Summer 8-11-2018

Ultrasound Visual Biofeedback and Accent Modification: Effects on Consonant and Vowel Accuracy for Mandarin English Language Learners

Courtney Armstrong

Recommended Citation

Armstrong, C. (2018). Ultrasound Visual Biofeedback and Accent Modification: Effects on Consonant and Vowel Accuracy for Mandarin English Language Learners (Master's thesis, Duquesne University). Retrieved from https://dsc.duq.edu/etd/1477

ULTRASOUND VISUAL BIOFEEDBACK AND ACCENT MODIFICATION: EFFECTS ON CONSONANT AND VOWEL ACCURACY FOR MANDARIN ENGLISH LANGUAGE LEARNERS

A Thesis

Submitted to the Rangos School of Health Sciences

Duquesne University

In partial fulfillment of the requirements for the
degree of Master of Speech-Language Pathology

By

Courtney Armstrong

August 2018

ULTRASOUND VISUAL BIOFEEDBACK AND ACCENT MODIFICATION: EFFECTS ON CONSONANT AND VOWEL ACCURACY FOR MANDARIN ENGLISH LANGUAGE LEARNERS

By

Courtney Armstrong

Approved July 9, 2018

Dr. Heather Leavy-Rusiewicz
Professor of Speech-Language Pathology
(Committee Chair)

Dr. Yang Chen
Professor of Speech-Language Pathology
(Committee Member)

Ms. Katie Micco
Assistant Professor of Speech-Language Pathology
(Committee Member)

Dr. Fevzi Akinci
Dean, The John G. Rangos, Sr. School of Health Sciences

Dr. Mikeal Kimelman
Chair, Department of Speech-Language Pathology

ABSTRACT

ULTRASOUND VISUAL BIOFEEDBACK AND ACCENT MODIFICATION: EFFECTS ON CONSONANT AND VOWEL ACCURACY FOR MANDARIN ENGLISH LANGUAGE LEARNERS

By

Courtney Armstrong

August 2018

Thesis supervised by Dr. Heather Leavy-Rusiewicz

The number of individuals in the United States who speak languages other than English continues to increase. With the increase of language diversity comes a potential rise in communication challenges for those who speak with non-mainstream American English accents as English language learners. A portion of these individuals may elect to seek accent modification services, perhaps due to decreased intelligibility or communication breakdowns. Thus, speech-language pathologists must research and provide effective techniques to enhance intelligibility of all American English speakers for optimal communication. Few approaches employ a variety of treatment methods to improve speech sound accuracy, naturalness and intelligibility to target accent modification. One of these methods is ultrasound biofeedback therapy. Ultrasound therapy relies on visual feedback for remediation of speech sound production errors for those with various etiologies and diagnoses. A single-subject ABAB withdrawal design was employed with two native Mandarin speakers to examine the effect of incorporating ultrasound visual biofeedback in the treatment of consonant and vowel targets as measured by perceptual, acoustic and visual analyses.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

Ultrasound visual biofeedback and accent modification: Effects on consonant and vowel accuracy for Mandarin English Language Learners

The United States census projects Mandarin to be the second-most commonly spoken language other than English (LOTE) in America by 2020 (Shin & Ortman, 2011, p. 12). These native Mandarin speakers will communicate in a country whose occupants speak a language with a significantly different phonetic inventory than their own which may impact the effectiveness of their professional and social communication exchanges. Such challenges can be addressed by speech-language pathologists (SLPs) who are aware of the speech sound mechanisms that support English Language Learners' (ELL) communication. However, empirical evidence about effective intervention for accent modification services provided by SLPs is limited. Thus, there is a call to expand investigations of the management of accents to best advocate for and implement evidence-based practice in the area of accent modification. One such treatment approach is ultrasound visual biofeedback. This technique, though novel, has been implemented with a variety of diagnoses and languages. However, employing ultrasound biofeedback as an accent modification approach has not yet been empirically studied. The following literature review examines the growth of Mandarin in the United States, the impact of having an accent, role of the SLP, current accent modification therapy approaches and ultrasound biofeedback therapy.

Chapter 1: Literature Review

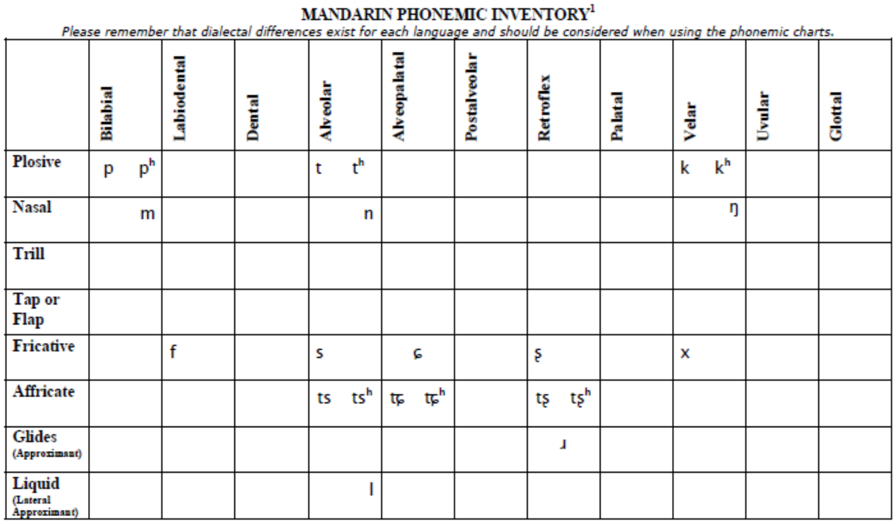
1.1 Mandarin Prevalence in the United States

The United States continues to be a melting pot not only of cultures, but also languages. Shin and Ortman (2011) stated, "the use of a language other than English at home increased by 148 percent between 1980 and 2009" (p. 1). Such exponential growth sparked a study from the United States census to project use of LOTE in 2020. Thirteen languages were included, each with more than 500,000 speakers in 2009 (p. 3). Various numerical equations were used for three projections and all determined LOTE to increase by 2020 (p. 4). Of the various languages studied, Mandarin was the second most spoken language, following Spanish, of all LOTEs in every projection (p. 12). LOTE will continue to increase in the United States and SLPs should consider the implications of these projections.

It is likely that nearly all Americans, including SLPs, will interact with native Mandarin speakers. This interaction may cause a decrease in intelligibility due to the notable phonetic differences between the two languages. For instance, Mandarin does not include consonant clusters or multisyllabic canonical shapes. It also does not contain many closed syllables and only two consonants, /n/ and /ŋ/, occur in the final position. Eight American English phonemes are not found in Mandarin including: /v/, /z/, /ʃ/, /ʒ/, /tʃ/, /ʤ/, /θ/ and /ð/ (ASHA, n.d.) (see Tables 1 & 2). Such differences inevitably cause a challenge when a native Mandarin speaker desires to speak American English. Common American English substitutions produced by Mandarin speakers include: /s/ or /f/ for / θ/, /d/ or /z/ for /ð/ and /f/ or /w/ for /v/ (ASHA, n.d.). Considering how often these phonemes are found in American English, such substitutions can impact optimal communication with a native American English speaker. It is notable, however, to point out that eight phonemes are found in both languages: /p/, /m/, / t/, /k/, /ŋ/, /f/, /s/ and /l/ (ASHA, n.d.). Although there are similarities, it is still likely that the contrasts will cause many Mandarin speakers to encounter challenges as they attempt to speak a phonetically different language.

Additionally, Mandarin and American English differ in their vowel inventories. American English contains 11 vowels; /i/, /ɪ/, /ɛ/, /e/, /o/, /ʌ/, /u/, /ʊ/, /ɔ/, /ɑ/ and /æ/ (Peterson & Barney, 1952). In Mandarin, there are only six vowels; /i/, /e/, / j/, /u/, /o/ and /ɑ/ (Chen, Robb, Gilbert and Lerman, 2001). There are vowels that are present in both languages, such as /ɛ/. However, several phonemes, such as /ɪ/, are unfamiliar to native Mandarin speakers.

Table 1. *Mandarin Phonetic Inventory (ASHA, n.d.).*

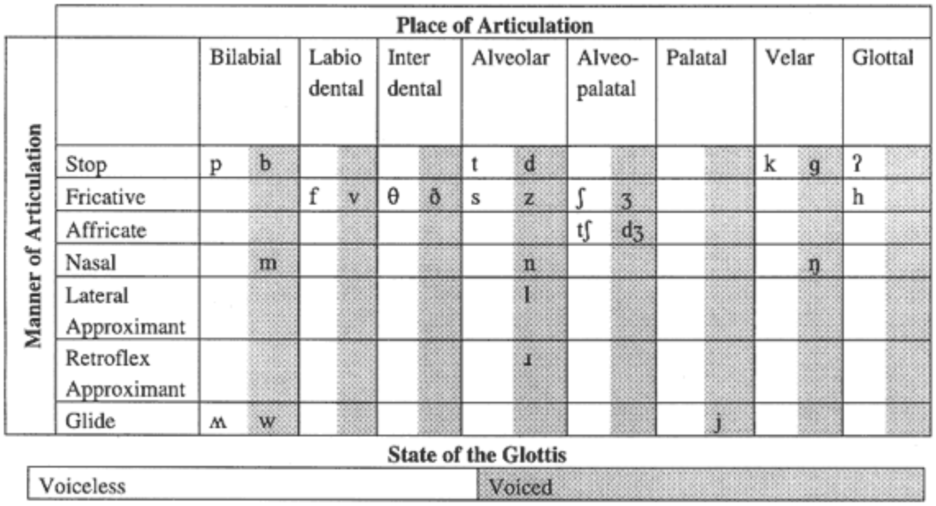


MANDARIN PHONEMIC INVENTORY[1]

Please remember that dialectal differences exist for each language and should be considered when using the phonemic charts.

	Bilabial	Labiodental	Dental	Alveolar	Alveopalatal	Postalveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Plosive	p pʰ			t tʰ					k kʰ		
Nasal	m			n					ŋ		
Trill											
Tap or Flap											
Fricative		f		s	ɕ		ʂ		x		
Affricate				ts tsʰ	tɕ tɕʰ		tʂ tʂʰ				
Glides (Approximant)							ɹ				
Liquid (Lateral Approximant)				l							

Table 2. *Consonants of Standard American English (Mihalicek & Wilson, p. 738, 2011).*



Place of Articulation

Manner of Articulation	Bilabial		Labio dental		Inter dental		Alveolar		Alveo-palatal		Palatal		Velar		Glottal	
Stop	p	b					t	d					k	g	ʔ	
Fricative			f	v	θ	ð	s	z	ʃ	ʒ					h	
Affricate									tʃ	dʒ						
Nasal		m						n						ŋ		
Lateral Approximant								l								
Retroflex Approximant								ɹ								
Glide	ʍ	w										j				

State of the Glottis

Voiceless	Voiced

Liquids are of particular interest in these languages because of the variations in pronunciation. Mandarin contains both /r/ and /l/ liquid phonemes; however, they differ from American English. According to Smith (2010), "Mandarin /l/ is a voiced apical denti-alveloar or apical alveolar lateral approximant and /r/ is an apical post-alveolar retroflex approximant" (p.

20). In contrast, Smith (2010) wrote, “in American English, /r/ is a voiced alveolar approximant and /l/ is a voiced alveolar lateral approximant” (p. 14). In addition, Smith (2010) continued, “unlike English /r/, Mandarin /r/ has little or no lip rounding and is produced with greater constriction, resulting in audible friction noise in some dialects and vowel contexts” (p. 20). In Mandarin, both can only occur in the syllable onset and neither can occur in a syllable coda. Although both languages have liquid consonants, articulation characteristics differ, thus creating challenges for a native Mandarin speaker to produce accurate speech. These subtleties create challenges for the non-native speaker because they are difficult to perceive. Anecdotally, an unfamiliar American English listener will perceive a vocalic /r/ as a derhotacized /r/. This perception will influence production so that vocalic /r/s are produced as derhotacized /r/s, thus impacting intelligibility.

1.2 **Underlying Mechanism of Accented Speech**

Although the production of more accurate or approximated productions of speech sound targets is emphasized in accent modification services, it is imperative to consider the interaction of speech perception and production to better understand the theoretical underpinnings of accented speech. Accented speech is thought to occur due to an inability to perceive phonemic differences rather than a lack of ability to produce motor movements of foreign phonemes. According to a study conducted by Shafer, Shucard, Shucard and Gerken (1998), “infants are sensitive to differences between the two language conditions and age is a factor” (p. 881). These authors analyzed infants' pacifier-sucking responses to new and unfamiliar sounds. Based on the findings, researchers concluded infants have the ability to perceive unfamiliar sounds from any language, both native and nonnative for the infants. Thus, as infants, all humans have the ability to perceive phonemic differences across all languages. However, Bernthal, Bankson and Flipsen

(2009) summarized, “as children are exposed to their native language and reach the end of their first year…their ability to discriminate nonnative sounds diminishes” (p. 85). In other words, as they become adults, humans lose their ability to discriminate between native and foreign phonemes. For example, a glottal and pharyngeal Hebrew fricative will be perceived as the same (i.e. a glottal fricative) to an American English speaker because these nuances do not exist in the native language. Moreover, a native Mandarin speaker will perceive a Mandarin /r/ and American English /r/ as the same because they are unfamiliar with the subtle perceptual differences.

Furthermore, Schmidt (1997) wrote of those participating in accent modification programs, “listeners had difficulty hearing the distinctions they were trying to learn to produce” (p. 2). Training individuals to perceive sounds they are unfamiliar with becomes a challenge of accent modification. Schmidt (1997) continued, “our speech perception systems have been influenced by the learning of our first language [so] that when we listen to the sounds of a foreign language, we do so using the categories of our first language” (p. 2). Following this, adult speakers of foreign languages have difficulty learning new sounds because they categorize sounds according to their native language and no longer have the ability to discriminate all foreign sounds as infants can. For example, while the /r/ phoneme differs significantly in Mandarin and American English, native Mandarin speakers may *produce* the American English /r/ similar to the Mandarin phoneme because they cannot *perceive* the difference. Therefore, it is important to train not only perceptual differences in accent modification to be sure the speaker understands the articulatory differences, but also the articulatory production of the speech sound targets.

1.3 Impact of Accent

According to Mihalicek and Wilson (2011), accent is a “systematic phonological variation” in *Language Files* (p. 409). Moreover, Cheng (2000) summarized that, “accents and variations have social, economic, emotional and political implications” (p. 132). On an individual level, an accent can influence a person’s involvement in society. According to the *Los Angeles Times*, “accent reduction students said they are self-conscious about how they sound and whether their accents are limiting their job opportunities or stunting their social lives” (p.2). Accented speech may hinder one’s social and vocational participation which can be the foundation for relationships, perhaps causing isolation in an unfamiliar community.

Although there are cultural rooting benefits of accents, accented speech may impact a person’s involvement in the workforce and their ability to prosper financially (Cheng, 2000, p. 132). Cheng called for SLPs to meet the “needs of the marketplace” (p. 133). In today’s global economy, individuals from various languages and cultures interact on a daily basis. Inevitably, they bring with them a dialect or accent that may be unfamiliar to their colleagues, to varying degrees. Consequently, Fitch (2000) commented that accent modification gives employees “an economic edge” (137). Recently, a study conducted by Hosoda and Stone-Romero (2017) explored the effects of foreign accents on employment-related decisions of 286 college students who spoke Standard American English accent, French with a strong accent and Japanese with a strong accent (p. 119). Participants were 17 to 48 years old and were asked to participate in recorded mock interviews (p. 118). The interviewer was a female native American English speaker who asked common interview questions. American English listeners were asked to review the taped interviews and resume packets, rate suitability for the job and make a decision about hiring the applicants. Listeners decided to hire participants based on accent, understandability, job status and communication demands after mock interviews. Results from this study found Japanese-accented applicants to have the least success in being considered for the position due to their accent. As Hosoda and Stone-Romero (2017) summarized, “they were evaluated more negatively when they applied for jobs that had high communication demands, regardless of job status” (p. 126). The authors conjectured that the negative results were potentially due to stereotypes about the culture in that, “Asians are often stereotyped as being quiet and reserved, lacking communication skills, being good at mathematics and lacking leadership skills” (p. 127). Unfortunately, these stereotypes were evoked when the participant spoke during the interviews. Subsequently, the researchers found certain accents evoked more negative reactions for certain jobs. Hosoda and Stone-Romero summarized, “there’s a hierarchy of preferences among different foreign accents such that a European-accent generally might be favored over an Asian-accent” (p. 127). Thus, accent may impact employment opportunities. Although SLPs cannot necessarily change stereotypes, they can work to improve production accuracy of foreign accents with different accent modification techniques. However, there are few evidence-based approaches to implement for effective therapy. SLPs, therefore, must explore therapy techniques that provide clients with therapy for optimal, fluent communication.

1.4 The Role of the Speech-Language Pathologist

SLPs are the professionals called on to ensure effective communication. As defined by the American Speech-Language-Hearing Association (ASHA, 2016), SLPs remediate communication challenges to help clients reach their full communication potential. Within this definition fall a wide variety of challenges that SLPs can help a client overcome but ultimately, they ensure optimal communication for their clients.

Accent modification falls within the Scope of Practice in Speech-Language Pathology (ASHA, 2016). Accent or dialect modification involves “address[ing] sound pronunciation, stress, rhythm and intonation of speech to enhance effective communication” (p. 11). Accent modification aims at enhancing communication for all speakers in the language of choosing. This includes immigrants electing to become more proficient at the language of their new country or first language natives with stronger dialects wanting more effective communication abilities for their job. Accent modification or remediating the “phonological characteristics of a language variety,” is fairly new to the SLPs’ professional practice (Muller et. al., 2000, p. 119). Such recent practice creates a dialogue and need for further evidence-based practice.

Terms used throughout literature to refer to changing an accent include: accent reduction, modification and management. Alison Behrman (2017) notes the term *reduction* as “bring[ing] the phonological and prosodic features closer to that of a native speaker” and *management* to be “used to encompass a broad range of strategies, including use of global strategies of communication enhancement...as well as traditional goals of reducing segmental and prosodic differences in the [second language]” (p. 1178). For this project, the term *accent modification* will be used in accordance with ASHA standards (n.d.) as well as the concern that the term *accent reduction* conveys eliminating an accent while the term *accent management* assumes maintaining or managing current speech abilities. This study aims to train new speech sounds or modify speech in a new way, rather than to eliminate an entire accent or maintain current abilities. Little evidence-based information is available on the effectiveness of specific accent modification therapy techniques, though there are a number of established approaches.

1.5 Current Accent Modification Treatment Approaches

1.5.1 Compton Approach. One technique available and frequently employed is the Compton Pronouncing English as a Second Language (ESL) Program developed by Dr. Arthur J. Compton in 1984. The goal of the program is to "create new speech habits, so new sounds will be produced automatically" with "50% or greater improvement" (p. 14). The manual provides outlines of hypothetical program schedules, intake forms and various materials with little evidence of beneficial use (Compton, 1984). Sessions are described to be "devoted to…learning to produce troublesome sounds and practicing specific accented sounds in words, phrases and sentences" after an initial analysis of the client's speech and introduction to the program (Compton, 1984, p. 2). Complete-word production, voice projection, short topical presentations, class discussions, role playing, work-related speaking situations, common sentences and phrases and tape-recorded and live conversational speech practice are all incorporated into a group therapy design (p. 2). The Compton Approach aims to include functional activities for non-native American English speakers. Various guiding tips are provided throughout the manual for the certified SLP, linguist or ESL teacher (p. 10). Although the goal of the approach clearly aims at improving accent, there are several limitations. For example, the approach bases training on group therapy and leaves little room for individualization. Current, supportive evidence is not readily available. Additionally, there are several materials to become acclimated to and cited articles are older than 10 years. It can be argued that this approach lacks efficiency and current evidence to support its objective. Moreover, the approach emphasizes pragmatic skills in addition to articulation. It seems that there is far too much that this program aims to modify. While pragmatics are important, a person's intelligibility as a function of speech sound production is not likely impacted by incorporating pragmatics while they are also attempting to learn new speech production patterns.

1.5.2 Articulation and phonological approaches. Traditional articulation and phonological approaches have also been implemented for accent modification. The traditional articulation approach treats a few phonemes that differ in articulation manner whereas the phonological approach treats a whole class of phonemes with the same articulation manner. A study conducted by Schmidt and Meyers (1995) explored effectiveness of both treatments for four male Korean university students (p. 829). Articulation treatment focused on training /s/, /z/, /ʃ/, /tʃ/ and /dʒ/ across 20 sessions with /s/, /ʃ/ and /tʃ/ targeted first until criteria was met for two participants. Treatment focused on describing correct production, details about place and manner of articulation, models and pointing to a picture of the sagittal view of the oral cavity. Articulation treatment increased accuracy of phoneme production (p. 834). The remaining two participants completed phonological treatment for all voiceless fricatives /f/, /ʃ/, /tʃ/, /s/ and / θ/ before treatment for voiced cognates /v/, /z/, /dʒ/, /ʒ/ and /ð/. Treatment focused on descriptions of acoustic characteristics, models, minimal pair drills and reference to a chart of common spellings. One reference to articulation manner was given by explaining sounds as more relaxed to encourage less lip rounding. Similar to the articulation treatment, phonological treatment also succeeded in improving percent accuracy of phoneme production (p. 836). While both treatments provided efficacy for accent modification, there was no comparison of which treatment was most effective. Generalization and maintenance were also not explored. Moreover, researchers commented that individual differences could have been the cause for the results rather than the treatment approach.

Franklin and McDaniel (2016) also studied phonological processes in two native Japanese adults. They examined final consonant deletion, cluster reduction, gliding, stopping, vocalization, prevocalic voicing, epenthesis and final consonant devoicing prevalence. Vocalization, cluster reduction, final consonant devoicing, final consonant deletion and stopping were all present, in that order of prevalence, in both speakers (p. 178). Findings of the study suggested evidence for phonological training in a cycles approach for non-native American English speakers. While it is evident that classes of sounds are produced differently in non-native speakers, there was no evidence that a phonological approach would be more beneficial than another approach that trains differences according to specific phonemes or suprasegmental characteristics.

