments across word positions, whereas the Japanese language does not permit clusters, except very rarely with palato-approximant /j/ word initially. Finally, the concept of loan words (or “Gairaigo”) was introduced to demonstrate influences of native language on pronunciation.

b) Description of how ultrasound works

The clinicians described how ultrasound works and gave an orientation to its images to each participant. Each participant also received an informational sheet on ultrasound technology and observation of the tongue using ultrasound (Appendix F). Discussion included information on how ultrasound works, why it may be useful for pronunciation training, an image demonstrating orientation of the tongue on ultrasound and a link to the UBC School of Audiology and Speech Sciences website (URL: <http://www.audiospeech.ubc.ca/research/child-phonology-phonetics-and-language-acquisition-lab/ultrasound-in-speech-training>) containing ultrasound videos.

c) Orientation to the ultrasound machine

Participants were shown the ultrasound machine and oriented to which side displays the tongue tip and which side displays the tongue root. The clinician demonstrated correct placement of the ultrasound probe under the chin and participants practiced probe placement to enable clear tongue visualization.

d) Teaching lingual components of /l/ and /ɹ/

Training began with the clinicians modelling the lingual components of /ɹ/ and /l/ using ultrasound. After modeling the correct lingual configurations for the segments, the participants held the ultrasound probe independently to view their own tongue positioning.

For /ɹ/, the retroflexed variant was the model because it is the more common variant (Espy-Wilson & Boyce, 1994). The mid-sagittal of ultrasound showed the tongue from near the tip to view the tongue tip up behind the alveolar ridge, the tongue back raised and the entire body of the tongue shifted posteriorly in the oral cavity. The coronal view provided information on medial grooving and lateral bracing against the upper back teeth. The focus of training for /l/ was the mid-sagittal orientation to ensure that the tongue blade was being used for articulation and that it was held for a sufficient amount of time in order to distinguish it from /r/.

e) Participant practice with ultrasound

Each individual was given opportunity to practice tongue positioning for /ɹ/ and /l/ with ultrasound and each progression of each training session depended on individual progress made during each session. See Appendix G for a list of words used in training. The general pattern of training sessions began with segments in isolation without sound (gestural components only) and increased in difficulty level as follows:

- a) Segments in isolation without voice
- b) Segments in isolation with voice
- c) Segments in word-initial position in open syllables (minimal pairs)

- d) Segments in word-final position in open syllables
- e) Segments in word-initial position in closed syllables
- f) Segments in word-final position in closed syllables
- g) Segments in word-medial position
- h) Segments in word-initial clusters
- i) Segments in word-final clusters
- j) Segments in a variety of word positions in simple carrier phrases
- k) Segments in complex sentences with multiple tokens

Table 3 shows how each individual progressed through each of these stages during training. In sessions two, three and four, it is estimated that each participant practiced /l/ and /ɹ/ approximately 200 times per session.

Participant #	Content of each session by session number							
	Isolation without voice	Isolation with voice	Open syllable	Closed syllable	WM	Cl	Words in carrier phrases	Words in complex sentences
1	1	2	2	2	3, 4	3, 4	4	4
2	1	1	1	2	2, 3	3	4	4
3	1	1	1	2	2	3	4	4
4	1	1	1	2	2, 3	3, 4	4	4
5	1	1	2	1, 2	2, 3	3, 4	3, 4	3, 4
6	1	1	2	2	3	3	3	4

Table 3. Individual participant progression throughout speech sessions. (1=session 1; 2 = session 2 and so forth; WM = word medial; Cl = Clusters). This table shows what each participant practiced by session number.

Participants each had four individual sessions and progressed at their own pace. Table 3 shows targets for each individual by session (1=Session 1, 2=Session 2 and so forth). Although there was slight variation based on segment (ex. /ɪ/ might be more advanced in certain word positions than /l/ at certain points during training), the general pattern of progression is shown in Table 3. Participant 1 did not add voice until session two. This is in contrast to participants 2, 3 and 4 who moved towards applying lingual positioning information learned into open syllables during the first session. Holding individual sessions allowed the clinicians to follow the client's lead and personalize training according to each individual's learning style and progression through training.

f) Home practice

Each participant was given the list of syllables or words that were used during training to practice at home. It was requested that each participant practice production of the word lists for at least 10 minutes each day.

Participants reported practicing between 3-20 minutes daily from the time between the last training session to the follow-up assessment (mean = 8.3 minutes daily).

Participants reported practicing speech production with word lists provided during therapy sessions, silently practicing lingual placement while walking, recording and imitating words from English television shows that contain target segments and being more conscious of lingual placement and movement during casual conversations in English.

2.5 Follow-up assessment

Two weeks following the final training session, participants were seen for an individual follow-up assessment. Although the researchers would have preferred to have a longer length of time to pass before assessing generalization of learned segments, the 2-week period was essential due to participants leaving for Christmas holidays and several participants moving back to Japan.

Extended perception task

The researchers concluded that the initial perception tasks did not contain an adequate number of tokens to accurately represent perceptual accuracy for target segments. Therefore, the number of minimal pairs was increased to 30 during the follow-up assessment. See section 2.5 for method of presentation. The extended minimal pairs word list can be found in Appendix H.

Practice with ultrasound

Participants practiced lingual components of /l/ and /ɹ/ with the author for a total of 5 minutes to re-orient themselves to the ultrasound machine prior to all assessment recordings. Practice was geared towards the level at which each individual was practicing during their final training session.

2.6 Outcome measures

2.6.1 Judgments of segmental accuracy

Audio and visual ultrasound recordings were made for segments /ɪ/ and /l/ in isolation with voice. Recordings were also done in CV open syllables in a variety of vowel contexts and target words recorded at baseline assessment were re-recorded with the same procedure. The words used in assessment were not targeted in the speech training sessions. For analysis in the current report, data were taken from the third elicitation because participants were more familiar with the word set and needed less cueing at the time of the third recording. As in the baseline assessment recordings, the ultrasound screen was not made visible to the participants during the follow-up recordings.

Twenty of the 40 elicited target words for each participant during assessment were semi-randomly selected to ensure that each word shape was included in the final analysis (eg. two word initial singleton words, two word medial singleton words and so forth). This means that randomization of word selection occurred within each word shape, rather than in the entire word list. The selected word list for analyses is included in Appendix D. These words were used during the baseline and follow-up assessment and were not trained or practiced during speech sessions. Therefore, this analysis suggests whether generalization of gains made during speech training occurred at the word level within a phrase. Although a variety of vowel contexts were systematically trained and assessed, there was an insufficient frequency of elicitations within each context to make a statement on significance.

Audio files of selected words were extracted from the carrier phrases with Audacity (Version 1.2.6). Both the pre-therapy and post-therapy productions of selected words for a total of 20 word pairs (40 words in total) for each participant. In total, 240 words were extracted by the author and used in the analysis across all participants (20 pre-therapy words + 20 post-therapy words x 6 participants = 240 words total). A visual schematic of this process is presented in Figure 2.

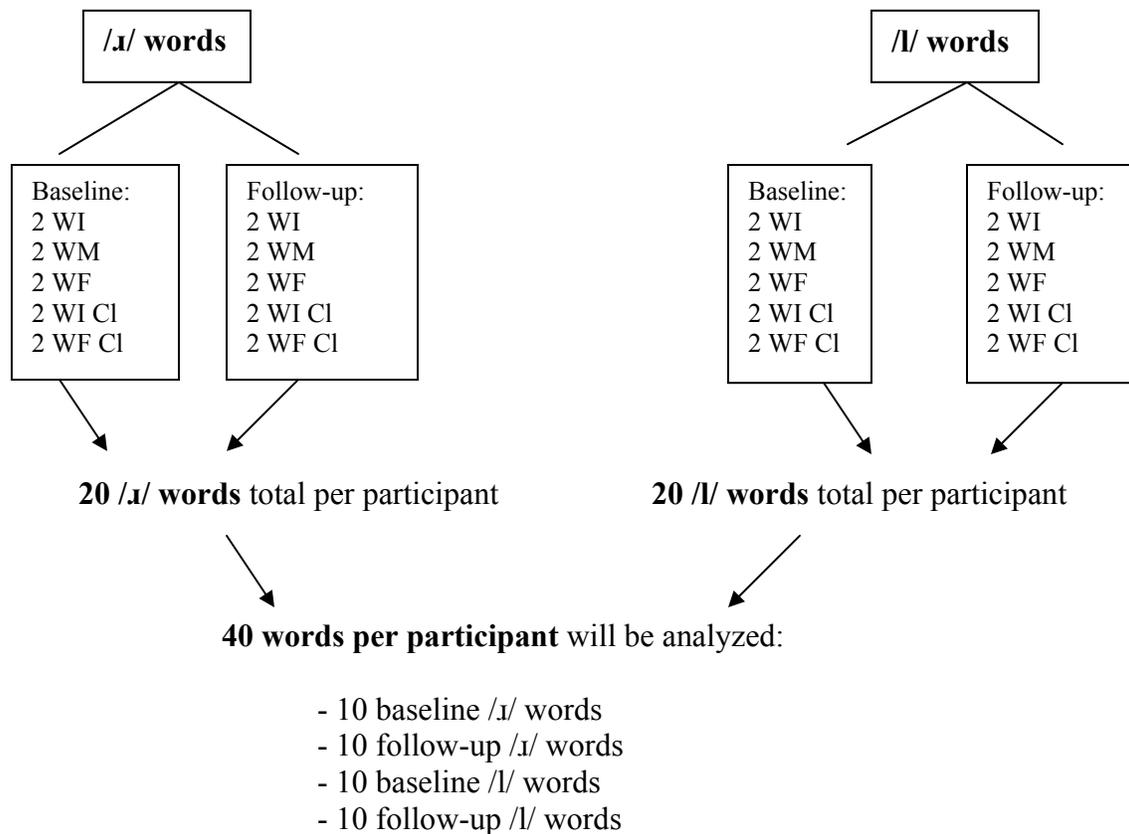


Figure 1. Schematic of word selection for analysis (WI=word-initial, WM=word-medial, WF=word-final, WI Cl=word-initial clusters, WF Cl=word-final clusters)

Two final year Masters students of Speech-Language Pathology who were native English speakers with normal hearing were recruited as judges for the recorded tokens.

One of the listeners had also been an English as a Second Language teacher with experience with native Japanese speakers learning English in Japan and Canada.

Listeners were informed of the general goals of the research study and the segments targeted. They were also informed that judgments should be made relatively, rather than on absolute accuracy of segments. For example, for words containing /ɪ/, listeners were asked to select which of the two tokens played one after the other had a more accurate rhotic quality, even if the /ɪ/ segment was not a flawless production (see details below). This method of evaluation allowed for an indication of improvement, even if the segments were not yet completely mastered in all word positions or phonological contexts (as in Bacsfalvi, 2010). Listeners were asked to disregard other consonant and vowel segments that were perceived as inaccurate and focus on the target segments. This has been done in previous listener judgment research in accent (Hazan et al., 2005). Additionally, listeners were given the list of target words to refer to if a word was unclear and could ask one of the researchers if they were unsure which word they were hearing.

Experienced listeners were chosen as judges for the study because of the need to notice subtleties for specific phonemes in accented speech. Because /ɪ/ and /l/ were the only phonemes targeted during therapy, it was crucial for listeners to make accuracy judgments based on those segments alone, and attempt to ignore other inaccurate segments in word productions. Thompson (1991) compared the use of experienced and inexperienced listeners in the judgment of accent accuracy in participants with Russian L1 and proficient English L2. The author reported that inexperienced raters generally perceived a higher degree of accent in an L2 and were less reliable and consistent than experienced listeners. The authors state that experienced listeners are perhaps better able

to perceive the effects of L1 phonology on L2 phonology as degrees on a continuum and subsequently be able to make more subtle relative accuracy judgments. The model of using expert listeners in speech therapy research is also supported by the literature for determining judgments of accuracy in the speech of adolescents with hearing impairments (Bacsfalvi & Bernhardt, 2011; Bacsfalvi, 2010).

