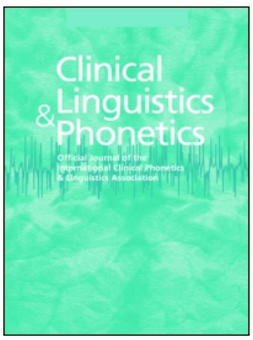

Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: http://www.tandfonline.com/loi/iclp20

Tongue shapes for rhotics in school-age children with and without residual speech errors

Jonathan L. Preston, Patricia McCabe, Mark Tiede & Douglas H. Whalen

To cite this article: Jonathan L. Preston, Patricia McCabe, Mark Tiede & Douglas H. Whalen (2018): Tongue shapes for rhotics in school-age children with and without residual speech errors, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2018.1517190

To link to this article: https://doi.org/10.1080/02699206.2018.1517190

Published online: 10 Sep 2018.

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2018.1517190

Taylor & Francis
Taylor & Francis Group

Tongue shapes for rhotics in school-age children with and without residual speech errors

Jonathan L. Preston[a,b], Patricia McCabe [c], Mark Tiede[b], and Douglas H. Whalen[d,e]

[a]Communication Sciences & Disorders, Syracuse University, Syracuse, NY, USA; [b]Haskins Laboratories, New Haven, CT, USA; [c]Communication Sciences & Disorders, The University of Sydney, Sydney, Australia; [d]Haskins Laboratories, New Haven, CT; [e]Speech-Language-Hearing Sciences, CUNY Graduate Center, New York, NY, USA

ABSTRACT

Speakers of North American English use variable tongue shapes for rhotic sounds. However, quantifying tongue shapes for rhotics can be challenging, and little is known about how tongue shape complexity corresponds to perceptual ratings of rhotic accuracy in children with residual speech sound errors (RSE). In this study, 16 children aged 9–16 with RSE and 14 children with typical speech (TS) development made multiple productions of 'Let Robby cross Church Street'. Midsagittal ultrasound images were collected once for children with TS and twice for children in the RSE group (once after 7 h of speech therapy, then again after another 7 h of therapy). Tongue contours for the rhotics in the four words were traced and quantified using a new metric of tongue shape complexity: the number of inflections. Rhotics were also scored for accuracy by four listeners. During the first assessment, children with RSE had fewer tongue inflections than children with TS. Following 7 h of therapy, there were increases in the number of inflections for the RSE group, with the cluster items *cross* and *Street* reaching tongue complexity levels of those with TS. Ratings of rhotic accuracy were correlated with the number of inflections. Therefore, the number of inflections in the tongue, an index of tongue shape complexity, was associated with perceived accuracy of rhotic productions.

ARTICLE HISTORY
Received 10 April 2018
Revised 24 August 2018
Accepted 25 August 2018

KEYWORDS
Rhotic; children; tongue; speech sound disorder; ultrasound

Rhotic phonemes, which include /ɹ, ɝ, ɚ/ in North American English, are among the last sounds to develop in children (Sander, 1972; Smit, Hand, Freilinger, Bernthal, & Bird, 1990). Rhotics are also among the most frequently misarticulated sounds in individuals with residual speech sound errors (RSE, Shriberg, 2009). RSE includes speech errors that persist past approximately the age of 8–9 years. One potential reason for the difficulty in acquiring this class of sounds is the complex articulatory configuration required (Boyce, 2015). Moreover, the specific complex tongue configurations used to achieve acoustically acceptable rhotic quality can be variable both within and across speakers (Mielke, Baker, & Archangeli, 2016; Westbury, Hashi, & Lindstrom, 1998; Zhou et al., 2008), making it difficult to characterize common features of lingual shapes in correct and erred productions.

Typical articulation of rhotics in adults requires the formation of three constrictions in the vocal tract, one with the lips and two with the tongue (Delattre & Freeman, 1968). A

CONTACT Jonathan L. Preston jopresto@syr.edu Department of Communication Sciences & Disorders, Syracuse University, 621 Skytop Rd Suite 1200, Syracuse, NY 13244

pharyngeal constriction is achieved via the tongue root retracting toward the posterior pharyngeal wall. An oral constriction is achieved with the tongue tip, blade, or anterior dorsum approximating the palatal or alveolar region. The oral constriction in particular varies significantly between speakers (and, sometimes, between word positions within a single speaker); however, the oral constriction can be broadly described as being achieved with a 'retroflex' tongue shape (tongue tip raising and angled up toward the palate) or with a 'bunched' tongue shape (Delattre & Freeman, 1968). The highly variable oral constriction contributes to the challenge of describing or quantifying canonical tongue shapes associated with correctly articulated rhotics. Evidence of these major constrictions (and their acoustic consequences) comes from magnetic resonance imaging of adult speakers of American English (Alwan, Narayanan, & Haker, 1997; Boyce, 2015; Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000; Tiede, Boyce, Holland, & Choe, 2004; Zhou et al., 2008).

Ultrasound imaging of the tongue has been used as a safe and effective method for describing tongue movement and shape (Bressmann, Harper, Zhylich, & Kulkarni, 2016; Cleland, Mccron, & Scobbie, 2013; Davidson, 2006; Gick, Campbell, Oh, & Tamburri-Watt, 2006; Stone, 2005; Tiede et al., 2004; Whalen et al., 2005). With respect to production of rhotics, ultrasound imaging has revealed that the major lingual constrictions are evident in most productions of American English rhotics, although the specific timing and magnitude of the lingual movements may vary somewhat as a function of word position (Campbell, Gick, Wilson, & Vatikiotis-Bateson, 2010).