1.5.3 Segmental and prosodic approaches. Segmental and prosodic approaches are employed to teach suprasegmental aspects of language. These approaches focus on training segmental features of speech, such as syllables or prosodic features, such as pitch, timing and loudness, rather than individual phoneme characteristics. Behrman (2014) compared both of these approaches among four adult native Hindi males. Segmental training focused on auditory stimulation, auditory discrimination training, articulator placement and sound production training with modeling and verbal feedback provided. Therapy worked through increasing levels of complexity, starting at isolation of the targeted phoneme and moving to conversational speech. Prosodic treatment incorporated auditory stimulation, auditory discrimination training and prosodic training with conversational practice, models and feedback. Rise-fall pitch in one-word utterances; rising, falling and rise-fall pitch intonation in three-word utterances; informational and yes/no questions; and prosodic rhythm of longer utterances were targeted with written stimuli, repetition, role-play, verbal tasks, conversations and monologues. Visual and melodic (e.g. tapping) cues were provided with prosodic treatment. Both treatments proved beneficial in improving accuracy but no explicit difference was noted for either approach (p. 556). Although both treatments yielded increased intelligibility, no generalization or maintenance was tested to provide evidence for extended use. Due to the nature of the prosodic treatment, it may be more functional to incorporate some tasks into accent management therapy for functional practice. Moreover, suprasegmental components of languages differ greatly and may be language-dependent. Thus, results from this study should be generalized to other languages with caution (p. 555).

1.5.4 Clear Speech approach. Behrman (2017) determined the effect of clear speech to increase native English speakers' ease of understanding. This study did not decrease "accentedness," rather it determined whether asking participants to speak clearer impacted the ability for listeners to understand Spanish-influenced speech. Findings suggested that there was an improvement in native English speakers' ability to understand foreign accented speech when the participants were asked to use clear speech (p. 555). However, no generalization or maintenance of skill was noted.

Furthermore, Lam and Tjaden (2013) explored clear speech instruction and its effectiveness for improving intelligibility of twelve native English speakers. This approach trained speech by asking speakers to talk as if in hypothetical situations. Speakers were asked to speak habitually, clearly, to over-enunciate and to talk as if to a person who was hearing impaired. Intelligibility percentages of all conditions were determined from 40 listeners' perceptions. The study concluded that asking speakers to over-enunciate produced the most intelligible speech, followed closely by asking to talk to a person with a hearing impairment (p. 1434). Findings suggested that intelligibility of speech can change based on type of instruction.

Of course, results are subject to variability based on individual differences. Generalization to other ages, populations and languages was not examined. Although this approach improved intelligibility for various conditions, there are other aspects of speech, such as articulation differences that heavily contribute to speech more than asking participants to speak with a different mindset.

Moreover, Smiljanic and Bradlow (2007) examined the effect of clear speech instruction provided to four native American English speakers who were judged by sixteen Croatian listeners (p. 2). American English speakers were instructed to read a stimuli sentence as if "talking to someone familiar" and as if they "were talking to a listener with a hearing loss or non-native speaker" (p.1). The results showed that clearer speech yielded greater intelligibility for the non-native listeners (p. 2). The same methods were applied a second time but with four native Croatian speakers and 40 American English listeners. The methods were applied a third time with all native Croatian speakers and listeners. For both the second and third trials, clear speech elicited greater intelligibility (p. 3). This study did not examine generalization or maintenance of the skill. While this study provided evidence that clear speech has an impact on overall intelligibility, it doesn't explore it as a therapy approach by implementing it in several sessions or measuring generalization to other contexts. Clear speech may be more beneficial if a client is limited by the number of sessions for therapy they are able to commit to or as a final remediation suggestion.

1.5.5 Biofeedback approaches. In addition to more traditional speech therapy techniques, biofeedback approaches are also employed as accent modification treatments. These approaches rely on external devices to provide feedback about various speech characteristics of an individual. For example, Brady, Duewer and King (2016) combined spectrogram with traditional articulation therapy to train American English vowels of a single 24-year-old Iranian male (p. 23). One and two syllable words, phrases and sentences were trained with feedback provided via a vowel chart and tongue illustrations. The clinician also began treatment by explaining vowel production in the oral cavity, a vowel quadrilateral to demonstrate characteristics of target sounds and bands on the spectrogram to orient the client to the therapy approach (p. 28). Maintenance was evaluated two weeks following treatment on two separate occasions. Results demonstrated that vowel training was effective with combined traditional and visual feedback (p. 30). Limitations included influence of vowel production in the L1 (p. 31). Thus, results might not be generalized to other languages as similarly. Neither the verbal articulation nor visual spectrogram feedback was separated to determine if either had a more significant influence on the participant's productions (p. 31). However, it is likely that some kind of verbal feedback about phoneme production would be given in addition to visual biofeedback in a realistic setting. Spontaneous speech data also was not collected to determine overall effectiveness and generalization to greater contexts (p. 32). However, it can still be concluded that visual biofeedback along with verbal instruction of production characteristics still resulted in an increase in intelligibility.

Although the previously mentioned studies focused on specific approaches that did not rely on visual feedback specifically, it was still incorporated in some way in several approaches to increase understanding and promote better instruction (Behrman, 2014; Brady, Deuwar, & King, 2016; Compton, 1984; Schmidt & Meyers, 1995). Thus, incorporating visual feedback of the tongue during articulation via ultrasound may also be an effective approach to accent modification. Though numerous studies have demonstrated the effectiveness of ultrasound visual biofeedback for the treatment of targets for individuals with a variety of speech sound disorders (see Table 3 for highlighted studies), there is a paucity of data on the use of this form of visual biofeedback for accent modification services.

Table 3. *Comparison of Accent Modification Approaches*

Citation	**Approach**	**Method(s)**	**Participants**	**Stimuli**	**Visual Feedback**	**Caveats/ Limitations**
Behrman, A. (2017)	Clear Speech	Participants asked to talk as if talking in a noisy environment, talking to a friend across the room or talk as if talking to a person with a hearing impairment	6, native Spanish speakers	25 anomalous phrases	No	No generalization or maintenance noted
Compton, A.J. *Compton P-ESL Program*, (1984).	Compton Approach	Word production, voice projection, short topical presentations, class discussions, role playing, work-related speaking situations, common sentences and phrases and tape-recorded and live conversational speech practice	Groups of ELL adults	Functional and individualized	Yes, pictures and videos	Limited recent evidence, bases therapy on groups, cumbersome materials
Schmidt, A.M., Meyers, K. A. (1995).	Articulation	Description of correct production, details about place and manner, models, pointing to oral cavity picture	2, Korean university students	/s/, /z/, /ʃ/, /tʃ/ and /dʒ/	Yes, picture of sagittal view of oral cavity	No evidence of generalization or maintenance, not distinguished from phonological

						treatment in study
Schmidt, A. M., Meyers, K. A. (1995).	Phonological	Description of acoustic characteristics, models, minimal pairs drills, reference to chart of common spellings	2, Korean university students	/f/, /ʃ/, /tʃ/, /s/ and / θ/	No	No evidence of generalization or maintenance, not distinguished from articulation treatment in study
Franklin, A. & McDaniel, L. (2016).	Phonological	Cycles approach	2, Japanese adults	Final consonant deletion, cluster reduction, gliding, stopping, vocalization, epenthesis, final consonant devoicing	No	Not compared to other approaches
Behrman, A. (2014).	Prosodic	Auditory stimulation, auditory discrimination training, prosodic training	4, adult males, native language: Hindi	Written stimuli, repetition, role-play, models, feedback	Yes, tapping hands with melody	No generalization or maintenance tested, individuals might not be stimulable

Behrman, A. (2014).	Segmental	Auditory stimulation, auditory discrimination training, articulator placement, sound production placement	4, adult males, native language: Hindi	Written stimuli, repetition, role-play, models, feedback	Yes, tapping hands with melody	No generalization or maintenance tested, individuals might not be stimulable
Lam , J. & Tjaden, K. (2013).	Clear Speech	Instructed to read sentences as if talking in different situations (i.e. to someone familiar, listener with hearing loss or non-native speaker)	12 American adult speakers, 40 American Adult listeners	Sample English sentences	No	No generalization to other ages, populations or languages examined
Smiljanic, R. & Bradlow, A. R. (2007).	Clear Speech	Instructed to read sentences as if talking in different situations (i.e. to someone familiar, listener with hearing loss or non-native speaker)	Trial 1: 4 American adults, 16 Croatian listeners Trial 2: 4 native Croatian adult speakers, 40 American adult listeners Trial 3: 4 Croation adult speakers, 4 Croatian adult listeners	Sample English sentence	No	No generalization examined

Brady, K. W., Duewer, N., & King, A. M. (2016).	Spectograph with Articulation	Train with spectrograph biofeedback and verbal articulation feedback	1, Iranian adult male	One and two syllable words, phrases, sentences	Yes, vowel chart and tongue illustrations	Influence of vowel production in the L1, generalization to other languages might not be effective, techniques were not separated

1.5.6 Limitations and caveats. Several current accent modification approaches lack sufficient support (Behrman, 2014; Compton, 1984; Lam & Tjaden, 2013; Smiljanic & Bradlow, 2007). Several of the studies also employed small sample sizes (Behrman, 2014; Brady, Duewer, & King, 2016; Franklin & McDaniel, 2016; Lam & Tjaden, 2013; Schmidt & Meyers, 1995; Smiljanic & Bradlow, 2007). These small sample sizes restrict generalization to larger populations. Moreover, several of the studies also implemented other types of cues or visualizations that confound the specific role of the specific therapy approach relative to the additional cues (Behrman, 2014; Brady, Deuwar, & King, 2016; Compton, 1984; Schmidt & Meyers, 1995) (see Table 3 for comparison of studies).

1.5.7 Extensions. To the knowledge of the researchers, the current study is the first to examine the impact of ultrasound visual biofeedback as an accent modification technique to improve accuracy of American English speech sounds produced by native Mandarin speakers. As mentioned previously, most of the approaches implemented some type of visual feedback or cues, whether it was pictures of the oral cavity or models (Behrman, 2014; Brady, Deuwar, & King, 2016; Compton, 1984; Schmidt & Meyers, 1995). Implementing an approach that relies on visual feedback improves the ability of participants to perceive differences in phoneme production of the therapy in the previously mentioned studies (Behrman, 2014; Brady, Compton, 1984l Deuwar, & King, 2016; Schmidt & Meyers, 1995). Moreover, Schmidt (1997) commented on the earliest accent modification strategies, "before the existence of books, it is likely that second language learners listened to, watched and imitated native speakers…visual methods of training were developed when imitation of a live native speaker was not possible" (p. 1). Following this, visual methods were the natural choice for non-native speakers to use before specific approaches existed. Following results from the previously mentioned studies and what clients historically used on their own, visual approaches should be implemented for optimal comprehension. The following approaches implemented ultrasound visual biofeedback for the remediation of speech sound deficits caused by a variety of diagnoses.

1.6 Roles of Ultrasound Visual Biofeedback in Speech-Language Pathology

1.6.1 Background. Ultrasound imaging relies on high-frequency sound waves emitted by a probe to create an image of the tongue (McAllister Byun, Hitchcock, & Swartz, 2014, p. 2118). Such technique allows clients to "learn visually" by looking at their tongue movements during speech production in real-time. Preston, McCabe, Rivera-Campos, Whittle, Landry and Maas (2014) described ultrasound as a technique in speech intervention that "allows the client and clinician to observe tongue position and shape to directly cue changes in tongue position or shape and to evaluate whether the client has achieved the intended changes" (p. 2102). Moreover, it provides information about articulation properties of various phonemes from two different positions, sagittal and coronal (see Figures 1 and 2), allowing individualization of the technology (Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005, p. 605-606). Ultrasound has been used for a variety of populations and positive outcomes have been noted. Preston, Holliman-Lopez and Leece (2018) noted that ultrasound has been used for the following disorders:

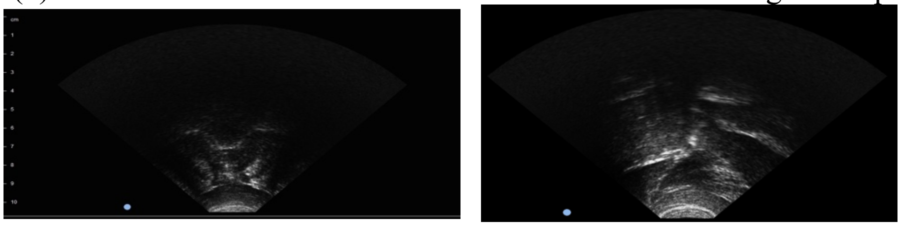
> “[T]hose with “persisting speech sound disorders (Adler-Bock et al., 2014; Bressman, Harper, Zhylich, & Kulkarni, 2016; Cleland, Scobbie, & Wrench, 2015; McAllister Byun et al., 2014; Preston, Leece, & Maas, 2017; Preston et al., 2014; Shawker & Sonies, 1985; Sjolie et al., 2016), Down syndrome (Fawcett, Bacsfalvi, & Bernhardt, 2008), childhood apraxia of speech (Preston, Brick, & Landi, 2013; Preston, Leece, et al., 2016; Preston, Leece, McNamara, & Maas, 2017; Preston, Maas, Whittle, Leece, & McCabe, 2016), hearing impairments (Bacsfalvi, 2010; Bacsfalvi & Bernhardt, 2011; Bacsfalvi, Bernhardt, & Gick, 2007; Bernhardt et al., 2008; Bernhardt, Bacsfalvi, Gick, Radanov, & Williams, 2005; Bernhardt, Gick, et al., 2005; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003), glossectomy (Blyth, McCabe, Madill, & Ballard, 2016), acquired apraxia of speech (Preston and Leaman, 2014) and cleft palate (Cleland, Crampin, Wrench, Zharkova, & Lloyd, 2017)” (p.1-2).

Moreover, in the study that examined 62 participants who had received ultrasound therapy, positive patient satisfaction and few negative side effects were noted (Preston, Holliman-Lopez, & Leece, 2018). Thus, therapy that implements ultrasound technology has positive effects regardless of the population.

Figure 1. Ultrasound Images of /l/ phoneme. This image shows a sample coronal view (a) and a sagittal view (b) with ultrasound visual biofeedback of the American English /l/ phoneme.

(a) *(b)*

Figure 2. Ultrasound Images of /r/ Phoneme. This image shows a sample coronal view (c) and a sagittal view (d) with ultrasound visual biofeedback of the American English /r/ phoneme.



(c) *(d)*

Feedback from the ultrasound is not limited to visual information. Acoustic information can also be extracted with some ultrasound equipment (Berhardt et al., 2005, p. 608). Multiple types of feedback and analyses provide the client with the best information about how their articulators work. Tactile/kinesthetic feedback, such as gestural cues, can also be paired with ultrasound biofeedback for more robust therapy (Preston, Leece, & Maas, 2016). Not only is this detailed feedback unique, but other benefits including less invasiveness, low cost, easy-to-comprehend displays and portability also exist (p. 614-615). These advantages supported the use of ultrasound biofeedback in the following treatment projects (see Table 4 for study highlights). Most frequently and most recently, ultrasound biofeedback was used to increase articulatory precision and articulation abilities of those with childhood apraxia of speech and residual speech sound errors (RSSEs).

1.6.2 Ultrasound and childhood apraxia of speech. There is growing empirical literature base on the role of ultrasound visual biofeedback for the management of childhood apraxia of speech (CAS) (Preston, Brick, & Landi, 2013; Preston, Leece, et al., 2016; Preston, Leece, McNamara, & Maas, 2017; Preston, Maas, Whittle, Leece, & McCabe, 2016). As a recent example, Preston, Leece and Maas (2016) implemented ultrasound visual biofeedback in an intensive speech therapy program for three children between ten and fourteen diagnosed with childhood apraxia of speech (CAS) for remediation of the /r/, /s/ and /tʃ/ phonemes (p. 2). The children participated in two 60-minute sessions per day Monday through Friday for two weeks. Hours were divided into 12-minute segments and every other segment included ultrasound treatment. Preston et. al (2016) stated results from the intensive study “revealed three unique profiles from the three participants” that could be “attributable to a number of factors” (p. 8). These unique profiles were made up of differing severity of diagnosis, stimulability and phonological processing skills. While three different profiles were examined, all participants increased speech production accuracy (p.8).

1.6.3 Ultrasound and residual speech sound errors: Importance of rhotics. Preston, Leece and Maas (2016) also examined the use of ultrasound visual biofeedback in the remediation of RSSEs affecting rhotics (p. 2). Ultrasound biofeedback was paired with principles of motor learning (PML) feedback to determine remediation and generalization of rhotic phonemes in “twelve children aged 10-16 with RSSEs affecting /ɹ/” (p. 6). This study employed an ABACA/ACABA framework to compare PML with and without ultrasound (US) feedback during two phases treating two syllable positions (p. 9). Treatment phases including seven sessions and two sixty-minute sessions were conducted per week. Like the previous study, sessions were divided into time periods with ultrasound therapy provided in every other period (p. 11). Findings suggested that ultrasound feedback resulted in remediation of rhotic phonemes and caused generalization to sentences.

Another study examined retention and generalization of RSSEs affecting rhotics was designed with a similar framework (Preston, McCabe, Rivera-Campos, Whittle, Landry, & Maas, 2014). This study included PML with and without ultrasound therapy. This study also concluded that ultrasound biofeedback is effective for remediation of RSSEs affecting rhotics. However, results were likely due to several approaches being implemented. Still, the authors concluded that the treatment, which included ultrasound visual biofeedback, resulted in the remediation and generalization of rhotic phonemes.

1.6.4 Ultrasound and examination of vowels. While an investigation of ultrasound visual biofeedback for vowel targets has yet to be conducted, ultrasound has been used to analyze vowel characteristics (Georgeton, Antolik, & Fougeron, 2016). Four native French speakers between 25 and 40 read twelve sentences that contained the one of the French vowels /i, e, ɛ, a, u, o/ in two prosodic conditions (p. 1577). Both visual and acoustic analyses were conducted (p. 1578-1579). The researchers found a correlation between "prosodic structuring" and "phonetic properties" of the vowels examined because of the implementation of ultrasound biofeedback (p. 1583). This study provides support for future investigations not only using ultrasound biofeedback for assessment purposes, but also for intervention.

1.6.5 Ultrasound and accent modification. The majority of research on ultrasound biofeedback focuses on disordered speech. More recently, ultrasound visual biofeedback has been implemented to manage accented speech. However, little research with this population exists. Tsui (2012) wrote on this topic, "research in the use of ultrasound with English L2 is sparse" (p. 26). Gick, Bernhardt, Bacsfalvi and Wilson (2008) concurred by stating, "[p]ossible applications of ultrasound to second language (L2) acquisition are only now beginning to be explored" (p. 309). Although novel, such research helps in studying second languages. For example, ultrasound has been used to examine articulation and tongue movement across various languages (Georgeton, Antolik, & Fougeron, 2016; Boyce, Hamilton, & Rivera-Campos, 2016). Additionally, there are two studies that employed ultrasound visual biofeedback with the aim of shaping nonmainstream American English accents by ELLs.

A one-session pilot study by Gick et al. (2008) found that ultrasound visual biofeedback useful in teaching three native Japanese linguistic students accurate production of American English phonemes /r/ and /l/ in word-initial, medial and final positions of CV, CVC or CVCV syllable shapes (p. 317). Word lists were randomized and repeated ten times within the carrier phrase "See X be" (p. 317). Unlike previous studies, this one conducted one 60-minute session consisting of pre and post-training recordings of the phonemes with the ultrasound with 30 minutes of training when participants compared videos of their productions (p. 317-318). After training, "all three participants were able to produce their target approximant successfully" (p. 319). While the three participants already understood language differences as linguistic students, accuracy of phoneme production still improved after only 60 minutes. These results sparked further investigation.

Tsui (2012) expanded upon the findings of this pilot study by "investigat[ing] the effectiveness of using two-dimensional tongue ultrasound to teach pronunciation of [/l/ and /ɹ/] to six adult native Japanese speakers" (p. ii). These phonemes were chosen as dependent variables due to the articulatory and acoustic challenges they present for Japanese speakers. Tsui (2012) stated: "The phonological inventory or the Japanese language does not contain the equivalent of English /l/ or /r/" (p. 2). This is similar to the Mandarin phonetic inventory. Four 45-minute sessions were conducted across two weeks and included initial, medial, final and cluster positions of 44 different words (p.ii, 34). Words were embedded in carrier phrases "I want to see____" and "I want to see ___ be" (p. 35). Analysis of change from pre-treatment and post-treatment was determined by perceived accuracy from novel listeners, acoustic analysis using a spectrogram and visual analysis of ultrasound images. At the end of the study "all participants, who were typical language learners, increased their accuracy of producing English

/l/ and /ɹ/ in a variety of word positions and phonetic contexts" (p. 76). Home practice was also encouraged in this study to augment results. This study was effective in examining the differences in an Asian phonetic inventory without the liquid /ɾ/ and /l/. These results provided a foundation for the present study.

The current study aims to extend these results (p. 76) to the speech production of ELLs who speak Mandarin as their first language. To reiterate, ultrasound was beneficial in providing treatment for several diagnoses (Preston, Leece, & Maas, 2016; Preston, Leece, & Maas, 2016; Georgeton, Antolik, & Fougeron, 2016; Gick, Bernhardt, Bacsfalvi, & Wilson, 2008; Tsui, 2012). This technique was also beneficial in analyzing differences of languages of various phonetic complexities (Georgeton, Antolik, & Fougeron, 2016). Ultrasound biofeedback also helped train phonetic differences in second languages with different phonetic inventories (Gick, Bernhardt, Bacsfalvi, & Wilson, 2008; Tsui, 2012). Following these findings, this study aimed to build upon this evidence to treat speech differences stemming from the phonetic inventory differences between Mandarin and American English previously mentioned (see Table 4 for comparison of studies mentioned).