Listeners were exposed to audio files from each participant consecutively. For instance, all of the pre- and post-therapy words were presented for one participant at a time. Once this was complete, all of the words from the next participant were presented. Additionally, listeners heard all words containing the /ɪ/ segment prior to listening to the words containing the /l/ segment. Presentation in this way was done to support the perception and accuracy judgments to focus of each segment and minimize the influence of judgments across participants or across the two segments.

The sound files containing the pre-therapy and post-therapy words were inserted into a custom-designed program to interface Microsoft Office PowerPoint 2003 (Version 11.8335.8341) and Microsoft Office Excel 2003 (Version 11.8346.8341). A screen shot of the program can be found in Appendix J. On each slide, participants clicked a button to play two words, one of which was a pre-therapy production and the other a post-therapy production by one speaker. The order of presentation was randomized, ensuring that listeners were blind as to which audio file was pre- or post-therapy. Since listeners clicked a button to play each word, it was certain that the listeners were prepared to listen to the audio file at each presentation. Listeners were asked to make an accuracy judgment for audio files and were permitted to play each word up to a maximum of three times in

order to facilitate thoughtful and informed ratings. However, the listener was asked to attempt to make her judgment after the first listening.

The listeners had four choices for submitting accuracy judgments for each pair of pre- and post-therapy words. The choices included selecting a button denoting that the first one was more accurate, the second word was more accurate, both were equally accurate (no difference, both accurate) or both were equally inaccurate (no difference, both inaccurate). The custom-designed program automatically recorded the listener's selection for each word pair. The task of listening to 20 pre-therapy and 20 post-therapy words for each participant took approximately an hour per listener.

2.6.2 Reliability

All responses from the two raters were compared for all words and across all participants to calculate inter-rater reliability. Responses were considered to be an agreement if both listeners either rated the pre-training word as more accurate, the post-training word as more accurate, both words equally accurate or both words equally inaccurate. Responses were considered to be not in agreement if either respondent rated a word differently than the other respondent.

Inter-rater reliability was 77% for words with the target /l/ segment and 80% for words with the target /ɪ/ segment. The majority of the inter-rater disagreements were in the instance of one rater judging a post-training word as more accurate and the other rater judging both words as equally accurate. It was very rare for one rater to judge a pre-training word as more accurate and the other rater to judge the post-training word as more accurate.

To determine intra-rater reliability, 20% of the tokens (4 words) for each participant were repeated during the rating tasks. The repeated words were randomly interspersed throughout the listening task for each participant. In other words, the repeated words were not presented consecutively during the listening task. After all listener judgment ratings were complete, the items that were repeated were compared to calculate intra-rater reliability. Intra-rater reliability was 95% for listener 1 and 93% for listener 2. The listeners were not informed that some of the words would be repeated during the rating task.

2.6.3 Acoustic and visual analysis

Two representative /ɪ/ words and two representative /l/ words from each participant were further analyzed. These included one word containing each segment that was considered to have a substantial improvement and one word containing each segment that did not improve based on the experienced listener ratings and the decision of the author.

2.6.4 Participant feedback survey

Participant or client satisfaction in speech-language pathology is often assumed based on the clinician's judgment that the client has improved whereas collecting feedback from individuals on their perception of improvement or the training sessions is uncommon (Manning, 2010). However, participant feedback is an important component of client-centered therapy and can provide the clinician with important information from the client's perspective.

The current study incorporated a participant feedback survey that each participant completed after each training session consisting of three questions (Figure 2). At the follow-up session, which was held approximately 10-14 days after the final training session, a longer questionnaire was completed by each participant (Appendix I). The interviews were conducted as part of a graduating project by Pasquini (2012), a second year speech and language pathology graduate student at The University of British Columbia. While all participants were willing to participate, one participant's responses were not received, making the total number of participants in the survey to be 5.

-
1. How was today's session? (circle one) EASY MODERATE DIFFICULT
 2. What did you find the most beneficial from today's session?
 3. Are there any other things you would like to suggest for the next session?
-

Figure 2. Optional questionnaire: post-session questions

For additional details on the qualitative portion of this study and a review on conducting client feedback within the field of speech-language pathology, please see Pasquini's (2012) graduating essay entitled "Measuring participant satisfaction concerning speech training with ultrasound for English /l/ and /ɹ/" from the School of Audiology and Speech Sciences at the University of British Columbia.

Chapter Three: Results

3.1 General outcomes and accuracy of /ɪ/

The words used for the analyses were elicited during the baseline and follow-up assessments only and were not targeted during training. Therefore, any improvement seen would likely be representative of generalization of gains made in therapy. A sign test of average ratings across participants for /l/ and /ɪ/ combined was significant ($p=.0063$ two-tailed). All outcome data are reported in Figures 3 and 4 for /ɪ/ and Figures 5 and 6 for /l/.

Both judges rated 52% of all selected words containing the target segment /ɪ/ to be more accurate in post-therapy productions. Both raters also judged seven words (15%) containing /ɪ/ as equally accurate pre- and post-therapy and two words (3%) as equally inaccurate by both raters. For a break-down of frequency of words rated in each category for each participant, see Table 4. As there were ten /ɪ/ word pairs (twenty /ɪ/ words in total) that each rater made an accuracy judgment on, each row totals ten.

Participant #	Average post-training words more accurate	Average pre-training words more accurate	Average words judged to have no change
1	5.5	3.5	1
2	7	1.5	1.5
3	5	3	2
4	8.5	1	0.5
5	3.5	2	4.5
6	5.5	2	2.5

Table 4. Frequency of /ɪ/ word ratings. Judgments are based on counts of words rated more accurate post-training, pre-training or no difference. Each row represents one participant. The values in the cells are averages calculated from both raters.

3.1.1 Accuracy of /ɪ/ by word position

See Figure 3 for accuracy of /ɪ/ in singletons by word position.

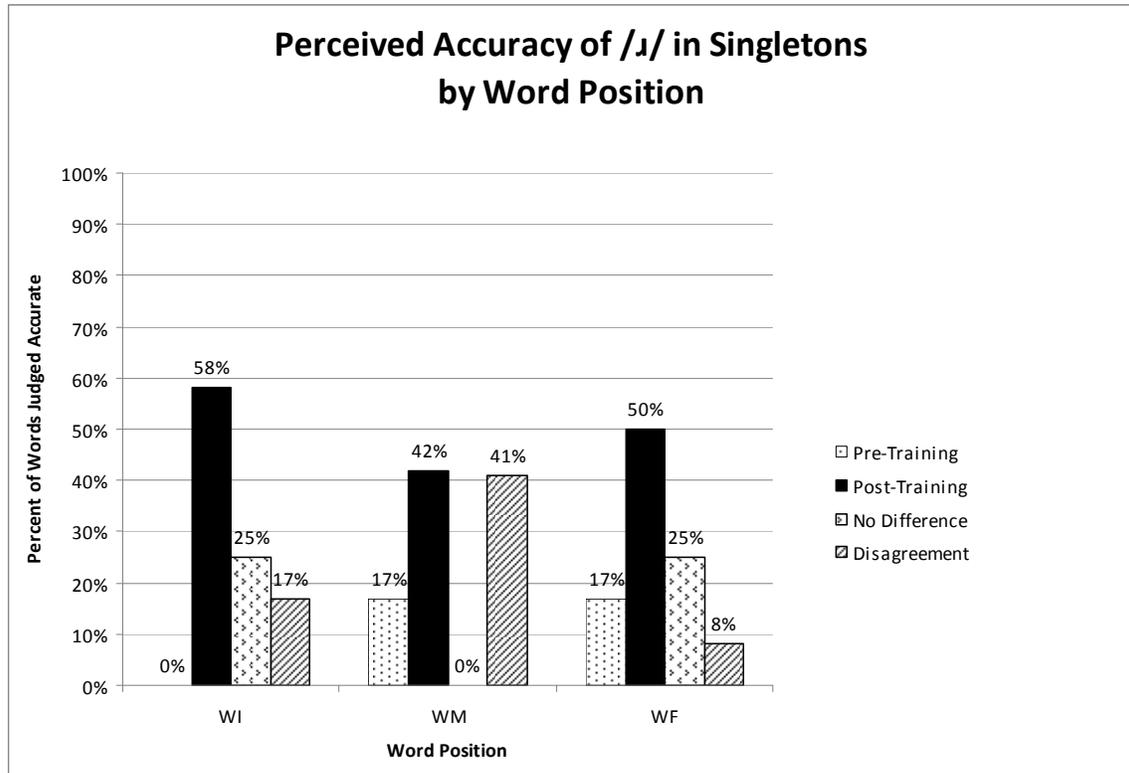


Figure 3. Perceived accuracy of /ɪ/ in singletons by word position. This figure shows the agreement of both listeners on which word (pre- or post-training) containing /ɪ/ was more accurate by word position (WI = word-initial; WM = word-medial; WF = word-final). It also shows the amount of words judged to be the same pre- and post-training (“no difference”) and the amount of inter-rater disagreement.

Word-initial and final /ɪ/ words are the contexts where the most improvement was observed across participants following training. For word-initial /ɪ/ words, 58% were judged as more accurate post-training. When the word pairs that were judged as equally accurate pre- and post-training are added to the analysis, word-initial segmental accuracy was judged to be 83% by both listeners.

For word-medial /ɪ/ words, 42% of were judged to be more accurate post-training by both raters. This was the word context with the highest amount of inter-rater

disagreement. For word-final /ɪ/ words, 50% were judged as more accurate post-training. When word pairs that were judged as equally accurate pre- and post-training are added to the analysis, word-final /ɪ/ segmental accuracy was judged to be 83% by both listeners.

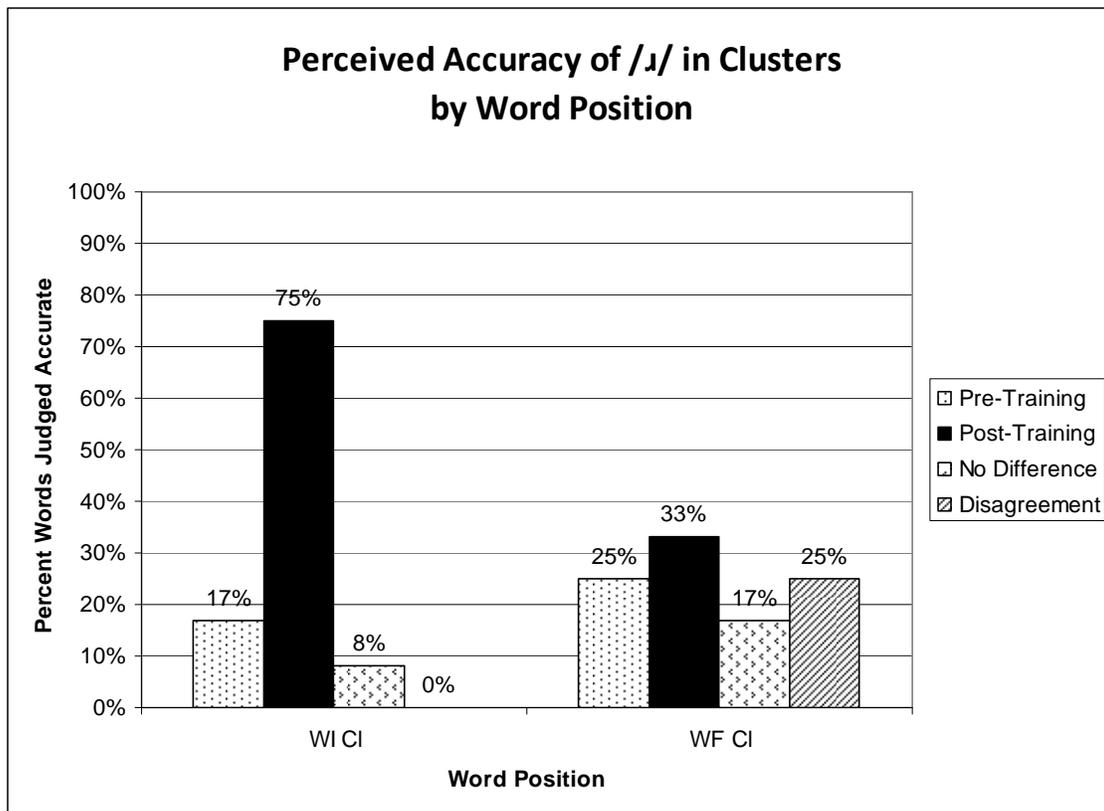


Figure 4. Perceived accuracy of /ɪ/ in clusters by word position. This figure shows the agreement of both listeners on which word (pre- or post-training) containing /ɪ/ was more accurate by word position (WI CI = Word-initial clusters, WF CI = Word-final clusters). It also shows the amount of words judged to be the same pre- and post-training (“no difference”) and the amount of inter-rater disagreement.