Ultrasound images of /ɹ/ productions may be useful in understanding misarticulations and, for children with RSE who produced distorted /ɹ/, ultrasound has been used as a biofeedback tool in speech therapy to teach lingual articulation (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Modha, Bernhardt, Church, & Bacsfalvi, 2008; Preston et al., 2017, 2014). As ultrasound imaging of the tongue becomes more common, it may be clinically advantageous to characterize tongue shape patterns associated with both correct and distorted rhotic productions. However, characterization of the complex and variable tongue shapes for correct and distorted rhotics from ultrasound images can be challenging. Klein, McAllister Byun, Davidson, and Grigos (2013) evaluated tongue shapes of two children ages 5–6 with /ɹ/ distortions who were participating in articulation therapy, as compared to three children ages 4–7 with typical /ɹ/ productions. Tongue shapes were coded as differentiated (bunched or retroflex) or undifferentiated, and the undifferentiated tongue shapes were quantified using two measures of tongue curvature location and curvature degree (Ménard, Aubin, Thibeault, & Richard, 2012). Undifferentiated productions that were most likely to be rated as perceptually incorrect were those in which the tongue was highly curved and had a posterior peak to the curve. Additionally, Gick et al. (2008) argued that children's errors on English liquids are commonly characterized by either (a) an omission of an articulatory gesture (i.e. loss of either the oral or pharyngeal constriction), or (b) 'stiffening' of the tongue, which can be thought of as the merging or averaging of two gestures (i.e. forming one single constriction between the two target locations of the anterior palate and oropharynx). In both cases, the result is a simplification of tongue shape and a lack of differentiation of the anterior and posterior tongue. Quantifying the difference between these simple and more complex tongue shapes is

an important goal for studying the relationship between tongue shapes and listeners' perceptions of speech sound accuracy.

Tongue shape measures that can be applied to a single tongue curve (i.e. which do not require reference to tongue location or coordinates of other curves) may be the most clinically applicable (Bressmann et al., 2016; Zharkova, Gibbon, & Hardcastle, 2015). For research purposes, a tongue image is often obtained relative to the palate or relative to other tongue images via head stabilization, but this technique is often clinically impractical. For some clinical populations, such as elderly speakers and those with neurological conditions such as cerebral palsy or Parkinson's Disease, it may be difficult to accommodate ultrasound imaging with head restraint. Therefore, for measures of tongue shape to be clinically useful, it may be advantageous to develop measures which do not require comparison between different shapes. The present investigation, therefore, addresses the utility of a reference-free measure of tongue shape complexity which can be computed on a single curve without reference to other images or structures (cf. Zharkova, Gibbon, & Lee, 2017). This study explores the validity of a new measure, the Number of INFLlections (NINFL), to characterize rhotics produced by children with and without RSE. Derived from a tongue contour tracing, NINFL reflects a count of tongue curvature changes whose values exceed a threshold (see the 'Methods' section for further detail), and therefore is intended to represent the relative complexity of the tongue shape.

Finally, children's realization of rhotics may be contextually influenced (Hoffman, Schuckers, & Ratusnik, 1977). It has long been known that, among children with RSE, /ɹ/ in clusters may be perceived to be more accurate than in non-clusters (Curtis & Hardy, 1959). It is possible that these differences in children are driven by different articulatory patterning in clusters (cf Mielke et al., 2016) and/or that listeners are more accepting of subtle phonetic variation in clusters.

Purpose and hypotheses

To avoid overreliance on qualitative descriptions of tongue images, there remains a need for objective measures which can quantify the complex tongue shapes used for rhotics. The purpose of the study was to characterize tongue shapes that correspond to correct and distorted productions of American English rhotics in school-age children with and without RSE, as well as to observe changes in tongue shapes in children with RSE following a brief period of treatment. It was hypothesized that the tongue shape of /ɹ/ in tokens in the RSE group would be less complex than /ɹ/ in tokens produced by a group of typical speakers, and that tongue shape complexity would increase following a brief period of speech therapy. Furthermore, it was hypothesized that there would be an association between tongue shape complexity and listeners' perceptions of rhotic accuracy. Finally, it was hypothesized that /ɹ/ production, and listeners' perceptions of /ɹ/ accuracy, would be influenced by the context of consonant clusters.

Method

Participants

Participants were native speakers of a rhotic dialect of General American English (all were from Connecticut) and were between the ages of 9 and 16 years. Participants had no

reported history of hearing problems and passed pure tone hearing screening at 20 dB HL at 500, 1000, 2000 and 4000 Hz bilaterally. Participants had no known developmental disabilities as reported by their parents. Two groups of participants meeting these criteria were recruited.

The first group consisted of 14 children (9 male, 5 female) with typical speech (TS) (mean age 11 years 11 months, SD 1.7 months). Children in the TS group had no history of speech or language disorders as reported by the parents, and none had received previous speech-language therapy. Children in the TS group achieved a standard score of at least 100 on the Goldman–Fristoe Test of Articulation-2 (GFTA-2, Goldman & Fristoe, 2000), indicating speech sound production skills within normal limits. Additionally, they were judged to have typical speech sound production in conversational speech, and they produced no more than two errors (as judged by a certified speech–language pathologist) on a 100-word reading list consisting of 25 word-initial /ɹ/ singletons, 50 word-initial /ɹ/ clusters and 25 vocalic /ɝ/.

