TRAINING JAPANESE L1 SPEAKERS IN AMERICAN ENGLISH L2 PRODUCTION OF THE PHONEME /R/ USING VISUAL BIOFEEDBACK IN THE FORM OF A SPECTROGRAM: A PRELIMINARY STUDY

By

IOMI PATTEN

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

UNIVERSITY OF FLORIDA

2012

To every traveler who has ever yearned to say more: it is my wish that you have every opportunity to share your own words, in your own voice, in a way that all may understand you; for these shared words make the whole world a far richer place.

And to my grandmother, Mary Elbert, for leading the way.

ACKNOWLEDGMENTS

I am deeply grateful to the Aphasia and Bilingualism laboratory at the University of Florida for the resources, guidance, and encouragement provided; without these, this project would not have been possible. My committee has been integral to this entire process: many thanks to Dr. Judith Wingate for the spark, and to Dr. Lisa Edmonds for opening the door.

I am grateful to all of the faculty and staff in the University of Florida Speech Language and Hearing Sciences department who have supported and encouraged curiosity. Idella King, Cassie Mobley, and Dalila Johnson were ongoing sources of support and help. Dr. J. Rosenbek, Dr. Jamie Reilly, and Yukari Nakamura Sensei provided important insights at various stages of this project's development. Thanks to Dr. Mehmet Yavas (FIU) for originally introducing me to phonology and phonetics (and particularly second language phonology), and to Dr. Kenneth Logan for demonstrating exactly how phonological acquisition training can be enjoyable for both instructor and learner. Thank you also to Angela LaGambina and Daniel Furnas for crucial support, advice, assistance, and so much more. Neal Musson: thank you for putting together all of the dusty pieces- over and over again until it worked.

To my friends and family: thank you for your patience and support while I disappeared from your lives for two years.

I appreciate all of the English language learners in my classrooms over the years; you propelled me toward this work, and I thought of you daily. Most of all, I am grateful to Participant 1 and Participant 2- you were both such a pleasure to know. Ganbatte!

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

C	Consonant
JL1	First Language Speaker of Japanese who is also a learner of English
L1	First Language
L2	Second Language
NNS	Non-Native Speaker of English
NS	Native Speaker of English
r	The concept of the sound “R”
ɹ	The rhotic North American /r/, how it is specifically pronounced
ɾ	Alveolar tap or flap (similar to what is produced by American English speakers for “t” in “butter” or “d” in “ladder”)
ð	Voiced interdental fricative (“th” as in “then” and “bathing”)
V	Vowel

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Arts

TRAINING JAPANESE L1 SPEAKERS IN AMERICAN ENGLISH L2 PRODUCTION OF THE PHONEME /R/ USING VISUAL BIOFEEDBACK IN THE FORM OF A SPECTROGRAM: A PRELIMINARY STUDY

By

Iomi Patten

May 2012

Chair: Lisa Edmonds
Major: Communication Sciences and Disorders

Background: Japanese speakers who are learners of English have difficulty both producing and perceptually identifying the phoneme /ɹ/.This phoneme is produced in variable ways by native speakers of English, so there is little concrete articulator movement or placement instruction that would be of help to these learners. However, the identifying feature of /ɹ/ is found in the acoustic signal: the third formant (F3) “dips” unusually low for this phoneme (below 1950 Hz; but productions under 2300 are also perceived as correct). Japanese speakers do not auditorily attend to this formant because it does not hold identifying linguistic information in the Japanese language, rendering auditory perception alone an unreliable cue as well. This study uses spectrographic visual biofeedback to encourage these learners to visually identify the low F3 and to replicate it in their own productions.

A single-subject multiple baseline design was implemented with two participants. Treatment stimuli included words with /ɹ/ in initial position which were presented in hierarchical phases of increased syllable shape complexity. Within each phase, cues were reduced from auditory-visual to visual only to delayed feedback. The treatment

protocol required participants to replicate native speaker model sounds of the trained syllables while monitoring a computer screen to try to achieve an F3 below 2300 Hz. Generalization trends were monitored with weekly probes on 71 untrained words containing /r/ in a variety of positions. Effect sizes (*d*) were also calculated to evaluate magnitude of change between the baseline phase (N = 5 probes) and the final probe for each participant. Potential changes in the perception of distinguishing /ɹ/ and /l/ in minimal pairs was also evaluated.

Both participants showed improvement in trained syllables with evidence of generalization to a number of untrained syllable shapes. Effect sizes for trained and untrained items ranged from 2 to 4.6, respectively for P1 and from 1 to 6 for P2. Perceptual discrimination of /ɹ/ and /l/ scores rose 6% for P1 and 15% for P2.

These preliminary results suggest that spectrographic visual biofeedback for /ɹ/ production may be an effective treatment approach for training /ɹ/ in Japanese speaking learners of English.

CHAPTER 1
BACKGROUND AND INTRODUCTION

Overview of English Language Learning

It would be very difficult to overstate the importance of second language (L2) spoken English in the world today. Though it is hard to get an accurate count, there are an estimated 1-2 billion people currently learning English as a second (i.e., in an English-speaking environment) or international/foreign (i.e., in a non-English speaking environment) language worldwide. In fact, non-native speakers of English (NNSs) now outnumber native speakers of English (NSs) (Jenkins, 2002). English frequently provides access to information and technology in one's home country, and lack of English can be a barrier to career advancement (Graddol 2006; Carlson & McHenry 2006). Increasingly, English proficiency is a requirement for entrance into institutions of higher education (Graddol, 2006).

In English Language Learner (ELL) instruction, there is a general trend toward ignoring pronunciation instruction in favor of vocabulary, grammar, and literacy skills. In fact, pronunciation training has been all but abandoned in many classrooms (See Saito & Lyster, 2010 and Derwing & Munro, 2005 for more details). This is due to a variety of factors: most teachers are NNSs, frequently with the same L1 influence on pronunciation that the students have; both NS and NNS teachers may feel underprepared or unsure of what to teach; there is a dearth of empirical evidence on how to teach pronunciation in the educational materials and literature, and when such research does exist, it is underutilized (Derwing & Munro, 2005; Jenkins, 2002). A recent survey of adult English learners who had been studying for an extended period of time revealed that only 8 in 100 had been taught anything about pronunciation (Derwing & Munro, 2005).

There are a number of components that comprise what we perceive as intelligible speech, including rate, word and syllable stress, intonation, and the individual phonemes of the words. Thus it is possible that any of these elements could be trained in isolation or in combination to improve a speaker’s intelligibility. In this study, we chose to start with phonology because we know that certain segmental errors can cause breakdowns in intelligibility (Lambacher, 2010).

Frequently, second language phonology is taught by articulator placement or by imitation. Articulator placement training is effective for many consonants, and for some vowels. Imitation training relies on the learner’s ability to perceive the target sound and then hypothesize a set of articulatory gestures to produce that sound. Both methods are benefitted by a learner’s ability to perceive whether he or she has matched the target through audition or proprioception. However, when articulator placement target is variable, includes more than one or two points of constriction, or includes articulators for which there is poor tactile and kinesthetic feedback, instruction becomes appreciably more challenging. Such is the case for /ɹ/, which is a challenge for many English learners, and most particularly for Japanese speakers who are learners of English (referred to hereafter as JL1s), for whom perception of this phoneme is also unreliable. This is the concept we will explore.

Japanese /r/

Japanese does not contain a phoneme that accurately resembles /r/. Japanese /r/, though often identified as “R”, is in fact produced as a voiced apico-alveolar tap or flap [ɾ] (Aoyama, Flege, Guion, Ahakahane-Yamada, & Yamada, 2004; Lambacher, 2010).

This phoneme is produced with the tongue tip in light contact with the alveolar ridge and has been called a “loose alveolar stop” (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975). This production is clearly quite dissimilar to NS /ɹ/ (and even /l/), and may actually be more similar to NS productions of flapped /t/ and /d/ (the way many NSs say the medial sounds in “butter” and “ladder”) (Aoyama et al., 2004). Style and pace of articulation may play a role. According to Flege, Takagi, and Mann (1995), Japanese /r/ particularly sounds like flapped /d/ when it is produced rapidly, and more like /l/ when it is produced emphatically, since there is lateral airflow. Native Japanese speaking children learning /r/ sometimes substitute /d/ (Vance 1987, as discussed by Flege et al., 1995). The flap also includes a good deal of allophonic variation (Lotto, Sato, & Diehl, 2004).

Japanese /r/ includes allophonic variations that overlap acoustically with American /ɹ/ and /l/ at its edges. Lotto, Sato, and Diehl (2004) demonstrated this concept clearly by mapping /ɹ/, /l/, and Japanese [ɾ] across F2 x F3 space (in a similar method to how vowels are typically mapped in F1/F2 space) (Figure 1-1).

They found that:

> [T]he flap distribution partly overlaps the optimal boundary between /l/ and /r/. Exemplars of both English liquids fall within the flap distribution in F2 x F3 space. That is, Japanese speakers would categorize some exemplars from these two distributions as members of a single category… (Lotto et al.,2004).

This overlap in distribution thus includes part of both phoneme categories (though slightly more in /l/ space than in /ɹ/ space). Importantly, while Japanese /r/ space overlaps both [ɹ] and [l], it also includes the SAE boundary between the two phonemes, as well as further space that does not include either phoneme or the boundary between them. The consequences of this overlap are clearly seen in JL1 productions of /ɹ/ and

/l/ in English- there is a poorly defined boundary between the two sounds (Lotto et al., 2004).

Thus, it is clear why NSs of English typically perceive the JL1 English /ɹ/ tokens, (which are frequently produced as [ɾ]) as [l], less frequently as [ɹ], sometimes as another English phoneme such as [d], or possibly as an ambiguous or uncategorizable sound in English (Flege et al., 1995). Past research has shown wide variation and many outliers in this perceptual identification, including clusters such as /gr/, /dl/ and /wl/, or even as poor exemplars of a retroflex /d/, or a trilled "Spanish" /r/ (Flege et al., 1995 discussing work of a variety of researchers). Aoyama and colleagues' 2004 study included a "confusion matrix" that found NS listeners perceiving mostly [ɹ] and [l] for JL1 /r/ productions, but occasionally /w/, /d/, /br/, and /bl/.

American English Native Speaker /ɹ/

The liquid /ɹ/ is notoriously variable. Kent and Read (2002) discuss how liquids, as a group, share properties not only with consonants, but also with glides. Thus, liquids are not the same as traditional consonants such as stops, but rather share some vowel-like qualities. This may be particularly true for /ɹ/, which has no distinct point of articulatory contact the way most consonants do. Also, /ɹ/ can be mapped in F2xF3 space the way vowels can be mapped in F1xF2 space- which is not the case for most consonants (this is discussed more later in this paper, but see Lotto and colleagues for more details). Allophonic variations on /ɹ/ are well known, though for the purposes of this study we choose to view /ɹ/ in all positions as a single entity. There is reason to believe that this unified approach to /ɹ/ is useful for treatment purposes (See Elbert & McReynolds, 1975). Between the two phonemes in the liquid category, /l/ and /ɹ/ differ from each other mainly in Third Formant (F3) properties: /ɹ/ exhibits low F3 while /l/ exhibits high F3 (Kent & Read, 2002).

The American /ɹ/ sound has historically been categorized into two groups in which there is narrowing, but not closure, in the palatal region. Though there is some variation in descriptions (and names) of these two positions, they can be broadly defined as: the “retroflex” /ɹ/ (raised tip/blade and lowered dorsum) and the “bunched” /ɹ/ (lowered blade/tip, raised dorsum) (sometimes called “front R” and “back R,” respectively).

The reasons one person produces retroflex position while another produces bunched are still unclear. Individual differences in vocal tract anatomy and size do not appear to be the deciding factor in whether an individual uses bunched or retroflex /ɹ/; Zhou, Espy-Wilson, Boyce, Tiede, Holland, and Choe (2008) describe a case of two male subjects who produce remarkably different bunched and retroflex shapes despite sharing very similar vocal tract shape, size, and overall anatomy.

Importantly, it has been demonstrated that listeners do not perceptually identify any difference between retroflex and bunched /ɹ/. This is due to the fact that the acoustic signal for F1, F2, and F3 is relatively stable for these two types of /ɹ/ (Guenther, Espy-Wilson, Boyce, Matthies, Zandipour, & Perkell, 1999; Zhou et al., 2008). The main acoustic/spectrographic difference between them is seen only in higher frequencies (F4, F5), which listeners do not attend to in identifying /ɹ/, thus for our purposes they are the same [ɹ] created in two strikingly different ways (Zhou et al., 2008).

From a Dichotomy to a Continuum

These two differing approaches to /ɹ/ production are well known to produce the same acoustic signal (low F3, or an F3 “dip,” Figure 1-2). However, a body of recent literature has revealed that these two points are not dichotomous but rather are merely general endpoints of what is actually a continuum of articulator movement in the production of American /ɹ/ (Guenther et al. 1999; Zhou et al. 2008; Westbury, Hashi, & Lindstrom 1998; Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000). As noted above, this continuum follows the same acoustic pattern throughout- the characteristic low F3 dip.

Such extreme articulator variability may at first seem counterintuitive- how can so many different lingual movements create the same sound? Articulation of /ɹ/ is not limited, however, to lingual movement and resultant palatal constriction. The primary articulatory gestures associated with /ɹ/ include three simultaneous supraglottal constrictions: labial (lip rounding), lingual/palatal (raising the tongue), and pharyngeal (narrowing of the pharynx) (Bradlow, 2008). Locations, shapes, and degrees of these constrictions vary widely and interact with one another dynamically (Zhou et al. 2008*;* Espy-Wilson et al., 2000). Each of these movements individually contributes to the lowering of F3, but all three are used by native speakers interactively to create /ɹ/. For example, a speaker's change in tongue height may necessitate an adjustment in pharyngeal constriction and/or lip rounding in order to maintain sufficiently low F3 height to produce /ɹ/ (See Guenther et al., 1999 for more details).

Further complicating the matter is the fact that while these three main constrictions are relatively well documented in the literature, additional vocal tract areas have been hypothesized to contribute to the lowering of F3, such as sublingual space and the pyriform sinuses (Espy-Wilson et al, 2000).

Interspeaker and Intraspeaker Variability

Westbury and colleagues (1998), found, in regards to lingual placement for [ɹ]:

> [T]here are as many kinds of /r/ as there are speaker-by-word combinations. No two speakers' tongue shapes were ever exactly the same, though many were closely alike.

Guenther and colleagues (1999), using an electromagnetic midsagittal articulometer (imaging of transducers adhered to the tongue) to measure articulator placement in 7 NS subjects, demonstrated that articulator position varied widely across subjects. All 7 participants exhibited strong trading relations between tongue back height and tongue front horizontal position- changes in one part of the tongue position (e.g. tongue back height) caused changes in another part of the tongue position (e.g. tongue front horizontal position). Importantly, wide variation was also seen within certain subjects, in different utterances. For example, some speakers used exclusively a more bunched /ɹ/ or a more retroflex /ɹ/ in all instances, while others used two or three different configurations depending on word phonetic context or prosodic variation (Espy-Wilson et al., 2000; Guenther et al., 1999). So, not only is there not one "way" to make an /ɹ/, there is also not one "way" to make (or not make) variation on an /ɹ/.

Whether using a single /ɹ/ position or multiple /ɹ/ positions, each speaker was shown to be highly consistent to his or her particular pattern of /ɹ/ production (Guenther et al., 1999). All of these variations continued to show the same acoustic cue: lowered F3. Thus, the F3 "dip" is very consistent, and the articulatory trading relations make it so.