Figure 4 shows the listener judgment outcomes for /ɪ/ clusters. For word-initial /ɪ/ clusters, 75% were judged to be more accurate post-training. When the word pairs that were judged to be equally accurate pre-therapy and post-therapy are added to the analysis, 83% word-initial cluster /ɪ/ segments were judged to be accurate. Finally, 33% of word-

final clusters were judged to be more accurate post-training, but 25% were judged to be more accurate post-training and there was a high degree of inter-rater disagreement as well.

3.2 General outcomes and accuracy of /l/

Of all selected words containing /l/, 62% were judged to be more accurate post-training by both raters across word positions. Both raters also judged five words (12%) as equally accurate pre-training and post-training and one word (2%) was judged to be equally inaccurate. For a break-down of frequency of words rated in each category for each participant, see Table 5. (Note that the combined average ratings were significant on a sign test as noted above [$p=.0063$].)

Participant #	Average post-training words more accurate	Average pre-training words more accurate	Average words judged to have no change
1	6	1	3
2	8	1	1
3	7	2	1
4	9	1	0
5	5	4	1
6	7.5	1	1.5

Table 5. Frequency of /l/ word ratings. Judgments are based on counts of words rated as more accurate post-training, pre-training or no difference. Each row represents one participant. The values in the cells are averages calculated from both raters.

3.2.1 Accuracy of /l/ by word position

See Figure 5 for a summary of accuracy of /l/ productions by word position.

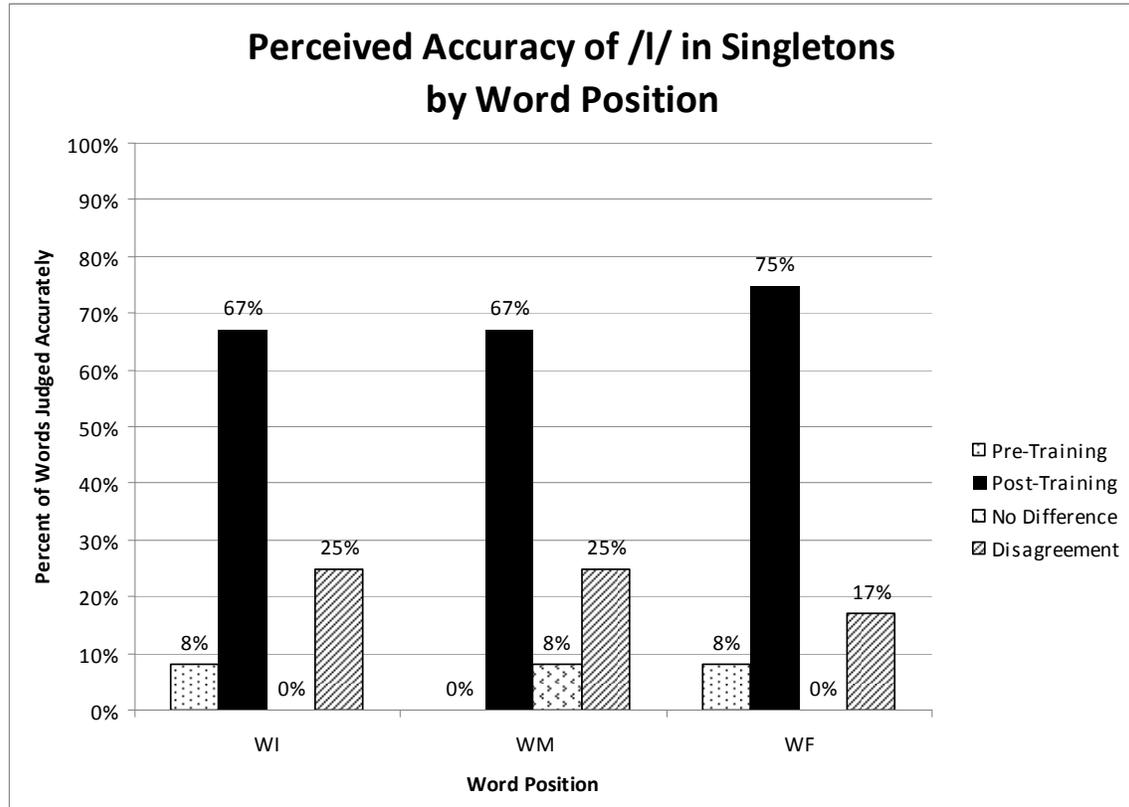


Figure 5. Perceived accuracy of /l/ in singletons by word position. This figure shows the agreement of both listeners on which word (pre- or post-training) containing /l/ was more accurate by word position (WI = word-initial; WM = word-medial; WF = word-final). It also shows the amount of words judged to be the same pre- and post-training (“no difference”) and the amount of inter-rater disagreement.

Of the /l/ segments in word-initial position, 67% were judged to be more accurate post-training. When /l/ segments that were judged as equally accurate pre- and post-training are included in the analysis, 83% of all word-initial /l/ segments were judged as accurate.

For word-medial /l/ words, 67% were judged to be more accurate post-training. When medial /l/ segments that were judged as equally accurate pre- and post-training are included in the analysis, 75% of all word-medial /l/ segments were judged as accurate. For word-final /l/ words, 75% were judged to be more accurate post-training. When final /l/ segments that were judged as equally accurate pre- and post-training are included in the analysis, 83% of all word-final /l/ segments were judged as accurate.

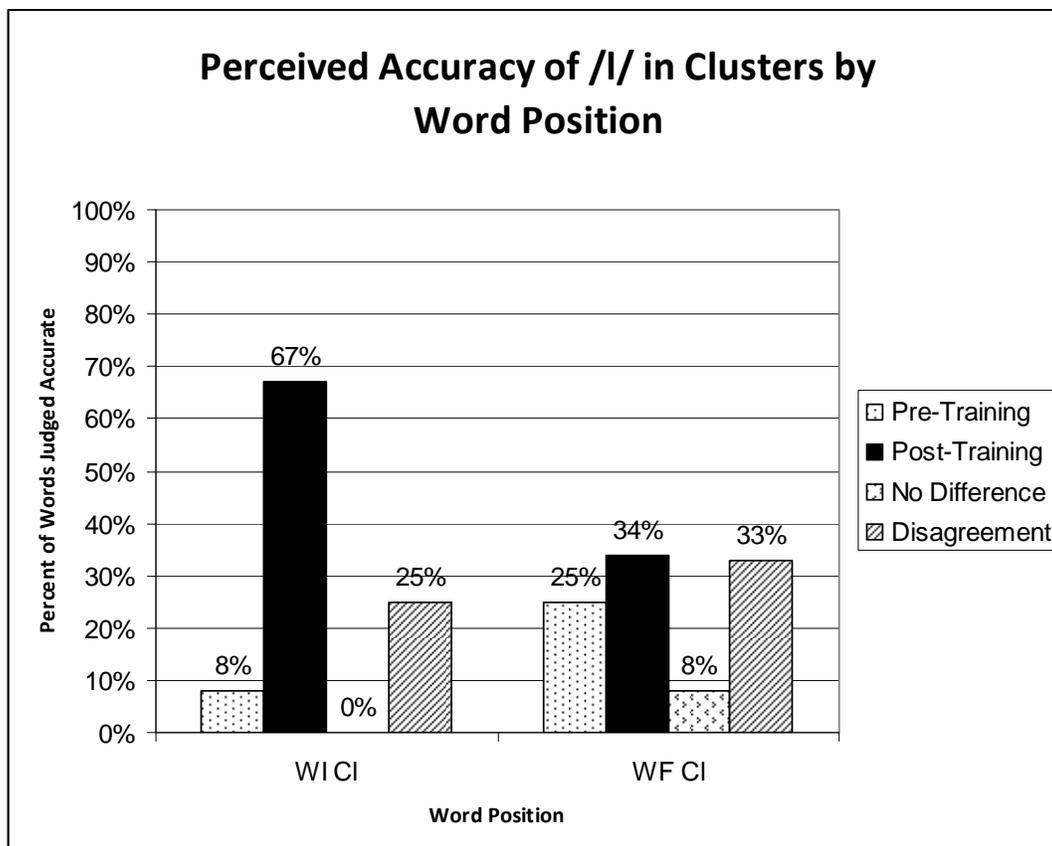
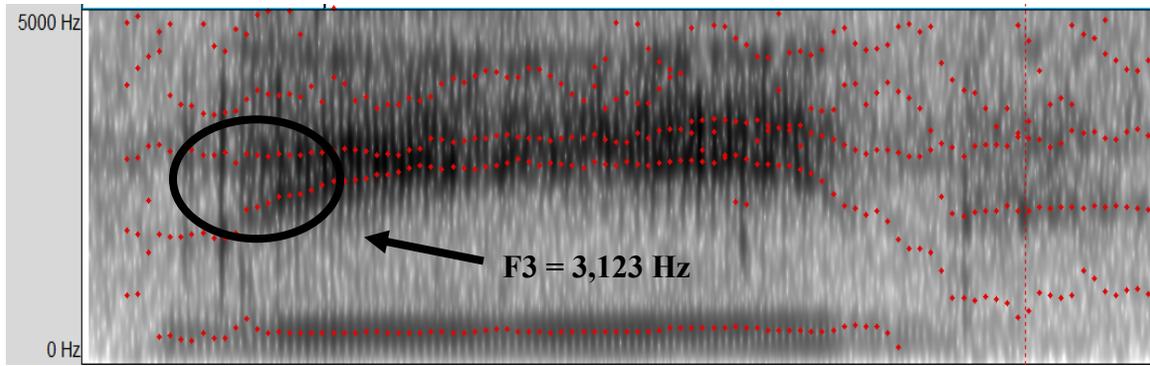


Figure 6. Perceived accuracy of /l/ in clusters by word position. This figure shows the agreement of both listeners on which word (pre- or post-training) containing /l/ was more accurate by word position (WI CI = word-initial clusters, WF CI = word-final clusters). It also shows the amount of words judged to be the same pre- and post-training (“no difference”) and the amount of inter-rater disagreement.

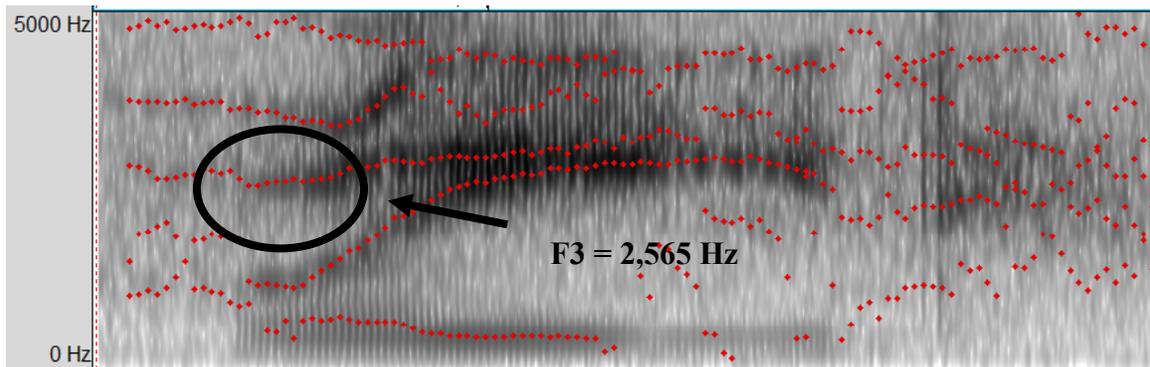
For word-initial /l/ clusters, 67% were judged to be more accurate post-training and 33% of word-final /l/ clusters were judged to be more accurate post-training.