The second group consisted of 16 children with RSE (14 male, 2 female, mean age 11.7 years SD 1.7 years). These participants were recruited as part of a series of treatment studies using ultrasound as a biofeedback tool to treat errors on rhotics (Preston, Leece, & Maas, 2017; Preston et al., 2014). Participants in the RSE group scored below a standard score of 75 on the GFTA-2 and also scored below 30% accuracy on the reading list with 100 /ɹ/ words, as judged by a speech–language pathologist. The RSE and TS groups did not differ in age (Independent samples *t*-test, t [28] = 0.30; $p = 0.77$, two-tailed) or gender (Fisher's exact test, $p = 0.20$).

Participants in the RSE group were scheduled to complete the experimental task described below on two separate occasions: Once after 7 h of ultrasound visual feedback therapy on /ɹ/ (Mid-treatment), and again after completion of an additional 7 h of visual feedback therapy (Post-treatment). Changes in /ɹ/ accuracy as a result of this treatment have been reported elsewhere (Preston et al., 2017; Preston, Maas, Whittle, Leece, & McCabe, 2016; Preston et al., 2014); the primary focus here is on the nature of the tongue shape changes observed following 7 h of therapy, and whether those changes correspond to perceived accuracy. Due to scheduling conflicts or attrition, 15 children in the RSE group completed the task at Mid-treatment and 13 completed the task Post-treatment (12 matched).

Ultrasound data acquisition and analysis

Participants read the sentence *Let Robby cross Church Street*. This sentence sampled an onset singleton /ɹ/ in *Robby*, a two-element lingual cluster /ɹ/ in *cross*, a three-element lingual /ɹ/ cluster in *Street* and a vocalic /ɝ/ in *Church*. This sentence was repeated a minimum of 12 times in four blocks of three repetitions for each participant. Between each block of three repetitions of the sentence were the individual words from the sentence presented one at a time in random order (e.g. *Church, let, Robby, street, cross*). The sentence and individual words were elicited by showing the participants the words displayed in a Powerpoint slideshow. A preliminary analysis showed no effects between tokens drawn from sentences vs. individual words; consequently tokens from both sentences and words were combined in subsequent analyses. In addition, because previous work has shown that raters typically give higher rhoticity scores to /ɹ/ produced in clusters (Curtis & Hardy, 1959; also confirmed in our own results described below), our analyses

bin words by whether their /ɹ/ occurs within a cluster; (i.e. *street* grouped with *cross*, vs. *Robby* grouped with *Church*). This division reflects the hypothesis that /ɹ/ coproduced in clusters is differs from /ɹ/ produced alone.

While reading, an Aloka SSD-100 ultrasound was used to collect midsagittal images of the tongue at a rate of 30 frames per second. Participants sat on a stool and rested their chin on the ultrasound probe, which was stabilized via microphone clamp attached to a microphone stand. Additionally, a head light was used to provide the participant with a visual display of head movement, and participants were instructed to keep the head light focused on a single target at the wall as they spoke in order to avoid head rotation. Ultrasound images were visually monitored by a researcher to ensure the landmark shadows of the hyoid and the mandible were visible during productions. However, participants did not view the ultrasound images while producing the stimuli. Feedback was provided by the researcher if the light deviated from the wall target, if the images were not capturing the landmark shadows, or if the ultrasound images became distorted due to reduced contact between the transducer and the skin. Data collection with the ultrasound took approximately 15 min. In addition to the video recordings with the ultrasound, audio was collected using a Sennheiser MKE-2 microphone. The audio and video data were mixed and recorded simultaneously using a DVD recorder. Previous tests had confirmed that there were no measurable delays in the delivery of the ultrasound video and that the two signals were temporally aligned.

Based on the acoustic signal, rhotics were identified for each production using waveforms and wideband spectrograms in Praat (Boersma & Weeninck, 2014). TextGrids were generated with tiers marking target words, and within each word markers were placed to identify target time points in the signal associated with the rhotic. For the words *Robby, cross* and *Street*, the onset of voicing was used as the acoustic cue for where the rhotic was expected to occur. For the word *Church*, the center of the vocalic segment was identified from the waveform and corresponding spectrogram. The time points identified in the acoustic signal were then used to identify the corresponding video frame for analysis.

An open-source Matlab procedure (GetContours; Tiede, 2015) was used to select the corresponding video frame for each rhotic that was marked in the text grid. From these still images, a research assistant traced the contour of the tongue by interactively placing between 6 and 8 anchor points defining a spline fit to the tongue surface using 100 equally distanced points along the curve. Where the TextGrid did not accurately identify a clearly measurable tongue image, the research assistant moved forward or back by up to two frames to label the image to characterize the rhotic.

