Published in final edited form as:

Am J Speech Lang Pathol. 2013 August ; 22(3): 540–553. doi:10.1044/1058-0360(2013/12-0137).

A Multidimensional Investigation of Children's /r/ Productions: Perceptual, Ultrasound, and Acoustic Measures

Harriet B. Klein^a, Tara McAllister Byun^a, Lisa Davidson^a, and Maria I. Grigos^a

^aNew York University, NY

Abstract

Purpose—This study explored relationships among perceptual, ultrasound, and acoustic measurements of children's correct and misarticulated /r/ sounds. Longitudinal data documenting changes across these parameters were collected from 2 children who acquired /r/ over a period of intervention and were compared with data from children with typical speech.

Method—Participants were 3 children with typical speech, recorded once, and 2 children with /r/ misarticulation, recorded over 7–8 months. The following data from /r/ produced in nonwords were collected: perceptually rated accuracy, ultrasound measures of tongue shape, and F3 – F2 distance.

Results—Regression models revealed significant associations among perceptual, ultrasound, and acoustic measures of /r/ accuracy. The inclusion of quantitative tongue-shape measurements improved the match between the ultrasound and perceptual/acoustic data. Perceptually incorrect /r/ productions were found to feature posteriorly located peaked tongue shapes. Of the children who were seen longitudinally, 1 developed a bunched /r/ and 1 demonstrated retroflexion. The children with typical speech also differed in their tongue shapes.

Conclusion—Results support the validity of using qualitative and quantitative ultrasound measures to characterize the accuracy of children's /r/ sounds. Clinically, findings suggest that it is important to encourage pharyngeal constriction while allowing children to find the /r/ tongue shape that best fits their individual vocal tract.

Keywords

acoustics; articulation; speech production; speech sound disorders

The phoneme /r/ is one of the last to be achieved in typically developing children (e.g., Smit, Hand, Freilinger, Bernthal, & Bird, 1990), is almost always found to be among the error sounds in children with phonological disorders (e.g., Hodson & Paden, 1991), and is misarticulated frequently by individuals with residual speech sound errors (e.g., Preston & Edwards, 2009). As a frequently occurring error sound, its remediation has presented a challenge to clinicians for decades (Bernthal, Bankson, & Flipsen, 2009). The difficulty that children experience in mastering rhotic sounds can be attributed at least in part to the

complexity of the articulatory configuration for /r/ (Gick et al., 2008). For most speech sounds, the tongue forms only one major constriction or narrowing of the vocal tract. By contrast, English /r/ is produced with two major lingual constrictions: one formed with the anterior tongue and one involving the tongue root; lip rounding also is generally present (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007).

Due to its complex character, the production of /r/ has been extensively studied using a range of methodologies. Technologies used include x-ray imaging (e.g., Delattre & Freeman, 1968; Westbury, Hashi, & Lindstrom, 1998), electromagnetic midsagittal articulometry (e.g., Boyce & Espy-Wilson, 1997; Guenther et al., 1999), magnetic resonance imaging (e.g., Alwan, Narayanan, & Haker, 1997; Tiede, Boyce, Holland, & Choe, 2004; Zhou et al., 2008), and ultrasound imaging (Gick et al., 2008). Most of these studies have paired articulatory descriptions with acoustic analyses in order to establish the relationship between tongue shapes and auditory-acoustic outputs for /r/. These studies have revealed an intricate relationship between perception and production: Multiple distinct tongue shapes, including the well-known *retroflex* and *bunched* configurations, converge on a characteristic acoustic signal featuring a lowered third formant (Delattre & Freeman, 1968; Westbury et al., 1998).

It is thought that differences in the tongue shape used to produce /r/ may be linked to individual variation in vocal tract morphology (e.g., Tiede et al., 2004). Given that children's vocal tracts diverge from adults' in both size and general configuration, there is reason to expect even greater differences in the context of children's tongue shapes for /r/. However, no study has systematically documented the tongue shapes associated with correct and incorrect /r/ production in child speakers or tracked longitudinal changes in tongue shape as children with /r/ misarticulation begin to produce /r/ accurately over the course of treatment. The present research was undertaken with the primary goal of using ultrasound imaging to observe articulatory changes in two children with /r/ misarticulation as they were followed through a period of intervention. Associated changes in perceptually rated accuracy and acoustics were evaluated in parallel. For comparison, measurements were also collected from three control participants with typical speech (TS).

Characteristics of English /r/

Variants of /r/

The American English rhotic phoneme /r/ has several variants. It can appear as the nucleus of a syllable, in which case it is transcribed /ɹ/ (as in *bird*) or /ɹ̥/ (as in *water*), or as a consonant in syllable-onset position (e.g., *red*, *tree*). Finally, /r/ can occur in postvocalic position (e.g., *care*, *fear*), where it has been characterized as the vocalic off-glide of a rhotic diphthong (e.g., McGowan, Nitttrouer, & Manning, 2004). Our study examined syllabic /ɹ̥/ and onset /r/, which will be referred to as *vocalic* /r/ and *consonantal* /r/ in the text to follow.

Acoustics

English /r/ is distinguished from other sonorant phonemes by the low height of the third formant, F3 (Delattre & Freeman, 1968; Hagiwara, 1995). In many cases, the third formant is so low that it appears to merge with the second formant, F2 (Boyce & Espy-Wilson, 1997;

Dalston, 1975). Research by Shriberg, Flipsen, Karlsson, and McSweeny (2001) and Flipsen, Shriberg, Weismer, Karlsson, and McSweeny (2001) reported that adolescents' correct and incorrect productions of /r/ could most sensitively be characterized by means of the distance between F2 and F3. Therefore, the acoustic measurements reported in this paper will focus on this parameter of F3 – F2 distance. The acoustic properties of /r/ are largely consistent across vocalic and consonantal variants, although F3 lowering may be more extreme in consonantal /r/ (McGowan et al., 2004; cf. Hagiwara, 1995). In child speakers, mastery of the acoustics of /r/ generally unfolds in a gradual fashion, with both F3 height and F3 – F2 distance decreasing steadily over the course of childhood (Lee, Potamianos, & Narayanan, 1999; McGowan et al., 2004).

Articulation of /r/ in adult speakers

As noted earlier, typical adult speakers produce /r/ with both an anterior lingual constriction, where the tongue approximates a point near the palate, and a posterior constriction, in which the tongue root retracts into the pharynx (Adler-Bock et al., 2007). Lip rounding is also generally present, at least in onset position. Tongue postures for /r/ are frequently divided into two major categories. In *retroflex* /r/, the tongue tip raises and curls back slightly in the vicinity of the alveolar ridge. In *bunched* /r/, the tongue tip lowers while the tongue body raises to approximate the hard palate. However, it is now known that typical adults may produce /r/ with tongue shapes that do not fit readily into either category (e.g., Tiede et al., 2004). It is argued that despite this variation in tongue shape across speakers, vocal tract shapes for /r/ are unified in featuring two lingual constrictions that separate the vocal tract into three resonating cavities, yielding the characteristic acoustic signature of /r/ (e.g., Alwan et al., 1997). Tiede et al. (2004) argued that the configuration of the tongue needed to produce the characteristic acoustic profile of /r/ is influenced by the volume and shape of the speaker's vocal tract. To understand individual differences in /r/ production, it is necessary to supplement perceptual and acoustic measures with articulatory data, such as ultrasound tongue-shape measurements.

