VISUAL FEEDBACK FROM ULTRASOUND IN REMEDIATION OF PERSISTENT /r/ ERRORS: CASE STUDIES OF TWO ADOLESCENTS

by

MARCY ADLER-BOCK

B.A. Simon Fraser University, 2002

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(School of Audiology and Speech Sciences)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 2004

THE UNIVERSITY OF BRITISH COLUMBIA FACULTY OF GRADUATE STUDIES

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Marcy Adler-Bock
Name of Author *(please print)*

06/10/2004.
Date (dd/mm/yyyy)

Title of Thesis: Visual Feedback from Ultrasound in Remediation of Persistent /r/ Errors: Case Studies of two Adolescents.

Degree: MSc.

Year: 2004

Department of Audiology & Speech Sciences
The University of British Columbia
Vancouver, BC Canada

ABSTRACT

This study examines the effectiveness of using visual feedback from ultrasound in remediation of persistent /r/ errors. Ultrasound provides the learner and the clinician with a dynamic sagittal or coronal image of the tongue during speech production. The participants in this study were two adolescent boys ages 12 and 14 who had not yet learned to produce an on-target North American /r/ in any context. Both participants had received at least one year of traditional /r/ therapy without improvement. Therapy was provided over 13 one-hour sessions using visual feed back from ultrasound. Initially, the /r/ was broken down into individual motor targets (tip, body, root); these components were then practiced in combination to produce /r/ in isolation, then in syllables, words, and phrases. Post-treatment improvements were analyzed through transcription, acoustic analysis, and tongue shape measurement. Both participants' /r/ productions were rated as having more tokens of on-target /r/ post-treatment. Acoustic results supported these findings with a lowering of the third formant post-treatment. Tongue shape measures indicated that the participants' tongue shapes were more similar to the modeled tongue shape post-treatment. It was concluded that visual feedback from ultrasound is beneficial in remediation of persistent /r/ errors.

TABLE OF CONTENTS

CHAPTER 2: Methodology

CHAPTER 3: Results

TABLES AND FIGURES

ACKNOWLEDGMENTS

I would like to acknowledge Penelope Bacsfalvi, Shaffiq Rahemtulla, and Bosko Radanov, who generously volunteered their time and knowledge to this project. This project would not have been possible without the enthusiasm of the participants, their families, and their interest in exploring speech therapy techniques using ultrasound.

I would also like to extend my gratitude to Barbara Bernhardt, who brought much needed research expertise and clinical experience to this project. This project would also not have been possible without Bryan Gick who not only allowed the use of technical equipment belonging to the Interdisciplinary Speech Research Lab, but provided his invaluable knowledge of articulatory-acoustic relationships and ultrasound analysis. Thank you to Karen Derry who read this work, and made suggestions through the eyes of a clinician.

Finally, I would also like to extend a special thanks to all of those who provided support, and words of encouragement throughout the course of this study, especially my grandmother Bertha Boch. This research was supported by a summer research grant from British Columbia Medical Services Foundation.

CHAPTER 1: A theoretical basis for using ultrasound in speech intervention

Section 1.1 A need for additional methods in /r/ therapy

While most children with phonological impairment show complete normalization, others continue to have difficulty with certain phonemes into adolescence and even adulthood (Ruscello, 1995b). The English /r/ was ranked by school speech-language pathologists as one of the most frequent phonemes that children struggle to learn (Ruscello, 1995a; Ruscello, 1995b; Shuster, Ruscello & Smith, 1992; Janzen & Shriberg, 1977; Shriberg, 1980; Shriberg, Flipsen, Karlsson, & McSweeny, 2001). There is a need for alternate methods of intervention for children with persisting speech sound errors when the traditional methods are not satisfactory (Ruscello, 1995b).

The /r/ is a very complex phone to produce because it requires the tongue to make two independent constrictions. Phones that have high articulatory complexity also require more cognitive resources to learn (Bernhardt & Stemberger, 1998). Motor learning theory advocates the importance of including a cognitive component within any intervention program (Fletcher, 1992). One way to target cognitive change is through augmented feedback from an external source (e.g. electropalatography or EPG, spectrography, and ultrasound). Bernhardt, Gick, Bacsfalvi, and Ashdown (2003) reported significant improvement in /r/ articulation for four hearing impaired adolescents after a 14-week intervention program using visual feedback from EPG and ultrasound. Based on Bernhardt et al.'s (2003) success, the treatment method employed in this study was predicted to promote improvement in /r/ production for two hearing adolescents.

Ultrasound sends out high-frequency sound waves. These sound waves are reflected back to the transducer when they reach a material or tissue of a different density. When used to show speech, the sound waves travel through the soft tissue of the tongue and are reflected back to the transducer when they reach bone or air (Stone, 1997). The ultrasound images the sagittal or coronal surface of the tongue as a white line (figures 1.1 and 1.2) and provides the clinician and the learner with a dynamic image of the tongue shape during speech sound production.

The remainder of this chapter addresses articulatory and acoustic characteristics of /r/, and reasons why /r/ might be a difficult sound to learn. It will also review traditional and alternative intervention programs for /r/ which provide grounds for using ultrasound in remediation of /r/ errors. This study explores the use of visual feedback from ultrasound as a method of intervention for persistent /r/ errors. Two adolescents aged 12 and 14 participated in 14 one-hour treatment sessions. Outcomes were evaluated perceptually, acoustically, and through tongue shape analysis.

Section 1.2 Articulation and acoustics of /r/

1.2.1 Typical articulation and acoustics of /r/

Although there are clear descriptions for two types of /r/, (tip-up retroflexed vs. tip-down bunched) individuals use a variety of tongue shapes for /r/ that fall between these two extremes (Delattre & Freeman, 1968; Guenther, Espy-Wilson, Boyce, Matthies, Zandipour & Perkell, 1999; Westbury, Hashi & Lindstrom 1998).

Speakers generally produce /r/ with three supralaryngeal constrictions (figure 1.1) (Alwan, Narayanan & Haker, 1997; Delattre & Freeman 1968; Gick et al., 2003; Westbury et al., 1998). The first is a labial constriction as the lips protrude. The second is an oral constriction, and involves tongue movement towards the palate. The /r/ is considered tip-up when the tongue tip stretches towards the palate (figure 1.1), or bunched if the tongue body approximates the palate. The third and final constriction is made as the tongue root retracts towards the pharyngeal wall. (Alwan et al., 1997; Delattre & Freeman, 1968). Alwan et al., (1997), Stone and Lundberg (1996), and Gick and Campbell (2003) found that when subjects used their tongue tip to create an anterior oral constriction (tip-up), a posterior mid-line lowering appeared behind the oral constriction (concave shaping). Finally, posterior lateral tongue bracing against the upper molars occurs. Figure 1.2 illustrates a coronal image of the posterior tongue during /r/ production. This image shows the lateral bracing, and the mid-line lowering.

Figure 1.1 Articulation of /r/ as viewed on ultrasound (sagittal section)

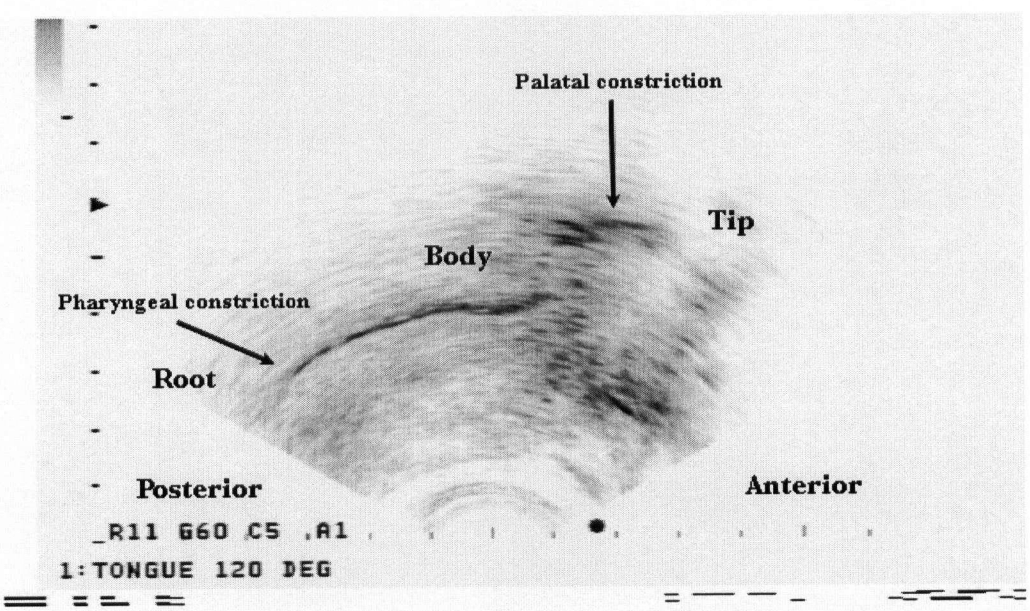
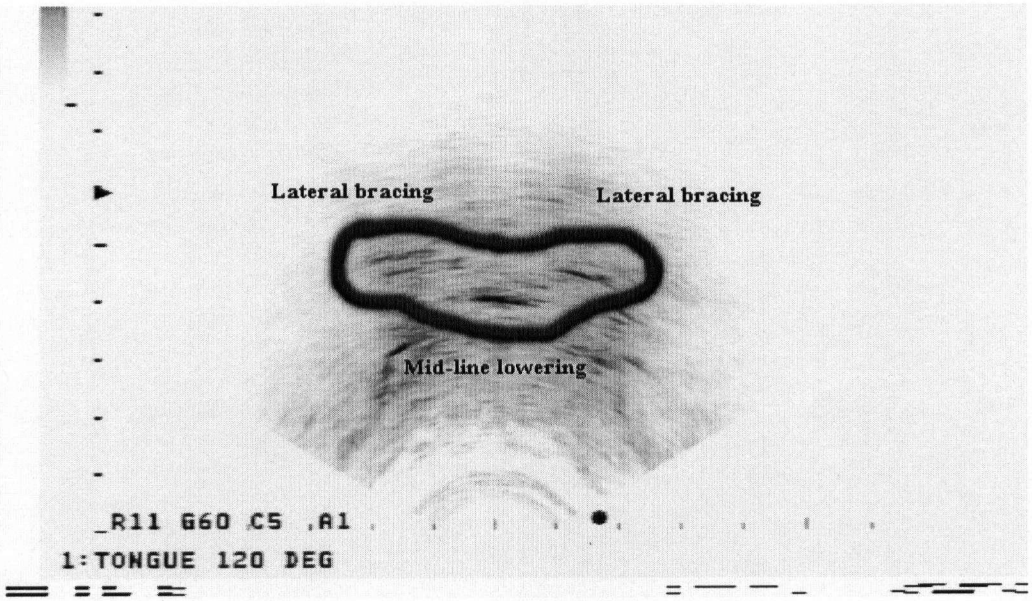


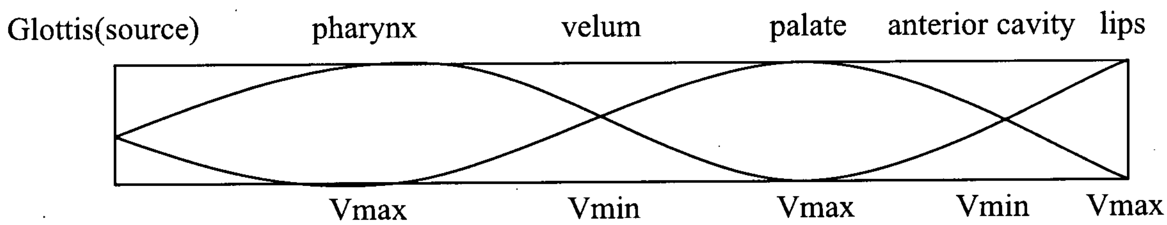
Figure 1.2 Articulation of /r/ (coronal section)



Although the place and degree of each constriction may vary amongst speakers, the resulting acoustic signal is relatively consistent. The acoustic signal for the different articulatory patterns of /r/ all have a similar dropping third formant (Delattre & Freeman 1968; Guenther et al. 1999; Westbury et al. 1998). "F3 is often low enough to approach and/or merge with F2 (Stevens, 1999, cited in Espy-Wilson et al., 2000, p. 344)."

A simple way of identifying the source of F3 lowering is through perturbation theory (Kent & Read, 1992). The basic premise of this theory is that the vocal tract acts as a quarter length resonator. When constrictions are made at points along the tube which have maximum velocity it serves to lower the formant frequency (Vmax for F3 are labeled on figure 1.3). If constrictions are made at points of minimum velocity it serves to raise the formant frequency (Vmin for F3 are labeled on figure 1.3). Points of maximum velocity for F3 are associated with specific anatomical features along the vocal tract (pharynx, palate, and lips). If constrictions are made at any of these points where maximum velocity occurs, F3 will lower. Kent and Read (1992) state that for /r/, constrictions are made at the lips, palate, and pharynx, consequently causing F3 to drop.

Figure 1.3 Model of Vmax and Vmin locations along the vocal tract for F3



According to Espy-Wilson et al. (2000) the simple tube model with constrictions at the pharynx, palate, and lips does not completely serve to explain the low F3 values that occur during /r/ production. They explore additional methods by which F3 may be lowered.

The lowered F3 value in /r/ is also thought to be a result of the resonance of the front cavity/sublingual area anterior to the palatal constriction (Alwan et al., 1997; Espy-Wilson et al., 2000; Guenther et al., 1999). F3 decreases as the length/volume of the anterior cavity increases. See Espy-Wilson et al. (2000) for more detailed information. Delattre and Freeman (1968) reported a correlation between the dip in the tongue dorsum (expansion at the velum), and F3 lowering. In order to achieve maximal F3 lowering during /r/ production, it is important to have constrictions at all three points along the vocal tract of maximum velocity (pharynx, palate, and lips), as well as expansion of the cavity size behind the palatal constriction (at the velum), and in front of the palatal constriction (front /sublingual cavity).

Typical male F3 and F2 values reported for /r/ are 1700 Hz and 1350 Hz respectively (Peterson & Barney, 1952). Flipsen, Shriberg, Karlsson, and McSweeny (2000) reported /r/ formant averages for typical adolescent /r/ productions. For the purposes of this study the first author averaged the F2 and F3 /r/ formant values for eight males ages 12-14 over nine different words (Flipsen et al., 2000). F2 was 1337 Hz and F3 was 1934 Hz. The averaged F3 value from Flipsen et al.'s (2000) adolescent data is higher than Peterson and Barney's (1952) averaged F3 value. This difference of a little more than 200 Hz is likely due to the fact that the vocal tract of adolescent males is shorter than adult males. Consequently, adolescent males should have slightly higher formant values.

Another identifying characteristic of /r/ is the separation between F3 and F2. This F3-F2 gap is small for /r/ compared to vowel sounds and other glides/liquids. Lee, Potamianos, and Narayanan (1999) reported averages and standard deviations (F3-F2) for /ɝ/ in the word 'bird' produced by children and adolescents ages 5-18. Male adolescents age 12 had an average F3-F2 difference of 477 Hz, (SD 160). Male adolescents age 14 had an average F3-F2 difference of 390 Hz, (SD 130).

Section 1.3 Articulation constrained by physiological, structural, and cognitive development

1.3.1 Articulatory complexity and phoneme development

There is a general progression in speech sound acquisition from articulatorily less complex phonemes to more complex ones or from unmarked to marked (Bernhardt & Stemberger, 1998, p.3; Kent, 1992). Kent (1992) analyzes the general order of consonant development within a framework of motoric complexity. He identifies four categories under which the sounds in development can be categorized. Each set is distinct from the others according to the complexity of the motor patterns required to produce the sounds. Some early acquired phones, such as stops, require rapid/ballistic movements of the articulators. These phones are articulated at a rapid duration, with fast acceleration and deceleration rates. It is easier to move the tongue as one unit either anterior, posterior, up, or down. The later appearing sounds in development are related to the child's ability to make finer adjustments in lingual position and shape (e.g. liquids) which are required to create multiple constrictions simultaneously within the vocal tract (e.g. /r/) (Kent, 1992; McGowan et al., (2003). It is not surprising, given the complex lingual control requirements for /r/, that it is one of the last sounds to be acquired during the course of phonological development. Children often do not develop a consistent and acceptable /r/ until the ages of 6;0-8;0 (Kent, 1992; Smit, Hand, Freilinger, Bernthal & Bird, 1990).