Table 4. *Comparison of Ultrasound Biofeedback Approaches*

Citation	Etiology	Participants	Design	Methods	Caveats/Limitations
Georgeton, L., Antolik, T. K., & Fougeron C. (2016).	Examination of vowels	4, 25-40 years		Read 12 sentences containing target vowels in 2 prosodic conditions	Small sample

Gick, B., Berhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008).	Pilot study: Teaching American English phonemes	3, adult linguistic university students	1 60-minute session	/r/ and /l/ phonemes presented in word-initial, medial and final positions of CV, CVC and CVCV shapes, word lists randomized and repeated 10 times	Speaker bias due to linguistics degree, short session, small sample
Preston, J., Leece, M. C., & Maas, E (2016).	Apraxia of Speech	3, aged 10-14 years	60 minute sessions, twice daily, Monday-Friday, 2 weeks, ABAB session design	Ultrasound administered, withdrawn, re-administered	Intensive study, small sample
Preston, J., Leece, M. C., & Maas, E. (2016).	RSSE: rhotics	12, aged 10-16 years	ABACA/ACABA	Ultrasound paired or withdrawn, with PML	Small sample
Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014).	RSSE: rhotics	8, children aged 10+		Ultrasound paired with PML	Several approaches likely influenced results, small sample
Tsui, H. M-L. (2012).	Teaching American English phonemes	6, adult native Japanese speakers	4, 45-minute sessions for 2 weeks	/r/ and /l/ phonemes in initial, medial, final and cluster positions in 44 words within carrier phrases	Small sample

1.7 Purpose and Hypothesis

This study proposes that implementation of ultrasound biofeedback will provide a means to resolving speech production difficulties of ELLS who seek services from SLPs. It aims to add to literature of ultrasound biofeedback, to continue investigation of accent modification services and to answer the following questions:

1. Does ultrasound visual biofeedback therapy improve speech sound accuracy of American English phonemes produced by native adult Mandarin speakers?

2. Does this treatment result in generalization to untreated targets?
3. Is maintenance of production accuracy evident after treatment is discontinued?

Chapter 2: Methods

2.1 Design

A single-subject ABAB withdrawal design across multiple baselines design was employed to determine the effect of ultrasound biofeedback on accent modification across sixteen sessions for two participants. Each phase consisted for four sessions. Phase A included baseline (A_1) and withdrawal (A_2) sessions during which no therapy with ultrasound biofeedback or verbal feedback was provided. Phase B sessions employed ultrasound biofeedback and verbal feedback as therapy (B_1 and B_2). Following this design, baseline data was gathered, treatment administered, treatment withdrawn and treatment re-administered. Participants returned six weeks after treatment ceased to assess maintenance of phoneme production accuracy.

2.2 Participants

Two participants who spoke Mandarin as their first language and American English as their second language were recruited at Duquesne University through the English as a Second Language Department. Recruitment occurred in the form of flyers and an email distributed to individuals in the university community. The investigator, a second-year graduate student, visited three English as a Second Language classrooms to explain the study and pass out flyers. Participants were instructed to email the graduate student if interested.

Two participants, referred to as participant 1 and participant 2, male and female, respectively, were enrolled in this study. These participants were selected from a total of four respondents due to self-reported accented phonemes of concern that were internal to the oral cavity and could be targeted with ultrasound equipment. Participant 1 was 23 years old and participant 2 was 30 years old. Neither participant spoke a third language, however participant 2 spoke an additional Chinese dialect. Neither received accent modification therapy in the past. Neither reported a having a history of hearing, neurological, speech or language deficits while in either China or the United States.

2.3 Equipment, Materials and Examiners

The primary instrumentation was the ultrasound device and supporting laptop. An Interson PI 7.5 MzH ultrasound transducer was placed under the mandible at the base of the tongue to transduce sonic waves through a small amount of Aquasonic gel on the probe during production of the phonemes similar to a study conducted by Preston et al. (2013). The ultrasound was connected to a Dell Latitude laptop with a 13-inch screen. Participants sat in front of the laptop to see the oral cavity display using SeeMore software (p. 3). The participants held the ultrasound probe during treatment sessions after initial orientation by the clinician. Participants sat approximately 18 inches from the screen. Distance between screen and participant were consistent during all screening, diagnostic and treatment sessions. All sessions were video- and audio-recorded. Audio-recordings were completed using Audacity 2.0 software via a head worn Micro Mic C 520 condenser microphone and modulated with a pre-amplifier (PreSonus Audio Box 22VSL) during all sessions. The head worn mic was approximately one inch from the participants' mouths. Participants also completed perceptual training exercises with recorded audio files at the beginning of each session. Audio files were inserted into a Powerpoint that was viewed using the laptop equipment.

Probe lists of American English words containing 10 words of two target phonemes were made. The probe lists were the same across all baseline and treatment sessions. Targets were selected after the initial screening/diagnostic session. Probes for participant 1 targeted final /r/

and medial /ɪ/ in multisyllabic words as well as medial /ɛ/ in monosyllabic and multisyllabic words. Probes for participant 2 targeted final /r/ and medial /ɪ/ in monosyllabic words as well as medial /ɛ/ in monosyllabic and multisyllabic words. The third targets were introduced after 75% accuracy for both targets during treatment Phase 1 and 2 sessions of Baseline Phase 2 were completed. These targets were probed during the remaining two sessions of Baseline Phase 2 and treated during Treatment Phase 2. In addition, after 75% accuracy of the initial two target phonemes was reached, more complex probes with the same targets were introduced for treatment. An additional list of untreated words with the target sounds were probed each session to determine generalization to untreated contexts. Carrier phrases containing probe words were also probed during the second half of Baseline Phase 2 and Treatment Phase 2 to determine generalization to more challenging contexts.

Similar to a study conducted by Adler-Bock, Bernhardt, Gick, & Bacsfalvi (2007), probe lists were presented randomly each session via PowerPoint on a laptop without verbal pronunciation from the clinician to diminish practice and retesting effects. Target word font remained the same throughout all sessions.

A second-year professional phase graduate student administered all screening, diagnostic and treatment sessions. A speech-language pathology faculty member oversaw and approved administration of all procedures. All data and assessment measures were recorded using participant numbers to de-identify information.

2.4 Procedures

All procedures were completed at the Duquesne University Speech-Language-Hearing Clinic in available treatment rooms. The same room was utilized for all screening, diagnostic and treatment procedures when available to diminish the effect of different testing environments. A different room was used once during Baseline Phase 1 session 3 and for the maintenance session for participant 2 due to scheduling conflicts. There is no reason that the room change should have impacted results. All procedures remained consistent. In the event of absences, participants were asked to record the probes on their own devices and email them to the graduate student. This only occurred twice during Baseline Phase 2 for both participants, once during Treatment Phase 2 when the recording equipment malfunctioned for participant 1 and once at the end of the maintenance session for participant 1 due to time constraints.

2.4.1 Screening and diagnostic procedures. Similar to the study conducted by Tsui (2012), participants passed hearing and vision screenings to be eligible. A number of assessment tools for speech production were used to guide target selection and to provide background information about the participants. Screening and diagnostic procedures took place during a 90-minute session for 3 participants. After diagnostic procedures, 2 participants were selected for based on assessment results.

The protocol and participant requirements were reviewed, English Language Experience surveys completed and informed consent documents signed. Voluntary participation and ability to end participation in the study was explained to each participant. Informed consent documents were kept in the Duquesne University Speech and Gesture Lab. All screening and diagnostic forms were de-identified for participant privacy. Hearing screenings were conducted using a pure-tone audiometer at 20 dB at 1000, 2000 and 4000 Hz. No hearing deficits were noted at the time of the screening.

Similar to the study conducted by Tsui (2012), participants completed a questionnaire about age of exposure to American English, length of time living in an English-speaking country, formal American English instruction, motivation to participate and self-rating of speech accuracy for chosen phonemes (p. 32) (See Appendix A). These results provided qualitative information about expected characteristics of their accent. For example, a later age of acquisition and less time living in an English-speaking country likely causes a stronger accent. Results of the questionnaire indicated that neither had been treated by an audiologist and consequently did not have concerns about their hearing. Both participants spoke Mandarin as their first language and English as their second language without a third language. Neither had been evaluated by an SLP at any time. Participant 1 was concerned about her English pronunciation but participant 2 had no concerns about his speech production or American English intelligibility. Neither had a history of medical, developmental or neurological problems. Participant 1 lived in an English-speaking country for 1 month. Participant 2 lived in the United States for 1 year. Both were first exposed to the English language at school in China; participant 1 at age 6 and participant 2 at age 11. Participant 1 was first immersed in an English-speaking environment, Pennsylvania, one year prior. Participant 2 had been immersed in an English-speaking country (Ghana) 4 years prior, where he worked for a year. Participant 1 had received instruction in English pronunciation at school for 2 years but participant 2 had not. Participant 1 rated that she spoke English 25% of her life daily whereas participant 2 spoke English during about 90% of his day. Both spoke English most often in school. The sounds “a” and “i” were rated as the easiest English sounds for participant 1 and “e” and voiced “th” were rated as the most challenging. Participant 2 rated the sounds “ing” and voiced and voiceless “th” as most challenging. He did not note easy sounds. Both were very motivated to participate in the study and both rated their English pronunciation as average, neither poor nor excellent.

Vision screenings were completed by asking the participants to identify objects on both static and dynamic ultrasound images after an initial orientation to the ultrasound. Participants were given a brief introduction to the ultrasound equipment to determine stimulability for use of the device. The graduate student explained pictures of the ultrasound while modeling use before handing the equipment to the participants and asking them to identify key points (e.g. top of tongue). No misunderstandings were detected during this screening. Both participants were judged to be appropriate clients and stimulable for ultrasound use. Preliminary pictures were taken of participants saying extended phonemes to obtain baseline measures for tongue placement.

Segmental and suprasegmental characteristics of speech were analyzed in a variety of ways to gain a comprehensive understanding of the participants' individual speech patterns and to guide target selection. Schmidt (1997) noted, "a good foreign accent assessment will offer a chance for the speaker to produce the sounds of English in contexts that help the clinician to see any patterns of differences that might occur" (p.5). Moreover, Sikorski (2005) suggested a valid assessment of foreign accent also include valid assessments of articulation, pitch variability, speech rate and stress.

A five-minute spontaneous speech sample was elicited at the beginning of the sessions as an informal assessment. Speech sound patterns in the participants' natural, conversational speech as well as prosodic characteristics were analyzed. Following recommendations by Sikorski (2005), the *Prosody-Voice Screening Profile (PVSP)* (Shriberg, Kwiatkowski, & Rasmussen, 1990) was administered to assess the prosodic characteristics (e.g., rate, prosodic stress, pitch, loudness, dysfluencies/hesitations, etc.) of the participants' speech. This assessment was proven beneficial in analyzing prosodic variations of several adult populations (McSweeny & Shriberg, 2001). Prosodic patterns helped determine overall speech differences in the participants. Due differences in rhythm, tone, stress and other speech patterns between Mandarin and American

English, prosodic differences were expected. The *PVSP* provided a systematic way to assess these potential differences by analyzing 24 utterances from the spontaneous speech sample.

Spontaneous speech samples were elicited by asking participants to talk about their homes in China. The *Prosody-Voice Screening Profile* (Shriberg, Kwiatowski, & Rasmussen, 1990) was used to systematically assess the participants' prosodic, voice, fluency and resonance characteristics. According to the guidelines of the PVSP, a total of 24 utterances from the spontaneous speech sample were coded for a total of 32 codes across 7 different parameters (i.e., phrasing/fluency, rate, stress, loudness, pitch, laryngeal quality and resonance quality). According to the *PVSP*, 20% or more utterances contain inappropriate prosody, voice or resonance features, the individual demonstrates challenges with prosodic and vocal characteristics and may warrant further management of these features. A total of 24 of 24 (100%) of participant 1's utterances were coded as inappropriate. The specific codes and parameters of concern are noted below. Overall, participant 1 did exhibit concerns with stress, phrasing and loudness (see Table 5). Errors were characteristic of Mandarin influence (i.e. different stress patterns, processing/rewording). Participant 1 also exhibited an increase in loudness and rate as sessions continued, likely due to comfort with the examiner. Errors were not targeted specifically in treatment sessions.

Table 5. Participant 1 *PVSP* Scores

	Number of Coded as Inappropriate	*Parameter of Concern (>20%)*	*Specific codes frequently used for each parameter*
Rate	3 (13%)		Slow articulation/pause time (5x)
Stress	16 (67%)	X	Reduced/equal stress (12x) Excessive/equal/misplaced stress (4x)

Phrasing	9 (38%)	X	Word repetition (2x) One word revision (1x) Repetition and revision (6x) Errors likely due to L2 processing/rewording rather than dysfluencies
Loudness	14 (58%)	X	Soft (14x)

For participant 2, a total of 23 of 24 (96%) utterances were coded as inappropriate. The specific codes and parameters of concern are noted below (see Table 6). Overall, participant 2 did exhibit atypical manners of stress, phrasing and loudness. Similar to participant 1, errors were characteristic of Mandarin influence (i.e. different stress patterns, processing/rewording) and were not targeted specifically in treatment sessions.

Table 6. Participant 2 *PVSP* Scores

Parameter	*Number of Coded as Inappropriate*	*Parameter of Concern (>20%)*	*Specific codes frequently used for each parameter*
Rate	0 (0%)		N/A
Stress	16 (67%)	X	Reduced/equal stress (2x) Excessive/equal/misplaced stress (14x)
Phrasing	11 (46%)	X	Sound/syllable repetition (3x) Word repetition (5x) One word revision (1x) More than one One word revision (1x) Repetition and revision (1x) Errors likely due to L2 processing/rewording rather than dysfluencies
Loudness	14 (58%)	X	Soft (14x)

Similar to Hack, Marinova-Todd and Bernhardt (2012), a standardized articulation assessment was administered to determine the participants' speech sound skills as related to normative data for their gender and age and to analyze speech sound differences. The *Photo Articulation Test (PAT)* (Lippke, Dickey, Selmar, & Soder, 1997) was given to determine articulation patterns of the participants at the word level. English vocabulary deficits were noted for several words. In these cases, the graduate student said the target word. Participant 1 obtained a raw score of 19 errors which was converted to a standard score of 62. Errors noted and relevant to this study included medial /ɛ/ and final /ɾ/. Participant 2 obtained a raw score of 21 which converted to a standard score of less than 60. Errors noted and relevant to this study included medial /ɪ/, final /ɾ/, /ɾ/ blends and medial /ɛ/ (See Table 7).

The *Assessment of Intelligibility of Dysarthric Speech* (Beukelman & Yorkston, 1984) was administered to determine intelligibility of various phonemes in addition to speaking rate in the event that the participant's intelligibility was perceived to be low. This procedure followed similar intelligibility assessment procedures from Fritz and Sikorski (2013). If the participants' intelligibility was perceived to be high, this assessment was not be administered. This assessment was only administered for the 2nd participant because intelligibility was judged to be at about 70%. Two naïve listeners were asked to write down word and sentence level utterances. Written responses were compared to what the participant read. Accuracy was averaged between the two listeners. Word level intelligibility was judged to be 51% and sentence level intelligibility was judged to be 89%. Word level intelligibility was likely lower due to the words being out of context (See Table 7).

Following similar frameworks from Lam and Tjaden (2013), Fritz and Sikorski (2013), a passage reading was administered to determine articulation patterns at the sentence level and to determine percent consonants correct (PCC) per recommendations by Schmidt (1997) and Sikorski (2005) and following implementation by McAllister Byun and Hitchcock (2012) and Morton, Brundage and Hancock (2010). *The Caterpillar Passage* (Patel, Connaghan, Franco, Edsall, Forgit, Olsen et. al, 2013) was selected because of its incorporation of prosodic contrasts and words of varying length and complexity as well as its contemporary theme. Patterns from this assessment augmented findings from the spontaneous speech and standardized assessments. PCC for participant 1 was 93%. Errors noted and relevant to the study were i/ɪ substitution, labialized and distorted /r/ in all positions and distorted vowels. Participant 2 had 87% PCC. Errors noted and relevant to this study included i/ɛ substitution, i/ɪ substitution and omitted or distorted /r/ in all positions (See Table 7).

The *Peabody Picture Vocabulary Test* (Dunn & Dunn, 2007) was administered to determine proficiency in English vocabulary. This assessment followed a similar receptive language procedure administered by Morton, Brundage and Hancock (2010). This test assessed vocabulary proficiency in American English relative to normative data for the participant's age and gender. Participant 1 obtained a raw score of 102 converted to a standard score of 38. Participant 2 obtained a raw score of 123 converted to a standard score of 48 (See Table 7). Target words were selected based on receptive vocabulary abilities. Although the scores are indicative of lower English receptive vocabulary, these abilities likely did not impede comprehension of the study. Target words were based on English receptive vocabulary abilities.

Table 7. *Participant Diagnostic Scores*

PAT	**Standard Score**	*1*	60	**Errors Noted**	*1*	Medial /ɛ/, final /r/
		2	<60		*2*	Medial /ɛ/, final /r/, /r/ blends, medial /ɪ/
AIDS	**Word Level Intelligibility**	*1*	51%	**Sentence Level Intelligibility**		89%
		2	N/A			N/A
Caterpillar Passage	**Percent Consonants Correct**	*1*	93%	**Errors Noted**	*1*	i/ɪ substitution, labialized and derhoticized /r/, vowel distortions
		2	87%		*2*	i/ɛ substitution, i/ɪ substitution, omitted and distorted /r/
PPVT	**Standard Score**	*1*	38			
		2	48			

Following, Behrman (2014) and Fritz and Sikorski (2013) the *Proficiency in Oral English Communication Screening (POEC)* (Sikorski, 2007) was administered due to its high validity for assessing foreign accent as noted by Morton, Brundage and Hancock (2010). Moreover, it was recommended via personal correspondence (May 30, 2017) by Dr. Alison Behrman, specifically for its ability to assess prosody in accented speech. The *POEC* (Sikorski, 2007) was not administered during the first screening and diagnostic session because it was not yet obtained at the time. It was administered during the following diagnostic and first baseline session. Subtests II, III, V and VI were given. Other subtests were omitted due to already obtaining similar data (e.g. single word level utterances) during previously administered tests. Participant 1 had 15% falling pitch contour errors, 0% rising pitch contour errors and 38% total stress errors during lengthier messages. Participant 1 also had 13% total stress errors during the contrastive intonation subtest. Participant 1 had 13% listening errors and 5% delayed responses during the auditory discrimination subtest. Participant 2 had a total of 46% falling pitch contour errors, 50% (1 out of 2) rising pitch contour errors and 19% total stress errors for lengthier messages. Participant 2 also had 7% total stress errors for contrastive intonation. Participant 2 had 5% listening errors with 10% delayed responses during the auditory discrimination subtest (See Table 8).

Table 8. *Participant POEC Scores*

	Falling Pitch Contour	**Rising Pitch Contour**	**Total Stress Errors in Lengthier Messages**	**Total Stress Errors in Contrastive Intonation**	**Listening Errors**	**Delayed Responses**
1	15%	0%	38%	13%	13%	5%
2	46%	50%	19%	7%	5%	10%

2.4.2 Target selection. Target phonemes were selected based on patterns from the various assessments. The three least accurately produced phonemes that were produced within the oral cavity and therefore appropriate for ultrasound treatment, were selected. Targets for participant 1 included multisyllabic final /ɾ/ and medial /ɪ/ words. Participant 2 targets included monosyllabic final /ɾ/ and medial /ɪ/ words. Both participants also had medial /ɛ/ targets in monosyllabic and multisyllabic contexts. Probe lists for each target were made. The master probe list was made up of 15 words for each target and randomly divided into five words to be treated during ultrasound treatment and five words to be treated during withdrawal sessions (see Appendix); 5 words treated and probed, 5 words probed but not treated to measure for generalization and 5 words treated but not probed. Words were randomly selected and divided into subsets within probe lists.

After Baseline Phase 2 was completed and at least 75% accuracy was reached for the initial target phonemes, carrier phrases with the original target words were probed to measure generalization. Additional more complex words for both targets were also introduced for both probing and treatment as well as 1-3 syllable medial /ɛ/ words.

2.4.3 Experimental procedures. Experimental procedures paralleled those conducted by Sjolie, Leece and Preston (2016). Given the ABAB withdrawal design, the experimental procedures consisted of sessions both with and without intervention. First, four baseline sessions were completed, followed by four treatment sessions, then four sessions with treatment withdrawn, followed by another four treatment sessions. After six weeks, the participants returned for a final maintenance session without treatment procedures.

All treatment sessions were 30 minutes and took place one to two times a week. The same speech-language pathology graduate student administered treatment under the supervision of a certified speech-language pathologist. Both treatment and withdrawal sessions began with administration of the generalization and training probes. Similar to Tsui (2012) words were presented in a random order on a PowerPoint during each session. During treatment sessions, the probe list was evaluated at the end after treatment with selected targets to determine acquisition of targets. Probes started at the identified target level and were replaced with more complex words when the participants reached greater than 75% accuracy of production.

The treatment phase sessions were divided into three time periods of treatment and a timer was used to be sure that the timing was adhered to. All withdrawal and treatment sessions began with perceptual training (5 minutes) during time period A. Next, during treatment sessions, time period B included treatment with the ultrasound for 20 minutes. Drill-like therapy was implemented and any amount or type of cueing was allowed during this time period. Finally, the last 5 minutes were spent probing treated and untreated words for future analysis (time period C). All probe list items were read three times with the ultrasound equipment but no verbal feedback was given.

2.4.4 Initial baseline phase (Baseline$_1$). Instructions regarding how to use the ultrasound were reiterated at the beginning of the initial baseline session. Perceptual training took place by asking participants to determine whether target phonemes were pronounced correctly or incorrectly in two consecutive words. Two pairs of correct and incorrect targets in words not included in the probe list were played to show the participant both correct and incorrect productions. Participants were instructed to listen for the target sound in the following pairs. Five pairs of each target phoneme/context were administered (i.e. 10 pairs altogether). The examiner replayed the recording if the volume was initially too low or if the participant asked for a repetition. No models were given by the examiner. Participants were rated on their ability to discriminate incorrect and correct productions.

Stimulability probes were administered three times to characterize pre-treatment accuracy. Participants were instructed to read the words naturally. No models were given.