Articulation of /r/ in children

When children misarticulate /r/, their output is commonly described in terms of substitution of another English phoneme. For consonantal /r/, the perceived substitute is typically /w/; for vocalic /r/, common substitutions include /ə/ and /ɔ/. However, a considerable literature has shown that children's /r/ errors are rarely true phonemic substitutions such as /w/ for /r/ (Hoffman, Stager, & Daniloff, 1983; Klein, Grigos, McAllister Byun, & Davidson, 2012). These intraphonemic errors are often described as distorted, derhotacized, or lacking in /r/ coloring (Shriberg & Kent, 2003). Close listening reveals a great deal of across- and within-child variation in these distorted realizations of /r/ targets (e.g., Ball, Lowry, & McInnis, 2006). The existence of such variability heightens the need for imaging data that will make it possible to describe the articulatory properties of children's incorrect /r/ sounds.

There is a lack of basic research studies using ultrasound or other imaging technologies to document tongue shapes for /r/ in young children. However, studies in which ultrasound imaging has been used to provide biofeedback intervention for /r/ misarticulation (e.g., Adler-Bock et al., 2007; Bacsfalvi, 2010; Bernhardt et al., 2008) have amassed images of

children's tongue shapes for /r/ with varying levels of perceptual accuracy. Gick et al. (2008) used ultrasound images from several previous treatment studies (Adler-Bock et al., 2007; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Modha, Bernhardt, Church, & Bacsfalvi, 2008) to discuss the characteristics of tongue shapes seen in children's correct and misarticulated /r/. Gick et al. reported that some child speakers produce retroflex or bunched tongue shapes that qualitatively resemble adult tongue shapes for /r/. They hypothesized that children's difficulty mastering /r/ reflects the gestural complexity of the sound, namely, the need to form simultaneous pharyngeal and oral tongue constrictions. These authors suggested that children's early attempts to produce /r/ tend to feature *undifferentiated* tongue shapes characterized by a single major constriction. Over the course of maturation, most children gain the ability to form two simultaneous lingual constrictions, and undifferentiated tongue shapes are replaced by normal /r/ shapes with two constrictions.

Although the above-described studies have made important contributions to our understanding of /r/ development in children, there remains a need for research systematically documenting the relationships among perceptual, ultrasound, and acoustic characteristics of /r/ sounds produced by child speakers. The present paper addresses this gap with a multidimensional look at both correct and misarticulated /r/ sounds produced by five children ages 5–8 years. In another novel contribution to the literature, the present study collected longitudinal data to document changes in tongue shape, perceptual ratings, and acoustic measurements of /r/ sounds produced by two children who acquired the /r/ phoneme through a period of intervention for /r/ misarticulation. This research is clinically relevant because clinicians rely on their own knowledge of the articulatory properties of /r/ when providing treatment for children who misarticulate this sound. However, the understanding of /r/ articulation that clinicians acquire through verbal descriptions and static images may be incomplete or inaccurate. In previous studies, practicing clinicians have reported that their exposure to ultrasound imaging of /r/ production enhanced their ability to provide traditional treatment by making them more adept at devising verbal cues to describe correct tongue placement for /r/ (Bernhardt et al., 2008).

Method

Participants

This study included a total of five child participants who were all male, Caucasian monolingual English speakers. Two of the children, who presented with impaired speech (IS) production, were recruited from the New York University Speech and Hearing Clinic. Before being accepted for treatment in the clinic, these children had completed a diagnostic evaluation that included an assessment of their hearing, phonology, and speech-motor behavior. Based on measures of the accuracy and intelligibility of their speech in single-word and continuous speech samples, these children were diagnosed with moderate-to-severe phonological impairment by a certified speech-language pathologist (SLP). Neither child demonstrated identifiable impairments in hearing, motor control, or cognition. Both children made progress in treatment, and at the start of this study, the children displayed only errors affecting vocalic and consonantal /r/, which they were judged to produce with 0% accuracy. They had not previously received intervention targeting the /r/ sound.

At the outset of this study, participant IS1 was age 5; 8 (years;months), and participant IS2 was age 6;11. A comparison group consisting of three children with accurate /r/ production (TS) was also recruited. The three TS children were ages 4;11, 7;9, and 7;11. All were drawn from the university community and had achieved /r/ production within expected limits. Although the TS children represented a range of ages, this study was not designed to measure changes in the properties of correct /r/ over the course of typical development. These children were recruited only to provide examples of the properties of perceptually correct child /r/ so as to serve as a point of comparison with the IS children.

Procedure

Stimuli—The stimuli, presented by the first author across all participants and sessions, featured both vocalic and consonantal variants of /r/ in phrases for imitation. Vocalic /r/ was elicited in isolation (/ɜː/), following the voiced plosives /b, d, g/ (/bɜː, dɜː, gɜː/) and preceding the same set of plosives (/ɜːb, ɜːd, ɜːg/). Consonantal /r/ was elicited as the singleton onset of a syllable whose nucleus was /æ/ and whose coda belonged to the set of voiceless plosives /p, t, k/ (/ræp, ræt, ræk/). All stimuli were elicited in the carrier phrase “Say _____ again.”

Participants repeated each carrier phrase and target 10 times, and three of those repetitions were randomly chosen for analysis. Targets were elicited in an order that varied the occurrence of syllable types. The number of productions in each category varied as follows: 123 tokens of consonantal /r/ in the CVC (consonant-vowel-consonant) context, 125 tokens of vocalic /r/ in the CV context, 126 tokens of vocalic /r/ in the VC context, and 42 tokens of isolated vocalic /r/ (/ɜː/). We chose to use nonword stimuli rather than more naturally occurring speech in order to control the phonetic environment of the target sounds and ensure that there were sufficient numbers of target productions for our measures.

Treatment—The intervention targeting /r/ production in participants IS1 and IS2 was administered weekly and was consistent with a sensorimotor approach (e.g., Bernthal et al., 2009). Intervention consisted of inter- and intrapersonal discrimination exercises, both before and interspersed with attempts at production. Production targets were initiated with /ɜː/ in isolation. After /ɜː/ was established in isolation, it was combined with a final consonant, then with an initial consonant. Generalization proceeded with monosyllabic words and phrases, followed by the introduction of syllable-initial /r/. When the children had difficulty matching /ɜː/ by imitation alone, a tactile cue for placement was sometimes provided through application of a *toothette* to the tongue tip. This device was placed against the tongue tip and was used to push the tongue back gently. With the tongue in a retracted position, the child’s /ɜː/ generally became a better perceptual match for the clinician’s model. The children were then encouraged to attempt to emulate the feeling of tongue retraction that was induced by the toothette. The children were never specifically told to bunch the tongue or retroflex the tongue tip; they were only told to make the tongue tense or to pull it back.