A custom Matlab procedure (freely available at https://osf.io/xzdb7/) was used to compute the number of inflections in each contour based on the signed curvature:

$$k = \frac{x'y'' - y'x''}{\left(x'^2 + y'^2\right)^{3/2}} \tag{1}$$

where primes denote derivatives with respect to offset along the curve, with any curl-over points (non-monotonic values of x with associated higher y values) deleted. The NINFL is the count of nonzero sign changes in trimmed curvature (values whose associated radius is $<thresh$ times the distance along the curve from the first to the last point, where *thresh* was a heuristically determined value of 0.3). NINFL values greater than 5 were removed from

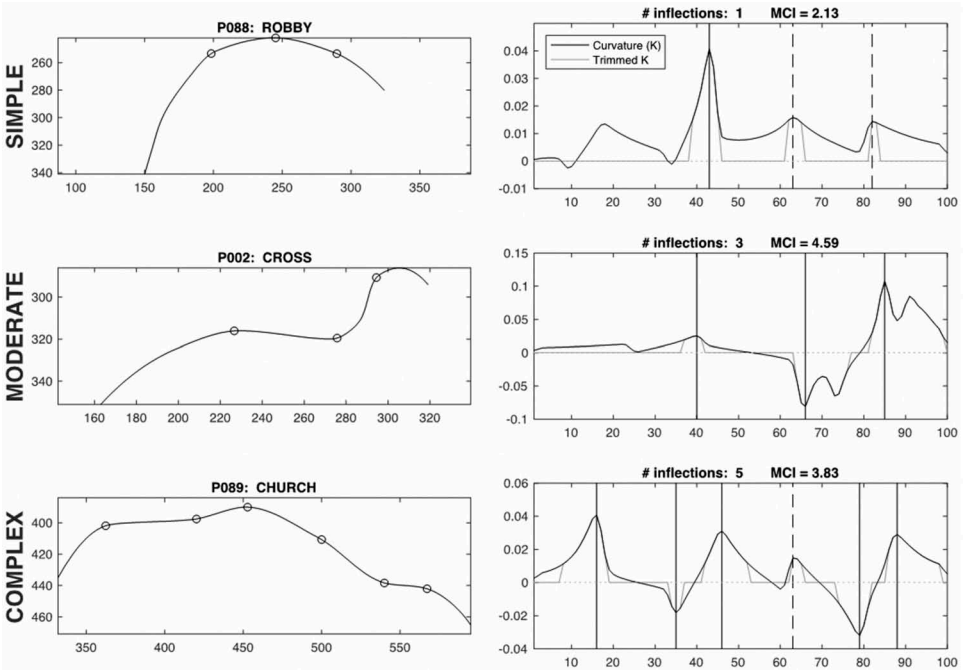

Figure 1. Examples of undifferentiated (simple), retroflex (moderate) and bunched (complex) tongue shapes for rhotics from the post-training group.

Note: In each row, circles on the tongue shapes (left) correspond to the inflection point candidates (right, vertical lines) in the unfiltered (black) and trimmed (grey) curvature; dashed lines show inflections having the same sign, ignored in computing the Number of INFLections (NINFL). The Modified Curvature Index (MCI) is included for comparison.

the analysis (n = 4 instances). Figure 1 provides examples of tongue shapes that correspond to NINFL values of 1, 3 and 5.

NINFL is related to a similar reference-free metric, the Modified Curvature Index (MCI) described in Dawson, Tiede, and Whalen (2016), which computes the integral of unsigned filtered curvature, and is included in Figure 1 for comparison. However, the NINFL metric outperformed MCI in distinguishing between our targeted groups, possibly because it distinguishes the most acoustically relevant inflections in tongue curvature from overall complexity, and accordingly NINFL was used as the dependent measure in these analyses. To test for reliability, 15% of all frames were relabeled by a second rater. Comparisons made on comparable frame contour NINFL counts resulted in values for Fleiss' kappa = 0.53 and Spearman's rho = 0.54, reflecting moderate agreement.

Perceptual judgements

Four trained listeners (one undergraduate research assistant, two graduate research assistants and one licensed speech–language pathologist) independently scored each production for /ɹ/ accuracy. Listeners were native speakers of General American English and minimally had training in articulatory phonetics and speech sound disorders. Each token was scored with a binary rating, 0 = inaccurate rhotic (off target) or 1 = perceptually

correct production of the rhotic, resulting in a mean score of 0 (all raters agreeing the token was incorrect) to 1.00 (all raters agreeing the token was correct). Raters were blind to other raters' scores, and were also blind to the NINFL values. They used only the audio recording to make judgements and did not view ultrasound images while making judgements. A total of 4680 tokens were judged by the four listeners.

Analysis

A generalized linear mixed effects model (Analysis 1) was used to test the hypotheses that tongue shape complexity (characterized by NINFL) would vary by Group (Mid-treatment RSE vs. TS) and by Cluster (no/yes). Data were fit by maximum likelihood (Laplace Approximation) using the package *lme4* in R (Bates, Maechler, Bolker, & Walker, 2015). The dependent variable NINFL follows a Poisson distribution and was modeled using *glmer* with Group and Cluster fit as fixed effects, and random intercepts by participant.[1] Model comparison using log-likelihood tests justified the inclusion of the interaction term but not random slopes by participant.

A related second model (Analysis 2) was used to test whether NINFL was responsive to treatment, with Session (RSE Mid- vs. Post-treatment RSE) and Cluster (no/yes) as fixed effects, and random intercepts by participant.[2] Neither the interaction nor random slopes by participant were supported for inclusion by model comparison. Because the Mid- and Post-treatment participants are paired, we also conducted a one-sided paired *t*-test of NINFL between matched Mid- and Post-treatment groups (averaged across unequal numbers of repetitions by Word and Group).