1.3.2 Articulatory complexity and cognitive resources

Bernhardt and Stemberger (1998) state that there is a psychological/cognitive reality to the complexity of phonemes (defined by the features of a single phone or sequence of features between two phones). A more complex motor behavior requires increased cognitive resources to learn and carry out the skill. Bernhardt and Stemberger (1998) propose that constraints on sound production are grounded in cognition as well as phonetics. More specifically, "all actions require the use of limited cognitive resources, and some actions require more resources than others (Bernhardt & Stemberger, 1997, p. 219)." Sounds that have high articulatory complexity such as /r/, demand more cognitive resources to learn. An underdeveloped physiological and cognitive system is biased towards sounds that are less complex. The developmental progression of speech sounds not only is influenced by physiological/structural development of the speech organs, but also by the complexity of the motor task, and the amount of cognitive resources required to learn the motor sequence.

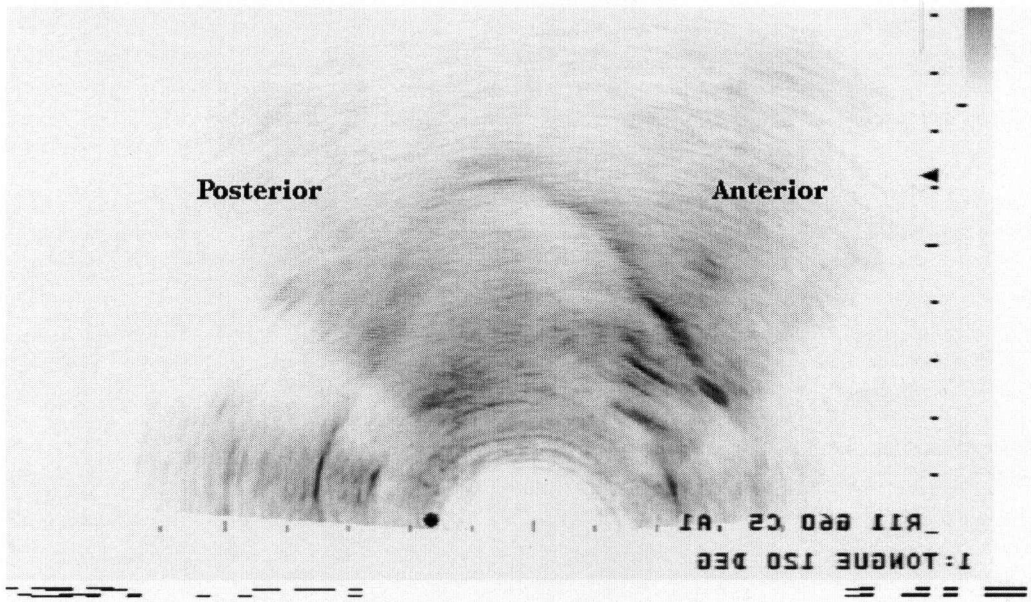
1.3.3 An explanation for residual speech errors

For children just acquiring a sound system both their physiological capacity and cognitive capacity influence their ability to learn and produce new sounds. For adults who have fully developed structural, physiological, and cognitive systems, we must ask why they are still having difficulty producing some phones. One hypothesis is that error patterns acquired at a young age due to motoric and cognitive resource constraints fail to resolve themselves (McGowan et al., 2004; Ruscello, 1995a). The individual's phonological representation for the sound remains the same as it was when their motor and cognitive systems were immature and unable to produce and/or represent the sound correctly. The distortions have been ingrained in their cognitive representation and motor control patterns, and are resistant to change (Shriberg et al., 2001).

1.3.4 /r/ substitution patterns - undifferentiation

When children produce a distorted or a de-rhotacized /r/, their tongues do not achieve the needed constrictions to produce the lowered F3 value for /r/. Shriberg (1980) classifies /r/, /ɝ/, and /ɚ/ substitutions as mid or high back vowels (e.g. /ʊ/, /ɔ/). Bernhardt and Stemberger (1998) state that gliding [w] is the most typical substitution pattern for word-initial /r/ with less typical being [j]. These substitutions are simplifications of the phonetic requirements for /r/. The /r/ requires three places of constriction labial, palatal, and pharyngeal. All of the substitutions, mid or high back vowels, [w], or [j] require only one or two places of constriction. Figure 1.4 illustrates a back vowel substitution for /r/.

Figure 1.4 Back vowel substitution for /r/



Acoustically, these substitutions all have higher F3 values than are expected for /r/. A [w] substitution has an F2 around 800 Hz, and an F3 at 2200 Hz; [j] has an F2 around 2200 Hz, and an F3 at 3000 Hz (Ferrand, 2000). For a high-back vowel substitution [ʊ] F2 lies at 1000 Hz, and F3 at 2250 Hz, and for a mid-back vowel substitution [ɔ], F2 lies at 850 Hz, and F3 at 2400 Hz (Kent, 1992). Shriberg et al. (2001) calculated *z*-scores for (F3-F2) using the /ɝ/ reference data provided by Lee et al. (1999) which were assumed to be representative of typical /ɝ/ production from a group of adolescents. To calculate the *z*-score one must take F3-F2 value from each token, subtract the group mean, and divide by the standard deviation. A *z*-score of 0 would mean that the token was equal to the mean production of the group. Shriberg et al. (2001) reported how far the productions of the participants in their groups fell from the means reported by Lee et al. (1999). One group consisted of adolescent speakers with speech delay plus residual rhotic distortions (group one). The de-rhotacized mean value for this group was 4.78 with a maximum *z*-score of 11.77 for tokens perceived as de-rhotacized /ɝ/ productions. Tokens perceived as correct /ɝ/ produced by group one had a mean *z*-score of 3.07. These scores can be compared to on target productions of /ɝ/ by a group of adolescents with no speech delay or /r/ distortions (group four) *z*(F3-F2)= 0.14. The *z*-score of 0.14 for group four indicates that their /ɝ/ (F3-F2) values lie close to the productions of the adolescents in Lee et al.'s (1999) study. The *z*-scores for group one indicate that both their good /ɝ/ productions and de-rhotacized /ɝ/ productions fall above the mean (F3-F2) reported in Lee et al. (1999).

The above substitutions for /r/ are less phonetically complex than the [ɹ] phone. The tongue appears to move as one entity, creating one constriction along the vocal tract. This is indicative of immature tongue control. Green, Moore, Higashikawa, & Steeve, (2000) stated that "limited independence of anatomically distinct segments is common in immature motor systems." As children learn to control the muscles of the tongue, movement patterns become increasingly differentiated. Differentiation is defined as "increased independence in control of the components involved in a motor task (Green et al., 2000)." Independent tongue movement control is outlined in Gibbon (1999) as the ability of the tongue tip and body to move independently from each other. The root must also be able to move independently from the body and tip achieve the palatal, pharyngeal, and lateral components of /r/.

1.3.5 Muscles required for /r/ production

Although it is unclear from the literature exactly which muscles are involved in /r/ production, we can make predictions based on actions of separate muscle groups. The following discussion is based on retroflexed /r/ production and may differ for bunched /r/.

The first muscle involved is the superior longitudinal muscle running anteroposteriorly along the surface of the tongue. The action of this muscle group raises the tongue tip and the lateral edges forming a concave dorsum as is required in retroflexed /r/ production (Palmer, 1993). The genioglossus is a fan-like muscle group that constitutes most of the medial volume of the tongue. It can be divided into anterior (tongue tip), middle (tongue dorsum), and posterior components (tongue root). When contracted, the anterior portion depresses the tongue tip, the middle group draws the superior surface of the tongue dorsum into a concave shape (Palmer, 1993), and the posterior portion pulls the root of the tongue anteriorly (Dickson & Maue-Dickson, 1982). To produce /r/, an individual must have independent control over the divisions of the genioglossus muscle. They must contract the middle portion of the muscle to pull the dorsum downward while keeping the anterior and posterior portions of the muscle relaxed so as not to depress the tongue tip, or advance the tongue root. Finally, the pharyngeal constrictor muscles may be involved in retracting the tongue root. An /r/ intervention program for residual errors must help learners achieve control over the independent tongue muscles required to produce the differentiated movements for /r/.

Section 1.4 Therapy methods in speech sound remediation

1.4.1 Previous therapy techniques

Traditional speech sound intervention techniques include imitation, contextual identification, shaping, phonetic placement, and moto-kinesthetic training (Ruscello, 1995a; Bernthal & Bankson, 2004). During imitation the clinician provides oral exemplars of the target phone for the client. Contextual identification is a technique whereby the target phone is placed in different phonetic contexts in order that features of a preceding or following phone may facilitate production of the target. For example, in placing /r/ in a /lr/ combination the /l/ facilitates the tongue tip placement for /r/. During shaping, the target phone is broken into component gestures (lips, tongue tip, tongue body) which are then re-combined into the target phone. For example in /r/ production, the learner could independently practice the tongue tip and tongue root components before attempting to put them together. In phonetic placement the articulatory positions for a given target phone are described to the client and even shown through pictures or drawings. Using a moto-elicitation technique the clinician manually manipulates the articulators so that they are in the correct position for target phone production. A speech-language pathologist typically uses some or all of these traditional techniques when teaching /r/. For example, Janzen and Shriberg's (1977) /r/ evocation and generalization techniques include ideas from all five of these traditional methods. Although these techniques help some people learn /r/, they fail to work for others (Ruscello 1995a; Ruscello 1995b; Shuster, Ruscello, & Smith 1992). In such cases, other techniques have been used in combination with traditional therapy methods for eliciting /r/.

Clark, Schwarz, and Blakeley (1993), Shuster et al., (1992), and Shuster, Ruscello, and Toth (1995) all attempted to use different forms of feedback (tactile and visual) to elicit the [ɹ] phone. Clark et al. (1993) used a speech appliance (somewhat like a retainer with a posteriorly placed wedge) that positioned the tongue in the correct shape for the [ɹ] phone. The 36 participants had received a minimum of six months of traditional therapy with no change, and were between the ages of 8-12. The program provided bi-weekly 15-minute sessions for six weeks. The group that used the speech appliance demonstrated significant improvement over the no-appliance group in their /r/ productions. Ruscello (1995a) hypothesized that the speech appliance exposed the subjects to internal tactile, and proprioceptive cues for correct tongue positioning that were not available without the appliance. The sensory cues that the appliance provided created a new awareness of the tongue position needed to produce the target phone.

Another tool used to elicit /r/ has been visual spectrographic feedback. Shuster et al. (1992), and Shuster et al. (1995) used spectrographic feedback to elicit the /r/ phone. Shuster et al. (1992) presented a case study of an adult who was not able to produce /r/ in pre- or post-vocalic positions but could produce it in some consonant clusters. Shuster et al. (1995) presented two case studies of adolescents who could not produce the /r/ phone in any context. Shuster et al. (1992) and Shuster et al. (1995) visually modeled the formants of [ɹ] and allowed the participants to practice the /r/ phone while visually monitoring their own formants to match the model provided. Both studies used contextual facilitation (e.g. /lr/, /ir/, /ɑr/, etc.) to elicit the /r/ in conjunction with spectrographic feedback. The participant in the study by Shuster et al. (1992) learned to produce the /lr/ during the first two sessions; by the fourth session he could produce /r/ in isolation. In the study by Shuster et al. (1995) one participant learned to produce the /r/ with the help of contextual facilitation by the sixth session, and subsequently learned /r/ in isolation by the eleventh session (after spectrographic feedback was discontinued). The other participant learned /r/ in isolation by the third session, and by the tenth session the participant could produce /r/ in words. Both participants continued to receive intervention for /r/ at school to work on generalization. The /r/ appliance (Clark et al., 1993) and spectrograph intervention techniques (Shuster et al., 1992; Shuster et al., 1995) provided the participants with augmented feedback. Similarly, Bernhardt et al. (2003) use feedback from EPG and ultrasound to improved /r/ articulation of four hearing impaired adolescents. The augmented feedback in the above studies brings otherwise unconscious information to a level of conscious control. Although it is not explicitly stated in the above intervention studies, or in traditional therapy techniques, they all have a common grounding in motor learning theory. In motor learning theory, forming a cognitive awareness and gaining conscious control over a new behavior are emphasized as being key components in successful motor learning.

1.4.2 Motor learning and cognitive theory: The theoretical basis for using ultrasound within the context of traditional speech therapy

Fletcher (1992), Ruscello (1993), Ruscello (1984), Ruscello and Shelton (1979), and Schmidt (1982), identify several key components for any program that targets new motor learning. These components are divided into pre-practice and practice. During pre-practice, the goal is for the learner to acquire a mental representation of the target motor behavior (Fletcher, 1992; Schmidt, 1982). A learner must focus on the target motor behavior, and engage in mental rehearsal. During practice the learner uses feedback to monitor and correct his or her productions towards the target behavior. Feedback consists of internal and external information that a learner is exposed to during a speech act. Both provide information that allow the speaker to monitor, verify, and adjust his or her articulatory postures and movements (Fletcher, 1992; Schmidt, 1982). Internal feedback consists of tactile, kinesthetic, and auditory information. External feedback is augmented and provides information about the learner's degree of success through two means: their knowledge of results (KR) and knowledge of performance (KP).

KR is quantitative information from another person or device (yes/no) and KP is qualitative feedback on quality of performance (Fletcher, 1992). For example, a clinician might say "not quite" for KR, and "your tongue was too far forward" for KP. "KR and KP supplement the information that the speaker derives from internal feedback [tactile, kinesthetic, and auditory]" (Fletcher, 1992). External feedback KR and KP support cognitive change in motor planning.

The practice component of therapy varies, depending on the learning stage. In the beginning, the execution and evaluation of the articulation is under conscious control and self-analysis. Here, mental rehearsal, imagery, and evaluation through augmented feedback are key. The new skill is learned with both a mental and a motor component (Ruscello, 1984). This is consistent with Bernhardt & Stemberger's (1998) proposal that learning is grounded in cognition. To change a motor behavior, one must also invoke change of the underlying cognitive representation. After this 'cognitive' stage, evaluation is handed over primarily to internal feedback and automatization and the phone is practiced within different linguistic contexts (isolation, syllables, words, phrases) (Ruscello, 1984; Shuster & Ruscello, 1992). The initial stages of an intervention program are cognitively focused and strive to develop a deep conscious awareness of articulatory positioning for the sound. The later stages strive for generalization of the sound.

Based on the previous studies that incorporate visual and tactile augmented feedback into their treatment programs (Bernhardt et al., 2003; Clark et al., 1993; Shuster et al., 1992;

Shuster et al, 1995) there is reason to believe visual feedback from ultrasound will work to teach participants the /r/ phone. The ultrasound can be used to achieve conscious control and cognitive awareness of the sound production during the initial stages of motor learning. The ultrasound enables us to break the target behavior into isolated components (tongue tip raising, midline lowering, lateral bracing, root retraction), in order to provide KR and KP for each of these. Once the sound is consistent using visual feedback the learner can then use his or her own internal feedback systems (tactile, kinesthetic, auditory) to monitor his or her productions.

Section 1.5 Predicted post-treatment changes

The current study evaluated outcomes of incorporating visual feedback from ultrasound into a traditional speech therapy setting to facilitate /r/ production. It was predicted that after a block of 13 treatment sessions[1] (Bernhardt et al., 2003) the participants would be able to produce the /r/ phone in isolation, and be able to practice and generalize the phone into words, phrases, and conversation.