Data collection—Ultrasound and audio data collection took place in a sound-treated booth in the communicative sciences and disorders department at New York University. All of the

participants were seen for two hour-long baseline sessions. The children's speech, language, and hearing skills were screened during the first baseline session. In the second baseline session, the children were engaged in playful interactions that were intended to familiarize them with the investigators, experimental tasks, and materials, including the ultrasound transducer. The children were then introduced to the target utterances, which were described earlier. Baseline data representing each child's current, typical production of /r/ were collected in this session. The TS children's participation in the project was completed after this session because they were already producing vocalic and consonantal /r/ correctly, and there was no expectation that their /r/ production would change significantly over the time course of the study. The IS children, however, continued to complete ultrasound data collection sessions every 4 weeks until the end of the treatment period (7–8 months, excluding summer months). IS1 and IS2 were recorded six and five times, respectively.

Ultrasound imaging was performed using a Sonosite Titan ultrasound with a 5- to 8-MHz transducer. Simultaneous acoustic recordings were made using a Marantz PMD-670 digital recorder with a Shure Beta 58A microphone positioned ~3 in. to the side of the speaker's mouth. children were seated on their parent's lap during all recordings. The ultrasound transducer was coated with water-based gel and was held in place beneath the child's chin by the parent; the experimenter then fine-tuned the placement of the transducer until an optimal image was achieved. The parent held the probe throughout the data collection procedure. Ultrasound images were saved to a computer using a Canopus ADVC-1394 capture card and Adobe Premiere software.

Analyses

Perceptual ratings—Each /r/ token was independently rated by two SLPs certified by the American Speech-Language-Hearing Association (the first and fourth authors) who have expertise working with children with speech sound disorders. Raters used a 3-point scale where a rating of 1 indicated a *standard* /r/ production, 2 represented an *approximation* of /r/, and 3 represented either a *severe distortion* with no rhotic quality or a *different phoneme* (e.g., [ɔ] or [w]). Reliability testing revealed that point-to-point agreement between the two raters was 85% (353/416 tokens). The kappa statistic was calculated to be 0.72 (95% CI [0.66, 0.78]; $p < 0.001$), indicating substantial agreement between the raters (Landis & Koch, 1977). Both raters reviewed the 63 tokens that were not initially agreed on, using a modification of Shriberg, Kwiatkowski, and Hoffmann's (1984) consensus procedure. Consensus was reached for 84% (53/63 tokens). The remaining 10 tokens were excluded from further analysis, yielding a data set of 406 tokens. Further details on the procedure for perceptual rating and reliability checking can be found in Klein et al. (2012).

Acoustic measurements—Acoustic measurements were carried out by a trained graduate student using Praat software (Boersma & Weenink, 2010). The student manipulated settings in Praat's automated linear predictive coding (LPC) formant tracking function until the automatically calculated formants were well matched to the visible areas of energy concentration in the spectrogram. She then selected a point that was judged to represent the minimum height of F3 in the /r/ target interval, avoiding points that appeared as outliers relative to adjacent points in the automatic LPC formant track. The height of F2

and F3 (Hz) was calculated for a 14-ms Hamming window around the selected point. For reliability purposes, >30% of the sample (148 tokens) was remeasured by a second trained graduate student. A *t* test for paired samples showed no significant difference in the two students' F2 and F3 measurements. The students were also asked to rate their degree of confidence that the measured values reflected the actual formants of the vocal tract for each token. The second author remeasured 106 tokens for which the students indicated low confidence, and 23 were discarded due to poor quality. Finally, tokens whose F2 or F3 fell >2 *SDs* away from the sample mean for that formant also were discarded. A total of 381 tokens remained following this processing. Additional details on the procedure used to obtain and verify acoustic measurements can be found in Klein et al. (2012).

Ultrasound measurements—For each of the ultrasound images collected, a trained research assistant assigned a qualitative tongue-shape coding using one of three different labels: *bunched*, *retroflex*, or *undifferentiated*. To qualify as a *bunched* shape, the tongue had to contain an anterior constriction and a retraction in the tongue root, with a concave shape between the two constrictions. For *retroflex* shapes, there had to be visual evidence of the tongue tip curling up; this is often shown on ultrasound images by a reflection above the top of the tongue tip where the tongue tip is touching the alveolar ridge. Finally, *undifferentiated* shapes had to be nearly semicircular shapes that did not contain any apparent alveolar constriction or retraction of the tongue root. An example of each of the three coding types is provided in Figure 1. In the images in Figure 1, the tongue appears as a bright white line on a black background. Two shadows are sometimes seen flanking the image of the tongue slice on either side; the left shadow is cast by the hyoid bone, and the right is cast by the jaw, because bone refracts the ultrasonic beam. In all of the images used here, the tongue tip is on the right, and the tongue root is on the left.

Of the 416 ultrasound files originally reviewed, 29 were discarded: 11 due to poor image quality, 10 due to errors in the computer algorithm used to track the edges of the tongue (see below), and eight due to experimenter error in which the original coding data were lost. In addition, 35 tokens for which ultrasound data were collected were eliminated because either perceptual or acoustic data were not available. The remaining 352 tokens (85 consonantal /r/ and 267 vocalic /r/) were used in the analyses reported in the following paragraphs. To measure reliability, the third author recoded a 329-item subset of the items for which complete data were available. The third author and original coder were in agreement for 303/329 items (92%). The second author then coded the 26 disputed items, and the label that was agreed on by two of the three coders was adopted.

For the images that were coded as undifferentiated, a quantitative analysis was carried out using methods reported in Ménard, Aubin, Thibeault, and Richard (2012). Ultrasound studies involving children generally use mobile probe placement, in which the transducer moves up and down with the motion of the jaw. However, deviations or translations in probe placement can occur, posing a challenge for accurate measurement. Ménard et al. assessed the validity of ultrasound measurements using mobile probe placement, specifically testing how horizontal, vertical, or rotational translation of the probe affects tongue measurements for children's vowel productions. Ménard et al. found that two measures, tongue curvature

degree and tongue curvature location, were relatively robust to translation in all dimensions. We adopted this method of measuring children's tongue shapes for use in our study.

The shape of each token that was labeled undifferentiated was tracked semi-automatically using EdgeTrak (Li, Kambhamettu, & Stone, 2005), which is a computer program designed to find the edges of ultrasound tongue curves. The track of each tongue shape was extracted in the form of x and y coordinates. MATLAB was then used to calculate perpendicular lines on the tongue curve, as shown in the bottom image in Figure 1. In that image, point A represents the tongue root, B represents the tongue tip, C represents the highest point of the curve of the tongue body, and D represents the base of a perpendicular segment connecting line AB to point C. *Curvature degree* is defined as the ratio of the segment CD divided by the segment AB. A higher curvature degree corresponds to a tongue shape that is more peaked. *Curvature location* is defined as the ratio of the segment AD divided by the segment DB. Higher values for curvature location correspond to more anterior constrictions.