A separate linear mixed effects model (Analysis 3) was used to predict the dependent variable of averaged listener ratings with fixed effects of Group (Mid-treatment RSE vs. TS) and by Cluster (no/yes), and random intercepts and slopes for cluster by participant (the interaction was not justified by model comparison).[3] In this case, the ratings followed an approximately Gaussian distribution and were modeled using *lmer* from the *lme4* package. *p*-Value estimates were based on Satterthwaite approximations for *F*-test denominator degrees of freedom as implemented in the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017).

A final model (Analysis 4) was used to predict averaged listener ratings with NINFL as a covariate and Cluster (no/yes) as a fixed effect, with random intercepts and slopes for cluster by participant.[4]

Results

Descriptive data

NINFL

The proportion of tokens with NINFL values (ranging from 1 to 5) according to group is shown in Figure 2. The general trend is for the TS group to have more tokens with higher

[1]NINFL ~ GROUP * CLUSTER + (1|ID).
[2]NINFL ~ GROUP + CLUSTER + (1|ID).
[3]RATING ~ GROUP + CLUSTER + (CLUSTER|ID).
[4]RATING ~ NINFL + CLUSTER + (CLUSTER|ID).

NINFL values than the RSE group (both at Mid-treatment and Post-treatment). Additionally, as shown in Figures 2 and 3, the RSE group showed an increase in NINFL values from Mid- to Post-treatment.

Results of mixed effects model comparing groups

Analysis 1 considered the group difference in tongue shape complexity, comparing NINFL values for the TS group with the first session from the RSE group (RSE-Mid) and their interaction with production within a cluster. Results of the model are presented in Table 1. They indicate a significant difference in NINFL values between the two groups ($z = 2.98$, $p < 0.003$) with the RSE group having fewer inflections in the tongue contour than the TS group. There was no main effect of Cluster, but there was a significant interaction between group and cluster, with systematically fewer inflections in clusters than in non-clusters in the TS group ($z = -2.88$, $p < 0.004$). These differences can be seen in Figure 3. Compared to the TS group, significantly lower NINFL values (corresponding to less complex tongue shapes) were observed in the RSE group in the rhotics in the non-clusters; that is, the RSE group did not show the same tendency as the TS group of greater tongue shape

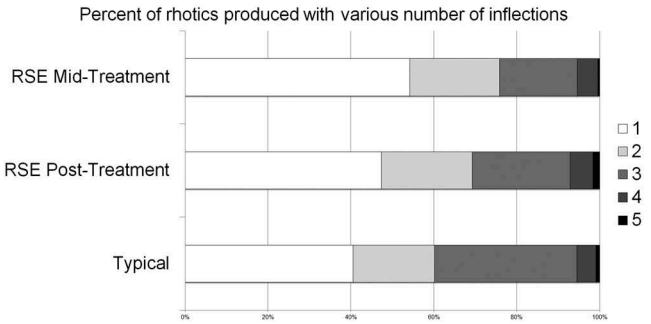

Figure 2. Per cent of rhotics produced with Number of INFLections (NINFL) by group.

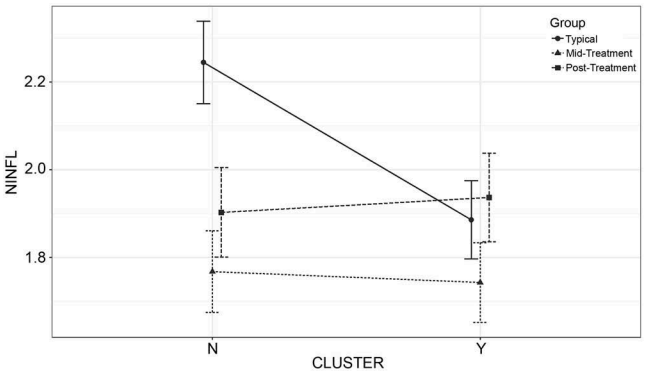

Figure 3. Mean Number of INFLections (NINFL) by cluster context and group.
Note: Error bars show standard error. TYP = typical speech group, RSE = residual speech error group.

complexity in the non-cluster words *Robby* and *Church*. A parallel analysis using the alternative tongue complexity measure proposed by Dawson et al. (2016), the MCI, as the dependent variable showed no significant group separation and no significant interaction, but a main effect also indicating less complexity for productions in cluster environments ($z = -3.32$, $p < 0.003$).

To determine whether NINFL values differed in the RSE group between the Mid- and Post-treatment sessions, a similar model testing fixed effects of Session and Cluster was applied to the data subset contrasting them (Analysis 2). The results indicate that there was a significant increase in NINFL values between the Mid- and Post-treatment sessions ($z = 3.08$, $p < 0.003$). There was no significant effect of Cluster ($z = 0.462$) and the interaction was not justified. The results are provided in Table 2 and visualized in Figure 3, in which NINFL values can be seen to be higher Post-treatment than Mid-treatment for the RSE group.