JL1s and /ɹ/

Despite the clear articulatory differences, JL1s often perceptually identify American /ɹ/ tokens as poor exemplars of Japanese /r/ (Aoyama et al., 2004). If prompted to identify the sound in English, /ɹ/ is identified primarily as either /ɹ/ or /l/ (the same is true for /l/) (Aoyama et al., 2004). Additionally, it has been demonstrated JL1s could not discriminate between recordings of their own productions of /ɹ/ and /l/ that had been judged as “correct” by native speakers (Goto, 1971). How can this be so? What are they listening to for identifying information? It has been shown that in the speech stream, Japanese speakers are attending most to F2, which holds high salience in Japanese to distinguish Japanese /r/, the language’s sole liquid, from the glide /w/. (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Ketterman, & Siebert, 2003; Yamada & Tohkura,1992; Lotto et al., 2004).

Studies (e.g. Iverson et al., 2003) have clearly shown that JL1s simply do not attend to F3 because it does not hold salience in Japanese for distinguishing one phoneme from the next (in just the same way that native speakers do not attend to F4 and F5 and thus fail to auditorily discriminate between retroflex and bunched /ɹ/). Miyawaki et al. (1975) demonstrated that Japanese speakers could identify and distinguish differences in F3 equally as well as native speakers do when F3 was synthetically removed from the speech stream. Thus, this is not an issue of whether or not JL1s can “hear” F3, but really that they are not attending to it because it does not hold any phonological information in Japanese.

Japanese speakers are often taught articulator placement for retroflex /ɹ /, if they are taught anything at all. This position is gesturally more similar to /l/ than to bunched /ɹ/ and even the most cursory understanding of articulator placement shows that this high front position is of little help in disambiguating /ɾ/ or /l/ from /ɹ/. Furthermore, as stated above, when /ɾ/ is produced rapidly or with emphasis by JL1s, it is more similar to English /l/ or /d/. Thus if a JL1 is creating muscle tension by excessively attending to tongue front shape and position, it would logically follow that further ambiguities and production of these erroneous substitutions could arise.

High variability in production of standard native speaker /ɹ/ (described above) creates obvious difficulty with the concept of articulator placement training. If each speaker produces /ɹ/ differently from one another, and some even produce it very differently for each context, how might we go about training a nonnative speaker in the “correct” production of /ɹ/? If native speakers with the same gender and similar vocal tract size and anatomy may naturally produce /ɹ/ with markedly different tongue positions (Guenther et al. 1999; Westbury et al., 1998), how may we lead a learner to one “correct” /ɹ/ target?

The very nature of the articulators and movements may provide further challenge. First, while lip movement may be very easy to visualize and describe, the constriction required for /ɹ/ intrinsically serves to conceal the movement of the tongue within (Lambacher, 2010). Next, the types of tongue positions used by native speakers vary widely, and even if one of the two endpoint positions is selected (e.g. retroflex or bunched), it does not then become an easy task: neither position includes true points of contact for tactile feedback, and placement cueing may even be detrimental, creating exaggerated movements (Schuster, Ruscello, & Toth, 1995). Then, pharyngeal constriction is all but impossible to describe, let alone to accurately monitor through proprioception and/or to intentionally control (and if the pyriform sinuses and sublingual space are key to producing /ɹ/, additional levels of difficulty are introduced). Combining these three primary movements at the same time and in just the right ways to create /ɹ/ is yet a further challenge.

So, what to do? One answer would appear to lie within the feature of /ɹ/ that is consistent across all native-speaker articulatory approaches: the lowered F3.

Summary

In essence, /ɹ/ is an acoustic lowering of F3. It is a highly variable and intricate sound produced by native speakers in a wide range of manners, both between and within individuals, creating a highly complex articulatory movement target for ELLs. The articulators and movements involved are not simple, easily definable, nor even completely understood. This is further complicated by the fact that each NS articulator movement interacts with other movements in a compound interdependent dance in order to maintain the most important acoustically identifying feature of /ɹ/: the low F3. Thus, articulator placement cueing is a poor option at best.

For JL1s, both perception and production of /ɹ/ are complicated by a natural lack of attendance to F3 in the speech stream. JL1s do not have a reliable metric by which to discern what is and is not an /r/ in either perception or production. This, along with the fact that the most common substitution is the JL1 phoneme /ɾ/ creates a situation in which JL1s are essentially grasping at sounds in an attempt to meet the target, and as a result, their comprehensibility is compromised. Self-perception alone is also, therefore, not an appropriate or reliable cue.

Defining the Task

Trying to make an American /ɹ/ without a dependable precept is something akin to trying to assemble a mysterious multi-dimensional puzzle with your hands concealed behind a curtain. How is one to know what the puzzle must look like? How does one know if one is moving closer to, or farther from, the goal? What is clearly needed here is a well-defined target, and reliable feedback on progress.

If the idea is that we need to create a reliable /ɹ/ from this tangle then we will need something concrete to reference. That thing is clearly F3. Saito and Lyster (2010) found that when JL1s do produce an /ɹ/ that is perceived as “correct” by NSs, the single most important factor is low F3 height.   But how do we ”train” for a low F3 in production?

Because a spectrogram displays a clear visual depiction of formant shapes and heights, it allows learners to visualize and understand their own productions of formants without relying on perception (which we know is problematic for JL1s) or articulator position (for which there is no clear, single target). Furthermore, the spectrogram provides an opportunity for learners to compare their own productions to those of native speakers in a single glance.

Spectrographic measurement of formants has been used in many studies as a measure of the relative correctness of /ɹ/ in JL1 production (See Saito and Lyster for an excellent example of this). Over time, efficacy and ease of use of this technology have combined with lowered cost to create a climate in which it is now a relatively reasonable tool for clinical use.

F3 Visual Feedback and /ɹ/: Previous Training Research

F3 training has been used previously in speech therapy for adult and pediatric populations. Schuster, Ruscello, and Smith (1992) describe a case study of a young man who misarticulated /ɹ/ in all contexts (except clusters) despite 12 years of traditional speech therapy treatment. With a combination of traditional articulator placement therapy and spectrographic biofeedback, the participant was able to accurately produce spectrographically correct vocalic and consonantal /ɹ/ which was also independently rated by two experienced judges as correct.

In a follow-up study, Schuster and colleagues (1995) provided similar treatment to two children (one male, one female) who misarticulated /ɹ/ despite two and four years of therapy respectively. Both participants produced /ɹ/ in all positions with 0% accuracy at baseline. The male, “Jerry”, though initially not as responsive to the biofeedback, was able to produce consonantal /ɹ/ and all vocalic /ɹ/s with >60% accuracy within ten sessions, when he was transferred back to traditional therapy methods and continued to improve until /ɹ/ was produced correctly in conversation. The female, “Tina,” was able to produce /ɹ/ with >80% accuracy within 8 sessions. Traditional therapy was then continued without biofeedback and she was able to attain correct /ɹ/ at the sentence level by the end of the study. More recently, McAllister and Hitchcock (In Press) produced a similar study with eleven children. Of the eleven, eight demonstrated measurably improved /ɹ/ production by post-test, as identified by independent judges and F3 formant height measurements. Of these eight, four demonstrated generalization of /ɹ/ to untreated items.

Spectrograms have also been used as visual biofeedback with language learners, primarily for vowel instruction. Brett (2004) described the in-progress development of a program to train Italian L1s on Standard British English vowels through formant biofeedback, using the author’s own development of a Macromedia Flash program that interfaces with PRAAT©. Motohashi-Saigo and Hardison (2009) used visual biofeedback in the form of waveform displays to train American students of the Japanese Language in perception of geminates (long sounds which are contrastive in Japanese but not in English). The results showed auditory-visual feedback training significantly improved perception for words (from 65% correct to 88% correct) and sentences (from 54% to 87% correct). This was found to be more effective than auditory-only feedback. Furthermore, this auditory-visual perception training showed some generalization to production.

Additionally, Lambacher (2010) used spectrogram training instructionally in the classroom for JL1s in Japan, for production of all phonemes in all positions (including /ɹ/). He mentions that low F3 is a salient visual cue to learners for the /ɹ/ phoneme, and for differentiating /ɹ/ from /l/. Though he provided excellent, detailed training instructions and reports that students using spectrogram training to improve productions (by better matching of spectrographic patterns to a teacher model), he did not provide a systematic analysis of the effect of the treatment on trained and/or untrained stimuli.

Thus, there is preliminary evidence that spectrogram training holds promise for improving individual phonemes in monolingual children with developmental articulation errors as well as adult language learners. However, to this author's knowledge, there have been no experimentally controlled studies evaluating the effect of /ɹ/ training with F3 feedback for JL1s, which is the aim of the current study.

Research Questions

Research Question 1: Does spectrographic visual biofeedback of F3 height improve production of /ɹ/ in trained words in initial position?

Research Question 2: Does spectrographic visual biofeedback of F3 height generalize to production of /ɹ/ in untrained words in initial, medial, final, and cluster positions?

Improvement in /ɹ/ will be measured by binary goodness ratings by trained native speaker judges in the University of Florida Aphasia and Bilingualism Lab. Our hypothesis is that production of /ɹ/ in the trained words will improve into an identifiable

/ɹ/. There is insufficient evidence in the literature to assert a strong hypothesis for generalization to untrained words and positions in JL1s. However, if the extensive training improves proprioception of an accurate /ɹ/ in the trained words and positions, it is conceivable that improvement to untrained words in the same positions would improve. Whether generalization will occur to untrained positions is less certain.

Research Question 3: Does spectrographic visual biofeedback of F3 height improve the participants' ability to perceptually discriminate /ɹ/ from /l/?

Previous literature is somewhat unclear on this topic, however, there is evidence that perception can improve in adult JL1 speakers. For example, a length of residence effect has been demonstrated such that JL1s who have resided in the US for an extended period (mean=21 years) have exhibited improved perception of the distinction between /ɹ/ and /l/ (Flege et al., 1995). Furthermore, Iverson, Hazan, and Bannister (2005) trained perception by synthetically enhancing F3 of both /ɹ/ and /l/ in the speech stream, and then used these stimuli to train JL1s. Discrimination of /ɹ/ and /l/ in initial position (which was trained) increased by 18%, but with little generalization to other positions. Whether F3 visual feedback (coupled with perception of their own production) improves their ability to perceive /ɹ/ is unknown, but there is some evidence that it is possible.

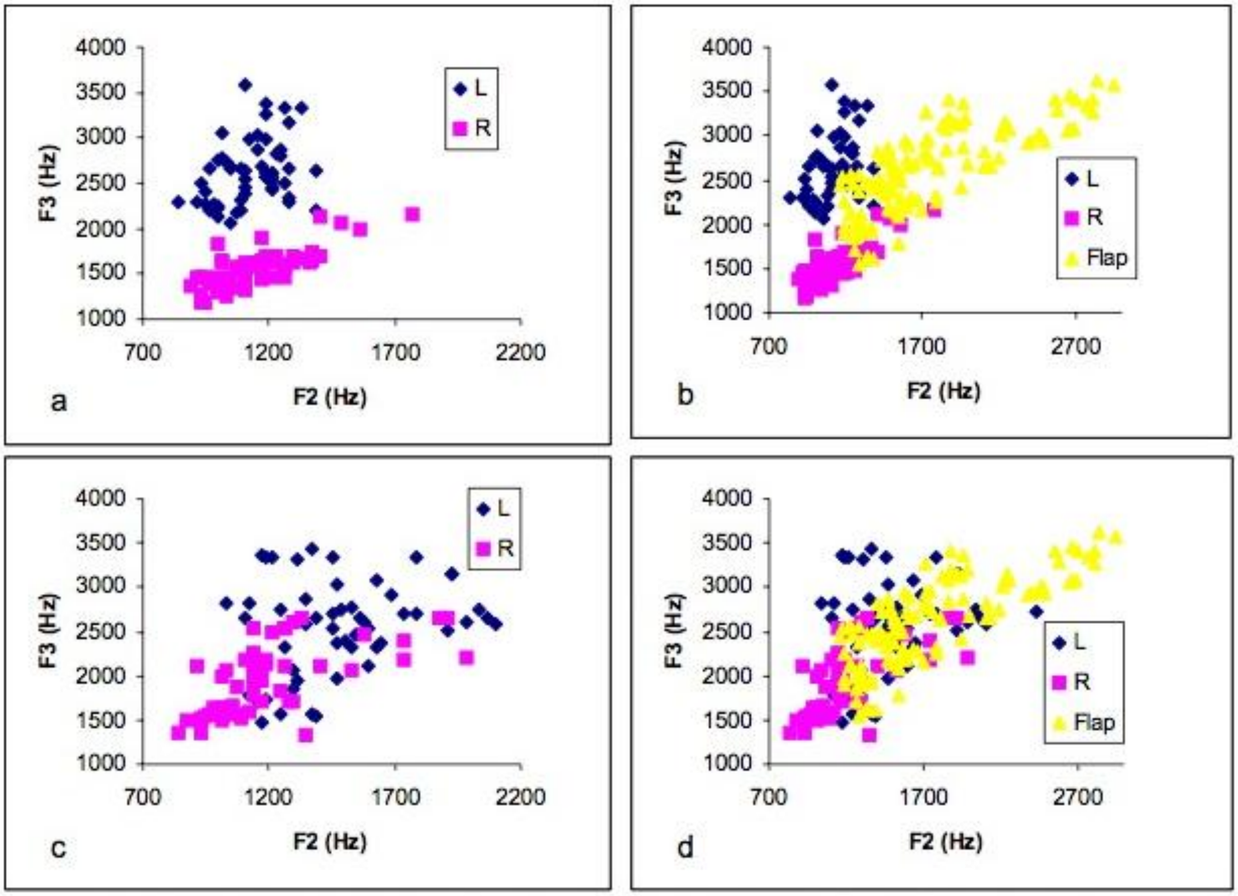


Figure 1-1. Japanese and American /r/ interaction. "Scatter plots of F2 and F3 onset frequencies obtained from productions of English /l/ and /r/ and the Japanese flap. a) Native English productions of /l/ and /r/; b) English productions of /l/ and /r/ with distribution of native Japanese productions of the flap; c) Native Japanese productions of /l/ and /r/; d) Native productions of flap with L2 productions of /l/ and /r/." Image and accompanying text from Lotto, Sato, and Diehl (2004). Reprinted with permission.

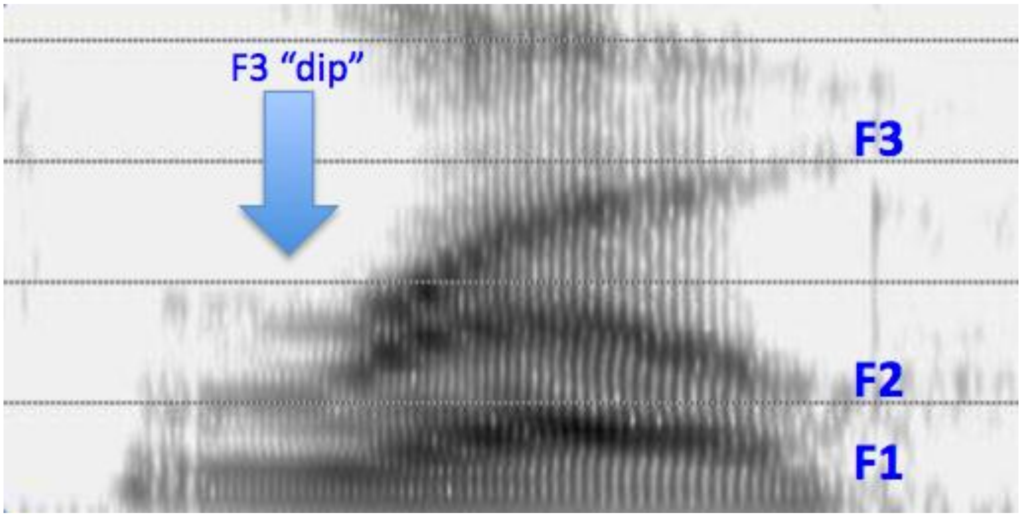


Figure 1-2. Spectrogram of a female NS saying "row." Time is visible on the Y axis, and formant height (Hz) in the X axis. The three formants are clearly visible, as well as the low F3 "dip" corresponding with the /r/ at the beginning of the word opening up at transition to a visibly higher F3 in the vowel. Note: this speaker's dialect does not include a final /w/ sound in this word, creating a CV syllable shape.