Statistical analyses—Regression models were created to explore the interrelations among perceptual ratings, tongue shapes, and acoustic measurements. The structure of these analyses was constrained by properties of the data. First, it was necessary to create separate models for data from the TS children, who were seen for only one session, and for data from the IS children, whose data spanned multiple sessions. The TS and IS samples also differ in that the TS children produced no retroflex tongue shapes; given the small sample size, this is not an anomalous finding. A complete tally of all of the participants' productions, broken down by tongue shape and perceptual rating, is presented in Figure 2.

Logistic regression was used to examine the relationship between tongue shape and experienced listeners' ratings of /r/ sounds produced by IS children. Because both bunched and retroflex tongue shapes can have a fully accurate perceptual quality, the tongue shapes were collapsed into a single *differentiated* category. Tongue-shape classification as differentiated or undifferentiated served as the outcome variable in the logistic regression. The predictor variable of primary interest was the perceptual rating category (ordinal, 1, 2, 3). Another predictor, /r/ variant, was included to assess whether the prevalence of differentiated tongue shapes differed across consonantal and vocalic /r/ targets. Fixed effects of child and session were also included to control for variation across individual participants and time points.

The significance of each factor was determined using the likelihood ratio test on the residual deviance statistic. Terms were added sequentially to the model in the following order—child, session, /r/ variant, perceptual rating—progressing from the factor of least interest to the factor of greatest interest. First-order interaction effects were also included. A similar logistic regression was used to analyze tongue-shape data from the TS children, but in this case, perceptual rating was not included as a predictor because 68 out of 69 TS tokens were rated 1 (accurate). Thus, the only relevant question was the relationship between /r/ variant and tongue shape. A fixed effect of child was included to control for between-subjects variation.

Multiple linear regression was used to examine the relationship between tongue-shape coding and acoustic measurements, with separate models constructed for IS and TS data. F3 – F2 distance, a continuous parameter, served as the dependent variable. As above, significance was determined by analysis of deviance, with terms added sequentially to the model. The predictors used for IS children were child, session, /r/ variant, perceptual rating, and tongue shape. The predictors used for TS children were child, /r/ variant, and tongue shape. First-order interactions were included.

Results

Relationships between perceptual, ultrasound, and acoustic measures

Based on previous research (e.g., Gick et al., 2008), we hypothesized that tongue shapes that were coded as undifferentiated would be rated perceptually less accurate and would be associated with significantly higher (less accurate) F3 – F2 distances than /r/ sounds that were coded with differentiated (retroflex or bunched) tongue shapes. Results of the logistic regression analyzing tongue-shape data from IS children were consistent with this prediction: There was a significant association between binarized tongue shape and perceptual rating category, where tokens with a lower rating (indicating *greater* perceptual accuracy) were significantly more likely to be coded with a differentiated tongue shape, $\chi^2(2) = 7.9, p = .02$. Specifically, a differentiated tongue shape was identified in 100/138 tokens rated 1 (72%), 24/67 tokens rated 2 (36%), and 6/78 tokens rated 3 (8%). Similarly, IS children's undifferentiated tongue shapes were associated with significantly larger F3 – F2 distances than tongue shapes that were coded as differentiated, $F(1, 271) = 7.6, p < .01$. The TS children also produced many tongue shapes that were coded as undifferentiated (44/68 tokens rated 1, or 65%). However, tongue shape was not a significant predictor of F3 – F2 variance in the data from the TS children.

Overall, the data from the IS children supported the validity of our tongue-shape codings because tongue shapes coded as undifferentiated were acoustically and perceptually less accurate than tongue shapes coded as differentiated. However, many acoustically and perceptually accurate tokens produced by the TS children were also coded as undifferentiated. This suggests that the undifferentiated category used in our first-pass codings was overly broad. In the next section, we use quantitative tongue-shape measurements to identify two subtypes within the undifferentiated category.

Differences between vocalic and consonantal /r/

Different patterns were observed in vocalic versus consonantal variants of /r/ in IS and TS speakers. For IS children, vocalic targets were more likely than consonantal targets to be coded as differentiated, $\chi^2(1) = 11.3, p < .001$, but /r/ variant was not a significant predictor of F3 – F2 distance. For TS children, the reverse was true: There was no effect of consonantal versus vocalic target on tongue-shape coding, but vocalic /r/ was associated with significantly smaller F3 – F2 distances than consonantal /r/, $F(1, 65) = 44.4, p < .001$. There were no significant interactions between tongue shape and /r/ variant. Figure 3 plots the mean F3 – F2 distance for each /r/ variant and tongue-shape category in both the IS and

TS participants; error bars represent 95% confidence intervals around the mean F3 – F2 distance.

Subcategories within the undifferentiated tongue-shape category

We hypothesized that quantitative measurements of curvature degree and curvature location (Ménard et al., 2012) could be used to divide undifferentiated tongue shapes into subcategories that align appropriately with acoustic and perceptual measures. In Figure 4, tokens have been plotted according to their curvature location (*x*-axis) and curvature degree (*y*-axis), with different point shapes representing different perceptual ratings. Recall that a higher value for curvature location indicates a more anterior point of constriction, and a higher value for curvature degree indicates a more peaked tongue shape. Only undifferentiated tokens are represented in Figure 4 because measurements of tongue curvature were not collected for tokens that were coded as retroflex or bunched. Also, to increase the salience of distinct subcategories within the set of tokens featuring undifferentiated tongue shapes, tokens that received the intermediate perceptual rating of 2 were omitted.

Data from both IS and TS children are included in Figure 4. This scatterplot is partially successful in identifying subsets within the set of tongue shapes that were classified as undifferentiated. In particular, the undifferentiated tongue shapes that received a rating of 3 (least accurate) form a well-defined cluster characterized by a low (posterior) curvature location value and a high degree of curvature. Undifferentiated tongue shapes that received a rating of 1 are more diffusely distributed, but in general, they have a lower degree of curvature and a higher (more anterior) curvature location than tokens rated 3.

Having shown that perceptually accurate and inaccurate undifferentiated tongue shapes do cluster separately along two dimensions of tongue curvature, we used linear discriminant analysis to partition the data into two classes. Linear discriminants are functions that maximize the distance between the means of two categories while minimizing the variance around the mean within each group. In this case, linear discriminants were calculated for the extreme perceptual rating categories (i.e., 1 and 3) and were then used to predict category membership for the full set of data, including tokens rated as 2. The newly formed categories were given the descriptive labels *posterior peak* (for the less accurate cluster with a high curvature degree and low curvature location) and *anterior constriction* (for the more accurate cluster with a lower curvature degree and higher curvature location). A graphic representation of the breakdown of points into the categories determined by linear discriminant analysis is provided in Figure 5.