A one-sided paired *t*-test was also used to compare NINFL values for matched Mid- and Post-treatment conditions, averaged across unequal numbers of repetitions by Word and Participant. The results again confirmed a significant increase in NINFL values from Mid- to Post-treatment ($t = 3.55$, $p < 0.001$, df = 55).

Perceptual ratings

Of the 4680 words rated, the four listeners were in unanimous agreement on 3521 (75% of tokens), rating 1200 tokens as incorrect and 2321 as correct. Of the tokens on which there was not unanimous agreement, 454 tokens were rated as incorrect by three listeners and correct by one listener, 324 tokens were split (two listeners scoring the token correct and two scoring incorrect) and 381 tokens were scored as correct by three listeners and incorrect by one listener. The mean rating of the four listeners for each

Table 1. Generalized linear mixed effects model testing the fixed effects of Group (Mid-training residual speech error (RSE) vs. typical speech (TS)), Cluster production context (No vs. Yes), and their interaction on the Number of INFLections (NINFL), with random intercepts by participant. Random slopes by participant were not justified by model comparison. Baseline (Intercept) is RSE:No.

	Analysis 1			
	Estimate	Std. error	*z*-Value	Pr(>|*z*|)
Intercept	0.531	0.060	8.873	< 0.001
Group: TS	0.256	0.086	2.978	< 0.003
Cluster: Yes	−0.012	0.041	−0.281	0.779
Group: TS * Cluster: Yes	−0.158	0.055	−2.879	< 0.004

Table 2. Generalized linear mixed effects model testing the fixed effects of training Session (RSE-Mid vs. RSE-Post), and Cluster production context (No vs. Yes) on the Number of INFLections (NINFL), with random intercepts by participant. The interaction and random slopes by participant were not justified by model comparison. Baseline (Intercept) is RSE-Mid:No.

	Analysis 2			
	Estimate	Std. error	*z*-Value	Pr(>|*z*|)
Intercept	0.562	0.059	9.540	<0.001
Session: RSE-Post	0.093	0.030	3.079	<0.003
Cluster: Yes	0.014	0.030	0.462	0.644

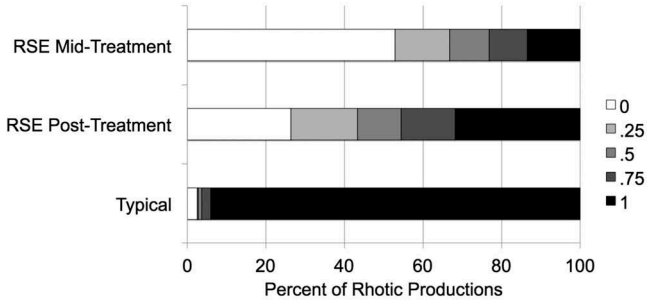

Figure 4. Distribution of perceptual ratings of rhotic accuracy among children with typical speech and children residual speech errors at midpoint of treatment and Post-treatment.

Table 3. Linear mixed effects model testing the fixed effects of session (RSE-Mid vs. RSE-Post), and Cluster production context (No vs. Yes) on listeners' perceptual ratings of rhotic accuracy in children with residual speech errors, with random slopes and intercepts by participant. The interaction was not justified by model comparison. Baseline (Intercept) is RSE-Mid:No.

	Analysis 3				
	Estimate	Std. error	df	*t*-Value	Pr(>|*t*|)
(Intercept)	0.259	0.063	13.3	4.11	<0.002
Session: RSE-Post	0.179	0.012	2312	14.82	<0.001
Cluster: Yes	0.164	0.056	12.7	2.91	<0.02

token (range 0–1.0) was used as the final value in subsequent analyses. The distribution of ratings by group is shown in Figure 4.

A linear mixed effects model was used to test whether perceptual ratings differed between the two sessions for the RSE group (Analysis 3). Session (Mid vs. Post) and Cluster (no/yes) were included as fixed effects, with random slopes and intercepts by Participant. Inclusion of the interaction was not justified by model comparison. Results for this model are given in Table 3. A highly significant main effect of Session shows that listeners rated Post-treatment productions as being significantly more accurate rhotics than Mid-Treatment (t = 14.82, $p < 0.001$), and this tendency is also illustrated in Figure 5. In addition, results show that /ɹ/ in clusters was rated more highly than in non-clusters (t = 2.91, $p < 0.02$).

A final model (Analysis 4) tested whether the perceptual ratings of rhotics for the RSE group could be predicted with NINFL as a covariate and Cluster as a fixed effect, with random intercepts and slopes for cluster by participant. The results are shown in Table 4. NINFL was positively correlated with perceptual rating (t = 2.24, $p < 0.03$), supporting the relationship between tongue shape complexity and the auditory percept of rhotic accuracy. Moreover, the perceptual ratings differed systematically by Cluster (t = 3.00, $p < 0.02$), with clusters being judged as more accurate than non-clusters.