CHAPTER 2
METHODS

Participants

JL1 Participants

Two participants were enrolled in the training program. Both participants are native speakers of Japanese that were living near the University of Florida. Both self-reported normal vision, hearing, and learning capabilities. Both were initially exposed to English in middle school in Japan but reported no significant increases in fluency until arrival in the US. Both participants were enrolled in English as a Second Language classes in the local community concurrent to treatment. Both participants travel to Japan annually or biannually for approximately one month, during which they exclusively speak Japanese. Please see information below and Appendix A for more detailed language information about each participant.

Participant 1: This participant, age 33 years, started learning English in middle school with exclusively JL1 instructors, but she was not consistently exposed to English until her arrival in the US four years prior to enrolling in this study. After 6 months in the US, she returned to Japan for approximately 1 year, after which she returned to the US and has resided here for the last 2.5 years. Since her arrival she has been inconsistently enrolled in local conversation classes sometimes taught by NSs. She has also worked with a language partner (a NS with whom she converses 50% in Japanese and 50% in English for to provide language practice for both individuals). Her current education involves <3 hours weekly of English Language conversational instruction in a group setting with a volunteer instructor. She reports that none of her English courses have included a significant pronunciation component. She currently speaks Japanese 100% of the time at home, and 50% Japanese 50% English in social situations.

Participant 2: Participant 2, age 43, also started learning English in middle school with exclusively JL1 instructors and was not consistently exposed to English until her arrival in the US 6.5 years ago. Since her arrival she has also been inconsistently enrolled in English Language courses at adult schools in group settings, and reports that there has been no specific focus on pronunciation in any of her courses. Her current educational situation includes concurrent enrollment in twice weekly general English Language group classes. She reports that she exclusively speaks Japanese at home, though it should be noted that some members of her household are bilingual with native-like fluency in American English, and these individuals frequently speak to her in English, though she exclusively responds in Japanese. Outside of the home, she typically speaks English.

Four additional participants were consented into the study, but they were not enrolled into the training program for a variety of reasons. Three participants' performance on the baselines was too high to allow for improvement and one was fluent in a third language.

Native Speaker Models

Two participants, one female (NS1- Production Model) and one male (NS2-Perception Model) were recruited by IRB approved flyers from the University of Florida campus. NS1 was an undergraduate student, and NS2 was a graduate student. Both participants are monolingual native speakers of American English standard dialects, and neither has worked, lived, or studied abroad for more than a few weeks. Neither has received more than two years of college-level instruction in a foreign language, nor spent extensive time with English-language learners.

Study Design

A multiple baseline single subject design was used (e.g., Kearns, 1986). Five baseline sessions were conducted. Thereafter, the effect of treatment to the production of various trained and untrained syllable shapes (described in Treatment Stimuli) was evaluated throughout the treatment phases (described in Treatment Procedure).

Stimuli Development

Treatment stimuli

A hierarchical treatment list of 15 words was developed. In an attempt to maximize utility of treatment and realism of items, real words or highly word-like nonwords were selected.

The words developed for Phase 1 of treatment consisted of 5 CV syllables with initial /ɹ/, utilizing vowels that are common to (or similar in) both Japanese and English. Phase 1 vowels included a variety of positions, from the low back lax vowel /a/ (e.g. “raw”) to the high front, tense vowel /i/ (e.g. “ree”). One diphthong was also included, “ray”. (Appendix B for full list of stimuli). These stimuli follow the recommendations of Lambacher (2010) with the exception of /ɹe/ which was changed in orthography to the diphthong “ray” for ease of orthographic transcription and real-word status. However, while participants were encouraged to match each model vowel, they were never penalized for inaccurate pronunciation of vowels (or incomplete diphthongs)- the focus was primarily on /ɹ/.

Phase 2 words consist of five r-initial CVC words and nonwords utilizing the same vowels used in Phase 1, with the addition of a final /m/ (e.g., ”rom”, “ream”). A final /m/

was chosen for this phase in order to add complexity with a phoneme that is both generally unmarked and has no lingual component, as well as being both common to and early learned in both languages (Preston & Seki, 2011).

Phase 3 words follow a similar pattern to Phase 2, with a /b/ replacing the /m/ in final position (e.g. “rob”, “reeb”), introducing a true stop. The phoneme /b/ was selected due to its similarity to /m/ and its commonality to both languages (Preston & Seki, 2011).

All training stimuli were recorded by the female native speaker participant. She was instructed to read each word individually and clearly in as natural a way as possible.

Probe stimuli

Seventy-one untrained real-word probes including /ɹ/ in a variety of positions (initial, medial, and final along with voiced and voiceless consonant clusters of varying complexity) were developed to measure any generalization of treatment to production of /ɹ/ in similar and dissimilar untreated words and positions. See Appendix C for complete list.

Control stimuli

Fifteen control stimuli included the voiced interdental fricative /ð/ in similar positions to those included in treatment stimuli. This phoneme was selected based on its relative difficulty and its absence in the Japanese language (Thompson, 2001). Complete matching of controls to treatment items was not possible due to the limited occurrence of /ð/ in common English words, but /ð/ was included in word-initial, word medial, and word final positions (Appendix D for complete list).

Perception stimuli: Listening task for pre- and post-treatment testing

Sixty simple r-l minimal pairs were developed in four positions (initial, medial, final, and cluster), for a total of 120 additional words. None of these words was included in either the treatment or probe stimuli. One half (/ɹ/ or /l/) of each minimal pair was randomly selected for each word such that half of the words included /ɹ/ and half included /l/. Each selected word (minimal-pair half) was recorded by the male native speaker (perception model) for a total of 60 items.

Suggestions list

A series of general suggestions was developed in order to aid participants in attempting to match the spectrogram. These included broad general suggestions such as “Go slow”, and “Experiment”. Very broad articulator suggestions were provided as well, but in such a way that no specific recommendation or bias was introduced, (e.g. “Change the front of your tongue- big or small, up or down, or just differently.”). See Appendix G for the complete list.

F3 Height Criterion Selection

For most speakers, the characteristic F3 “dip” is below 2000 Hz. Many references describe F3 to be under 1950 Hz for most NSs, regardless of gender, age, or vocal tract size (Westbury et al. 1998; Saito & Lyster, In Press). It is important to note, however, that Hagiwara (1995) describes three NS female participants produced /ɹ/ with F3s in the 2200 range. Though these measurements may first appear to be outliers, there is little research including normative data on F3 production of /ɹ/ in females as a group (typically the measures include males). These F3 measurements, though outside the typical range reported by others, still demonstrated the characteristic “dip” /ɹ/ and vowel; a pattern of relative difference between the F3 of /r/ and the F3 of the adjacent vowel was maintained. This finding is also supported by Saito and Lyster's (in press) results, in which /ɹ/ with F3 in the 2200-2300 range was perceptually judged as "correct" /ɹ/ by NS judges. With these results in mind, we selected 2300 as a reasonable limit of "correctness" for F3 height in this study.

Procedure

Pre-Test

Baseline

Participants completed a total of 5 baseline sessions each over a two-week period. During each session, all 101 stimuli were presented on index cards in randomized order (15 treatment stimuli + 71 probe stimuli+15 control stimuli =101). All subjects were tested individually by a single examiner in a quiet room in the Aphasia and Bilingualism Lab at UF.

Recordings were made using the PRAAT© program, downloaded from http://www.fon.hum.uva.nl/praat/. Investigator 1 opened the program, selected "Record Mono Sound" and clicked the radial for "11025 Hz". This frequency was selected based on agreement with the Visi-Pitch™ program settings.

Participants were seated at a normal desk a comfortable distance from the microphone (Shur model SM-78) and asked to read each index card. No instructions were given as to pronunciation of the elicited words and sounds. Participants were given breaks periodically throughout recording and were informed that they could request a break at any time. Recording the 101 stimuli took approximately 10-25 minutes. Each participant was asked (and agreed) not to practice or attempt to learn any of the words in any way throughout the duration of the study.

During baseline collection, no feedback related to pronunciation accuracy of any kind was provided. In the early sessions, positive feedback of task performance was periodically provided (i.e. “good job”, “that’s great”). When participants indicated during early sessions that they did not know how to say a word, they were instructed in a friendly manner that they should try their best and that this would meet the requirements of the task to satisfaction.

Perception task- minimal pairs

The perception task was performed after the baselines were complete during session for Baselines 3 and 4. Before beginning the listening task, participants were provided with a short 4-item familiarization task that matched the target task, but with a different (female) speaker.

The 60 minimally paired words developed for this study were presented in two equal halves during consecutive baseline recording sessions. Participants were presented with a recording of one-half of a minimal pair (e.g. either “rock” or “lock”) and asked to write what they heard. In order to reduce non-target cognitive load and any unrelated variation in vowel perception or spelling errors, stimuli were presented with an accompanying answer sheet upon which were typed cloze-type (gap fill) written versions of each word (e.g. “___ock”). Some of the spellings were altered into nonword spellings of the real-word targets to equalize the likelihood of either an “R” or “L” answer, and to minimize any effect of word spelling on letter choice (e.g. the minimal pair ‘tool/tour’ was written as ‘too__e’ on the answer sheet). No instructions were given as to how many sounds or which sounds might be written into the spaces provided, and the two target phonemes were never mentioned explicitly (i.e. participants were told simply, “listen to the recording and write what you hear.”). Thus, this was not a forced-choice task between /ɹ/ and /l/ but rather a true phoneme identification task. (Appendices E and F for full list of minimal pairs and corresponding answer sheet).

Treatment Procedure

Task familiarization

Treatment was introduced in treatment session 1. First, a brief overview of the /ɹ/ phoneme was given verbally, without any specific articulatory instruction. Participants were informed that native speakers produce the /r/ sound in a variety of ways (with a variety of articulatory gestures), but that all of these methods produced a certain spectrographic image. The Visi-Pitch™ Real-Time Spectrogram module was then opened, and the investigator produced the utterance “hello” into the top screen, and the participant was asked to repeat the utterance into the bottom screen. A very brief orientation followed, in which the beginning and ending of the word was pointed out, and it was noted that each utterance may look a little bit different even though it is still correct.

Subsequently, participants were familiarized with the target spectrogram through a series of printed images highlighting a) the target formant (F3) b) the treatment target area (below 2300 Hz) and c) how these look together. A series of “practice” images depicting high-contrast “correct” and “incorrect” /ɹ/ spectrograms was reviewed with high verbal feedback until the participant was familiar with the task (Appendix I).

A period of “free practice” was then introduced, in which the participant was asked to “try anything” and make sounds into the microphone with continuous live spectrographic feedback. Periodically, the investigator stopped the live spectrographic stream to point out formants and formant shapes. This period lasted a few minutes until the participant indicated that she understood the relationship between making sounds and seeing a spectrographic image.

Treatment

Participants were seated as above with the same microphone, computer, and speaker. For treatment, the Real-Time Spectrogram module of VisiPitch™ IV model 3950 was used for live spectrographic display. This module provides a horizontal split screen, allowing for display of a target in one half of the screen and trials in the other half.

Throughout treatment, participants were provided feedback exclusively related to the visual feedback. The native-speaker spectrogram model was displayed in the top half of the screen as a visual feedback guide, which participants attempted to match in the bottom half of the screen through repeated trials. Improvement was defined by a low F3 (under, overlapping, or directly contacting the 2000 Hz mark, visually approximating an “under 2300” F3) and smooth lines (i.e. no “breaks” which would indicate a tap). Throughout the entire treatment, the experimenter was seated next to the participant and provided verbal feedback on the spectrograms (e.g. “that looks lower than the last one, let’s try to get it even lower if we can”, “You matched the shape of the example well, let’s try to get it even lower” “that one looks a little flat, see how the example has an arch in it”). Participants quickly understood the task and began to self-monitor effectively, with comments like “that was too high” or “I see a break” (Figure 2-1 for a visual depiction).

At no time was explicit articulatory instruction given. If participants requested assistance or instruction on how to “move the line down” they were referred to the suggestions sheet (Appendix G) and reminded that, because native speaker /r/ is produced differently by different individuals, there is no one “correct” way to reach the goal (essentially, they were simply told to experiment, or “try something different”).

In addition, at no time was the phoneme /l/ mentioned in any way by the experimenter, and no contrast or mention was made when participants produced /l/ during treatment (it was simply defined as an F3 that was “too high”).

Stimulus Progression

The treatment protocol was hierarchical with 3 Phases and 3 stages of reduced cueing within each phase. Each phase included 5 hierarchically ordered stimuli, with incremental increases in task difficulty. For example, Phase 1 began with the stimulus item “ra”, a low back, open, lax vowel, in order to reduce possible interference of tongue height and tension associated with the primary substitution errors (specifically, /l/ and the voiced alveolar tap). Following this logic, Phase 1 stimuli presentation subsequently progressed upward through high back, rounded vowels, then low front vowels, and finally to the high front, tense vowel /i/, in the nonword stimulus “ree”. (Appendix B).

It is important to note that because it took more time to meet criteria at the beginning of treatment, the phases did not take an equal amount of time. Phase 1, Stage 1 lasted considerably longer than subsequent stages and phases (4 weeks/12-13 sessions of treatment were dedicated solely to Phase 1, Stage 1, compared to around three sessions for subsequent phases). This provided ample opportunity for practice of the initial 5 stimuli, which were repeated in varying order throughout the first phase. Subsequent stage 1s included an initial group of 40 trials of each word for practice (whether or not criteria was met, to ensure adequate practice of newly-introduced words) followed by a second group of trials with the same words in which the participant attempted to meet criteria.

Each phase added one incremental increase in task complexity. Phase 2 added the continuant /m/ in final position to the same CV syllables used in Phase 1. Phase 3 followed the same pattern, but with /b/ replacing /m/ in the final position. (Appendix J for complete list and order of stimuli.)

Stages: Reduced cueing

In order to move toward more independent production, each phase consisted of three stages of reduced cueing conditions, each of which repeated the same set of 5 stimuli.

Stage 1: Auditory and Visual Condition. Each Stage 1 presented that phase's stimuli (in order) in three ways: orthographically (index cards), auditorily (NS example recording), and visually (corresponding spectrogram of the NS example recording, as well as real-time spectrographic display of the participant's own productions). First, the stimulus was presented orthographically on an index card. Then, the auditory native speaker model recording was opened and played in the top half of the window, providing a simultaneous spectrographic visual example across the screen. Participants were then asked to attempt to match the picture above by recording and viewing their own attempt in the bottom half of the screen. (Figure 2-1 for examples). The native speaker recording was played once before every participant attempt. When criteria were met for four of the five words, the participant moved to the next stage.

Stage 2: Visual Only Condition. Stage 2 presented the same stimuli in the same order as in stage 1, but this time without the auditory example (orthographic and visual cues only).

Stage 2 presented the same stimuli orthographically and visually in the same order as in stage 1. However, the spectrographic example image was static and the native speaker model recording was not played. Participants continued with trials in order, and criteria were the same.

Stage 3: Delayed Feedback Condition. This final stage presented only an orthographic cue and delayed visual feedback. In stage 3, the same 5 words were presented orthographically only, and recorded on PRAAT©. After recording, the investigator asked the participant to give a binary up/down judgment on correctness (i.e. “Do you think that was right?”). Then, the recording was viewed and measured by the investigator, and the participant was able to view only her own corresponding spectrogram and results (without any NS example spectrogram for comparison).