3.3 Acoustic and visual analysis

The acoustic properties of the most improved words were analyzed using Praat version 5.3.22 (Boersma & Weenink, 2012). For the /l/ words that did not show improvement, there was no change in the F3 values between the words. In fact, the average values were an average of 15 Hz higher in post-training productions. No significant differences were observed in tongue configuration during a comparison of images from these productions. For most improved /l/ words, the most notable acoustic change was that F3 onset values dropped an average of 544 Hz across the examples of the participants' most improved words. The male productions dropped from 2,814 Hz in pre-training to 2,120 Hz in post-training, which is closer to values seen in male native English speakers, which is approximately 1,548 Hz (Dalston, 1975). The F3 in the female participants also dropped from 2,834 Hz in pre-training to 2,438 Hz in post-training. This is also closer to female native English speaker values in F3 for /l/, which are typically around 2,078 Hz (Dalston, 1975). See Figure 7 for a pre-training and post-training spectrogram of the word "Read" spoken by Participant 1, demonstrating a drop in F3 onset value.



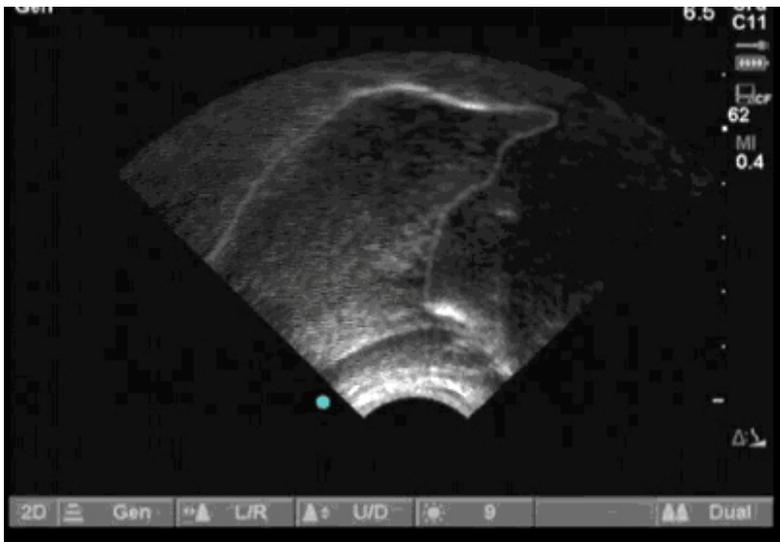
Pre-training production of /ɪ/ in "Read"



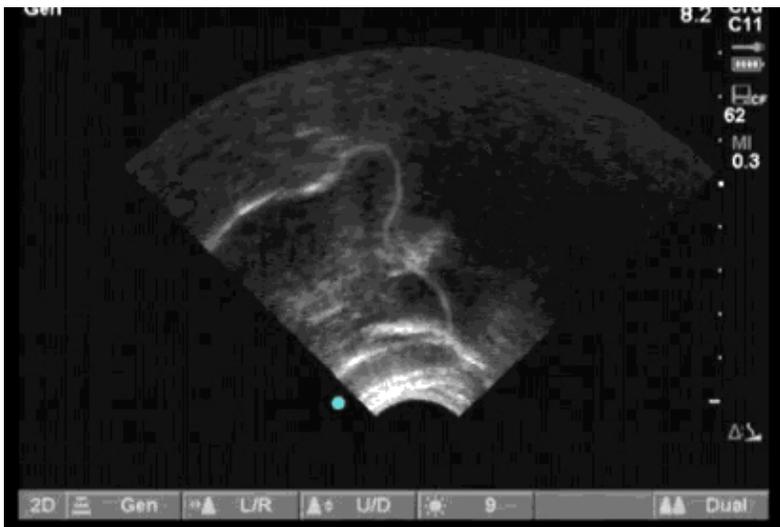
Post-training production of /ɪ/ in "Read" by same speaker

Figure 7. Spectrograms of one participant's improved /ɪ/ word

A visual analysis of ultrasound images of most improved /ɪ/ words showed all of the lingual components for /ɪ/ present in the post-training session videos, demonstrating that the articulatory patterns taught during the training phase were positively affecting the acoustic properties and perceived accuracy of the spoken word. In several of the pre-training videos analyzed, the tongue tip was raised to the alveolar ridge, but was not retroflexed, or curled back. An example of the ultrasound images for the most improved /ɪ/ word spoken by Participant 1 is presented in Figure 8 and shows the tongue blade up and retroflexed in the post-training production. This is the same word by the same speaker that was represented in the spectrograms in Figure 7.



Pre-training production of /ɪ/ in “Read”



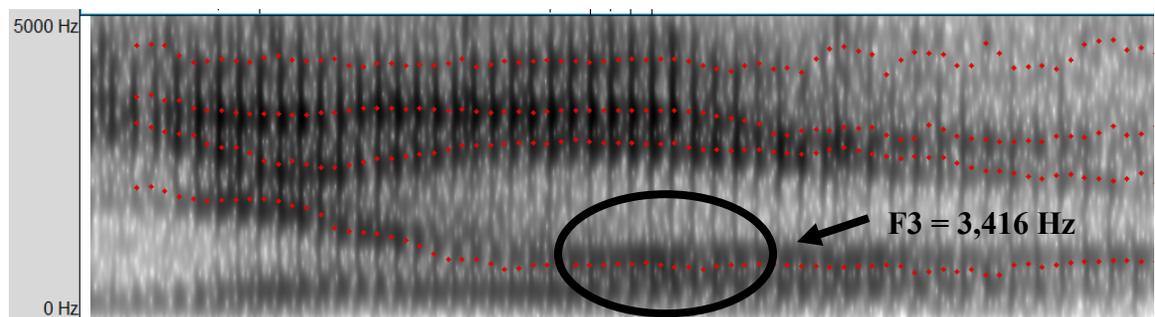
Post-training Production of /ɪ/ in “Read” by the same speaker

Figure 8. Ultrasound images of one participant’s improved /ɪ/ word

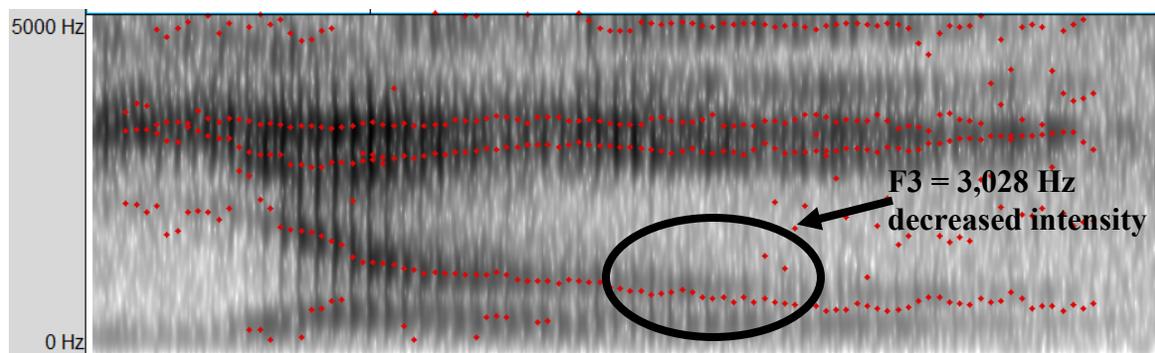
For the words where /ɪ/ did not show improvement, there was no difference between pre-training and post-training productions either in acoustic or visual analysis. For the most improved words containing /ɪ/, the F3 frequency dropped by an average of 567 Hz. For the male Japanese participants, this value dropped from an average of 2,809 Hz in pre-training to 2,242 Hz in post-training. This is within the average range for male

native English speakers, which is typically around 2,523 Hz (Dalston, 1975). For the female Japanese participants, the F3 value also dropped from 3,078 Hz in pre-training to 2,830 Hz in post-training productions. This is also within the average range for female native English speakers, which typically produce /l/ with an F3 frequency of approximately 2,935 Hz.

The analysis of participant two's productions of the word "yellow" are included here. An acoustic analysis revealed a decrease in intensity in F3 during /l/ production and a slight decrease in frequency from 3,416 Hz to 3,028 Hz (Figure 9).



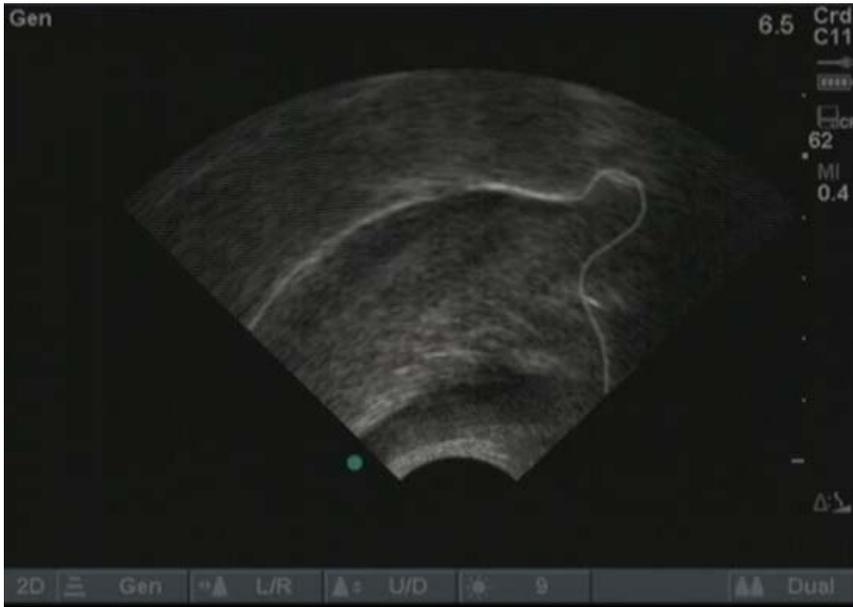
Pre-training production of /l/ in "Yellow"



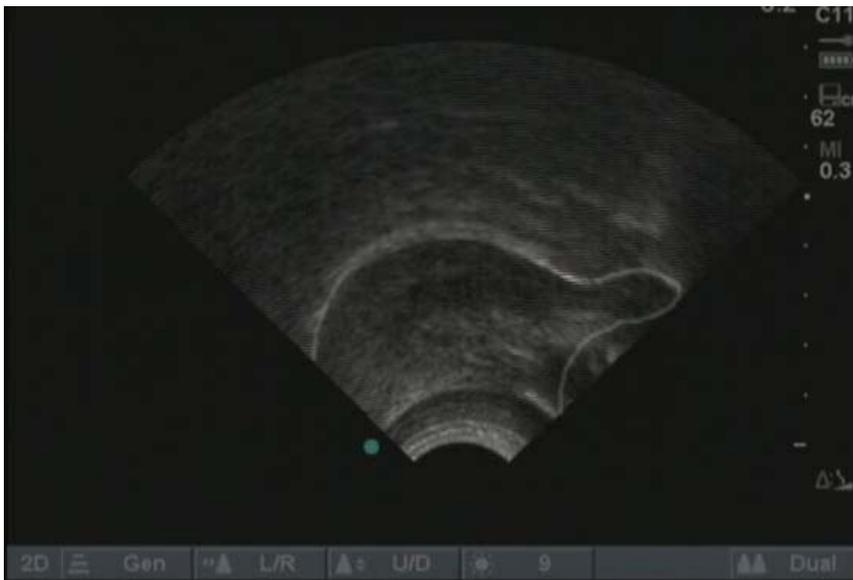
Post-training production of /l/ in "Yellow"

Figure 9. Spectrograms of one participant's improved /l/ word.

For the ultrasound images of the same word by the same speaker, more of the tongue's anterior surface was used in post-training productions, including the tongue blade. The tongue also appeared flatter during the post-training production (Figure 10).



Pre-training production of /l/ in “Yellow”



Post-training production of /l/ in “Yellow”

Figure 10. Ultrasound images of one participant's improved /l/ word.

3.4 Perception task

The perception task was exploratory and modified throughout the study in order to increase the reliability of the data collected. For this reason, statistical analysis was not performed on this data. However, the raw results from the perception task for each participant can be found in Figure 11.