Based on the cognitive aspects of the motor learning approach, the participants were predicted to demonstrate knowledge gains for /r/ in being able to explicitly state the tongue shape requirements for /r/ production. Perceptually, the participants' /r/ in words and phrases was expected to be transcribed by a trained listener with more rhotic quality post-treatment. Acoustically, it was expected that the participants' overall F3 values for /r/ would be lower in frequency post-treatment. This F3 value would be comparable to Peterson and Barney (1952) and Flipsen et al.'s (2000) averaged F3 values. Additionally, *z*-scores for /ɝ/

productions in 'her' were predicted to be similar to those scores Shriberg et al. (2001) reported as on target /ɝ/ productions by the group who had prior speech delay (group one). In order to maximize F3 lowering, the participants would learn to match their sagittal tongue shape to resemble the model tongue shape (figure 1.1) creating two points of constriction (palatal & pharyngeal), and two expansions (dorsum lowering & front/sublingual cavity expansion). Based on typical substitution patterns of vowels such as /ʊ/, /ɔ/, or glide /w/ (Bernhardt & Stemberger, 1998; Shriberg, 1980) the following articulatory changes would be expected:

If pre-treatment substitutions had a high component but no tongue root retraction such as in /ʊ/ or /w/ where the main constriction is uvular, then the expected changes post-treatment would be:

(a) The tongue tip would increase in height as it is lifted towards the palate creating the anterior constriction. As a result the size of the sublingual/anterior cavities should increase.

(b) The tongue root would retract towards the pharyngeal wall to form the posterior constriction.

(c) Tongue body lowering would occur as a result of the tongue tip and tongue root stretching, there by creating an expansion at the velum.

However, if the pre-treatment substitution was a vowel such as /ɔ/ which already has tongue root retraction then the expected changes post-treatment would be:

---

[1]Bernhardt et al., (2003) reported significant change in participants' speech production after 14 sessions.

(a) The tongue tip would increase in height as it was lifted towards the palate creating the anterior constriction. As a result the size of the sublingual/anterior cavities should increase.

(b) The tongue root would not retract towards the pharyngeal wall, as there is already pharyngeal constriction.

(c) Tongue body lowering would occur as a result of the tongue tip and tongue root stretching, there by creating an expansion at the velum.

CHAPTER 2: Methodology

Section 2.1 Participants

The participants in this study were two adolescent males. Both were referred to the study after receiving /r/ speech intervention through traditional means with negligible improvement. VF was 14, and ML was 12. Both spoke English as their only language. Previous audiology reports indicated that both ML and VF have normal hearing.

Both participants had typical gross motor skills and excelled at athletics; VF was a ski racer, and ML was a swimmer. VF and ML came from mid-SES families and both sets of parents had university degrees. In an oral motor assessment both ML and VF could produce alternating motion rates ('pa', 'ta', 'ka') within typical limits (see table 2.1) but demonstrated difficulty with initially sequencing speech gestures for sequential motion rates (SMRs) ('pataka') (Kent, Kent, & Rosenbek, 1987). Both participants produced SMRs with a typical number of syllables per second, but their initial productions were not in the correct order. ML and VF produced the sequences of sounds incorrectly for the first several SMRs (e.g. 'papaka,' 'pututka'). Also noted during the oral motor assessments was that both ML and VF benefited from using a mirror for visual feedback for tasks such as raising their tongue tips, or tongue lateralization.

Table 2.1 ML's and VF's AMR and SMR rates compared with typical values summarized in Kent et al. (1987)

Typical syll/sec	/pʌ/ 6.3*, 5.0†	/tʌ/ 6.2*, 4.8†	/kʌ/ 5.8*, 4.4†	/pʌtʌkʌ/ 5.0*, 3.6†
ML	6.0	5.6	5.4	4.8
VF	5.2	5.2	5.6	4.2

* Median values reported in Kent et al. (1987).
† Minimum values reported in Kent et al. (1987).

Each participant has received speech language services since they were two (ML) and three (VF) years of age targeting various speech sounds. The /r/ articulation was the last phone they needed to acquire. Previous speech-language therapy reports indicated that VF received two years of /r/ intervention from school services, and ML received one year of /r/ private speech intervention. VF's speech has been labeled as dysarthric and dyspraxic, and several years ago he participated in the Beckman Oral Motor Program (Beckman, 1975). In addition to speech sound distortions both participants had histories of phonological awareness, reading, and writing difficulties, and at the time of the study received extra support for learning. ML and VF both had trouble holding auditory information in sequence, leading to confusion when attempting to repeat words with a larger number of syllables. According to an audiology report dated April 2003, VF also had difficulty with word discrimination in the context of background noise. VF had older twin brothers who also have a long history of reading and writing difficulties.

Section 2.2 Apparatus and Stimuli

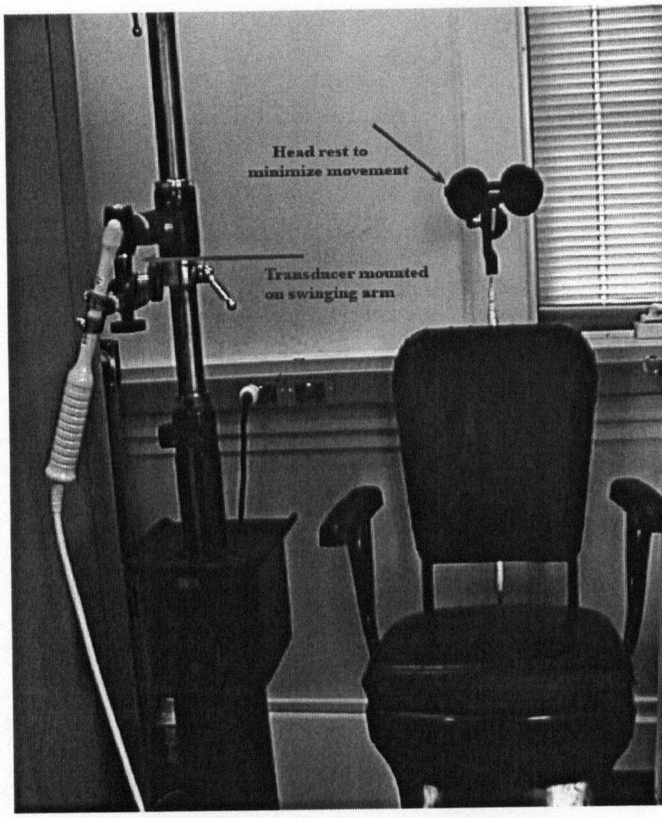
2.2.1 Apparatus and set up for data collection

Audio-recordings were taken using a TASCAM 202MK111 recorder during the pre- and post-treatment sessions. Dual channel microphones (Shure 5M58, and Beyerdynamic TGX58) were placed six inches from the participant's mouth.

Ultrasound recordings were taken at pre- and post-treatment sessions using a stationary Aloka Pro-Sound SSD-5000 ultrasound with a 6 MHz transducer series M00196. The audio-signal was captured using a Pro-Sound YU34 unidirectional microphone. Both the audio signal and the ultrasound image were recorded onto digital videotape at a rate of 30 frames per second with a JVC Super VHS ET Professional recorder. Head cups on the assessment chair stabilized the participant's head, and the transducer was held firmly under the participant's chin with an extension arm attached to the chair (figure 2.1). Data were collected displaying the mid-sagittal section of the tongue as is illustrated in figure 1.1. Note that the pre- and post-treatment assessments were completed without the participants viewing their tongue shape on ultrasound. Coronal data were not collected due to the inability to be consistent in coronal probe placement for /r/.

Software used to analyze the acoustic data and ultrasound images were Final Cut Express 2.0, Adobe Photoshop, and Praat 4.0.49. The ultrasound recordings were digitized using Final Cut Express, and still images of the /r/ productions were exported to Adobe Photoshop for measurement. Sound files were exported to Praat for acoustic analysis.

Figure 2.1 Ultrasound assessment chair



2.2.2 Apparatus and set up for /r/ intervention

A Sonosite 180 Plus portable ultrasound machine with a C15/4-2 Mhz MCX transducer probe was used for therapy purposes in addition to the stationary Aloka ultrasound mentioned above. Coronal and mid-sagittal images were viewed during therapy sessions (figures 1.1 and 1.2). The viewing screen was placed at eye level in front of the participant and the clinician. The transducer was hand-held in treatment and could be used by both the clinician and the participant.

2.2.3 Stimuli design

ML's baseline data were collected using the standard /r/ wordlist. VF's baseline data were collected using stimuli words from CAPES (Masterson & Bernhardt, 2001). VF's baseline data are from the CAPES wordlist because at the time of baseline data collection, it was not known that VF would be participating in this study.

1. The standard word list for ultrasound elicitation included words with /r/ in different syllable positions and phonetic contexts (Appendix A). A total of 29 /r/ words were on the list. Words were read in the carrier phrase "say ___ again" for initial, consonant cluster, and medial /r/ words, and "say ___ day" for final /r/ words. The carrier phrase was changed for word-final /r/ because there was concern that the participants' vowel substitutions for /r/ would be difficult to differentiate from the initial schwa in 'again.' The stop in 'day' made a clear cut off point for the word-final /r/.
2. An /r/ perceptual discrimination tape was created. This tape consisted of 25 tokens of /r/ in different syllable positions and phonetic contexts randomly selected from the /r/ stimuli word list and audio-recorded by the first author. The /r/ in each word was produced as either (a) an on-target /r/, (b) an /r/ distortion, or (b) a vowel/glide substitution.

Section 2.3 Data collection

2.3.1 Time-line and design

The study followed case study design consisting of traditional treatment followed by a no-treatment baseline with 13 subsequent sessions of intervention using ultrasound. The therapy blocks began with one session using only traditional elicitation techniques before ultrasound was introduced. This was to ensure that the participants were not readily stimulable for /r/.

The participants received different intervention schedules (intensive vs. distributed). This was because VF lived several hours away from the treatment site, and could only come on some weekends; ML lived only minutes away and could come for therapy several times per week.

Table 2.2 Time-line for /r/ intervention with VF and ML

VF	Two years of traditional /r/ therapy through school S-LP	No-treatment baseline (five months), then pre-treatment assessment	Distributed therapy: 14 hours of intervention over a five-month period	Post-treatment assessment
ML	One year of /r/ therapy provided by a private S-LP	No-treatment baseline, (two months) then pre-treatment assessment	Intensive therapy schedule, 14 hours over a one and a half month period	Post-treatment assessment

2.3.2 Data collection procedure

1. During the baseline assessment each participant completed a CAPES phonological assessment (Masterson & Bernhardt, 2001).
2. The pre-treatment assessment consisted of (a) a standard oral motor exam; (b) a case history; (c) a CAPES assessment to determine any other phonological patterns in the participants' speech; (d) an /r/ discrimination task (see above); (e) oral reading of the /r/ word list for an audio-recording (once each); (f) oral reading of the /r/ word list for an ultrasound recording (ten times per word) (g) a connected speech sample.
3. During the post-treatment assessment the participants read (a) the /r/word list for an audio recording (one time each); (b) the /r/ word list for an ultrasound recording (ten times per word) (c) a connected speech sample. Additionally, single-word samples were recorded without the carrier phrase.

Section 2.4 Therapy procedure for /r/ intervention

2.4.1 Traditional elicitation

Traditional elicitation techniques as outlined in table 2.3 were used during the first session to identify if the participants were stimulable for /r/ without visual feedback from ultrasound. An /r/ elicitation program similar to Shriberg's (1975) was also used during the first therapy session.

Table 2.3 Traditional /r/ intervention techniques used with VF and ML

Auditory imitation	Phonetic placement	Visual feedback	Contextual facilitation	Shaping
√ - The /r/ phone was modeled and the child was asked to repeat the sound.	√- Verbal description and depiction of where the tongue is placed for the /r/ sound was provided	√ - Used mirrors to view the lips and anterior vocal tract.	√ - The /r/ target sound was placed in different phonetic contexts /i/, /ɑ/, /k/, /t/. Certain features of the preceding or following phoneme may facilitate production.	√ - The /r/ was shaped from a different sound. For example, elicit /r/ from /l/ as in Shriberg's (1975) elicitation program.

2.4.2 Therapy sessions

The sessions consisted of the clinician and the participant sharing the ultrasound. The transducer was held under the chin to display either a sagittal or coronal view of the tongue. The sagittal view provided an image like the one illustrated in figure 1.1. This view was useful in identifying height and backing of the tongue tip, body, and root. The coronal view provided a cross-sectional image of the tongue and helped for viewing the lateral bracing of the tongue and the mid-line groove (figure 1.2). Markers were set on the ultrasound display to provide the participant with reference points and targets to reach when practicing activities such as raising the tongue tip. The /r/ was taught through a hierarchy of steps starting from learning the components of /r/ in isolation without phonation to using the phone in words and phrases (see Appendix B for more details). The therapy sessions included the following goals.

1. Knowledge goals, awareness of the /r/ tongue shape: The participant was oriented to the ultrasound image, and the /r/ target. This was accomplished through discussion of the /r/ components, and modeling and sketching the tongue in /r/ position.

2. Motor and production goals, establishing the components for /r/: After the target components were identified, the participants used visual feedback to practice each component in isolation, and then in combination. Contextual facilitation was also used to elicit /r/ production.

At the end of each session the participants were given activities to practice for ten minutes at home at the level of success during the therapy session.

Section 2.5 Data analyses

2.5.1 Transcription analyses procedure

Data were phonetically transcribed for each participant at baseline, pre-treatment, and post-treatment. The /r/ was narrowly transcribed and categorized as:

1. A complete substitution (usually a vowel) (VS)
2. A vocalic substitution with some rhotic quality (RQ)
3. An on-target /r/ ([ɹ])

All of the post-treatment data were transcribed twice by the first author within an interval of several weeks between transcriptions, and 20% of the data were transcribed by another speech-language pathologist. For the first author's transcriptions, if there was a disagreement between two transcriptions, the less /r/-like transcription was selected. For example, a VS would be chosen over a RQ transcription. In addition to transcribing, from each set of ten repetitions per word, the best and worst productions were coded. Only the best and the worst tokens that matched across the two transcriptions were used in the data analysis comparison for best and worst tokens.

Of the 275 transcribed post-treatment /r/ tokens by the first author for VF, 245 tokens matched between the two transcriptions. Of the 30 that were non-matching, all were off by a single step, meaning VS and RQ were interchanged, and RQ and on-target /r/ were interchanged. Overall, there was 89% agreement on the transcribed tokens for VF. A similar trend occurred for ML where 40 out of 285 first author's transcribed post-treatment /r/ tokens were non-matching. All of the non-matching tokens were off by a single step. Overall there was 86% agreement on transcribed tokens for ML.

There was 80.1% agreement between the two transcribers. Of the 14 non-matching post-treatment tokens, the other speech-language pathologist transcribed all but three with more /r/ quality than the first author's transcriptions. All non-matching tokens were off by a single step.

2.5.2 Ultrasound analyses procedure

After articulatory ultrasound data were captured in Final Cut Express, still frames were extracted at 'max /r/' for each repetition. Max /r/ was defined perceptually and visually when the tongue reached the point of maximum /r/ for each token. For both participants, max /r/ was reached as tongue passed through the point of maximum height and backness.

In Adobe Photoshop, the still /r/ tongue shapes were measured at several different points in order to capture change quantitatively. Height and distance from centre of the probe were measured at (a) tongue root (R), (b) max tongue body height (B), and (c) tongue tip (T). These points of measure were selected to maximally capture changes in tongue shape towards the target /r/ as therapy focused on (a) raising tongue tip, (b) lowering the body of tongue, and (c) tongue root retraction. Tokens with unclear points of measurement were discarded. Overall, 34 images were discarded from ML's data, and 20 from VF's. Measure points of pre- and post-treatment /r/ are illustrated in figures 2.2 and 2.3.