2.4.5 Treatment phase I (Treatment$_1$). Five words randomly selected for treatment with the ultrasound and an additional set of five words with the target sound that were not in the probe list were targeted during this phase. Data was recorded on number of probes and target words produced each session (see Appendices C and D). During time period B, both verbal and visual feedback was given. Sagittal and/or coronal views were used at the discretion of the clinician. Following Sjolie, Leece and Preston (2016), cues were based on the participant's accuracy of constricting the anterior tongue (e.g., "lift the back of your tongue"), lateral elevation of the sides of the tongue (e.g., "lift the sides of your tongue") and inhibit incorrect movement (e.g. "keep the body of your tongue down"). Similar to McAllister Byun and Hitchcock (2012) and Preston et. al (2013), only ultrasound visual biofeedback and verbal articulation feedback was given during treatment sessions.

2.4.5.1 Verbal feedback. Treatment incorporated verbal feedback was based on principles of motor learning described by Maas, Robin, Austermann Hula, Freedman, Wulf, Ballard, et al. (2008). Principles of motor learning are based on the thought that speech movements require similar skills needed for gross motor movements. Maas et al. (2008) stated: "learning cannot be directly observed but rather must be inferred from changes in performance over time" (p.278). Change in performance results from improving capability for the learned skill. One of the ways to encourage understanding of change in performance and consequently influence capability is to provide verbal feedback. Maas et al. (2008) described two types of verbal feedback that were implemented to encourage improvement of motor skills during this study: knowledge of performance (KP) and knowledge of results (KR). Maas et al. (2008) described knowledge of performance feedback: “the nature or quality, of the movement pattern” (p. 288). In contrast, they define knowledge of results as “information about the movement outcome” (p. 288). Thus, KP feedback was given early in treatment to enhance understanding and production of correct motor skills (e.g. “pull your tongue back”) before switching to KR feedback when the movement became more learned and specific information was not needed. Both KP and KR verbal feedback are equally effective (Maas et al., 2008).

2.4.5.2 Perceptual training. Based on the theories of underlying mechanisms that influence accent mentioned previously and Van Riper’s Complexity Staircase Model (1996), perceptual training was implemented during the beginning five minutes of each session to increase understanding of articulation differences between the languages and correct auditory perception of target phonemes. As a “warm-up,” auditory bombardment occurred in the form of negative practice by asking participants to identify correct and incorrect productions of 10 target phoneme pairs. Pairs were randomized each session to control for retesting effects. Specific verbal feedback was given concerning accuracy of perception. Participants did not produce target phonemes during this time.

To complete these measures, six individuals who spoke Mandarin as their first language from (4 female and 2 male) were asked to read 40 words related to the target phonemes (i.e. “poor,” “give”) in a 15-minute session (see Appendix B). The examiner explained that they would not be judged on accuracy of pronunciation. Words paralleled the complexity and context of targeted phonemes (i.e. monosyllabic final /r/ words). No models were given. Students were also asked to pronounce the “sound /r/ makes” and the “sound I makes” in addition to reading through short vocalic /r/ words twice. Models were only given when asking students to pronounce the sound /r/ makes due to the students pronouncing the letter instead of the sound.

Perceptual training recordings were broken into words and saved as audio files for every individual word. Not all words were saved from original recordings. Words that included incorrect productions of more than one phoneme that were not the target phoneme were not included. Only productions with incorrect or correct productions of only the target phoneme were saved to decrease confusion and minimize effect of hearing other incorrect phonemes. Perceptual training pairs were randomized each week so that no words produced by the same speakers were heard twice. The number of incorrect and correct pronunciations also varied each week according to the audio files selected for the pairs (i.e. 8 correct, 12 incorrect).

2.4.6 Withdrawal phase (Baseline$_2$). Following the first phase of intervention, ultrasound visual biofeedback and other treatment strategies were removed. The procedures and time allotment were identical to the initial baseline phase. Both participants reached 75% accuracy with original probe words. To continue treatment and determine generalization to more challenging contexts, more complex probe words were selected. These words were probed during the second baseline phase. Original probe words were still probed during the remainder of the study to determine maintenance and probed in addition to the original probe words during the second baseline phase. In addition original probe words were placed in the carrier phrase "say___ again" similar to a study analyzing vowels conducted by Chen, Robb, Gilbert and Lerman (2001). Generalization from the word level to the sentence level was determined by rating accuracy of probe words in a carrier phrase with a continuum scale.

Due to schedule conflicts, three Baseline$_2$ sessions were completed for the second participant and two Baseline$_2$ sessions were completed in the clinic for the first participant. The remaining baseline sessions for this phase were recorded at home with cell phone recording applications and emailed to the examiner. Thus, a total of four Baseline$_2$ sessions were completed. Ultrasound pictures of tongue placement were taken only during the first baseline session of this phase.

2.4.7 Treatment phase II (Treatment$_2$). Procedures for Treatment$_1$ were replicated for Treatment$_2$. Based on treatment from the first phase and level of accuracy reached, treatment for the first two targets ceased but were still probed to analyze retention. More complex phonetic environments for both targets in addition to a third target, /ɛ/, were introduced for treatment. Carrier phrases were also probed to measure for generalization. Verbal feedback continued to be implemented with ultrasound biofeedback.

2.4.8 Maintenance session. A one-hour follow-up session was conducted to determine maintenance and generalization six weeks after treatment sessions end. A spontaneous speech sample, *The Caterpillar Passage* (Patel, Connaghan, Franco, Edsall, Forgit, Olsen et. al, 2013), *PAT* (Lippke, Dickey, Selmar, & Soder, 1997)*, POEC* (Sikorski, 2007) *and PVSP* (Shriberg, Kwiatowski, & Rasmussen, 1990) were re-administered to assess potential change in segmental and suprasegmental skills of the participants. A new list with the target phonemes in varying word lengths and contexts were read three times by the participants to measure generalization. Identical probe words administered in treatment sessions were read in random order three times to determine maintenance of skills learned during the study.

Additionally, spontaneous speech samples were elicited to note general errors and intelligibility as well as to assess prosodic, voice, fluency and resonance characteristics with the *PVSP* (Shriberg, Kwiatowski, & Rasmussen, 1990). These results were compared to diagnostic session results. A total of 14 of 24 (58%) of participant 1's utterances were coded as inappropriate, a 42% decrease from the diagnostic session. The specific codes and parameters of concern are noted below. Overall, participant 1 did exhibit concerns with stress and phrasing.

Stress and phrasing errors were characteristic of Mandarin influence (i.e. different stress patterns, processing/rewording). Participant 1's decrease in loudness errors was likely due to comfort with the examiner. Errors were not targeted specifically in treatment sessions. However, decrease in percentages can be correlated with the overall increase in intelligibility noted after treatment (see Table 9).

Table 9. *Participant 1 PVSP Scores Maintenance Session*

Parameter	*Number of Coded as Inappropriate*	*Parameter of Concern (>20%)*	*Specific codes frequently used for each parameter*
Rate	0 (0%, 13% decrease from diagnostic)		N/A
Stress	12 (50%, 17% decrease from diagnostic)	X	Excessive/equal/misplaced stress (12x)
Phrasing	8 (33%, 5% decrease from diagnostic)	X	Word repetition (6x) One Word Revision (2x) Errors likely due to L2 processing/rewording rather than dysfluencies
Loudness	0 (0%, 58% decrease from diagnostic)	X	N/A

A total of 19 of 24 (79%) of participant 2's utterances were coded as inappropriate, a 17% decrease from diagnostic measures. The specific codes and parameters of concern are noted below. Overall, participant 2 exhibited atypical manners of stress and phrasing. These percentages were maintained from the diagnostic session. Errors were characteristic of Mandarin influence (i.e. different stress patterns, processing/rewording) and were not targeted specifically in treatment sessions. Maintenance of errors was possibly due at least in part to the fact that no explicit instruction of these parameters was provided. The decrease in loudness errors is likely due to the participant being cued to lower his voice throughout treatment sessions (see Table 10).

Table 10. *Participant 2 PVSP Scores Maintenance Session*

Parameter	*Number of Coded as Inappropriate*	*Parameter of Concern (>20%)*	*Specific codes frequently used for each parameter*
Rate	0 (0%, maintained from diagnostic session)		N/A
Stress	16 (67%, maintained from diagnostic session)	X	Excessive/equal/misplaced stress (16x)
Phrasing	10 (42%, 4% decrease from diagnostic session)	X	Sound/syllable repetition (1x) Word repetition (4x) One word revision (1x) More than one word revision (2x) Repetition and revision (2x) Errors likely due to L2 processing/rewording rather than dysfluencies
Loudness	0 (0%, 58% decrease from diagnostic)		N/A

The *PAT* assessment was re-administered to determine changes in articulation abilities at the word level (Lippke, Dickey, Selmar, & Soder, 1997). Of the errors noted, participant 1 only had errors with /ɾ/ blends, while /ɾ/, /ɛ/ and /ɪ/ were not in error. Of errors noted, participant 2 only had errors with /ɪ/.

Both completed subtests, II, III, V and VI of the *POEC* a second time. Participant 2 had errors that were characterized by atypical stress and intonation. There were few hesitations and errors during the auditory discrimination task. Errors and hesitations occurred when both words were the same. Participant 1 had 5% auditory discrimination errors, a 7% decrease from the diagnostic session. Participant 2 had a 5% auditory discrimination errors but 1% hesitations, a 9% decrease from the diagnostic session.

The Caterpillar Passage was administered a second time to note paragraph reading ability and to note PCC (Patel, Connaghan, Franco, Edsall, Forgit, Olsen et. al, 2013). Participant 1 read with 96% PCC during the maintenance session read, a 3% increase from the diagnostic session. Less i/ɪ substitutions and /ɾ/ distortions were noted during the second time. Other errors included but not relevant to the study were devoicing, vowel distortions and medial or final consonant deletion. Participant 2 read with PCC was 91%, a 4% increase from the diagnostic session. Participant 2 also displayed errors characterized by the addition of shwas at ends of words, omitting syllables in longer words, substitutions including i/ɪ, z/voiced "th," s/voiceless "th," rounded /l/ and derhoticized /r/ intermittently. See Table 11 for a review of study phases.

Table 11. *Overview of Study Phases*

Screening Session	Diagnostic Session	Experimental Sessions
· 1 hour · Informed consent · Hearing screening · Vision screening · Introduction to ultrasound equipment · Questionnaire about second language	· 1 ½ hours · Spontaneous speech sample · *Peabody Picture Vocabulary Test* administered · *Proficiency of English* or *Compton Phonological Assessment of Foreign Accent* administered · *Photo Articulation Test* administered · *Assessment of Intelligibility of Dysarthric Speech* administered depending on perceived intelligibility of participant · *PVSP* · Caterpillar passage reading	· 30 minutes · 4 baseline sessions, one-two times a week · Probe lists administered 3 times during baseline sessions · 8-10 treatment sessions · Five words treated · All probe words read three times · Second probe list administered when accuracy of first list reaches 75%

2.5 Data Collection and Analysis

Data was collected and recorded by the graduate student clinician, filed on excel documents and paper across sessions and de-identified for further analysis. Data was also saved onto a USB drive and kept within the Speech and Gesture Lab in 413 Fisher Hall along with signed informed consent documents. Effect of treatment was measured quantitatively, visually and acoustically. Configuration of the tongue shape via ultrasound images was also be analyzed to determine accuracy of motor movement.

2.5.1 Perceptual rating procedures. The data were captured using a video and audio recording system available within the speech and language clinic. Five pre-professional phase speech-language pathology students, who speak American English as their first language, served as naïve listeners and rated accuracy of phoneme production across sessions provided in random order. Perceptual production analysis of target accuracy was conducted by asking naïve listeners to rate speech productions similar to Sjolie, Leece and Preston (2016). Listeners scored probed productions dichotomously as correct or incorrect. The percentage of probes scored correct was averaged and used as the percent of probes correct for the final visual and quantitative analysis. It is possible that probing after treatment “primed” individuals for more accurate production. However, unprobed targets speak to effect of treatment. Moreover, the purpose of the study was to examine effect of treatment after implementation. Thus, probing occurred after treatment was given to determine effect. Participants were also asked to score overall accuracy of each probe set as well as carrier phrase productions along a continuum, (i.e. very accurate or not at all), by marking an “x” to the closest representation (See Appendix E). Carrier phrases were noted to measure generalization for initial targets final /r/ and medial /ɪ/. Probes were not treated within carrier phrases. Rather, they were probed during Baseline$_2$ through Treatment$_2$ and during the maintenance session. Higher numbers corresponded with more accurate productions.

2.5.2 Visual analysis. Perceptual analysis data from each session were plotted on line graphs for visual analysis between baseline, treatment and withdrawal phases following procedures outlined by Byiers, Reichle and Symons (2012) Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf and Shadish (2010). Data was examined for changes in two parameters; level, variability and trend (slope). Level allowed comparison of data points between phases. Variability showed amount of change between sessions. Trend depicted the overall improvement of phoneme accuracy during the study. Visual analysis of data determined the strength of relationships between implementation of ultrasound biofeedback and improvement of accuracy of American English phonemes.

2.5.3 Quantitative analysis. Quantitative analyses served as the primary means for determining effect of treatment and were based on the perceptual dichotomous ratings. Quantitative analyses were modified from studies by McAllister Byun (2017) and Behrman (2014) and included analyses of data before and after intervention periods to determine each participant's response to the ultrasound treatment. Descriptive data for means and standard deviations of accuracy for all conditions (targeted treated, untreated, total and non-targeted items) were presented.

Standard mean difference (SMD) effect sizes and percent non-overlapping (PND) data were completed to deduce similarities in performance across and between treatment phases. As noted by Olive and Smith (2005), SMD is a simple, beneficial analysis for single-subject design studies. Olive and Smith stated, "this method utilizes data from the mean performance during baseline as well as mean performance during intervention" (p. 322). Following the Olive and Smith study (2005), SMD was calculated to compare the participants' performance in Treatment_2 and Baseline_1. SMD was calculated by finding the difference between the means of the first baseline and second treatment sessions divided by the standard deviation of the scores in the first baseline phase.

Olive and Smith (2005) also noted the benefit of percent non-overlapping data (PND) as an additional analysis for studies interested in either decreasing or increasing target behaviors. PND was calculated by finding the highest baseline point and the number of intervention points that fell above the highest baseline to determine effect of ultrasound implementation and improving the accuracy of American English phonemes.

2.5.4 Acoustic analysis. Acoustic analysis was performed as a secondary analysis to examine the feasibility and value for future studies. Analysis of formant frequency characteristics from spectrograms followed procedures similar to those outlined by Chen, Robb, Gilbert, and Lerman (2001), Georgeton, Antolik, and Fougeron (2016), McAllister Byun (2017), and Tsui (2012) using Kay Multispeech software. Acoustic analysis with spectrogram data allows an additional means of examining production change with implementation of the ultrasound. All phonemes have specific formant characteristics that are evident through acoustic analysis. For example, McAllister Byun (2017) implemented such analysis and noted: "the acoustic hallmark of rhoticity is a significant lowering of the third formant (F3), the second formant (F2) is relatively high in rhotics, resulting in a small distance between the two formants," (p. 1176). Change in production accuracy was determined by analyzing specific characteristics of the target phonemes across baseline and intervention sessions. Specific analyzing methods followed those similar to Chen et al. (2001) but utilized spectrogram analysis rather than LPC waveform coding. One treated probe word said three times each during every baseline and treatment session for each target sound was randomly selected. The F1 and F2 frequencies of each target phoneme in addition to F3 and the F2 to F3 distance for /r/ were determined. Acoustic signals were digitized at 44.1 kHz sampling rate using a speech software package (Kay CSL 4300B). Following Chen et al. (2011),"[o]nce the word was displayed as an amplitude-by-time waveform, a 50 msec window was imposed at the mid-point of the vowel segment." Then, waveform within the window was transformed into a spectrogram using the software. The cursor was placed on the center frequencies which represented F1, F2, and, only in the case of /r/, F3. Formants from each word were averaged for each session to compare change between sessions and phases as well as to compare to norms for American English and Mandarin

(See Appendices F, G, & H). Data were plotted with histograms and analyzed visually. Means were compared to American English and Mandarin established means by Chen et al. (2001) and Hagiwara (1995).

Following reliability measures paralleled those by Sjolie, Leece, and Preston (2016), a single rater was given written guidelines and a brief training on how to conduct acoustic analysis measures. A total of 20% of the total trials were measured by the rater who was blinded to the session number and unaware of baseline or treatment phase for interrater reliability. Scores were compared with the graduate student's analysis to determine the degree of agreement. The absolute mean difference value between the formant data for the first and second rater was determined for acquisition, maintenance, and generalization for each participant.

2.5.5 Analysis of ultrasound images. Similar to acoustic analysis, ultrasound image analysis was performed as a secondary analysis to examine the feasibility and value for future studies. Visual analysis of ultrasound images augmented quantitative and visual analyses to determine accuracy of tongue placement across baseline and treatment sessions as well as whether productions were typical for American English phoneme placement. This considered visual analysis of ultrasound images completed by Tsui (2012). Ultrasound images were analyzed using a visual analog scale similar to those implemented for voice analysis with the *Consensus-Auditory-Perceptual Evaluation of Voice* (Kempster, Gerratt, Verdonlini Abbott, Barkmeier-Kraemer, & Hillman, 2009). Sagittal pictures of the participants producing sustained target phonemes were taken at the beginning of the first sessions of Baseline_1 and Baseline_2, all treatment sessions and at the maintenance session. Ultrasound images acquired during the target phoneme productions were analyzed further to determine improvement of motor patterns across treatment sessions (e.g. tongue tip placement, retroflexed, bunched, etc.). The sagittal view was chosen because it was cued most for both participants during treatment sessions. Pictures were randomized and saved on flash drives. Flash drives were distributed to three individuals familiar with ultrasound imaging of the tongue. The individuals were asked to rate the tongue configuration along a continuum; i.e. high to low for targets /ɪ/ and /ɛ/ and bunched/retroflexed to undifferentiated for /r/ (See Appendix I). Individuals were asked to place an “x” closest to the configuration that most represented the picture, similar to the *CAPE-V* (2009). Kempster et al. (2009) noted “visual analog scales are easy for raters to use and appear to have become more commonplace in voice research in the past 2 decades” (p. 126). For this reason, a visual analog scale was used to determine tongue height change over time. Distance from the beginning of the line to the “x” in millimeters was measured. Lower numbers correspond to more accurate tongue placement. “Gold Standard” images from the sessions were also included as a reference for the scorers (See Appendix J). Similar to perceptual analysis procedures, visual analysis were completed for ultrasound image analysis.

2.5.6 Treatment fidelity. Treatment fidelity refers to the methodological strategies used to monitor and enhance the reliability and validity of behavioral interventions. Assuring optimal treatment fidelity also may decrease the costs of a study and help the research team explain findings. Similar to a study conducted by Rusiewicz and Rivera (2017) and Sjolie, Leece and Preston (2016), 25% of treatment sessions were viewed by an individual unfamiliar with the purpose of the study. This individual checked for use of KP or KR verbal cues, number of probes targeted and implementation of visual biofeedback with the ultrasound.

Chapter 3: Results

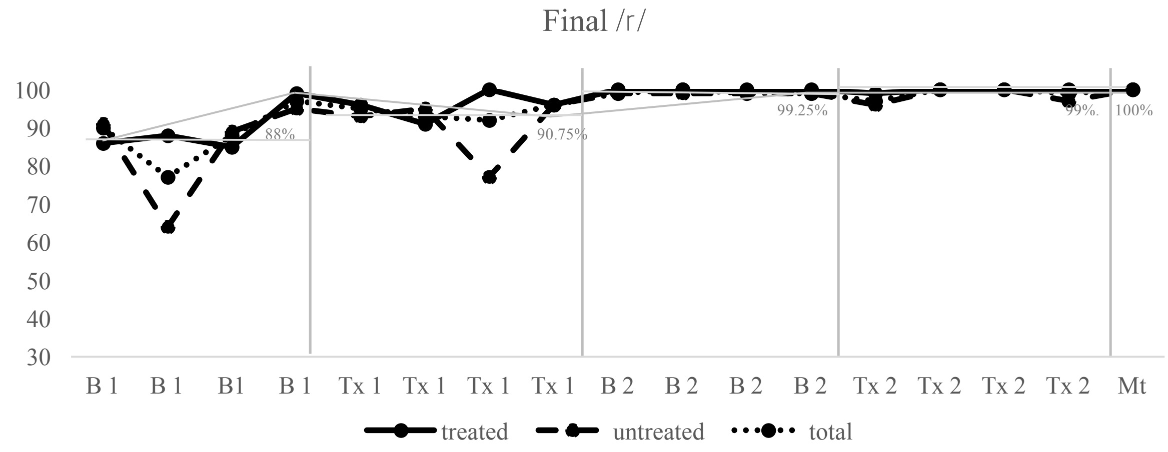
3.1 Participant 1

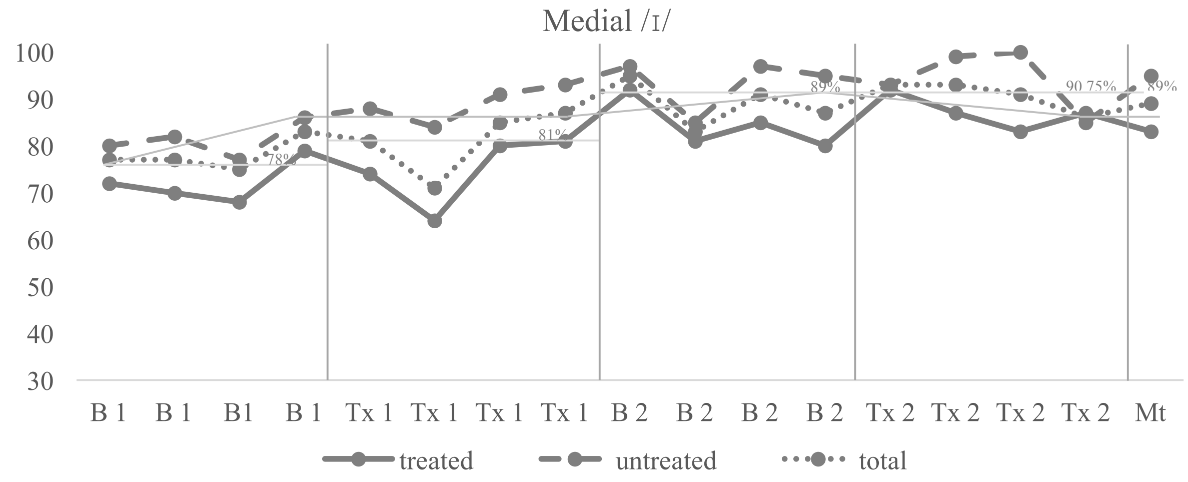
3.1.1 Visual analysis. In addition to the descriptive analyses, visual analyses were also completed to compare productions of target phonemes across all baseline and treatment phases. Visual analyses included level (i.e. mean performance within phases) to compare the data points between phases, trend (i.e. slope within phases) to depict the overall improvement of accuracy during the study and variability to determine stability of performance within phases. As stated by Rusiewicz and Rivera (2017) "[a] causal relationship is supported when data across the phases show at least three demonstrations of effect at three separate points in time" (p. 1240). These measures supported the quantitative analysis described above.