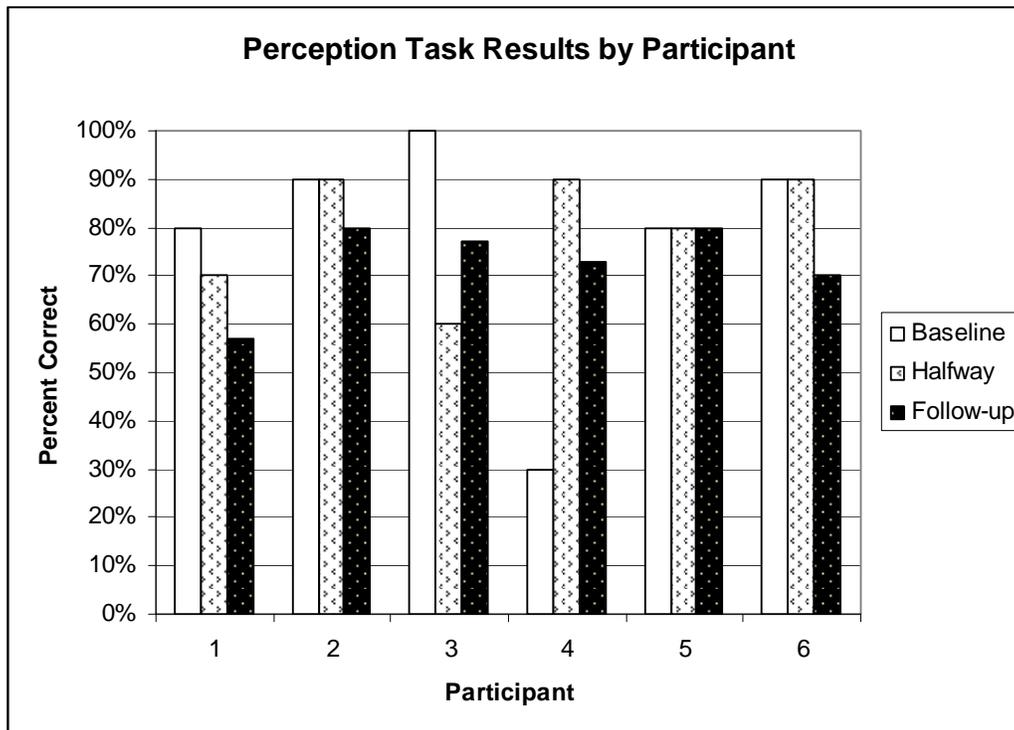


Figure 11. Perception task results by participant at three elicitation points (baseline, halfway throughout training sessions and at a two week follow-up). It should be noted that there were ten minimal pairs at baseline and halfway points and thirty minimal pairs at the follow-up session.

It should be noted that 10 minimal pairs were presented at the baseline and halfway point (after the second training session), whereas 30 minimal pairs were presented at follow-up. This was due to the retrospective consideration that 10 minimal

pairs was an insufficient sample to judge perceptual ability and resulted in the data collected at baseline and halfway to be highly variable both within and across subjects. Data collected at follow-up (dark bar with white dots) included more minimal pairs and was more consistent across participants. There was no difference in segment or word position on perceptual accuracy. This may also be due to the small number of minimal pairs presented to the participants.

3.5 Participant feedback questionnaire

All participants rated their pre-training spoken accuracy at “somewhat on target” for both /l/ and /ɪ/. After training, all participants increased their self-reported accuracy ratings to “almost on target”, indicating that all participants believed that they had made gains in training.

Chapter Four: Discussion

The current study demonstrates that typical native Japanese speakers can be trained to increase their accuracy in production of /l/ and /ɭ/ in a variety of word positions, both as singletons and in clusters, even if perceptual accuracy is not flawless. These results establish the benefits of using visual feedback through the use of ultrasound to teach tongue configuration and placement of non-native segments for Japanese speakers. Participants used ultrasound during all training sessions for learning and practicing both /l/ and /ɭ/ and were able to improve production immediately after the first session and continue to extend improvement to a variety of syllable and word contexts throughout the four training sessions. All participants maintained gains made in training and generalized to new words during the follow-up assessment session that occurred two weeks after the final training session. Participants had very positive feedback about the use of ultrasound in speech training and all self-judged their own productions to have improved from pre-training accuracy levels. Furthermore, the pattern of individual progress throughout the sessions demonstrates the importance of providing individualized assessment and training.

4.1 Production of /ɭ/

4.1.1. Word-initial, medial and final /ɭ/

For word-initial and medial /ɭ/ words, 58% and 42% of words were judged to be more accurate than pre-training productions, respectively. Speech training in the current study focused on the retroflexed variant of /ɭ/, although it can be produced as retroflexed or bunched (Bernhardt & Stemberger, 1998). The ultrasound was used in the mid-sagittal orientation to view the tongue tip up behind the alveolar ridge, the tongue back raised and

the entire body of the tongue shifted posteriorly in the oral cavity. The ultrasound was used in the coronal view orientation to visualize medial grooving and lateral bracing against the upper back teeth.

For word-medial, there was a high degree of inter-rater disagreement at 41%. The medial position of the word may be influencing what aspect of the segment that listeners are attending to. The segment itself may be influenced by the preceding or following vowel and listeners may be differentially attending to different aspects of /ɪ/ depending on whether they are attending to the beginning or end of the segment. The duration of word medial segments is often shorter than when in initial or final position (Oller, 1973), influencing the difficulty of attending to the segment inter-vocally and judging its accuracy. These factors may have contributed to the high amount of inter-rater disagreement for word medial /ɪ/.

Production improvements in word-initial and medial /ɪ/ may be facilitated by their ease in perception as well. Both Mochizuki (1981) and Sheldon and Strange (1982) found that Japanese listeners had more difficulty in perceptual categorization tasks of /l/-/ɪ/ when they were part of a consonant cluster. Hazan et al. (2005) also found greater improvement in perceptual categorization of the same segments in word-initial in singletons as compared to word-initial and medial clusters. These studies suggest that perception of these novel segments may be easier as singletons which may contribute to improvements in production accuracy. However, it should be noted that participants in the current study also showed highly significant improvement in word-initial clusters for both /l/ and /ɪ/.

For word-final /ɹ/, there was no difference in accuracy ratings between pre-training and post-training tokens. The absence of improvement may reflect the training sessions that focused only on retroflexed /ɹ/ and did not distinguish between retroflexed /ɹ/, bunched /ɹ/ or rhotacized vowels. Although the difference between retroflexed and bunched /ɹ/ is a difference in tongue shape, the difference between these and rhotacized vowels is more about articulatory timing than tongue position. Specifically, the tongue root for postvocalic /ɹ/ tends to move towards the palate sooner than prevocalic /ɹ/ (Secord, 2007). Additionally, lip rounding is often present in prevocalic /ɹ/ (Bernhardt & Stemberger, 1998) and was taught during training in the current study. Differentiating between different articulatory patterns for /ɹ/ in various contexts was not the focus of training in the current study and possibly contributes to the lack of improvement in word final /ɹ/ words.

Additionally, the syllable structure of Japanese only rarely contains codas and ‘superheavy’ syllable structures containing long vowels and codas are considered marked segments and are only found rarely in loanwords from other languages such as /kooN/ (‘cone’) (Ota & Ueda, 2007). Loanwords are, however, pronounced with repair strategies to prevent coda consonants through the use of epenthetic vowels, namely [u] after all consonants except /t, d/ and /tʃ, dʒ/ in which case [o] or [i] is typically inserted, respectively. Therefore, the combination of producing a novel segment (/ɹ/) within the context of a novel position (syllable coda) and a novel configuration (consonant cluster) undoubtedly contributes to the difficulty the participants had in production of word-final /ɹ/.

4.1.2 /ɹ/ clusters

For /ɹ/, greatest improvement was noted across participants in word-initial clusters. These findings are consistent with previous research that also found the most improvement in word-initial cluster production after a speech training program for native Japanese speakers, despite segments in this position being the most difficult to identify perceptually and to produce during training (Massaro & Light, 2003).

It is intriguing that production of /ɹ/ improved more in clusters than as a singleton, especially considering that Japanese phonology disallows consonant clusters, with the exception of a rare syllable onset cluster with /j/ in second position. Coupled with the absence of liquids in the Japanese phonological inventory, consonant clusters containing /ɹ/ should be very difficult to acquire for adult L1 Japanese speakers. However, /ɹ/ in word-initial clusters was determined to be the most accurate position in the follow-up assessment and the word position where the most gains in accuracy were made.

Although Japanese phonology does not contain any liquid or clusters with plosive segments, it does possess consonants /t/ and /d/ in word-initial and medial positions (Ota & Ueda, 2007). Interestingly, one technique for teaching production of the /ɹ/ segment to native English speaking children is to start with production of /d/ to achieve alveolar tongue placement and slowly move the tongue back to encourage retroflexion of the tongue tip immediately beyond the alveolar ridge. Starting with the /d/ segment also helps initiate the lateral edges of the tongue to contact the molars, or medial grooving that is also an important lingual component for production of /ɹ/. This method encourages learning /ɹ/ within a homorganic cluster after the /d/ segment to attain the multiple lingual components and appropriate tongue placement for /ɹ/. Since native Japanese speakers do

have experience with /d/ and /t/ as singletons, producing a cluster may have facilitated tongue placement for /ɹ/ despite Japanese phonology disallowing clusters.

Within word-final clusters, /ɹ/ showed a small amount of improvement across all participants. There are several reasons why this is not a surprising result, including differences between Japanese and English: phonological inventories, syllable structure and the general focus of the training program of the current study. As mentioned, Japanese phonology does not contain clusters with the exception of a rare /j/ in second position and these unique clusters are only permissible in syllable onset position. Retroflexed /ɹ/ was also the main focus of the training sessions and co-articulation effects of /ɹ/ within word-final clusters were only addressed sparingly with some participants.

4.2 Production of /l/

For all participants, /l/ showed greater improvement than /ɹ/ in all word positions with the exception of word-final clusters. However, this may be due to the very low accuracy judgments of /l/ at baseline and therefore there room for major improvement during training. This is in contrast to some studies that have found greater improvement of /ɹ/ than /l/ (Ogitsu, 2009; Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). However, it should be noted that neither of these studies implemented a speech production training program. Ogitsu (2009) examined the perceptual characteristics of native Japanese speaker's /l/ and /ɹ/ productions whereas Aoyama et al. (2004) studied the effects of immersion on production accuracy, not on specific training programs. Therefore, while Aoyama et al. (2004) confirmed their predictions that /ɹ/ would improve more than /l/ because /l/ is more similar to Japanese /ɹ/, the authors did not attempt to

specifically target /l/ or /ɹ/ production. In their study, none of the native Japanese children (9 years old) or adults improved production accuracy of the /l/ segment after immersion in an English-speaking environment for one year.

The current study found highly significant improvement of /l/ in all word positions with the exception of word-final clusters. The focus of training for /l/ was using the ultrasound in the mid-sagittal orientation to ensure that the tongue blade was being used for articulation. Prior to training, it was observed that all participants were using the tongue tip to articulate /l/ and did not sustain tongue placement, contributing to their productions resembling an alveolar /ɹ/ than English /l/. The clinicians modeled English /l/ production using ultrasound and highlighted the importance of using the tongue blade and holding the tongue in the position for a longer period of time to avoid the sharp movement associated with tap production. The participants proceeded to use ultrasound to practice tongue positioning and movement for English /l/.

4.2.1 Word-initial, medial and final /l/

Both word initial and medial /l/ were judged to be more accurate post-training. This is especially encouraging for these word-medial segments because none of these segments were judged to be more accurate in the pre-training recordings.

Due to its extremely gradual improvement from immersion (Flege et al., 1995; Ayoama et al., 2004) and its feature similarity to /ɹ/, the results for /l/ are encouraging. One participant stated during a speech session that /l/ was very difficult to pronounce, but learning about the difference between using the tongue tip and blade was very useful in distinguishing /l/ from other sounds.

Of word-final /l/ words 75% were rated as more accurate post-training, making this segment in this position to have the largest increase in accuracy. Although /l/ in word final position is typically velarized in English, this difference was not directly mentioned to the participants, but was modelled by the clinicians during training with ultrasound.