Using the newly defined categories of anterior constriction and posterior peak, we can revisit the relationships among tongue-shape categories, perceptual ratings, and acoustic measurements. Figure 6 shows each perceptual rating category broken down by tongue shape, with data combined across TS and IS participants. Assuming that the anterior constriction category can be treated as an instance of an appropriate tongue shape for /r/, >90% of tokens that received a perceptual rating of 1 are now correctly classified as having an appropriate shape.

Subdividing the class of undifferentiated tongue shapes also improves the goodness of category fit relative to acoustic measurements. Figure 7 shows the mean F3 – F2 distance for each of the four tongue-shape categories, combined across TS and IS participants. The posterior peak category has by far the highest mean F3 – F2 distance, whereas the F3 – F2 distance for the anterior constriction category falls closer to that of the two differentiated tongue shapes. There is no overlap in the 95% confidence intervals around the mean for the anterior constriction and posterior peak categories, indicating that linear discriminant analysis successfully partitioned the undifferentiated class into one more accurate and one less accurate subcategory. Given this success, the categories of anterior constriction and posterior peak will be used to describe IS children's trajectories of /r/ acquisition in the next section.

Longitudinal changes in /r/ production over the course of treatment

This section documents intercorrelated longitudinal changes in tongue shapes, F3 – F2 distance, and perceptually rated accuracy of /r/ productions from the two IS children who were followed over the course of treatment. Both participants made significant progress during their period of intervention. Differences in the patterning of consonantal and vocalic /r/ variants over time are also reported.

We begin with IS1, a 6-year-old boy. Due to an error in microphone configuration, no audio record of IS1's first session was captured, and so all data from this session had to be discarded. Therefore, IS1's sessions are numbered 2–7, and the first usable data (Session 2) were collected roughly 1 month after the onset of treatment targeting /r/. IS1 made rapid gains in intervention. In Session 2, all of his productions received a perceptual rating of 3 (least accurate), but by his next probe 1 month later, 50% of his consonantal /r/ tokens and 61% of his vocalic /r/ tokens were judged fully accurate. In the final three data collection sessions, perceptual ratings of IS1's /r/ productions approached ceiling-level accuracy (92%–96% 1 ratings with all tokens combined). Overall, IS1's vocalic /r/ received higher perceptual ratings (1 rating assigned to 71% of all tokens) than his consonantal /r/ (1 rating assigned to 59% of all tokens).

Figure 8 depicts a similar progression in the acoustic domain. In Session 2, IS1's mean F3 – F2 distance, combined across vocalic and consonantal targets, was very high, at 2325 Hz ($SD = 297$ Hz). By the next session, IS1's mean F3 – F2 distance had decreased to 1227 Hz ($SD = 395$ Hz), and in Session 4, it decreased again to 962 Hz ($SD = 431$ Hz). By the final session, there was substantial overlap in the 95% confidence interval around IS1's mean F3 – F2 distance and the confidence interval around the mean for TS speakers; this was true of both his vocalic and consonantal /r/. Comparing across variants of /r/, we see that IS1's vocalic /r/ productions reliably featured a lower mean F3 – F2 distance and a narrower confidence interval than his consonantal /r/ productions. A multiple linear regression (dependent variable: F3 – F2 distance; independent variables: /r/ variant, perceptual rating, tongue shape, and session) confirmed that the acoustic difference between IS1's consonantal and vocalic /r/ was significant, $\beta = -236.1$, $t = -3.7$, $p < .001$.

Changes in IS1's tongue shape for /r/ over the course of treatment are depicted in the images in Figure 9. IS1 did not produce any tongue shapes that were classified as bunched; once

accurate tongue shapes for /r/ emerged, they all fell into the retroflex or anterior constriction coding categories. In Session 2, nearly all of IS1's productions featured the least accurate posterior peak tongue shape. In subsequent months, an interesting dissociation in tongue shape between IS1's consonantal and vocalic /r/ became apparent. In Sessions 3–4, IS1's vocalic /r/ targets were dominated by retroflex tongue shapes, and his consonantal /r/ targets were more likely to fall into the anterior constriction category. In Session 5, both consonantal and vocalic /r/ were realized primarily with retroflex tongue shapes, but in Sessions 6–7, the dissociation reemerged, with vocalic and consonantal /r/ dominated by retroflex and anterior constriction tongue shapes, respectively.

In the case of IS2, no data collection sessions were discarded, but only five sessions were completed in total. IS2 made slower progress in treatment than IS1. None of his /r/ productions received a rating of 1 (fully correct) until Session 4. By Session 5, however, 89% of IS2's /r/ productions were judged to be fully correct. Differences in the relative perceptual accuracy of vocalic and consonantal /r/ targets were also less clear-cut for IS2 than for IS1. In the first three data collection sessions, 75%–100% of IS2's consonantal /r/ tokens received the intermediate rating of 2, whereas 57%–100% of his vocalic /r/ tokens received the least accurate rating, 3. However, the fully accurate rating (1) was first observed in Session 4 for both consonantal and vocalic /r/, and perceptual ratings were more evenly distributed across /r/ variants in Sessions 4–5. In the final session, the 1 rating was assigned to 100% of IS2's vocalic and 94% of his consonantal /r/ targets.

These perceptual impressions were corroborated by the acoustic data in Figure 10. For vocalic /r/, IS2's mean F3 – F2 distance was generally high in the first two data collection sessions and declined steadily over the next three sessions. However, his F3 – F2 distance remained significantly elevated in the final recording session, with no overlap between the 95% confidence interval around IS2's mean value and that of the TS participants. Thus, even though IS2's vocalic /r/ productions in the last session were judged to be perceptually accurate, they were still acoustically distinct from the productions of TS participants in the same age range. We hypothesize that if we had used a finer grained or continuous rating scale (e.g., Munson, Edwards, Schellinger, Beckman, & Meyer, 2010), we might have observed differences in perceived accuracy between IS2's productions in Session 5 and those of TS children.

A somewhat different trajectory was seen for IS2's consonantal /r/ targets. IS2's mean F3 – F2 distance for consonantal /r/ was notably low in Session 1, and his F3 – F2 distance remained significantly lower for consonantal than vocalic /r/ in Sessions 2–3. Subsequently, IS2's F3 – F2 values declined less steeply for consonantal than vocalic /r/, and his 95% confidence intervals around the mean for the two /r/ variants overlapped in Sessions 4 and 5. A multiple linear regression (dependent variable: F3 – F2 distance; independent variables: /r/ variant, perceptual rating, tongue shape, and session) indicated that IS2's F3 – F2 distance was significantly higher for vocalic than consonantal targets, $\beta = 194.6$, $t = 4.3$, $p < .0001$.