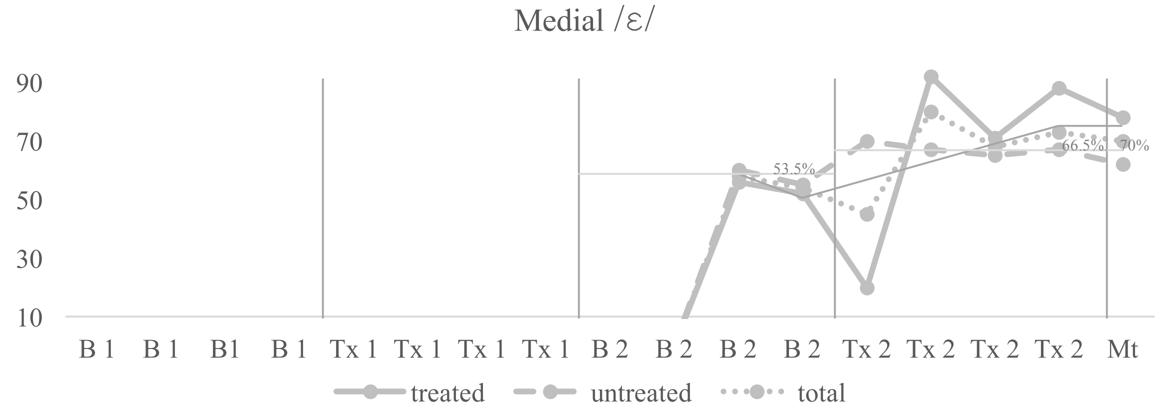
Improvement for all targets, treated and untreated, was observed as the study progressed, with the greatest accuracy typically noted for Treatment$_2$ (see *Figure 3*). A greater increase from Baseline$_1$ to Treatment$_2$ and a lesser increase from Baseline$_2$ to Treatment$_2$ was noted by both the clinician and naïve listeners. Maintenance numbers were typically higher than Treatment$_2$ numbers. In cases when the mean was lower than Treatment$_2$, it was still higher than Baseline$_1$ and Treatment$_1$. For the mean judgments of all treated and untreated final /ɾ/ productions, an increase from Baseline$_1$ to Treatment$_2$ was noted by the clinician and naïve listeners. For treated final /ɾ/ productions, an increase from Baseline$_1$ to Treatment$_2$ was noted by the clinician and naïve listeners with the greatest increase from Baseline$_1$ to Treatment$_1$. For untreated final /ɾ/ productions, an increase from Baseline$_1$ to Treatment$_2$ was noted by the clinician with the greatest increase from Baseline$_1$ to Treatment$_2$. For the naïve listeners, an increase from Baseline$_1$ to Treatment$_2$ and near perfect performance from the end of Treatment$_2$ until the Maintenance sessions were noted.

For the total treated and untreated mean judgments of medial /ɪ/ productions, both the clinician and naïve listeners noted an increase from Baseline$_1$ to Treatment$_2$ (see *Figure 3*). For treated medial /ɪ/ productions, the clinician and naïve listeners noted an increase from Baseline$_1$ to Treatment$_2$ with the greatest increase from Baseline$_1$ to Treatment$_1$. There was less of an increase from Baseline$_2$ to Treatment$_2$ due to high accuracy being reached during this baseline phase. For untreated medial /ɪ/, an increase from Baseline$_1$ to Treatment$_2$ was noted by both the clinician and naïve listeners. A greater increase from Baseline$_1$ to Treatment$_1$ for both the clinician and naïve listeners was also noted.

Figure 3. Participant 1 Visual Analysis Naïve Listeners. This figure includes trend and level lines as well as total means for each phase.







For all total treated and untreated, treated only and untreated only mean judgements of medial /ɛ/, an increase from the baseline to treatment phases was noted by the clinician and naïve listeners (see Table 12). A greater increase was noted by the clinician for the total treated and untreated as well as untreated productions. Lesser mean accuracies were noted by the final treatment phase.

Table 12. *Participant 1 Clinician and Listener Mean Percent Accuracies*

		Total Final /r/	Treated Final /r/	Untreated Final /r/	Total Medial /ɪ/	Treated Medial /ɪ/	Untreated Medial /ɪ/	Total Medial /ɛ/	Treated Medial /ɛ/	Untreated Medial /ɛ/
B$_1$	*Clinician*	22.5	26.5	24.75	47.25	19.5	56.5	0	0	0
	Listeners	88	89.5	84.75	78	70	83.25	0	0	0
Tx$_1$	*Clinician*	68	79.5	66.25	56.5	46.5	63	0	0	0
	Listeners	90.75	95.75	90.25	81	74.25	91.25	0	0	0
B$_2$	*Clinician*	79.75	90	69.5	62.25	53.25	71.5	8.25	26	3.25
	Listeners	99.25	100	99	89	84.75	92.5	53.5	54	57.5
Tx$_2$	*Clinician*	79	91.5	71.5	69	66.5	74	59.25	76.5	49.75
	Listeners	99	99.75	98.25	90.75	85.5	94.25	66.5	67.75	67.25
Mt	*Clinician*	80	93	60	70	60	80	60	60	60
	Listeners	100	100	100	89	83	95	70	78	62

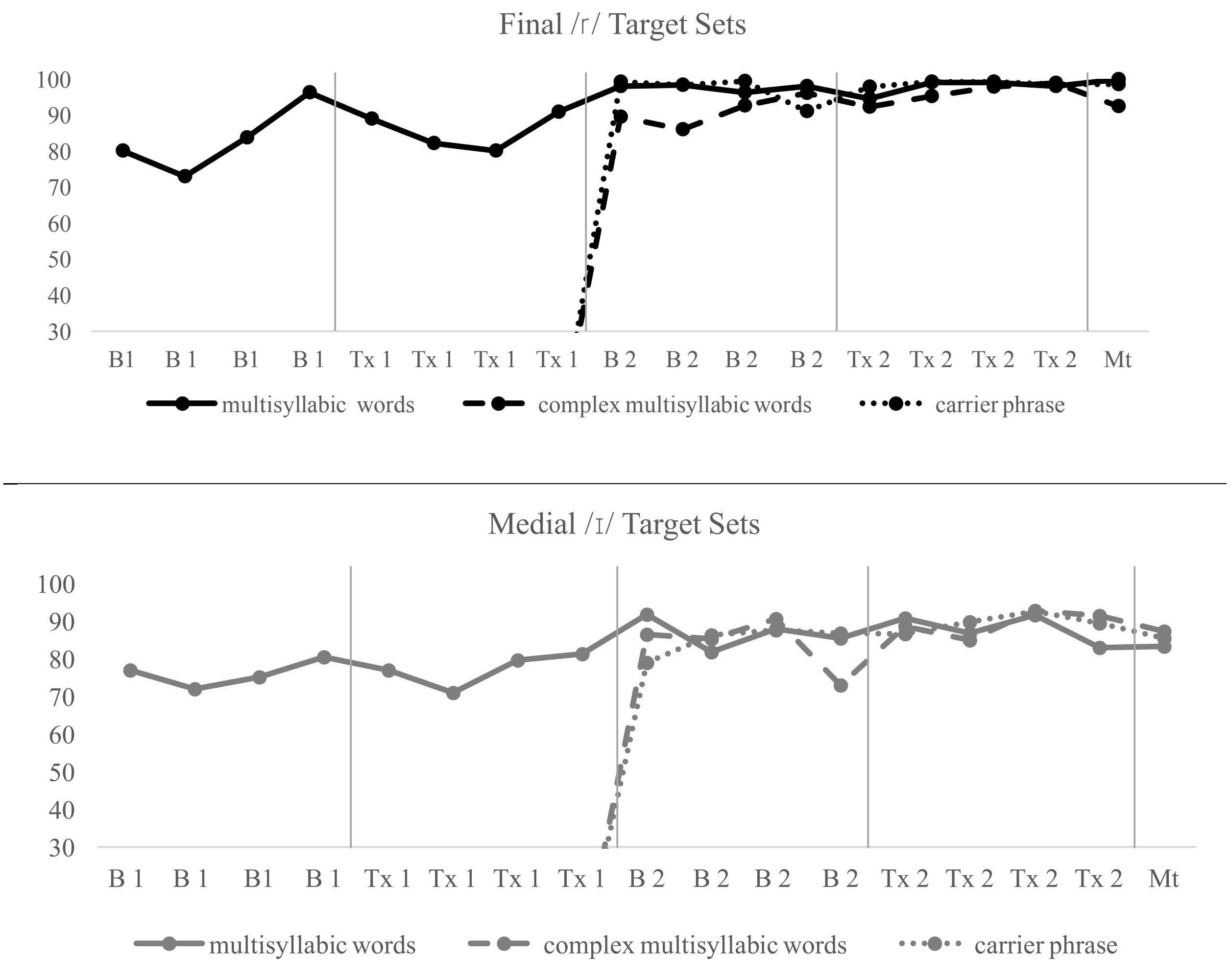
For the total final /r/ productions, the naïve listeners' ratings contained the greatest variability during the initial baseline phase (i.e. 77%-97%). High accuracy was reached and maintained by the end of Treatment$_1$ and minimal variability was noted during the Treatment$_2$ phase showing retention of skills through the remainder of the study (i.e. 97%-100%). For the treated productions, the naïve listeners' ratings showed the greatest variability during the initial baseline and treatment phases (i.e. 85%-99%). High accuracy was reached and less variability was noted through the rest of the study to Treatment$_2$ (i.e. 99%-100%). The naïve listeners' ratings of untreated productions showed the greatest variability during the initial baseline phase (i.e. 64%-95%). High accuracy was reached and less variability was noted from Treatment$_1$ through to Treatment$_2$ (i.e. 96%-100%) showing retention of skills.

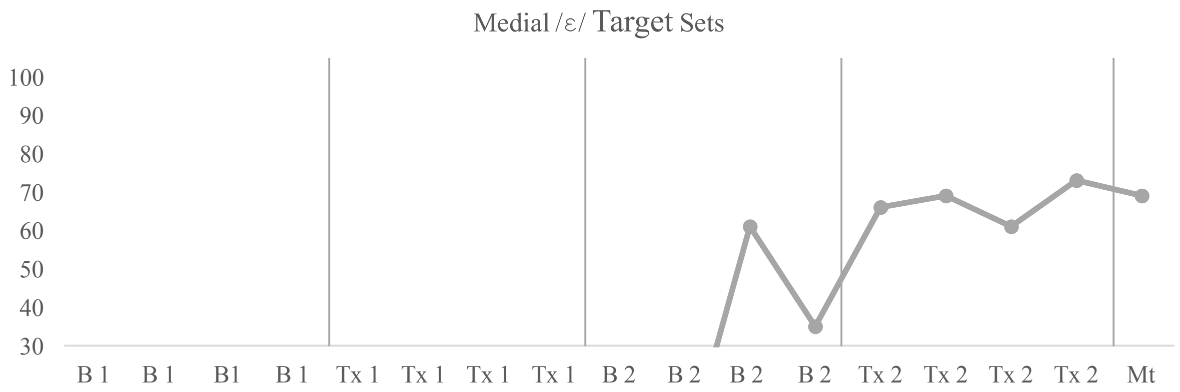
The naïve listeners' ratings showed greater variability of performance during Baseline$_1$ for medial /ɪ/ productions (i.e. 75%-83%). By the end of Treatment$_2$, relatively high accuracy was reached and slightly less variability was noted (i.e. 86%-93%). For treated medial /ɪ/ productions, the naïve listeners noted greater variability during the initial baseline and treatment phases (i.e. 68%-79%). Relatively high accuracy and less variability was noted during the second baseline and treatment phases (i.e. 83%-92%). For the untreated medial /ɪ/ productions, the naïve listeners noted variability during the initial baseline and treatment phases (i.e. 77%-86%). A relatively high accuracy was reached and maintained with somewhat low variability during the remainder of the study (i.e. 85%-100%). Overall, the most variability was seen during the first baseline and treatment phases as productions improved, likely due to treatment implementation. The less variability noted during the final phases is likely a result of the treatment and ability of participant 1 to generalize the skills successfully.

For all, treated and untreated /ɛ/ productions, variability was noted during the initial baseline phases by the naïve listeners (i.e. 49%-58%). Both the clinician and naïve listeners also noted variability during the treatment phase as high accuracy was reached by the final two sessions (i.e. 73%-68%). The naïve listeners noted the least variability of untreated medial /ɛ/ productions during the treatment phase. These productions also reached the least accuracy of all the groups. For treated productions, there was no variability between the two baseline points (i.e. 26%). Low variability was noted for the final two Treatment$_2$ sessions (i.e. 71%-88%). For untreated productions, there was low variability during baseline sessions (i.e. 55%-60%). There was also a slightly greater variability for treatment sessions (i.e. 62%-70%).The higher variability was likely due to only one treatment phase.

Visual analog scores for overall perceptual accuracy was recorded for each set of targets and carrier phrases. Targets showed a general improvement and maintenance of accuracy by the end of the study and corresponded with dichotomous ratings (see *Figure 4*). There was a smaller variability between scores for participant 1 (i.e. 61%-99.4%). Medial /ɛ/ showed the least amount of improvement. Final /ɾ/ in 2-3 syllable words showed the greatest improvement. Likewise, carrier phrase probes were scored for generalization to more challenging contexts. Accuracy of production within carrier phrases started relatively high for both targets and improved over time. Visual analog scores corresponded with dichotomous ratings, noting improvement from baseline with a smaller improvement for medial /ɛ/.

Figure 4. Participant 1 Visual Analog Analysis





3.1.2 Quantitative analysis. Descriptive data for the clinician and naïve listeners' judgements of final /ɾ/, medial /ɪ/ and medial /ɛ/ are included (see Table 13). These analyses included standard mean difference (SMD) effect sizes and percent non-overlapping data (PND). SMD determined degree of change from the initial baseline phase to the final treatment phase (i.e. greater number shows a greater degree of change). PND depicted difference of data points between the initial baseline and final treatment phases (i.e. greater percentage shows a greater degree of change). Both numbers determined degree of treatment implementation.

The standard mean difference (SMD) effect size was determined for both the clinician and naïve listeners (see Table 13). All data showed a clinically relevant effect of treatment. The clinician's ratings yielded greater SMD numbers for final /ɾ/ productions and therefore a greater effect than the naïve listeners' ratings. However, ratings for all groups of productions still showed notable change from Baseline$_1$ to Treatment$_2$. The naïve listeners' ratings of medial /ɪ/ yielded greater SMD numbers for all production groups than for all production groups of final /ɾ/. The naïve listeners' ratings yielded the largest effect sizes for treated medial /ɛ/. Medial /ɛ/ was introduced during the second baseline phase and only treated during the second treatment phase. The clinician's SMD for treated medial /ɛ/ could not be computed due to a SD of 0 during

Baseline_1. PND was judged to be 100% for all naïve listener data except total treated and untreated productions of final /ɾ/ (see Table 13). High PND numbers further supported an improvement from Baseline_1 to Treatment_2 due to the implementation of treatment.

Table 13. *Participant 1 Quantitative Analysis*

		Total Final /ɾ/	Treated Final /ɾ/	Untreated Final /ɾ/	Total Medial /ɪ/	Treated Medial /ɪ/	Untreated Medial /ɪ/	Total Medial /ɛ/	Treated Medial /ɛ/	Untreated Medial /ɛ/
SMD	*Clinician*	7.27	6.89	4.7	0.83	4.98	1.04	10.3	Can't compute	5.06
	Listeners	1.33	1.6	0.96	3.68	2.23	2.92	2.04	4.88	2.75
PND	*Clinician*	100	100	100	75	100	100	100	100	75
	Listeners	75	100	100	100	100	100	100	100	100

3.1.3 Fidelity. The individual blinded to session number checked for use of KP or KR verbal cues, number of probes targeted and implementation of visual biofeedback with the ultrasound in 25% of sessions (see Table 14). Ultrasound was implemented 100% of the time. Ten probes of each target were treated up to 12 times each. Final /ɾ/ probes were treated 10 times each during Treatment_1. Medial /ɪ/ probes were treated between 7 and 10 times each. The same measures were determined for Treatment_2. Final /ɾ/ probes were treated between 8 and 12 times each and medial /ɪ/ probes were treated between 8 and 10 times each. Medial /ɛ/ probes were treated between 9 and 10 times each. KR verbal feedback was implemented more often than KP feedback. Both types of cues were implemented more often during Treatment_1. No cues were given most often during Treatment_2.

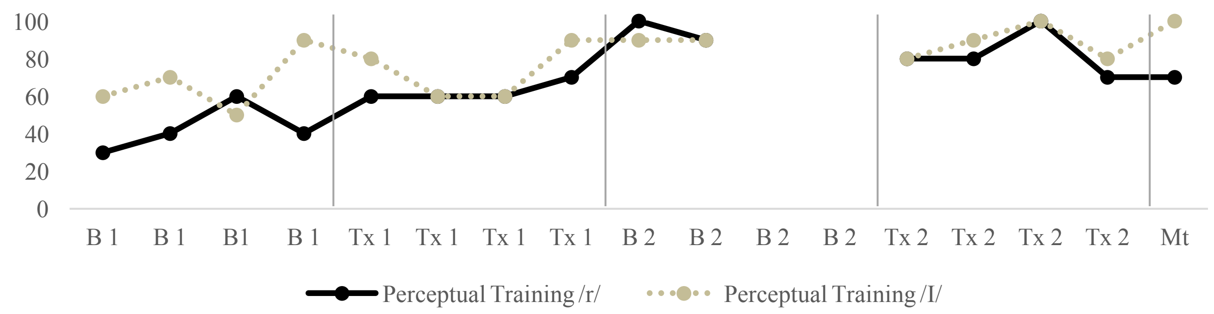
Table 14. *Participant 1 Fidelity*

		KP	**KR**	**Both**	**No Cues**
Treatment_1	*Final /ɾ/*	1%	82%	12%	5%
	Medial /ɪ/	10%	67%	5%	18%

Treatment$_2$	*Final /r/*	6%	77%	0%	17%
	Medial /ɪ/	1%	65%	6%	28%
	Medial /ɛ/	3%	63%	2%	32%

3.1.4 Perceptual training. Perceptual training accuracy was recorded across all treatment and baseline sessions with the exception of two home Baseline$_2$ sessions (see *Figure 5)*. Perception of correct and incorrect production improved as the study progressed. Slight declination of performance was noted for medial /ɪ/ during Treatment $_1$. This corresponds with perceptual accuracy ratings.