4.2.2 /l/ clusters

Post-training word-initial /l/ clusters were rated as significantly more accurate than pre-training productions. Word-final clusters containing /l/ did not show a significant difference between pre-training and post-training words. Insight into why this was a difficult position to generalize to can come from loanwords in the Japanese language that have been borrowed from other languages, such as /mi.ru.ku/ (for ‘milk’), that demonstrate the rules of Japanese phonology to disallow clusters. The loanword breaks the /lk/ cluster into two distinct, open syllables. The /l/ is absent in the Japanese phonological inventory and is subsequently changed to /r/, whereas /k/ remains due to its presence in both English and Japanese phonological inventories. A comparison of Japanese and English phonologies, including segmental inventories and syllable structure, lends insights into the difficulties the participants of the current study encountered in learning to produce /l/ and /ɹ/ in word-final clusters. The combination of the novel segment (/ɹ/ or /l/) within a novel structure (consonant clusters) in a novel phonological environment (syllable coda position) for native Japanese speakers learning English contribute to the small amount of improvement for this word position. Also, velarization was not discussed or trained during the speech sessions, which may have contributed to

the difficulty of word-final clusters. Additional sessions may have been beneficial in this regard.

4.3 Acoustic and visual analysis

The acoustic analysis of each individual's most improved /ɪ/ words demonstrated that the third formant (F3) was lower in post-training productions of /ɪ/ at the onset of the segment. The average F3 onset value for pre-training words was 2964.28 Hz, which is slightly higher than the values found in Japanese speakers in Lotto et al. (2004). The average post-training F3 value was 2457.48 Hz. The lower F3 values correspond more closely with typical acoustic values for L1 American English speakers (Flipsen et al., 2000), but are still slightly higher than values from native English speakers which typically show values at or below 2000 Hz (Dalston, 1975; Lotto et al., 2004). This is an important consideration for using experienced listeners who may be better able to perceive more subtle improvements in speech, even if the acoustic signal is not entirely equal to native speaker productions.

The visual analysis of ultrasound tongue images demonstrate that the lingual components taught during training were present during post-training productions when viewing the tongue in the mid-sagittal orientation. In pre-training productions of /ɪ/, all participants were using the tip of their tongue at the alveolar ridge and did not demonstrate retroflexion. Training included teaching medial grooving if the participant was lacking this lingual component in the assessment. However, the post-training assessment recordings of ultrasound images were done in the mid-sagittal view and therefore it is unclear whether medial grooving was maintained at follow-up. In the mid-

sagittal view, it was observed that participants were able to incorporate the lingual components of /ɹ/ in their productions even when the visual feedback was removed and they could not view the ultrasound screen.

The acoustic analysis for each individual's most improved /l/ words also showed a drop in F3 from pre-training to post-training. Frequency of F3 in pre-training productions of /l/ was within the range of 2,000 – 3,400 Hz, which is similar to the range found by Japanese speakers in Lotto et al. (2004). The average F3 frequency for pre-training /l/ was 3,087 Hz and the average for post-training /l/ was 2,536 Hz. This lower frequency is closer to the typical F3 onset frequency value seen in typical native English speakers, which is approximately 2,400 Hz (Lotto et al., 2004).

A visual analysis of /l/ productions showed less use of the tongue tip, as observed during pre-training assessments, and a more distributed use of the tongue blade during /l/ productions. This is reflective of how /l/ was taught during training.

4.4 Perception

Research on native Japanese speaker's difficulties in learning English /l/ and /ɹ/ is heavily focused on the ability to perceive the difference and aims to increase auditory discrimination abilities between the two segments. There has been a recent trend towards computer-assisted training programs (CAPT) that provide binary accuracy feedback to listeners (right or wrong) in order to training perceptual abilities in this contrast.

The current study included an exploratory assessment of perceptual accuracy of distinguishing /l/ and /ɹ/ at baseline and follow-up sessions. However, perception was not targeted in the speech training sessions. Although the number of minimal pairs used in

the auditory discrimination task was retrospectively deemed to be too small to adequately represent a participant's actual ability, the perception task did provide the researchers with some insight in this area prior to speech training because all of the participants had errors on some minimal pairs both prior to training and at follow-up post training. There were no specific words that showed particular difficulty for perception.

None of the participants performed without errors on perceptual distinction tasks, but all participants were still able to make significant production gains on both /ɪ/ and /I/ segments in a variety of word positions. The clinical implications are that perception does not need to be perfected in order to train production of segments for second language learners. This is consistent with other research studies with native Japanese speakers that have also found improvement in speech production of /I/ and /ɪ/ in the presence of remaining perceptual errors of the same contrast (Sheldon & Strange, 1982).

Although our perceptual accuracy data is too small to draw conclusions about the impact of production training on perception of the same pair, there is scant and conflicting evidence in the literature whether production training can have positive effects on perceptual accuracy for native Japanese speakers learning English. One study found no effect on production training on perceptual accuracy for the /ɪ/-/I/ contrast although significant gains were made in production accuracy for all six participants (Sheldon & Strange, 1982). An electropalatography (EPG) study of the same phonemic contrast states that their two native Japanese speaking participants self-reported better perceptual accuracy after learning the articulation distinction even though perception was not the focus of therapy (Gibbon, Hardcastle, & Suzuki, 1991). However, this was self-reported and not systematically tested. Based on the participant feedback questionnaire,

some participants of the current study also reported that the perception task was beneficial to include as part of the speech production training.

Hazan, Sennema, Ida and Faulkner (2005) examined the perception-production relation in the reverse direction in their large study with 62 native Japanese learners of English. Training was focused on the perceptual distinction using both audio only and audio-visual stimuli with the Baldi talking head computer software. The authors report significant gains made in production accuracy of /l/ and /ɪ/ in their participants in the absence of specific production training. The authors also note that visualizing the facial gestures of the talker, even with a synthetic face, can positively affect production accuracy, especially when the articulators within the oral cavity are made visible. The current study used ultrasound of the participant's own tongue, which may provide even greater benefits due to the personalization of feedback provided because ultrasound allows visualization of articulators that would not normally be visible due to their location within the oral cavity. Using feedback with the participants' own tongues also helps the clinician to individualize training.

4.5 Participant satisfaction questionnaire

Participants were given the option of completing a participant satisfaction questionnaire at the time of the follow-up assessment (Appendix I). The questionnaire and analysis were conducted with each participant by Pasquini (2012), a final year graduate student in speech-language pathology at the University of British Columbia.

All of the participants reported the orientation and use of the ultrasound during speech training to be helpful and commented that they thought about the images of the

tongue during speech tasks in everyday conversation for speech production accuracy. Many participants also found that the perception tasks were useful to include in speech training, despite perception not being taught during training or feedback about accuracy provided. Some participants stated that they would have appreciated receiving feedback about their perception accuracy from minimal pairs which could be included in future speech training research.

4.6 Limitations and future research

The results of this study are highly encouraging for using ultrasound in second language speech training. However, there are a number of factors that can be considered in future research to improve the empirical design of such research. The inclusion of a control group who does not receive any training would be beneficial to validate that the outcome measures were in fact a result of the training. It would be interesting to also include a third group who receives training without visual feedback of ultrasound and compare the gains made in training. The participants in the current study made gains very quickly and it would be useful to compare whether the trajectory and speed of learning is distinct from traditional non-technological forms of teaching.

The use of the repeated carrier phrase “*I want to see ...*” has some drawbacks. Although it did provide a context for each target word to be elicited, its ending with a high front vowel may have interacted with /l/ and /ɪ/ production, especially in word-initial position. Its repetition for all words also resulted in some nonsensical sentences being produced, especially when the target word was a verb such as “*I want to see read*”. Although repetition of the carrier phrase in its exact form is beneficial for controlling any

extraneous variables interacting differentially with the target segments, it may eventually become more automatic with several repetitions. Additionally, since the vowels for assessment and training words were carefully selected to represent a wide range of contexts, there were not a high number of different words within each vowel context and therefore, vowel type was controlled for but not studied as a contributing factor to production accuracy.

Much of the research in second language learning involves mention of the interaction between perception and production, but results are often nebulous. The current study retrospectively deemed the perception task to be too simplistic at baseline assessment. According to the participant feedback survey, all participants felt that the perception task positively benefited production of English /l/ and /ɹ/. In the future, it would be useful to include a more substantive assessment of minimal pair recognition with at least 30 minimal pairs to obtain a more representative sample of each participant's ability. Ultrasound may be a unique tool to use with perception tasks by pairing audio and visual tongue information. Whether this is a useful training technique to teach perception of non-native speech contrasts remains to be seen.

It is possible that if synthetic models of human speech production can improve second language learners in production and perception of novel segments (Hazan, Sennema, Iba, & Faulkner, 2005), visual feedback of the speaker's own tongue movements and positioning, with guidance from a trained clinician, could possibly be even more beneficial in learning novel speech sounds. Future research could determine the impact of the individualized visual support provided by the ultrasound that is viewed in real-time during speech production and its potential to support the perceptual

development of non-native speech contrasts. A comparison of the visual feedback provided with ultrasound could be compared with a general synthetic model, such as the Baldi talking head, to determine the impact of receiving feedback that is particular to each speaker.

Our focus was on native Japanese speakers and the English /l/- /ɹ/ contrast because of its unique difficulty in second language acquisition and to maintain a homogenous sample with which training sessions could be organized. To the author's knowledge, this is the only study to date that has used ultrasound for second language speech training, with the exception of the informal pilot study by Gick et al. (2008). It would be beneficial to extend the investigation to other language families to determine whether a similar trajectory of learning is achievable. Additionally, all of the participants in the current study were in their mid-twenties, which is common for speech training research. However, it would be interesting to use ultrasound in second language training of other ages, including younger adolescents and older adults.

4.7 Conclusion

The current study demonstrated very promising benefits of using ultrasound in teaching Japanese speakers to produce the English /l/- /ɹ/ contrast. In a period of only four sessions over two consecutive weekends, all participants, who were typical language learners, increased their accuracy of producing English /l/ and /ɹ/ in a variety of word positions and phonetic contexts. The visual, real-time feedback of the tongue provided during speech production tasks allowed participants to practice the specific lingual configurations for these sounds in order to better distinguish them into separate phonemic

categories. One-on-one assessments and training allowed the researchers to individualize the training sessions to each participant's progress while still moving through the same order of tasks. In addition to demonstrating significant gains made in a variety of word positions for both /l/ and /ɹ/ segments, which were measured through expert listener ratings, all participants increased their self-reported accuracy ratings, stated that they enjoyed using the ultrasound and found it extremely useful for learning to pronounce these difficult sounds.

References

- 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 / American Speech-Language-Hearing Association, 16*(2), 128-139.
- Asher, I. I., & Garcia, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal, 53*, 334-41.
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233-250.
- Audacity Team (2008): Audacity (Version 1.2.6) [Computer program]. Retrieved September 11, 2011, from <http://audacity.sourceforge.net/>.
- Badin, P., Youssef, A. B., Bailly, G., Elisei, F., & Hueber, T. (2010). *Visual articulatory feedback for phonetic correction in second language learning*. Unpublished paper presented at Second Language Studies: Acquisition, Learning, Education and Technology. Tokyo, Japan, pp. 1-10.
- Bacsfalvi, P. (2010). Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech-Language Pathology and Audiology, 34*(3), 206-217.
- Bacsfalvi, P., & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for several adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics and Phonetics, 25*(11-12), 1034-1043.

- Bernhardt, B., Bacsfalvi, P., Adler-Bock, M., Schimizu, R., Cheney, A., Giesbrecht, N., O'Connell, M., Sirianni, J., & Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics and Phonetics*, 22(2), 149-162.
- 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 and Phonetics*, 17(3), 199-216.
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6-7), 605-617.
- Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.
- Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.23, retrieved 1 August 2011 from <http://www.praat.org/>.
- Bradlow, A. R., Akahame-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61 (5), 977-985.
- Bradlow, A. R. (2008). Training non-native language sound patterns. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). Philadelphia: John Benjamins.