Progression

For Stage 1s of each phase, participants were first provided at least 40 repetitions of each word in the set, to familiarize themselves with new words as they were introduced. After this had been completed, the word set was re-started and attempts were made toward reaching criteria. For stages 1 and 2, participants were given 40 opportunities to meet criteria (a consecutive string of 8/10 correct) on each individual word. If criteria were met within 40 trials, the participant was moved to the next consecutive word in the set. This process continued until all words in the set had been attempted, or until the session time had elapsed.

For Stage 3, participants were given 4 attempts at each word. Criteria to move on to the next phase an F3 measurement of 2300 or less during /r/ production of 8/10 total for all 5 words, with at least one correct answer from each word category.

It is important to note that during the very first stage of treatment (Phase 1, Stage 1) participants were encouraged to meet the stricter criteria of an F3 under 2000 Hz if they could, though the criteria for stage progression were the same (under 2300 Hz).

This stage lasted considerably longer than the other stages (approximately 1 month compared to a day or two) as the participants learned how to form the phoneme and attempted to meet criteria. During this first stage, the same 5 initial words were repeated in the auditory-visual feedback condition until criteria were met, as described above.

Participant preference/request

In order to reduce potential frustration when a word was difficult or repeated many times, participants were provided with three options: Move, Stay, or Free Practice. Any of these options were permitted at any time throughout treatment.

"Move" allowed participants to move on to the next word in the set. The word was returned to later in the session or in a successive session. "Stay" allowed participants to continue further practice on the same word. Participants typically selected this when incremental improvements were being seen but criteria had not yet been met (e.g. if the F3 shape was moving gradually lower in successive trials, but had not yet met criteria). In "Free Practice," participants were allowed unlimited continuous attempts at production with live spectrographic visual feedback (i.e. without the customary pause required for measurement, data collection, and replaying of the auditory example). During Free Practice data was not collected, and participants were encouraged to experiment freely and watch the F3 in order to experiment with their own articulatory hypotheses.

Probes

Once per week (or every third session) a recording was made of the full set of treatment, non-treatment, and control words (101 total items). Probes were always administered before the treatment began. Just as in baseline testing, no feedback was given.

Treatment scoring and reliability

Trained native speaker judgments were used for scoring baseline and probe items. Listener judgments are considered the gold standard for the evaluation of L2 pronunciation (Derwing & Munro, 2009; Saito & Lyster, In Press). After a period of working together to define rules and ensure good inter-rater-reliability, three trained native speaker judges (the author, the author's supervisor, and a trained research assistant) listened to individual recordings of each baseline and probe utterance and assigned one of three judgments: 1) correct, in-category /ɹ/, 2) incorrect, out of category (another sound such a tap, /l/, or other), or 3) ambiguous, impossible to judge (infrequently assigned). See Appendix G for details on the scoring guidelines the raters used.

Composite scores for each word were provided based on a) unanimous agreement or b) 2/3 agreement. Any scores with a three-way disagreement (correct, incorrect, ambiguous) were re-judged in a separate meeting of all three judges, who listened simultaneously and judged the items again. The raters did not know their original rating when they re-scored these words. The new scores were combined in the same way as above unless a tie persisted, in which case the ambiguous rating was automatically assigned. Scoring reliability averaged 96% for P1 (unanimous plus 2/3) with an average of 62% unanimous ratings (4% of ratings resulted in a 3-way tie). For P2, reliability averaged 98% with 56% unanimous ratings (2% of ratings resulted a in 3-way tie).

A trained research assistant observed one-third of the training sessions live and recorded whether the clinician was adhering to the prescribed protocol with special attention on cueing to check for the use of any specific cues other than looking at the spectrogram (such as articulatory cues). Reliability was 99.2% for adhering to the training protocol for item presentation (one word presented out of order once), and 100% for avoidance of articulatory cueing or instruction.

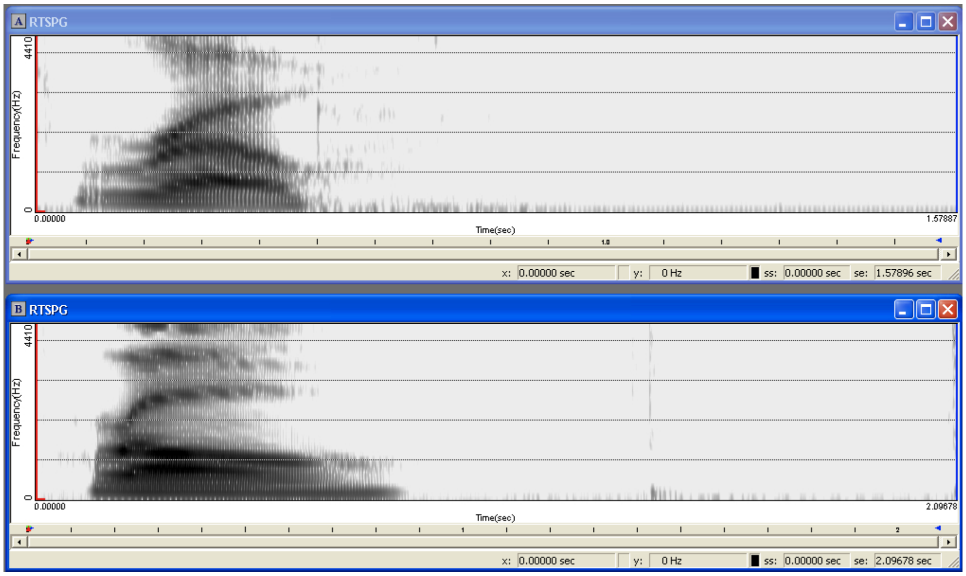


Figure 2-1. Training screen shot A. NS example is visible in the top half of each screen, with a participant trial below. A. “row” with a correct participant production (counts toward criteria).

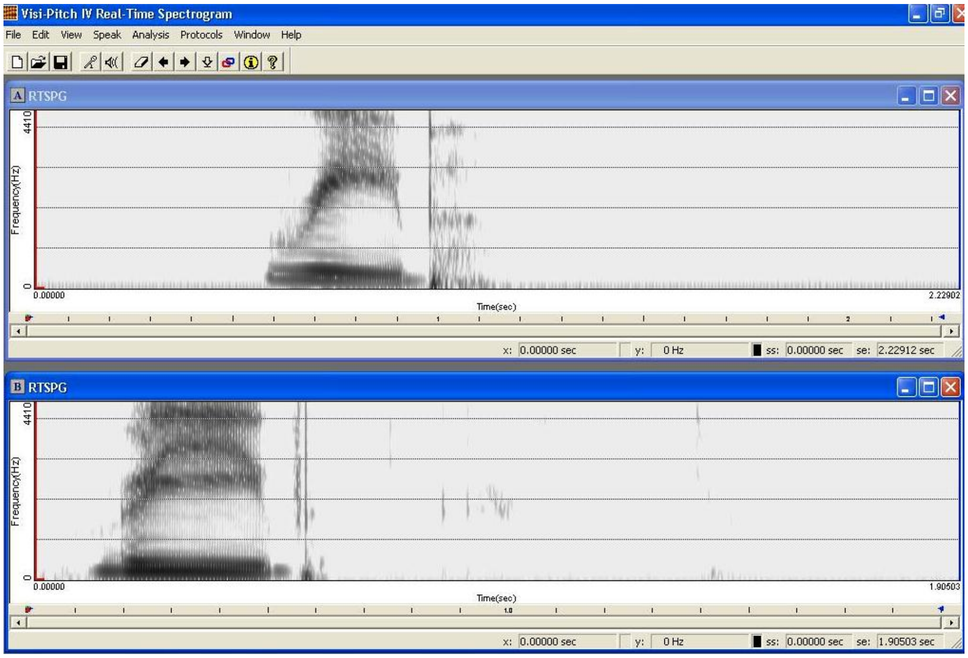


Figure 2-2. Training screen shot B. “reeb” with incorrect production (F3 is high, does not count toward criteria).

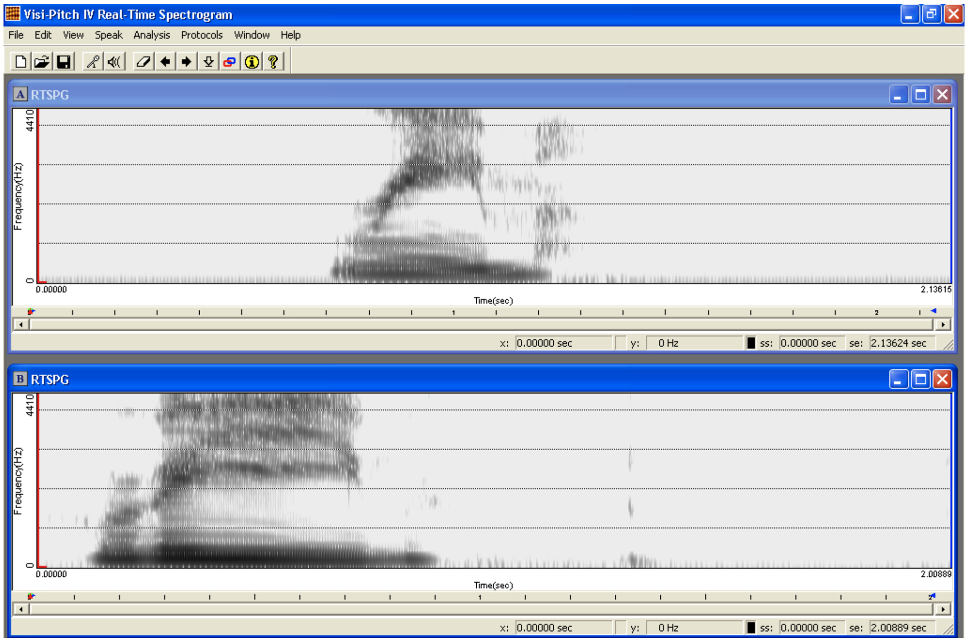


Figure 2-3. Training screen shot C. “ream” with an incorrect production (“break” indicating a tap or flap is visible, does not count toward criteria).

CHAPTER 3
RESULTS

To evaluate Research Question1 and Research Question 2 (magnitude of change), effect sizes (d) (Beeson & Robey, 2006) were calculated to get an “index of durability” (Beeson & Robey, 2006, p. 167) from baseline to last data point in the treatment phase for both participants. Effect sizes were calculated by determining the difference between the final data probe point and the average of the baseline points (N = 5) and diving that difference by the standard deviation of the baseline. To this author’s knowledge, there are virtually no published studies with effect size benchmarks on /r/ training results in Japanese English Language Learners, so the findings in the current study will provide preliminary benchmarks. Saito & Lyster (in press) calculated effect sizes for their production treatment study using F3 height measurements, which is dissimilar from the NS judgment paradigm used here (however in “word reading”- which is most similar to our task- they had an effect size of 0.59). Also, results can be compared within participants across stimuli in order to gauge relative improvement, particularly with trained and untrained stimuli.

The results for Research Question 1, 2, and 3 are described below. See Figures 3-1, 3-2, and 3-3 for visual representation of the data and Tables 3-1 and 3-2 for the raw data.

Participant 1

Research Question 1: P1’s baseline productions of /ɹ/ were rated as correct or incorrect, and stable (i.e., non-ascending) baselines were established for all syllables. She then participated in seven weeks of treatment with three sessions per week (total of 13 sessions including introductory session). Before treatment began on the last day of treatment for each week, P1 read the same words she had read for the baseline sessions in order to gauge potential improvement. It was intended for P1 to continue through the entire protocol; however, she had to discontinue the training due to travel requirements. Consequently, P1 was trained on five r-initial syllables ending in different vowels (Phase 1 stimuli), so the remaining 10 intended treatment items (from Phase 2 and 3) were added to initial position generalization measures for a total of 81 probe items (Research Question 2). The average baseline for Participant 1 was 2.0, and Probe 4 was 4.0. Thus, the calculated effect size between the baseline results and the last treatment phase probe (Probe 4) was 2.0.

Research Question 2: Research Question 2 investigated generalization to untrained ɹ-initial words as well as to untrained words in untrained syllable shapes with /ɹ/ in various positions. The untrained ɹ-initial baseline average was 10.80 (N = 27), and Probe 4 was 26.0 with an effect size of 4.65. Similar to the r-initial syllable shape is the medial shape where the baseline average was 0.8 (N = 12) with a Probe 4 average of 5 resulting in an effect size of 5.02. The cluster shape baseline average was 4.4 (N = 19) with a Probe 4 of 8 resulting in an effect size of 3.16. Final position started with an average of 4.5 (N = 15) in baseline with a Probe 4 of 5, resulting in an effect size of 0.22. Finally, the control condition (ð) had the lowest effect size (0.06) of all conditions, indicating that experimental control was exhibited.

Research Question 3: In the pre-treatment perception task P1 scored 33/60 for 55% correct. In post-treatment (when Probe 4 was collected), she scored 37/60 for 61.67%.

Participant 2

Research Question 1: P2 had five baseline sessions where her production of /ɹ/ was rated as correct or incorrect. She then participated in 3 phases of treatment. The first phase was 5 weeks on the syllable ɹV, the second was one week on ɹVm, and the third was one week on ɹVb. During each phase, the syllable shapes trained in the other phases were kept in baseline in order to evaluate the potential effect of treatment on them. The average baseline for Phase 1 stimuli (ɹV) was 2.2, and at the end of Phase 1 (Probe 5), the total was 4, resulting in an effect size of 1.6. At the end of this phase, the syllable shape ɹVm had reached ceiling (100%), average at baseline, with an effect size of 2. However, according to protocol, ɹVm was trained in phase two, and P1 reached criteria within one week, and her probe was maintained at 100%. For phase 3 stimuli (ɹVb), there appeared to be a slight linear increase during Phase1 treatment, but this was not maintained in Phase 2. However when the ɹVb syllable was trained, in Phase 3, production returned to 80%, with an effect size of 1.5 over baseline.

The average baseline for Phase 2 stimuli was 3 and probe 7 was 4 for an effect size of 1. The average baseline for Phase 3 stimuli was 1.8 and probe 7 was 4 for an effect size of 1.5.

Research Question 2: Research Question 2 investigated generalization to untrained r-initial words as well as to untrained words in untrained syllable shapes with /ɹ/ in various positions. Even though there were three phases trained with P2, they were all r-initial syllables, so generalization will look at the entire span of probes for each participant.

The untrained r-initial baseline average was 13 (N = 19), and Probe 7 was 18 with an effect size of 1.71. Similar to the r-initial syllable shape is the medial shape where the baseline average was 4.8 (N = 11) with a Probe 7 score of 9 resulting in an effect size of 1.6. The cluster shape baseline average was 2.8 (N = 16) with a Probe 7 of 5 resulting in an effect size of 1.5. Final position started with an average of 4.2 (N = 13) in baseline with a Probe 7 of 15, resulting in an effect size of 6, and performance at ceiling. Finally, the control condition (ð) had no effect size at all (0.0), indicating that experimental control was exhibited.

Research Question 3: In pre-treatment, the perception task was administered to P2 with an score of 37/60 for 62% correct. In mid-treatment (collected before treatment during session 15), she scored 46/60 for 77%.