Figure 5. Participant 1 Perceptual Training

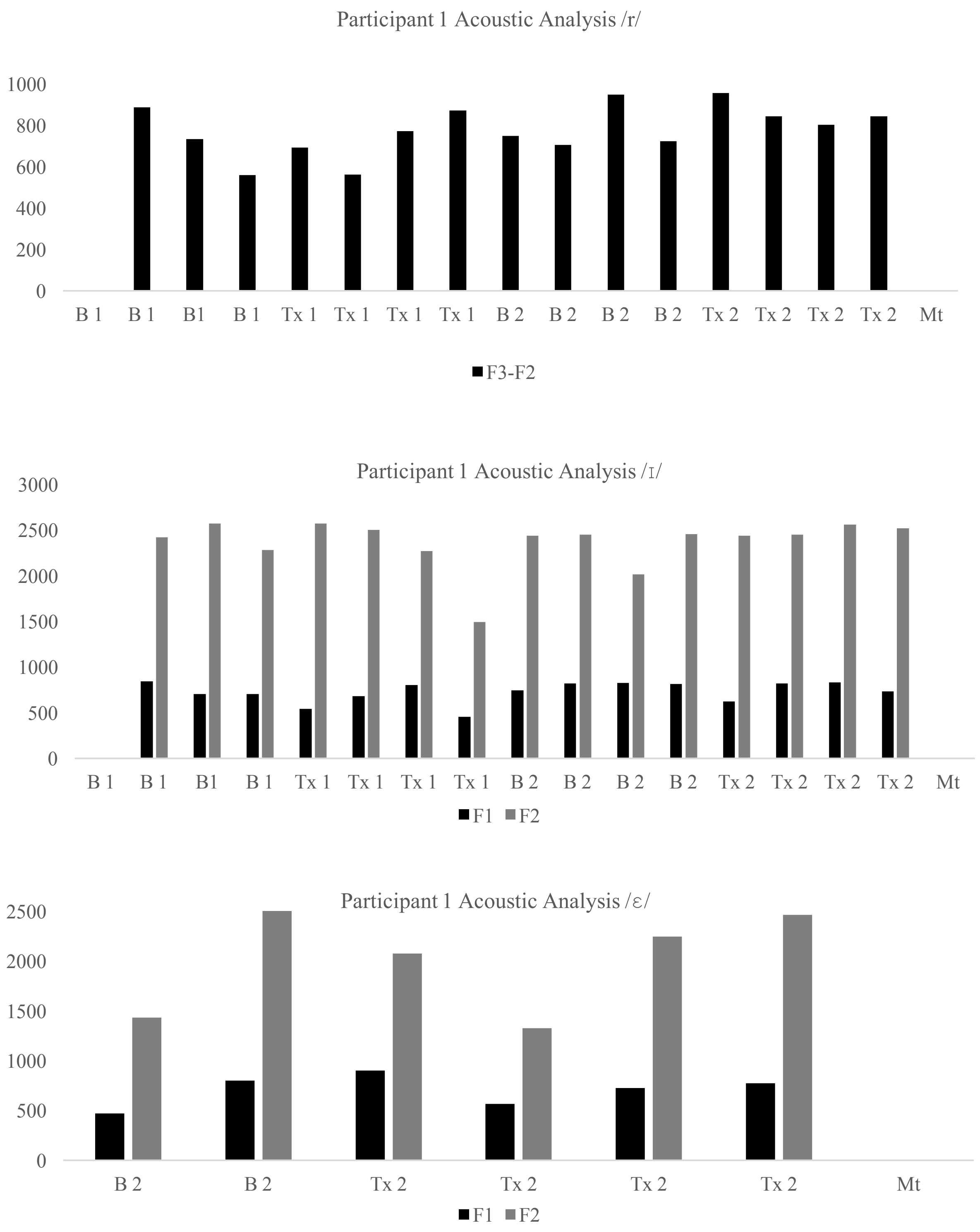


3.1.5 Acoustic analysis. Formant frequencies for participant 1 were compared to means from previous studies. Distance between F2 and F3 was considered most relevant for rhotics as described by McAllister Byun (2017). F1 and F2 vowel frequencies were compared to means productions of Mandarin and American English female speakers. Change was analyzed between sessions visually. Due to recording differences, some sessions were not able to be analyzed due to light spectrograms causing the formants to be difficult to distinguish. Hagiwara (1995) noted means for American English /r/ phoneme production by females to be 532 Hz for F1, 1628 Hz for F2, and 2198 Hz for F3 (p.70). Distance between F2 and F3 decreased slightly from the initial baseline session based on visual analyses (see *Figure* 6). This analysis differed from quantitative analysis findings due to the focus of ultrasound biofeedback treatment. This technology focused primarily on lingual configuration, not labial rounding. As stated previously, lip rounding is characteristic of American English /r/ production and absent in Mandarin /r/ production (Smith, 2010, p. 20). Following this, acoustic signals likely were effected and not as characteristic of American English phoneme production because this was not an element of treatment.

Norms for medial /ɪ/ in Mandarin and American English are close. Chen et al. (2001) noted means of American English phonemes produced by Mandarin speakers. American English female means were noted to be 492 Hz for F1 and 2267 Hz for F2 (p. 433). Chen et al. (2001) noted Mandarin female means for to be 434 Hz for F1 and 2444 Hz for F2 (p. 432). Moreover, Chen et al. (2001) noted that Mandarin females typically produced F1 of medial /ɪ/ lower than American English speakers. There was an overall decrease seen with F1 however, F1 never reached either American English or Mandarin norm (see *Figure 6*). F2 remained relatively consistent and was closer to established Mandarin means.

Chen et al. (2001) noted female American English means for medial /ɛ/ to be 737 Hz for F1 and 2141 Hz for F2 (p. 433). For Mandarin females, the norms were determined to be 762 Hz for F1 and 2078 Hz for F2 (p. 432). Both of these formants were noted to be produced similar to American English means. By the end of treatment, F1 and F2 were more characteristic of American English phoneme production (see *Figure 6*). Acoustic data from Baseline$_1$ and Treatment$_1$ is not available because treatment of this phoneme did not start until the second phase.

Figure 6. Participant 1 Acoustic Analysis


Participant 1 Acoustic Analysis /r/

Participant 1 Acoustic Analysis /ɪ/

Participant 1 Acoustic Analysis /ɛ/

Absolute mean difference and standard deviation were computed for all formants and the difference between F2 and F3 for /ɾ/. /ɾ/ showed the greatest difference between raters with F3 for /ɾ/ showing the most difference (see Table 15). The greatest mean difference and standard deviation was for the F2 of /ɛ/.

Table 15. *Participant 1 Reliability*

		Absolute Mean Difference	*Standard Deviation*
/ɾ/	**F1**	35.83	23.47
	F2	190.5	161.94
	F3	148.3	166.21
	F2-F3	86.83	83.90
/ɪ/	**F1**	75.3	21.35
	F2	53	50.86
/ɛ/	**F1**	70	38.08
	F2	266.3	352.92

3.1.4 Analysis of ultrasound images. Similar to the perceptual rating analysis, visual analyses were also completed for ultrasound image analysis to compare analysis of pictures of target phoneme productions across baseline and treatment phases. Only one baseline picture was taken so PND and SMD effect size could not be completed. Visual analyses included level and trend. Mean percent accuracies were also noted. These measures augmented the perceptual and acoustic analyses described above.

The average ratings for Basline$_1$ were relatively high for final /ɾ/ (see Table 16). There was a decrease between Baseline$_1$ and Treatment$_1$. Baseline$_2$ decreased when treatment was discontinued and increased again during Treatment$_2$. For medial /ɪ/, there was a decrease from

Baseline$_2$ to Treatment$_2$. The first Treatment$_2$ session yielded the lowest ratings overall. After treatment was implemented for the remainder of Treatment$_2$, an increase in mean percentages was noted.

The mean percentage for medial /ɛ/ during Treatment2 was 43.75%. There were no baseline images taken for medial /ɛ/. However, there was an increase from the initial image to the final image. The greatest amount of change was noted between the third and fourth Treatment$_2$ sessions where there was a decrease in ratings. Although there was an increase, the final Treatment$_2$ point was still lower than the initial Treatment$_2$ point.

Table 16. *Participant 1 Mean Ultrasound Accuracies Across Phases*

	Final /ɾ/	Medial /ɪ/	Medial /ɛ/
Basleine$_1$	80.66	41	0
Treatment$_1$	49.92	49.33	0
Baseline$_2$	9.33	46.67	29.67
Treatment$_2$	64.84	28.25	43.75
Maintenance	17	14.33	20

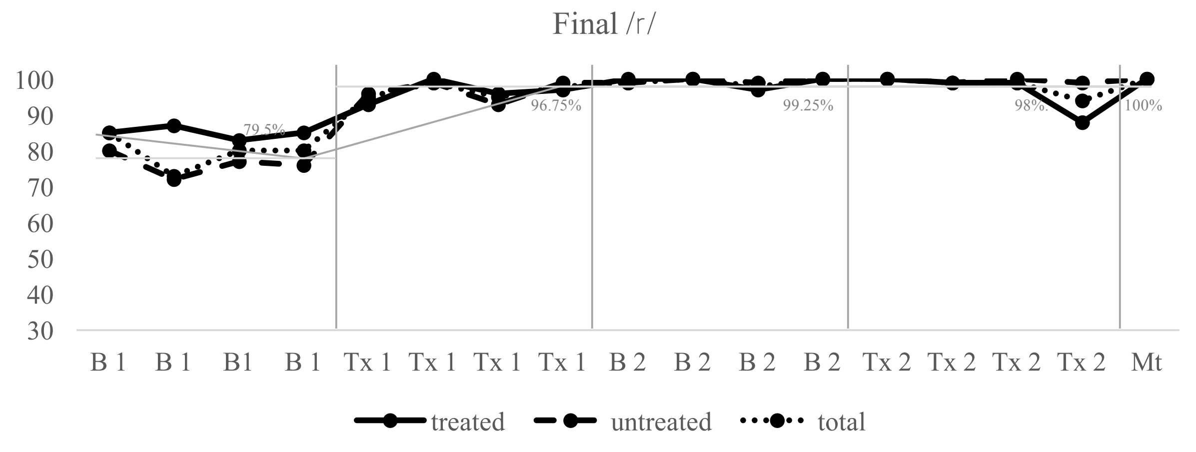
3.2 Participant 2

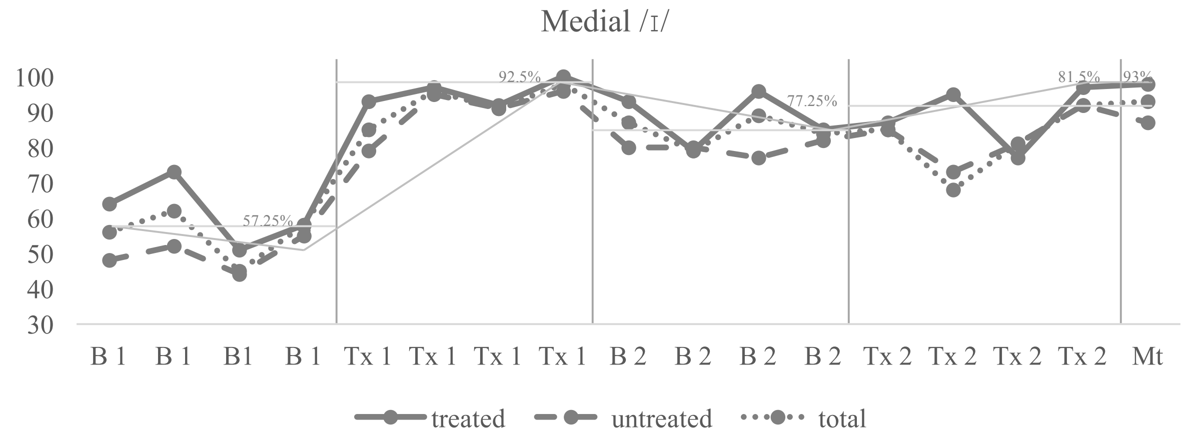
3.2.1 Visual analysis. In addition to the descriptive analysis replication, visual analyses including level, trend and variability were also replicated for participant 2 to compare productions of target phonemes across all baseline and treatment phases. For the total treated and untreated mean judgments of final /ɾ/ productions, the clinician and naïve listeners noted an increase from Baseline$_1$ to Baseline$_2$ and a slight decline from Baseline$_2$ to Treatment$_2$ (see *Figure 7)*. For treated final /ɾ/ productions, the clinician and naïve listeners noted an increase from Baseline$_1$ to Baseline$_2$ and a slight drop from Baseline$_2$ to Treatment$_2$. For untreated productions, the clinician noted an increase from Baseline$_1$ to Baseline$_2$ and a slight drop from

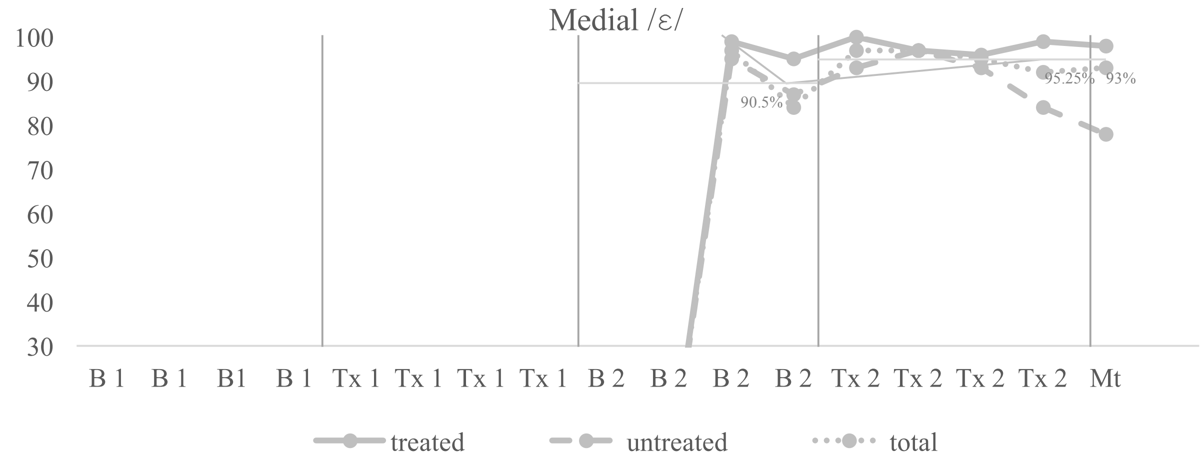
Baseline$_2$ to Treatment$_2$. For the same productions, the naïve listeners noted an increase from Baseline$_1$ to Baseline$_2$ and a maintenance of accuracy through Treatment$_2$.

Mean accuracies of productions were noted for medial /ɪ/ as well (see *Figure 7*). For the total treated and untreated medial /ɪ/ productions, the clinician noted an increase from Baseline$_1$ to Baseline$_2$ and a slight declination of accuracy from Baseline$_2$ to Treatment$_2$. For the same productions, the naïve listeners noted an increase from Baseline$_1$ to Treatment$_1$, decline from Treatment$_1$ to Baseline$_2$ and an increase from Baseline$_2$ to Treatment$_2$. For treated medial /ɪ/ productions, both the clinician and naïve listeners noted an increase from Baseline$_1$ to Baseline$_2$ and a slight decrease in accuracy from Baseline$_2$ to Treatment$_2$. For the untreated medial /ɪ/ productions, both the clinician and naïve listeners noted an increase from Baseline$_1$ through to Treatment$_2$.

Figure 7. Participant 2 Visual Analysis Naïve Listeners. This figure includes trend and level lines as well as total means for each phase.







For all the total, treated and untreated mean medial /ɛ/ productions, the clinician and naïve listeners both noted an increase from baseline to treatment phases (see Table 17). Both the clinician and naïve listeners noted increases from baseline to treatment phases across almost all groups. High accuracy was noted for treated /ɛ/ productions so less of an increase was noted for this group.

Table 17. *Participant 2 Clinician and Listener Mean Percent Accuracies Across Phases*

		Total Final /r/	Treated Final /r/	Untreated Final /r/	Total Medial /ɪ/	Treated Medial /ɪ/	Untreated Medial /ɪ/	Total Medial /ɛ/	Treated Medial /ɛ/	Untreated Medial /ɛ/
B_1	*Clinician*	26.25	28.25	25	35.5	43	28	0	0	0
	Listeners	79.5	85	76.25	57.25	61.5	49.75	0	0	0
Tx_1	*Clinician*	61.5	54.75	61.25	64.75	71.25	56.5	0	0	0
	Listeners	96.75	93.5	96.75	92.5	85	80.5	0	0	0
B_2	*Clinician*	95	96.75	93.25	70.75	86.25	56.25	38	49.5	26.5
	Listeners	99.25	98.5	99.5	77.25	92	81	90.5	97	91
Tx_2	*Clinician*	86	81.25	89.5	66.75	73	60	74	86	56.5
	Listeners	98	96.5	99.5	81.5	89	82.75	95.25	98	91.75
Mt	*Clinician*	100	100	100	80	86	80	90	100	80
	Listeners	100	100	10	77	98	87	93	98	78

The naïve listeners noted the greatest variability during Baseline$_1$ for total final /r/ (i.e. 73%-85%). High accuracy was reached and maintained with little variability through the end of Treatment$_1$ (i.e. 94%-100%). For treated productions, little variability was noted by the naïve listeners during Baseline$_1$ (i.e. 83%-87%). Accuracy was maintained and variability was low from Treatment$_2$ to the end of the study (i.e. 99%-100%). For untreated productions, the naïve listeners noted variability during Baseline$_1$ (72%-80%). High accuracy and low variability was reached during Treatment$_2$ (i.e. 99%-100%).

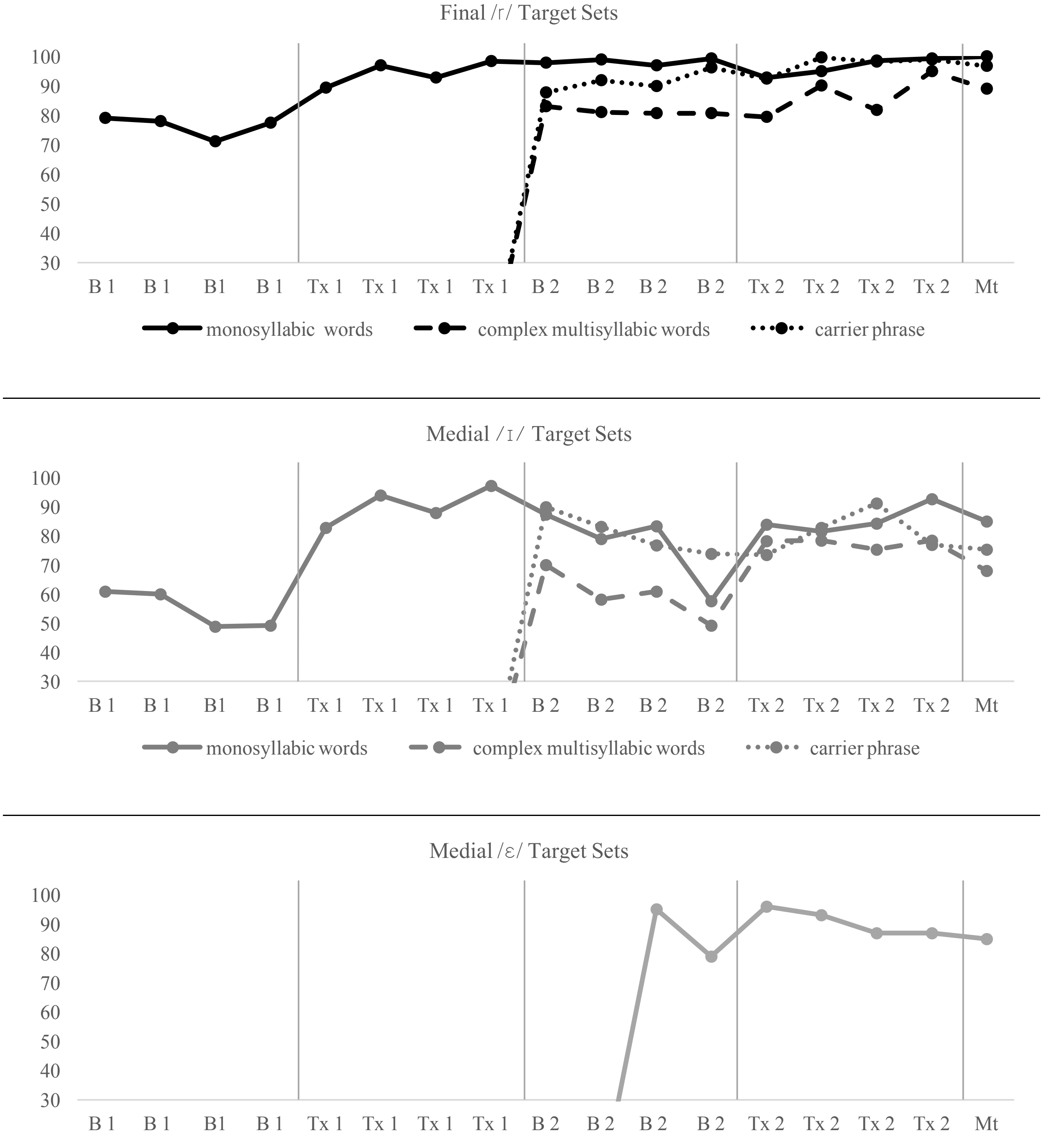
The naïve listeners noted the greatest variability during Baseline$_1$ and Baseline$_2$ for all productions of medial /ɪ/ (i.e. 45%-64%). The naïve listeners noted a greater variability during Baseline$_1$ for treated phonemes (i.e. 51%-73%). A relatively high accuracy was reached and maintained through the remainder of Treatment$_2$ (i.e. 87%-97%). The naïve listeners noted low accuracy and little variability during Baseline$_1$ for untreated productions (i.e. 44%-55%).

Variability during Treatment_1 was greater as treatment was introduced, showing generalization of skill (i.e. 51%-95%). Baseline_2 showed little variability (i.e. 77%-80%). Treatment_2 showed greater variability than Baseline_2 (i.e. 73%-92%).

The naïve listeners noted less variability during baseline and treatment points for all groups of medial /ɛ/ (i.e. 33%-43%). However, these were the only two baseline points. High accuracy was noted by the naïve listeners during the treatment phase in all total, treated and untreated groups (i.e. 92%-97%).

Visual analog scores for overall perceptual accuracy were replicated for participant 2 for each set of targets. Targets showed a general improvement and maintenance of accuracy by the end of the study and corresponded with dichotomous ratings (see *Figure 8*). Overall scores for participant 2 had a greater variability than those for participant 1 (i.e. 48.8%-100%). Final /r/ in monosyllabic words showed the greatest amount of improvement. Medial /ɪ/ in more complex multisyllabic words showed the least amount of improvement. Generalization to carrier phrases was also noted for participant 2. Generalization was depicted with relatively high initial numbers which remained somewhat consistent through the end of treatment. Visual analog scores corresponded with dichotomous ratings, noting improvement from baseline.

Figure 8. Participant 2 Visual Analog Analysis



3.2.2Quantitative analysis. All quantitative analyses including SMD and PND were replicated for participant 2. SMD effect size was calculated to noted change from Baseline$_1$ to

Treatment_2 with higher numbers evidencing a greater degree of change (see Table 18). The highest numbers were noted for both treated and untreated final /r/ ratings by the naïve listeners, however, all numbers evidenced change. The greatest change for medial /ɛ/ noted by the naïve listeners was noted for the untreated productions. Treated productions of this the final /r/ phoneme showed the greatest effect size. Medial /ɛ/ showed the least amount of change. PND was judged to be 100% for all total, treated and untreated final /r/ and medial /ɪ/ by both the clinician and naïve listeners (see Table 18). PND was judged to be lower for the naïve listeners' ratings of medial /ɛ/.