- 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. *The Modern Language Journal, 54*(7), 477-481.
- Dalston, R. M. (1975). Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America, 57*(2), 462-469.
- Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review, 59*(4), 547-566.
- Eckman, F. (2008). Typological markedness and second language phonology. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 95-115). Philadelphia: John Benjamins.
- Eckman, F., & Iverson, G. (1994). Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavas (Ed.), *First and Second Language Phonology* (pp. 251-266). San Diego: Singular Press.
- Elliott, R. E. (1995) Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal, 79*, 356-371.
- Espy-Wilson, C., & Boyce, S. (1994). Acoustic differences between “bunched” and “retroflex” variants of American English /r/. *Journal of the Acoustical Society of America, 95*(5), 2823.

- Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34(2), 153-175.
- Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, 91(1), 370-389.
- 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., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech*, 38(1), 25-55.
- Flege, J. E., Yeni-Komshian, G. F., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Flipsen, P., Jr., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeney, J. L. (2000). Acoustic data for American English /r/ and /ɜ~/ in typically speaking adolescents (Technical Report No. 10). Madison: University of Wisconsin—Madison, Phonology Project, Waisman Center.
- Gibbon, F., & Hardcastle, W. (1987). Articulatory description and treatment of "lateral /s/" using electropalatography: A case study. *International Journal of Language and Communication Disorders*, 22(3), 203-217.
- Gibbon, F., Hardcastle, W., & Suzuki, H. (1991). An electropalatographic study of the /r/, /l/ distinction for Japanese learners of English. *Computer Assisted Language Learning*, 4(3), 153-171.
- Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In Edwards, J.G. & Zampini, M.L.

- (Eds.), *Phonology and second language acquisition* (pp. 309-322). Philadelphia: John Benjamins Publishing Company.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, 9(3), 317-323.
- Hancin-Bhatt, B. (2008). Second language phonology in optimality theory. In Edwards, J.G. & Zampini, M.L. (Eds.), *Phonology and second language acquisition* (pp. 117-146). Philadelphia: John Benjamins Publishing Company.
- Hayes, B. (2009). *Introductory Phonology*. Massachusetts: Wiley Blackwell.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.
- Howard, S., & Varley, R. (1995). EPG in therapy using electropalatography to treat severe acquired apraxia of speech. *International Journal of Language and Communication Disorders*, 30(2), 246-255.
- Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, 15(2), 255-274.
- Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native-like English 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* (pp. 41-62). Philadelphia: John Benjamins.

- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, 118, 3267.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.
- Jokisch, O., Koloska, U., Hirschfeld, D., & Hoffmann, R. (2005). Pronunciation learning and foreign accent reduction by an audiovisual feedback system. *Affective Computing and Intelligent Interaction*, 3784, 419-425.
- Labrone, L. (2012). *The Phonology of Japanese*. New York: Oxford University Press.
- Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Massachusetts: Wadsworth.
- Ladefoged, P., & Johnson, K. (2010). *A Course in Phonetics* (6th ed.). Massachusetts: Wadsworth.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly*, 39(3), 369-377.
- 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/. *From Sound to Sense*, 50, C381-C386.
- Long, M. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition*, 12(3), 251-285.
- MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371-390). Amsterdam: North Holland.

- Manning, W. H. (2010). Evidence of clinically significant change: The therapeutic alliance and the possibilities of outcomes-informed care. *Seminars in Speech and Language, 1*, 207- 216.
- Marx, N. (2002). Never quite a 'native speaker': Accent and identity in the L2 and the L1. *Canadian Modern Language Review, 59*(2), 264-281.
- Massaro, D. W., & Light, J. (2003). Read my tongue movements: Bi-modal learning to perceive and produce non-native speech /r/ and /l/. Paper presented at the Eighth European Conference on Speech Communication and Technology. Geneva, Switzerland.
- Masuda, H., & Arai, T. (2010). Processing of consonant clusters by Japanese native speakers: Influence of English learning backgrounds. *Acoustical Science and Technology, 31*(5), 320-327.
- McNeil, M., Doyle, P., & Wambaugh, J. (2000). Apraxia of speech: A treatable disorder of motor planning and programming. In S.E. Nadeau, L.J. Gonzales Rothi, & B. Crosson (Eds.), *Aphasia and Language: Theory to Practice* (pp. 221-266). New York: Guilford Publications.
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Attention, Perception, & Psychophysics, 18*(5), 331-340.
- Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9*, 293-303.

- Moyer, A. (1999) Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition, 21*, 81-108.
- Nordberg, A., Carlsson, G., & Lohmander, A. (2011). Electropalatography in the description and treatment of speech disorders in five children with cerebral palsy. *Clinical Linguistics and Phonetics, 25*(10), 831-852.
- Odisho, E. Y. (2003). *Techniques of teaching pronunciation in ESL, bilingual and foreign language classes*. Munich: Lincom Europa.
- Ogitsu, K. (2009). *Acoustic differences between English /l/ and /r/ in Japanese speakers*. Unpublished manuscript, University of Aizu, Aizu-Wakamatsu, Japan.
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54*, 1235-1247.
- Ota, M., & Ueda, I. (2007). Japanese Speech Acquisition. In S. McLeod (Ed.), *The International Guide to Speech Acquisition* (pp. 457-471). New York: Thomson Delmar Learning.
- Öster, A. M. (1998). Spoken L2 teaching with contrastive visual and auditory feedback. Paper presented at the *Fifth International Conference on Spoken Language Processing*. Sydney, Australia.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*(3), 261-283.
- Pasquini, K. (2012). *Qualitative study: Measuring participant satisfaction concerning speech training with ultrasound for English /l/ and /r/*. Unpublished manuscript, University of British Columbia, Vancouver, Canada.

- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in L2: a review. *Journal of Phonetics*, 29, 191-215.
- Presson, N., & MacWhinney, B. (2010). The Competition Model and language disorders. In J. Guendouzi, F. Loncke & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes* (pp. 31-47). New York: Psychology Press.
- Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: a reexamination. *Language Learning*, 30(2), 271-287.
- Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and identifiability and accuracy. *Studies in Second Language Acquisition*, 20, 213-243.
- Saito, K., & Lyster, R. (in press). Effects of Form - Focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*.
- Scovel, T. (1988). *A time to speak: a psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House/Harper & Row.
- Schmidt, A., & Beamer, J. (1998). EPG treatment for training Thai speakers of English. *Clinical Linguistics & Phonetics*, 21, 389-403.
- Secord, W. (2007). *Eliciting sounds*. (2nd ed.). New York: Thomson Learning.
- 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(03), 243-261.
- Shibuya, Y., & Erickson, D. (2010). Consonant cluster production in Japanese learners of English. *Proceedings of Interspeech*. Makuhari, Japan.

- Smit, A. B. (1993a). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech and Hearing Research*, 36, 533-547.
- Tahta, S., Wood, M., & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, 24(3), 265-272.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41(2), 177-204.
- Tsujimura, N. (2007). *An introduction to Japanese Linguistics*. Massachusetts: Wiley-Blackwell.
- Underwood, G. N. (1988). Accent and Identity. In Thomas, A.R., (ed.) *Methods in Dialectology* (pp. 406-427). Clevedon: Multilingual Matters.
- 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.
- Zhang, Y., Kuhl, P., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, 46, 226-240.

Appendix A. Recruitment poster

THE UNIVERSITY OF BRITISH COLUMBIA



School of Audiology and Speech Sciences
Faculty of Medicine
Room 443, Friedman Building
2177 Wesbrook Mall
Vancouver, BC Canada V6T 1Z3

Tel: (604) 822-5591 Fax: (604) 822-6569

Do you want to improve your English speaking skills?

Is your first language Japanese?

Title of Research Project: The effectiveness of ultrasound as articulatory visual feedback device for teaching English speech sounds to speakers of English as a Second Language

Principal Investigator: Dr. B. May Bernhardt
Professor, UBC School of Audiology and Speech Sciences

Description:

The UBC School of Audiology and Speech Sciences is currently recruiting participants for a study examining the effectiveness of using ultrasound to teach perception skills and pronunciation of "l" and "r" in November, 2011. Eligible participants will take part in 6 sessions (one pre and one post-training assessment session of one hour each and 4 40-minute speech training sessions) led by a Masters student in speech-language pathology or a registered speech-language pathologist.

We are currently offering this opportunity for individuals whose first language is Japanese and are between the ages of 19-50 years old. If you have any questions about the study or would like to sign up, please contact **Haley** at: hmtsui@interchange.ubc.ca or Dr. Bernhardt at xxx-xxx-xxxx.

Appendix B. Consent form

THE UNIVERSITY OF BRITISH COLUMBIA



School of Audiology and Speech Sciences
Faculty of Medicine
Room 443, Friedman Building
2177 Wesbrook Mall
Vancouver, BC Canada V6T 1Z3

Tel: (604) 822-5591 Fax: (604) 822-6569

Consent Form

Study Title: The effectiveness of ultrasound as an articulatory visual feedback device for teaching English speech sounds to speakers of English as a Second Language

Principal Investigator Dr. Barbara May Bernhardt
Professor, UBC School of Audiology and Speech Sciences
Phone: (xxx) xxx-xxxx

Co-Investigators Dr. Penelope Bacsfalvi
Registered Speech-Language Pathologist

Dr. Stefka Marinova-Todd
Assistant Professor, UBC School of Audiology and Speech Sciences

Haley Tsui
M.Sc. candidate, Speech-Language Pathology

Purpose

Ultrasound of the tongue has shown promise in teaching of speech sounds both in English second language learners and traditional speech therapy. This study aims to determine the effectiveness of using tongue ultrasound in the instruction of perceiving and producing English segments "l" and "r" in native Japanese speakers.

This study is being conducted as part of a M.Sc. thesis at the School of Audiology and Speech Sciences at UBC for Haley Tsui.

Study Procedures

Participants will attend 6 individual sessions in total at the Friedman Building (Room 342) of the UBC School of Audiology and Speech Sciences (2 assessment sessions before and after training of one hour each) and four speech training sessions of 40 minutes each. Pronunciation of "l" and "r" will be practiced under the guidance of a registered speech-language pathologist or a supervised graduate student in speech-language pathology. Practice homework will be given. In the assessment sessions, participants will complete a short questionnaire on their language background, and name pictures as single words and in sentences, during which time audiorecordings of their speech will be made. In addition, there will be a short task involving listening to pairs of words containing /l/ and /r/ for identification of words and speech sounds. In the

pre-training assessment session, a hearing screening will be done, plus a set of ultrasound images will be captured. Participant data, including audio files, will not be saved with any form of identifying information to ensure anonymity of participants.

Potential Risks

There are no known risks above those encountered in everyday life.

Potential Benefits

Participants will receive individualized attention and feedback with regards to English pronunciation and perceptual skills. Participants will also be offered a certificate of completion in this research project.

Confidentiality

All records, including consent forms, self-evaluation questionnaires and audio recordings, will be identified only by numerical code in a locked filing cabinet in Room 436, Friedman Building, UBC. Audio files will be named using numerical codes only and will be kept on a password protected computer. Subjects will not be identified by name in any reports of the completed study.

Contact for information about the study

If you have any questions or desire further information with respect to this study, you may contact Dr. Bernhardt at (xxx) xxx-xxxx or email Haley Tsui, xxxxx@interchange.ubc.ca.

Contact for concerns about the rights of research subjects

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598 or if long distance e-mail to RSIL@ors.ubc.ca or toll free 1-877-822-8598.

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without consequence.

Consent

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without consequence.

The researchers may wish to use some of your anonymous data for teaching purposes in the speech therapy program, or at conference presentations. They may also wish to use the data in the future.

Your signature below indicates that you have received a copy of this consent form for your own records and indicates that you consent to participate in this study and to secondary uses of your anonymous data.

Signature

Date

Printed Name

Appendix C. Language experience questionnaire

Name: _____

Age: _____

Gender: Male / Female

1. How many years have you lived in an English-speaking country or environment?
2. At what age were you first exposed to the English language? Where was this?
3. At what age were you first immersed in an English speaking environment? Where was this?
4. Have you ever had instruction in the English language or in pronunciation before? If so, for how long and what did it involve?
5. How often do you speak English in your daily life?
100% of the time 75% 50% 25% or less
6. In what contexts do you speak English? At home, school, work?
7. What is your level is spoken English for the sound “L”?
Exactly on Target Almost on Target Somewhat Not at all
What makes it hard for you?