Discussion

This study investigated whether tongue shape complexity, as measured by the number of inflections in the curvature of the tongue contour (NINFL), was associated with accuracy

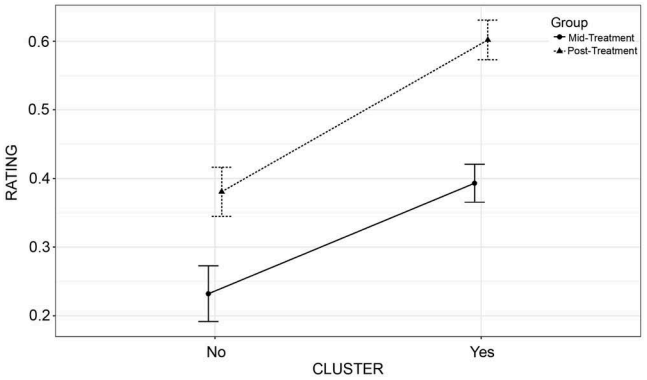

Figure 5. Perceptual ratings of rhotic accuracy by Cluster in the Residual Speech Error group at the midpoint of treatment vs. Post-treatment.

Note: Error bars show standard error.

Table 4. Linear mixed effects model predicting listeners' ratings with NINFL as a covariate and the fixed effect of Cluster production context (No vs. Yes), with random slopes and intercepts by participant. Baseline (Intercept) level of Cluster is No.

	Analysis 4				
	Estimate	Std. error	df	*t*-Value	Pr(>|*t*|)
(Intercept)	0.322	0.063	14.2	5.08	<0.001
NINFL	0.015	0.007	2322	2.24	<0.03
Cluster: Yes	0.166	0.055	12.7	3.00	<0.02

of rhotic productions in children with and without RSE. At the group level, children with RSE had significantly less complex tongue shapes during productions of rhotics than children with TS. Moreover, as the children with RSE progressed through treatment, perceptual judgements of rhotic accuracy improved as tongue shape complexity increased. Finally, listeners' perceptions of rhotic accuracy were significantly and positively correlated with NINFL values, suggesting that complex tongue shapes may contribute to the acoustic consequences of perceived rhoticity.

Overall, the RSE group produced fewer tongue contour inflections than controls. However, it was also the case that the words containing clusters (*cross* and *Street*) had systematically fewer inflections than the non-clusters (*Robby* and *Church*) in the TS group. Phonetic environment, particularly the gesture of the adjacent lingual consonant, likely influenced the complexity of the tongue shape used to achieve the rhotic. That is, the tongue is freer to accommodate a range of tongue shapes for rhotics in onset singletons and nucleus than in clusters containing lingual stops (Curtis & Hardy, 1959; Mielke et al., 2016; Westbury et al., 1998). However, this context-dependent lingual complexity may be difficult for children with RSE to master. As seen in Figure 3, children with RSE, even following treatment, did not show the same pattern of increased complexity in the non-cluster words (*Robby* and in *Church*) as their typically speaking peers. Therefore, continued treatment may be necessary to achieve more 'natural' tongue configurations and more accurate productions in contexts that involve more complex tongue shapes. These context-specific effects should

be interpreted with caution, however, due to the limited set of words and phonetic environments sampled here.

The observed relationship between NINFL and perceptual rating provides some general guidance for the articulatory goals in remediating rhotic distortions. The dissociation of the anterior and posterior portions of the tongue may help to achieve appropriate oral and a pharyngeal constrictions; therefore, to the extent that greater NINFL values reflect a desirable articulatory feature of rhoticity, it may be helpful to convey to children with RSE (in an age-appropriate manner through verbal descriptions or images) that they might aim to achieve a tongue shape with one or more 'turns', 'twists' or 'curves' along the sagittal dimension. That is, particularly when visual feedback is made available during treatment, strategies and cues to achieve multiple 'bends' within the tongue may help to convey the concept of complex tongue curvature.

Caveats and limitations

In this study, we avoided gestalt descriptions of 'bunched' and 'retroflex', in part due to the fact that perceptually accurate rhotics may be produced with a variety of tongue shapes (some of which are neither classically bunched nor retroflexed). Therefore, NINFL may be a useful quantitative supplement to such terminology as it is intended to capture features that may be present in both bunched and retroflexed tongue shapes (cf. the MCI described by Dawson et al., 2016). However, it is important to acknowledge that the associations between tongue shape complexity and listeners' perceptions of rhotic accuracy are related, but there is not a one-to-one correspondence between NINFL and listeners' ratings. Indeed, a number of 'accurate' rhotics were achieved with NINFL values of 1. This may be due to a variety of factors, including phonetic environment (with less complex tongue shapes featured in lingual clusters), individual preference in tongue shapes used to achieve rhoticity (Mielke et al., 2016), parasagittal complexity not captured by midsagittal ultrasound, lip configuration or a lack of head stabilization which limits the precision of the midsagittal images collected. Moreover, it is also feasible that some speakers achieve a complex tongue shape with multiple inflections (i.e. higher NINFL values), yet produce a distorted rhotic quality due to improper positioning of the tongue constrictions within the vocal tract (Boyce, 2015).

Additionally, NINFL is a measure applicable to sagittal images of the tongue, yet there are likely other important articulatory requirements for rhoticity which are not captured by this measure, such as elevation of the lateral margins of the tongue and lip positioning. Thus, quantification of sagittal images alone may not fully reflect the necessary articulatory requirements to achieve acceptable rhotic quality.

Summary and conclusions

This study explored the validity of NINFL as a reference-free method to characterize tongue shape. As anticipated, children with TS produced rhotics with greater accuracy and with more complex tongue shapes than children with RSE. Following 7 h of speech therapy for the children with RSE, there was an increase in both the perceived accuracy of rhotics as well as the NINFL values. NINFL therefore captures an element of tongue shape that may be related to phonetic distortions of American English rhotics. The NINFL

measure is also capable of capturing change in tongue shape over time as well as context-specific variation in typical speakers. Useful information about tongue shape can therefore be extracted from ultrasound images even without head restraint or other means of correction.