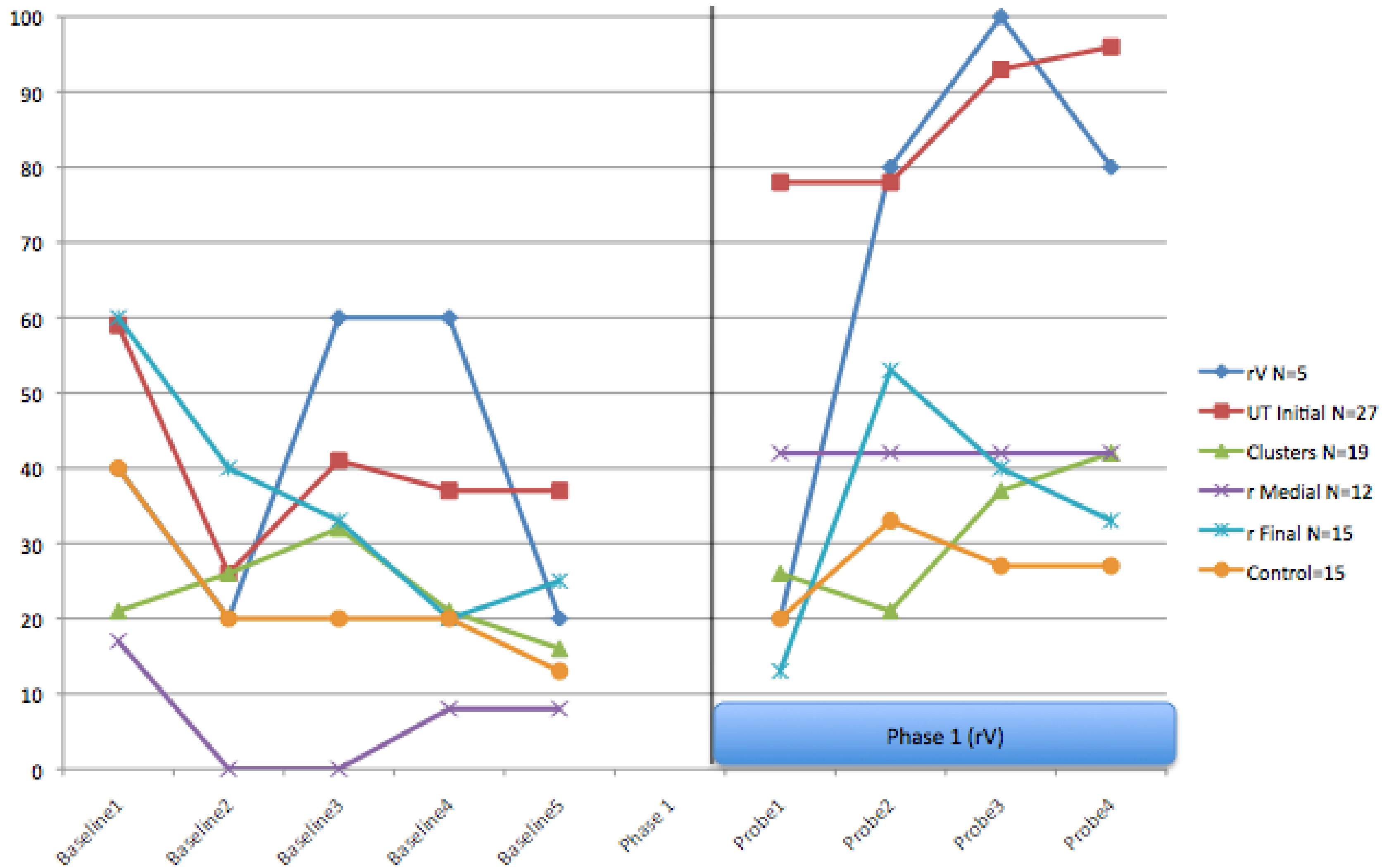


Figure 3-1. Participant 1: trained and untrained item results by percentage correct.

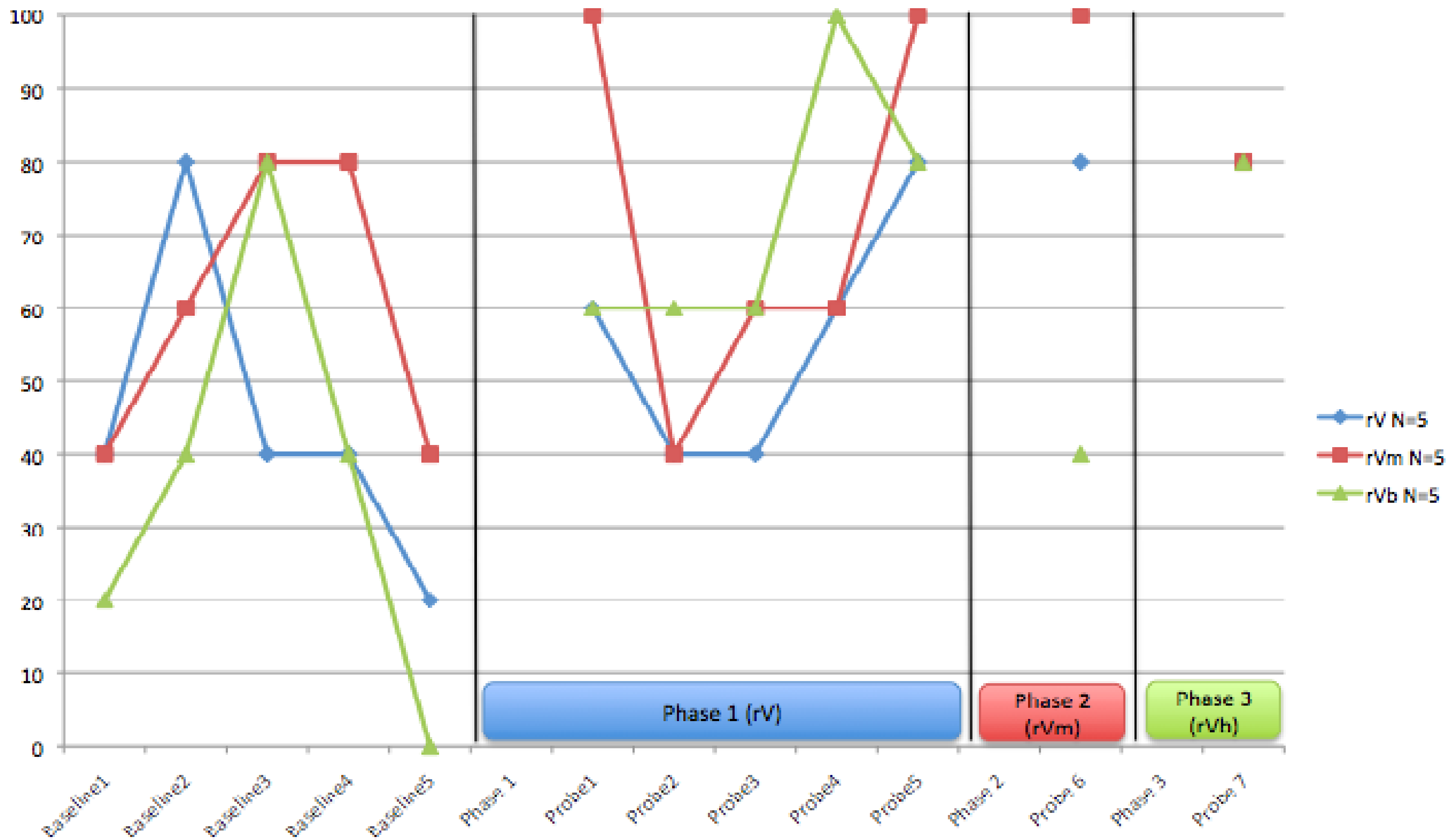


Figure 3-2. Participant 2, A: trained item results by percentage correct.

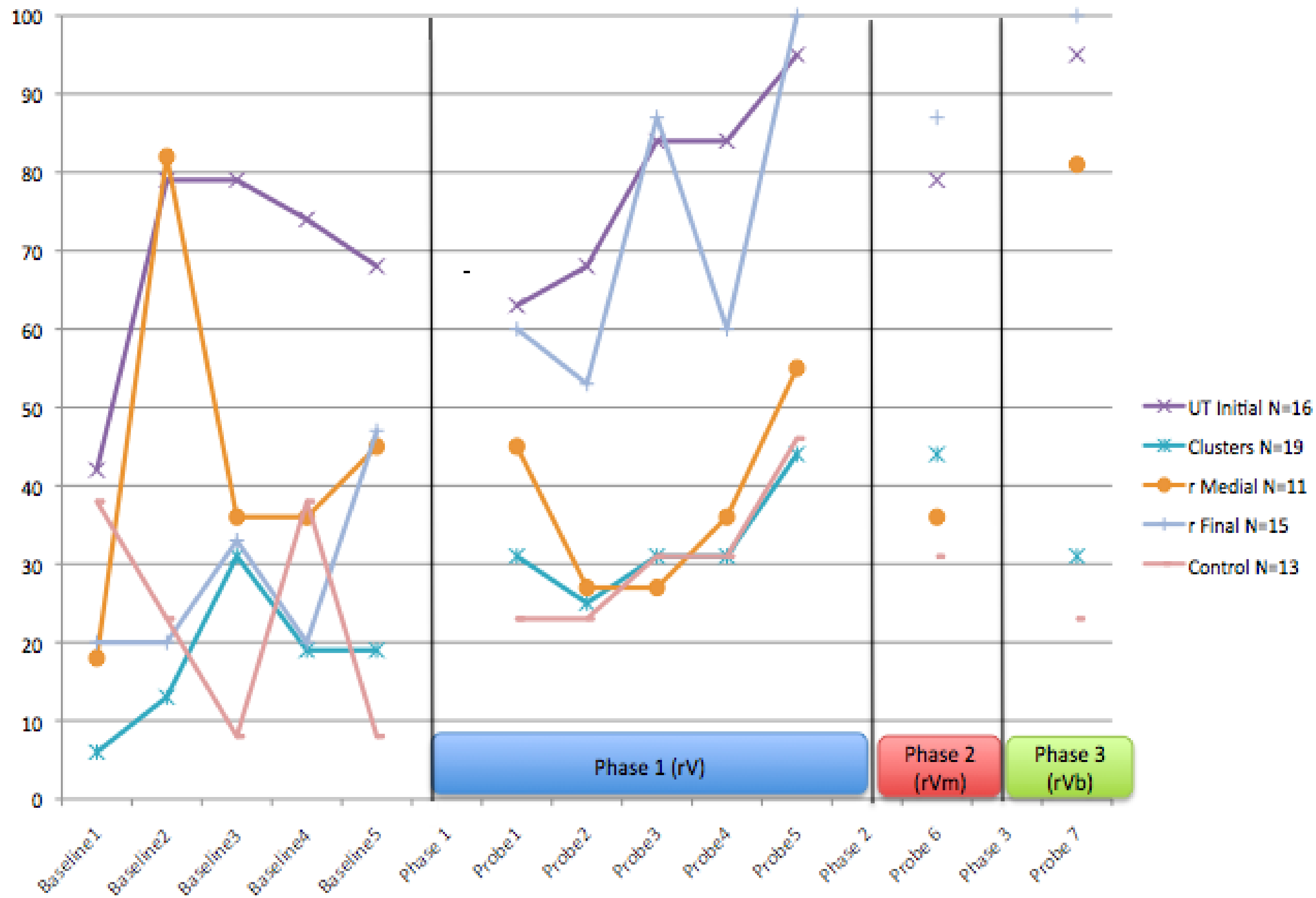


Figure 3-3. Participant 2, B: untrained item and control results by percentage correct.

Table 3-1. Participant 1 production effect size calculations.

P1	Probe 4	Baseline Average	Baseline Standard Deviation	*d*
Trained				
Phase 1	4	2	1	2
Untrained				
UT Initial	26	10.8	3.27	4.65
Clusters	8	4.4	1.14	3.16
Medial	5	0.8	0.84	5.02
Final	5	4.5	2.30	0.22
Control	4	3.4	10.11	0.06

Table 3-2. Participant 1 raw production data.

P1		Training	Generalization				Control
		Phase 1	UT Initial	UT Clusters	UT Medial	UT Final	
		ɹ	ɹ	ɹ	ɹ	ɹ	/ð/
	N	5	27	19	12	15	15
Baseline 1	Raw	2	16	4	2	9	6
	Percent	40%	59%	21%	17%	60%	40%
Baseline 2	Raw	1	7	5	0	6	3
	Percent	20%	26%	26%	0%	40%	20%
Baseline 3	Raw	3	11	6	0	5	3
	Percent	60%	41%	32%	0%	33%	20%
Baseline 4	Raw	3	10	4	1	3	3
	Percent	60%	37%	21%	8%	20%	20%
Baseline 5	Raw	1	10	3	1	4	2
	Percent	20%	37%	16%	8%	25%	13%
Training							
Probe 1	Raw	1	21	5	5	2	3
	Percent	20%	78%	26%	42%	13%	20%
Probe 2	Raw	4	21	4	5	8	5
	Percent	80%	78%	21%	42%	53%	33%
Probe 3	Raw	5	25	7	5	6	4
	Percent	100%	93%	37%	42%	40%	27%
Probe 4	Raw	4	26	8	5	5	4
	Percent	80%	96%	42%	42%	33%	27%

* Three words were removed from UT Initial N because of lost recordings from a probe or baseline: read, rude, and rusk.

** Five words were removed from UT Final N because they were frequently pronounced as disyllabic with medial /ɹ/: dire, ire, mere, spare, and spire. Note that no pronunciation instruction whatsoever was provided for untrained words.

Table 3-3. Participant 2 production effect size calculations.

P2	Probe 7	Baseline Average	Baseline Standard Deviation	*d*
Trained				
Phase 1	4	2.2	1.1	1.64
Phase 2	4	3	1	1
Phase 3	4	1.8	1.48	1.48
Untrained				
UT Initial	18	13	2.92	1.71
Clusters	5	2.8	1.48	1.48
Medial	9	4.8	2.59	1.62
Final	15	4.2	1.79	6.03
Control	3	3	2	0

Table 3-4. Participant 2 raw production data.

P2		Training			Generalization				Control
		Phase 1 ɹ	Phase 2 ɹ	Phase 3 ɹ	UT Initial ɹ	UT Cluster ɹ	UT Medial ɹ	UT Final ɹ	/ð/
	N	5	5	5	16	19	11	15	13
Baseline 1	Raw	2	2	1	8	1	2	3	5
	Percent	40%	40%	20%	42%	6%	18%	20%	38%
Baseline 2	Raw	4	3	2	15	2	9	3	3
	Percent	80%	60%	40%	79%	13%	82%	20%	23%
Baseline 3	Raw	2	4	4	15	5	4	5	1
	Percent	40%	80%	80%	79%	31%	36%	33%	8%
Baseline 4	Raw	2	4	2	14	3	4	3	5
	Percent	40%	80%	40%	74%	19%	36%	20%	38%
Baseline 5	Raw	1	2	0	13	3	5	7	1
	Correct	20%	40%	0%	68%	19%	45%	47%	8%
Training									
Probe 1	Raw	3	5	3	12	5	5	9	3
	Correct	60%	100%	60%	63%	31%	45%	60%	23%
Probe 2	Raw	2	2	3	13	4	3	8	3
	Correct	40%	40%	60%	68%	25%	27%	53%	23%
Probe 3	Raw	2	3	3	16	5	3	13	4
	Correct	40%	60%	60%	84%	31%	27%	87%	31%

Table 3-4. Continued

P2		Training			Generalization				Control
		Phase 1 ɹ	Phase 2 ɹ	Phase 3 ɹ	UT Initial ɹ	UT Cluster ɹ	UT Medial ɹ	UT Final ɹ	/ð/
	N	5	5	5	16	19	11	15	13
Probe 4	Raw	3	3	5	16	5	4	9	4
	Correct	60%	60%	100%	84%	31%	36%	60%	31%
Probe 5	Raw	4	5	4	18	7	6	15	6
	Correct	80%	100%	80%	95%	44%	55%	100%	46%
Probe 6	Raw	4	5	2	15	7	4	13	4
	Correct	80%	100%	40%	79%	44%	36%	87%	31%
Probe 7	Raw	4	4	4	18	5	9	15	3
	Correct	80%	80%	80%	95%	31%	82%	100%	23%

* 7 words were removed from various position Ns because of lost recordings of a probe or baseline: deary, rusk, spar, drink, tray, teething, and them.

** Five words were removed from UT Final N because they were frequently pronounced as disyllabic with medial /ɹ/: dire, ire, mire, and spire. Note that no pronunciation instruction whatsoever was provided for untrained words.