Figure 11 shows a breakdown of IS2's tongue shapes for /r/ over time. Unlike IS1, IS2 favored a bunched tongue shape in his accurate /r/ productions. For the first three data

collection sessions, IS2's tongue shapes for vocalic /r/ were dominated by the least accurate posterior peak category. His tongue shapes for consonantal /r/ in these sessions fell predominantly in the anterior constriction category, with roughly equal numbers representing these two categories in the third session. In Session 4, IS2's vocalic /r/ tokens fell primarily in the anterior constriction category, whereas most of his consonantal /r/ tokens were judged to feature a bunched shape. Finally, Session 5 showed convergence between the two /r/ variants, with both vocalic and consonantal /r/ featuring a majority of bunched tongue shapes and a minority of anterior constriction shapes.

Discussion

Several major findings emerged from this research. First, we found significant associations between qualitative tongue-shape codings and perceptual and acoustic measures of accuracy in /r/ sounds produced by children receiving treatment for /r/ misarticulation. We were able to observe parallel changes across all three of these measures over the course of a successful period of intervention. Second, we found that consonantal and vocalic variants of /r/ were associated with different perceptual, acoustic, and articulatory properties, and the two variants followed different trajectories of change over the course of treatment. Finally, we found that the undifferentiated tongue shape that we initially considered indicative of an incorrect production was overly broad because it encompassed some correct as well as incorrect productions. Using quantitative measures of tongue-shape curvature, we found that the least accurate /r/ productions were associated with a highly curved, posteriorly located tongue shape. In this section, we discuss the clinical implications of these findings. In particular, we propose that the properties of the posterior peak category can inform clinicians' decisions regarding cues provided in traditional intervention for /r/ misarticulation. We also suggest that our findings regarding the properties of consonantal and vocalic /r/ have implications for treatment planning.

Tongue shape requirements for perceptual accuracy

In Figure 4 and the associated discussion, we established that the subset of tokens that was perceived to be least accurate was characterized by a high curvature degree and a low (posterior) curvature location, giving rise to the label posterior peak. This is consistent with qualitative characterizations of children's incorrect /r/ tongue shapes found in previous literature. Boyce, Combs, and Rivera-Campos (2011) reviewed ultrasound images from 37 children with persistent /r/ misarticulation. They found that 19/37 speakers exhibited a tongue shape that they termed "humped." This tongue shape was described as having a single, highly peaked constriction located "further back than that for the typical 'bunched' shape for normal productions of American English /r/." Figure 12 provides an example of the humped tongue shape category from Boyce et al. This can be compared to the bottom image in Figure 1, which is an example of our posterior peak category. Based on the similarities in both of these images and the qualitative description of the categories, we conclude that our posterior peak label can be viewed as equivalent to the humped tongue category described by Boyce et al.

Boyce et al. (2011) argued that the humped tongue shape was characterized by the absence of a tongue root constriction. Acoustic modeling suggests that some degree of constriction in the pharynx must be present to produce perceptually accurate /r/, at least for the range of anterior constriction shapes attested in studies to date (Zhou et al., 2008). Because ultrasound imaging does not provide a view of the posterior pharyngeal wall, and because the angle of the ultrasound probe could vary from trial to trial, the degree of pharyngeal constriction could not be directly ascertained as part of the present study. However, it is generally possible to draw inferences about the presence or absence of a pharyngeal constriction based on the shape of the tongue in the oral cavity. When some of the volume of the tongue is pushed back into the pharyngeal space to form a tongue root constriction, the tongue body cannot maintain a highly peaked shape; it necessarily flattens out to some extent. This flattening of the tongue body in the palato-velar region takes place even if there is an anterior constriction, either retroflex or bunched. By contrast, when all of the volume of the tongue forms a single constriction in the oral cavity, the resulting shape of the tongue is highly curved. Therefore, like Boyce et al., we maintain that tongue shapes with a posterior peak or hump are characterized by the absence of a tongue root constriction. This can be contrasted with the anterior constriction tongue shape, which features a lower curvature degree than the posterior peak category. We hypothesize that the less curved shape of these productions indicates that part of the volume of the tongue is retracted into the pharynx, creating some degree of pharyngeal constriction. Combined with some amount of constriction at a more anterior point in the vocal tract, this tongue shape yielded an output that tended to be perceptually and acoustically more accurate than that of the posterior peak shape.

This finding has implications for clinical practice. Many clinicians deemphasize the pharyngeal constriction during /r/ treatment, focusing instead on the lips or the anterior lingual constriction (Boyce et al., 2011). However, these components may not represent the primary source of difficulty for children with /r/ misarticulation. In the present study, the productions that were both acoustically and perceptually least accurate were produced with a highly curved tongue shape that lacks a pharyngeal constriction. This suggests that clinicians should make it a priority to encourage the formation of a constriction with the tongue root during articulatory treatment for /r/. The use of ultrasound imaging to provide biofeedback (e.g., Adler-Bock et al., 2007; Bacsfalvi, 2010; Bernhardt et al., 2003, 2008) represents one way to address the tongue root constriction in treatment for /r/ misarticulation. However, it is also possible to target pharyngeal constriction using conventional articulatory methods. For example, in our study, IS1 and IS2 received verbal and occasional tactile cues encouraging them to adopt a retracted tongue position. It is important to note that the children were not directed toward a particular tongue shape, such as bunching or retroflexion, in order to achieve retraction.

Consonantal and vocalic /r/ targets

In their longitudinal study of the acoustics of /r/ sounds produced by children ages 14–26 months, McGowan et al. (2004) found that adultlike F3 – F2 distances were slower to emerge for consonantal than vocalic /r/ targets. This finding can be compared against the acoustic data collected in our study. It is not possible to describe the longitudinal emergence

of correct /r/ in the group of TS children, who were seen for only one data collection session. However, the mean F3 – F2 distance in their vocalic /r/ productions was 512.9 Hz ($SD = 159.9$ Hz), whereas the mean F3 – F2 distance in their consonantal /r/ productions was 916.0 Hz ($SD = 440.5$ Hz). This difference was significant in the linear regression reported previously. Thus, although we cannot say whether perceptually accurate /r/ emerged earlier for consonantal or vocalic /r/ in the TS speakers, the finding that vocalic /r/ was acoustically more adult-like than consonantal /r/ is generally consistent with the findings reported by McGowan et al.