Is “L” easier to say in some English words over others? If so, which ones are easier for you? Which words do you think are harder?
8. What is your level is spoken English for the sound “R”?
Exactly on Target Almost on Target Somewhat Not at all
What makes it hard for you?

Is “R” easier to say in some English words over others? If so, which ones are easier for you? Which words do you think are harder?
9. Is there anything else about English pronunciation or the English language that you find particularly difficult?
10. Have you ever had speech therapy or been diagnosed with a speech, language disorder?
11. Do you have any hearing loss that you know of?

12. How motivated are you to participate and practice?

Extremely
Motivated

Very
Motivated

Slightly
Motivated

Not very
Motivated

13. Why you are interested in these sessions? What are your expectations?

Appendix D. Word list for assessment

/ɪ/ Word List

Word Initial Clusters

Prize
Pray
*Bread
Brother
*Tree
Tractor
Draw
Drink
Crown
Crib
Grew
Grow

Word Final Clusters

*Arm
Fork
*Heart

Word Initial

Row
*Rat
*Read

Word Medial

*Arrow
*Carry
Arrive

Word Final

*Beer
*Hair
Pour

/l/ Word List

Word Initial Clusters

*Please
Play
Blog
Blast
*Clown
Clip
Glue
Globe

Word Final Clusters

*Film
Hulk
*Salt

Word Initial

*Low
*Leaf
Law

Word Medial

*Pillow
*Yellow
Bowling

Word Final

Fall
*Bowl
*Peel

* = words used in analysis

Appendix E. Participant handout of language differences

Information to remember about the Japanese and English languages:

1. Syllable Structure

C = Consonant, V = Vowel, () = optional

Japanese:

Usually: (C)V

Rarely: (C)(y)V(V)(C)

More restricted what sound can be at the end of a syllable (ex. try saying “book” or “writer” in Japanese)

English:

Usually: (C)V(C)

Can be: (C)(C)(C)V(C)(C)(C)

Many sounds in English can occur at the end of a syllable, including many together in a cluster can have 35+ variations (sp, sk, ks, lp, lb, lt, lk, lf etc...).

3. Consonant Clusters

Japanese:

No clusters, except with “y”
Ex. Japanese word for “today”
(/kyoo/)

English:

Many clusters with 2-4 consonants:
Pl, bl, tr, dr, spl, mps, rsts etc.

3. “Gairaigo”

Japanese has some “loan words” from many languages, including English. These words resemble English words, but have Japanese language principles incorporated in their pronunciation.

For example: ミルク - **miruku** – **milk**

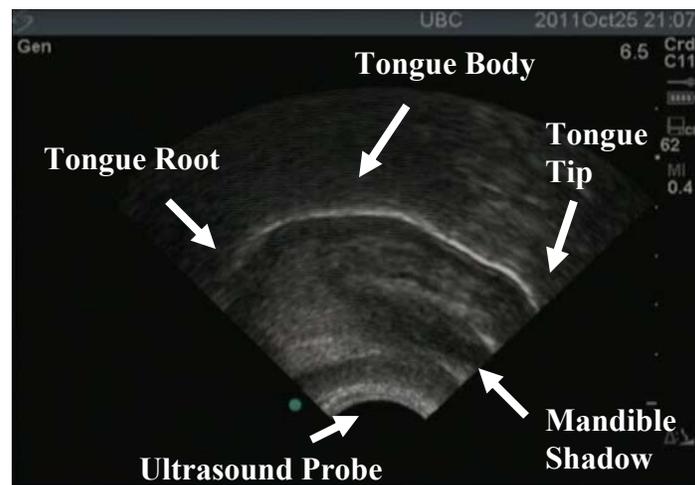
The Japanese word “miruku” resembles the English word for “milk”, but the “LK” cluster has been broken up into two syllables because the Japanese language does not include words with an “LK” cluster.

Appendix F. Participant handout describing ultrasound

Information about Ultrasound for Participants

Ultrasound is a safe and non-invasive method of observing the movements of the tongue within the mouth.

Images of the tongue are depicted from the use of the probe placed under the individual's chin that emits *high frequency sound waves* (outside of the range that humans can hear). There is no exposure to radiation during the use of ultrasound. These sound waves are emitted through the tongue and reflected back off the upper surface of the tongue. The ultrasound machine measures the time it takes for the sound waves to reflect back to the probe and then plots an image:



**Mid-Sagittal Ultrasound Image of the Tongue at rest
Image courtesy of Haley Tsui, UBC S.A.S.S. Student**

Ultrasound can be used to provide real-time images of tongue movement during speech and can also be used to demonstrate complex tongue positioning of speech sounds. Speakers of English as a second language may benefit from being able to see the complex tongue configurations used for North American “l” and “r” and practicing articulating speech sounds while viewing their tongue movements.

For videos on tongue shapes for “l” and “r”, please review:

UBC School of Audiology and Speech Sciences website:

<http://www.audiospeech.ubc.ca/research/child-phonology-phonetics-and-language-acquisition-lab/ultrasound-in-speech-training>

Ultrasound in Speech Training Presentation by Penelope Bacsfalvi, Ph.D., CCC-SLP(C), and Bosko Radanov M.Sc., SLP(C)

Appendix G. Word lists used in training

<p>Word-Initial /l/ Let Leg Live Late Last Long Light Leave Listen Little</p>	<p>Word-Medial /l/ Only Alive Hello Salad Alone Collect Family Believe Balloon</p>	<p>Word-Final /l/ All Tell Fill Call Fool Able Apple Table People Trouble</p>
<p>Word-Initial /l/ Clusters Black Blend Block Glow Please Cloud Flag Slide</p>	<p>Word-Final /l/ Clusters Bulb Bald Bold Cold Curled Field Collar Healer Similar Smaller Golf Self Shelf Milk Bulk Silk Help Scalp Else Impulse Aisles Animals Bills Deals Falls Nails Adult Belt Default</p>	<p>Melt Health Wealth Adults Dissolve Twelve</p>

Word-Initial /ɪ/ Red Run Rest Rain Real Wrong Write	Word-Medial /ɪ/ Very Marry Story Sorry Hurry Carrot Orange Around Tomorrow	Word-Final /ɪ/ Or Are Far Door Near More Sure Their Before Appear Poor	
Word-Initial /ɪ/ clusters Bring Cry Freeze Press Broke dry	Word-Final /ɪ/ Clusters Absorb Afford Bored Card Yard Hard Keyboard Landlord Poured Scared Scored Postcards Yards Scarf Bark Park Dark Mark Fork New York Question mark Shark Alarm Dorm Farm Form Harm Perform Storm Warm	Worm Barn Born Corn Horn Popcorn Torn Worn Harp Sharp Tarp Course Horse Resource Airs Dares Doors Ears Floors Guitars Hours Ours Stairs Theirs Forced Pierced Airport Apart Art Dart Part Smart	Report Fourth North Sport

Minimal Pairs Word List

[l, ɹ] Word-Initial Position	Syllable final [l, ɹ] in Medial Position	Others (not minimal pairs) syllable initial
Lay ray Laid raid Lake rake Lain rain Lace race Late rate Leap reap Lease Reese Leak reek Leaf reef Leach reach	Palace Paris Ceiling searing Kneeling nearing Peeling peering Palate Parrot Pilot Pirate Belated berated Elect erect Belly berry Elect erect	target tailgate carpet tell tale starfish fall down sellout
	[l, ɹ] in Final Position	
Life rife Lime rhyme Lied ride Lice rice Light right Lies rise	Ail air Pale pair Hail hair	
Lobe robe Load road	Meal mere Steal steer Kneel near Deal dear Teal tear	
Loom room Loot root Lewd rude	Dial dire Tile tire File fire	
Lash rash Lag rag Lamb ram Lap rap Lack rack	Tool tour Pool poor	
Led red	Owl our Towel tower	
Lip rip Lid rid List wrist Limb rim	Stale stare Fair fair Snail snare Wail wear Scale scare	
Lug rug Lush rush		

[l, ɹ] in Word Initial Clusters

	/bl/, /bɹ/	/kl/, /kɹ/	/fl/, /fɹ/	/g/ clusters	/p/ clusters
/eɪ/	blake – brake			glade – grade	play – pray
/i/	bleed – breed		flees – freeze	glean - green	
/oʊ/	bloke – broke	cloak – croak			
/u/	blue – brew	clue – crew	fruit – flute	glue – grew	
/ɛ/					pleasant – present
/æ/	bland – brand	clam – cram		glass – grass	plank – prank
/ɑ/	blonds – bronze		flock – frock		

/l/ Initial Clusters		/ɹ/ Initial Clusters
[b]	Led sled	[bɹ]
Lead bleed	Low slow	Raid braid
Lock block	Lot slot	Rat brat
Land bland	Link slink	
	[spl]	[kɹ]
[k]	Lash splash	Rack crack
Lamb clam		Rhyme crime
Lick click		Rose crows
Loud cloud		
		[dɹ]
[fl]		Rain drain
Lie fly		Red dread
Low flow		Raw draw
Lute flute		
Lap flap		[fɹ]
Led fled		Ride fried
		Reed freed
[gl]		
Loss gloss		[gɹ]
Laze glaze		Rid grid
Land gland		
Listen glisten		[pɹ]
		Rank prank
[pl]		Ride pride
Ledge pledge		
Lace place		[tɹ]
		Rate trait
[sl]		Rim trim
Lab slab		Rot trot

Appendix H. Word list for perception task

- | | | |
|------|---------|---------|
| 1. | lobe | robe |
| 2. | lip | rip |
| 3. | peeling | peering |
| *4. | pirate | pilot |
| *5. | arrive | alive |
| *6. | leaf | reef |
| *7. | grew | glue |
| *8. | file | fire |
| 9. | rush | lush |
| *10. | call | car |
| *11. | lock | rock |
| *12. | belly | berry |
| 13. | flute | fruit |
| 14. | lamb | ram |
| 15. | grew | glue |
| 16. | grass | glass |
| *17. | pray | play |
| 18. | lock | rock |
| 19. | pray | play |
| 20. | call | car |
| 21. | hair | hail |
| 22. | arrive | alive |
| 23. | flight | fright |
| 24. | collect | correct |
| 25. | towel | tower |
| 26. | leaf | reef |
| *27. | flight | fright |
| 28. | collect | correct |
| 29. | file | fire |
| 30. | belly | berry |

* = 10 pairs used in initial assessment

Appendix I. Participant satisfaction questionnaire

Please circle one: SA (strongly agree), A (agree), N (neutral), D (disagree), SD (strongly disagree)

The speech training was explained in a way I could understand. SA A N D SD
The environment was comfortable and pleasant. SA A N D SD
The speech training helped me make speech sounds accurately. SA A N D SD
I would seek out this type of service again if needed. SA A N D SD
I would recommend this service to others. SA A N D SD

Did you find the introduction to the ultrasound helpful in the initial session?
Yes _____ No _____.

If yes, what specifically did you find useful?

Did you find using the ultrasound helpful? For /l/? _____ For /r/? _____

If so, what specifically did you find useful from the ultrasound?

How difficult did you find the perception tasks?

Easy _____ Moderate _____ Difficult _____.

Did you find them helpful for hearing the difference between /l/ and /r/?
Yes _____ No _____.

Production of /l/ and /r/? Yes _____ No _____

If yes, how were they helpful?

Do you think you are better at hearing the difference between /l/ and /r/ after these sessions?

Yes ___ No ___

If yes, what helped in order of usefulness (1-4)?

_____ seeing tongue shapes
_____ practice similar words (e.g. lake/rake) using ultrasound
_____ doing the perception tasks
_____ other (please specify) _____

Did you follow the practice provided between sessions?

Yes _____ No _____

If yes, how was the practice helpful?

Did you think about the ultrasound images (i.e. what your tongue looked like) during your home practice? If yes, how was this helpful?

What was most helpful from the sessions?

What was least helpful from the sessions?

Has your speech production changed compared to before?

/l/: Exactly on target Almost on target Somewhat Not at all

/r/: Exactly on target Almost on target Somewhat Not at all

Any other comments?

Appendix J. Screen shot of listener judgment program