Figure 2.2 Locations of tongue measurement for /r/ pre-treatment

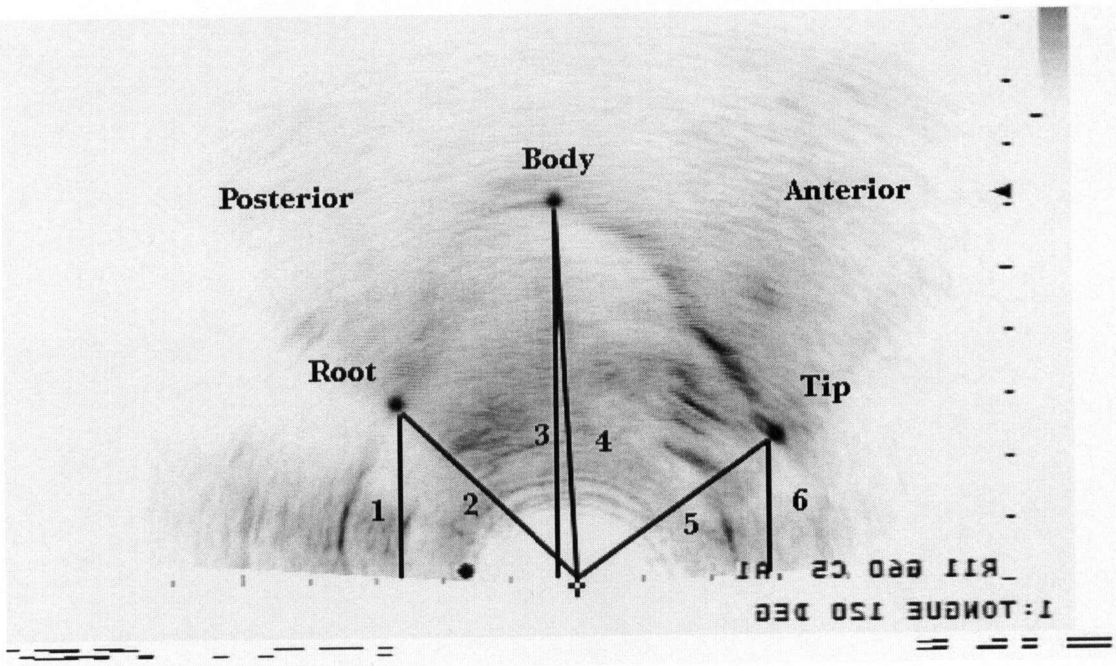
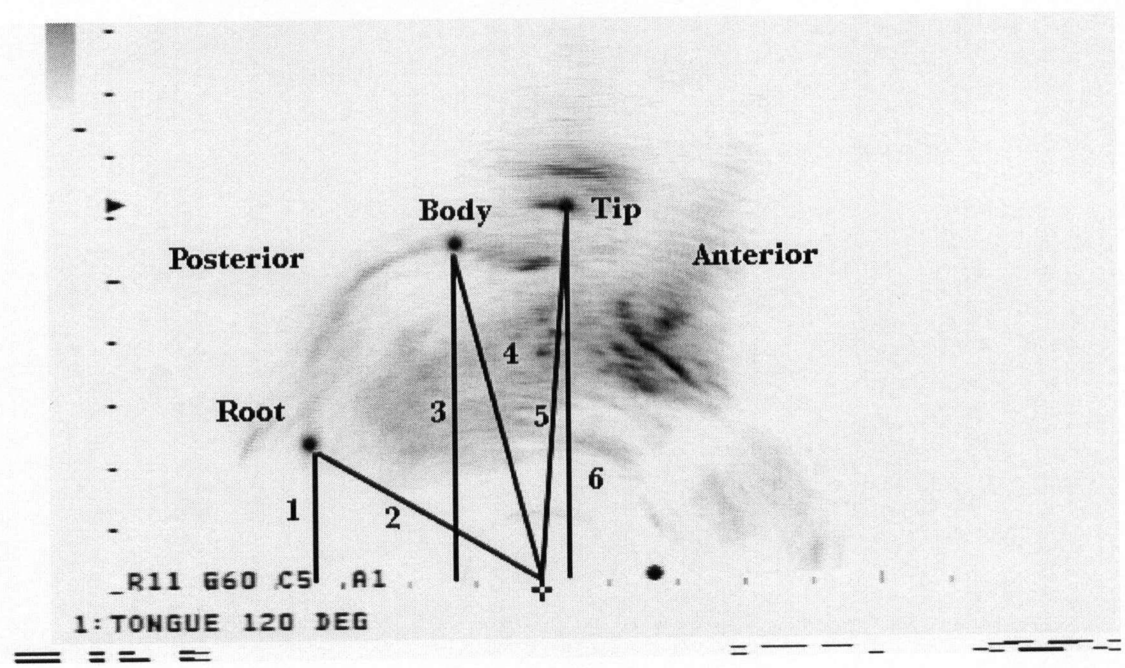


Figure 2.3 Locations of tongue measurement for /r/ post-treatment



1 = Height root (HR)
2 = Distance root (DR)
3 = Height body (HB)
4 = Distance body (DB)
5 = Distance tip (DT)
6 = Height tip (HT)

After the ultrasound images were measured, a translation (vertical and horizontal) and rotation of pre-treatment measurements was completed to correct for any difference in transducer positioning between the pre- and post-treatment sessions.

Inter-speech rest positions were used to correct for the pre- and post-treatment differences in transducer placement. Inter-speech rest position is a stable consistent posture within a speaker that occurs just before the onset of speech (Gick, Wilson, Koch & Cook, 2004). This inter-speech rest position was captured between word repetitions (e.g. "say ____ again" rest position "say ____ again"). Twenty tokens of the tongue at inter-speech rest position were taken from pre- and post-treatment tapes for both participants. Means and standard deviations of these measures are displayed in tables 2.4 and 2.5. The goal was to match pre- and post-treatment inter-speech resting position images through vertical and horizontal transposition, and angle rotation (see Appendix C for more details). The pre-treatment inter-speech resting position was matched to the post-treatment inter-speech resting position. For VF this required shifting the pre-treatment tip and body measures .8 mm along the vertical axis, and -16.9 mm along the horizontal axis with 10.04 degrees of upward rotation from the fixed root point. For ML this required translation of -20.32 mm along the horizontal axis, and -11.59 mm along the vertical axis, followed by an upward rotation of 18.3 degrees. These same calculations were applied to ML's and VF's pre-treatment /r/ data.

Note that these adjustments do not factor out extraneous head movement during data collection; they only adjust for differences in static transducer placement.

Table 2.4 Height and distance measures at inter-speech resting position ML

ML	DT	DB	DR	HT	HB	HR
Pre Mean	72.60	71.51	46.82	25.04	62.70	41.48
Std. Deviation	3.84	3.69	3.87	4.74	3.14	3.87
Post Mean	65.22	68.74	51.57	42.64	68.12	29.89
Std. Deviation	2.47	2.06	3.83	3.13	2.21	3.28

Table 2.5 Height and distance measures at inter-speech resting position VF

VF	DT	DB	DR	HT	HB	HR
Pre Mean	68.18	68.28	45.70	31.15	62.55	30.04
Std. Deviation	1.32	2.92	2.13	2.55	2.30	2.86
Post Mean	67.09	74.21	59.94	49.58	73.62	30.80
Std. Deviation	2.13	2.02	4.22	4.61	2.10	4.54

A research assistant was given the video ultrasound recordings and was asked to define the tongue at max /r/ for 7% of the tokens. He then marked and measured the tip, body, and root measures on pre- and post-treatment data. The correlations between the first author's measures and the research assistant's measures ranged from r = .954-.980. This indicates that the procedure of extracting max /r/ from the video, and marking the T, B, and R points along the tongue surface were sufficiently similar across experiments.

2.5.3 Acoustic analyses procedure

Sound files were extracted from the ultrasound video and formant values were analyzed with Praat 4.0.49. As stated in the introduction, the dropping F3 towards F2 is the most prominent acoustic feature of /r/. Due to the nature of the participants' speech, a low F3 point was not always present on the spectrograms. One acoustic cue that was consistent for both participants' pre- and post-treatment attempted /r/s was a fall in F2. This fall in F2 corresponded with (a) a rise in F3, (b) a steady F3, or (c) a dropping F3. F2 was used to guide the selection of /r/ midpoint (McGowan et al., 2003). Due to the differences in the participants' speech samples, the analysis procedure had to be modified for each participant's speech.

1. For VF, when /r/ was in word-initial position, measures were taken at F2 minimum. One problem that arose was that the second formant during /r/ was often long and steady with no obvious minimum point. Minimum F2 was found

by selecting the entire low steady F2 and measuring at 50% (figure 2.4). For ML, word-initial /r/ measures were taken just after the onset of /r/ phonation. ML's speech contained a pause between the /eI/ of 'say' and the onset of /r/ (figure 2.5).

2. When /r/ was word-medial, measurements were made in the same manner as for VF's word-initial /r/.
3. When /r/ was in word-final position, measurements were made before the closure of the /d/ for 'day,' where F2 and F3 minima were visually observed.
4. When /r/ was in word-initial consonant clusters, measurements were made after the initial consonant at F2 and F3 minima. Words with initial voiceless stops (kr, tr, pr) were eliminated from acoustic analysis because the aspiration occluded the formants for /r/.

Figure 2.4 VF /r/ acoustic measures post-treatment: word initial /r/

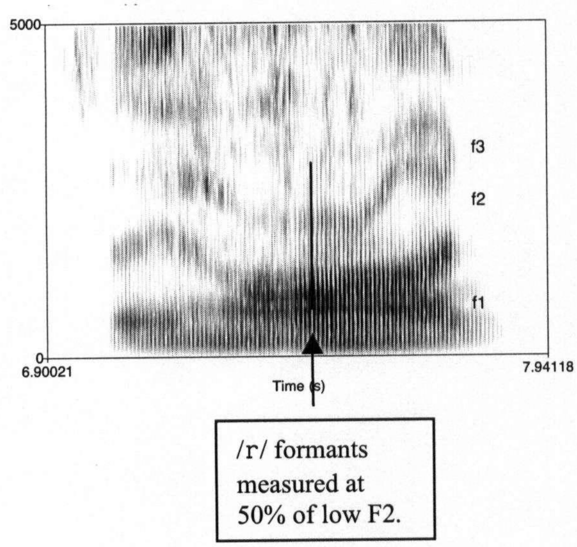
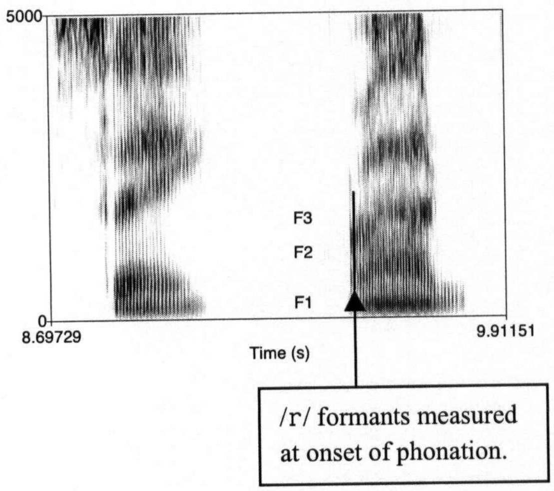


Figure 2.5 ML acoustic measures post-treatment: word initial /r/



Tokens with unclear /r/ formants were excluded from analysis. Each point of measure was hand marked, and measures were automatically extracted. For VF, 218/247 pre-treatment tokens, and 229/246 post-treatment tokens were measured. For ML, 144/254 pre-treatment tokens, and 232/257 post-treatment tokens were measured. ML's pre-treatment number is low because the word-initial tokens (n = 100) could not be measured due to their fricative quality.

CHAPTER 3: Results

This chapter identifies pre- to post-treatment changes in the participants' /r/ knowledge and performance in terms of: (a) knowledge goals, (b) transcriptions, (c) acoustic analyses, and (d) measurements of tongue shape.

Section 3.1 Traditional elicitation techniques and perceptual discrimination

Prior to introduction of the ultrasound, several traditional /r/ elicitation methods were attempted. Neither participant could produce /r/ with techniques listed in table 2.3.

ML and VF demonstrated that they could perceptually differentiate between a good /r/ and a de-rhotacized /r/ production. ML scored 25/25 on the perceptual discrimination task, and VF scored 24/25.

Section 3.2 Knowledge goals

When initially asked what they knew about /r/ tongue shape, neither participant was aware of the posterior lateral bracing, mid-line lowering, or the tongue root retraction for /r/. ML could explicitly talk about the components of /r/ after the first two sessions. VF could do this after four sessions. In the final sessions, both VF and ML were asked to instruct their parent or teachers how to produce /r/. Both participants could clearly tell the listener what the tongue shape should look like for /r/, and identify the important components on the ultrasound monitor.

Section 3.3 Trained listener transcription results

3.3.1 Transcriptions of /r/ word list stimuli

Transcription symbols are as follows:

[̜] = unrounded or delabialized
1. [ʊ̜] = high back unrounded vowel
2. [əɔ̜] = mid central-back unrounded vowel
3. [l] = liquid lateral
4. [β] = voiced labial fricative
5. [w̜ʰ] = delabialized glide with excess aspiration
6. [w̜] = delabialized glide

Tables 3.1 and 3.2 contain baseline, and pre- and post-treatment /r/ transcriptions for VF and ML. There was little change in the participants' /r/ production during the baseline period (baseline assessment → pre-treatment assessment). At baseline and pre-treatment assessment VF used vowel or glide substitutions for /r/ in all word positions (e.g. rid → [ʊ̜id], her→ [həɔ̜], story→ [stɔi], and tray→[tw̜eɪ]) At the time of ML's baseline and pre-treatment assessments his word initial (WI), word final (WF), and word medial (WM) /r/s had no rhotic quality (RQ). He used vowel substitutions (VS) in place of WM and WF /r/ (e.g. ear→ [iəɔ̜], hairy → [hɛəɔ̜i]) and a bilabial fricative in place of WI /r/ (e.g. row→ [βow]). ML's /r/ in clusters was noted to have some RQ (4/9) at baseline. This was comparable with his performance at the pre-treatment assessment where 5/9 clusters were perceived to have RQ.

The second and third columns (pre-treatment assessment → post-treatment assessment) in tables 3.1 and 3.2 indicated that there was improvement for /r/ in all word positions during the treatment period for both participants. VF improved the most when /r/ was in WI, WF, and WM positions. He made only one vowel substitution for /r/ in WI position, two for /r/ in WM position and none for /r/ in WF position. The /r/ in clusters was still a challenge for VF; he made vowel or glide substitutions (VS) 44% of the time (e.g. gray→ [gw̜eɪ]). Overall, at the post-treatment assessment 76% of VF's attempted /r/ words contained either an on-target /r/, or RQ /r/. This was an improvement from his pre-treatment assessment where all of his productions were substitutions without rhotic quality.

ML made progress in all categories. At pre-treatment assessment he produced RQ /r/s in only clusters (17% of all /r/ words); at the post-treatment assessment he produced 93% of all /r/ stimuli words with on-target /r/ or RQ. ML had the most difficulty with WF /r/ (e.g. air→ [ɛəɔ̜]). At the post-treatment assessment ML was asked to try his WF /r/s without the carrier phrase "say ___ day." In single words ML produced all WF tokens with an on-target /r/ (24/24). Typical substitutions that VF and ML used for /r/ in different word positions are illustrated in tables 3.1 and 3.2.

Table 3.1 /r/ accuracy VF
Baseline single word assessment, pre-treatment assessment of /r/ words in standard phrases, and post-treatment assessment /r/ words in standard phrases.

VF	Baseline			Pre-treatment sample			Post-treatment sample		
	RQ	[ɹ]	typical substitutions	RQ	[ɹ]	typical substitutions	RQ	[ɹ]	typical substitutions
WI	0/5	0/5	[ʊ̜]	0/10	0/10	[ʊ̜]	5/10	4/10	[ʊ̜]
WF	0/10	0/10	[əɔ̜]	0/6	0/6	[əɔ̜]	3/6	3/6	
WM	-	-	-*	0/4	0/4	[əɔ̜], [l]	2/4	1/4	[əɔ̜]
#Cr	2/20	0/20	[w̜], [ʊ̜]	0/9	0/9	[w̜], [ʊ̜]	3/9	2/9	[w̜], [ʊ̜]

*No WM samples of /r/ were collected at baseline.

Table 3.2 /r/ accuracy ML
Baseline single word assessment, pre-treatment assessment of /r/ words in standard phrases, and post-treatment assessment /r/ words in standard phrases.

ML	Baseline			Pre-treatment sample			Post-treatment sample		
	RQ	[ɹ]	typical substitutions	RQ	[ɹ]	typical substitutions	RQ	[ɹ]	typical substitutions
WI	0/10	0/10	[β]	0/10	0/10	[β]	4/10	4/10	[ʊ̣̜]
WF	0/6	0/6	[əɔ̣̜]	0/6	0/6	[əɔ̣̜]	5/6	1/6	
WM	0/4	0/4	[ə]	0/4	0/4	[əɔ̣̜]	1/4	3/4	[əɔ̣̜]
#Cr	4/9	0/9	[ẉ̜ʰ], [ʊ̣̜]	5/9	0/9	[ẉ̜ʰ], [ʊ̣̜]	1/9	8/9	

3.3.2 Activity level and /r/ performance

1. Differences were observed in ML's and VF's ability to produce /r/ accurately in connected speech vs. single words (figures 3.1 and 3.2).