Also consistent with McGowan et al.'s (2004) acoustic findings was the longitudinal trajectory that IS1 exhibited over the course of intervention. From Session 3 on, IS1 reliably produced a higher percentage of perceptually accurate /r/ sounds for vocalic than consonantal targets, and his productions of vocalic /r/ were associated with significantly lower, more adult-like F3 – F2 values. A somewhat different pattern was exhibited by IS2. For the first three sessions, IS2's productions of consonantal /r/ tended to receive higher perceptual ratings than his productions of vocalic /r/. On its own, this finding is inconclusive because there is evidence that listeners may apply a less stringent standard in their perceptual rating of consonantal /r/ relative to vocalic /r/ sounds (Klein et al., 2012; McAllister Byun & Hitchcock, 2012). Furthermore, the 1 rating (indicating a fully accurate /r/) was first observed in the same session for both vocalic and consonantal /r/, and vocalic /r/ received a slightly higher percentage of 1 ratings in the final two sessions. However, IS2 also demonstrated more adult-like acoustics in his productions of consonantal /r/, and this difference emerged as significant in the multiple linear regression reported earlier. Thus, although there is general agreement between the present results and McGowan et al.'s finding that child speakers tend to master the acoustics of /r/ earlier in the context of a vocalic target, IS2 serves as a reminder that there can be individual variation within this larger trend.

We can also evaluate a second claim made by McGowan et al. (2004) regarding the emergence of consonantal versus vocalic /r/. Although they did not collect tongue-shape data as part of their study, these authors hypothesized that the delayed emergence of acoustically accurate consonantal /r/ could be attributed to what they characterized as the greater articulatory complexity of consonantal /r/. More specifically, McGowan et al. proposed that postvocalic and syllabic variants of /r/ are “produced with a single tongue constriction” (p. 881), whereas onset /r/ requires two tight lingual constrictions. This claim contrasts with previous studies reporting that postvocalic and syllabic /r/ are produced with two lingual constrictions (Alwan et al., 1997), and that the pharyngeal gesture may be even larger in syllable-final /r/ than in onset /r/ (Gick & Goldstein, 2002).

We can compare these competing hypotheses by examining perceptual ratings of the posterior peak category, which was argued earlier to be the only one of the four tongue-shape categories coded here that is produced with just one major constriction. If Gick and Goldstein (2002) are correct, we predict that posterior peak tongue shapes should not receive the perceptually accurate 1 rating, whether the target is consonantal or vocalic. If McGowan and colleagues' (2004) account holds true, the posterior peak category should not occur in

the set of 1-rated consonantal /r/ tokens, but it might be observed in the set of 1-rated vocalic /r/ tokens.

In our data, 19 consonantal /r/ tokens were coded in the posterior peak category; of these, 6 (31%) were rated 1, 6 (31%) were rated 2, and 7 (36%) were rated 3. By contrast, out of 78 vocalic /r/ tokens in the posterior peak category, 11 (14%) were rated 1, 9 (27%) were rated 2, and 58 (84%) were rated 3. Fisher's exact test (nonparametric) indicated that the difference between these two distributions was significant ($p = .02$). Thus, our findings are not consistent with the hypothesis that consonantal and vocalic /r/ differ in articulatory complexity, or that perceptually accurate vocalic /r/ can be produced with only a single major lingual constriction. This finding has implications for clinical practice. As noted earlier, many clinicians deemphasize the tongue root constriction when providing articulator placement cues for /r/. McGowan and colleagues' (2004) hypothesis that perceptually accurate vocalic /r/ can be produced with only a single lingual constriction might lead some clinicians to assume that the pharyngeal constriction can be ignored altogether when the target is vocalic /r/. This contrasts with our finding that >90% of perceptually accurate /r/ sounds were classified into one of the three tongue-shape categories that feature at least some degree of pharyngeal constriction. Furthermore, the two participants who received intervention for /r/ misarticulation in our study were explicitly cued to adopt a retracted tongue position when practicing vocalic /r/. As reported earlier, they demonstrated improvement across perceptual, acoustic, and tongue-shape measures of accuracy over the course of This intervention.

Conclusion

This paper has made a novel contribution to our understanding of the nature of /r/ production by using perceptual, tongue-shape, and acoustic measurements to track longitudinal changes in /r/ production accuracy over the course of intervention for /r/ misarticulation. We found that qualitative codings of tongue shape had a statistically significant association with both perceptual and acoustic measures, supporting the validity of ultrasound imaging as a means of characterizing children's /r/ productions. Quantitative measurements of tongue-shape curvature can improve the accuracy of the match between tongue-shape categories and perceptual or acoustic measurements of /r/ accuracy. Our ultrasound measurements suggest that a constriction in the pharyngeal region is essential for perceptually correct production of either vocalic or consonantal variants of /r/. The two children whose progress in intervention was documented here received cues encouraging a retracted tongue posture, and they demonstrated gains across perceptual, acoustic, and ultrasound dimensions. However, the children did not receive any specific cues for the shape of the anterior tongue (e.g., bunched vs. retroflex). Together with previous evidence that various tongue postures in the same child or across children result in a perceptually accurate /r/, the present results suggest that children should be allowed to discover their own most natural posture for matching the target /r/. Our results also suggest that ultrasound studies of articulation can be of value even to clinicians who do not have access to this technology in their own practice because viewing ultrasound images and videos can increase clinicians' understanding of articulation in a way that may allow them to give more accurate cues during articulatory intervention.

Acknowledgments

The authors gratefully acknowledge the contributions of the following individuals: for ultrasound data coding, Ryan Bennett; for acoustic measurements, Elizabeth Varall and Corey Silverstein; for statistical analysis, Peter Halpin; and for input on ultrasound measurement and interpretation, Suzanne Boyce. We also thank our participants and their families for their cooperation throughout the study.

Aspects of this research were presented at Ultrafest IV (2007), the International Child Phonology Conference (2009), and the annual convention of the American Speech-Language-Hearing Association (2009).

References

- Adler-Bock M, Bernhardt BM, Gick B, Bacsfalvi P. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*. 2007; 16:128–139. [PubMed: 17456891]
- Alwan A, Narayanan S, Haker K. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America*. 1997; 101(2):1078–1089. [PubMed: 9035399]
- Bacsfalvi P. Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech Language Pathology and Audiology*. 2010; 34:206–217.
- Ball MJ, Lowry O, McInnis L. Distributional and stylistic variation in /r/-misarticulations: A case study. *Clinical Linguistics & Phonetics*. 2006; 20(2–3):119–124. [PubMed: 16428227]
- Bernhardt B, Bacsfalvi P, Adler-Bock M, Shimizu R, Cheney A, Giesbrecht N, Radanov B. Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics*. 2008; 22(2):149–162. [PubMed: 18253873]
- Bernhardt B, Gick B, Bacsfalvi P, Ashdown J. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*. 2003; 17(2):199–216. [PubMed: 12858839]
- Bernthal, JE.; Bankson, NW.; Flipsen, P, Jr. *Articulation and phonological disorders: Speech sound disorders in children*. 6th ed.. Boston, MA: Pearson; 2009.
- Boersma, P.; Weenink, D. Praat: Doing phonetics by computer (Version 5.1.2.9) [Computer software]. 2010. Retrieved from <http://www.praat.org/>
- Boyce S, Combs S, Rivera-Campos A. Acoustic and articulatory characteristics of clinically resistant /r/. *Journal of the Acoustical Society of America*. 2011; 129(4):2625.
- Boyce S, Espy-Wilson C. Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America*. 1997; 101:3741–3753. [PubMed: 9193061]
- Dalston RM. Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults. *Journal of the Acoustical Society of America*. 1975; 57(2):462–461. [PubMed: 1117100]
- Delattre P, Freeman DC. A dialect study of American r's by X-ray motion picture. *Linguistics*. 1968; 6(44):29–68.
- Flipsen P, Shriberg LD, Weismer G, Karlsson HB, McSweeney JL. Acoustic phenotypes for speech-genetics studies: Reference data for residual /3/ distortions. *Clinical Linguistics & Phonetics*. 2001; 15(8):603–630.
- Gick B, Bacsfalvi P, Bernhardt BM, Oh S, Stolar S, Wilson I. A motor differentiation model for liquid substitutions: English /r/ variants in normal and disordered acquisition. *Proceedings of Meetings on Acoustics*. 2008; 1:1–9.
- Gick B, Goldstein L. Relative timing in the three gestures of North American English /r/. *Journal of the Acoustical Society of America*. 2002; 111:2481.
- Guenther FH, Espy-Wilson CY, Boyce S, Matthies ML, Zandipour M, Perkell JS. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*. 1999; 105(5):2854–2865. [PubMed: 10335635]
- Hagiwara R. Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics*. 1995; 90:1–187.