Acknowledgements

The authors thank Jessica Whittle and Ahmed Rivera-Campos for assistance with data collection and to Emily Wen-Hsin Ku, Rebecca Medwin and Corinne Walker for assistance with data coding.

Declaration of interest

The authors report no conflicts of interest.

Funding

This work was supported by the National Institute on Deafness and Other Communication Disorders [R01DC013668, R03DC012152, R15DC016426].

ORCID

Patricia McCabe http://orcid.org/0000-0002-5182-1007

References

Adler-Bock, M., Bernhardt, B., 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*(2), 128–139. doi:10.1044/1058-0360(2007/017)

Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America*, *101*(2), 1078. doi:10.1121/1.417972

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Boersma, P., & Weeninck, D. (2014). *Praat v 5. 3.82*: www.praat.org.

Boyce, S. E. (2015). The articulatory phonetics of/r/for residual speech errors. *Seminars in Speech and Language*, *36*(4), 257–270. doi:10.1055/s-0035-1562909

Bressmann, 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. doi:10.3109/02699206.2016.1140227

Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson, E. (2010). Spatial and temporal properties of gestures in North American English/r/. *Language and Speech*, *53*, 49–69. doi:10.1177/0023830909351209

Cleland, J., Mccron, C., & Scobbie, J. M. (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics & Phonetics*, *27*(4), 299–311. doi:10.3109/02699206.2012.759626

Curtis, J. F., & Hardy, J. C. (1959). A phonetic study of misarticulation of/r/. *Journal of Speech & Hearing Research*, *2*(3), 244–257. doi:10.1044/jshr.0203.244

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, *120*(1), 407–415.

Dawson, K. M., Tiede, M. K., & Whalen, D. H. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*, *30*(3–5), 328–344. doi:10.3109/02699206.2015.1099164

Delattre, P., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, *6*(44), 29–68. doi:10.1515/ling.1968.6.44.29

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English/r/. *The Journal of the Acoustical Society of America*, *108*(1), 343–356. doi:10.1121/1.429469

Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2008). *A motor differentiation model for liquid substitutions in children's speech*. Paper presented at the Proceedings of Meetings on Acoustics.

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, *34*(1), 49–72. doi:10.1016/j.wocn.2005.03.005

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe test of articulation - second ed*. Circle Pines, MN: AGS.

Hoffman, P. R., Schuckers, G. H., & Ratusnik, D. L. (1977). Contextual-coarticulatory inconsistency of/r/misarticulation. *Journal of Speech & Hearing Research*, *20*, 631–643. doi:10.1044/jshr.2004.631

Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's/r/productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology*, *22*(3), 540–553. doi:10.1044/1058-0360(2013/12-0137)

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model. *Folia Phoniatrica Et Logopaedica*, *64*(2), 64–72. doi:10.1159/000331997

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex/ɹ/. *Language*, *92*(1), 101–140. doi:10.1353/lan.2016.0019

Modha, G., Bernhardt, B., Church, R., & Bacsfalvi, P. (2008). Ultrasound in treatment of/r/: A case study. *International Journal of Language & Communication Disorders*, *43*(3), 323–329. doi:10.1080/13682820701449943

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 & Communication Disorders*, *52*(1), 80–94. doi:10.1111/1460-6984.12259

Preston, J. L., Maas, E., Whittle, J., Leece, M. C., & McCabe, P. (2016). Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, *30*(3–5), 363–381. doi:10.3109/02699206.2015.1052563

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., … Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *Journal of Visualized Experiments*, *2017*(119), e55123. doi:10.3791/55123

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*(6), 2102–2115. doi:10.1044/2014_JSLHR-S-14-0031

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech & Hearing Disorders*, *37*(1), 55–63. doi:10.1044/jshd.3701.55

Shriberg, L. D. (2009). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul & P. Flipsen (Eds.), *Speech sound disorders in children*. San Diego, CA: Plural Publishing.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–797.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6), 455–501. doi:10.1080/02699200500113558

Tiede, M. (2015). GetContours, software supporting tongue contour extraction from Ultrasound images. https://github.com/mktiede/GetContours

Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English/r/using MRI and ultrasound. *The Journal of the Acoustical Society of America*, *115*(5), 2633–2634. doi:10.1121/1.4784878

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English/ɹ/. *Speech Communication*, *26*(3), 203–226. doi:10.1016/S0167-6393(98)00058-2

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language & Hearing Research*, *48*(3), 543–553. doi:10.1044/1092-4388(2005/037)

Zharkova, N., Gibbon, F. E., & Hardcastle, W. J. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics*, *29*(4), 249–265. doi:10.3109/02699206.2015.1007528

Zharkova, N., Gibbon, F. E., & Lee, A. (2017). Using ultrasound tongue imaging to identify covert contrasts in children's speech. *Clinical Linguistics & Phonetics*, *31*(1), 21–34. doi:10.1080/02699206.2016.1180713

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English/r/. *Journal of the Acoustical Society of America*, *123*(6), 4466–4481. doi:10.1121/1.2902168