Figure 3.1 Connected speech /r/ production vs. /r/ in isolated words: ML

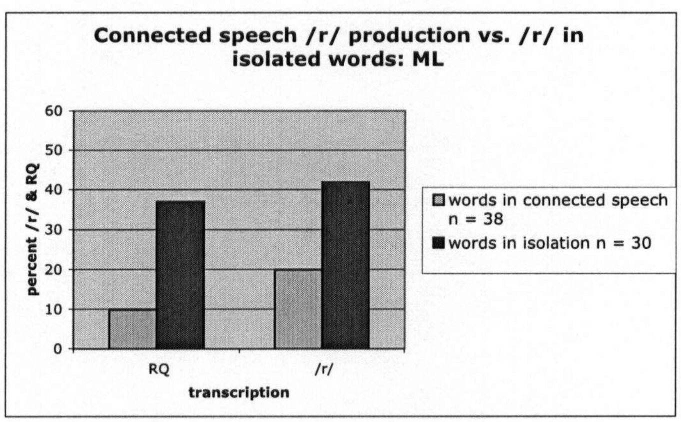
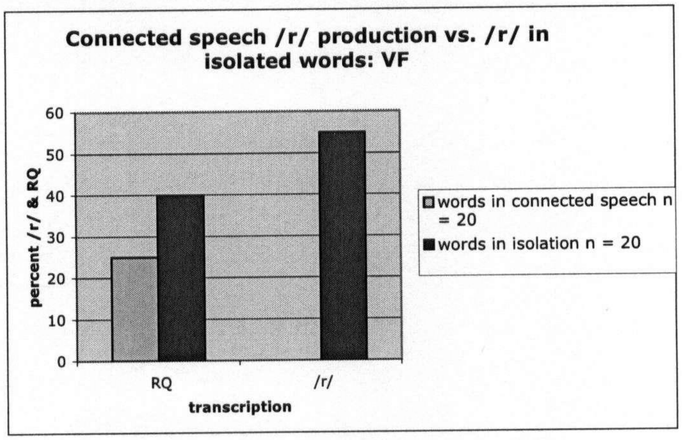


Figure 3.2 Connected speech /r/ production vs. /r/ in isolated words: VF



ML produced /r/ words in isolation with more accuracy than when /r/ words were in connected speech. In single words 42% of the words had an on-target /r/ and 37% had RQ where in connected speech only 20% of ML's /r/ words contained an on-target /r/, and 10% had RQ. VF also produced /r/ words in isolation with more accuracy than when /r/ words were in connected speech. In single words 55% of the words had an on-target /r/ and 40% had RQ where in connected speech none of VF's /r/ words contained an on-target /r/ and 25% had RQ.

2. In phrases vs. isolated words there was also a difference in both participants' /r/ accuracy. ML's /r/ production was better when the words were in isolation. Most words that had RQ or VS when they were in phrases were produced with an on-target /r/ in isolation. In isolation 27/29 words were produced with an on-target /r/, and 2/29 with RQ; none were substitutions. When the same words were in phrases, ML produced 16/29 with an on-target /r/, 11/29 with RQ, and 2/29 were substitutions.

When VF produced words in isolation, 20/29 were on-target /r/, 6/29 had RQ, and three were a substitutions. When the same words were in phrases, 10/29 were on-target, 13/29 had RQ, and 6/29 were substitutions.

Section 3.4 Qualitative analyses during treatment sessions

The first author recorded data throughout the treatment sessions to track when the participants began using certain structures that facilitated /r/ production. The checkmarks in the below tables are not indicative of a certain mastery criterion, but when the participants began using these structures to successfully produce an on-target /r/ during the sessions.

Table 3.3 /r/ production in different contexts over treatment sessions: VF

VF	Knowledge **	/r/ TBRG*	Iso /r/	/lr/	/ɑr/	/ir/	/tr/	/dr/	/ʃr/	Sylls	Words ***	Phrases ***
1			Traditional elicitation- session one									
2		√I										
3							√	√				
4	√	√C										
5												
6				√	√							
7												
8												
9												
10												
11												
12			√							√I,F	√I,F	
13											√C,M	
14												√I,M,F

Table 3.4 /r/ production in different contexts over treatment sessions: ML

ML	Knowledge **	/r/ TBRG*	Iso /r/	/ɪr/	/ɑr/	/ir/	/tr/	/dr/	/ʃr/	Sylls	Words ***	Phrases ***
1			Traditional elicitation- session one									
2	√	√I										
3		√C					√					
4				√								
5									√			
6			√			√		√		√I,F		
7												
8												
9										√M,C		
10											√I,F,C	
11											√M	√I,F,C
12												√M
13												
14												

*TBR = Tip, Body, Root, Groove components in isolation (I), then combined (C) without phonation.

** Knowledge goal explicitly stating the components of /r/ articulation.

*** I=initial, F=final, M=medial, C=consonant cluster /r/.

3.4.1 Contexts that facilitated /r/

Throughout the therapy sessions, contexts that facilitated /r/ production were noted. For both VF and ML the following sounds facilitated /r/ production, and were the first contexts where ML and VF produced an on target /r/ as noted in tables 3.3 and 3.4.

1) /ɪr/: The participant could visually monitor that his tongue tip did not drop from the roof of the mouth while retracting the entire tongue body.
2) /ɑr/: The participant would hold the /ɑ/ sound while raising the tongue tip into the retroflex position. (/ɑ/ worked for VF only)
3) /ir/: The participant would hold the lateral tongue contacts while bringing the tongue tip up and retracting the tongue body.
4) /tr/ & /dr/: These consonants provided the tongue tip placement for /r/ The participant then moved the tongue body back while holding the tip and lateral contact.
5) /ʃr/: The participant held the /ʃ/ lateral contact while retracting the tongue body.

3.4.2 Final note on ML

At the times of the baseline and pre-treatment assessment, ML had difficulty with the /ɑr/ combination and produced it as /əɔ/ for example, 'are' → /əɔ/. This distortion was noted throughout ML's connected speech and in single words. Post-treatment ML could correct this vowel distortion when prompted, but in conversation the distortion of the /ɑr/ combination was still observed.

Section 3.5 Acoustic analyses results

The most prominent feature of /r/ quality is the dropping F3 towards F2 (figure 2.4). Peterson and Barney (1952) reported average formant values for /r/, F3 = 1690 Hz, and F2=1350 Hz for men. Averages from Flipsen et al.'s (2000) data set for male adolescents ages 12-14, were F3=1934 Hz, and F2=1337 Hz. F2 and F3 values were measured in the manner described in methods section 2.5.3. The mean hertz values from each participant were taken from all available tokens. For a paired sample *t*-test, variables were inspected for outliers using boxplot and scatterplot graphs. Extreme outliers were excluded from analysis. Split half averages were calculated for ML's and VF's data. F3 pre-treatment split half averages were 2506 Hz and 2486 Hz for ML, and 2794 Hz, and 2738 Hz for VF. F3 post-treatment split half averages were 2483 Hz and 2414 Hz for ML, and 2142 Hz and 2130 Hz for VF. Therefore, pairs with missing tokens were also excluded from analysis without skewing the results.

Table 3.5 VF pre- and post-treatment averages of F3 and F2 across WI WF WM and CC /r/

VF		F3	F2
Pre-treatment N = 225	mean Hz SD	2768.89 203.26	1155.70 254.76
Post-treatment N = 228	mean Hz SD	2134.05 309.05	1065.50 116.15

As reported in table 3.5 the F3 value decreased from pre- to post- treatment, 2769 Hz to 2134 Hz for VF. Spectrograms in Appendix D (figures D1 and D2) illustrate the difference between VF's pre- and post-treatment F3. F2 values remained relatively stable. This pre- to post-treatment difference in F3 values was statistically significant in a paired sample *t*-test $t(193) = 25.84$; $p<.000$.

Table 3.6 VF post-treatment formant values based on transcription

VF		F3	F2
/r/ N = 86	mean Hz SD	1918.47 245.51	1071.7 123.51
RQ N = 93	mean Hz SD	2170.18 230.93	1041.60 96.284
Substitution N = 49	mean Hz SD	2443.85 225.81	1099.90 129.58

When the formant frequencies were analyzed according to perceptual transcription there was a decline in F3 from substitution (2444 Hz) to on-target /r/ productions (1918 Hz). Again, the F2 values remained relatively stable across transcriptions and the F3 dropped increasingly closer to F2 as the transcriptions improved towards an on-target /r/. The same trend is observed in table 3.7 based on the best and worst /r/ ratings for each token. For VF's /r/ productions rated as 'best' the F3 value drops near to Peterson and Barney's (1952) male average, and below the average for adolescent males ages 12-14 (Flipsen et al., 2000).

Table 3.7 VF post-treatment formant values based on best and worst ratings

VF		F3	F2
Best	mean Hz	1884.99	1053.50
N = 15	SD	311.75	98.77
Worst	mean Hz	2472.84	1116.00
N = 24	SD	252.172	160.48

Table 3.8 ML post-treatment formant values based on transcription

		F3	F2
/r/	mean Hz	2236.63	1266.13
N = 61	SD	209.52	148.87
RQ	mean Hz	2390.60	1243.90
N = 67	SD	262.44	170.21
Substitution	mean Hz	2592.01	1229.00
N = 105	SD	291.91	185.99

When ML's target /r/, RQ, and substitution productions were separated, ML's on-target /r/ productions had a lower F3 frequency value than his substitution or RQ productions. His best and worst productions showed an even greater contrast in F3 values (table 3.9).

Table 3.9 ML post-treatment format values based on best and worst ratings

		F3	F2
Best	mean Hz	2155.78	1272.88
N = 19	SD	166.06	134.60
Worst	mean Hz	2628.69	1194.96
N = 35	SD	273.23	152.64

The F2 between his best and his worst productions remained relatively stable and the F3 in his best productions dropped towards the second formant. The difference between F3 values of ML's best and worst productions was significant in an independent sample *t*-test (equal variance not assumed) $t(50.26) = -7.81$; $p < .000$. The F3 of ML's best /r/ productions is still higher than Peterson and Barney's (1952) male F3 average, but closer to the averages of 12-14 year old adolescents (Flipsen et al., 2000).

As was reported in the perceptual results section, ML was most successful when /r/ was in isolated words. Single word /r/ samples were collected at the post-treatment assessment and formant frequencies were measured. The values were averaged over ten stimuli words with four tokens each (n = 40) F3=1644 Hz, and F2=1037 Hz.

When ML produced /r/ without the carrier phrase he was consistent in his ability to use an on-target /r/ every time. Perceptually, these productions were rated at 100% on-target /r/. For words in isolation ML's F3 value is similar to the value outlined by Peterson and Barney (1952) and lower than Flipsen et al.'s (2001) adolescent 12-14 year old averages. Figures D3 and D4 in Appendix D illustrate the acoustic differences between ML's post-treatment /r/ in phrases, and in isolation.

Finally, ML's and VF's pre- and post- treatment F3-F2 scores for target word 'her' were converted to *z*-scores so that they could be compared to Shriberg et al.'s, (2001) on-target /ɝ/ and de-rhotacized /ɝ/ *z*-scores for adolescents. The data presented in Shriberg et al. (2001) represented a group of adolescents (group one) who had prior speech delay and produced de-rhotacized /ɝ/ with a mean *z*(F3-F2) score of 4.78, or they produced on-target /ɝ/ with a mean *z*(F3-F2) score of 3.07. In contrast, a group with no history of speech delay (group four) produced on-target /ɝ/ with a mean *z*(F3-F2) score of 0.14. ML and VF had pre-treatment 'her' *z*(F3-F2) scores of 6.06, and 11.55 respectively. This indicates that their pre-treatment F3-F2 values are larger than is typically expected during /ɝ/ production. Post-treatment, VF's 'her' in phrases z(F3-F2) score was 3.8. VF's mean z-score is similar to those perceived as on-target /ɝ/ produced by Shriberg et al.'s (2001) group one. ML's score for 'her' in phrases post-treatment was above the mean z-score for on-target /ɝ/ production (8.72). When ML's z(F3-F2) score was calculated for 'her' in isolation, mean z(F3-F2) was 0.59. This score falls close to what is expected for an individual with typical /ɝ/ production.

Section 3.6 Ultrasound measurement results

The reported averages in tables 3.10-3.12 are taken from all available tokens. Prior to running the paired sample t-tests, variables were analyzed for outliers through inspection of scatterplot and boxplot graphs. Extreme outliers were excluded from analysis. Split-half averages were calculated for ML's and VF's data. All of the split-half averages for T, B, and R measures fell within 1.4 mm of each other. Therefore, pairs with missing tokens could be excluded from analysis without skewing the results.

The ultrasound measurement values are reported in terms of distance(D) in (mm[1]) from the probe centre to the T, B, and R points hand-marked along the surface of the tongue. The overall goal was for the participants' tongue shape to approximate the sagittal tongue shape for /r/ that was modeled for them during the therapy sessions (figure 1.1). Figures D5 and D6 in Appendix D illustrate examples of VF's tongue shapes for /r/ at pre- and post-treatment. Table 3.10 outlines the gross differences between pre- and post-treatment tongue shapes for VF when tokens were averaged over all productions. Based on the observed pre-treatment /r/ tongue shapes in table 3.11 for VF, the majority of substitutions had high back tongue shapes (no root

[1] Scale is 1.38:1, or, 13.8mm of reported change is equivalent to 10mm of change within the oral cavity.

retraction) in all syllable positions. Therefore, post-treatment changes should be similar to those predicted for /ʊ/ in chapter one section 1.5: (a) tongue tip raising, (b) body lowering, and (c) root retraction.

Table 3.10 VF tongue distances from probe centre averaged over all tokens

	DR	DB	DT
Pre-treatment mean(mm)	60.7	83.00	59.64
SD	5.11	2.73	5.33
N	250	250	250
Post-treatment mean(mm)	68.85	77.46	80.95
SD	4.55	3.18	4.07
N	264	272	231
Paired sample *t*-test P<.000 *significant	N= 230 T=21.11*	N=243 T=-29.36*	N=201 T=51.76*

According to the DT values, VF's tongue tip significantly increased its distance from the probe centre an average of 21.31 mm post-treatment indicating that his tongue tip reached up to form the anterior oral constriction for /r/. VF's DR values also significantly increased post-treatment; this indicated that at post-treatment the root of his tongue was retracting more towards the pharyngeal wall, creating the posterior oral constriction. VF's DB post-treatment was significantly less than the pre-treatment assessment value; the height of the body/dorsum dropped, creating an expansion at the velar area. Table 3.11 identifies change in tongue shape by syllable position. Tongue shape changes for /r/ in each syllable position were predicted according to pre-treatment substitution tongue shape as observed on ultrasound.

Table 3.11 VF tongue shape changes by syllable position

VF	Substitutions Pre-treatment (perceptual)	Most frequent Pre-treatment tongue position (observed)	Required changes based on pre-treatment tongue shape	Significant change? p<.000	Df & *t*-value
WI Post N = 95	[ʊ̹]=54 [l]= 11 [w̹]=29	High back ≈ 100%	Tip raising Y Body lowering Y Root retraction Y F3 lowering? On-target /r/ n = 17	Y✓ Y✓ Y✓ Y 2249Hz	60, 28.9 76, -14.0 75, 26.4 68, 11.6
Syllable final (WM/WF) Post N = 97	[əɔ̹]=38 [əʊ̹]=10 [a]=10	High back ≈ 100%	Tip raising Y Body lowering Y Root retraction Y F3 lowering? On-target /r/ n = 68	Y✓ Y✓ Y✓ Y 2016Hz	81, 43.9 88, -16.4 82, 10.2 82, 18.3
CC Post N = 87	[w̹]=87	High back ≈ 100%	Tip raising Y Body lowering Y Root retraction Y F3 lowering? On-target /r/ n = 19	Y✓ Y✓ Y✓ Y 2155Hz	64, 26.2 76, -25.3 79, 12.3 49, 11.9

✓ = change matched prediction based on pre-treatment observed tongue shape

Based on VF's frequent pre-treatment observed tongue shape of a high back vowel or glide in all syllable positions, required changes post-treatment for maximal F3 lowering were similar to those predicted for high back vowel /ʊ/. These changes were all significant for /r/ and F3 was significantly lowered in each category.