- Hodson, B.; Paden, E. Targeting intelligible speech: A phonological approach to remediation. Austin, TX: Pro-Ed; 1991.
- Hoffman PR, Stager S, Daniloff RG. Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders*. 1983; 48(2):210–215. [PubMed: 6621013]
- Klein HB, Grigos MI, McAllister Byun T, Davidson L. The relationship between inexperienced listeners' perceptions and acoustic correlates of children's /r/ productions. *Clinical Linguistics & Phonetics*. 2012; 26(1):628–645. [PubMed: 22690718]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]
- Lee S, Potamianos A, Narayanan S. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*. 1999; 105:1455–1468. [PubMed: 10089598]
- Li M, Kambhamettu C, Stone M. Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*. 2005; 19(6–7):545–554. [PubMed: 16206482]
- McAllister Byun T, Hitchcock ER. Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*. 2012; 21:207–221. [PubMed: 22442281]
- McGowan RS, Nitttrouer S, Manning CJ. Development of [r] in young, midwestern, American children. *Journal of the Acoustical Society of America*. 2004; 115(2):871–884. [PubMed: 15000198]
- Ménard L, Aubin J, Thibeault M, Richard G. Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica*. 2012; 64(2):64–72. [PubMed: 22212175]
- Modha G, Bernhardt BM, Church R, Bacsfalvi P. Case study using ultrasound to treat /r/. *International Journal of Language & Communication Disorders*. 2008; 43(3):323–329. [PubMed: 17852527]
- Munson B, Edwards J, Schellinger SK, Beckman ME, Meyer MK. Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*. 2010; 24(4–5):245–260. [PubMed: 20345255]
- Preston JL, Edwards ML. Speed and accuracy of rapid speech output by adolescents with residual speech sound errors including rhotics. *Clinical Linguistics & Phonetics*. 2009; 23(4):301–318. [PubMed: 19382016]
- Shriberg LD, Flipsen PJ, Karlsson HB, McSweeney JL. Acoustic phenotypes for speech-genetics studies: An acoustic marker for residual /ɜ:/ distortions. *Clinical Linguistics & Phonetics*. 2001; 15(8):631–650.
- Shriberg, LD.; Kent, R. *Clinical phonetics*. 3rd ed.. Boston, MA: Allyn & Bacon; 2003.
- Shriberg LD, Kwiatkowski J, Hoffmann K. A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*. 1984; 27(3):456–465. [PubMed: 6482415]
- Smit AB, Hand L, Freilinger JJ, Bernthal JE, Bird A. The Iowa Articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*. 1990; 55(4):779–798. [PubMed: 2232757]
- Tiede M, Boyce S, Holland C, Choe K. A new taxonomy of American English /r/ using MRI and ultrasound. *Journal of the Acoustical Society of America*. 2004; 115:2633.
- Westbury J, Hashi M, Lindstrom M. Differences among speakers in lingual articulation for American English. *Speech Communication*. 1998; 26:203–226.
- Zhou X, Espy-Wilson CY, Boyce S, Tiede M, Holland C, Choe A. A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *Journal of the Acoustical Society of America*. 2008; 123(6):4466–4481. [PubMed: 18537397]

Figure 1.

Ultrasound images illustrating children's tongue shapes for /r/. The top image is an example of a retroflex /r/ that was produced by IS1. The arrow pointing to the reflection above the main tongue body indicates where the tip of the tongue is curled back and touching the alveolar ridge. The middle image is an example of a bunched /r/ that was produced by TS1. The bottom image is an example of an undifferentiated /r/ that was produced by IS2. This image is annotated to illustrate the metric used to measure curvature degree and location, based on Ménard et al. (2012).

Figure 2.

Frequency of occurrence of tongue shapes across rating categories for all participants. Note that the scale of the y-axis differs for children with impaired speech (IS1, IS2) versus typically developing children (TS1, TS2, TS3).

Figure 3.

Mean F3 - F2 distance for each /r/ variant and tongue-shape category in both the IS and TS participants. Error bars represent 95% confidence intervals around the mean.

Figure 4.
Scatterplot of 1 - and 3-rated undifferentiated tongue shapes by curvature location and degree for all participants.

Figure 5.
Scatterplot of all undifferentiated tokens, categorized by curvature location and degree through linear discriminant analysis.

Figure 6.
Frequency of occurrence of new tongue shapes across rating categories for all participants.

Figure 7.
Mean F3 - F2 distance associated with new tongue-shape categories for all participants.
Error bars represent 95% confidence intervals around the mean.

Figure 8.

Changes in the acoustic properties of IS1's productions of vocalic and consonantal /r/ over the course of treatment. Data from TS children are provided for comparison.

Figure 9.

Changes in the occurrence of tongue-shape categories in IS1's /r/ productions over the course of treatment. The left image shows vocalic targets only; the right image shows consonantal targets only. Data from TS children are provided for comparison.

Figure 10.

Changes in the acoustic properties of IS2's productions of vocalic and consonantal /r/ over the course of treatment. Data from TS children are provided for comparison.

Figure 11.

Changes in the occurrence of tongue-shape categories in IS2's /r/ productions over the course of treatment. The left image shows vocalic targets only; the right image shows consonantal targets only. Data from TS children are provided for comparison.

Figure 12.

Example of a “humped” tongue shape. Reprinted from “Acoustic and Articulatory Characteristics of Clinically Resistant /r/,” S. Boyce, S. Combs, and A. Rivera-Campos, 2011, *Journal of the Acoustical Society of America*, 129(4), p. 2625. Copyright 2011 by S. Boyce. Reprinted with permission.