Table 18. *Participant 2 Quantitative Analysis*

		Total Final /r/	Treated Final /r/	Untreated Final /r/	Total Medial /ɪ/	Treated Medial /ɪ/	Untreated Medial /ɪ/	Total Medial /ɛ/	Treated Medial /ɛ/	Untreated Medial /ɛ/
SMD	*Clinician*	5.13	5.33	4.01	6.25	1.58	3.2	5.09	7.37	1.57
	Listeners	3.75	7.06	7.05	2.84	2.82	6.89	0.3	0.35	0.13
PND	*Clinician*	100	100	100	100	100	100	100	100	50
	Listeners	100	100	100	100	100	100	0	25	25

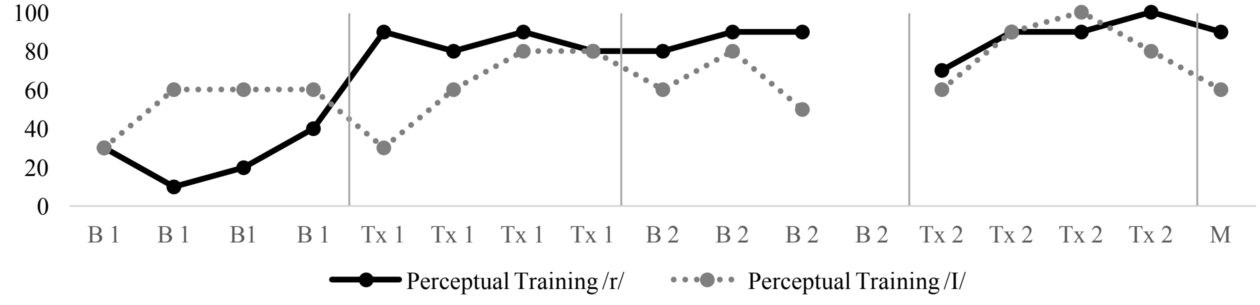
3.2.3 Fidelity. Fidelity measures were replicated for participant 2 for 25% of sessions (see Table 19). Ultrasound was implemented 100% of the time. For final /r/ probes, ten probes were treated between 9 and 13 times each during Treatment_1. For medial /ɪ/, technical issues resulted in the recording cutting short during the 8^{th} probe so fidelity could only be recorded for 8 probes. All 8 probes were treated between 6 and 11 times each. The same measures were determined for Treatment_2. Ten probes of each target were treated up to 13 times each. Final /r/ probes were treated between 9 and 13 times each and medial /ɪ/ probes were treated between 9 and 10 times each. Medial /ɛ/ probes were treated between 7 and 11 times each. KR cues were given more often than KP cues. No cues were given more often during Treatment$_2$ after behavior was learned. No cues were given more often for participant 2 due to quick probe production. Participant 2 was often encouraged to slow down for the clinician to provide verbal feedback.

Table 19. *Participant 2 Fidelity*

		KP	KR	Both	No Cues
Treatment$_1$	*Final /ɾ/*	10%	62%	11%	17%
	Medial /ɪ/	20%	26%	10%	44%
Treatment$_2$	*Final /ɾ/*	12%	21%	2%	65%
	Medial /ɪ/	9%	28%	2%	61%
	Medial /ɛ/	7%	26%	2%	65%

3.2.4 Perceptual training. Perceptual training accuracy was recorded across all treatment and baseline sessions with the exception of one home Baseline$_2$ session (see *Figure 9)*. Perception of correct and incorrect production improved as the study progressed with slight declines noted for medial /ɪ/ during Treatment$_1$ and Baseline$_2$. The overall improvement aligns with perceptual accuracy ratings. However, while there was drop during some baseline perceptual ratings, there continued to be improvement through baseline sessions because no change in perceptual training was implemented.

Figure 9. Perceptual Training Participant 2

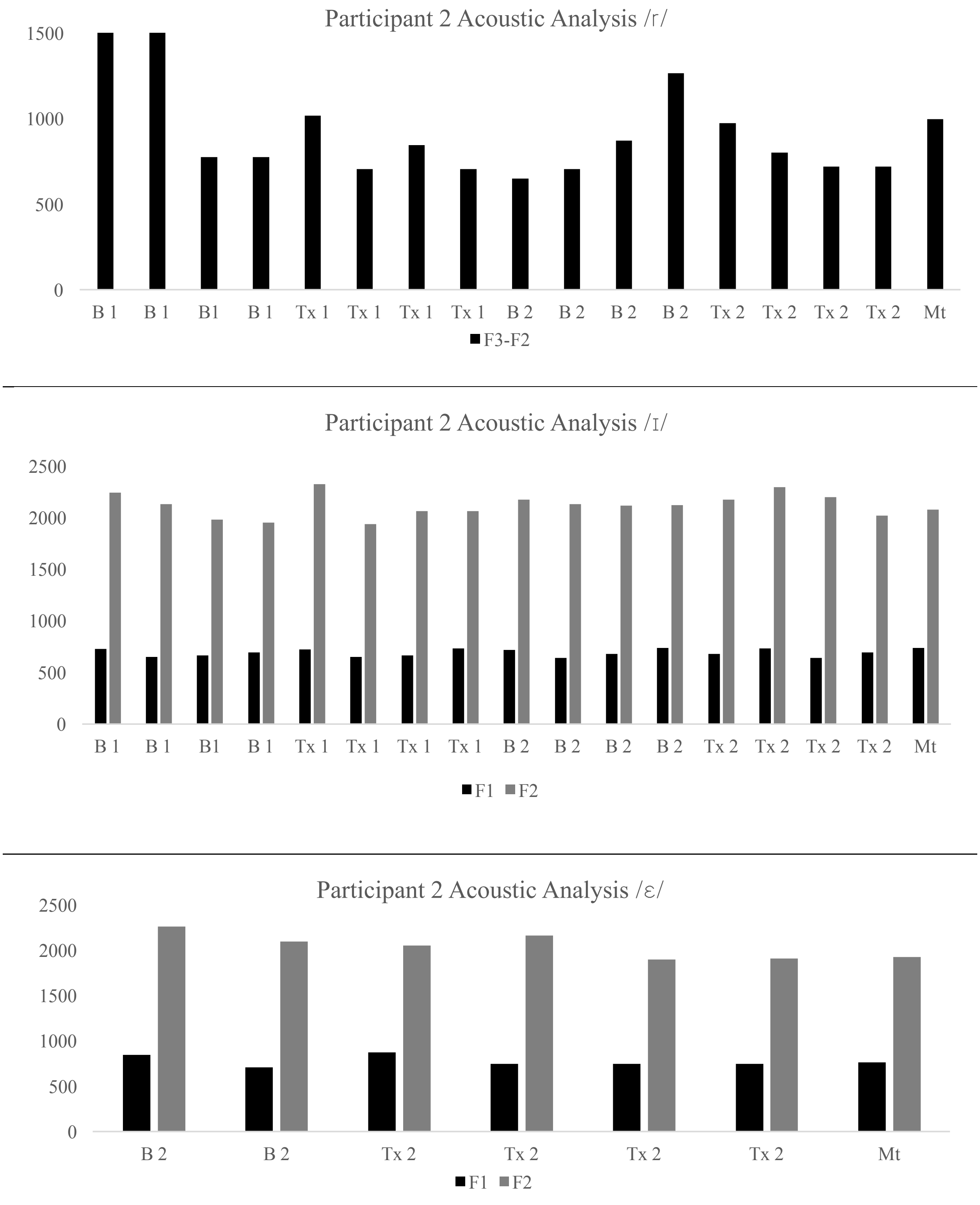


3.2.5 Acoustic analysis. Acoustic analyses were replicated for participant 2 and formant frequencies were compared to established male means by Chen et al. (2001) and Hagiwara (1995). The examiner noted an overall decrease for all /ɾ/ formants from Baseline1 to Treatment2 reflecting similar American English characteristics (see *Figure 10*). However, this distance never reached what the mean established by Hagiwara (1995).

Chen et al. (2001) noted American English male means for medial /ɪ/ to be 432 Hz for F1 and 1864 Hz for F2 (p. 432). Chen et al. (2001) noted Mandarin male means for medial /ɪ/ to be 412 Hz for F1 and 2046 Hz for F2 (p. 433). Moreover, Chen et al. (2001) noted that Mandarin males typically produced F2 with a higher frequency that American English males. Similar to participant 1, participant 2 produced F2 that was characteristic means from the referenced study (see *Figure 10*).

Chen et al. (2001) noted American English male means for medial /ɛ/ to be 578 Hz for F1 and 1793 Hz for F2 (p. 432). Chen et al. (2001) noted Mandarin male means for medial /ɛ/ to be 606 Hz for F1 and 1823 Hz for F2. Chen et al. (2001) also noted that both frequencies were typically produced similar to American English speakers. F2 remained relatively consistent and was closer to established means. Medial /ɛ/ formants were consistently high but decreased overall from initial baseline to the end of treatment (see *Figure* 10).

Figure 10. Participant 2 Acoustic Analysis



Absolute mean difference and standard deviation were replicated for participant 2. F2 for /ɪ/ yielded the greatest absolute mean difference (see Table 20). F1 for /ɛ/ yielded the greatest standard deviation.

Table 20. *Participant 2 Reliability*

		Absolute Mean Difference	*Standard Deviation*
/r/	**F1**	106.3	82.67
	F2	95	52.32
	F3	89.33	39.02
	F2-F3	94.33	105.79
/ɪ/	**F1**	54.67	27.23
	F2	166.67	100.26
/ɛ/	**F1**	106	111.30
	F2	72.5	37.79

3.2.6 Analysis of ultrasound images. Visual analysis of ultrasound images was replicated for participant 2. Similar to participant 1, there was a decrease from Baseline_1 to Treatment_1 and an increase from Treatment_1 to Baseline_2 for final /r/ productions (see Table 21). There was a slight decrease from Baseline_1 to Treatment_2. A greater an overall increase from Treatment_1 to Treatment_2 was noted. For medial /ɪ/, there was a decrease from Baseline_1 to Treatment_1. There was an increase from Treatment_1 to Baseline_2 and a decrease from Baseline_2 to Treatment_1. The final Treatment_2 mean was higher than the initial Treatment_1 mean. For medial /ɛ/, there was a decrease from the baseline to the treatment phase. The mean accuracy of medial /ɛ/ for the maintenance session was higher than the mean accuracy of this same phoneme in the initial baseline phase.

Table 21. *Participant 2 Mean Ultrasound Accuracies Across Phases*

	Final /r/	Medial /ɪ/	Medial /ɛ/
Basline$_1$	86	68.33	0
Treatment$_1$	46.67	48.25	0
Baseline$_2$	80	68.33	54
Treatment$_1$	71.42	52.75	32
Maintenance	73.33	50.67	78

Chapter 4: Summary and Discussion

Quantitative and visual analyses served as the primary means of analysis and showed improvement of speech sound accuracy of American English phonemes produced by native Mandarin speakers. Improvement from Baseline$_1$ to Treatment$_2$ was noted for both participants as SMD effect sizes all evidenced change similar to previous studies (Gick et al., 2008; Tsui, 2012). Effect sizes were greater for treated and untreated medial /ɪ/ for participant 1 and treated medial /ɛ/ showed the greatest change from Baseline$_1$. Both vowels required less tongue manipulation than final /r/ so acquisition was less challenging, resulting in greater change for these phonemes. All final /r/ SMD numbers were greatest for participant 2 with treated final /r/ showing the greatest improvement from Baseline$_1$. Medial /ɛ/ change was less for participant 2 than for participant 1. Participant 2 demonstrated difficulty recognizing the difference in tongue height for this phoneme. PND numbers were high and reflected change from Baseline$_1$ to Treatment$_2$ for all phonemes and both participants except for medial /ɛ/ for participant 2. Again, this was likely due to the challenge this phoneme presented for this particular participant. Mean accuracies were analyzed with Treatment$_2$ typically yielding higher means for all groups of both participants. Higher means by the second treatment phase reflected change from the initial baseline means and effectiveness of treatment. Both participants also told the clinician that they noticed improvement through the course of treatment and to the end of the study. Visual analog scores corresponded with dichotomous ratings, noting improvement from baseline.

This treatment resulted in generalization to untreated targets and more challenging contexts similar to studies conducted by Preston et al. (2013); Sjolie, Leece, & Preston (2016); and Tsui (2012). This was evidenced by large effect sizes for untreated phonemes and high visual analog scores for carrier phrases. Generalization to untreated and more challenging contexts was crucial to determine in order to propose this method as a potential treatment for ELLs. Anecdotally, both participants frequently expressed that the skills were carrying over and generalizing to everyday life.

This study also sought to determine maintenance of skill, unlike previous studies. All analyses noted high numbers for measures six weeks post-treatment. Moreover, both participants were able to cue themselves and explain ultrasound imaging to the clinician by the end of the study. Both participants were very receptive to ultrasound biofeedback treatment and demonstrated ability to maintain skill.

Principles of motor learning were considered as verbal feedback cues were implemented throughout treatment phases for both participants (Preston, McCabe, Rivera-Campos, Whittle, Landry, & Maas, 2014; Preston, Leece, & Maas, 2016). Interestingly, cues changed more between treatment phases for participant 1. KR cues were given most often for both participants showing skill was learned and retained quickly. Specific feedback decreased during Treatment$_2$ for both participants once the skill was learned and became habituated. Participant 2 also produced probes quickly and required reminders to slow down to benefit from verbal feedback. Although KR cues were given most often, the clinician often started Treatment$_1$ sessions with a longer explanation resembling KP that the participants were very receptive to. Moreover, treatment focused on ultrasound imaging so less specific feedback was needed during drill-like trials following the initial explanation. Both participants were able to give themselves both KP and KR cues by the end of the second treatment phase.

This study was novel by considering underlying mechanisms of accent and incorporating perceptual training. Perceptual training improvement paralleled the increases in quantitative and visual analyses. Participants also were able to perceive their own incorrect productions more accurately by the end of the study. In fact, participant 2 shared a few anecdotes about noticing perceptual differences in fellow ELL classmates' speech as he became more aware of these differences. Moreover, participant 2 occasionally looked away from the ultrasound and relied on his perception of phoneme production before looking back at the ultrasound to note tongue change.

Acoustic analysis and ultrasound image analyses were performed to examine the feasibility and value for future studies. Less change was noted with acoustic analyses. However, small change was noted for all phonemes. All phonemes were characteristic of means from previous studies by the end of treatment (Hagiwara, 1995; Chen et al., 2011). Acoustic analysis for final /r/ for participant 1 did not parallel quantitative findings, likely due to the treatment focusing on lingual configuration rather than labial rounding which is characteristic of this phoneme and potentially affects acoustic signals. Vowel formants showed the least amount of change with F2 remaining most consistent. Absolute mean difference and standard deviations were similar to those from a previous study (Georgeton, Analik, & Fourgon, 2016). Participants were unaware of acoustic analyses so these measures were not noted anecdotally by them.

In addition to acoustic analyses, ultrasound image analysis augmented quantitative and visual analyses and was based on analysis conducted by Tsui (2012). Ultrasound image analysis differed slightly from quantitative and visual analyses because the images were of single sustained phonemes rather than of the phoneme within probe productions. However, maintenance of skill was noted with high mean accuracies during the follow-up session for both participants. In addition, both participants were very adept at independently cueing their tongue movements visualized with the ultrasound by the end of the study. Both participants enjoyed using the ultrasound and noted that they benefited from the technology.

Future projections concerning LOTE in the United States and potential impacts of accent sparked a need for investigating current accent modification approaches and proposing this study (Cheng, 2000; Hosoda & Stone-Romero, 2017; Shin & Ortman, 2011). Although Mandarin is projected to be one of the most commonly LOTE spoken, very little evidence is available for accent modification approaches targeting those who speak this language, making this study novel and unique in its investigation. American English and Mandarin phonetic inventories were analyzed to determine differences which could cause a need for intervention (ASHA, n.d.; Peterson & Barney, 1952). In particular, differences between rhotics and vowels were considered. There is limited evidence of ultrasound technology for vowel treatment so this study’s investigation of vowels was also novel.

Perceptual training was included in this treatment based on the theories of accented speech (Berthnal, Bankson, & Flipsen, 2009; Schmidt, 1997; Shafer et al., 1998). Given the potential impact of accent on everyday life, improving perception was imperative to consider for effective treatment. To the knowledge of the author, this had not been implemented in previous accent modification approaches.

Ultrasound biofeedback has been effective for various populations and continued to be effective for the participants in this study (e.g., Adler-Bock et al., 2014; Bressman, Harper, Zhylich, & Kulkarni, 2016; McAllister Byun et al., 2014; Preston, Leece, & Maas, 2017; Preston et al., 2014; Preston, Brick, & Landi, 2013; Preston, Leece, et al., 2016; Preston, Leece, McNamara, & Maas, 2017; Preston, Maas, Whittle, Leece, & McCabe, 2016; Shawker & Sonies, 1985; Sjolie et al., 2016). Given the positive results of this treatment, it is likely that this technology will continue to evolve and be implemented for accent modification services.

4.1 Limitations & Future Directions

There are a number of limitations of this current investigation. For instance, the protocol of treatment was not representative of everyday speech. Treatment procedures were drill-like, only practicing targets at the single word level. While generalization to carrier phrases was analyzed, it was not treated. Actual results of the participants' speech outside of the study may not present the same observations through analysis even though the participants felt that their skills were generalizing. However, researchers can still examine stimulability and responsiveness to ultrasound biofeedback therapy as well as generalization and maintenance of skills from the results.

Baselines were not as stable as anticipated. Higher final Baseline$_1$ points could have been due to testing effects because the same probes were used every session. This also could have been due to the participants' knowledge of the purpose of the study and a high motivation to improve phoneme production accuracy.

Given the underlying mechanism of accent and the implementation of perceptual training, treatment results may have been due to perceptual training rather than ultrasound biofeedback treatment. No explicit correlation between the two were examined. Future studies should consider the implications of perceptual training to determine true cause of treatment results.

Given the nature of this study, a language barrier was noted with both participants throughout the study. Studies of non-native English speakers mentioned previously had no evidence for controlling effects of language barriers such as clearer instructions or incorporating pictures to assure comprehension (Georgeton, et. al, 2016; Gick, et. al, 2008; Tsui, 2012). This concern was considered in the current proposal. More detailed explanations with clearer language were given occasionally during the screening, diagnostic and treatment procedures. This likely did not change the expectations of performance of the participants. More explicit explanations only augmented the participants' understanding and ability to implement the procedures appropriately.

This study was also limited by sample size. This was true of several studies noted previously (Georgeton, Antolik, & Fougeron, 2016; Gick, Bernhardt, Bacsfalvi, & Wilson, 2008; Preston, Leece, & Maas, 2016; Preston, Leece, & Maas, 2016; Tsui, 2012). Despite this, results were still positive for improving accuracy for phoneme production in this study and the mentioned studies. Sample size was based on the intensity of the study. Only two participants were studied to analyze individual characteristics of phoneme production. Further exploration with larger sample sizes should be considered for greater efficacy.

The Hawthorne effect may have been present as participants were aware that they are participating in a study. The same probe list may have caused learning or test practice results which could have conflicted with the ultrasound therapy. However, probe words were randomized each session when collecting data to minimize this effect. When possible, the same treatment room was used for screening, diagnostic and treatment sessions to minimize the effect of different testing conditions. However, due to the existing speech clinic schedule, rescheduling caused different rooms to be used in a few instances. Distance between the client and laptop as well as orientation in room (e.g. facing a mirror) remained the same in all treatment rooms to control for environment effects.

This study also points to the need for continued research in the areas of (1) accent modification and (2) ultrasound biofeedback, especially for vowel objectives. Limited research is available for both of these areas. Continued research in accent modification would improve SLPs' ability to meet the communication needs of ELLs. This study also suggests that ultrasound biofeedback is beneficial for improving vowel production. Vowel production with ultrasound biofeedback should continue to be explored to provide more evidence for treatment of this class of phonemes.

As mentioned, future studies should consider larger sample sizes, not only with native Mandarin speakers but also with native speakers of other languages. Different phonetic inventories may contain phonemes that are more stimulable to ultrasound treatment. Therefore, more populations could benefit from this technology.

More in-depth analysis of ultrasound images should be considered for future studies. This study focused on analyzing sustained target phonemes, however this was not representative of the treatment probes. Future studies should consider analyzing ultrasound image analysis that more closely aligns with treatment probes.

In addition, future studies should consider correlating ultrasound image and acoustic analyses with quantitative and visual analyses. No explicit correlation was noted for this study. Although perceptual analysis with is typically the "gold standard," future studies should consider whether these analyses can be correlated to improve treatment protocol. In addition, acoustic analysis should analyze F2 to F3 distances within Mandarin /r/ productions and be compared with American English norms to determine differences for potential treatment targets.

Chapter 5: Conclusion

This study considered projections of LOTE spoken in the United States, impacts of an accent and current treatment approaches to propose a novel investigation of accent modification services for Mandarin speakers. Current accent modification approaches and theory about mechanisms underlying accent set the foundation for the targets and population. Evidence from various ultrasound biofeedback studies provided evidence for its implementation as a novel accent modification treatment approach. Perceptual, acoustic and ultrasound image analyses provided support for the effectiveness ultrasound biofeedback as a treatment for ELLs. Generalization and maintenance of skills provided further efficacy of treatment. Results from this study contributed novel evidence to existing literature about both accent modification services and ultrasound biofeedback and continued to spur the movement of study in these areas. The author also recommended additional factors to consider in future investigations. In conclusion, this study proposed ultrasound biofeedback as an effective treatment for improving production accuracy of American English phonemes for ELLs seeking to decrease impacts of foreign accent

References

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

American Speech-Language Hearing Association (2009). Consensus auditory-perceptual evaluation of voice (CAPE-V). *American Journal of Speech-Language Pathology, 18*, 124-132

American Speech-Language-Hearing Association. (n.d.). Accent management. Retrieved 18 May 2017 from http://www.asha.org/public/speech/development/Accent-Modification/

American Speech-Language-Hearing Association. (n.d.). Phonetic inventories across languages: Facts on Mandarin phonology. Retrieved 23 April 2017 from http://www.asha.org/uploadedFiles/practice/multicultural/MandarinPhonemicInventory.pdf

American Speech-Language-Hearing Association. (2016). Scope of Practice in Speech-Language Pathology [Scope of Practice]. Retrieved 23 April 2017 from http://www.asha.org/policy/SP2016-00343/

Bacsfalvi, P. (2010). Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Journal of Speech-Language Pathology and Audiology, 34*(3), 206-217.

Bacsfalvi, P., & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for three adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics, 25* (11-12), 1034-1043.

Bacsfalvi, P., & Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *International Journal of Speech-Language Pathology, 9*(1), 36-45.