Figure D7 in Appendix D is an example of ML's /r/ substitution pre-treatment. Based on the observed tongue shapes in Table 3.12 for ML, he used a different substitution in WI position vs. syllable final, or CC positions. It therefore does not make sense to average his tongue shape changes over all productions as different predictions were made depending on the pre-treatment tongue shape (chapter one, section 1.5). For WI /r/ ML used high tongue shapes with no pharyngeal component, therefore predictions for change would be similar to high back vowel /ʊ/:

(a) tongue tip raising, (b) tongue body lowering, and (c) root retraction. For /r/ in syllable final, and CC positions, ML commonly had a high tongue shape with a pharyngeal component, therefore predicted changes were:(a) tongue tip raising, (b) body lowering, and (c) no root retraction. Table 3.12 identifies significant changes in tongue shape by syllable position.

Table 3.12 ML tongue shape changes by syllable position

ML	Substitutions Pre-treatment (perceptual)	Most frequent Pre-treatment tongue position (observed)	Required changes based on pre-treatment tongue shape.	Significant change? P<.000	Df & *t*-value
WI Post N = 97	[β]= 94 [w]=3	High front/mid/ back ≈83%	Tip raising Y Body lowering Y Root retraction Y F3 Lowering? On-target /r/ n = 34	Y✓ Y✓ Y✓ -* 2325Hz	81, 15.8 81, - 6.8 83, 18.5 -
Syllable final (WM/WF) Post N = 96	[əɔ̹] = 56 [əʊ]=30 [ɔ]=10	High front/mid/ back with pharyngeal constriction ≈70%	Tip raising Y Body lowering Y Root retraction N F3 Lowering? On-target /r/ n = 2	Y✓ N✗ Y✗ N 2655Hz	78, 15.3 81, 5.9 77, -9.0 80, 1.2
CC Post N = 88	[w̹ʰ]=14 [w̹]=51, [w]=13 [w˞]=12	High front/mid/ back with pharyngeal constriction ≈86%	Tip raising Y Body lowering Y Root retraction N F3 Lowering? On-target /r/ n = 24	Y✓ Y✓ Y✗ N 2316Hz	76, 23.2 75, -15.7 76, -19.3 47, -.57

* Significance could not be calculated as pre-treatment WI /r/ had fricative quality and acoustic analysis was not completed.
✓ = Change matched predictions for /r/ based on observed pre-treatment tongue shape
✗ = Change did not match predictions for /r/ based on observed pre-treatment tongue shape

Changes at post-treatment matched predictions except for body and root in the syllable final category, and root in the CC category. An interesting point to note is that for syllable final, and for consonant clusters the tongue root was significantly pulled forward post-treatment: for syllable final, DR mean of 56.90 mm pre-treatment → mean of 51.56 mm post-treatment, for consonant clusters, DR mean of 63.81 mm pre-treatment → mean of 53.96 mm post-treatment.

F3 for syllable final /r/ and CC /r/ did not significantly change post-treatment (WI could not be tested). For syllable final /r/ no significant F3 lowering was expected as the low number of on-target tokens in this category (n=2). This may be due to the tongue not achieving required body/dorsum lowering, and loss of root retraction. For CCs, F3 also was not significantly lower post-treatment. Similar to syllable final /r/, there was a loss of root retraction observed for CC productions.

CHAPTER 4: Discussion

Section 4.1 Review of theoretical basis for using ultrasound in speech therapy

This study incorporated visual feedback from ultrasound into a traditional treatment program to teach two adolescents the /r/ phone. The treatment program consisted of breaking /r/ down into its component gestures (tip, body, root) and teaching each individually to the participants. The goal was for the participants to gain an awareness of the articulatory requirements for /r/ and subsequently to learn and combine the motor components to form the tongue shape for /r/. Results were evaluated perceptually, acoustically, and through tongue shape measurement.

According to motor learning theory, the best practice conditions for acquiring a new skill are when (a) the learner forms a mental image and a conscious awareness of the target behavior before attempting the task, and (b) augmented feedback is provided to the learner that states the performance result (KR) 'yes' or 'no', and what it was about the motor behavior that contributed to this result (knowledge of performance KP) (Fletcher, 1992; Ruscello, 1993; Ruscello & Shelton, 1979; Schmidt, 1982). For speech sounds such as /r/ these conditions are difficult to fulfill. The /r/ is a very complex articulation with multiple components, three of which are occluded within the oral cavity (palatal and pharyngeal constrictions, and mid-line grooving). Ultrasound allowed us to break this visibility barrier by providing sagittal and coronal images of the tongue shape during speech production. Incorporating visual feedback from ultrasound into a traditional speech therapy setting (table 2.3) provided the learner with articulatory information that was required to modify his current /r/ motor program.

Section 4.2 Integration of results

Two adolescent participants in this study were expected to learn the /r/ phone after 13 one-hour sessions of speech intervention using visual feedback from ultrasound. Both participants had received one and two years of traditional /r/ therapy, and had a several month no-treatment baseline without /r/ improvement. Pre-treatment transcription, acoustic, and ultrasound measures indicated that prior to ultrasound intervention neither participant could produce an on-target /r/ in any context. Post-treatment measurements supported the hypothesis that providing the participants with visual feedback from ultrasound was a useful tool in teaching the /r/ phone. Note that during the pre- and post-treatment assessments, the participants were not allowed to see the ultrasound image. The results are reflective of the participants' /r/ performance using internal feedback mechanisms (tactile, auditory, kinesthetic).

Transcription results at post-treatment assessment indicated that both participants produced the /r/ phone with more tokens falling in the on-target /r/ and RQ categories than at the pre-treatment assessment. At pre-treatment most of the participants' /r/ tokens fell within the substitution category. Task complexity was found to be a factor affecting both participants' performance. Words in isolation contained a higher percentage of on-target /r/ productions than words in phrases or connected speech. The participants needed a controlled and structured environment in order to be successful. The motor behavior of producing the /r/ phone was not yet an automatic task for either participant.

Acoustic results support the transcriptions. Post-treatment, both participants' F3 dropped towards F2 as is expected during /r/ production. This was supported by the averaged F3 values, and mean *z*(F3-F2) scores which were comparable to on-target /r/ productions (Flipsen et al., 2000; Peterson & Barney, 1952; Shriberg et al., 2001). However, as stated above, this occurred at different activity levels for each participant. VF showed a significant difference between his pre- and post-treatment F3 values when words were in standard phrases. VF's *z*(F3-F2) score post-treatment for 'her' in phrases was close to the on-target /ɝ/ productions of group one in Shriberg et al.'s (2001) study. ML's productions did not show a significant difference between F3 values pre- and post-treatment when the measures were averaged across all tokens in standard phrases. However, when ML's best and worst productions were extracted, a difference in F3 frequency values was identified with the best production F3 values falling just slightly higher than Peterson and Barney's (1952) F3 average male values for /r/, and male adolescent F3 values (Flipsen et al., 2000). Finally, when formant values were measured for ML's words in isolation, the F3 average values fell below those supplied by Peterson and Barney (1952) and Flipsen et al. (2000). ML's *z*(F3-F2) score for 'her' in isolation was comparable to the expected *z*(F3-F2) score for an on-target /r/ production produced by the group of children with typically developing speech (Shriberg et al., 2001).

Acoustics and phonetics cannot be correlated on a one-to-one basis; however, some general patterns can be extracted. The lowered F3 of /r/ as predicted by perturbation theory (Kent & Read, 1992) is a result of constrictions at the lips, palate, and pharynx. Low F3 is also thought to be related to resonance of the front/sublingual cavity

(Alwan et al., 1997; Espy-Wilson et al., 2000; Guenther et al., 1999) where F3 lowering is associated with increased front cavity /sublingual length and volume. Finally, F3 lowering has been correlated with a dip in the tongue dorsum (expansion at the velum) (Delattre & Freeman, 1968). Based on VF's pre-treatment high back tongue shape across syllable positions (table 3.11) predicted changes to achieve maximum F3 lowering were: (a) tongue tip raising, (b) tongue dorsum lowering, and (c) tongue root retraction. When all tokens were analyzed together (table 3.10), and across syllable position (table 3.11) VF demonstrated a dramatic increase in DT as the tongue tip lifted to create the palatal constriction for /r/. In bringing the tongue tip up, the size and length of the front/sublingual resonance cavities would likely increase (Alwan et al., 1997; Espy-Wilson et al., 2000; Guenther et al., 1999). VF also demonstrated significant tongue body lowering, and tongue root retraction. An overall lowering of F3 across /r/ productions when words were in phrases coincided with achievement of the predicted constrictions (palatal and pharyngeal), and cavity expansions (frontal/sublingual and velar).

ML's tongue shape changes could not be averaged over all tokens because he had different tongue shape substitutions in WI vs. syllable final and CC positions. Based on ML's most common pre-treatment tongue shape for WI (high but no pharyngeal component) the predicted changes to achieve maximum F3 lowering were: (a) tongue tip raising, (b) tongue body lowering, and (c) tongue root retraction. Based on ML's typical pre-treatment tongue shape for /r/ in syllable final position, and CCs (high with retracted root) (table 3.12) predicted changes to achieve maximum F3 lowering were: (a) tongue tip raising and (b) tongue body lowering. Tongue root retraction was not predicted. WI tongue shape changes matched the predictions and ML achieved the required constrictions and expansions to lower F3 in WI position. When /r/ was in syllable final position, the tongue tip prediction was matched, tongue body lowering did not occur, and ML's tongue root actually moved forward post-treatment. His F3 values in syllable final position did not significantly lower indicating that for /r/ in this position, his tongue did not achieve the required constrictions and expansions to cause F3 lowering. Finally, when /r/ was in CC position tip and body predictions were matched. However, like syllable final /r/, the root actually moved forward post-treatment. F3 did not significantly lower post-treatment for /r/ in clusters. This could be because ML lost the required pharyngeal constriction. The genioglossus muscle is responsible for the actions of dorsum/body lowering, tongue tip lowering, as well as pulling the root forward. Perhaps ML, in attempting to achieve dorsum lowering (contraction of the middle genioglossus) simultaneously contracted the posterior portion of this muscle drawing the tongue root forward. (Note that ML's tongue shape changes for words in isolation could not be quantified as inter-speech resting position could not be obtained. Inter-speech resting position was required in order to 'normalize' the tongue on a grid system.)

A trend for both participants was that their tongue gestures for /r/ became increasingly differentiated from pre- to post-treatment. Differentiation is defined as "increased independence in control of the components involved in a motor task (Green et al., 2000)." Green et al., (2000) stated that "limited independence of anatomically distinct segments is common in immature motor systems," and will decrease with maturation and training. For both ML and VF their pre-treatment undifferentiated /r/ productions (tongue moving as a whole unit) were learned at an early age when their motor and cognitive systems were immature. This early acquired undifferentiated motor pattern for /r/ failed to change as their motor and cognitive systems matured. However, through treatment both participants learned to produce the independent gestures required for /r/. VF learned to retract his tongue root as his tongue tip moved up to create the palatal constriction, while pulling his tongue body down. ML learned to move his tongue tip up to create the palatal constriction (WI, syllable final, and CC), and to pull his tongue body down (WI and CC), but his pharyngeal constriction was variable. In WI position ML could achieve both the pharyngeal and the palatal constrictions. When /r/ was in syllable final, or CC positions the tongue root retraction he had prior to treatment was lost as he gained tongue tip raising. This demonstrates that ML still might be having some difficulty moving his tongue root independently from the rest of his tongue body and tip.

Overall, both ML's and VF's /r/ productions qualitatively approximated the modeled sagittal target shape for some productions of /r/ (figure 1.1). At some level, both showed significant changes as predicted for tongue tip, tongue body, and tongue root positioning within the oral cavity between pre- and post-treatment (all positions for VF, and WI for ML when words were in phrases). The acoustic consequence of this change in tongue shape was a decrease in the F3 value of VF's /r/ productions. ML did not show overall acoustic change but when his best and worst productions were factored out, lowering of F3 occurred during his best productions. In addition, ML's words in isolation contained low F3 values. Factors hypothesized to be responsible for limiting acoustic change of ML's overall /r/ productions were loss of tongue root retraction during syllable final, and CC productions, as well as no tongue body lowering during syllable final productions.

According to the data, VF could produce /r/ with more accuracy than ML when the words were in phrases. One reason for VF's ability to produce /r/ in phrases exceeded ML's might be due to the fact that VF received distributed intervention over a five-month period. ML received intervention over a period of a month and a half. For this reason, VF had more time to practice his assigned /r/ homework. In addition to more practice opportunities, VF also had more time to consolidate the articulatory information he learned about /r/ between sessions.

Section 4.3 Controlled vs. automatic processing

According to the World Health Organization's International classification of Functioning, Disability and Health (ICF), speech intelligibility is a measure of activity level functioning (McLeod & Bleile, 2004). ML and VF produced /r/ with varying accuracy depending on the structure of the task. Although speech intelligibility was not formally quantified, it is logical that ML's and VF's speech was more or less intelligible depending on the linguistic structure of the task (single words, phrases, conversation). Both produced on-target /r/ consistently in isolated words consequently increasing their speech intelligibility, and inconsistently in phrases and conversation (fluctuating effects on speech intelligibility).

It is apparent from the post-treatment assessment data that the /r/ phone was not yet under automatic control for either participant. Automatic control is defined as a process that occurs without intention, and does not require conscious awareness or introspection (Poser & Snyder, 1975). The reverse is true for a conscious processing task, which is defined as a process that occurs with intention, is open to introspection, and draws upon an individual's pool of attentional resources (Poser & Snyder, 1975). Both participants had to concentrate on the target motor behavior in order to be successful in their /r/ production.

At the initial stages of motor learning the goal is to establish conscious control (conscious processing) of the target behavior using augmented feedback; however, the later stages of motor learning strive for generalization and automatic production of the target behavior using internal feedback (Ruscello, 1993). ML and VF appeared to be at the initial stages of generalization. They were both able to use internal feedback mechanisms to produce the phone in small linguistic units but still require a structured environment to be successful. The next step for them will be to integrate the /r/ sound into larger units and conversation through drill and rehearsal. ML and VF will both receive further speech intervention during the school year.

Using ultrasound was helpful in the initial stages of learning the /r/ sound. This was a skill that neither ML or VF had acquired with previous speech intervention. The ultrasound provided ML and VF with the underlying cognitive knowledge and ability to monitor their tongue positioning visually for the /r/ phone to bring the behavior under conscious control. These components were absent in previous intervention. Using ultrasound for speech intervention proves to be efficacious in teaching the North American /r/ to two adolescents who struggled to learn the sound through other speech therapy techniques.

Section 4.4 Challenges in this field of research

Such innovative research is not without its flaws and challenges. First, tongue shape measurement proved to be difficult due to the nature of the ultrasound imaging technique. Although transducer placement was controlled for through angle measurements on the probe cuff, and stable head positioning was ensured, it was clear that the pre- and post-treatment ultrasound images for ML and VF were recorded at different vertical/horizontal and rotational positions. For this reason the inter-speech rest position (Gick et al., 2004) adjustments had to be made before the distances from probe centre could be compared pre- to post-treatment. In addition to static probe placement differences, small head movements during the recording can also affect the ultrasound image measures.

This tool is by no means a quick fix requiring only one dose of treatment. This study provided evidence that the ultrasound is useful during the initial stages of speech intervention. However, after 13 one-hour sessions the participants were at a performance level where they still needed to practice the sound in a controlled linguistic environment in order to be successful. The ultrasound allowed initial changes in the motor program, but it is still no easy task to break a habit that one has consistently repeated hundreds of times per day for more than a decade. To overcome the use of an old motor behavior there must be abundant practice of the new one.

Access to this sort of equipment was also a challenge for speech-language pathologists who work in the school district, health units, or privately. However, the cost benefit of having access to an ultrasound may far outweigh the amount of dollars that are spent for a speech-language pathologist to work with children who have persistent /r/

distortions using traditional means. After one and two years of intervention neither ML nor VF respectively had acquired the /r/ sound in any context. Thirteen sessions with the ultrasound machine taught these two participants to produce an on-target /r/ in a controlled environment. We have found that in VF's case distributed exposure worked to elicit the /r/ sound. Therefore, even if a speech-language pathologist had access to this equipment once a month it would likely be beneficial in therapy. The speech-language pathologist could initially use the ultrasound to help the child or adolescent become aware of the articulatory requirements for /r/. In the following sessions when they did not have access to ultrasound the child or adolescent could then practice the components of /r/, and try /r/ in different phonetic contexts.