CHAPTER 4
DISCUSSION

The purpose of this study was to evaluate the effect of spectrographic visual biofeedback on the production of /ɹ/ by JL1s. The results of this preliminary study suggest an overall positive effect of the training on the production and perception of /r/ by JL1s.

Research Question 1

Research Question 1 asked whether treatment improved production of /ɹ/ in trained the words. P1 only participated in one phase of treatment on ɹ V syllable shapes (N=5). The trained items were variable (but did not rise) in baseline, suggesting an unstable representation of /ɹ/. The first probe did not reflect improvement, but the three final probes demonstrate more accuracy and increased stability. Based on visual inspection and the effect size of change, it appears that the trained items improved. However, one syllable, "ree," continued to be challenging throughout the treatment phase. This syllable shape was predicted a priori to be the most difficult because of the high-front tongue position and tension involved serve to interfere with disambiguation from /l/ and flap production.

Participant 2 engaged in three phases of treatment and thus was trained in ɹV, ɹVm, and ɹVb syllables, respectively. For each treatment phase, the effect of treatment on one syllable was evaluated on the other syllables. P2 also showed high variability in trained items during baseline (similar to P1). For probe 5, when only Phase 1 stimuli (ɹ V) had been introduced, Phase 1 gains were *d*=1.6 (N=5). There is a linear trend upward toward probe 5. At probe 7, the only item missed was "ree," following the same pattern as P1. At Probe 6, for which Phase 2 stimuli (ɹVm) had been introduced, Phase

2 gains were *d*=2 (N=5). Phase 3 stimuli (ɹ Vb) were introduced before Probe 7, and results demonstrate Phase 3 gain of *d*=1.5 (N=5). There was no improvement to control items, indicating maintenance of experimental control (i.e. observed improvement was not a result of stimulating the entire language system but was likely due to the treatment itself).

Research Question 2

Research Question 2 asked whether this treatment demonstrated generalization from trained items (ɹ V) to untrained items in initial, medial, and final and cluster positions. Both participants demonstrated a trend of improvement from trained items to untrained items.

P1's increased stability of production due to the five trained items is also reflected in the generalization measures. The items most similar to the treated items- initial position- demonstrated immediate increases and remained high (over 78% correct) throughout the remainder of treatment. Medial position began with a low baseline and immediately rose to 42% correct, which remained stable throughout the remainder of treatment, for the largest effect size that P1 achieved (5.0).

P2 also demonstrated generalization to untrained items, with large effect sizes for each untrained position. Initial position (which was most similar to trained items) demonstrated an effect of 1.7. Dissimilar untreated items demonstrated gains as well, with final position being the largest (*d*=6), followed by medial position (*d*=1.6) and clusters (*d*=1.4). Furthermore, for P2 during probes 1-5, Phase 2 and 3 stimuli may be viewed as generalization measures because they had not been introduced yet. During that time increases were seen in Phase 2 stimuli (*d*=2) and Phase 3 stimuli (*d*=1.5).

However it is also important to note that P2's final position scores did rise on the final baseline measure, so results should be interpreted cautiously.

Research Question 3

An overall increase was demonstrated in the discrimination abilities of both participants. It should be noted that because NS Model 2 (male) was instructed to produce the items in as natural a way as possible, the utterances were challenging (the rate was rapid and no attempt was made toward particular articulatory clarity). However, one of the excluded participants (who had lived in the US for an extended period) was able to judge them with 100% accuracy, thus the task is possible for a highly experienced JL1. Furthermore, because participants were instructed to "write whatever [they] heard", and the phoneme /l/ was never mentioned, this was a true phoneme identification task (not forced choice); evidenced by the fact that the participants sometimes wrote in sounds other than "R" or "L": another letter, two letters, or a "null" mark indicating that they had heard nothing or only the adjacent sound. P1's improvement (6 percentage points) may appear modest, but when combined with P2's improvement (15 percentage points), these results suggest an overall positive influence of production training on perceptual discrimination abilities.

General Discussion

Both participants demonstrated a general linear upward trend in all trained and most untrained items over the course of treatment, suggesting that visual feedback of F3 may be a promising approach to training /ɹ/ in JL1s. Control was maintained, indicating that all effects were due to the treatment. To this author's knowledge, this is the first study to use single subject design on /ɹ/ production with spectrographic biofeedback for JL1s. Thus there are few appropriate benchmarks (besides possibly

Saito & Lyster’s in press finding of 0.58 as described above). However, visual inspection and percentage increases suggest improvement.

Instability in baseline productions of /ɹ/ generally smoothed out over the course of treatment, suggesting an overall solidification of the phoneme from a volatile pre-treatment state. This follows the concept that JL1s may be hypothesizing about the parameters of /ɹ/ in pre-treatment, sometimes hitting the mark and sometimes missing it entirely, but with little ability to self-monitor their own productions for accuracy (following Goto, 1971). This treatment appeared to assist the participants in formulating a more reliable parameter for /ɹ/ by allowing them to test self-generated articulatory hypotheses with reliable feedback.

This treatment appears to hold promise as an alternative to articulator placement instruction, as participants were free to hypothesize about and choose any combination of constrictions that created the effect they saw onscreen, with immediate knowledge of their results. While it was not possible to know which lingual position or positions were being used by the participants or how they may have changed over time as their approximations became spectrographically and acoustically closer to native-speaker /ɹ/, there is at least some indication that bunched /ɹ/ may have been a factor in correct productions during later stages of treatment. P1 noted during her exit interview that using the “back” of her tongue more than the “front” was of help in achieving correct productions. Around Probe 5, P2 began to intermittently make a new type of substitution error in treatment trials: the back consonant /g/ began to appear occasionally in place of /ɹ/ (there had only ever been front consonant substitutions before). Taken together, these suggest the possibility of the participants beginning to utilize a more back or bunched-type lingual position for /ɹ / as treatment progressed.

Anecdotally, all three judges independently noted qualitative improvements in production by both participants for both “correct” and “incorrect” categories. A token marked “incorrect” in both baseline and probe typically sounded much closer to in-category (i.e. a distorted /ɹ/ rather than an entirely different phoneme) during probes than treatment, and tokens marked “correct” in both baseline and probe- though judged as in-category for both conditions- typically sounded more native-like during probes.

While both participants produced gains in generalization measures, they did so in markedly different ways. P1 improved most in medial position, with large gains also seen in UT initial position. P2, on the other hand, improved most in final position, with relatively small gains in medial position. This pattern suggests that treatment affected each participant differently- P1 generalized most to consonantal /ɹ/, while P2 generalized most to r-colored vowels (final position or dark /ɹ/). Both demonstrated gains in medial position, which could support this idea, as medial position /ɹ/ can be produced more similarly to either a consonantal/initial /ɹ/ or an ɹ-colored vowel depending on word stress and style of production. Thus, P1’s generalization gains could be attributed mainly to consonantal /ɹ/ categorization improvement, and P2’s generalization gains could be largely attributed to r-colored vowel improvement. This variability in generalization results follows Elbert and McReynolds’ (1975) findings that training /ɹ/ in initial position can create carryover to any other /ɹ/ allophone, with different levels of generalization to different allophones varying by participant with no particular pattern noted.

Pre-treatment, the /ɹ/ seemed to be “unstable” for both participants, which follows logically from Lotto and colleagues’ findings of overlapping category boundaries in production (Figure 1-1). Both reported that before treatment, NS listeners would “sometimes and sometimes not” understand their attempted /ɹ/ productions, and the participants couldn’t figure out why. This was particularly salient for P1, whose name includes the letter “R” in English orthography. Initial variability was also evidenced in the fact that both participants were able to achieve 8 consecutive tokens of /ɹ/ with F3 under 2000 Hz in the introductory session with the stimulus “raw” and then were unable to reproduce these results for another few sessions. Both reported that before treatment they “didn’t know how far away [they were]” from NS productions and credited the treatment with clarifying the situation. Over time, each sound appeared to ‘get worse’ before it got better- this may be due to a reorganization of the mental representation of the phoneme, occurring gradually word by word or sound by sound, and some of this variability can be seen clearly in the line graphs. Some words or categories appeared to improve incrementally, and then as a new challenge was introduced, dip down before finally rising.

Affective and educational differences may have played a role in the variation between the results of P1 and P2. P1 reported that over the course of her English-language education she had never been taught any particular articulatory positioning for /ɹ/, rather her instructors asked her to mimic their own productions. P2, on the other hand, was instructed in strict retroflex articulator positioning, and informed by instructors that this phoneme would always be very difficult or impossible for JL1 speakers. She was further instructed that the retroflex position for /ɹ/ could only be achieved with high muscle tension and a great deal of attention to articulator placement during every attempted production (at one point, she even asked Investigator 1 about surgical frenulum clipping, which she had been told would facilitate /ɹ/ production: this was, of course, vehemently discouraged by the investigator).

These past educational and affective differences also may have played a role in how P1 and P2 differed in their tactile realization of what was causing a “break” (alveolar tap or flap) during their /ɹ/ productions. As noted elsewhere, during treatment no articulatory cueing was provided. This meant that when a tap, flap, or other stop was produced, the only feedback the participants received was a blank space in the spectrographic visual feedback. These were called “breaks” because of their visual representation, and both participants were encouraged to avoid “breaks” and to try to “make the line as smooth as possible.” Both participants focused on this frequently during “free practice” times. P1 realized and mentioned within a few sessions that when her tongue touched her alveolar ridge, it caused a “break.” During her exit interview, she also mentioned that this was one of the best “tips” she had learned from treatment about how to make a correct /ɹ/, and that if she used the back of her tongue more than the front, she had better success with the spectrogram (suggesting a bunched /ɹ/ position after treatment). P2 struggled with this for a longer time, exhibiting frustration and describing different hypotheses about the “break” during the first few weeks of treatment, possibly because she had received articulator placement instruction for a tense retroflex position in production of /ɹ/. Just after Probe 5 she began to verbally propose the involvement of lingual contact in creating the “break” and her production of taps or flaps during treatment sessions began to decline more precipitously than before.

There is at least some indication that through the course of treatment, both participants experienced a reorganizing or reshaping of their concept of the phoneme. Notably, P2 indicated in the early weeks that when the spectrogram showed a correct production of /ɹ/, it “didn’t feel any different” than the incorrect productions. Then, as her in-treatment productions were becoming markedly more consistent and correct, she mentioned that the correct /ɹ/ productions felt strange to her, “like an animal sound” that had nothing to do with language. Between Probes 4 and 5, she mentioned that they didn’t “feel like animal sounds anymore,” but rather like a new “language sound.” Furthermore, the results of P2’s Probe 5 appear to reflect this conceptual shift. This supports the idea that F3 is not included in what is considered “language” for JL1s, and that even correct productions may not “feel” correct at first, supporting Goto’s 1971 findings about lack of self-perception of correct tokens. These subjective comments provide fascinating insight into the participant experience of gaining a new phoneme parameter, and are echoes of remarks made by all three of the young NS participants trained by Schuster and colleagues that the phoneme didn’t feel or sound “right” when produced correctly (Schuster et al. 1992; Schuster et al., 1995).

Investigation of perception was limited and interpretation should be cautious. P2’s higher scores in the perception task may have been reflective of the one additional session of treatment that she received before testing. Alternately, there may have been influence of the fact that she receives more NS-like auditory input in her daily life, and thus potentially had more opportunity to generalize perception than P1 did. These results appear to support Sheldon and Strange’s (1982) finding that production can precede perception, as well as that individual results may vary (also see Catford & Pisoni 1970).

Limitations and Future Directions

One limitation was that only two females participated; investigation with more male and female participants is warranted due to some differences in F3 across gender. Further participant demographic diversity in areas such as age, length of study, country of residence, and even L1 would provide additional insight. Second, a more task-specific design for the computer program could provide participants more control of the practice with a user-friendly interface, or color-coded target prompts (e.g. a red F3 center formant frequency, a blue line for the 2300 Hz target). Automatic data collection could provide a richer data set and fine-grained analysis, such as F3 formant heights, response time measures in the no-feedback condition, session data tracking, etc. Third, while we feel confident with the target F3 threshold of 2300 Hz, it would have been helpful to have more normative data on how women's speech production may differ from men's (though Hagiwara, 1995, provides an excellent discussion of this topic). Fourth, trained NS judges, while considered the "gold standard," many not provide full indication of how productions are perceived by naïve NS listeners or NNSs; further exploration into any corresponding increases in overall speech intelligibility with these listener populations is warranted in order to most fully provide for the types of audiences L2 speakers of English encounter in their daily lives. Fifth, detailed error analyses of incorrect productions would likely show more detailed changes in changes over time. For example, substitutions of /l/, while relatively frequent in baseline, precipitously reduced once treatment commenced; by the final probe for each participant utterances marked “incorrect” by the judges were more commonly /ɹ/ distortions rather than substitutions.

Variations in treatment design may also reveal more specific information on how to most effectively provide this type of treatment. There may well be a way to train this sound in a shorter amount of time, and with more efficiency, for example if erroneous tap or /l/ production were explicitly described, with fewer stimuli, or with variations to frequency and intensity of treatment. Semi-cognates and loanwords (e.g. “rue” and “more”) may have had a positive or negative effect on treatment efficacy, further exploration is warranted. While Elbert & McReynolds (1975) describe high efficacy for generalization of treated initial /ɹ/ to other allophones in NS children, further exploration into stimuli selection and progression order are indicated. Furthermore, a large group study of spectrographic biofeedback with the addition of articulator movement measures (such as with a midsagittal articulometer) could provide knowledge of which articulator positions are being used effectively by the learners and subsequently be extrapolated into more effective articulator placement instruction models for new learners in classroom and clinical environments. Overall, we believe that this application shows promising results and could have great potential for instructional use.