Berharddt, B., Bacsfalvi, P., & Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics, 22*(2), 149-162.

Berharddt, B., Bacsfalvi, P., Gick, B., Radanov, B., & Williams, R. (2005). Exploring the use of electropalatography and ultrasound in speech habilitation. *Journal of Speech-Language Pathology and Audiology, 29*(4), 169-182.

Berharddt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics, 19*(6/7), 605-617.

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

Behrman, A. (2017). A clear speech approach to accent management. *American Journal of Speech-Language Pathology, 26*(4), 1178-1192.

Behrman, A. (2014). Segmental and prosodic approaches to accent management. *American Journal of Speech-Language Pathology, 23*(4), 546-561.

Bernthal, J. E., Bankson, N. W., & Flipsen Jr., P. (2009). *Articulation and phonological disorders: Speech sounds disorders in children*. Boston, MA: Pearson.

Beukelman, D. R. & Yorkston, K. M. (1984). *Assessment of intelligibility of Dysarthric speech* [Assessment manual]. Austin, TX: Pro-Ed.

Blyth, K., M., McCabe, P., Madill, C., & Ballard, K. J. (2016). Ultrasound visual feedback in articulation therapy following partial glossectomy. *Journal of Communication Disorders, 61*, 1-15.

Boyce, S. E., Hamilton, S. M., Rivera-Campos, A. (2016). Acquiring rhoticity across languages: An ultrasound study of differentiating tongue movements. *Clinical Linguistics & Phonetics, 30*(3-5), 174-201.

Brady, K. W., Duewer, N., & King, A. M. (2016). The effectiveness of a multimodal vowel-targeted intervention in accent modification. *Contemporary Issues in Communication Science and Disorders, 43*, 23-34.

Bressman, T., Harper, S., Zhylich, I., & Kulkarni, G. V. (2016). Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors. *Clinical Linguistics & Phonetics, 30*(3-5), 345-362.

Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology, 21*, 397-414

Cheng, L.L. (2000). A multicultural millennium. *Advances in Speech-Language Pathology, 2*, 131-133.

Chen, Y., Robb. M., Gilbert, H., & Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics & Phonetics, 15*(6), 427-440.

Cleland, J., Crampin, L., Wrench, A. A., Zharkova, N., Lloyd, S. (2017). *Visualising speech: Using ultrasound visual biofeedback to diagnose and treat speech disorders in children with cleft lip and palate*. Paper presented at the Royal College of Speech and Language Therapists Conference, Gasgow, Scotland. Retrieved from http://strathprints.strath.ac.uk/61041/

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics, 29*(8-10), 575-597.

Compton, A.J. *Compton P-ESL Program*, Carousel House, San Francisco, CA (1984).

Compton, A. J. (2008). *Compton Phonological assessment of foreign accent* [Assessment manual]. San Francisco, CA: Institute of Language and Phonology.

Dunn, L. B. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.) [Assessment manual]. San Antonio, TX: Clinical Assessment.

Fawcett, S., Bacsfalvi, P., Bernhardt, B. (2008). Ultrasound as visual feedback in speech therapy for /r/ with adults with Down syndrome. *Down Syndrome Quarterly, 10*(1), 4-12.

Fitch, J. (2000). Accent reduction: A corporate enterprise. *Advances in Speech-Language Pathology, 2*, 135-137

Franklin, A. & McDaniel, L. (2016). Exploring a phonological process approach to adult pronunciation training. *American Journal of Speech-Language Pathology, 25*, 172-182.

Fritz, D. R., Sikorski, L. D. (2013). Efficacy in accent modification services: Quantitative and qualitative outcomes for Korean speakers of American English. [Spechial Interest Group 14 Perspectives] Retrieved May 15, 2017 from http://sig14perspectives.pubs.asha.org/

Georgeton, L., Antolik, T. K., & Fougeron C. (2016). Effect of domain initial strengthening on vowel height and backness contrasts in French: Acoustic and ultrasound data. *Journal of Speech, Language and Hearing Research*, *59*, 1575-1586.

Gick, B., Berhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M.L. Zampini (Eds.), *Phonology and Second Language Acquisition* (309-322). British Columbia: John Benjamins.

Hack, J., Marinova-Todd, S. H., & Bernhardt, B. M. (2012). Speech assessment of Chinese-English bilingual children: Accent versus developmental level. *International Journal of Speech-Language Pathology, 14*(6), 509-519.

Hagiwara, R. (1995).Acoustic realizations of American /r/ as produced By women and Men. UCLAWorking Papers in Phonetics, 90, 1–187.

Kay Elemetrics, 1994, *CSL Computerized Speech Lab: Model 4300B, Software ver, 5.X.* (Lincoln Park, NJ: Kay Elemetrics Corporation).

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

Kempster, G.B., Gerratt, B.R., Verdonlini Abbott, K., Barkmeier-Kraemer, J. & Hillman, R.E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*, 124-132.

Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J. R., Odom, S.L., Rindskopf, D.M., & Shadish, W.R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lam , J. & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language and Hearing Research*, *56, 14*29-1440.

Lippke, B. A., Dickey, S. E., Selmar, J. W. & Soder, A. L. (1997). Photo Articulation Test. [Assessment manual]. Austin, TX: Pro-Ed.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*, 277-298.

McAllister Byun, T. & Hitchcock, E. R. (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology, 21*, 207-221.

McAllister Byun, T., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultraounsd biofeedback intervention. *Journal of Speech, Language and Hearing Research, 57*, 2116-2130.

McAllister Byun, T. (2017). Efficacy of visual-acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language and Hearing Research, 60*, 1175-1193.

McSweeny, J. L. & Shriberg, L. D. (2001). Clinical research with the prosody-voice screening profile. *Clinical Linguistics & Phonetics, 15*(7), 505-528.

Mihalicek, V. & Wilson, C. (Eds.). (2011). *Language files: Materials for an introduction to language and linguistics*. Columbus, OH: The Ohio State University Press.

Muller, N., Ball, M. J., & Guendouzi, J. (2000). Accent reduction programmes: Not a role for speech-language pathologists? *Advances in Speech-Language Pathology*, *2*, 199-129.

Morton, E. S., Brundage, S. B. & Hancock, A. B. (2010). Validity of the Proficiency of Oral English Communication screening. *Contemporary Issues in Communication Science and Disorders, 37*, 153-166.

Olive, M. L. & Smith, B. W. (2005). Effect sizes calculations and single subject designs. *Educational Psychology, 25*(2-3), 313-324.

Patal, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell (2013). "The caterpillar": A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology, 22*, 1-9.

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

Preston, J. L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology, 22,* 627- 643.

Preston, J. L., Holliman-Lopez, G., Leece, M. C. (2018). Do participants report any undesired effects in ultrasound speech therapy?*American Journal of Speech-Language Pathology, 27*, 813-818.

Preston, J. L., & Leaman, M. (2014). Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology, 28*(3), 278-295.

Preston, J. L., Leece, M. C., & Maas, E (2016). Intensive treatment with ultrasound visual biofeedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience*, *10*, 1-9.

Preston, J. L., Leece, M. C., & Maas, E. (2017). Motor-based treatment with and without ultrasound feedback for residual speech-sound errors. *International Journal of Language and Communication Disorders, 52*(1), 80-94.

Preston, J. L., Leece, M. C., McNamara, K., & Maas, E., (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology, 26*(3), 840-852.

Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language and Hearing Research*, *57*, 2102-2115.

Rusiewicz, H.L., DeLisio, P., & Maize, K. (2016, November). *Ultrasound biofeedback & the treatment of two school-age girls with speech sound disorders*. Poster presented at the American Speech-Language-Hearing Association Annual Convention, Philadelphia, PA.

Rusiewicz, H. L., & Rivera, J. L. (2017). The effect of hand gesture cues within the treatment of/r/for a college-aged adult with persisting childhood apraxia of speech. American Journal of Speech-Language Pathology, 26(4), 1236-1243.

Schmidt, A. M. (1997). Working with adult foreign accent: Strategies for intervention. *Journal of Speech and Hearing Research*, n.p.

Schmidt, A. M., Meyers, K. A. (1995). Traditional and phonological treatment for teaching English fricative and affricates to Koreans. *Journal of Speech and Hearing Research, 38*, 828-838.

Shawker, T. H., & Sonies, B. C. (1985). Ultrasound biofeedback for speech training: Instrumentation and preliminary results. *Investigative Radiology, 20*(1), 90-93.

Shin, H. B., & Ortman, J. (2011, April 21). *Language projections: 2010 to 2020.* Paper presented at the Federal Forecasters Conference, Washington, DC. Retrieved from https://www.census.gov/hhes/socdemo/language/data/acs/Shin_Ortman_FFC2011_paper. Pdf

Shriberg, L. D., Kwiatkowski, J. and Rasmussen, C. (1990). *The Prosody-Voice Screening Profile: Scoring forms and training materials.* Communication Skill Builders.

Shriberg, L. D., Kwiatkowski, J. and Rasmussen, C. (1990). *The Prosody-Voice Screening Profile* (Tucson, AZ: Communication Skill Builders).

Sikorski, L. D. (2005). Regional accents: A rationale for intervening and competencies required. *Seminars in Speech and Language, 26*(2), 118-125.

Sikorski, L. D. (2007). *Proficiency in Oral English Communication mauual* (4th ed., electronic ed.). Santa Ana, CA: LDS & Associates.

Sjolie, G. M., Leece, M. C., & Preston, J. L. (2016). Acquisition, retention and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders, 64*, 62-77.

Smiljanic, R. & Bradlow, A. R. (2007). Clear speech intelligibility: Listener and talker effects. *ICPhS XVI, 661-664.*

Smith, J. G. (2010). *Acoustic properties of English /l/ and /r/ produced by Mandarin Chinese speakers* (Master's forum paper). Retrieved from Academia. http://www.academia.edu/4225207/Acoustic_Properties_of_English_l_and_%C9%B9_Produced_by_Mandarin_Chinese_Speakers

Tsui, H. M-L. (2012). Ultrasound speech training for Japanese adult learning English as a second language (Unpublished thesis). University of British Columbia, Vancouver.

Van Riper, C. & Erickson, R. (1996). *Speech correction: An introduction to speech pathology and audiology* (9th ed.). Boston: Allyn & Bacon. Reprinted with permission.

Appendix A

Effectiveness of Visual Biofeedback in the Improving Intelligibility of American English Accent

Eligible________ Not Eligible______

Participant's Name:__________________________

Phone #:______________________Alt Phone#:_______________________

Email address:_________________________________

DOB:________________________Current Age________________________

How did you hear about the study?_________________________________

Are you between 18;0 and 30;11 years of age? _____YES _____NO

Do you speak Mandarin as your first language? _____YES _____NO

Do you speak English as your second language? _____YES _____NO

Do you speak any other languages? _____YES _____NO

If yes, please specify third language: ____________________________

Have you been evaluated/treated by an SLP? _____YES _____NO

Are there concerns about your speech production or American English intelligibility? _____YES _____NO

If so, what are they? ____

Have you been seen/treated by an audiologist? _____YES _____NO

Are there concerns about your hearing? _____YES _____NO

Any history of medical (cleft palate, etc), developmental (MR, etc) or neurological problems (cerebral palsy)? _____YES _____NO

If yes to above explain:____

What are the best times to be contacted? ___________________________

What are the best times to schedule the appointments (days, am or pm):

Monday____________ Tuesday____________ Wednesday_____________

Thursday________________ Friday_______________

English Language Experience

How many years have you lived in an English-speaking country?________

At what age and where were you first *exposed* to the English language?

At what age and where were you first *immersed* in an English-speaking environment?

____

Have you ever had instruction in English pronunciation before? If so, for how long?

How often do you speak English in your daily life?

100% 75% 50% 25% or less

Where do you speak English most often?

Home School Work Other(please specify_____________)

Is/are there an English sound(s) that is easiest for you?

____

Is/are there an English sound(s) that is hardest for you?

____

How motivated are you to participate and practice (1-not motivated, 10-very motivated)?

1 2 3 4 5 6 7 8 9 10

How would you rate your English pronunciation (1-poor, 10-excellent)?

1 2 3 4 5 6 7 8 9 10

What are your expectations for participating in this study?

For researchers: Initial appointment Date and Time:

____

Appendix B

Perceptual Training Words

1. Simmer
2. Finger
3. Dollar
4. Father
5. Mother
6. Clammer
7. Fatter
8. Better
9. Beware
10. Sever
11. Poor
12. Fair
13. Fear
14. Near
15. Dare
16. Cure
17. Star
18. Store
19. Care
20. Far
21. Kiss
22. Give
23. Quick
24. Fix
25. Live
26. Zip
27. Whip
28. Big
29. Pit
30. Sip
31. Wedding
32. Picnic
33. Fitness
34. Mitten
35. Credit
36. Quickest
37. Practice
38. Classic
39. Visit
40. Rabbit

Appendix C

Participant 1 Probes Session_________ Sample Data Sheet

Final /ɾ / Multisyllabic

Treatment/Baseline	Trials (+/-)			Errors
1. Summer				
2. Singer				
3. Doctor	▨	▨	▨	
4. Feather				
5. Matter	▨	▨	▨	
6. Hammer	▨	▨	▨	
7. Ladder				
8. Letter				
9. Before				
10. Never	▨	▨	▨	
11. Alligator				
12. Flower				
13. Admirer				
14. Explorer	▨	▨	▨	
15. Anywhere				

Perceptual Training		Trials (+/-) Response **Accuracy**	
1. Simmer			
2. Finger			
3. Dollar			
4. Father			
5. Mother			
6. Clammer			
7. Fatter			
8. Better			
9. Beware			
10. Sever			

Medial /ɪ / Multisyllabic

Treatment/Baseline	Trials (+/-)			Errors
1. Running				
2. Finish	▨	▨	▨	
3. Liquid				
4. Kitten				
5. Listen				
6. Ticket	▨	▨	▨	
7. Active				
8. Fabric	▨	▨	▨	
9. Metric				
10. Tennis				
11. Friendship	▨	▨	▨	
12. Gymnastics				
13. Typical				
14. Analysis	▨	▨	▨	
15. Symphony				

Perceptual Training		Trials (+/-) Response **Accuracy**	
1. Wedding			
2. Picnic			
3. Fitness			
4. Mitten			
5. Credit			
6. Quickest			
7. Practice			
8. Classic			
9. Visit			
10. Rabbit			

BOLD-probed and treated
Italicized-not probed, treated
Regular-probed, not treated

Appendix D

Participant 2 Target Words, Session ________ Sample Data Sheet

Final /r/ Monosyllabic

Treatment/Baseline	Treatment Trials with Feedback(+/-)										Errors
1. Singer											
2. *Doctor*											
3. *Matter*											
4. *Hammer*											
5. Ladder											
6. Before											
7. *Never*											
8. Alligator											
9. Admirer											
10. *Explorer*											

Medial /ɪ/ Monosyllabic

Treatment/Baseline	Treatment Trials with Feedback(+/-)										Errors
1. *Finish*											
2. Liquid											
3. Listen											
4. *Ticket*											
5. *Fabric*											
6. Metric											
7. *Friendship*											
8. Gymnastics											
9. *Analysis*											
10. Symphony											

Medial /ɛ/ Multisyllabic

Treatment/Baseline	Treatment Trials with Feedback(+/-)										Errors
1. *Guest*											
2. Pesky											
3. Wreck											
4. *Realm*											
5. *Chest*											
6. Read											
7. *Weather*											
8. Message											
9. *Healthy*											
10. Leathery											

Appendix E

Session a/b #____

Overall Accuracy

Instructions: Mark an X closest to the perceived accuracy of the probe.

Legend: NC-not at all correct
C-correct

Final /ɾ/ Monosyllabic ____ ____/100

NC **C**

Medial /ɪ/ Monosyllabic ____ ____/100

NC **C**

Final /ɾ/ Multisyllabic ____ ____/100

NC **C**

Medial /ɪ/ Multisyllabic ____ ____/100

NC **C**

Medial /e/ ____ ____/100

NC **C**

Final /ɾ/ Multisyllabic, carrier phrase

____ ____/100

NC **C**

Medial /ɪ/ Multisyllabic, carrier phrase

____ ____/100

NC **C**

Appendix F

Acoustic Analysis Participant 1 Sample Data Analysis Sheet

Final /ɾ/ Multisyllabic

Summer

	B$_1$ a15			B$_1$ a13			B$_1$ a11			B$_1$ a9		
F1												
F2												
F3												
Avg F1												
Avg F2												
Avg F3												
F2 F3 Dist												

	Tx$_1$ a18			Tx$_1$ a16			Tx$_1$ a14			Tx$_1$ a12		
F1												
F2												
F3												
Avg F1												
Avg F2												
Avg F3												
F2 F3 Dist												

	B$_2$ a7			B$_2$ a5			B$_2$ a3			B$_2$ a1		
F1												
F2												
F3												
Avg F1												
Avg F2												
Avg F3												
F2 F3 Dist												

	Tx$_2$ a10			Tx$_2$ a8			Tx$_2$ a6a			Tx$_2$ a4		
F1												
F2												
F3												
Avg F1												
Avg F2												
Avg F3												
F2 F3 Dist												

	Mt		
F1			
F2			
F3			
Avg F1			
Avg F2			
Avg F3			

Appendix G

Acoustic Analysis Participant 1 Sample Data Analysis Sheet

Medial /ɪ/ Multisyllabic

Active

	B$_1$ a15			B$_1$ a13			B$_1$ a11			B$_1$ a9		
F1												
F2												
Avg F1												
Avg F2												

	Tx$_1$ a18			Tx$_1$ a16			Tx$_1$ a14			Tx$_1$ a12		
F1												
F2												
Avg F1												
Avg F2												

	B$_2$ a7			B$_2$ a5			B$_2$ a3			B$_2$ a1		
F1												
F2												
Avg F1												
Avg F2												

	Tx$_2$ a10			Tx$_2$ a8			Tx$_2$ a6a			Tx$_2$ a4		
F1												
F2												
Avg F1												
Avg F2												

	Mt		
F1			
F2			
Avg F1			
Avg F2			

Appendix H

Acoustic Analysis Participant 1 Sample Data Analysis Sheet

Medial /ɛ/

Head

	B_2 a3			B_2 a1		
F1						
F2						
Avg F1						
Avg F2						

	Tx_2 a10			Tx_2 a8			Tx_2 a6a			Tx_2 a4		
F1												
F2												
Avg F1												
Avg F2												

	Mt		
F1			
F2			
Avg F1			
Avg F2			

Appendix I

Ultrasound Image Analysis Sample Analysis Sheet

Instructions: Mark an X closest to the perceived accuracy of the probe.

Picture #_____

/ɾ/ ____ ____/100

Retroflexed/ Bunched **Undifferentiated**

Picture #_____

/ɪ/ ____ ____/100

Low **High**

Picture #_____

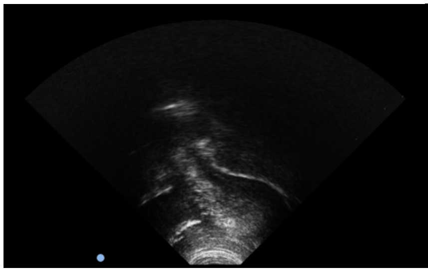
/ɛ/ ____ ____/100

Low **High**

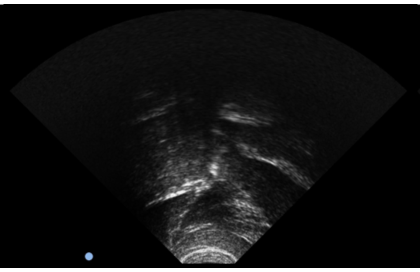
Appendix J

“Gold Standards”

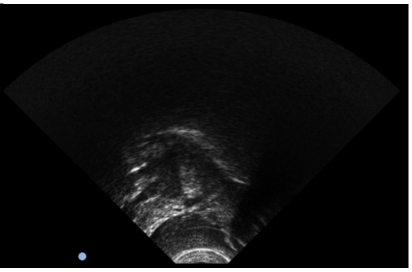
/r/



Retroflexed

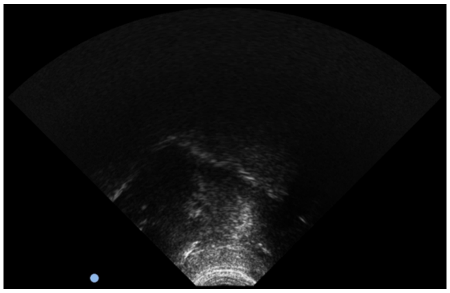


Bunched

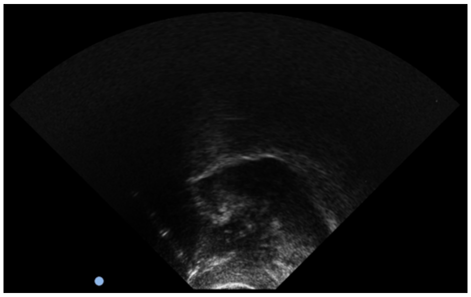


Undifferentiated

/ɪ/

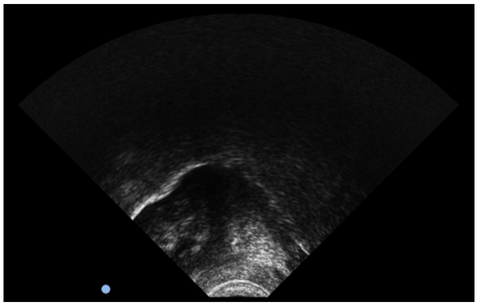


Low

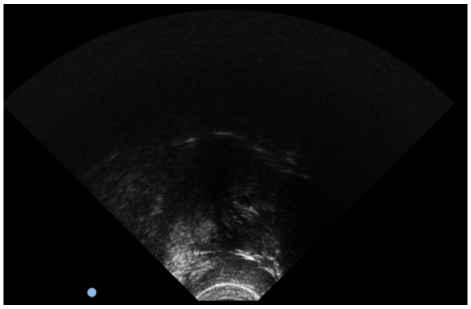


High

/ɛ/



Low



High