Section 4.5 Advancing this field of research

There are many questions that need to be answered in this new field of speech intervention. The participants in this study were young adolescents who produced the /r/ distortion throughout their life times. One question that arises is how early in speech intervention could we use visual feedback from ultrasound as a tool in therapy? Would younger children equally benefit from visual feedback, or would it be too complicated for them to understand that the abstract image of the tongue on the monitor was a representation of their own tongue? With early intervention for these distortions we could correct these sounds before they become an ingrained motor habit resistant to change. Another question is how much exposure to the visual feedback is needed in order for it to be effective? These participants came to 13 one-hour sessions of therapy, but by the 10th session both were using the ultrasound only for warm-up, and in some difficult vowel/consonant contexts. Might less exposure to visual feedback be equally effective in teaching the /r/ sound? Finally, this study did not use coronal section ultrasound recordings for assessment purposes. It would be interesting to identify the changes for /r/ that occur coronally from pre- to post- treatment as the lateral tongue bracing, and midline lowering were discussed and practiced in therapy sessions.

The next step in this field of research would be to run a control group, or use a staggered baseline design where the ultrasound would not be introduced until after the participants received a certain number of controlled traditional therapy sessions. A school district may be targeted in such as study design where there are a number of needy participants for this type of therapy, and the intervention could continue throughout the year. It would also be interesting to compare the effectiveness of using visual feedback from ultrasound for eliciting /r/ to other types of visual feedback such as spectrographic feedback (Shuster et al., 1992; Shuster et al., 1995), or EPG (Bernhardt et al., 2003). Long-term outcomes also should be evaluated to discover the effects of this type of intervention on generalization to conversation months/years after intervention.

REFERENCES

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

Beckman, D. (1975) http://www.beckmanoralmotor.com/about.htm. September 15, 2004.

Bernhardt, B., & Stemberger, J. (1997). Optimality theory. In M. Ball, & R. Kent. (Eds.), *The new phonologies development in clinical linguistics* (pp.211-245). San Diego, CA: Singular Publishing Group Inc.

Bernhardt, B., & Stemberger, J. (1998). *Handbook of phonological development. From the perspective of constraint based nonlinear phonology.* San Diego, CA: Academic Press.

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

Bernthal, J., & Bankson, N. (2004). *Articulation and phonological disorders, fifth edition.* Boston, USA: Allyn and Bacon.

Clark, C.E., Schwarz, I.E., & Blakeley, R.W. (1993). The removable R-appliance as a practice device to facilitate correct production of /r/. *American Journal of Speech Language Pathology, 2*, 84-92.

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

Dickenson, D.R., & W. Maue-Dickenson. (1982). *Anatomical and physiological bases of speech.* Boston, USA: Little, Brown, and Company Inc.

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

Ferrand, C. (2000). *Speech science: An integrated approach to theory and clinical practice.* Needham Hights, MA: Allyn and Bacon.

Fletcher, S. (1992). *Articulation, a physiological approach.* San Diego, CA: Singular Publishing Group Inc.

Flipsen, P., Jr., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeny, J. L. (2000). *Acoustic data for American English /r/ and /ɝ/ in typically speaking adolescents* (Tech. Rep. No. 10). Phonology Project, Waisman Center, University of Wisconsin-Madison. http://www.waisman.wisc.edu/phonology/bib/bib.htm.

Flipsen, P., Shriberg, L., Weismer, G., Karlsson, H. B., & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics: Reference data for residual /ɝ/ distortions. *Clinical Linguistics and Phonetics, 15*(8), 603-630.

Gibbon, F. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*(2), 382.

Gick, B., Wilson, I., Koch, K., & Cook, C. (2004). Language-specific articulatory settings: Evidence from inter-utterance rest position. *Unpublished manuscript.*

Gick, B. & Campbell, F. (2003). Intergestural timing in English /r/. *Unpublished manuscript.*

Gick, B., Iskarous, K., Whalen, D., & Goldstein, L. (2003). Constraints on variations in the production of English /r/. *Proceedings of the 6th International Seminar on Speech Production,* Sydney, December 7-10.

Green, J., Moore, C., Higashikawa, M., & Steeve, R. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43,* 239-255.

Guenther, F., Espy-Wilson, C., Boyce, S., Matthies, M., 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.

Janzen, K., & Shriberg, L. (1977). *How to evoke and generalize "R", A compendium of 36 evocation and phonetic context cues.* Madison, Wisconsin: The University Bookstore.

Kent, R. (1992). The Biology of Phonological Development. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development models, research , implications* (pp. 65-90). Timonium, MD: York Press.

Kent, R., Kent, J., & Rosenbek, J. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders, 52*, 367.

Kent, R., & Read, C. (1992). *The acoustic analysis of speech.* San Diego, CA:

Singular Publishing Group Inc.

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

Masterson, J., & Bernhardt, B. (2001). *CAPES.* San Antonio, TX: The Psychological Corporation.

McGowan, R., Nittrouer, S., & Manning, C. (2004). Development of [ɹ] in young, Midwestern, American children. *Journal of the Acoustical Society of America, 115*(2), 871-884.

McLeod, S., & Bleile, K. (2004). The ICF: A framework for setting goals for children with speech impairment. *Child Language Teaching and Therapy, 20*(3), 199-219.

Palmer, J. (1993). *Anatomy for speech and hearing (4th ed.).* Boston: Allyn & Bacon

Peterson, G., & Barney, H. (1952). Control of methods used in a study of vowels. *Journal of the Acoustical Society of America, 24,* 175-184.

Poser, I., & Snyder, R. (1975). Facilitation and inhibition in the processing of signals. In P. M. Rabbit & S. Dornic (Eds.), *Attention and Performance V* (pp. 669-682). New York: Academic Press.

Ruscello, D. (1984). Motor learning as a model for articulation instruction. In J. Costello (Eds.), *Speech disorders in children, recent advances* (pp. 129-156). San Diego, CA: College Hill Press.

Ruscello, D. (1993). A motor skill learning treatment program for sound system disorders. *Seminars in Speech and Language, 14*(2),106-118.

Ruscello, D. (1995a). Speech appliances in the treatment of phonological disorders. *Journal of Communication Disorders, 28,* 331-353.

Ruscello, D. (1995b). Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders, 28*, 279-302.

Ruscello, D., & Shelton, R. (1979). Planning and self-assessment in articulatory training. *Journal of Speech and Hearing Disorders, XLIV*, 504-512.

Schmidt, R. (1982). *Motor control and learning, a behavioral emphasis.* Champaign, IL: Human Kinetics Publishers.

Shuster, L., Ruscello, D., & Smith, K. (1992). Evoking /r/ using visual feedback. *American Journal of Speech-Language Pathology, May*, 29-34.

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

Shriberg, L. (1975) A response evocation program for /ɝ/. *Journal of Speech and Hearing Disorders, 40*(1), 92-105.

Shriberg, L. (1980). An intervention procedure for children with persistent /r/ errors. *Language, Speech, and Hearing Services in Schools, 11*, 102-110.

Shriberg, L., Flipsen, P., Karlsson, H., & McSweeny, J. (2001). Acoustic phenotypes for speech-genetics studies: An acoustic marker for residual /ɝ/ distortions. *Clinical Linguistics & Phonetics, 15*(8), 631-650.

Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *American Speech-Language-Hearing Association, 55*, 779-798.

Stone, M. (1997). Experimental phonetics. In W. J. Hardcastle (Ed.), *The handbook of phonetic sciences* (pp. 11-32). Cambridge, Massachusetts: Blackwell Publishers Inc.

Stone, M., & A., Lundberg (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America, 99*(6), 3728-3737.

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

Appendix A: /r/ stimuli word list

rid
read
red
rad
ray
run
row
root
rook
raw

ear
her
air
or
poor
are

tray
cray
pray
dray
grey
bray
fray
shred
thread

heary
hurry
hairy
story

Appendix B: Therapy goals and methods

1. Knowledge goals, establishing awareness of the /r/ tongue shape through:

a) Discussion of where the tongue tip, body and root were displayed on the monitor (sagittal image), and orientation to the coronal image lateral tongue raising and mid-line lowering.

b) Clinician modeling of the /r/ phone for the participant and discussing the three components of the sound: tongue tip raising and tongue root retraction (sagittal view), and tongue lateral bracing and the midline groove (coronal view). Participants watched the clinician model the sound and identified the components of the /r/ tongue shape.

c) Freezing the /r/ coronal and sagittal images on the monitor: The participants sketched the tongue, and labeled the components. They compared their /r/ tongue shapes to their parents' /r/ and the clinician's /r/ to identify differences and similarities in articulation.

2. Motor goals, establishing the gestures for /r/: After the target gestures were identified, the participants used visual feedback to practice each gesture in isolation without and with phonation, and then in combination. Crosshairs were arranged on the ultrasound viewing screen so the participants could identify where their tongue tip should be reaching, and where their tongue body should be centered. These markers also allowed the participants to self-monitor their productions for accuracy. The clinician frequently

modeled the target motor goals during these activities. When the participants demonstrated control over each of the gestures they then practiced them in combination.

3. Production goals, production of the /r/ gestures in context: Each therapy session the clinician documented which phonetic contexts were easier for the participants and which were more difficult. Once /r/ was elicited, the speech hierarchy was used to guide further practice. Participants practiced /r/ in the following order, first with visual feedback and then without:

1. In isolation
2. In syllables (syllable-initial, syllable-final, syllable-medial, and in consonant clusters). Whenever possible, short one-syllable real words were introduced to make practice as functional as possible.
3. In words, participants were asked if they had any /r/ words they would like to practice.
4. In short phrases.

The participants practiced /r/ in at the level where they were successful for each context. For example, /r/ in consonant clusters 'dr' and 'tr' were easier for VF and so he practiced these at phrase level, while /r/ in the context of back rounded vowels was more difficult and so he practiced these at syllable level.

At the end of each session the participants were given activities to practice for ten minutes at home at the level of success during the therapy session. For example, at the initial stages of therapy the homework given was to practice the individual components of /r/ without phonation, whereas towards the end of the therapy program homework was to practice /r/ in words and phrases.

Appendix C: Inter-speech rest position adjustments

The below Figure C1 illustrates VF's inter-speech resting positions at pre- and post-treatment. The goal was to match pre- and post-treatment inter-speech resting position images through vertical and horizontal transposition, and angle rotation. The pre-treatment inter-speech resting position was matched to the post-treatment inter-speech resting position. The process required two steps as is illustrated through figures C2 & C3. First, the root points of the tongues were matched. This required shifting the whole pre-treatment tongue along the vertical axis .8mm, and along the horizontal axis –16.9 mm. The resultant position is illustrated in figure C2. With the root points matched the rotational difference was calculated through finding the degree of rotation required to match the tip points. For VF this was 11.03 degrees of upward rotation from the fixed root point. A rotation of 11.03 degrees was also applied to the body point and the resulting pre-treatment inter-speech resting position is illustrated in figure C3. The similarity between the pre- and post-treatment inter-speech resting positions after translation of (0.8, -16.9), and rotation of +11.06 degrees suggests that any vertical, horizontal, or rotational differences between pre- and post-treatment transducer placement have been accounted for. ML's data required a translation of -20.32 mm along the horizontal axis, and -11.59 mm along the vertical axis, followed by an upward rotation of 18.3 degrees in order to match the tip and body points. These calculations were applied to all of ML's and VF's pre-treatment /ɹ/ data. Note that these adjustments do not factor out any extraneous head movement during data collection. They only adjust for differences in static transducer placement.

Figure C1 pre- and post-treatment inter-speech rest position: VF

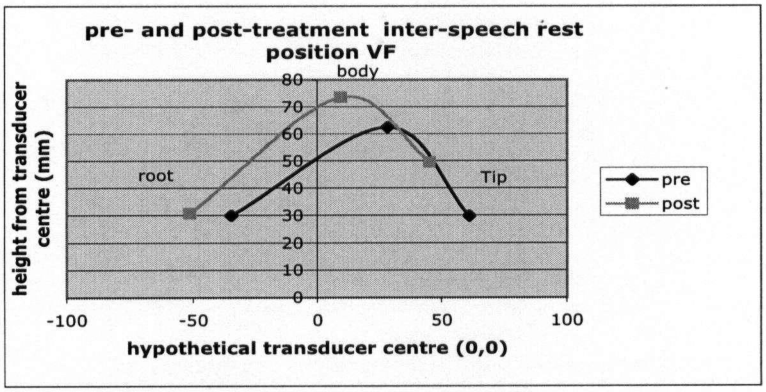


Figure C2 adjusted inter-speech rest position VF

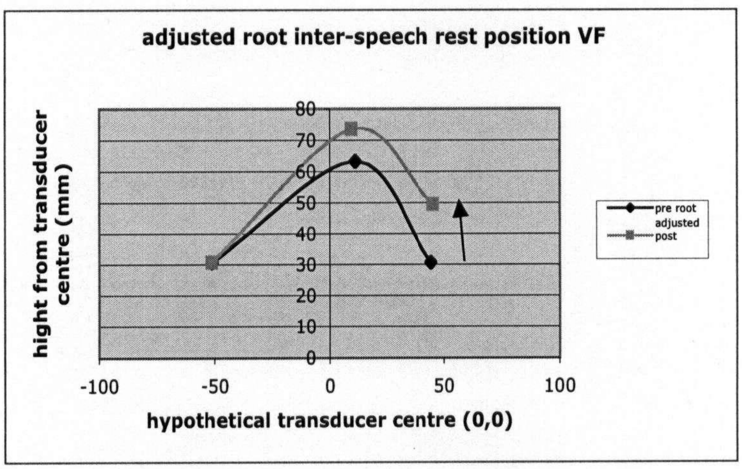
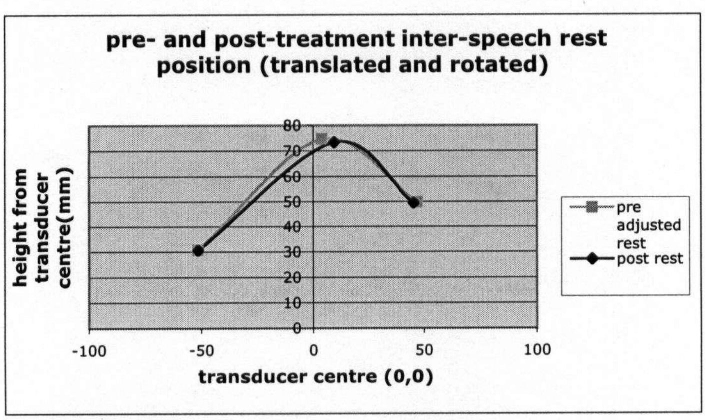


Figure C3 pre- and post-treatment inter-speech rest position (translated and rotated)