APPENDIX A
PARTICIPANT SELF-RATINGS

Participant 1	Please indicate your own level of proficiency in English:						
	Nonfluent						Native Fluency
Reading	1	2	**3**	4	5	6	7
Writing	1	2	**3**	4	5	6	7
Listening	1	2	**3**	4	5	6	7
Speaking	1	2	**3**	4	5	6	7
	In your opinion, how much of a foreign accent do you have in English?						
	Heavily-accented						Native-like
	1	2	**3**	4	5	6	7
	Please rate how often others identify you as a non-native speaker based on your accent in English:						
	Always						Never
	1	2	3	4	5	6	7

Participant 2	Please indicate your own level of proficiency in English:						
	Nonfluent						Native Fluency
Reading	1	2	**3**	4	5	6	7
Writing	1	2	**3**	4	5	6	7
Listening	1	**2**	3	4	5	6	7
Speaking	1	**2**	3	4	5	6	7
	In your opinion, how much of a foreign accent do you have in English?						
	Heavily-accented						Native-like
	1	2	3	4	5	6	7
	Please rate how often others identify you as a non-native speaker based on your accent in English						
	Always						Never
	1	2	3	4	5	6	7

APPENDIX B
TREATMENT STIMULI

Phase	Stimuli
Phase 1 ɹV	1. /ɹɑ/ (raw)
	2. /ɹe/ (ray)
	3. /ɹo/ (row)
	4. /ɹu/ (rue)
	5. /ɹi/ (*ree)
Phase 2 ɹVm	1. /ɹɑm/ (rom)
	2. /ɹem/ (*rame)
	3. /ɹom/ (roam)
	4. /ɹum/ (room)
	5. /ɹim/ (ream)
Phase 3 ɹVb	1. /ɹɑb/ (rob)
	2. /ɹeb/ (*rabe)
	3. /ɹob/ (robe)
	4. /ɹub/ (rube)
	5. /ɹib/ (*reeb)

APPENDIX C
UNTRAINED PROBE STIMULI

UT Initial (20)	UT Medial (12)	UT Cluster (19)	UT Final (20)
Roy	airy	bray	air
rye	Ari	Bree	are
raid	arrow	brew	ear
ram	eerie	bro	ire
read	era	draw	or
rep	deary	drew	mar
rhyme	marry	drink	mare
rib	Perry	gray	mere
rim	sorry	grew	mire
ripe	tarrow	grow	more
road	sparrow	pro	bar
rod	starry	tray	bear
rub		tree	beer
rude		true	bore
rum		try	dire
rasp		abrade	spar
risk		abridge	spare
roast		agree	spear
roost		aground	spire
rusk			spore

APPENDIX D
CONTROL STIMULI

CONTROL (15)
they
though
thy
them
then
these
thine
those
thou
bathe
smooth
teethe
bathing
seething
teething

APPENDIX E
PERCEPTION MINIMAL PAIRS

Perception Minimal Pairs (60)

	/ɹ/	/l/		/ɹ/	/l/
1	care	kale	31	rate	late
2	race	lace	32	core	coal
3	rude	lewd	33	pour	pole
4	rake	lake	34	rag	lag
5	star	stall	35	raced	laced
6	tire	tile	36	car	call
7	raid	laid	37	rode	load
8	reek	leek	38	near	Neil
9	Ross	loss	39	rot	lot
10	rite	light	40	tour	tool
11	rice	lice	41	dire	dial
12	sear	seal	42	sore	sole
13	root	loot	43	deer	deal
14	dare	dale	44	pre*	plea
15	stare	stale	45	cray	clay
16	ran	lan	46	rad	lad
17	mart	malt	47	tear*	tale
18	tar	tall	48	Ron	lawn
19	pyre*	pile	49	rock	lock
20	craw	claw	50	rope	lope
21	pry	ply	51	read	lead
22	pray	play	52	store	stole
23	pair/pear	pail/pale	53	rest	lest
24	tier*	teal	54	red	led
25	peer	peal	55	par	pall
26	door	dole	56	gore	goal
27	crew	clue	57	tore	toll
28	rug	lug	58	rain	lane
29	rid	lid	59	rust	lust
30	reap	leap	60	steer	steal

APPENDIX F
PERCEPTION ANSWER SHEET

Perception Answer Sheet

1	ka__e	31	__ate
2	__ace	32	co__e
3	__ude	33	po__e
4	__ake	34	__ag
5	sta__	35	__aced
6	ti__e	36	ca__
7	__aid	37	__oad
8	__eek	38	nea__
9	__oss	39	__ot
10	__ight	40	too__e
11	__ice	41	di__e
12	sea__	42	so__e
13	__oot	43	dea__
14	da__e	44	p__ee
15	sta__e	45	c__ay
16	__an	46	__ad
17	ma__t	47	ta__e
18	ta__	48	__on
19	pi__e	49	__ock
20	c__aw	50	__ope
21	p__y	51	__ead
22	p__ay	52	sto__e
23	pa__e	53	__est
24	tea__	54	__ed
25	pee__	55	pa__
26	do__e	56	go__e
27	c__ew	57	to__
28	__ug	58	__ane
29	__id	59	__ust
30	__eap	60	stee__

APPENDIX G
SUGGESTIONS SHEET FOR JL1 PARTICIPANTS

Ideas about what to try:

*Try anything! It's ok to get it "wrong" a lot of times and try different things. We'll try a lot of things and then look at the machine to see what happens. We can try as long as we want to. The purpose is to find what works for YOU, and we don't know what that is yet! Please try anything you can think of, it's ok to try and try again. Here are some ideas:

1. Go slow.
2. Experiment.
3. Change one thing about your mouth.
4. Change two things about your mouth.
5. Try small changes.
6. Try big changes.
7. Change everything- try something new.
8. Relax your tongue.
9. Relax your jaw.
10. Change the front of your tongue- big or small, up or down, or just differently.
11. Change the back of your tongue- big or small, up or down, or just differently.
12. Change the shape of your tongue (more flat, more bent, or just different).
13. Move your jaw up or down or just differently.
14. Move your throat open or closed or just differently (maybe this is hard to feel- if so, don't worry about it).
15. If you can think of new things to try, please try them!
16. Don't forget to breathe.
17. Have fun!

APPENDIX H
INSTRUCTIONS FOR NATIVE SPEAKER JUDGES

R 3-category listening judgments:

1- It's an R. For sure. That's what you heard.
Diphthongs that are glided so strongly as to become disyllabic: i.e. "mire" becomes "meyer" = as long as the R is correct and in the correct position (final), mark it correct.

0- It's not an R. For sure. Code these like in the next column (C) like this:

- Touch (For tap/touch of tongue on the palate- this category includes voiced alveolar taps of all varieties, including abbreviated trills, flaps, taps, etc.).
- L (It's an L)
- D (It's a D)
- W (It's a W*)*
- Two (Two sounds or some extra consonant-like sound in there: not just because they added a vowel or syllable elsewhere in the word, but the actual R location has two sounds in it.)
- Unfinished (The R might be started, but never finished or just trails off. Also use this for a deleted R or a vowelized R.)
- Epenthesis: If an /r/-cluster is broken by a vowel (i.e. "pre" becomes "puh-ree" etc), mark it incorrect even if the /r/ sounds right- it is not being produced correctly in the called-for position.
- Other- just try to give a little note on what you heard/what sounds incorrect if you can, but don't be afraid to mark it "other" if it's just wrong but you can't explain why.

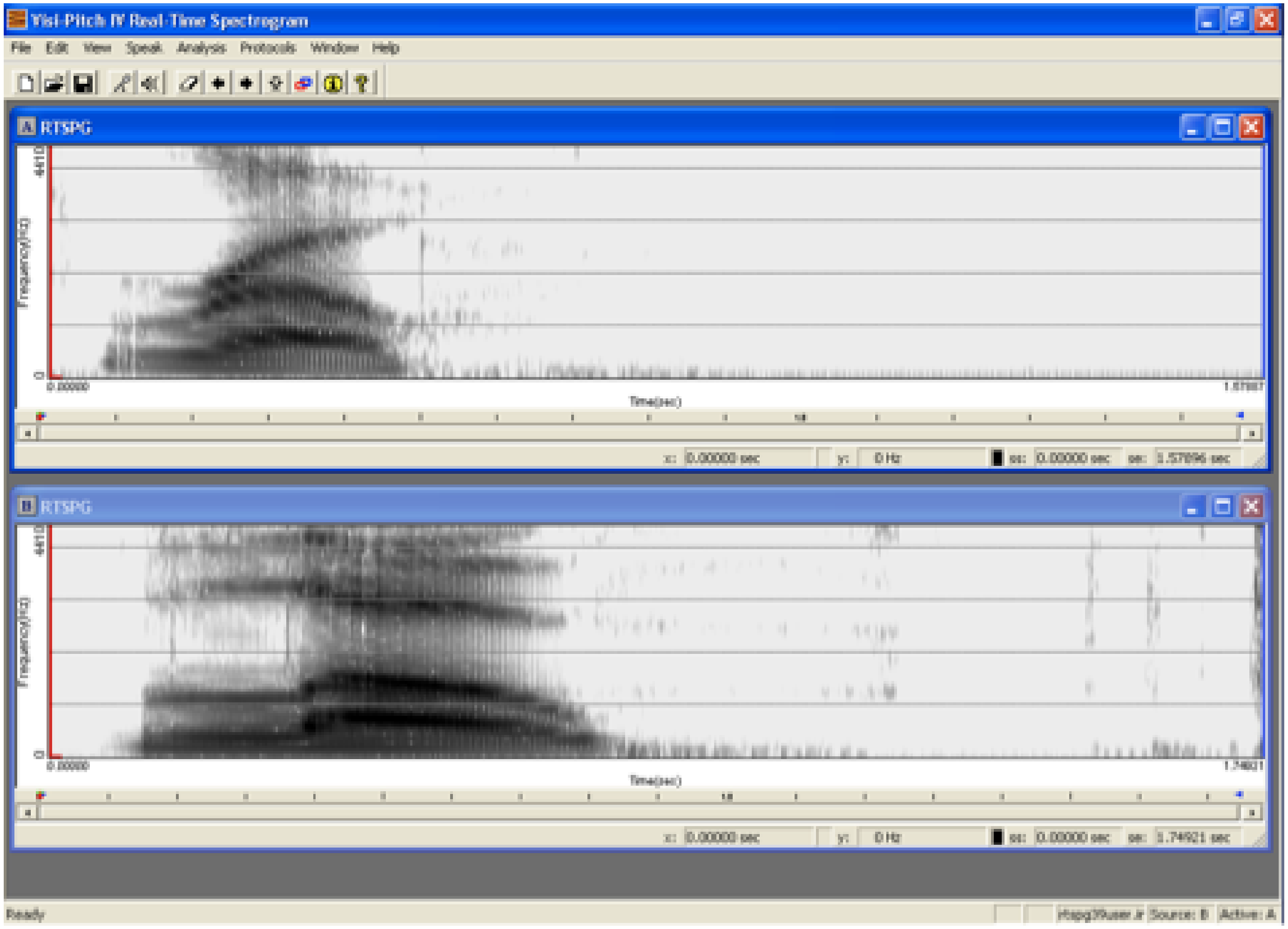
?- Ambiguous: it's not clear whether it's an R or not, maybe it took a few listens and still you aren't sure; it sounds right but kind of wrong, or you can't tell; maybe it's too short to decide or otherwise unclear. Not a definite yes, but not a definite no. Give this a question mark in column B. If you have a comment on why it seems unclear, write it in column C.

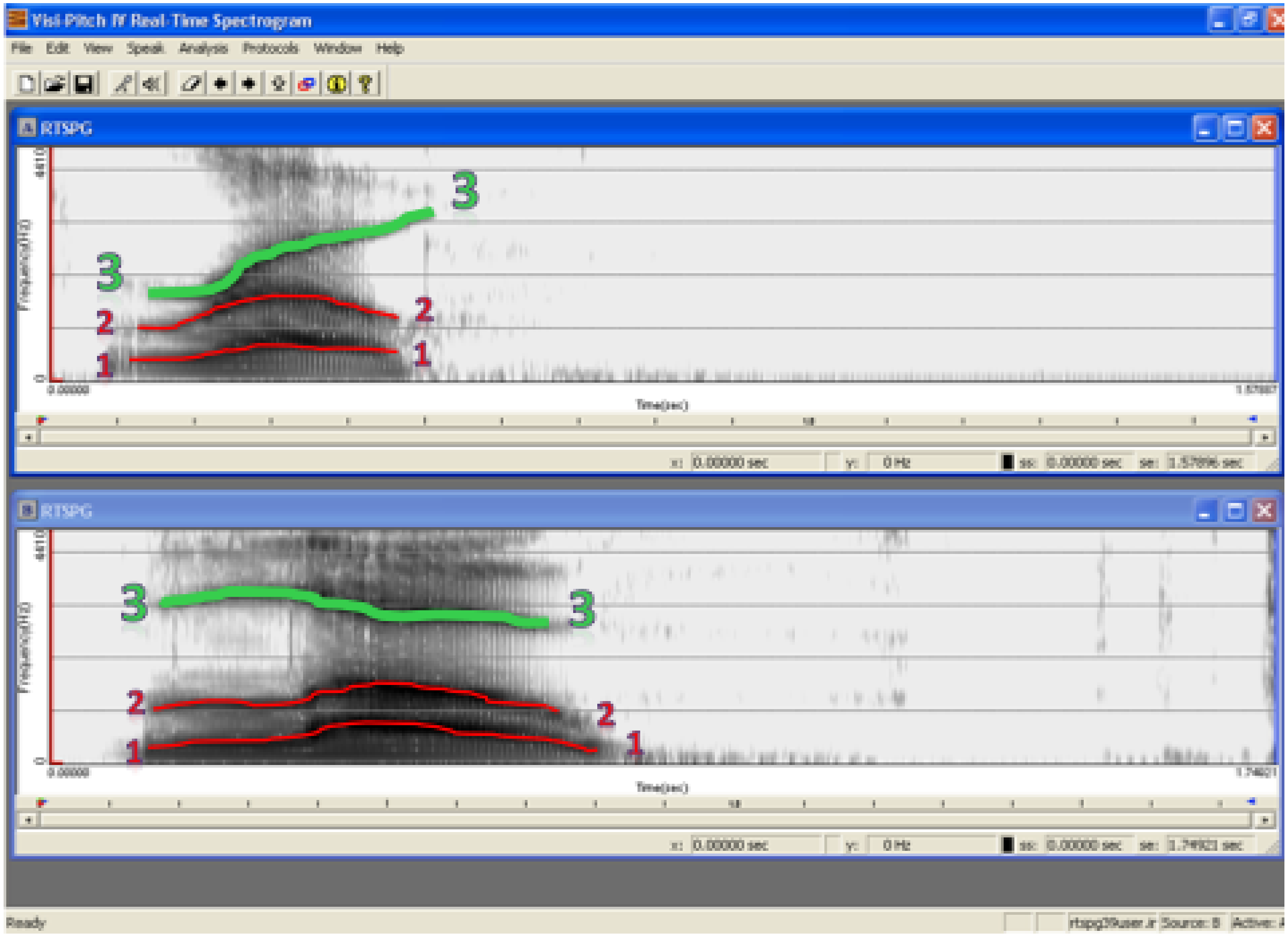
CONTROL Listening Judgment
Incorrect:

- Voiceless
- Z
- S
- D
- Deletion
- Anything else that is NOT a voiced interdental fricative

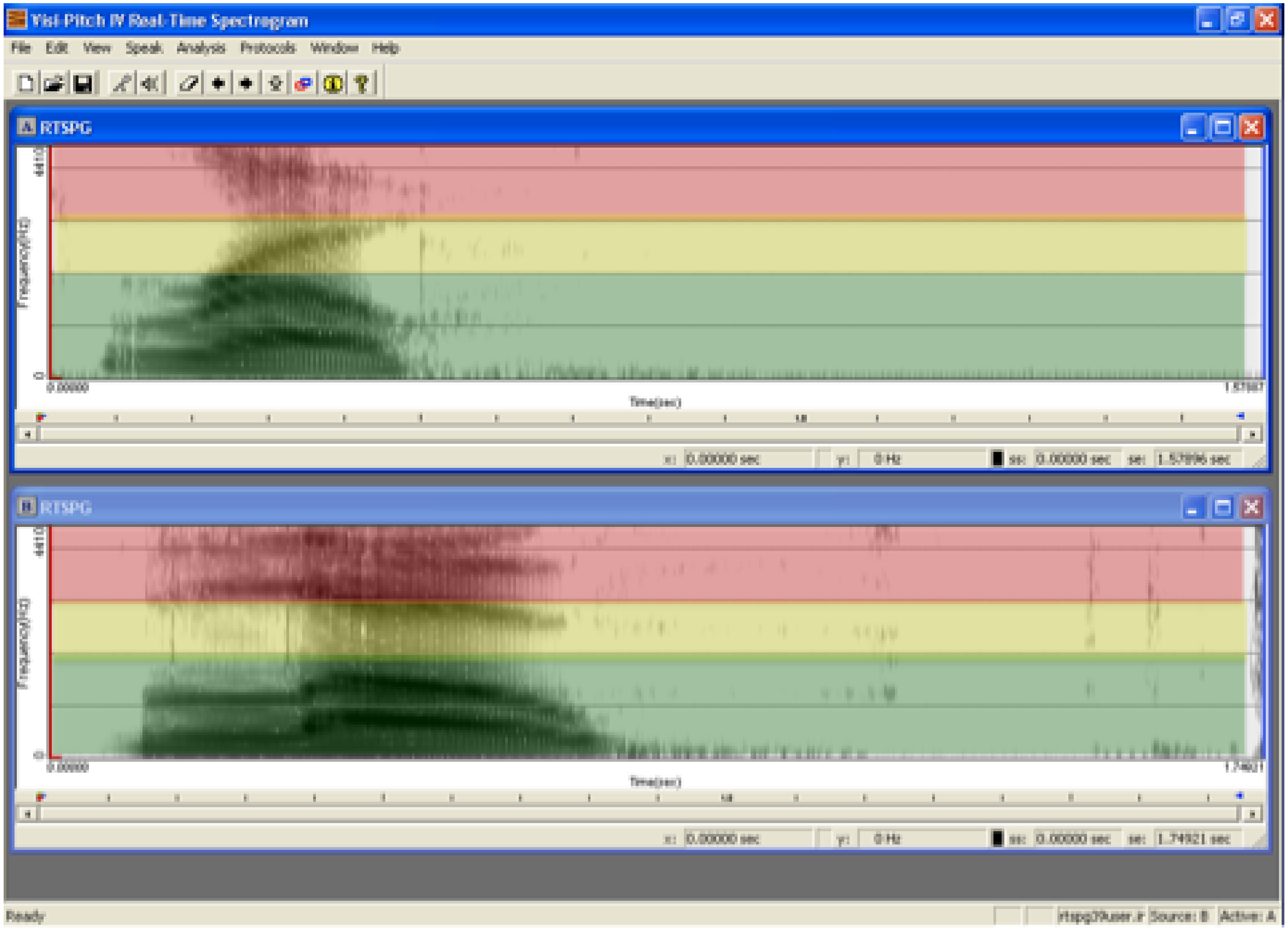
As above, mispronunciation of the word/vowels is allowed if it does not affect the position of the target phoneme.