Appendix D: Spectrograms and tongue images for VF's and ML's /r/ production

Figure D1 VF 'rad' pre-treatment
F3 at /r/ = 2753 Hz
F2 at /r/ = 1112 Hz

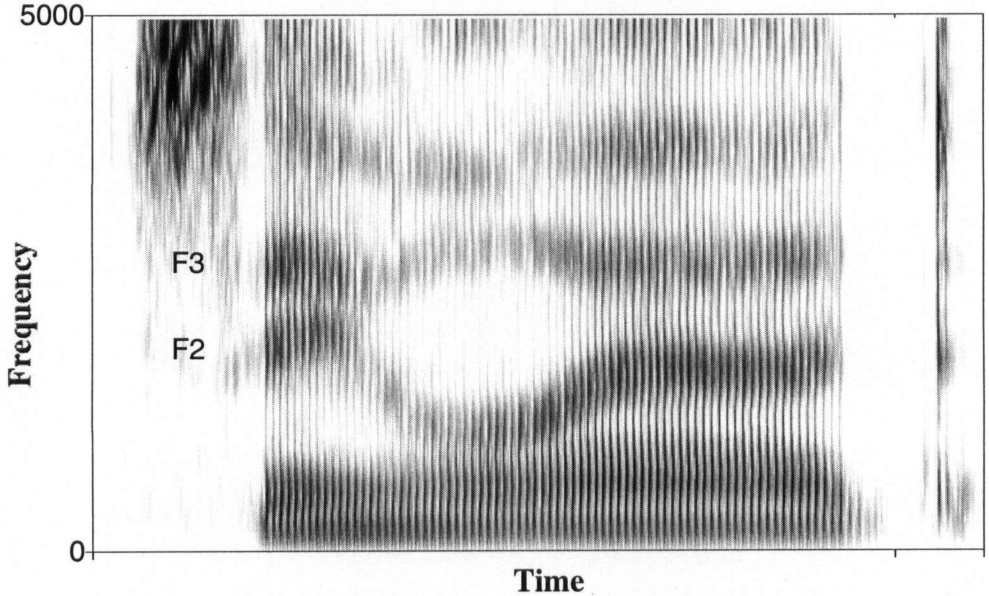


Figure D2 VF 'rad' post-treatment
F3 at /r/ = 1987 Hz
F2 at /r/ = 1166 Hz

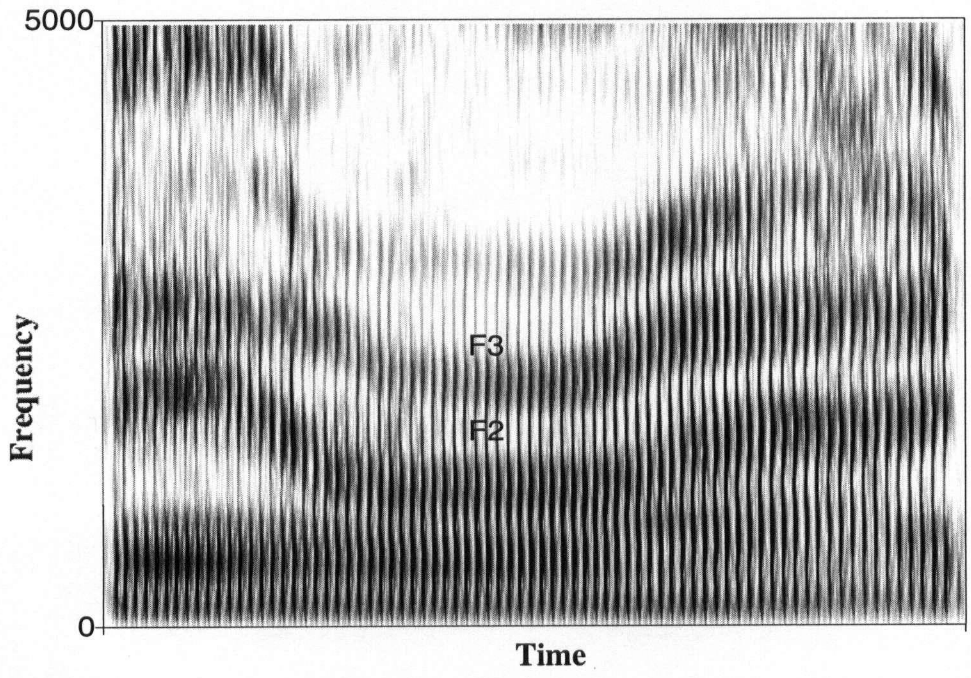


Figure D3 ML 'are' in phrase post-treatment
F3 at /r/ = 2424 Hz
F3 at /r/ = 1002 Hz

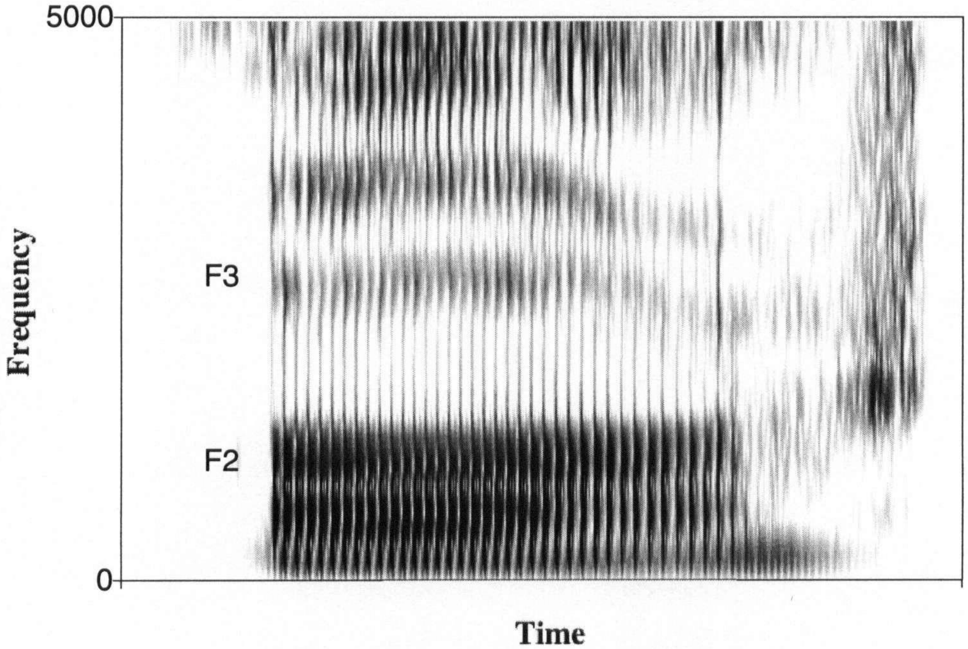


Figure D4 ML 'are' in isolation post-treatment
F3 at /r/ = 1549 Hz
F2 at /r/ = 1002 Hz

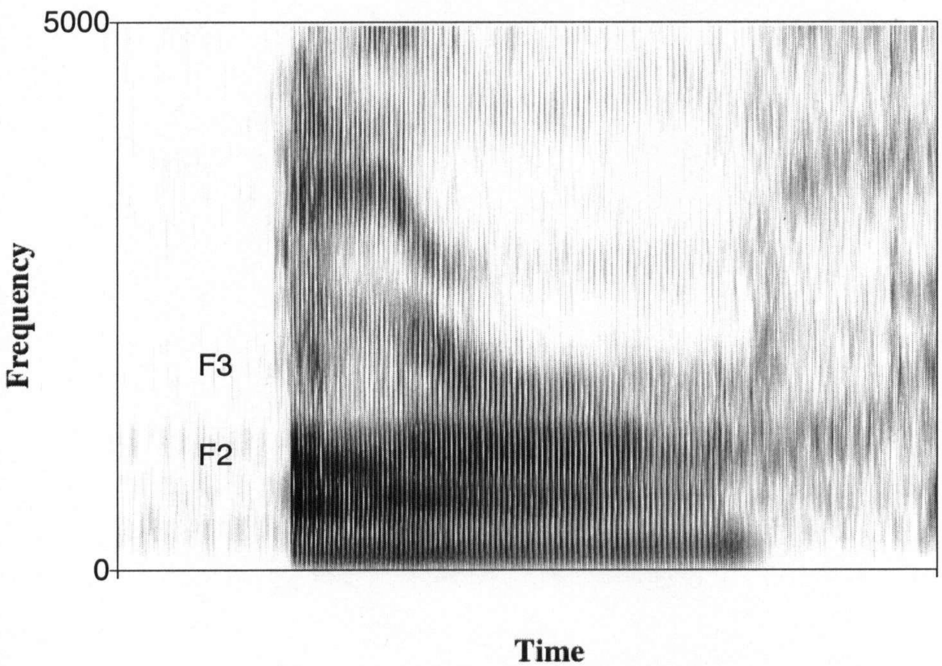


Figure D5 VF /r/ tongue shape pre-treatment

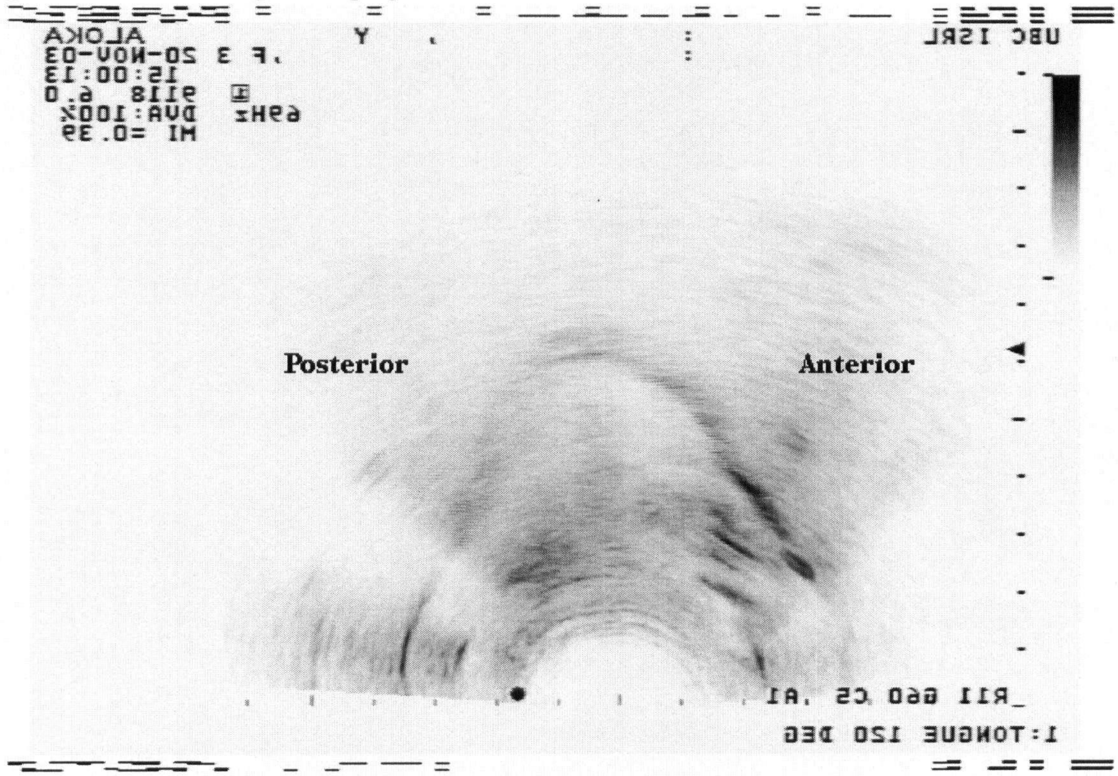


Figure D6 VF /r/ tongue shape post-treatment

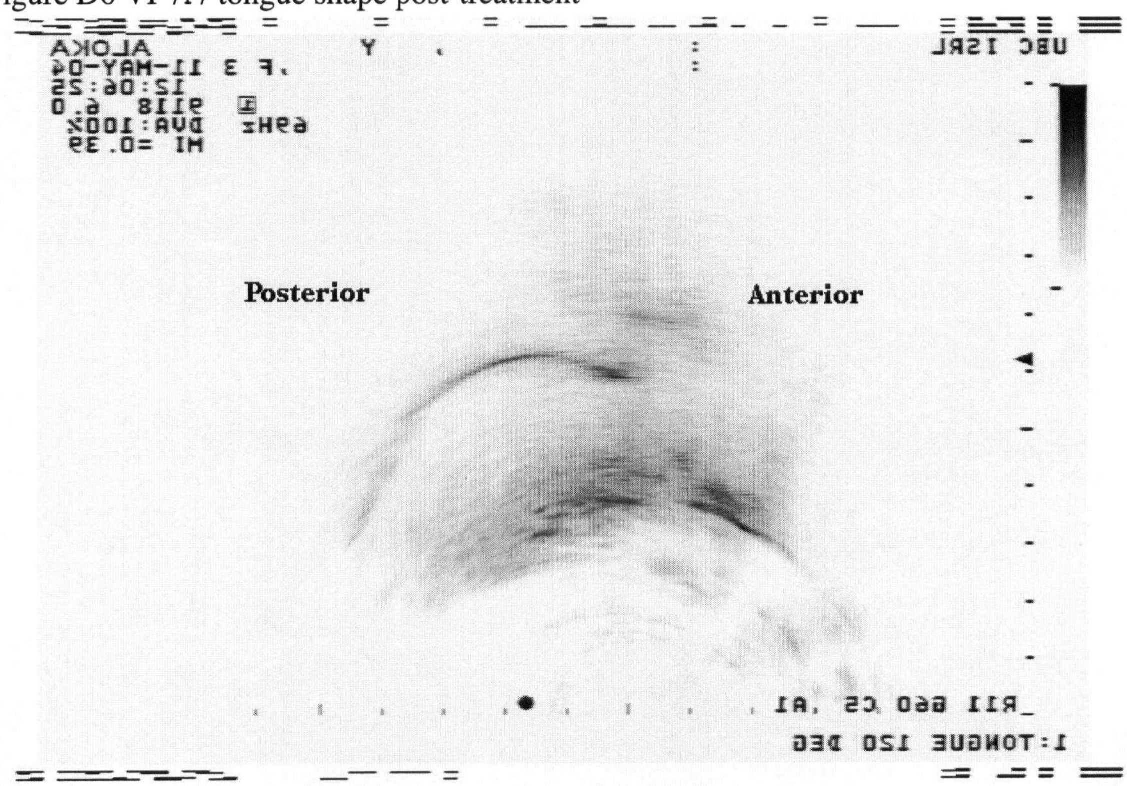


Figure D7 ML pre-treatment /r/ tongue shape

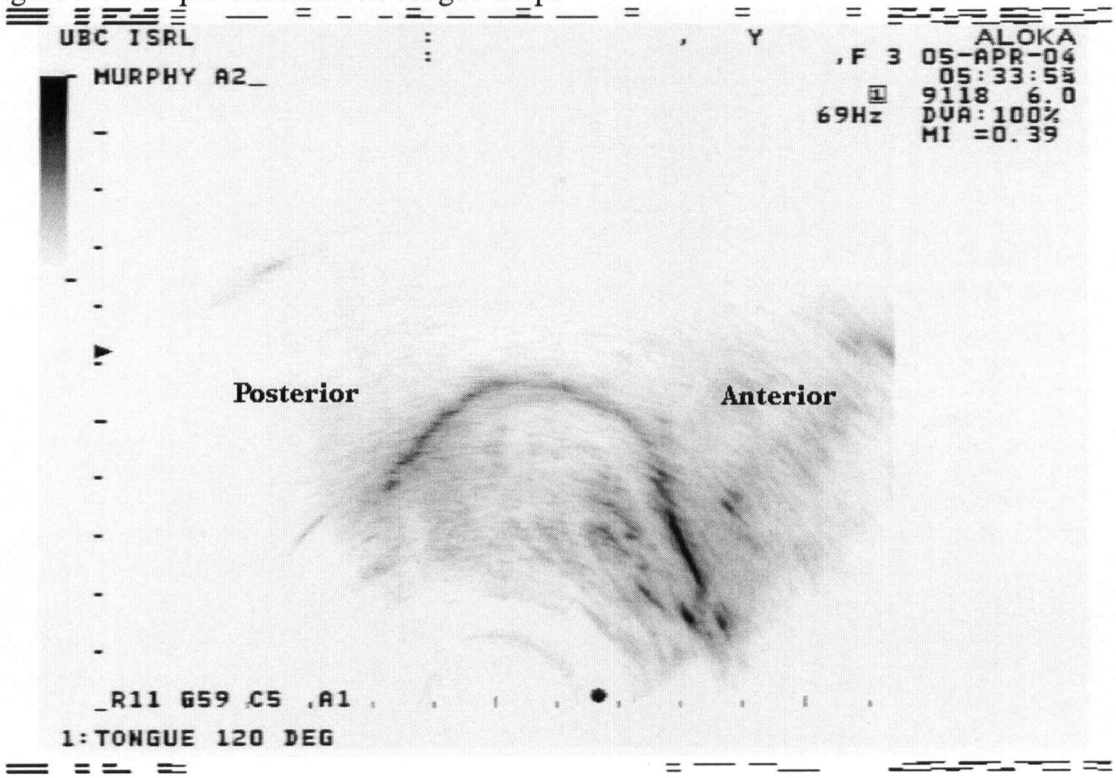


Figure D8 ML post-treatment /r/ tongue shape