APPENDIX I
INTRODUCTORY TRAINING MATERIALS

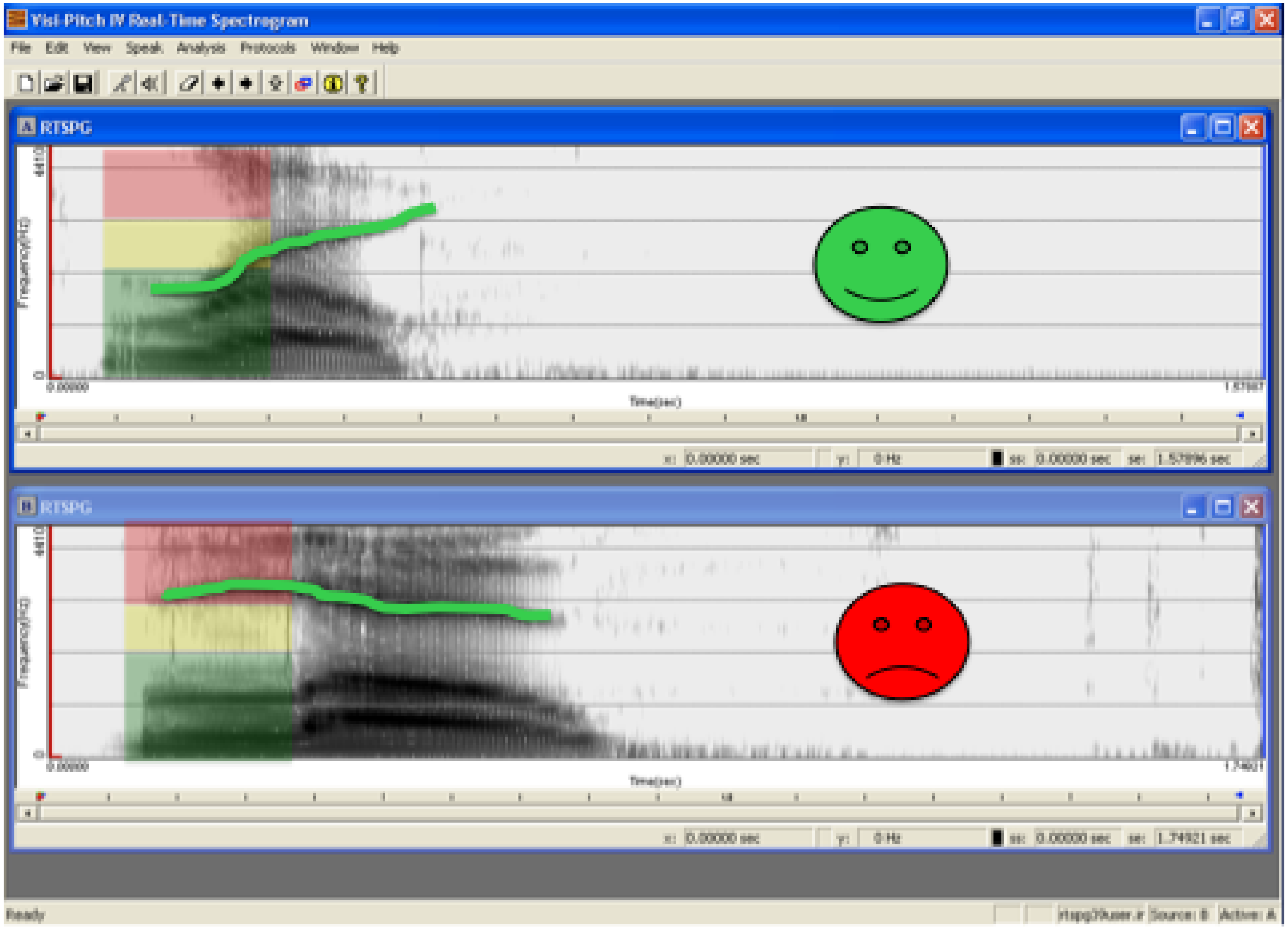

Visi-Pitch IV Real-Time Spectrogram
File Edit View Speak Analysis Protocols Window Help
RTSPG
Frequency(Hz)
Time(sec)
RTSPG
Frequency(Hz)
Time(sec)
Ready

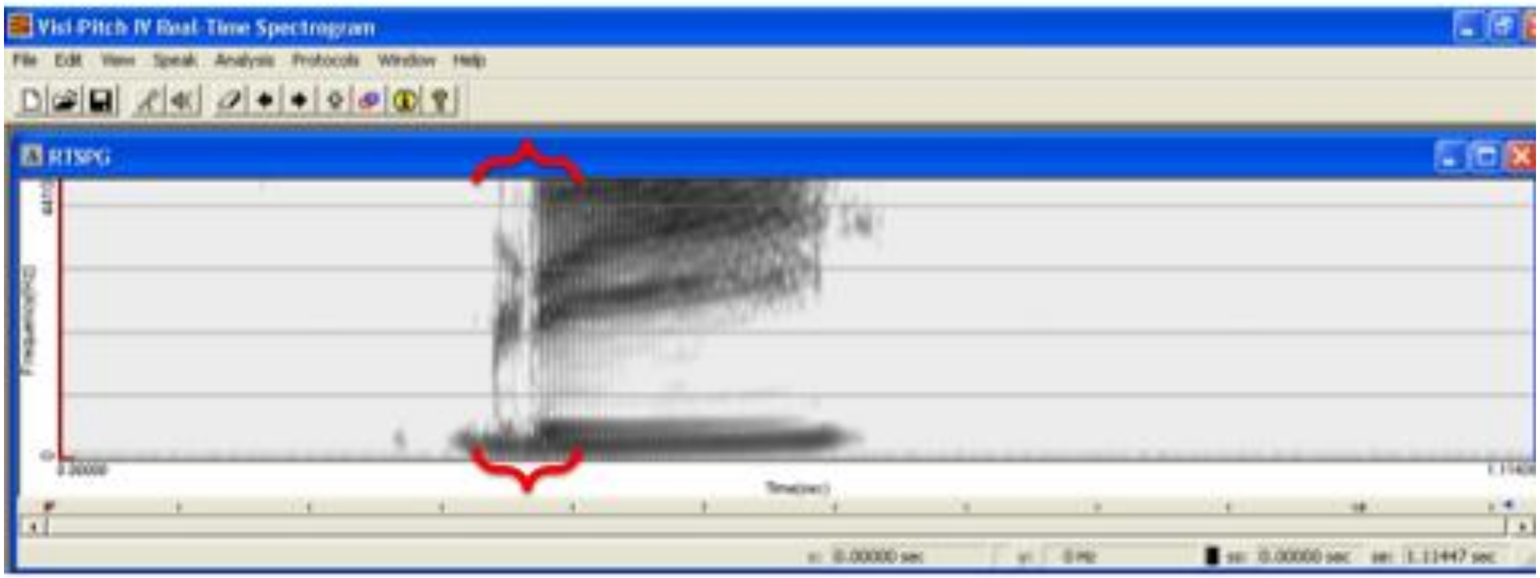


Visi-Pitch IV Real-Time Spectrogram
File Edit View Speak Analysis Protocols Window Help
RTSPG
3
3
2
2
1
1
Time(sec)
Ready



Visi-Pitch IV Real-Time Spectrogram
File Edit View Speak Analysis Protocols Window Help
RTSPG
Frequency(Hz)
Time(sec)
RTSPG
Frequency(Hz)
Time(sec)
Ready


Visi-Pitch IV Real-Time Spectrogram
File Edit View Speak Analysis Protocols Window Help
RTSPG
Frequency(Hz)
Time(sec)
RTSPG
Frequency(Hz)
Time(sec)
Ready


Visi-Pitch IV Real-Time Spectrogram
File Edit View Speak Analysis Protocols Window Help
RTSPG

APPENDIX J
TREATMENT PROGRESSION AND CUEING HIERARCHY

Phase	Stage	Cueing Level	Program	Stimuli	Word Move Criteria	Stage Move Criteria
Phase 1 CV	1	Repetition- listening to recordings or verbal cue	Visipitch	1. /ɹɑ/ (raw) 2. /ɹe/ (ray) 3. /ɹo/ (row) 4. /ɹu/ (rue) 5. /ɹi/ (*ree)	8/10 under 2300 Hz	4/5 words meet word criteria
	2	Reading- written prompt	Visipitch	1. /ɹɑ/ (raw) 2. /ɹe/ (ray) 3. /ɹo/ (row) 4. /ɹu/ (rue) 5. /ɹi/ (*ree)	8/10 under 2300 Hz	4/5 words meet word criteria
	3	Self- assessment: Reading, goodness judgment	PRAAT	1. /ɹɑ/ (raw) 2. /ɹe/ (ray) 3. /ɹo/ (row) 4. /ɹu/ (rue) 5. /ɹi/ (*ree)	Four repetitions	10 total under 2300 with at least one from each word
Phase 2- CVC /m/	1	Repetition- listening to recordings or verbal cue	Visipitch	1. /ɹɑm/ (rom) 2. /ɹem/ (*rame) 3. /ɹom/ (roam) 4. /ɹum/ (room) 5. /ɹim/ (ream)	8/10 under line	4/5 words meet word criteria
	2	Reading- written prompt	Visipitch	1. /ɹɑm/ (rom) 2. /ɹem/ (*rame) 3. /ɹom/ (roam) 4. /ɹum/ (room) 5. /ɹim/ (ream)	8/10 under line	4/5 words meet word criteria
	3	Self- assessment: Reading, goodness judgment	PRAAT	1. /ɹɑm/ (rom) 2. /ɹem/ (*rame) 3. /ɹom/ (roam) 4. /ɹum/ (room) 5. /ɹim/ (ream)	8/10 total, at least one from each category	10 total under 2300 with at least one from each word
Phase 3- CVC /b/	1	Repetition- listening to recordings or verbal cue	Visipitch	1. /ɹɑb/ (rob) 2. /ɹeb/ (*rabe) 3. /ɹob/ (robe) 4. /ɹub/ (rube) 5. /ɹib/ (*reeb)	8/10 under line	4/5 words meet word criteria
	2	Reading- written prompt	Visipitch	1. /ɹɑb/ (rob) 2. /ɹeb/ (*rabe) 3. /ɹob/ (robe) 4. /ɹub/ (rube) 5. /ɹib/ (*reeb)	8/10 under line	4/5 words meet word criteria
	3	Self- assessment: Reading, goodness judgment	PRAAT	1. /ɹɑb/ (rob) 2. /ɹeb/ (*rabe) 3. /ɹob/ (robe) 4. /ɹub/ (rube) 5. /ɹib/ (*reeb)	8/10 total, at least one from each category	10 total under 2300 with at least one from each word

LIST OF REFERENCES

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and l2 speech learning: the case of Japanese /r/ and English /l/ and /r/. Journal of Phonetics, 32, 233-250.

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

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*(4), 161-169.

Bradlow, A. R. (2008). *Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast*. (pp. 287-308). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Bradlow, A.R., Pisoni, D.B., Yamada, R.A. & Tohkura, Y. (1997*).* Training Japanese listeners to identify English /r/ and /l/. *Journal of the Acoustical Society of America, 101*(4), 2299-23.

Brett, D. W. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL, 1,* 103-113.

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

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.

Elbert, M., & McReynolds, L. V. (1975). Transfer of /r/ across contexts. *Journal of Speech and Hearing Disorders, 40*, 380-387.

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

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

Gick, B., Bernhardt, B, P. Bacsfalvi P., & I. Wilson, I. (2008) *Ultrasound imaging applications in second language acquisition.* In J. Hansen & M. Zampini (eds.) *Phonology and Second Language Acquisition. Ch. 11,* pp. 309-322. Amsterdam: John Benjamins.

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

Graddol, D. (2006). *English next.* London: British Council.

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

Hagiwara, Robert. (1995). WPP, No. 90: Acoustic Realizations of American /r/ as Produced by Women and Men. UC Los Angeles: Department of Linguistics, UCLA. Retrieved from: http://escholarship.org/uc/item/8779b7gq

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495-522.

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

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*(5), 3267-3278.

Kearns, K.P. (1986). Flexibility of single subject designs. Part II. Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders, 51,* 204-213.

Kent, R. D., & Read, C. (2002). *Acoustic Analysis of Speech* (2nd ed., pp. 179-183). Albany, NY: Singular Thomson Learning.

Knock, T. R., Ballard, K. J., Robin, D.A., & Schmidt, R.A. (2000): Influence of order of stimulus presentation on speech motor learning: A principled approach to treatment for apraxia of speech, *Aphasiology, 14*(5-6), 653-668.

Lambacher, S. (2010): A CALL Tool for Improving Second Language Acquisition of English Consonants by Japanese Learners, *Computer Assisted Language Learning, 12*(2),137-156.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1962). A motor theory of speech perception. *Proceedings of the Speech Communication Seminar,* Stockholm.

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

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

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

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/. J. Slifka J., Manuel, S., & Matthies, M. (Eds.), From Sound to Sense: 50+ Years of Discoveries in Speech Communication. Electronic conference proceedings.

Marian, V., Blumenfield, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940-967.

McAllister Byun, T., & Hitchcock, E. (In press). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. To appear in *American Journal of Speech-Language Pathology.*

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

Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays*. Language Learning & Technology, 13*(2), 29-47.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Preston, J. L., & Seki, A. (2011). Identifying residual speech sound disorders in bilingual children: a Japanese-English case study. *American Journal of Speech-Language Pathology, 20*, 73-85.

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. To appear in *Language Learning.*

Schuster, L. I., Ruscello, D. M., & Smith, K. D. (1992). Evoking [r] using visual biofeedback. *American Journal of Speech-Language Pathology, 1*, 29-34.

Schuster, L. I., Ruscello, D. M., & Toth, A. R. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology, 4*, 37-44.

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

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

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation of American English /ɹ/. *Speech Communication, 26*, 203-226.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics, 52*(8), 1992.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M. L., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NuroImage, 26,* 703-720.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. Journal of the Acoustical Society of America, 123(6), 4466-4481.

BIOGRAPHICAL SKETCH

Iomi Patten is from the San Francisco Bay Area. She received a Bachelor of Arts degree from the Friends World Program at Long Island University, including three years of study abroad. She taught English to adult speakers of other languages for several years before earning a Master of Arts in Speech-Language Pathology in 2012.